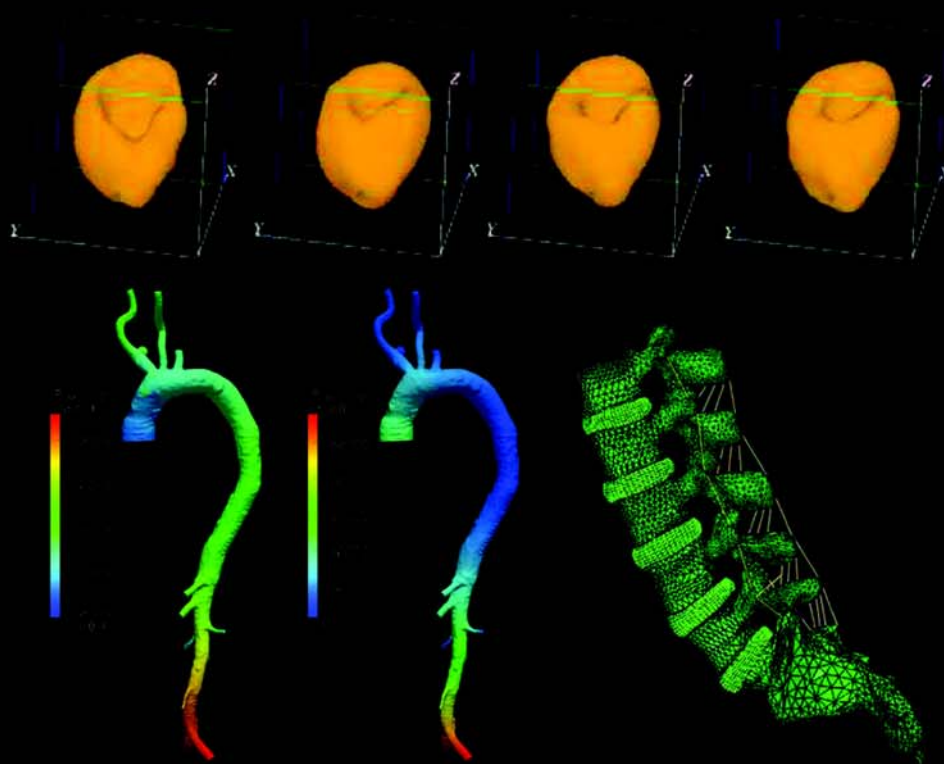


Scientific Computing Applied to Medicine and Healthcare

Current State and Future Trends at the INCT-MACC the Brazilian National Institute of Science and Technology in Medicine Assisted by Scientific Computing

Raúl A. Feijóo, Artur Ziviani and Pablo J. Blanco (Eds.)



Scientific Computing Applied to Medicine and Healthcare

**Current State and Future Trends at the INCT-MACC the Brazilian National Institute of
Science and Technology in Medicine Assisted by Scientific Computing**



Raúl A. Feijóo
Artur Ziviani
Pablo J. Blanco (Eds.)

Scientific Computing Applied to Medicine and Healthcare

**Current State and Future Trends at the INCT-MACC the Brazilian
National Institute of Science and Technology in Medicine
Assisted by Scientific Computing**

Raúl A. Feijóo

HeMoLab - Hemodynamics Modeling
Laboratory
INCT-MACC - National Institute of
Science and Technology in Medicine
Assisted by Scientific Computing
LNCC/MCTI - National Laboratory for
Scientific Computing, Brazil

Artur Ziviani

MARTIN Laboratory
INCT-MACC - National Institute of
Science and Technology in Medicine
Assisted by Scientific Computing
LNCC/MCTI - National Laboratory for
Scientific Computing, Brazil

Pablo J. Blanco

HeMoLab - Hemodynamics Modeling
Laboratory
INCT-MACC - National Institute of
Science and Technology in Medicine
Assisted by Scientific Computing
LNCC/MCTI - National Laboratory for
Scientific Computing, Brazil

Scientific Computing Applied to Medicine and Healthcare

ISBN: 978-85-99961-17-9

© INCT-MACC/LNCC, Brazil, 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broad-casting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the Brazilian Copyright Law in its current version, and permission for use must always be obtained from INCT-MACC. Violations are liable to prosecution under the Brazilian Copyright Law.

The use of general descriptive names, resgistered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover desing: Carla Maduro

Cover image: Carla Maduro from book's images

Typesetting with \LaTeX

Printed in Brazil



National Institute of Science and Technology in Medicine Assisted by Scientific Computing and National
Laboratory for Scientific Computing, Av. Getúlio Vargas 333,
Quitandinha, 25651-075, Petrópolis, RJ, Brazil

Preface

In recent decades we have witnessed the enormous impact of scientific computing in medicine and healthcare. For instance, computational modeling and numerical simulation fields are allowing a better understanding of the behavior and interaction of several physiological systems that play a role in the functioning of the human body both in normal and altered states by diseases or medical procedures, as well as increase the efficiency in the daily medical practices. All this has significantly contributed in the diagnosis, treatment and surgical planning of medical procedures as well as in the training and education of health professionals by incorporating techniques based on scientific computing, visualization and virtual reality. Moreover, they have contributed in a decisive way in the construction and integration of geographically distributed databases containing the most diverse patients' medical information, enabling and facilitating decision making as well as in the generation of new knowledge through data mining and related techniques.

Aware of the socio-economic cost and strategic importance for the development and welfare of the population, the Brazilian government initiated a vigorous program to support the development and consolidation of these activities through the creation of the INCT-MACC the National Institute of Science and Technology in Medicine Assisted by Scientific Computing (<http://macc.lncc.br>). Supported by Brazilian institutions such as the MCTI - Ministry of Science, Technology and Innovation, CNPq, FAPERJ and CAPES, the INCT-MACC is characterized by being a cooperative and interdisciplinary network composed of researchers in various areas of human knowledge (engineering, mathematics, computing, medicine and biology among others) from 23 Associate Laboratories placed in 11 states of Brazil. It also has the collaboration of researchers from countries such as Argentina, United States of America, Canada, England, Portugal, Spain, France, Switzerland and Italy.

Headquartered in the National Laboratory for Scientific Computing (LNCC/MCTI), Petrópolis, Rio de Janeiro, the INCT-MACC develops R&D and human resource training activities focused on the following fields:

- Computational modeling and simulation of the human cardiovascular system and bone trauma and its applications to diagnosis, treatment and planning of several medical procedures.
- Advanced medical image processing including visualization and three dimensional reconstruction of structures of medical relevance and its applications in computational modeling and simulation and novel automated computer methods for image-based diagnosis.
- Collaborative environments of virtual and augmented reality for human resource training and surgical planning.
- Information systems in healthcare, with applications in emergency medical treatment and public health surveillance.
- High performance distributed computing cyber-environments for medical applications.

In September 2012 the INCT-MACC completes three years of activities. For this reason we now present this book that brings together 13 chapters presenting the current state of the R&D activities reached by INCT-MACC in the aforementioned areas. Written by researchers from the Associate Laboratories of the INCT-MACC specially invited in recognition to their scientific leadership, these contributions not only show the interaction achieved within the network but also the future trends of development within the INCT-MACC.

We would like to thank the authors for their dedication and enthusiasm with which they received the task of preparing these chapters. Our appreciation extends to all members of INCT-MACC, which somehow have contributed to the growth and consolidation of our institute. At last but not least, we thank the institutions supporting science and technology in Brazil among which we mention the MCTI, CNPq, CAPES and FAPERJ. Without the support and confidence given by these institutions anything reported in these chapters would be possible.

LNCC/MCTI, Petrópolis,
June 2012.

Raúl A. Feijóo
Artur Ziviani
Pablo J. Blanco

Contents

1	Development and Calibration of a Finite Element Model for the Biomechanical Behaviour of the Human Lumbosacral Spine	1
	Borja M. Yllera, Carlos R. M. Roesler, Marcelo Simões, Ari D.O. Moré and Eduardo A. Fancello	
2	Trends in the computational modeling and numerical simulation of the cardiovascular system	29
	Pablo J. Blanco, Sansuke M. Watanabe, Rafael A. B. Queiroz, Paulo R. Trenhago, Luciano G. Fernandes and Raúl A. Feijóo	
3	Implementation issues of large strain formulations of hyperelastic materials for the modeling of arterial wall mechanics	79
	Santiago A. Urquiza, Pablo J. Blanco, Gonzalo D. Ares and Raúl A. Feijóo	
4	Black-box decomposition methods to model blood flow in the cardiovascular system using dimensionally-heterogeneous models . . .	121
	Pablo J. Blanco, Jorge S. Leiva and Gustavo C. Buscaglia	
5	Mathematical and Statistical Models for Medical Image Processing .	165
	Gilson A. Giraldi, Paulo S. S. Rodrigues, Denise Guliato and Antonio Lopes Apolinário Jr.	
6	Computational methods for medical image quantification in SPECT .	197
	Marco A. Gutierrez, Ramon A. Moreno and Marina S. Rebelo	
7	MRI volumetric measurements of Amygdala and Hippocampus: evaluation of automatic image segmentation methods	229
	Flávio L. Seixas, Andrea S. de Souza, Fernanda Tovar-Moll, Aura Conci, and Débora C. Muchaluat-Saade	
8	Adaptive mesh generation for fluid flow simulation in arteries	247

	Alexandre De Lacassa, Antonio Castelo Filho, Mario A. Lizier, Ícaro L. L. da Cunha	
9	Virtual Reality Applied to Medicine	276
	Jauvane C. de Oliveira, Selan R. dos Santos, Bruno M. de Carvalho, Liliane dos S. Machado, Ronei M. Moraes, Rosa M. E. M. da Costa, Vera M. B. Werneck, Márcio S. Pinho, Fátima L. S. Nunes, Hélio Pedrini, Ildeberto Rodello, Marcio E. Delamaro and Romero Tori	
10	Model-driven development of healthcare applications	315
	Regina Maria Maciel Braga, Luciana Tricai Cavalini, Carlos Eduardo Cirilo, Timothy Wayne Cook, Bruno de Souza Pinto Marques Correa, Sérgio Miranda Freire, Antônio Tadeu Azevedo Gomes, Vinícius Moreira Macedo, Anderson Menezes, João Luís Cardoso de Moraes, Antonio Francisco do Prado, Wanderley Lopes de Souza, Iuri Malinoski Teixeira, and Artur Ziviani.	
11	A decision support system for Alzheimer's disease diagnosis	355
	Flávio L. Seixas, Bianca Zadrozny, Jerson Laks, Aura Conci, and Débora C. Muchaluat-Saade	
12	Global Teleradiology services and education: Brazilian projects to improve service and education	373
	Alexandra Maria Vieira Monteiro, Cláudio Márcio Amaral de Oliveira Lima and Alair Augusto Sarmet dos Santos	
13	Distributed Cyberenvironments Supporting Computing Assisted Medical Applications	391
	Bruno Schulze, Giacomo V. McEvoy, Antonio R. Mury, Ramon Costa, Felipe J. Fernandes, Mariza Ferro, Fabio Licht, Matheus Bandini, Henrique Kloh, Daniel Yohoyama, Luis de Bona, Marcos Castilho, Marcos Sunye, Eduardo Todt, Daniel Weingaertner, Jose Neuman de Souza, Rossana Andrade, Helder Borges, Maria Cristina S. Boeres, Vinod Rebello, Fabio Porto	

Chapter 1

Development and Calibration of a Finite Element Model for the Biomechanical Behaviour of the Human Lumbosacral Spine

Borja M. Yllera, Carlos R. M. Roesler, Marcelo Simões, Ari D.O. Moré and
Eduardo A. Fancello

Abstract This chapter discusses the development of FE-model of the lumbosacral spine including the intervertebral discs, vertebrae, and the 8 major spinal ligaments. The model should be able to properly simulate the complex mechanical behavior in healthy situation and in different pathological stages. To achieve this goal, a laborious systematic calibration of geometry and mechanical properties was performed based on experimental data found in literature. The final results was satisfactory and the model is reliable enough to be used for different applications such as implants design or the study of lumbar pathologies.

1.1 Introduction

The spine shows a characteristically non-linear mechanical response when spinal segments are loaded at physiological magnitudes. Ideally, surgical solutions such as arthroplasty or dynamic stabilization devices should consider this behavior in their kinematics and load sharing characteristics. Several finite elements (FE) models found in the literature have been developed to investigate this mechanical behavior. Such models are expected to assist in the design of spinal implants, and therefore, it is of great importance that they properly predict the pattern of non-linear displace-

Borja Martinez Yllera
Carlos Rodrigo de Mello Roesler
Marcelo Simões
Ari Digiácomo Ocampo Moré
Eduardo Alberto Fancello
LEBm - Biomechanical Engineering Laboratory - University Hospital an Associated Laboratory
of INCT-MACC
GRANTE - Design and Analysis Group - Mechanical Engineering Department
Federal University of Santa Catarina, Campus Universitário sn, Florianópolis, SC Brazil, e-mail:
borja@gmail.com; rroesler@hu.ufsc.br; atendimento@cirurgiadacoluna.
com.br; arimore@terra.com.br; fancello@grante.ufsc.br

ment versus load. The aim of this study is to develop a FE-model of the lumbosacral spine, in order to properly simulate its highly non-linear behavior in healthy and different altered stages. The model considers the segments L1 to L5, the sacrum, intervertebral discs and ligaments. The initial geometry of vertebrae and sacrum was obtained from a commercial model and it was modified with CAD tools, according to anthropometric data obtained from the literature [54]. The intervertebral discs and ligaments were designed by the authors. One of the highlights of this FE-model is that its properties were established through a careful calibration process where the non-linear stiffness of the discs and ligaments were adjusted within physiological limits, in order to make the model able to reproduce, as accurately as possible, the experimental *in vitro* results not only for different anatomical reductions [19] but also for the whole sacrolombar segment [38, 24]. This method ensures a correct response of the model when subjected to structural excisions or resections commonly practiced in surgery, and allows its use for further applications such as implant design or the study of lumbar pathologies. The commercial Finite Element Code ABAQUS was chosen to perform the FE modeling and simulations.

1.2 Geometric Model of the Spine

The process aimed to build a model of the intact spine, without any pathology or implants. The spine is composed of hard tissues, vertebrae and sacrum, and soft ones, intervertebral discs, ligaments and muscles.

1.2.1 Geometric Model of Vertebrae and Sacrum

A geometric description of the five lumbar vertebrae (L1 to L5) and sacrum was commercially acquired (Zygo Media Group - American Fork, UT, USA) and modified by considering lordosis angles, orientation of the vertical and horizontal axes, dimension and morphology of the vertebrae and intervertebral spaces obtained from several anthropometric studies ([21, 4, 11, 25, 2]). Geometry modification was applied to each vertebra by defining new dimensions, relative position and adapting them to the shape of the intervertebral spaces. For the geometric characterization of each individual vertebra, it was used the anthropometric studies by [28, 25] and particularly [54], who worked with computerized tomography measures of a large sample of men. All dimensions from the model are within standard deviation of the corresponding measures. The resulting model is displayed in Figure 1.1.

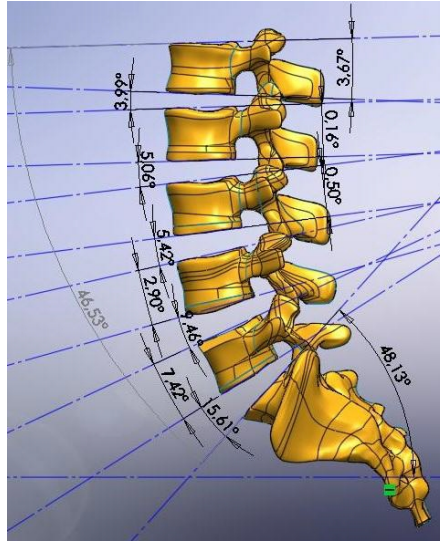


Fig. 1.1 Lateral view of the modified geometric model with wedge angles of each delimited vertebra and disc. It is also shown the lordosis and sacrum angles.

1.2.2 Geometric Model of Intervertebral Discs

The geometric construction of the discs was defined by their anterior, middle and posterior heights, its profile in the transverse plane, lateral curvature and division between the annulus fibrosus and nucleus pulposus. Thus, almost all of its geometry has been defined by the vertebrae positioning, remaining to define the lateral curvature and separation between annulus and nucleus. The cross-sectional area of the nucleus represents 44% of the disc total area, and its center is located approximately at 3.5 mm posterior to the center of the disc.

The annulus fibrosus consists of laminar and concentric bands with helical arrangement, with fibers oriented in the same direction within each band, but in the opposite direction at adjacent bands. The fibers are disposed at an angle of approximately 30° with the horizontal plane of the disc and, therefore, 120° with the fibers from adjacent bands. Internal fibers are bonded to the cartilaginous endplates, while the most peripheric ones are connected directly to the cortical bone. By a general rule, fibers located at the dorsal part of the disc have a more vertical direction (55°) than those in the ventral zone (24°) [40]. The annulus fibrosus was modeled with a two-component structure, a matrix reinforced by a network of fibers designed like non-linear bar elements [40]. This composite arrangement is frequently used and provides good results. Another possible approach, less present in the literature, is a non-linear continuous model of a single anisotropic component that does not ignore the interaction between fibers and matrix [10, 15].

1.2.3 Geometric Model of Ligaments

The lumbosacral ligaments were modeled with non-linear one-dimensional elements connecting adjacent vertebrae. The insertion points were defined according to the morphology from each ligament [19, 27, 31].

1. Anterior Longitudinal Ligament (ALL)- It is a strong band attached to the anterior wall of the vertebral bodies. Each fiber joins two or three adjacent vertebrae. In the model, it was implemented by three non-linear unidimensional elements, one in the midline and one at each side.
2. Posterior Longitudinal Ligament (PLL)- It has the same shape than ALL but it is smaller and attached to the posterior part of the vertebral bodies.
3. Intertransverse Ligaments (ITL)- They connect transverse apophyses and have a very small area. They were implemented with a non-linear unidimensional element between the extremes of the transverse apophyses.
4. Capsular Ligaments (CL)- They connect articular facets edges with fibers perpendicular to the articular surfaces. In the model, they are simulated by four non-linear unidimensional elements attaching these surfaces in a perpendicular way and parallel to the horizontal level.
5. Yellow Ligament (YL)- They are positioned between the lamina of adjacent vertebrae. In the model they are simulated by six non-linear unidimensional elements, three on each side of the lamina.
6. Interspinous Ligaments (ISL)- They connect the adjacent spinous apophyses from its origin in the plate until near the posterior edge. It was represented by three non-linear unidimensional elements that take the same direction observed on *in vitro* studies.
7. Supraspinous Ligament (SSL)- It connects the spinous apophyses tips. It was modeled as a single unidimensional element between the tips of the adjacent apophyses.
8. Iliolumbar Ligaments (ILL)- Its implementation is one of the highlights of the model, and it is addressed separately in Section 1.6.

1.3 Modeling of Vertebrae

A mesh of tetrahedral elements was used to represent the vertebrae, taking advantage of those elements ease in the representation of complex geometries. The vertebrae body was modeled as an orthotropic elastic material with the vertical direction stiffer than the other two, i.e. E_1 and $E_2 = 140 \text{ MPa}$ and $E_3 = 200 \text{ MPa}$ [40, 22, 7]. The distinction between trabecular and cortical bone has been achieved by applying a mesh of two-dimensional elements covering the surface of the vertebrae and assigning to it the thickness and properties of the cortical bone (about 0.35 mm and $E = 12000 \text{ MPa}$) [20, 22, 41]. The endplates were modeled like the cortical bone, with a thickness of 1 mm in the region near the boundary and 0.5 mm in the center

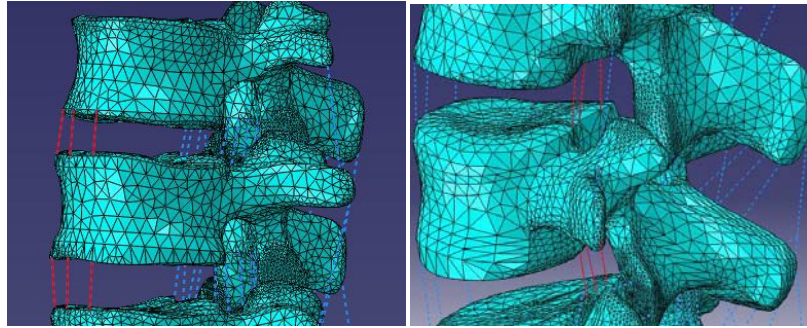


Fig. 1.2 View of the unidimensional elements simulating the anterior longitudinal ligament (left) and the posterior longitudinal ligament (right). Both ligaments are shown in red.

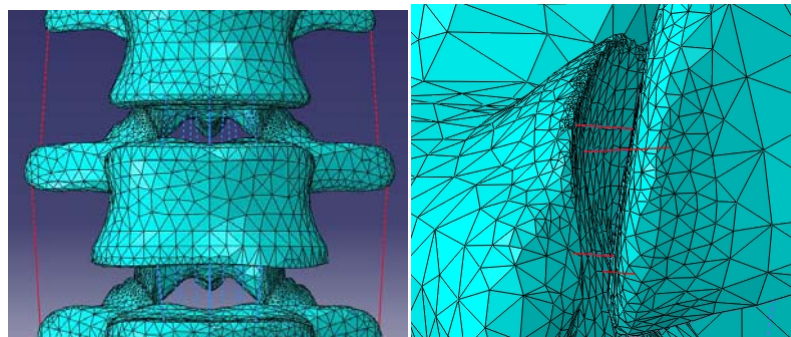


Fig. 1.3 View of the unidimensional elements simulating the intertransverse ligaments (left) and capsular ligament (right).

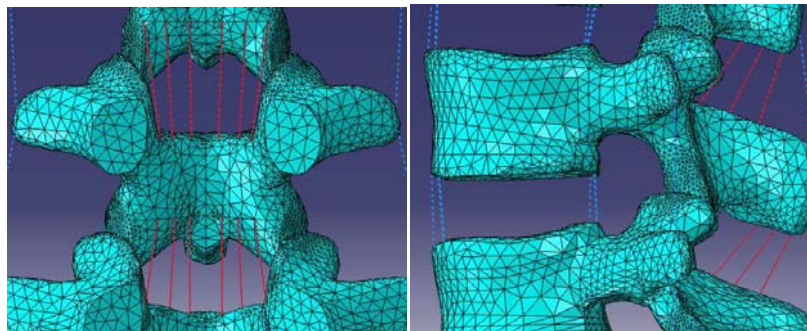


Fig. 1.4 View of the unidimensional elements simulating the flavum ligament (left) and interspinous ligament (right).

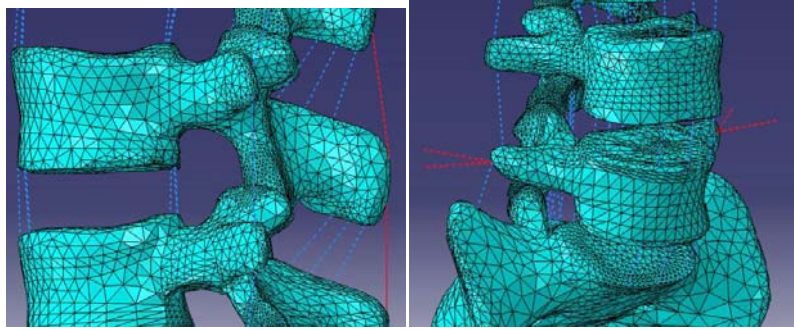


Fig. 1.5 View of the non-linear unidimensional elements simulating the supraspinous ligament (left) and iliolumbar one (right).

[12, 36]. The posterior elements of the vertebrae were modeled with an isotropic material and intermediate stiffness ($E = 3500 \text{ MPa}$), since in that region the percentage volume occupied by the cortical bone is bigger [37, 41]. Figure 1.6 shows the vertebra mesh of L1.

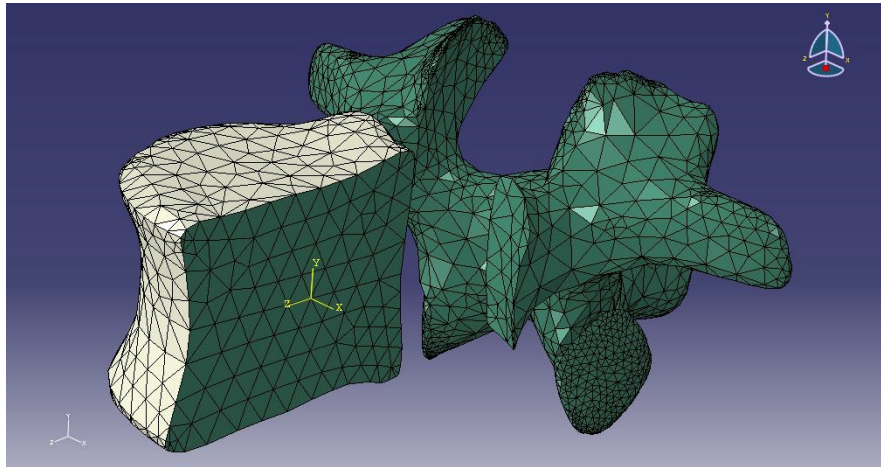


Fig. 1.6 Cut of L1 vertebra with mesh. Green tetrahedral elements simulate the trabecular bone while white triangular elements simulate cortical bone and endplates.

1.4 Modeling and Calibration of Segment L4-L5 Soft Tissues

The difference in stiffness between bones and soft tissues is so large that the vertebrae behave almost as rigid solids, and most of the whole movement is due to deformations of the soft components. Therefore, the calibration of their properties is much more relevant than that of the vertebrae. Unfortunately, the material properties reported in the literature present appreciable differences not only due to the intrinsic variations from person to person, but also due to the different measurement methods used to obtain them. As a consequence, the use of material properties calculated by averaging values obtained in the literature does not ensure the model will behave in a realistic way. A more appropriate procedure in these cases is carrying out a calibration process in order to enforce the model to reproduce behaviors similar to those obtained in several studies in Vitro [40]. Heuer [19] measured the movements of eight segments L4-L5 in flexion, extension, lateral bending and torsion, setting L5 and applying pure moments of 2.5 Nm , 5 Nm , 7.5 Nm and 10 Nm to the L4 superior endplate. Measurements were performed with an intact segment and after removing several structures, one by one (SSL, ISL, YL, CL, facet joints, PLL, ALL, nucleus pulposus) so that, in the last measurement, it remained the vertebrae and annulus fibrosus only. By this experiment it was possible to verify quantitatively the mechanical contribution each structure provides to the movement of the spine.

The model calibration began with the segment L4-L5 considering both vertebrae and the annulus fibrosus only. The other structures of this segment were sequentially added until it reached its original configuration. At each added structure, its material parameters were adjusted to obtain a difference of less than 5%, in comparison to the behavior curves obtained by Heuer in all directions and magnitudes. The structures were incorporated into the model in the sequence presented by the next subsections.

1.4.1 Annulus Fibrosus

In the model, the annulus fibrosus is represented by three-dimensional elements simulating the matrix and a network of one-dimensional elements simulating the fibers. For the matrix, it was used a hyperelastic Mooney-Rivlin model ($c_1 = 0.18$, $c_2 = 0.045$). The fibers were implemented with non-linear stiffness, with magnitudes varying according to their radial and circumferential position in the annulus. Several studies have evaluated fibers orientation and stiffness in different parts of the annulus [20, 9, 55]. It is an agreement that their stiffness is greater in the anterior part than the posterior one, and also in external layers than the internal ones. In order to quantify those differences in the model, the annulus was divided in 8 radial and 5 circumferential sections, creating a total of 40 sub-regions, each one with different material properties, according to their position in the annulus and the proportion between fibers and matrix inside each volume. Using experimental data found in [20, 9, 10, 43], parameterized curves for each one of the 40 subregions of the annulus were defined. The calibration was performed by modifying the para-

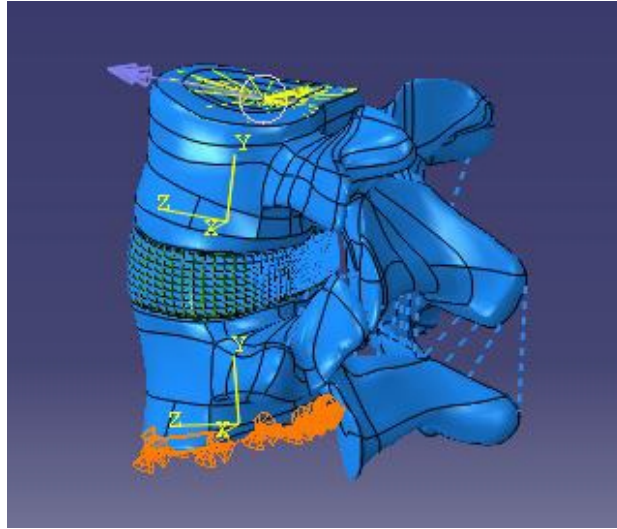


Fig. 1.7 Reduced model of L4-L5 segment with contour conditions, interactions, local coordinate systems of each vertebra and load on the superior endplate

metrical curves and the hyperelastic Mooney-Rivlin material constants, so that the segment could reproduce movements recorded in the experimental study of Heuer ([19]). The final calibration result is shown in Fig. 1.10. The graphs show that using the four load cases ($2.5Nm$, $5Nm$, $7.5Nm$ and $10Nm$) in flexion, extension and torsion, the model was capable of reproducing three of them within the limits of Heuer, while in lateral bending the model fulfilled all of them.

1.4.2 Inclusion and Calibration of the Nucleus Pulposus

The nucleus was modeled as an incompressible Mooney-Rivlin material with coefficients $C01 = 0.12$, $C10 = 0.09$ [40], resulting in curves similar to Heuer's, without needing any calibration.

1.4.3 Inclusion and Calibration of Ligaments

Calibration of the ligaments was performed with the same method used for the annulus, except that each ligament is defined by a single stiffness curve. Behavior curves obtained from many experimental and numerical studies were considered [6, 26, 37, 43, 34]. From all of them, it was chosen those curves whose simulation results were closer to the values measured by Heuer.

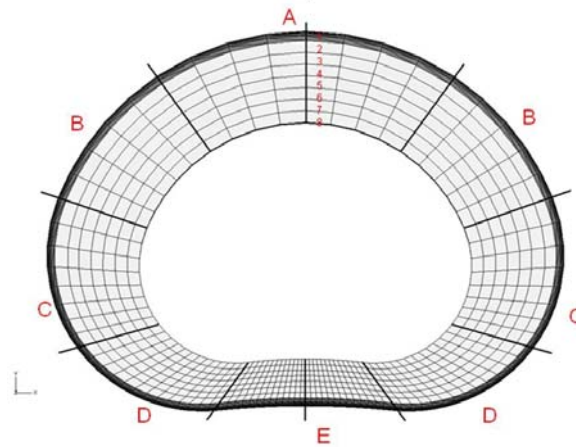


Fig. 1.8 Annulus division in circumferential and radial sections.

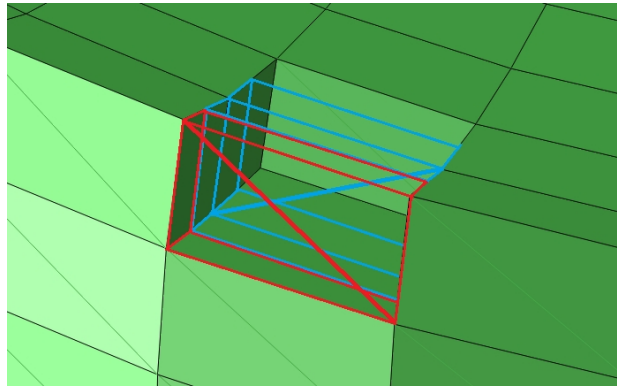


Fig. 1.9 Modeling of annulus fibers by unidimensional elements (thick red and blue lines)

1.4.3.1 Anterior Longitudinal Ligament (ALL)

ALL has the main task of resisting extension, task that is shared with the facet joints and the small resistance to compression offered by some posterior ligaments. Thus, ALL supports heavy loads and is the most rigid ligament in the spine. By including the ALL to the model, extension is reduced by 60%. For the other movements, however, it has little influence (see Fig. 1.11).

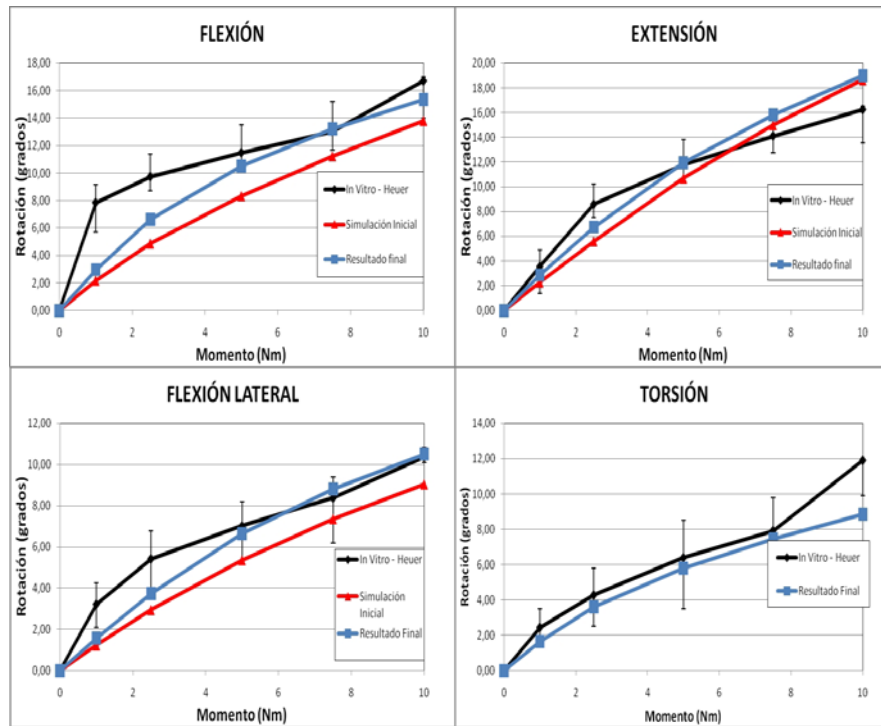


Fig. 1.10 Comparison among final calibration results (blue line), initial ones (red line) and in vitro results of Heuer (black line)

1.4.3.2 Posterior Longitudinal Ligament (PLL)

PLL offers resistance particularly in flexion but, unlike ALL, its function is shared with other ligaments of the posterior part of the spine. Additionally, it lies very close to the instantaneous axis of rotation (IAR) and, therefore, the lever arm of its resistance moment is the smallest from all of the considered ligaments (see Fig. 1.12).

1.4.3.3 Facet Joints (FJ)

Their function is to constraint torsion movements, extension in broader movements and relative displacement in the anterior direction of each vertebra with respect to the lower one. In the model FJ were implemented without the capsular ligaments (CL), inclusion that was considered in the next stage. This resulted in no influence for flexion movements. The torsion movement, however, was considerably limited due to contact of the articular apophyses. In extension, despite the facet joints are in

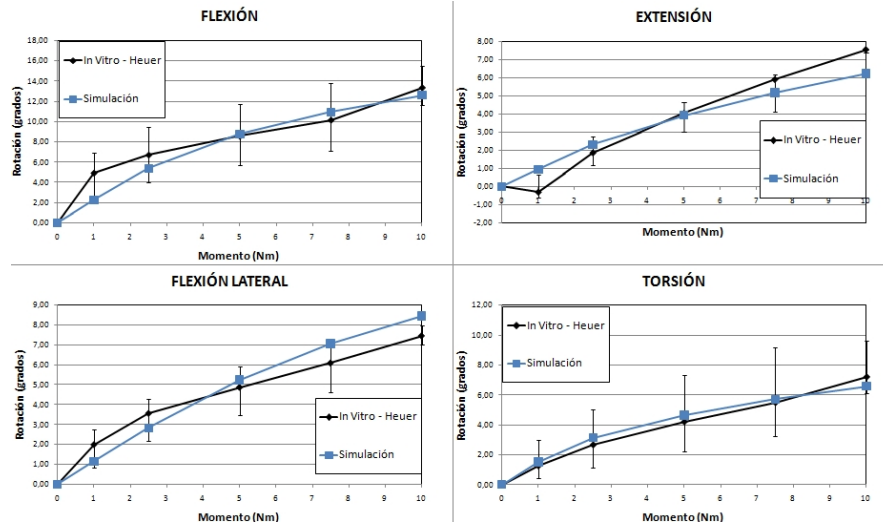


Fig. 1.11 Comparison of the rotation obtained after ALL calibration in flexion, extension, lateral bending and torsion compared to the rotation experimentally measured by Heuer. Error bars indicate Heuer measures range

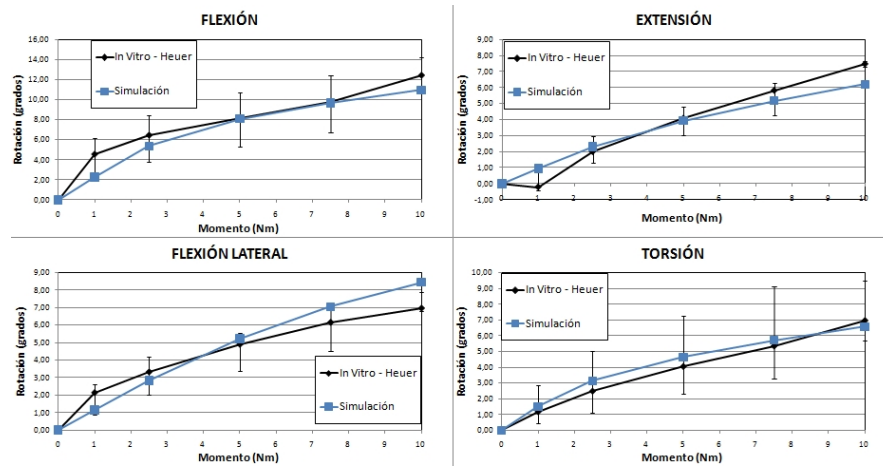


Fig. 1.12 Comparison of the rotation obtained after PLL calibration in flexion, extension, lateral bending and torsion with experimental results of Heuer.

contact, their relative movement is mostly tangential, offering very little resistance to this movement.

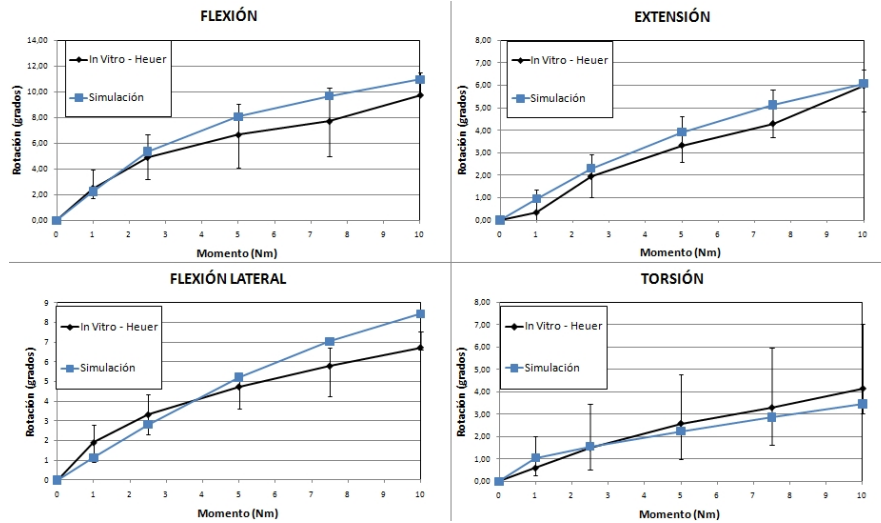


Fig. 1.13 Comparison of rotation obtained after the facet joints calibration in flexion, extension, lateral bending and torsion with experimental results of Heuer

1.4.3.4 Capsular Ligaments (CL)

These ligaments restrict mainly the flexion movement and only slightly the others. The curves obtained with the model are very similar to those presented by Heuer in flexion and extension, with minor differences. The Force-Elongation curve of these ligaments is the second more stiff of the spine. It is only overcome by the ALL, which shows the importance of those ligaments to limit flexion.

1.4.3.5 Yellow Ligament (YL)

Its biomechanical influence is very small, with some effects on anterior and lateral bending. The FE-model was able to reproduce its flexion effect, but not the lateral bending. Even so, it did not represent significant variation in the calibration results.

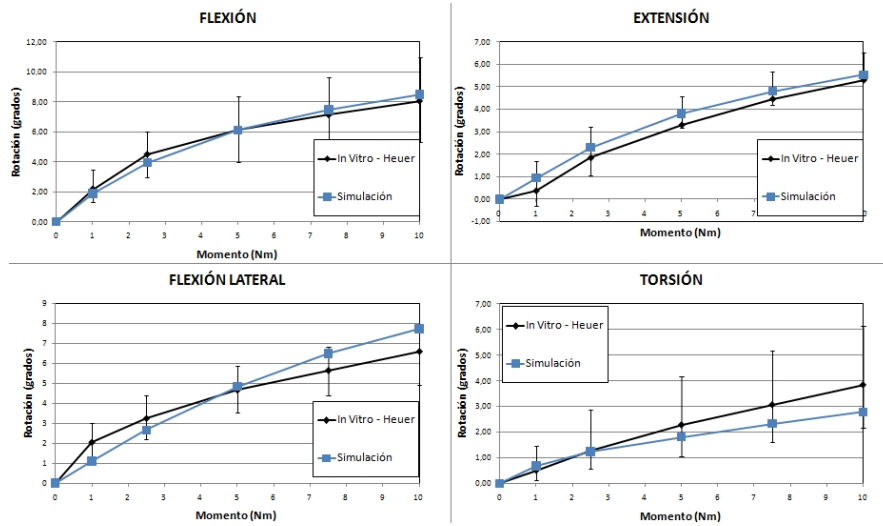


Fig. 1.14 Comparison of the rotation obtained after the CL calibration in flexion, extension, lateral bending and torsion.

1.4.3.6 Interspinous Ligament (ISL)

It joins the spinous apophyses and its lever arm in relation to the IAR is bigger than the lever arm of the ligaments already added to the model. Still, according to Heuer, the movement constrained by it is negligible, even in flexion. Considering its position, the little effect on the spine mechanics means it must have a very low stiffness or it is relaxed within the applied loading conditions.

1.4.3.7 Supraspinous Ligament (SSL)

It joins the extremities of the spinous apophyses, possessing the greatest arm in relation to the IAR. Its influence on the movements, however, is minimal. Like the ISL, it must have a very low stiffness or be relaxed within the applied loading conditions.

1.4.3.8 Intertransverse Ligament (ITL)

Several studies indicate that the ITL at the lumbar region has a low stiffness due to its small cross-sectional area [6, 27] and it is not included in the study of Heuer. The ITL is attached to the transverse apophyses and its line of action is quite different from the IAR. Therefore, it does not need great stiffness to reduce excessive lateral bending and it is the least rigid of all the ligaments.

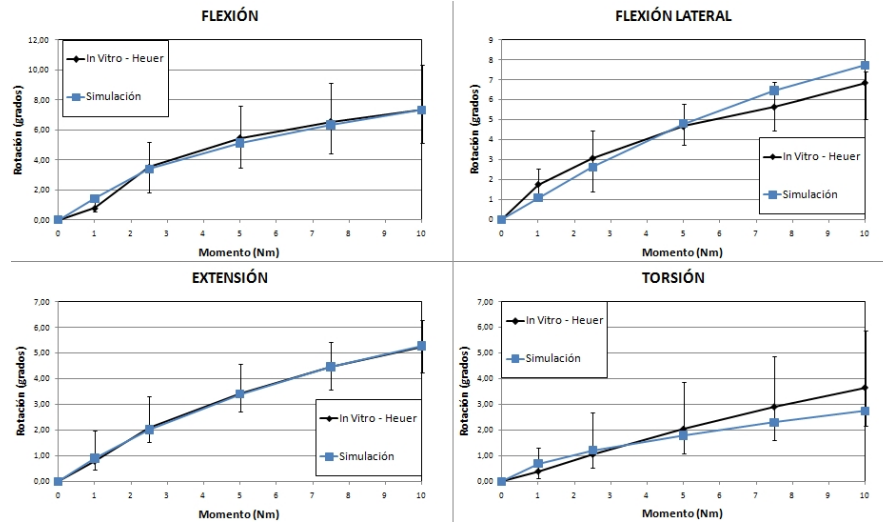


Fig. 1.15 Comparison of the rotation obtained after the ISL calibration on flexion, extension, lateral bending and torsion.

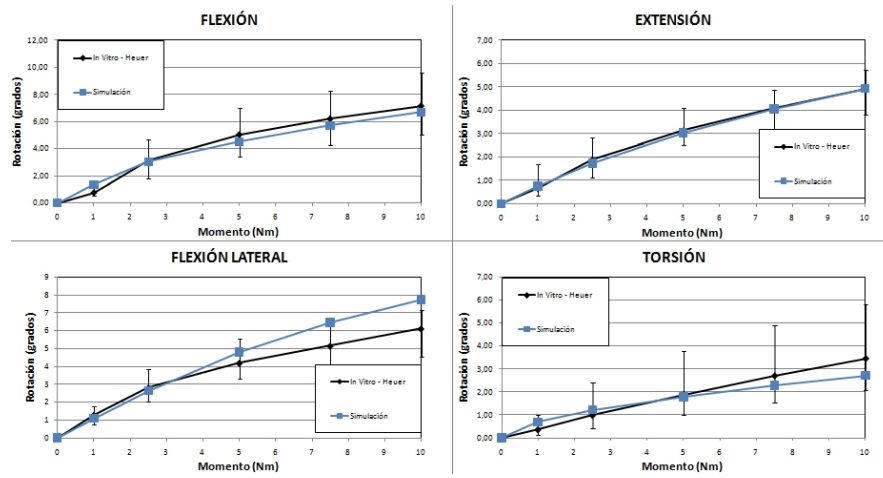


Fig. 1.16 Comparison of the rotation obtained after the SSL calibration on flexion, extension, lateral bending and torsion.

1.5 Modeling and Calibration of the Complete Lumbosacral Segment (L1-S1)

An important limitation of Heuer's study is that focuses L4-L5 segment only, while the material properties vary significantly depending on the segment level. Since no other study as complete as this one was found to estimate the mechanical contribution from each structural component of the spine, it was decided to extrapolate the properties obtained from L4-L5 to the other segments, in a parameterized form. After applying the properties to the model, its behavior was evaluated with several experiments *in Vitro*, reproducing their load cases and boundary conditions and comparing the numerical results with those experimentally obtained. As expected, due to differences in methodology and populations studied by the several authors, results are quantitatively different, but shows similar trends. Details on these comparisons are the subject of the following items.

1.5.1 Boundary Conditions and Load Cases For the Complete Model

The loads acting on the spine *in Vivo* are not precisely known, what made the researchers look for alternative types of loads that could be applied in experiments *in vitro*. Most of them recommend applying a pure moment on the superior vertebra, while the lower vertebra of the multisegmental specimen is fixed, transmitting the same moment to all segments. In the FE-model, this load case was reproduced fixing the sacrum and applying a pure moment to the superior endplate of the L1 vertebra. Another type of load introduced in the *in vitro* tests is the follower load. It consists of a compressive preload passing through the IAR of each vertebra, with the mechanical function of stabilizing the spine for the test with no additional moments. In the experiment is applied by two guided cables, one on each side of the vertebra, attached to a support placed in the first vertebra and stressed by a dead weight. In the FE-model, non-linear unidimensional elements join the IAR from the adjacent vertebrae. Therefore, those nodes receive a compressive load equal to the dead weight used *in vitro*. This model presents contact conditions in the facet joints only. To this aim, a contact model of non-linear penalty type was used. In other words, the cartilage layers of 0.2 mm were not simulated by physical elements but through a nonlinear curve that relates the contact pressure to the penetration between facets. In this curve, the initial resistance corresponds to an elastic modulus of 11 MPa, and at the point where the bones should enter in contact, the resistance is set to 12000 MPa, equal to that of cortical bone [41]. The initial gap between facets was adjusted in such a way the force transmitted by them agree with the studies *in vitro* [49, 24].

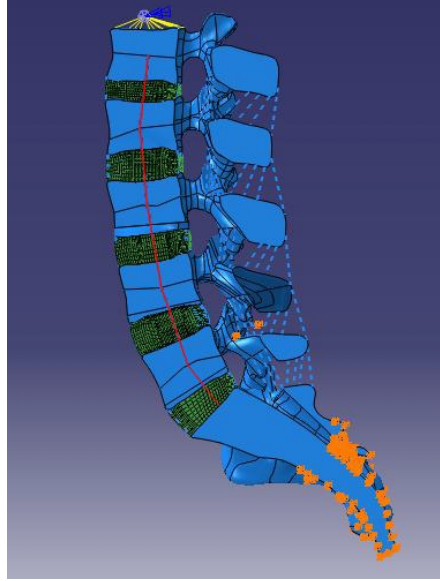


Fig. 1.17 Midsagittal section of the FE model. The pure moment is represented by two blue arrows applied on the master node of the coupling. The lower half of the sacrum is fixed, restriction represented by the orange triangles. The axial element simulating the follower load are in red color.

1.5.2 Comparison of Numerical Results with *in vitro* Results

The *in vitro* studies used for comparison calibration and validation of the FE model were those by Panjabi [29], Yamamoto [51], Guan [14] and Rohlmann [38], summarized in Table 1.1.

Simulations reproducing each one of the experiments mentioned above were carried out, and the corresponding results were compared. Generally, the model behaved in a more rigid way during flexion in nearly all simulated conditions. Behavior in extension, lateral bending and torsion showed a stiffness at some isolated segments with differences close to 50% of the experimental values. This was expected due to the important differences in mechanical properties of soft tissues at different spinal levels. A new parametric calibration based on the values of L4-L5 level was then performed.

	Panjabi 1994	Yamamoto 1989	Guan 2007	Rholman 2001
Specimens (number)	L1-S1 (5) L2-S1 (4)	L1-S1 (6) L1-S1 (4)	L1-S1 (10)	L1-S5 (10)
Load Conditions	10 Nm	10 Nm	4 Nm	7.5 Nm
Compressive Preload	100 Nm	-	-	100 Nm

Table 1.1 Experimental studies used for calibration of the spine model

1.5.3 Calibration of lumbosacral segment

In order to approximate numerical responses to the experimental ones, a calibration of the ligaments properties at different segments was carried out, keeping unmodified the L4-L5 calibration and taking into account Pintar's study [35] to get an idea of stiffness variation among the segments. The contact forces are a good indicator of the facet joints effects to in the spine biomechanics. In certain positions, the extent of those forces on the FE-model helps to know if it behaves close to reality. To complete the setup of FE model properties, the distance between articular surfaces from facet joints was modified until contact forces reach values next to those observed in experimental studies *in vitro* [49, 25]. Even if the amendments made on each ligament are quantitatively different, they all show the same trends. Ligaments rigidity decreases from L1-L2 until L5-S1, and the properties of this last segment are more flexible than the others. After modifying the ligament properties, flexion obtained on segments L1-L2, L2-L3 and L3-L4 adjusted quite well to the study of Guan's. In all loading conditions, movements are within the mean standard deviation and close to the average values. In general, the behavior in extension was excellent, with the initial properties well adjusted to the study of Guan's. Facet joints closes, creating contact forces in all spinal segments with magnitudes very similar to those obtained *in vitro* ([49, 24]). Lateral bending was corrected by modifying the ITL stiffness, approaching the model behavior to the experimental values of Guan's and Panjabi's. It was not necessary to reset the properties of ligaments for torsion, because initial results were already similar to those obtained *in vitro* by Wilson and Niosi ([49, 24]). Facet joints contact forces also produced values of the same order of magnitude of those from *in vitro* experiments. Contact forces on facets and displacements obtained with loads up to 4 Nm, matched the results obtained by Guan's. However, for larger loads the displacements were distant from those obtained by Panjabi with loads of 10 Nm. Rohlmann et.al. [38] held a study to test the influence of the follower load. The comparison between the adjusted model and Rohlmann's experiment demonstrates that the numerical model is able to reproduce the rotations obtained in this study with satisfactory accuracy.

1.6 Iliolumbar Ligament (ILL)

Very few studies involving FE-models had considered the ILL and most of them simplifies too much its morphology and mechanical characteristics[44, 53, 1, 42]. However, its modeling is essential to properly simulate the mechanical behavior of the lumbosacral spine. The ILL originates at the tip of the transverse apophyses from L5 (in some cases from the transverse apophyses of L4, too) and splits into two bands, anterior and posterior, inserted in the superior part of the pelvis (Fig. 1.18). Its mechanical function is to control lateral bending, flexion, extension and torsion in the lower lumbar spine [50, 18]. Some issues related to the ILL biomechanics, like morphological changes related to low back pain or its mechanical function over the

sacroiliac joint, remain unclear. ILL modeling using FE-models may help to clarify such questions, and assist in the design and evaluation of implants to be used in the lumbosacral spine. In [17] the ILL was studied with magnetic resonance, defining its anterior and posterior portions and quantifying the morphology from each part. No studies measuring the ILL mechanical properties were found. In [50] a *in vitro* study was held, applying pure moments of 10 Nm and measuring the displacements of L5-S1 segments in specimens with intact ligaments and after the extirpation of one and both parts, in order to verify the influence of ILL on that segment mechanics. The method used to obtain ILL stiffness in the model was similar to that carried out *in vitro* by [50, 51], i.e., following a reverse process. First, pure moments pursuing Panjabi's conditions with the a compressive preload of 100 N , were applied to the model. Then, four non-linear unidimensional elements were attached to the extremes of the transverse apophyses of L5. The geometry of these elements was taken from the studies of [17, 42]. The element that represents the anterior ligament fascicle has 30 mm length, the posterior has 25 mm and the angle between them has 33° . The element ends has their six degrees of freedom fixed, simulating its attachment to the pelvis. Anterior and posterior ILL stiffness was setup through a calibration procedure similar to that of the other ligaments. The stiffness curve was supposed to be equal to that of the ALL of segment L4-L5 and scaled by a coefficient (one for the anterior and one for the posterior ILL) that were modified in order to match the displacements obtained by [50].

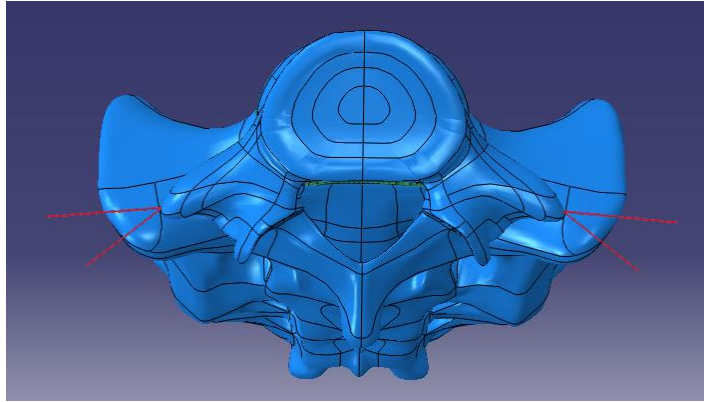


Fig. 1.18 View of L5 and corresponding IL representation (in red).

1.7 Muscular Forces in Vivo

The application of pure moments is a convenient procedure for model calibration since it allows the comparison with *in vitro* experiments. In order to simulate the load conditions to which the spine is subjected in the human body, it is necessary to include in the model the gravitational and muscular forces governing the movement. The neuromuscular control system not only produces movements, but also stabilizes the trunk. Without the muscles, the spine is extremely unstable, even with all ligaments structures intact. Nevertheless, very few FE-models include muscular forces. Even with several studies on the subject, present knowledge is still limited. The muscle forces were estimated experimentally for some basic exercises such as standing, trunk extension and flexion movement, lateral bending, torsion and holding some load [3, 8, 16, 23, 30, 46]. The FE method was also used for its determination as well as to investigate its influence in the spine behavior [37, 52]. Trunk muscles can be divided into global and local (intrinsic). The global muscles are larger; they belong to the musculature of the body that control the movements of the whole spine. The local muscles are smaller and constitute the intrinsic musculature of the spine. They are localized between the spinal segments and control the positioning of the vertebrae.

The model is subjected to the weight force, local muscular forces and general muscular forces represented by the spine erector and rectus abdominis .

The body weight was modeled by a force applied on the center of gravity of the trunk that, according to [5, 45], lies about 200 mm cranial and 30 mm anterior to the center of the disc T12-L1. The force magnitude is 260 N directed downwards, which corresponds to a body weight of 56 kg, the same used in the studies of [37, 37] and [46, 47] and it is attached by a rigid connector to the disc T12-L1, moving with the model. Although the force vector is constant in amplitude and direction, it produces a variable moment depending on the position of the disc T12-L1.

The forces produced by local muscles may change significantly with the position and movement of the torso. Their action is simulated by a compressive load that follows the spinal curvature and, ideally, passes through the IAR of each spinal segment (follower load). That simplification enables more realistic loading conditions for *in vitro* experiments, since it increases the intradiscal pressure to values similar to those measured *in vivo*, without the addition of undesirable moments [38, 33]. In the model, it is represented by unidimensional elements joining the vertebrae centers with a compressive load of 200 N.

Although many muscles are involved in the flexion and extension of the spine, the model uses a simplification in which the flexor muscles are represented by the force of the rectus abdominis and the extensor muscles are represented by the force of the erector spinae. The rectus abdominis acts on a node located 153 mm ventral to the center of the disc T12-L1, and the erector spinae on a node 40 mm dorsal to the center of the disc T12-L1 ([47], Fig. 1.19). The lever arm of those forces do not change with the spine rotation, but its magnitude and direction do.

When a person leans forward in flexion or backwards in extension, the movement combines the rotation not only of the torso, but also of the pelvis. In this work a set

of muscular forces were identified for different body rotations of spine and pelvis: neutral positions, 15, 30 and 50 degrees in flexion, 15 and 25 degrees in extension. The method used for this identification follows the work of [47] and [37], consisting of the following steps:

1. Apply a pure moment on L1 vertebra in order to achieve a fixed rotation angle on it.
2. Enable the weight load, follower load and general musculature.
3. Since the weight load and follower load have defined values for each spine position, identify the global muscular (rectus abdominis or erector spinae) forces needed to eliminate the moment initially applied to L1, keeping the L1 angle unmodified.

This iterative identification process provided the results shown in Table 1.2.

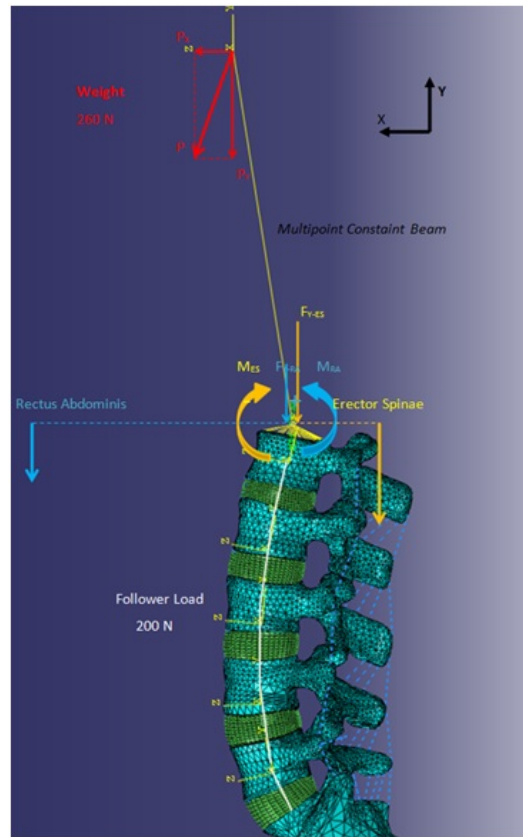
It should be noted that the force exerted by the erector spinae is much larger than the rest of the musculature during flexion and it increases non-linearly with rotation. The force of the rectus abdominis is much smaller than that of the erector spinae, but its lever arm is about 153 mm against 40 mm of the erector. In addition, the trunk center of gravity is also anterior to the spine which results in a flexion moment for the neutral position, reducing the need for loads in the rectus abdominis and increasing that on the erector.

Rotation (degrees)	Weight	Local Muscles	Rectus Abd.	Erectus Sp.
Total / Hip / Spine		(compression)		
50/30/20 Flexion	$F_x = 130N$ $F_y = -225N$	200N		$F_x = -305.9N$ $F_y = -940.5N$
30/20/0 Flexion	$F_x = 89N$ $F_y = -244N$	200N		$F_x = -134.5N$ $F_y = -762.93N$
15/10/5 Flexion	$F_x = 45N$ $F_y = -256N$	200N	$F_x = -1.74N$ $F_y = -19.92N$	$F_x = -48.5N$ $F_y = -555N$
0/0/0 Neutral	$F_x = 0N$ $F_y = -260N$	200N	$F_x = 0N$ $F_y = -20N$	$F_x = 0N$ $F_y = -168N$
15/5/10 Extension	$F_x = -23N$ $F_y = -259N$	200N	$F_x = 13.54N$ $F_y = -76.79N$	
25/10/15 Extension	$F_x = -45N$ $F_y = -256N$	200N	$F_x = 26.04$ $F_y = -147.72N$	

Table 1.2 Weight and musculature forces applied to the model at each load case.

1.8 Analysis of in Vivo Load Case

After calibration and muscular forces application, the relative motion of each segment, the contact forces in the facet joints and forces in ligaments were calculated to each posture condition: neutral position, flexion of 15, 30 and 50 degrees, and extension of 15 and 25 degrees. The results obtained for each case are displayed in Tables 1.3-1.5. The results show that the general trend expected is satisfied: forces and relative rotation within segments increase with global rotation; however, some



cur for flexion. In the model it is possible to see that these conclusions are difficult to generalize due to the complex combination of movements that may counteract among them. For example, although flexion causes separation of facet surfaces, anterior migration of the superior vertebra also present during flexion tends to join them. In extension happens exactly the opposite; the facet closing imposed by the angular momentum is compensated by the translational posterior displacement of the superior vertebra. Compressive loads on the spine reduce the height of the disc and approximate the facets, increasing the joints contact forces in a different way on each level. In L5-S1, the extension increases contact forces, since the segment rotation influences more than the posterior migration of L5 over the sacrum. In the other segments, such a clear trend does not appear. The highest contact forces are found in L4-L5 and L5-S1, probably due to a bigger initial proximity of the facets in L4-L5, and the large rotation in L5-S1. In L1-L2, there is no register of contact, maybe because there are only small rotations in flexion, causing separation in the articular surfaces. Unlike facet joints contact forces, ligaments show clearer trends of behavior. Since the ALL works solely in extension, compression of the spine relaxes this ligament to the point of getting no forces in it, even with some extension. In the PLL, forces are practically constant, since their proximity to the disc makes compression to have more influence than the flexion. Except in 50° of flexion, ISL, SSL and YL ligaments work in the presence of compression forces and that increases with the spine extension. The ILL operates almost exclusively in flexion, withstanding very high loads, which confirms its importance in stabilizing the inferior segments of the spine. In flexion and extension, most of the load is supported by the posterior ILL. The anterior ligament, due to its orientation, may be loaded for lateral bending, not tested in present simulations.

	Rotation (degrees)	Contact Force(N)		Ligament Force (N)			
		Left	right	ALL	PLL	CL(l)	CL(r)
L1L2	4.97	0.00	0.00	0.00	-0.38	0.17	0.82
L2L3	4.62	0.00	0.00	0.00	-3.11	8.41	8.29
L3L4	4.54	0.00	0.00	0.00	-4.81	0.00	0.00
L4L5	1.24	19.84	16.82	0.00	-3.98	0.00	0.00
L5S1	4.62	42.83	43.12	0.00	-1.35	0.00	0.00

	Ligament Forces (N)						
	FL	ISL	SSL	ITL(l)	ITL(r)	ILL(ant)	ILL(pos)
L1L2	18.14	2.38	48.28	7.27	9.14		
L2L3	10.61	2.85	24.26	3.88	3.42		
L3L4	9.77	0.15	13.72	0.33	0.51	19.32	285.79
L4L5	1.84	-3.76	-3.08	0.00	0.00		
L5S1	0.84	0.00	0.00	0.00	0.00		

Table 1.3 Rotations, contact forces and ligament forces for 50 degrees in flexion.

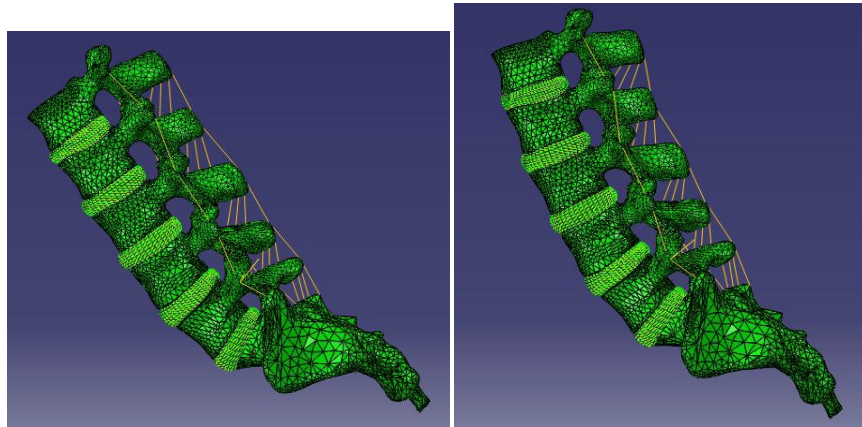
	Rotation (degrees)	Contact Force(N)		Ligament Force (N)			
		Left	right	ALL	PLL	CL(l)	CL(r)
L1L2	1.21	0.00	0.00	0.00	-2.27	0.00	0.00
L2L3	-0.48	0.00	0.00	0.00	-4.92	15.04	13.92
L3L4	-1.37	8.76	9.08	0.00	-7.09	15.47	14.79
L4L5	-1.53	34.79	35.02	0.00	-4.04	0.00	0.21
L5S1	2.16	50.79	51.85	0.00	-0.90	0.00	0.00

	Ligament Forces (N)						
	FL	ISL	SSL	ITL(l)	ITL(r)	ILL(ant)	ILL(pos)
L1L2	10.37	0.04	3.16	0.00	0.00		
L2L3	1.93	-6.09	-11.37	0.00	0.00		
L3L4	-1.11	-10.75	-20.22	0.00	0.00	8.28	79.56
L4L5	-0.66	-11.49	-12.21	0.00	0.00		
L5S1	0.86	-0.39	0.00	0.00	0.00		

Table 1.4 Rotations, contact forces and ligament forces for neutral position.

	Rotation (degrees)	Contact Force(N)		Ligament Force (N)			
		Left	right	ALL	PLL	CL(l)	CL(r)
L1L2	1.16	0.00	0.00	0.00	-0.99	2.55	1.82
L2L3	-2.15	4.39	6.95	4.88	-4.85	7.61	44.86
L3L4	-4.49	12.07	12.89	68.31	-8.65	75.70	75.51
L4L5	-4.32	52.15	60.83	61.78	-5.72	32.46	38.40
L5S1	-5.25	82.95	88.56	14.37	-1.83	35.89	34.45

	Ligament Forces (N)						
	FL	ISL	SSL	ITL(l)	ITL(r)	ILL(ant)	ILL(pos)
L1L2	11.40	0.07	3.83	0.00	0.00		
L2L3	-0.60	-10.04	-21.19	0.00	0.00		
L3L4	-6.19	-17.92	-37.98	0.00	0.00	5.77	0.00
L4L5	-5.36	-20.49	-25.63	0.00	0.00		
L5S1	-1.79	-5.77	-11.98	0.00	0.00		

Table 1.5 Rotations, contact forces and ligament forces for 25 degrees in extension.**Fig. 1.20** lumbosacral spine subject to weight, local and global muscles in 50 and 30 degrees of flexion.

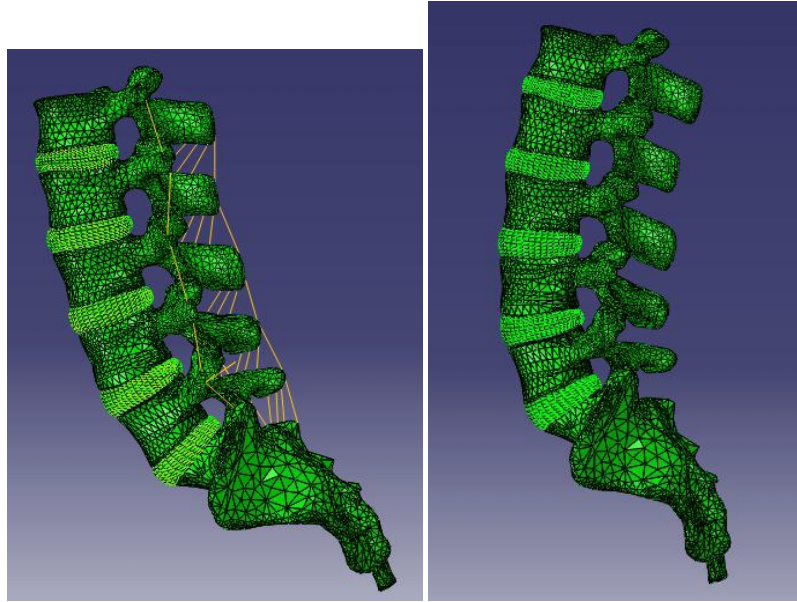


Fig. 1.21 lumbar spine subject to weight, local and global muscles at 15 degrees of flexion and neutral position.

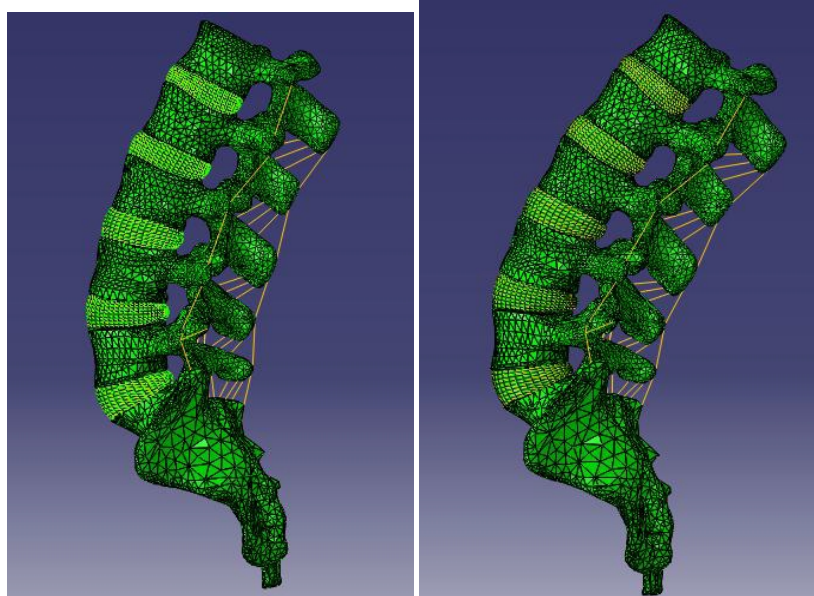


Fig. 1.22 lumbar spine subject to weight, local and global muscles in 15 and 25 degrees of extension.

1.9 Final Remarks

A non-linear model of the lumbosacral spine was developed considering vertebrae, intervertebral discs and the 8 major lumbosacral ligaments. The tissues properties were established by a calibration process in which non-linear stiffness were modified within physiological limits, so that the model is able to reproduce as accurately as possible the results of *in vitro* experiments for different anatomical reductions. The model validation was based on *in vitro* experiments for maximal moments of 4 Nm, being able to be used as a tool in the determination of muscular forces, pathologic states, implants analysis and other situations from biomechanical investigation of the lumbosacral spine. Despite of these encouraging results, the validation process does not allow valuing clear quantitative results, but mainly to use the model in order to establish behavior trends, understand better the biomechanical function of each component from the lumbosacral spine and elaborate comparison of performance among different prosthesis designs, which is the subject of further works.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. The support of these agencies is gratefully acknowledged.

References

- [1] Anderson, A.E., et al., Subject-Specific Finite Element Model of the Pelvis: Development, Validation and Sensitivity Studies. *Journal of Biomechanical Engineering*. **127**(3), 364-373 (2005).
- [2] Been, E., et al., Vertebral Bodies or Discs: Which Contributes More to Human-like Lumbar Lordosis? *Clinical Orthopaedics and Related Research*. **468**(7), 1822-1829 (2010).
- [3] Bogduk, N., J.E. Macintosh, and M.J. Percy, A Universal Model of the Lumbar Back Muscles in the Upright Position. *Spine*. **17**(8), 897-913 (1992).
- [4] Boulay, C., et al., Sagittal alignment of spine and pelvis regulated by pelvic incidence: standard values and prediction of lordosis. *European Spine Journal*. **15**(4), 415-422 (2006).
- [5] Calisse, J., A. Rohlmann, and G. Bergmann, Estimation of trunk muscle forces using the finite element method and in vivo loads measured by telemeterized internal spinal fixation devices. *Journal of Biomechanics*, **32**(7), 727-731 (1999).
- [6] Chazal, J., et al., Biomechanical properties of spinal ligaments and a histological study of the supraspinal ligament in traction. *Journal of Biomechanics*, **18**(3), 167-176 (1985).
- [7] Chen, S.-H., et al., Biomechanical comparison between lumbar disc arthroplasty and fusion. *Medical Engineering & Physics*, **31**(2), 244-253 (2009).
- [8] Dolan, P. and M.A. Adams, The relationship between EMG activity and extensor moment generation in the erector spinae muscles during bending and lifting activities. *Journal of Biomechanics*. **26**(4-5), 513-522.
- [9] Ebara, S., et al., Tensile Properties of Nondegenerate Human Lumbar Anulus Fibrosus. *Spine*, **21**(4), 452-461 (1996).

- [10] Eberlein, R., G.A. Holzapfel, and C.A.J. Schulze-Bauert, An Anisotropic Model For Annulus Tissue and Enhanced Finite Element Analyses of Intact Lumbar Disc Bodies. *Computer Methods in Biomechanics and Biomedical Engineering*, **4**(3), 209-229 (2001).
- [11] Eijkelkamp, M.F., Klain, J.P., Veldhuizen, A.G., van Horn, J.R., Verkerke, G.J., The geometry and shape of the human intervertebral disc. (2007).
- [12] Grant, J.P., T.R. Oxland, and M.F. Dvorak, Mapping the Structural Properties of the Lumbosacral Vertebral Endplates. *Spine*, **26**(8), 889-896 (2001).
- [13] Guan, Y., et al., Moment-rotation responses of the human lumbosacral spinal column. *Journal of Biomechanics*, **40**(9), 1975-1980 (2007).
- [14] Guan, Y., et al., Validation of a clinical finite element model of the human lumbosacral spine. *Medical and Biological Engineering and Computing*, **44**(8), 633-641 (2006).
- [15] Guerin, H.A.L. and D.M. Elliot, Degeneration affects the fiber reorientation of human annulus fibrosus under tensile load. *Journal of Biomechanics*, **39**(8), 1410-1418 (2006).
- [16] Guzik, D.C., et al., A Biomechanical Model of the Lumbar Spine During Upright Isometric Flexion, Extension, and Lateral Bending. *Spine*, **21**(4), 427-433 (1996).
- [17] Hammer, N., et al., Description of the iliolumbar ligament for computer-assisted reconstruction. *Annals of Anatomy - Anatomischer Anzeiger*, **192**(3), 162-167 (2010).
- [18] Hartford, J.M., et al., The Iliolumbar Ligament: Three-dimensional Volume Imaging and Computer Reformatting by Magnetic Resonance: A Technical Note. *Spine*, **25**(9), 1098-1103 (2007).
- [19] Heuer, F., et al., Stepwise reduction of functional spinal structures increase range of motion and change lordosis angle. *Journal of Biomechanics*, **40**(2), 271-280 (2007).
- [20] Holzapfel, G. and M. Stadler, Role of facet curvature for accurate vertebral facet load analysis. *European Spine Journal*, **15**(6), 849-856 (2006).
- [21] Labelle, H., et al., The importance of Spino-Pelvic Balance in L5-S1 Developmental Spondylolisthesis: A review of Pertinent Radiologic Measurements. *Spine*, **30**(6S), S27-S34 (2005).
- [22] Lu, Y.M., W.C. Hutton, and V.M. Gharpuray, Do Bending, Twisting, and Diurnal Fluid Changes in the Disc Affect the Propensity to Prolapse? A Viscoelastic Finite Element Model. *Spine*, **21**(22), 2570-2579 (1996).
- [23] Macintosh, J.E., N. Bogduk, and M.J. Percy, The Effects of Flexion on the Geometry and Actions of the Lumbar Erector Spinae. *Spine*, **18**(7), 884-893 (1993).
- [24] Niosi, C.A., et al., The Effect of Dynamic Posterior Stabilization on Facet Joint Contact Forces: An In Vitro Investigation. *Spine*, **33**(1), 19-26 (2008).
- [25] Nissan, M. and I. Gilad, Dimensions of human lumbar vertebrae in the sagittal plane. *Journal of Biomechanics*, **19**(9), 753-755, 757-758 (1986).
- [26] Nolte LP, P.M., Oxland TR Biomechanical properties of lumbar spinal ligaments. Elsevier, Heidelberg, **9**, 6 (1990).
- [27] Panjabi, A.A.W.a.M.M., *Clinical Biomechanics of the Spine*. Lippincott Williams & Wilkins, (1990).
- [28] Panjabi, M.M., et al., Human Lumbar Vertebrae: Quantitative Three-Dimensional Anatomy. *Spine*, **17**(3), 299-306 (1992).
- [29] Panjabi, M., et al., Mechanical behavior of the human lumbar and lumbosacral spine as shown by three-dimensional load-displacement curves. *J Bone Joint Surg Am*, **76**(3), 413-424 (1994).
- [30] Panjabi, M., et al., Spinal Stability and Intersegmental Muscles Forces: A Biomechanical Model. *Spine*, **14**(2), 194-200 (1989).
- [31] Panjabi M.M., G.G., Duranceau J., Nolte L.P., Three-dimensional quantitative morphology of lumbar spinal ligaments. *Journal of Spinal Disorders*, **4**(1), 9 (1991).
- [32] Panjabi M.M., G.G., Duranceau J., Nolte L.P., Three-dimensional quantitative morphology of lumbar spinal ligaments. *Journal of Spinal Disorders*, **4**(1), 9 (1991).
- [33] Patwardhan, A.G., et al., A Follower Load Increases the Load-Carrying Capacity of the Lumbar Spine in Compression. *Spine*, **24**(10), 1003 (1999).

- [34] Pingel, T.H., Beitrag zur Herleitung und numerischen Realisierung eines mathematischen Modells der menschlichen Wirbelsule. Mitteilungen aus dem Institut für Mechanik, (1991).
- [35] PPintar, F.A., et al., Biomechanical properties of human lumbar spine ligaments. *Journal of Biomechanics*, **25**(11), 1351-1356 (1992).
- [36] Roberts, S., et al., Does the thickness of the vertebral subchondral bone reflect the composition of the intervertebral disc? *European Spine Journal*, **6**(6), 385-389 (1997).
- [37] Rohlmann, A., et al., Determination of trunk muscle forces for flexion and extension by using a validated finite element model of the lumbar spine and measured in vivo data. *Journal of Biomechanics*, **39**(6), 981-989 (2006).
- [38] Rohlmann, A., et al., Influence of a Follower Load on Intradiscal Pressure and Intersegmental Rotation of the Lumbar Spine. *Spine*, **26**(24), E557-E561 (2001).
- [40] Schmidt, H., et al., Application of a calibration method provides more realistic results for a finite element model of a lumbar spinal segment. *Clinical Biomechanics*, **22**(4), 337-384 (2007).
- [40] Schmidt, H., et al., Application of a new calibration method for a three-dimensional finite element model of a human lumbar annulus fibrosus. *Clinical Biomechanics*, **21**(4), 337-384 (2006).
- [41] Sharma, M., N.A. Langrana, and J. Rodriguez, Role of Ligaments and Facets in Lumbar Spinal Stability. *Spine*, **20**(8), 887-900 (1995).
- [42] Shim, V.B., et al., The use of sparse CT datasets for auto-generating accurate FE models of the femur and pelvis. *Journal of Biomechanics*, **40**(1), 26-35 (2007).
- [43] Shirazi-Adl, A., A.M. Ahmed, and S.C. Shrivastava, A finite element study of a lumbar motion segment subjected to pure sagittal plane moments. *Journal of Biomechanics*, **19**(4), 331-350 (1986).
- [44] Sims, J.A. and S.J. Moorman, The role of the iliolumbar ligament in low back pain. *Medical Hypotheses*, **46**(6), 511-515 (1996).
- [45] Stokes, I.A.F. and M. Gardner-Morse, Analysis of the interaction between vertebral lateral deviation and axial rotation in scoliosis. *Journal of Biomechanics*, **24**(8), 753-759 (1991).
- [46] Wilke, H.-J., et al., Is It Possible to Simulate Physiologic Loading Conditions by Applying Pure Moments? A Comparison of In Vivo and In Vitro Load Components in an Internal Fixator. *Spine*, **26**(6), 636-642 (2001).
- [47] Wilke, H.-J. et al., ISSLS Prize Winner: A Novel Approach to Determine Trunk Muscle Forces During Flexion and Extension: A Comparison of Data From an In Vitro Experiment and In Vivo Measurements. *Spine*, **28**(23), 2585-2593 (2003).
- [48] Wilke, H.-J., et al., Stability Increase of the Lumbar Spine With Different Muscles Groups: A Biomechanical In Vitro Study. *Spine*, **20**(2), 192-197 (1995).
- [49] Wilson, D.C., et al., Accuracy and repeatability of a new method for measuring facet loads in the lumbar spine. *Journal of Biomechanics*, **39**(20), 348-353 (2006).
- [50] Yamamoto, I., et al., The role of the Iliolumbar Ligament in the Lumbosacral Junction. *Spine*, **15**(11), 1138-1141 (1990).
- [51] Yamamoto, I., et al., Three-Dimensional Movements of the Whole Lumbar Spine and Lumbosacral Joint. *Spine*, **14**(11), 1256-1260 (1989).
- [52] Zander, T., et al., Estimation of muscle forces in the lumbar spine during upper-body inclination. *Clinical Biomechanics*, **16**(Supplement 1), S73-S80 (1989).
- [53] Zander, T.A. Rohlmann, and G. Bergmann, Influence of different artificial disc kinematics on spine biomechanics. *Clinical Biomechanics*, **24**(2), 135-142 (2009).
- [54] Zhou, S.H., et al., Geometrical dimensions of the lower lumbar vertebrae - analysis of data from digitised CT images. *European Spine Journal*, **9**(3), 242-248 (2000).
- [55] Zhu, D., et al., Micro-structure and mechanical properties of annulus fibrosus of the L4-5 and L5-S1 intervertebral discs. *Clinical Biomechanics*, **23**(Supplement 1), S74-S82 (2008).

Chapter 2

Trends in the computational modeling and numerical simulation of the cardiovascular system

Pablo J. Blanco, Sansuke M. Watanabe, Rafael A. B. Queiroz, Paulo R. Trenhago, Luciano G. Fernandes and Raúl A. Feijóo

Abstract In the present work we present several developments that are characterized by being the current achievements in terms of cardiovascular modeling attained at the Hemodynamics Modeling Laboratory (HeMoLab). Specifically, the following topics are addressed: (i) the description of a closed-loop self-regulated model of the cardiovascular system, (ii) the construction of anatomically highly detailed models of the arterial network and (iii) the automatic construction of peripheral beds. For each of these fields, several examples of application ranging from academic to potentially-clinical are simulated and discussed. The developments presented here, when combined altogether are intended to set up a comprehensive modeling framework to address widest range of cardiovascular scenarios observed in both physiological and pathophysiological conditions due to diseases and or human interventions.

Pablo Javier Blanco

Rafael Alves Bonfim de Queiroz

Paulo Roberto Trenhago

Raúl Antonino Feijóo

HeMoLab - Hemodynamics Modeling Laboratory, INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing and LNCC - National Laboratory for Scientific Computing, Av. Getúlio Vargas 333, 25651-075, Petrópolis, Brazil, e-mail: `\{pjblanco,sansuke,rqueiroz,trenhago,feij\}@lncc.br`

Sansuke Maranhão Watanabe

Rural Federal University of Pernambuco, Academic Campus of Garanhuns, Garanhuns, HeMoLab - Hemodynamics Modeling Laboratory, INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing and LNCC - National Laboratory for Scientific Computing, Av. Getúlio Vargas 333, 25651-075, Petrópolis, Brazil, e-mail: `sansuke@uag.ufrpe.br`

Luciano Gonçalves Fernandes

Institute of Biology, Department of Physiological Sciences, Rural Federal University of Rio de Janeiro and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, 23890-000, Seropédica, Brazil, e-mail: `fernandeslg@gmail.com`

2.1 Introduction

Over the past three decades the advances in computational modeling and numerical simulation techniques, together with the increasing computational power offered by computers, have allowed researchers to study, develop and solve highly accurate mathematical models capable of giving insight about the physical phenomena taking place in complex physiological systems, such as the Human Cardiovascular System (CVS). In this context, not only the mere study of the CVS itself, but also the analysis of the onset and progress of cardiovascular diseases, the outcomes of surgical procedures and the possibility of medical training have been also benefitted with this modeling-based approach.

The CVS can be considered as the integration of different parts at different geometric/time scales and different physiological mechanisms into a *whole* with rather complex structural and functional behavior. These different components, scales and functionalities operate in an integrated manner by virtue of the interdependence of the local and global phenomena. Although some analyses are possible by segregating components in such a way that the local phenomena may be isolated from the rest of the system, there are many other situations where it is impossible to apply this approach due to the fact that local and systemic phenomena are inseparably coupled as a result of their mutual interactions. Consequently, there is little doubt that in the upcoming years modeling efforts must be directed to account for different levels of integration present in the CVS. This will consider the interplay between different geometrical and time scales, and the interaction between physiological and molecular processes with the mechanical and fluid-dynamical aspects of blood flow. A deep understanding of these issues is expected to impact on the ability of computational modeling to assist physicians in facilitating decision-making process. In this way, numerical simulations of the CVS will contribute for a better insight of the onset and development of vascular diseases such as stenotic plaques, perfusion insufficiency, aneurism growth and rupture, elevated arterial pressure, atherosclerosis, among others, as well as, to assist planning medical interventions and more efficient drug delivery strategies.

In the context of the modeling the hemodynamic aspects of the cardiovascular system we identify the following levels of detail (also called levels of integration): (i) the hemodynamics of large arteries, (ii) the local circulation in specific vessels, (iii) the peripheral circulation, (iv) the venous circulation and (v) the cardiac/pulmonary circulation. Such levels of circulation sometimes refer to a certain geometrical scale (for instance the blood flow in large arteries, or in specific districts), and sometimes refer to a given vascular entity (the heart, or a given peripheral bed).

Consequently, several models have been devised that take into account the relevant phenomena at each level of integration and at each geometrical scale. Sometimes, these models are determined by the precision attainable by the medical imaging equipment. As a matter of fact, very acceptable geometric reconstructions of large vessels can be obtained from MRI, whereas their microvascular counterpart is not yet feasible in routine medical practice. In this way, large vessels may be treated as full 3D models whereas microvascular networks must be treated as lumped 0D

models. Other times, the number of geometric scales employed in a model of the CVS depends on the point of view of the user, i.e. it depends on the kind of information that is required to retrieve from different vascular districts. In this sense, the user may select the level of complexity of each component based on the type of information to be obtained, e.g. mean pressure and flow (0D, 1D models) or detailed aspects of the velocity and pressure fields among others (3D models). On the other hand, in certain circumstances the selection and treatment of the type of models used to represent the vascular network are dictated by the physiological information available. Also, it is important to note that full 3D models of the whole vascular tree result exceedingly costly because the calculation requirements are far beyond the computational capabilities of current hardware resources.

In the literature there have been several approaches to integrate different levels of circulation in the sense introduced in the previous paragraphs. Mostly, models based on lumped representations were employed to accomplish this task [21, 41, 45, 63, 70, 82], incorporating 0D models to simulate flow in the larger arteries, veins and cardiac circulation. As well, distributed models for simulating the blood flow in compliant vessels has been an exhaustive area of research through the last decades [1, 39, 55, 64, 29, 84, 33]. More recently, in [24], 1D models of the arterial circulation have been coupled to 0D models of the venous-cardiac-pulmonary circulation to study the influence of arterial stenoses on the wave propagation. Particularly, the 1D model employed in [24] was taken from [29] and is comprised of 55 arterial segments and a 0D lumped representation for the peripheral/venous/pulmonary and cardiac circulations. In turn, valves are modeled using an ideal model of a diode, not allowing for back flow to occur. This last point has been addressed in [21], in which phenomenological models of the cardiac valves are proposed to model more accurately the opening and closing phases of the valves, being capable of modeling certain pathological conditions like valve regurgitation and stenosis. In this last work these valve models are coupled to a purely 0D representation of the closed-loop cardiovascular system.

All above models are based on simple mathematical relations governed by ordinary differential equations or simple partial differential equations. In turn, more detailed models, where the full Navier-Stokes equations have to be solved, have provided another level of understanding in the way haemodynamics factors can be correlated with some relevant phenomena such as arterial wall thickening, solute transport and other physical processes that are better expressed in terms of the local scales.

Although progress has been made in this direction, the computer resources are still insufficient for tackling the problem with entire full 3D models of the cardiovascular system. Moreover, if such computational power was available, the task of handling the huge volume of data needed for setting up those models would still constitute a barrier difficult to overcome. In this sense, it has been recognized that at least for the next years multiscale modeling will be the primary methodology opening up the possibility of simulation studies of the cardiovascular system for investigating, among others, the oxygen transport to the tissues and brain, as well as for assisting on the design of more effective procedures for drug administering

and presurgical computer assisted planning [15]. In the field of modeling blood flow in specific vessels, several works have dealt with the use of heterogeneous representations in order to couple local and global hemodynamics phenomena. This has been mostly carried out using 3D and 1D (or 0D) models to couple blood flow in complex arterial geometries with either full or partial geometrical descriptions for the systemic dynamics [3, 4, 5, 15, 20, 49, 31, 32].

With these scenarios in mind, several questions appeared in situations for which 3D or 1D standalone models may not give accurate answers. The need for studying the interaction between local and global scales, as well as for performing simulations in real physiological regimes and under more realistic conditions led researchers to combine the previous models. This new generation of models derived from coupling simple 1D-0D with full 3D models has led researchers to study new mathematical settings as well as to explore new problems [3, 13, 21, 49, 31].

However, it is clear also the necessity of a new class of models capable to integrate not only all these advances [13, 19] but also capable to allow the interaction between the CVS and systems such as autonomous nervous, respiratory, digestive, endocrine and lymphatic among others [14], which influence in a significant manner the behavior of the CVS under normal or altered conditions due to diseases or human intervention. In this context, we are convinced that models with this level of integration will contribute in the near future to make modeling-based diagnoses, therapy and surgical planning strategies more accurate and appropriate for patient-specific.

At this point it is interesting to remark that all above mentioned previous works (and the references there in) addressed the topology of the arterial network from a simplified point of view, that is by removing smaller arteries and simplifying connectivity among vessels as well as their relation with vascular territories. Nevertheless, there are specific applications which require to take into consideration the true topology of the cardiovascular system, redundancies in the circulation, anastomoses, etc. in order to account for a proper blood flow distribution as well as to be able to model certain physiological and/or pathophysiological scenarios. Surgical procedures like radial harvesting for coronary revascularization and its impact on the blood flow supply to the arteries in the hand [35] and the study of the impact of arterial and cerebrospinal pulsations on spinal cord [47] are clear examples of medical interest in detailed cardiovascular models. Other instances are procedures related to reconstructive surgery, in which knowing the blood supply to the different vascular territories may be of potential clinical relevance in order to perform tissue transplantation successfully [69, 99].

The previous paragraph sets clear examples to motivate the development of closed-loop detailed anatomical models of the cardiovascular system as well as its integration with the other physiological systems. This is the reason why at the HeMoLab group¹ we have developed an unprecedented highly detailed model of the CVS. This model incorporates a total of 1584 arteries described in the medical lit-

¹ The HeMoLab Hemodynamics Modeling Laboratory (<http://hemolab.lncc.br>) is an associated laboratory of the INCT-MACC the Brazilian National Institute of Science and Technology in Medicine Assisted by Scientific Computing (<http://macc.lncc.br>).

erature [15, 50, 53] from which 252 segments correspond to head and neck, 164 to the brain, 64 for each arm and 104 for each leg. The rest is on the torax region. At the terminal districts of this detailed arterial Windkessel models are incorporated in order to simulate the behavior of arterioles and capillaries. Furthermore the inflow and outflow locations of this 1D detailed model are coupled with proper lumped parameter descriptions (0D model) of the remainder part of the circulatory system, closing the network. In turn, the whole peripheral circulation converges to the venous system through the upper and lower parts of the body (using the ideas proposed in [24]). These two main compartments are represented using lumped models for the venules, veins and cavas (inferior and superior). The right and left heart circulation, as well as the pulmonary circulation are also modeled by means of 0D models. Particularly we point out the modeling of the four heart valves, which is carried out by using a non-linear model which allows for the regurgitation phase during the valve closing, following [21]. Finally, the 0D model of the left ventricle is coupled with the inflow point in the 1D model, closing the cardiovascular loop. Furthermore and when necessary, in order to incorporate to the above detailed cardiovascular model the information of the vessel trees associated to each of the human organs we also developed a variant of the Constrained Constructive Optimization (CCO) which allows the construction of such vascular system based on certain optimality criteria [72, 76, 78, 79, 80].

From the above considerations our aim in this work is to present the current developments attained at the HeMoLab Laboratory in the computational modeling and numerical simulation of a comprehensive model of the CVS and its integration with other physiological systems. In particular in Section 2.2 we focus in the presentation of the computational model of the baroreflex mechanism. In Section 2.3 and due to space constraints, we present the development and setting of only a part of our entire detailed cardiovascular model. Particularly, in this section, we focus on the arterial functioning and blood supply to the vascular territories of the arm. In Section 2.4 we present some results obtained with our implementation of the CCO Method with emphasis in the liver vasculature. The final remarks are given in Section 2.5.

2.2 Closed-loop auto-regulated model of the cardiovascular system

Closed-loop representations of the entire cardiovascular system are mandatory when trying to understand the complex interactions between systemic, venous, cardiac and pulmonary circulations and the implications of certain cardiac/valvar pathological conditions over the cardiovascular response and viceversa. Works that have contributed with developments in this direction are [21, 24]. Nevertheless, the integration of closed-loop models with full 3D models have not been addressed in the literature yet. This comprises the next step in order to characterize in a more accurate way the different levels of interactions that lead to the complex behavior of the cardiovascular system. In fact, this kind of 3D-1D-0D closed-loop model allows to study

the influence of global conditions derived from arterial-venous-cardiac-pulmonary interactions on the local blood flow in specific districts. Nevertheless, when some adjustment (or user intervention) in the parameters of the system is needed in order to simulate, say, a pathological condition, none of the models mentioned above is able to give an accurate response in the new equilibrium state of the system, but just to obtain a sort of sensitivity in the solution, that is, a trend in the behavior of the system. This is because these models do not incorporate the control mechanisms that are able to mediate the pressure through a feedback, leading the system to a new homeostatic condition after such intervention.

In this direction, a coupled 0D-3D model including the baroreflex has been presented recently in the literature [33]. In this model, the closed-loop cardiovascular system is represented through 0D compartments. Nonetheless, the number of compartments that model the systemic arteries and the rest of the system is small, and the topological description of the arterial tree is quite poor, as occurs in most of the cardiovascular models employed by authors that develop the models that represent the control mechanisms [59, 60, 91, 92, 93] (and references therein). This is important, among other things, in the proper differentiated vasomotor control activity.

Several neural mechanisms interact to maintain arterial blood pressure and overall cardiovascular function within a narrow range of variation [12]. The neural cardiovascular reflexes control the efferent autonomic activity to blood vessels, heart, kidneys and other organs affecting the cardiovascular function [85, 101]. The baroreceptor represents the main neural mechanism involved in short term cardiovascular regulation. Baroreceptors in the carotid arteries and the aortic arch are activated by beat-to-beat fluctuations in systemic blood pressure. The activation of baroreceptors by a rise in blood pressure leads to an increase of parasympathetic outflow and a decrease of sympathetic firing, leading to a decreased vascular resistance, venous return, heart rate and cardiac contractility [24].

In this section we present a heterogeneous 3D-1D-0D closed-loop model of the cardiovascular system on top of which we develop a 0D model of the homeostatic control mechanism corresponding to the baroreceptor reflex. As examples of application we firstly simulate an acute hemorrhage in the abdominal aorta artery and observe how the hemodynamic quantities are affected by the forcing action of the control. Secondly we simulate a valvular disease, specifically aortic insufficiency, and study the blood flow in a cerebral aneurism analyzing the characteristics of the flow when the baroreflex control is switched on and switched off.

2.2.1 Governing equations

2.2.1.1 Model for the closed-loop cardiovascular network

The topological model of the 1D arterial network considered here follows [1]. The peripheral beds are modeled through Windkessel terminals as in [29, 70]. The entire system is split into two regions, the upper and the lower parts of the body, and these

regions converge to the venules, venous and cavas connective networks as in [24]. The four cardiac chambers and the pulmonary circulation are considered following [24], as well. For the four cardiac valves a model inspired in [21] is used. This integrative model has been described in detail in [8].

In Figure 2.1 we provide a scheme with all the elements present in our model. Observe that any 3D model can be considered in this closed-loop model. The equations modeling each part will be described in the following sections. Each terminal arterial vessel is coupled to a Windkessel terminal to model the peripheral beds. The set of all Windkessel terminals contains 61 elements, and is denoted by \mathcal{W} , and we identify the subset of terminals that correspond to the perfusion of the brain by \mathcal{W}_b , which contains 8 elements. The set $\mathcal{W} \setminus \mathcal{W}_b$, with 53 elements, is the set of peripheral beds that will undergo resistance regulation by the baroreflex. In addition, the set of venous compartments that will undergo compliance regulation is denoted by \mathcal{V} and contains 6 elements (inferior and superior venules, veins and cavas). Finally, the four cardiac chambers will undergo contractility and cardiac period regulation, this set being denoted by \mathcal{C} .

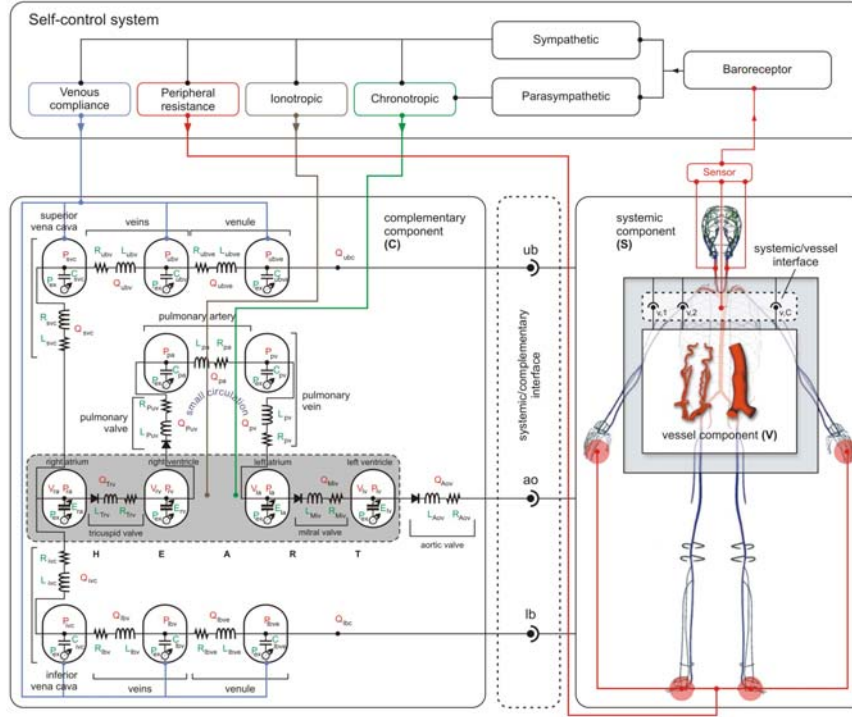


Fig. 2.1 Closed-loop self-regulated model of the cardiovascular system.

The dimensionally-heterogeneous model for describing the closed-loop cardio-vascular model which is used for integrating the model of the baroreflex control mechanism is summarized in Table 2.1. In such table the vascular entities considered in the model together with the corresponding governing equations, suitable references and the nomenclature are given. The corresponding notation is given in Table 2.2.

The vascular entities are grouped in the following sets:

- Arterial circulation: \mathcal{A} consisting of 128 elements
- Peripheral circulation: \mathcal{W} consisting of 61 elements
- Venous circulation: \mathcal{V} consisting of 6 elements
- Pulmonary circulation: \mathcal{P} consisting of 2 elements
- Cardiac chambers: \mathcal{C} consisting of 4 elements
- Cardiac valves: \mathcal{T} consisting of 4 elements
- Specific vessels: \mathcal{S} with as many elements as needed

At junctions among arteries continuity of pressure and flow rate is considered. At interfaces between arterial and complementary circulation continuity of pressure and flow rate is also considered.

2.2.1.2 Model for the baroreflex mechanism

Firstly, we assume that for the physiological and pathophysiological conditions considered in the present work the baroreceptors located in the aortic arch (denoted by aa) as well as those located in both carotid sinus (denoted by lc and rc for the left and right carotids, respectively) behave the same. For simplicity, the activation level of the afferent nervous fiber attached to these three locations by means of the average value of the blood pressure throughout a cardiac cycle, that is

$$f_{a,x} = \frac{1}{T} \int_{-T}^0 P_{sa,x} dt, \quad (2.1)$$

where $f_{a,x}$ is the afferent fiber activation frequency, T is the cardiac period and $P_{sa,x}$ is the pressure at a single point in the 1D segment representing the corresponding arteries (see topology of the arterial tree in [1]), with $x \in \{aa, lc, rc\}$. The fact that we compute the level of activation in the fiber using information in the interval $(-T, 0)$ is due to the fact that for the starting cardiac cycle we compute the evolution of the control quantities using information obtained at the previous heart beat.

Finally, a representative afferent fiber activation frequency, which is sent to the central nervous system, is computed as follows

$$f_a = \frac{1}{3}(f_{a,aa} + f_{a,lc} + f_{a,rc}), \quad (2.2)$$

standing for the contributions of the activations retrieved from the aortic arch and from both carotid sinus.

Table 2.1 Summary of the models for the different vascular entities used in the closed-loop model of the cardiovascular system.

Vascular entity	Governing equations	Set
Arterial circulation [1, 34]	$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left(\beta \frac{Q^2}{A} \right) = -\frac{A}{\rho} \frac{\partial P}{\partial x} - f_r \frac{\rho \pi D \bar{u} \bar{u}}{8\rho}$ $\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0$ $P = P_o + \frac{h_o E_c}{R_o} \left(\sqrt{\frac{A}{A_o}} - 1 \right) + \frac{K h_o}{R_o} \frac{1}{2\sqrt{A A_o}} \frac{dA}{dt}$	\mathcal{I}
Peripheral circulation [29]	$\frac{dQ_i}{dt} = \frac{1}{R_c R_a C_a} [R_a C_a \frac{d}{dt} (P_i - P_o) + (P_i - P_o) - (R_c + R_a) Q_i]$ $Q_i = Q_o$	\mathcal{W}
Venous/Pulmonary circulation [24]	$L \frac{dQ_o}{dt} + R Q_o = P_i - P_o$ $C \frac{d}{dt} (P_i - P_{ex}) = Q_i - Q_o$	$\mathcal{V} \setminus \mathcal{P}$
Cardiac chambers [24]	$P_i - P_{ex} = (E_A e(t) + E_B)(V_{ch} - V_{0, ch}) + \alpha_{ch} P_i \frac{dV_{ch}}{dt}$ $e_v(t) = \begin{cases} \frac{1}{2} [1 - \cos(\pi \frac{t}{T_{vc}})] & 0 \leq t \leq T_{vc} \\ \frac{1}{2} [1 + \cos(\pi \frac{(t-T_{vc})}{T_{vr}})] & T_{vc} < t \leq T_{vc} + T_{vr} \\ 0 & T_{vc} + T_{vr} < t \leq T \end{cases}$ $e_a(t) = \begin{cases} \frac{1}{2} [1 + \cos(\pi \frac{(t+T-t_{ar})}{T_{ar}})] & 0 \leq t \leq t_{ar} + T_{ar} - T \\ 0 & t_{ar} + T_{ar} - T < t \leq t_{ac} \\ \frac{1}{2} [1 - \cos(\pi \frac{(t-t_{ac})}{T_{ac}})] & t_{ac} < t \leq t_{ac} + T_{ac} \\ \frac{1}{2} [1 + \cos(\pi \frac{(t-t_{ar})}{T_{ar}})] & t_{ac} + T_{ac} < t \leq T \end{cases}$ $\frac{dV_{ch}}{dt} = Q_i - Q_o$	\mathcal{C}
Cardiac valves [21]	$L \frac{dQ_o}{dt} + R Q_o + B Q_o Q_o = \Xi (P_i - P_o)$ $\Xi = \frac{(1 - \cos \theta)^4}{(1 - \cos \theta_{max})^4}$ $I \frac{d^2 \theta}{dt^2} + k_F \frac{d\theta}{dt} =$ $k_P (P_i - P_o) + k_Q Q_o \cos \theta + \begin{cases} k_V Q_o \sin(2\theta) & \text{if } P_i \geq P_o \\ 0 & \text{if } P_i < P_o \end{cases}$ $\theta = \begin{cases} \theta_{min} & \text{if } \theta < \theta_{min} \\ \theta_{max} & \text{if } \theta > \theta_{max} \end{cases}$	\mathcal{T}
Specific vessels [31]	$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho (\mathbf{u} \cdot \nabla) \mathbf{u} - \mu \Delta \mathbf{u} + \nabla P = 0$ $\text{div } \mathbf{u} = 0$ $P - P_o = \frac{E_e h_o}{R_o^2} \delta + \frac{K h_o}{R_o^2} \frac{d\delta}{dt}$ $\mathbf{w} = \delta \mathbf{n}$ $\mathbf{v} = \frac{\partial \mathbf{w}}{\partial t}$ $Q_c = \int_{\Gamma_c} \mathbf{u} \cdot \mathbf{n} d\Gamma_c$ $P_c = ((P\mathbf{I} - \mu(\nabla \mathbf{u})^s) \mathbf{n}) _{\Gamma_c}$	\mathcal{I}

The averaged responses mediated by the sympathetic and parasympathic systems are sigmoid functions of the afferent fiber activation frequency, and consequently of the mean arterial pressure [60, 37]. Hence, the sympathetic and parasympathic activities (n_s and n_p , respectively) are described by the following expressions

$$n_s = \frac{1}{1 + \left(\frac{f_a}{\mu} \right)^v} \quad n_p = \frac{1}{1 + \left(\frac{f_a}{\mu} \right)^{-v}}, \quad (2.3)$$

Table 2.2 Notation for governing equations in Table 2.1.

Q	flow rate	P	blood pressure
\bar{u}	mean blood velocity	A	vessel lumen
ρ	blood density	μ	blood viscosity
f_r	Darcy friction factor	β	convective acceleration factor
R	vessel radius	h	arterial wall thickness
E	effective Young modulus	K	viscoelastic modulus
$(\cdot)_o$	reference value for (\cdot)	R_c	capillaries resistance
R_a	arterioles resistance	C_a	arterioles compliance
$(\cdot)_i$	input value of (\cdot) at 0D pipe	$(\cdot)_o$	output value of (\cdot) at 0D pipe
L	pipe inertance	R	pipe resistance
C	pipe compliance	P_{ex}	external reference pressure
V_{ch}	chamber volume	$V_{0,ch}$	reference chamber volume
E_A	maximum elastance	E_B	baseline elastance
e_v	ventricular activation function	e_a	atrial activation function
α_{ch}	chamber viscoelastic coefficient	T	cardiac period
T_{vc}	ventricular contraction period	T_{vr}	ventricular relaxation period
T_{ac}	atrial contraction period	T_{ar}	atrial relaxation period
B	pipe coefficient for convective effects	Ξ	orifice coefficient
θ	opening angle	I	moment of inertia
k_F	friction moment coefficient	k_P	pressure moment coefficient
k_Q	inertia moment coefficient	k_V	vortex moment coefficient
θ_{min}	minimum opening angle	θ_{max}	maximum opening angle
\mathbf{u}	blood velocity	\mathbf{v}	frame of reference velocity
δ	normal component of displacement	\mathbf{w}	wall displacement
\mathbf{n}	wall outward normal	Γ_c	3D-1D coupling interface
$(\cdot)_c$	value of (\cdot) at coupling interface Γ_c		

where μ is the baseline fiber activation frequency imposed by the sensor and ν is a parameter related to the slope of the responses of n_s and n_p at the central point.

The efferent responses are governed by a first order ordinary differential equation as in [60]. The same model is employed for the different organs (actuators). Such equation depends upon a linear combination of the sympatic and parasympatic activities, and reads

$$\frac{d}{dt}x_i = \frac{1}{\tau_i} \left(-x_i + \sigma_i^b(f_a) \right) \quad i \in \mathcal{E}, \quad (2.4)$$

where the index i ranges the set $\mathcal{E} = \{H, E_{A,x}, R_{a,y}, R_{c,y}, C_{v,z}\}$ of organs undergoing regulation. Here $H = \frac{1}{T}$ stands for the frequency of the four cardiac chambers, $E_{A,x}$ is the maximum value of the elastance of the cardiac chamber $x \in \mathcal{C} = \{RA, RV, LA, LV\}$, $R_{a,y}$ and $R_{c,y}$ represent the arterioles and capillaries resistances, respectively, for the set of peripheral beds $y \in \mathcal{W} \setminus \mathcal{W}_b$ which undergo regulation, while $C_{v,z}$ represents the venous compliances of the set of venous compartments $z \in \mathcal{V}$. The characteristic time constant τ_i , $i \in \mathcal{E}$, is specified for each organ accordingly. In addition, we have

$$\sigma_i^b(f_a) = \alpha_i n_s - \beta_i n_p + \gamma_i \quad i \in \mathcal{E}, \quad (2.5)$$

where α_i and β_i are weights for the sympathetic and parasympathic activities in the actuator i , $i \in \mathcal{E}$, and γ_i is the basal activation level of such actuator, whenever the autonomic nervous system is disconnected. These parameters are specified below.

Let us sum up the actuators involved in the cardiovascular model. In each cardiac chamber the cardiac frequency H (chronotropic effect) and the maximum level of contractility E_A (ionotropic effect) are regulated. In the 1D model all the terminal segments are coupled to the corresponding peripheral bed, this set was denoted by \mathcal{W} (standing for set of Windkessel terminals). From this set, the subset that undergoes regulation is the set $\mathcal{W} \setminus \mathcal{W}_b$, consisting of 53 elements, recalling that the vessels perfusing the brain are not innervated. Finally, the systemic venous network is described through 6 compartments, which form the set \mathcal{C} and undergo regulation of the vessel compliance.

2.2.2 Model parameters

As said, the closed-loop model of the cardiovascular system has already been presented in [8]. Therefore, all the involved parameters are in accordance with the data reported in [1, 29, 21, 24].

The parameter calibration process was performed following an open loop approach. Firstly, the parameters μ and ν corresponding to the central nervous system were determined. Secondly, the characteristic times τ_i , $i \in \mathcal{E}$ were adjusted. Thirdly, the parameters α_i , β_i and γ_i , $i \in \mathcal{E}$ were calibrated through physiological data concerning the threshold and saturation levels of each actuator.

For μ the value $\mu = 94.3 \text{ mmHg} = 125733 \text{ dyn/cm}^2$ was considered, being this the mean pressure value at which the system remains at equilibrium. For ν the value considered in [60] was adopted here $\nu = 7$.

For the characteristic times τ_i we adopted the values reported in [87]. These values are given in Table 2.3.

Table 2.3 Characteristic times τ_i , $i \in \mathcal{E}$, for the different actuators.

Actuator ($i \in \mathcal{E}$)	$\tau_i [\text{s}]$
H	4.0
$E_{A,x}$, $x \in \mathcal{C}$	10.0
$R_{a,y}$, $R_{c,y}$, $y \in \mathcal{W} \setminus \mathcal{W}_b$	15.0
$C_{v,z}$, $z \in \mathcal{V}$	30.0

The values of the parameters α_i , β_i and γ_i , $i \in \mathcal{E}$ for the components $H = \frac{1}{T}$, $E_{A,x}$, $x \in \mathcal{C}$, $R_{a,y}$, $R_{c,y}$, $y \in \mathcal{W} \setminus \mathcal{W}_b$ and $C_{v,z}$, $z \in \mathcal{V}$ are reported in normalized form with respect to the corresponding baseline parameters in Table 2.4 (i^b denotes the baseline, or equilibrium, value for the actuator i), and were determined using threshold and saturation values physiologically admissible (see Table 2.5, [60]). As before, index b denotes baseline value.

Table 2.4 Parameters for the efferent pathways normalized with respect to the corresponding base-line value.

Actuator ($i \in \mathcal{E}$)	α_i/i^b	β_i/i^b	γ_i/i^b
H	1.15	0.34	0.59
$E_{A,x}, x \in \mathcal{C}$	0.4	0	0.8
$R_{a,y}, R_{c,y}, y \in \mathcal{W} \setminus \mathcal{W}_b$	0.8	0	0.6
$C_{v,z}, z \in \mathcal{V}$	-0.2	0	1.1

Table 2.5 Threshold and saturation values for actuators.

Normalized parameter	Threshold	Saturation
H^b/H	0.25	1.75
$E_{A,x}^b/E_{A,x}, x \in \mathcal{C}$	0.80	1.20
$R_{a,y}^b/R_{a,y}, R_{c,y}^b/R_{c,y}, y \in \mathcal{W} \setminus \mathcal{W}_b$	0.60	1.40
$C_z^b/C_z, z \in \mathcal{V}$	0.90	1.10

Particularly, the modification of the heart rate H entails changing the cardiac period T , and therefore we must change the different characteristic times that define the elastance activation functions, that is $T_{vc}, T_{vr}, t_{ar}, T_{ar}, t_{ac}$ and T_{ac} (see Table 2.1). This is carried out scaling these quantities in the same proportion as T .

2.2.3 Numerical approximation

The finite element method is used for the spatial discretization of the 1D and 3D problems. For the 1D model the system of equations is rewritten along the characteristics and GLS stabilization terms are incorporated to deal with the dominant convection. For the 3D model the formulation of the mini element is used to deal with the incompressibility constraint, and SUPG stabilization is also used to treat dominant convection. In view of our simple model for the 3D arterial wall, the 3D fluid-structure problem is solved monolithically. The interested reader can find these techniques in [3, 31]. Time discretization (in all cases, that is 3D, 1D and 0D) is performed using an implicit second order finite difference scheme. All the nonlinearities in the problem are treated using Picard iterations. The originally coupled problem is transformed into an interface problem using a domain decomposition approach developed for this kind of heterogeneous modeling [43]. Such interface problem is solved through an iterative Broyden method [6, 23].

The coupling between the hemodynamics problem and the baroreflex control is carried out in an explicit way, that is, once a cardiac cycle is finished the quantities involved in the control are computed (mean pressure, sympathetic and parasympathetic activities) and the parameters that define the behavior of the cardiovascular system (efferent pathways) are modified as a function of the last cardiac beat. Then, the next cardiac cycle begins.

When running a 3D-1D-0D simulation of the self-regulated cardiovascular system we follow a three-step procedure:

1. alter the conditions of the 1D-0D system according to the physiological scenario;
2. run the 1D-0D simulation until the baroreflex find the new equilibrium point in the system (no 3D computations are performed here);
3. embed the 3D model within the 1D-0D system and run three cardiac cycles to have the local hemodynamics corresponding to such new equilibrium point.

In doing this we are assuming that the evolution of the control mechanism for the 3D-1D-0D system coincides with the evolution of the 1D-0D system. That is, the presence of the 3D vessel does not influence in the new equilibrium point determined by the baroreflex (which is reasonable due to the small 3D vessels we will consider here). Such hypothesis should be carefully revisited when coupling 3D models of larger vessels, for instance a 3D model of the entire aorta.

2.2.4 Applications

2.2.4.1 Acute hemorrhage at the abdominal aorta artery

An hemorrhagic episode taking place at the abdominal aorta is simulated. This is performed using an outflow terminal to induce an outgoing blood flow which is a function of the terminal resistance and of the pressure in the 1D segment corresponding to the abdominal aorta. This could be seen as the simulation of a partial rupture in vessel. Four cases are considered (i) the normal case (named *healthy*), (ii) a mild hemorrhage episode (named *hemorrhage*), (iii) an acute hemorrhage episode (named *acute hemorrhage*) and (iv) an acute hemorrhage episode suppressing the baroreflex control (named *no control*). Here, “acute” is used as a synonym of “severe”. Then, in cases (i)-(iii) the baroreflex is acting over the system and in case (iv) the baroreflex action is suppressed. The loss of blood volume in case (ii) is 135 cm³, and in cases (iii)-(iv) is 370 cm³.

The hemorrhagic event starts at time $t_s = 10$ s and ends at $t_e = 20$ s. The entire simulation goes up to $t_T = 45$ s such that the control mechanism restored the system to the new periodic state.

Variation of the heart rate are observed although it never reaches lower values than in the *healthy* case, due to a reduction in n_p . This entire process is shown in Figure 2.2. In such figure the evolution of defining parameters for the left ventricle like the heart rate (this is a common parameter for the four chambers) and the maximum elastance ($H = \frac{1}{T}$ and $E_{A,LV}$, respectively, see Table 2.2) are shown, from which we appreciate the different characteristic times for those actuators.

In Figure 2.3 the evolution of the pressure at the aortic arch is shown before, during and after the hemorrhagic episode. The *healthy* case is compared with the *hemorrhage* and *acute hemorrhage* scenarios in the left column, whilst the *acute hemorrhage* case is compared in the situations with and without control in the right column. Note the lack of sensitivity of the minimum diastolic pressure to the hemorrhagic event when the control is activated. In view of the changes produced by the

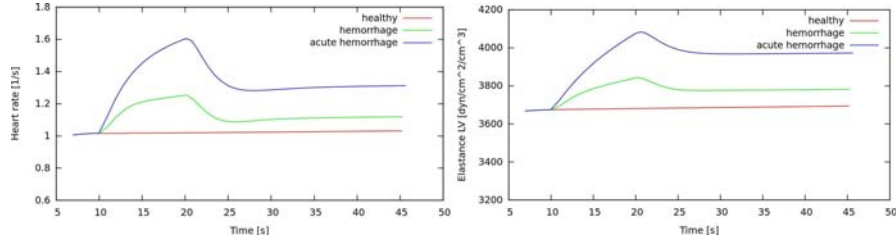


Fig. 2.2 Alteration of heart parameters (heart rate H on the left and maximum contractility $E_{A, LV}$ on the right) for the left ventricle when affected by the baroreflex in the three scenarios of interest.

baroreflex in the values of the actuators the significant change in the way in which reflections interact with each other gives rise to different pressure pulses. Particularly, note that the dicrotic notch is more accentuated as the hemorrhagic episode is more severe.

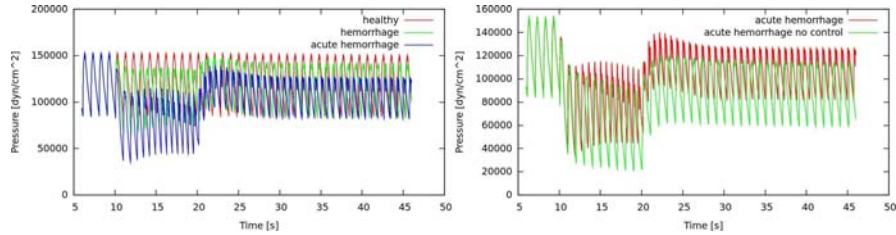


Fig. 2.3 Pressure at the aortic arch before, during and after the hemorrhagic episode. In the left figure we compare the normal scenario against two hemorrhagic episodes, while in the right figure the acute hemorrhagic case is compared with and without the baroreflex control.

The increase in n_s and the reduction in n_p after a hemorrhagic event modify the pressure-volume loops for the cardiac chambers. As observed in Figure 2.4 left, even with the baroreflex acting over the system, hemorrhage reduces volemia and therefore venous return to the heart affecting preload. The stroke volume SV (difference between end diastolic and end systolic volumes) and the ejection fraction EF (stroke volume/end diastolic volume) are reduced, and then the cardiac output CO (stroke volume \times heart rate H) is greatly decreased, as seen in Figure 2.4 right. This causes the heart to completely alter its work regime. Table 2.6 summarizes these parameters in the three cases of interest.

In Figure 2.5 the pressure and flow rate (evolution throughout the simulation) at the 1D segment representing the abdominal aorta artery are shown, specifically at a proximal location with respect to the location of the hemorrhagic point. Note that the dicrotic notch in the pressure pulse remains more accentuated in the hemorrhagic episodes, while we also observe a decrease in the maximum flow rate while keeping almost invariant the minimum backflow flow reached during a cardiac cycle once the system regain its periodic state.

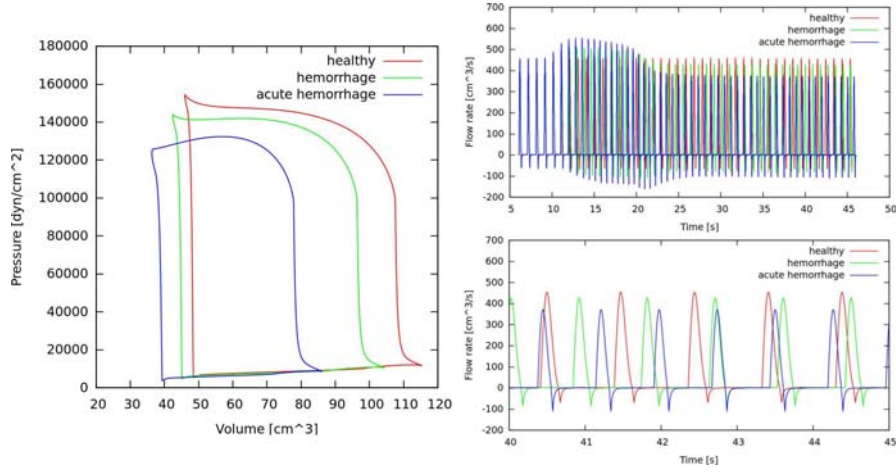


Fig. 2.4 Pressure-volume diagrams and resulting aortic flow for the three cases of analysis.

Table 2.6 Work regime for the left ventricle for the three scenarios of interest after the hemorrhagic episode when the periodic state has been restored.

Scenario	$SV^* [cm^3] (\%)$	$EF (\%)$	$CO^+ [cm^3/s] (\%)$
healthy	58.3 (100%)	0.531 (100%)	58.3 (100%)
hemorrhage	48.95 (84%)	0.504 (95%)	55 (94%)
acute hemorrhage	35.5 (61%)	0.444 (84%)	46.7 (80%)

$$^*SV = \frac{CO}{H} \quad ^+CO = \frac{1}{T} \int_t^{t+T} Q(\tau) d\tau \quad (T = \frac{1}{H})$$

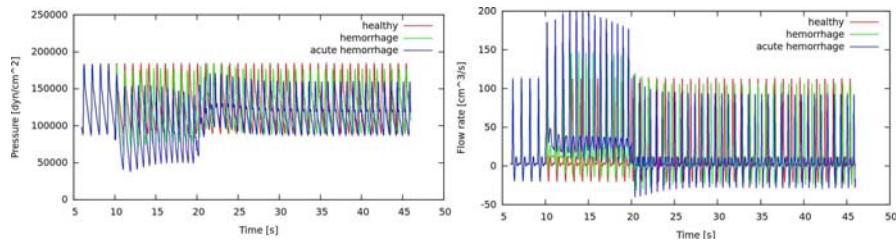


Fig. 2.5 Pressure and flow rate at the location of the hemorrhage in the abdominal aorta artery.

Figure 2.6 presents the same results at the common carotid artery, from where we can see the different hemodynamic environments triggered by the action of the baroreflex control at different locations of the arterial tree.

2.2.4.2 Local aneurysmal hemodynamics for aortic insufficiency

In this second example we simulate the impact of aortic insufficiency in global and local hemodynamic quantities and compare the result obtained with and without the baroreflex control acting over the system. This is a first attempt to account for

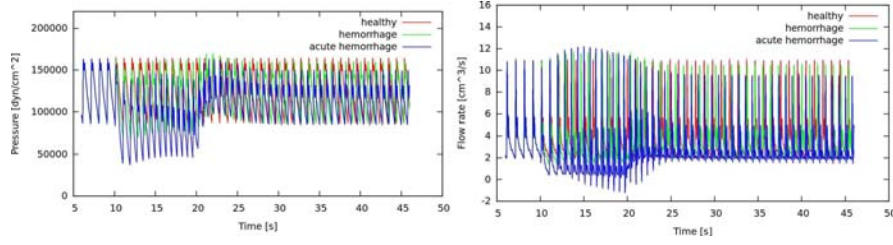


Fig. 2.6 Pressure and flow rate at the common carotid artery.

coupled global-local phenomena while taking into account multiple time scales, as stated in the introduction of this work.

The aortic insufficiency is modeled increasing the minimum opening angle $\theta_{min,Aov}$ (see [21]), Aov standing for aortic valve (see Table 2.2). Specifically we consider a level of severity given by $\theta_{min,Aov} = 30^\circ$ (the value corresponding to the healthy case is $\theta_{min,Aov} = 5^\circ$) from a certain instant in the simulation given by $t = 10$ s. The simulation goes up to $t_T = 45$ s. Three cases are compared (i) the normal case (named *healthy*), (ii) the regurgitant aortic valve with control case (named *regurgitant valve*) and (iii) the regurgitant aortic valve without control case (named *regurgitant valve no control*). Furthermore, we consider the presence of a saccular aneurism in the middle cerebral artery.

In the left part of Figure 2.7 the pressure-volume diagram of the left ventricle is shown. Due to the regurgitation, the cardiac output is reduced substantially: in the *regurgitant valve* case the CO falls to the 81% of the value corresponding to the *healthy* case, whereas in the *regurgitant valve no control* case the CO falls to the 68% of the same reference value. In addition, in the right part of Figure 2.7 the evolution of the pressure pulse at the aortic arch is shown as well as a detail of the resulting pulse once the system reached the new periodic state. Also, note that the accentuation of the diastolic notch is produced in the *regurgitant valve* case, when the baroreflex control is acting over the system. As a matter of fact, in the *regurgitant valve no control* case, when the baroreflex is removed, no accentuation is observed.

In Figure 2.8 a detail of the global hemodynamic quantities (pressure and flow rate) of interest at the middle cerebral artery are shown for the three scenarios of interest once the system reached the periodic state under the action of the baroreflex control. Note the larger fluctuations in the flow rate in the *regurgitant valve* case as a consequence of the increase in the sympathetic activity, which in turn affects the heart rate and cardiac contractility (among other actuators).

The hemodynamic indexes WSS and OSI are presented in Figure 2.9. The WSS index diminishes in the pathological condition in view of the reduced mean value of blood flow through the artery. Particularly, the WSS index is underestimated when the baroreflex control is not acting over the system (compare Figures 2.9(b) and 2.9(c)). Regarding the OSI index we observe that in the *regurgitant valve* case the region of higher OSI becomes more concentrated in the anterior view as it changes its structure as well, while it becomes more diffuse in the posterior

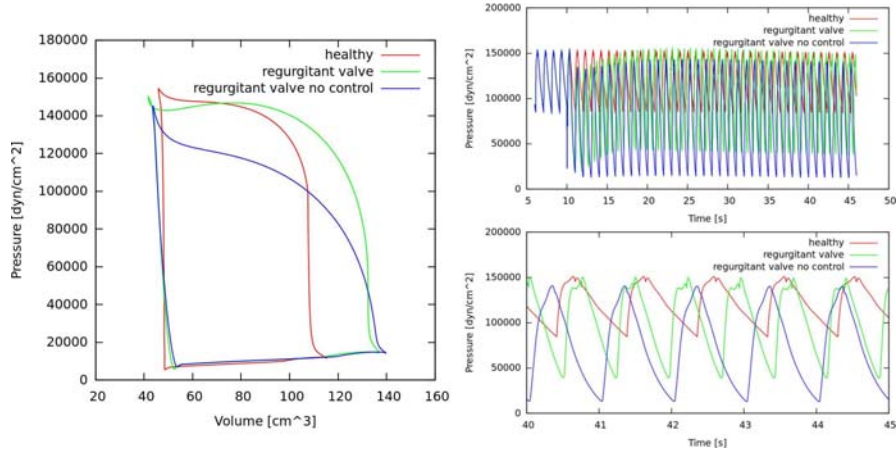


Fig. 2.7 Pressure-volume diagram for the left ventricle in the three cases of interest (left), evolution of the pressure pulse at the aortic arch throughout the simulation (top-right) and resulting pressure pulse after the system reached the new periodic state (bottom-right).

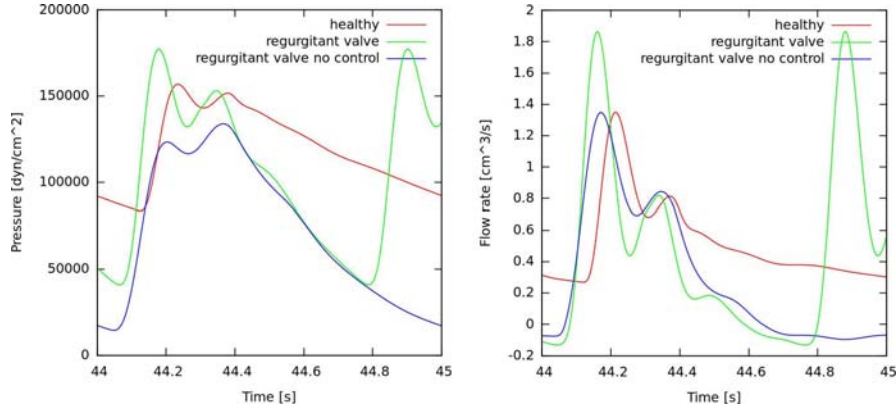


Fig. 2.8 Pressure (left) and flow rate (right) at the proximal location of the three-dimensional geometry of the cerebral aneurism.

view with respect to the *healthy* case. When compared with the *regurgitant valve* case, the *regurgitant valve no control* case rendered a similar distribution for the OSI.

2.3 Anatomically detailed models of arterial networks

In view of the large amount of data involved in the set-up of the model, previous works [29, 33, 1, 64] addressed the topology of the arterial network from a simplified point of view, that is by removing smaller arteries and simplifying connectivity

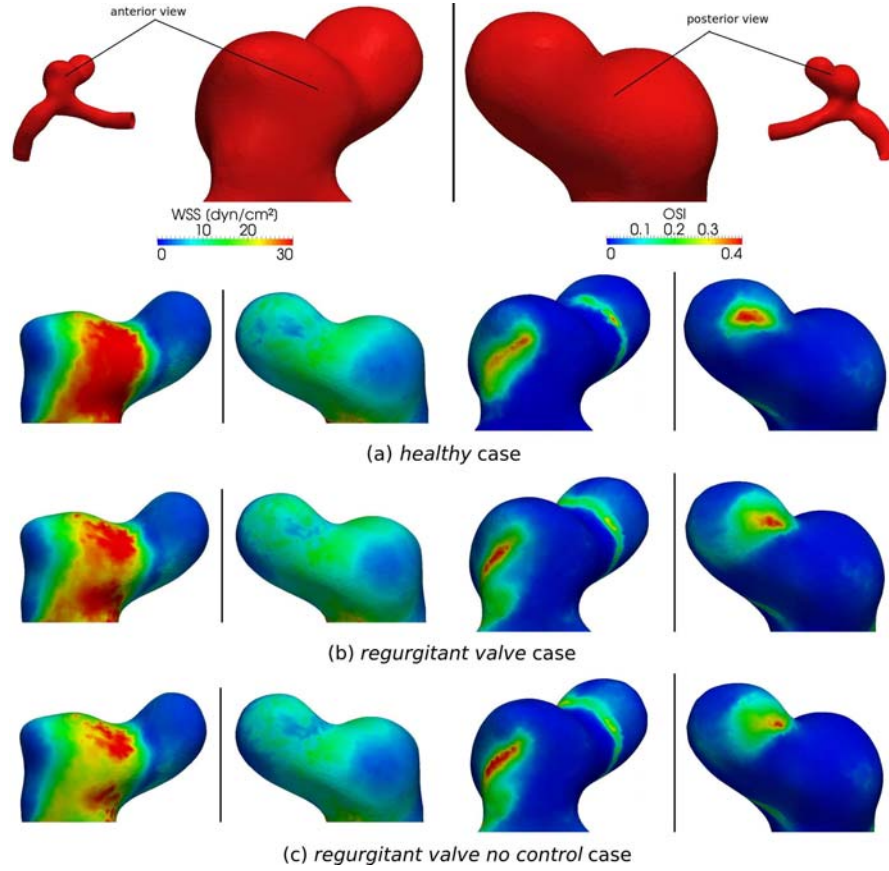


Fig. 2.9 WSS and OSI indexes in the three cases of interest.

among vessels as well as their relation with vascular territories. Nevertheless, there are specific applications which require to take into consideration the true topology of the cardiovascular system, redundancies in the circulation, anastomoses, etc. in order to account for a proper blood flow distribution as well as to be able to model certain physiological and/or pathophysiological scenarios. Surgical procedures like radial harvesting for coronary revascularization and its impact on the blood flow supply to the arteries in the hand is a clear example of medical interest [35]. Other instances are procedures related to reconstructive surgery, in which knowing the blood supply to the different vascular territories may be of potential clinical relevance in order to perform tissue transplantation successfully [69, 99].

Motivated by the preceding considerations and making use of one-dimensional models, the authors of this contribution have been working in the development of an arterial network as close as possible to the human cardiovascular system reported in medical literature. Previous works related to computational hemodynamics have dealt with arterial trees comprising at most 130 vessels as shown in Figure 2.10-A

[1]. In turn, the model we are developing is substantially detailed with more than 1200 arteries as can be seen in Figure 2.10-B,C and D. Nevertheless, due to space constraints, we present in this work the development and setting of only a part of this entire network. Hence, in the present contribution we focus on the arterial functioning and blood supply to the vascular territories of the arm. The novelty of the present work is the set-up of an anatomically detailed model for the arterial network of the arm following stringent anatomical and physiological considerations. The final goal is to establish the data infrastructure (geometrical and mechanical parameters) to be able to perform 1D blood flow simulations.

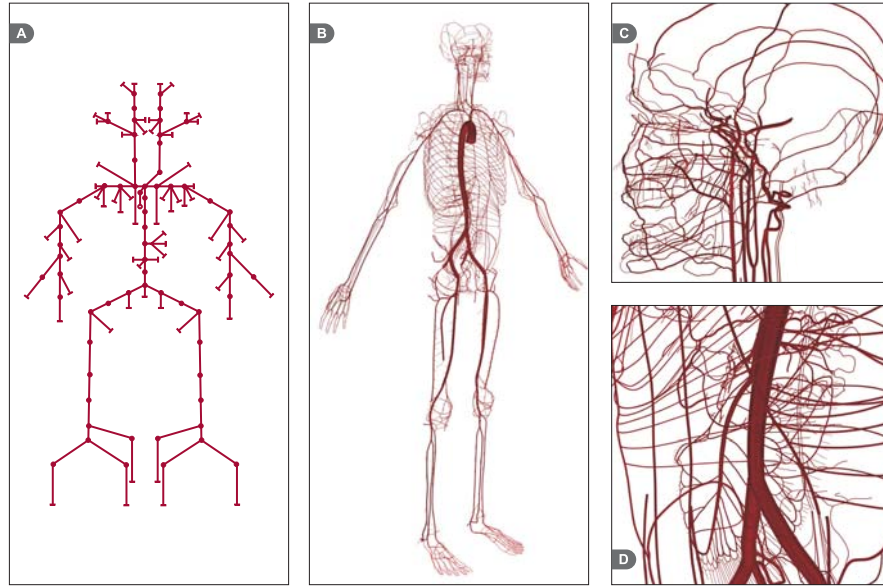


Fig. 2.10 A: Avolio's arterial tree with 128 vessels; B: overview of the anatomically detailed model in development; C: details of the arterial network of the head; D: details of the abdominal vessels.

2.3.1 Parameter estimation

In order to feed the governing equations, we need a set of geometrical and mechanical parameters for arterial vessels. Nevertheless, in most cases, the external diameters of arteries are the only information reported in the medical literature. Moreover, gathering the data related to the external diameters of all the main arteries turns to be an arduous task. We have carried out such task, therefore, from this point onwards we consider that the external arterial diameters are known, and then we discuss the approaches used to assess the *internal radius*, the *wall thickness* and the *effective*

Young's modulus. Furthermore, there were few cases in which the external arterial diameters were not found in the literature. There, we used either the Murray's law or the magnitude order of surrounding vessels to estimate them.

2.3.1.1 Wall thickness

A table with 76 values of internal radii and their respective wall thickness for 128 arterial segments is reported in [1]. Hence, for each internal radius r , we evaluated the ratio h/r , where h is the correspondent wall thickness. The 76 points were fitted by the curve given by

$$\frac{h}{r} = ae^{br} + ce^{dr} \quad (2.6)$$

with $a = 0.2802$, $b = -5.053$, $c = 0.1324$ and $d = -0.1114$. Therefore, given the lumen radius r for each artery, we use equation (2.6) to assess its wall thickness h .

2.3.1.2 Internal radius

As said, most of arterial calibers were found in morphometrical assessments whose purpose was to measure external diameters. However, we are concerned in the arterial lumen radius. Thus, denoting by D the external diameter of an artery, we obtain the reference value for its internal radius r by solving the following nonlinear equation

$$2r \left(1 + \frac{h}{r} \right) = D \quad (2.7)$$

where the ratio h/r is given by expression (2.6).

2.3.1.3 Elastic properties

Based on empirical data, an exponential expression for the ratio Eh/r as a function of the lumen radius was provided in [54], where E is the effective Young's modulus. So, this parameter E can be written as a function of (r, h) , yielding

$$E(r, h) = \frac{r}{h} \left(k_1 e^{k_2 r} + k_3 \right), \quad (2.8)$$

where $k_1 = 1.4 \times 10^7 \text{ dyn} \cdot \text{cm}^{-2}$, $k_2 = -22.53 \text{ cm}^{-1}$ and $k_3 = 6.055 \times 10^5 \text{ dyn} \cdot \text{cm}^{-2}$ are constants. However, once we have relation (2.6), then expression (2.8) can be seen as a function of a single variable r . We have adopted this approach to estimate the effective Young's moduli throughout the arterial network of the arm.

The values for k_1 and k_3 have been slightly modified from [54] (original values were $k_1 = 2.0 \times 10^7 \text{ dyn} \cdot \text{cm}^{-2}$ and $k_3 = 8.65 \times 10^5 \text{ dyn} \cdot \text{cm}^{-2}$), in order to fit the

pressure amplitude (diastole-to-systole amplitude) according to a physiologically significant range.

2.3.1.4 Bifurcation law

In order to estimate the internal radius for arterial segments not found in the specialized literature, we made use of Murray's Law of bifurcations (see [81]), which is given by the expression

$$r_o^3 = r_1^3 + r_2^3 \quad (2.9)$$

where r_o denotes the radius of a parent vessel and r_1, r_2 the daughter radii.

2.3.2 Anatomical and physiological data

2.3.2.1 Features of arterial connections in the arm

The main blood supply of the arm comes from the axillary artery, whose prolonging is the brachial artery. This last one courses until the bifurcation that gives rise to the radial and ulnar arteries close to distal end of the humerus. The brachial artery and its main branches nourish the upper arm, whereas radial and ulnar arteries as well as their branches account for the blood supply to the forearm and to the hand. Previous works including arm vessels as [1, 29, 31, 64] encompass only those main four arteries plus the interosseous artery, an important branch of the ulnar artery. In turn, terminals are placed at the distal end of the radial, ulnar and interosseous arteries to simulate the resistive and compliant effects of the vascular beds beyond the ending points. Such a simplifying procedure prematurely interrupts the pressure wave propagation, changing the way in which the blood flows and yielding a loss of information about wave reflections.

As already stated before, a novel arterial topology for the arm based on anatomical and physiological considerations was built. This arterial network consists in 108 segments, where 64 of them are the arteries listed in Table 2.7 and the other 44 are the so-called perforator vessels. The three-dimensional structure of the arterial topology was accurately designed taking into account the pathways of vessels through muscles, connective tissue and bones. In order to carry out this task, we have used anatomical references largely adopted by the medical community. Particularly, we were preferentially guided by the Netter's Atlas of Human Anatomy [53], the Pocket Atlas of Human Anatomy by Feneis [15] and the Clinically Oriented Anatomy [50]. Finally, the three-dimensional vessel design was also guided by a 3D digital skeleton structure with an average height of approximately 170 cm. Figure 2.11 presents several perspectives of the the vessels connectivity along the arm. Numbers are related to the data summarized in Table 2.7.

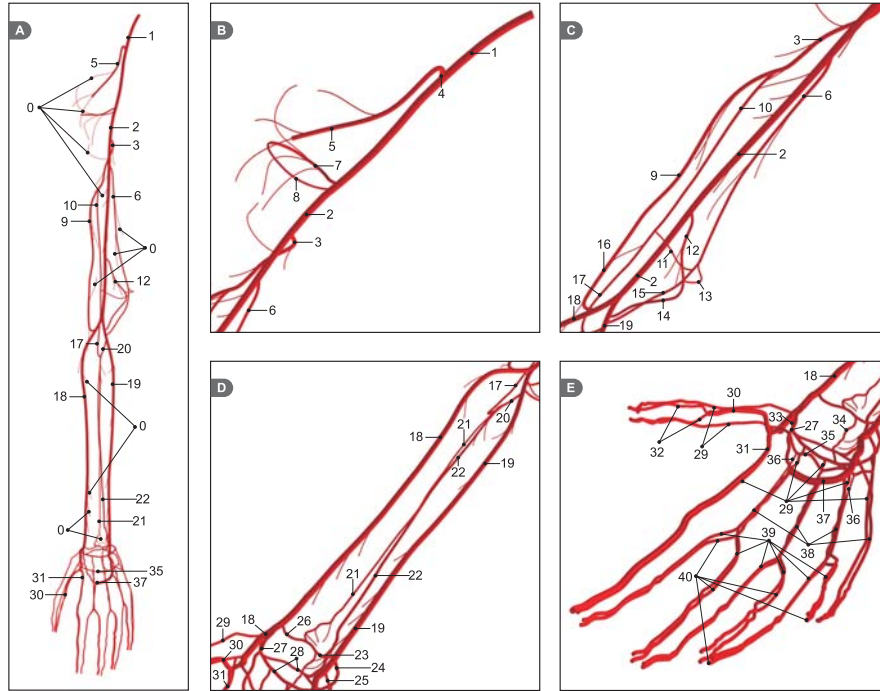


Fig. 2.11 Anatomically detailed arterial network of the arm. A: overview of the entire network. B and C: details of the upper arm. D: details of the forearm. E: details of the wrist and hand.

2.3.2.2 Morphometrical data

As explained before, available medical data provides arterial calibers in terms of external diameters. In addition, such studies are carried out using different populations, and some inconsistencies may arise. Thus, the external diameters which are part of the present model have been picked from the available literature such that they are consistent with surrounding neighbouring vessels. That is, a parent vessel cannot be smaller than its prolonging and bifurcating ones. In this sense the data retrieved from the literature (external diameters) is coherent.

Geometrical parameters of arterial vessels

Table 2.7 presents geometrical data of arterial vessels obtained from the literature and making use of (2.6), (2.7) and (2.9) to estimate the internal radii (see Section 2.3.1.2). The name of arteries followed the international nomenclature adopted in [15]. The number in brackets after each artery's name stands for how many times this artery takes place in the network so that the summation of all arteries must be equal to 64 segments. We point out that the lengths of arteries resulted from the

three-dimensional outline of vessels taking into account the skeletal and muscular structure and guided by anatomical references, as it was shown in Figure 2.11. The last column in Table 2.7 presents the bibliographical sources (or other criterion, when necessary) used to retrieve the external vessel diameters.

No.	Name	L (cm)	ED (cm)	LR (cm)	BS
1	Axillary (1)	12.003	0.560	0.230	[2], [71], [97]
2	Brachial (1)	22.312	0.510	0.208	[67], [89]
3	Profunda brachial (1)	6.221	0.156	0.058	(*)
4	Thoracoacromial (1)	0.653	0.379	0.150	[1]
5	Thoracoacromial deltoid branch (1)	7.539	0.260	0.100	(*)
6	Superior ulnar collateral artery (1)	15.151	0.175	0.066	[62]
7	Posterior circumflex humeral (1)	5.086	0.275	0.106	[17]
8	Anterior circumflex umeral (1)	4.690	0.115	0.042	[17]
9	Radial collateral (1)	11.511	0.125	0.046	[11]
10	Middle collateral (1)	15.207	0.125	0.046	[11]
11	Middle collateral network of elbow branch (1)	3.877	0.100	0.037	(**)
12	Inferior ulnar collateral (1)	5.065	0.145	0.054	[100]
13	Inferior ulnar collateral network of elbow branch (1)	2.637	0.100	0.037	(**)
14	Ulnar recurrent anterior branch (1)	4.835	0.160	0.060	[62] (***)
15	Ulnar recurrent posterior branch (1)	6.148	0.160	0.060	[62]
16	Radial recurrent (1)	6.784	0.200	0.076	[100]
17	Recurrent interosseous (1)	5.383	0.100	0.044	[100]
18	Radial (1)	30.216	0.350	0.138	[5], [89]
19	Ulnar (1)	26.904	0.357	0.141	[5], [89]
20	Common interosseous (1)	1.627	0.250	0.096	[14]
21	Anterior interosseous (1)	22.312	0.070	0.025	[69]
22	Posterior interosseous (1)	23.169	0.180	0.068	[69]
23	Ulnar palmar carpal branch (1)	1.678	0.065	0.024	[25]
24	Ulnar dorsal carpal branch (1)	1.481	0.115	0.042	[104]
25	Ulnar deep palmar branch (1)	1.918	0.260	0.100	[89] (***)
26	Radial palmar carpal branch (1)	1.578	0.075	0.027	[25]
27	Radial superficial palmar branch (1)	4.511	0.170	0.064	[5]
28	Dorsal carpal network (2)	3.689	0.065	0.024	[56]
29	Dorsal metacarpal (7)	8.737	0.065	0.024	[56]
30	Princeps pollicis (1)	4.654	0.193	0.073	[5]
31	Radial indicis (1)	12.261	0.153	0.057	[5]
32	Princeps pollicis branch (2)	3.960	0.155	0.058	(*)
33	Radial dorsal carpal branch (1)	0.873	0.135	0.050	[104]
34	Palmar carpal network (1)	3.293	0.065	0.024	[56] (***)
35	Deep palmar arch (1)	5.349	0.260	0.100	[89]
36	Palmar metacarpal (4)	2.941	0.065	0.024	[56] (***)
37	Superficial palmar arch (1)	5.175	0.280	0.108	[89]
38	Common palmar digital (4)	3.133	0.175	0.066	[18]
39	Proper palmar digital (6)	7.127	0.110	0.040	[89]
40	Dorsal digital (6)	7.242	0.052	0.019	(*)

Table 2.7 *Arterial network data.* (n): number of occurrences; L: lenght; ED: external diameter; LR: lumen radius; BS: bibliographical sources. (*) Murray's law was employed; (**) The order of magnitude of an anatomically similar artery was used; (***) The order of magnitude of surrounding arteries found in the bibliographical source was used.

2.3.2.3 Features of blood supply to vascular territories

Source arteries are vessels whose pathway is usually parallel and sometimes close to the bone skeleton. This kind of vessels mainly accounts for the blood transport. Examples of such vessels are the brachial artery, deep brachial artery, ulnar artery and radial artery. Narrower arteries emerge from these source vessels in order to nourish both the deep fascia and the superficial fascia. Roughly speaking, the deep fascia is composed by muscles and the superficial fascia by the subcutaneous fat and the skin. According to Taylor [86], a cutaneous perforator is any vessel that branches from any source artery, goes through the deep fascia and reaches the superficial fascia in order to feed a particular vascular territory.

On the other hand, by vascular territory we mean the portion of muscle, fat and skin that is fed by perforators that arise from some specific source vessels which play the role of blood suppliers. In other words, each vascular territory is associated to a source artery. We have based the present approach on [86, 51] about the placement of vascular territories and the sets of associated perforators. Particularly, the vascular territories of the arm and the corresponding source arteries are retrieved from [13]. Chen et al. have identified that perforators of the upper extremity emerge from 16 source arteries and, therefore, the upper extremity can be divided into 16 vascular territories. Figure 2.12 shows the division map of these 16 vascular areas of the arm.

Table 2.8 lists the source artery for each territory, the number of perforators per territory, the perforator's diameters and the fraction of area occupied by each vascular territory. All those values in Table 2.8 are mean values based on [13]. Finally, we used the mapping given by [69] in order to locate places from which perforators arise from their source arteries.

Vascular territory	Source artery	Area fraction	No. perforators	PMD (cm)
A1	Posterior circumflex humeral	7.16%	3	0.10
A2	Thoracoacromial (Deltoid branch)	5.27%	3	0.08
A3	Superior ulnar collateral artery	10.86%	2	0.09
A4	Profunda brachial	7.73%	4	0.07
A5	Brachial	8.39%	6	0.07
A6	Radial collateral	4.74%	2	0.10
A7	Inferior ulnar collateral	5.93%	2	0.08
A8	Radial recurrent	2.90%	2	0.07
A9	Radial	11.93%	5	0.06
A10	Ulnar	11.29%	7	0.06
A11	Posterior interosseous	3.86%	5	0.05
A12	Anterior interosseous	0.98%	3	0.05
A13	Dorsal arch and dorsal metacarpal	5.84%	7	0.05
A14	Princeps pollicis	2.51%	4	0.11
A15	Common palmar digital	1.81%	5	0.07
A16	Superficial palmar arch	8.80%	7	0.11

Table 2.8 Main data of the vascular territories. PMD: perforator mean diameter.

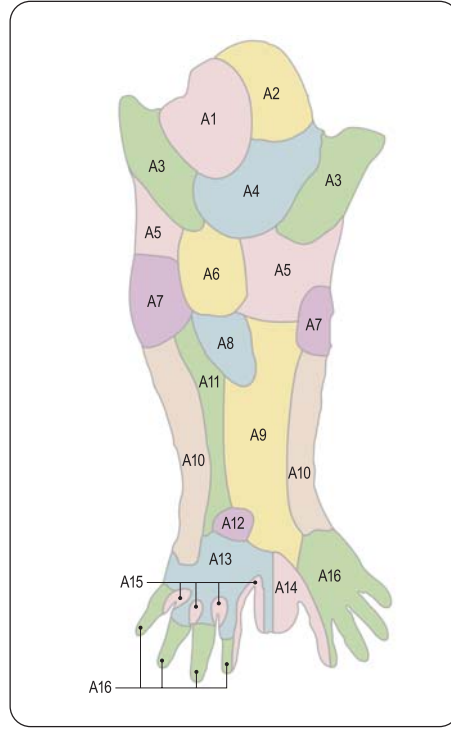


Fig. 2.12 *Vascular territories*. Adapted from [13].

Terminal parameters

We assessed the terminal resistances by taking into account the approximated muscle mass of each territory. For this aim, we assumed that the percentual area of each vascular territory, as listed in Table 2.8, provides a measure of the percentual of the related muscle mass. Then, data published in [27] allow us to evaluate the total muscular mass of the arm, while the work of Mahabir [46] allows us to estimate the muscle vascular resistance of each territory as a function of such mass. Finally, these 16 resistances were proportionally scaled in order to obtain physiological results in terms of the mean pressure. Moreover, the estimation of resistive parameters is carried out under the hypothesis that the drop pressure in the main arteries is negligible when compared with the drop pressure in terminal resistances. In total, there are 67 terminal models distributed throughout the network.

2.3.3 Simulations and results

In order to carry out the simulations, an inflow boundary condition is considered at point **A** shown in Figure 2.14. This flow rate curve corresponds to the result obtained at the proximal point of the axillary artery using an entire model of the cardiovascular system [3].

The general parameters are $\mu = 0.04$ P, $\rho = 1.04$ g/cm³ and $P_o = 133333$ dyn/cm². As said, the finite element method was used to approximate the mathematical formulation. The element size in all the simulations is $\Delta x = 0.25$ cm and the time step is $\Delta t = 0.0025$ s. Three cardiac cycles have been simulated until the periodic state was reached. For this setting a satisfactory mass conservation is obtained (relative error 0.3%).

Figure 2.13 shows a comparison between the pressure cuves given in [58] and the ones provided by the model developed here. Clearly, even with a very simple constitutive behavior and the inflow condition is not provided in [58], the model manages to reproduce the main pressure pulse features of patient records.

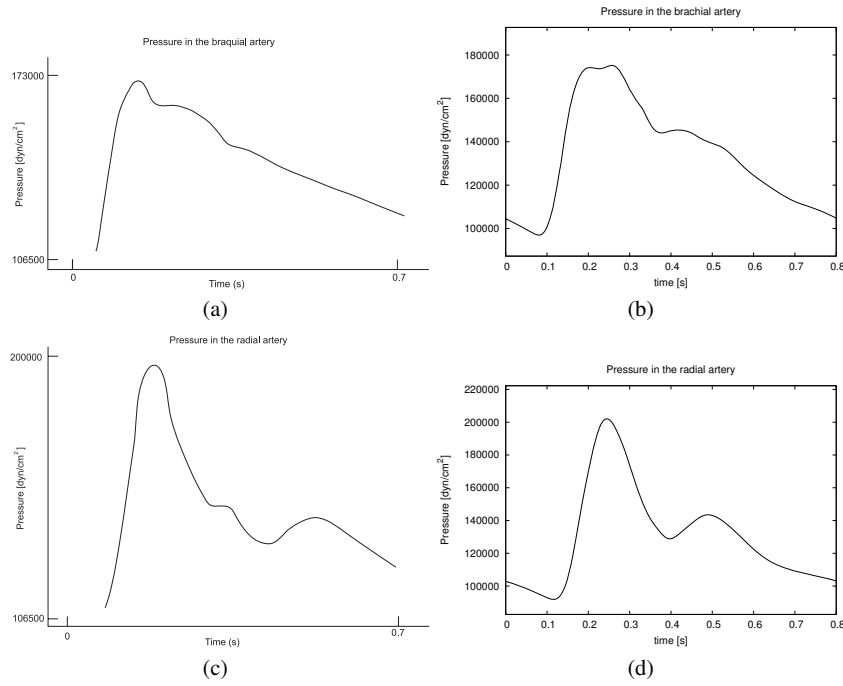


Fig. 2.13 Pressure in the brachial artery: (a) found in [58] p.184 and (b) provided by the model. Pressure in the radial artery: (c) found in [58] p.184 and (d) provided by the model.

Figure 2.14 presents the inflow curve used as the proximal boundary condition (see point A in such figure) and the flow rate curves in six arterial districts of the arm. An important feature of the blood flow in the arm with potential impact on surgical procedures is the determination of vessel dominance in terms of blood supply to the hand [89]. In the present model, the mean blood flow through the ulnar artery results in $0.7069\text{cm}^3/\text{s}$, while through the radial artery the mean blood flow is $0.6627\text{cm}^3/\text{s}$. So, the ulnar artery prevails over the radial artery as a blood supplier to the vessels in the wrist and hand, which is the most common situation encountered in practice.

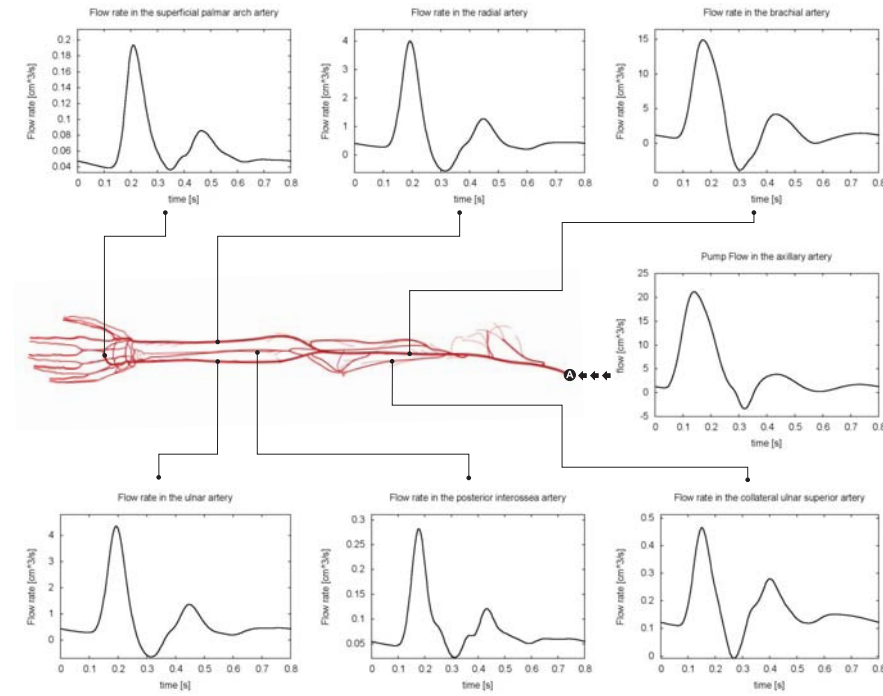


Fig. 2.14 Blood flow rate at several locations in the arterial vasculature of the arm.

2.3.4 Applications

Dimensionally heterogeneous-modeling

Dimensionally-heterogeneous modeling has been applied in a wide variety of applications in order to provide insight on the local blood flow patterns and its relation to global hemodynamic environments [8, 5, 4, 14, 20, 31]. In this section we apply the detailed model of the arterial network of the arm developed in

Sections 2.3.1 and 2.3.2 to study the blood flow in a branching site. Regarding the description of the numerical methodologies to address the solution of the 3D Navier-Stokes equations in compliant domains and its coupling with the 1D model the reader is referred to [3, 6, 23].

The 3D model matches the geometrical and mechanical parameters of the 1D model in which it is embedded. The bifurcation under analysis corresponds to the branching of the brachial artery into the ulnar and radial arteries. Also, the ulnar recurrent anterior and posterior branches, as well as the branches joining with the common interosseous and the radial recurrent vessels are present (see Figure 2.11 and Table 2.7 for a reference to these vessels). Figure 2.15 displays the location of the 3D model into the 1D network and the results in terms of pressure and flow rate at the coupling interfaces. Positive flow rate is always defined in the proximal-to-distal direction. Notice the change in the pressure wave form as a result of the vessel distensibility, and recall that blood flow distribution is mostly dictated by resistive parameters representing the vascular territories.

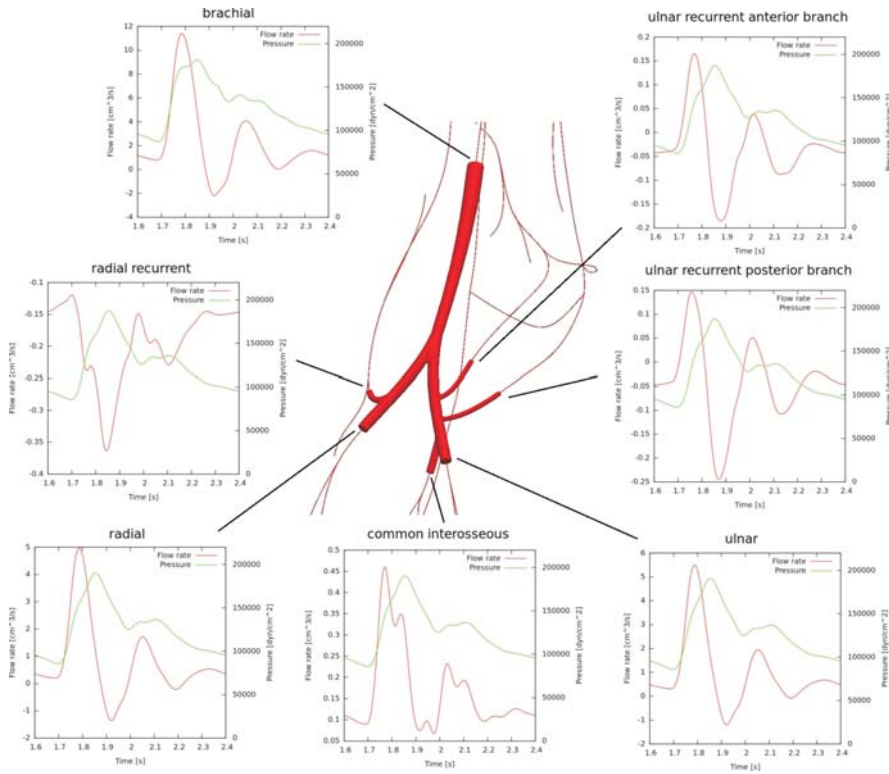


Fig. 2.15 Pressure and flow rate at coupling interfaces between the 3D and the 1D models.

In Figure 2.16 the pressure field in the 3D geometry is presented at four time instants in the cardiac cycle. Observe the inversion of the pressure gradient right

after systole. This change in the direction of the pressure gradient is responsible for the presence of Womersley-like velocity profiles during diastole, as seen in the same figure. Also, observe the high complex pattern of blood flow at the posterior and anterior recurrent ulnar branches, a common feature during the entire cardiac cycle. This feature is absolutely neglected when making use of 1D modeling, which may cause an impact on the flow rate wave form and the mean flow direction as observed in Figure 2.15. In such figure, it is observed that the mean flow rate through these ulnar branches is negative (outgoing the 3D bifurcation), pointing out that these are natural anastomotic branches present in the elbow which form a graft with the recurrent interosseous artery.

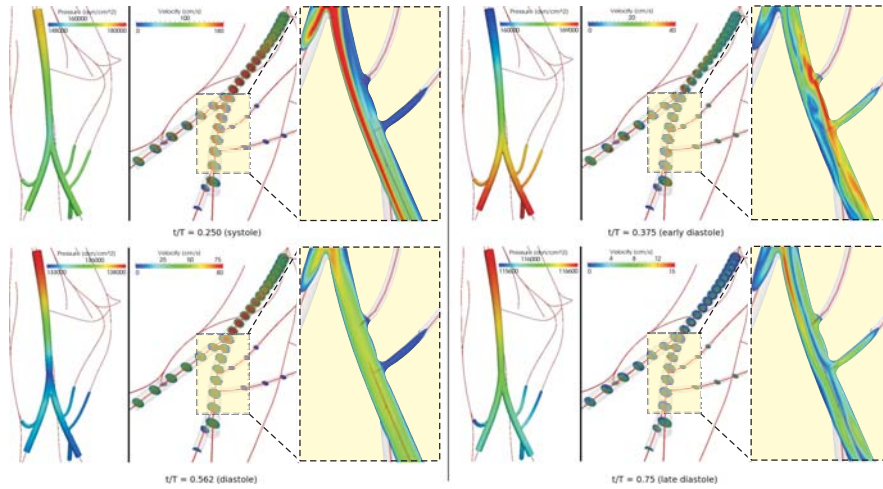


Fig. 2.16 Pressure and velocity profiles in several slices at different time instants during the cardiac cycle. Also, a detail of the velocity field is shown at the ulnar artery and its recurrent branches.

2.4 Constructive methods of peripheral vascular networks

The purpose of the arterial network is to convey blood to tissues and organs. They represent highly complex branching structures that have been found to follow certain optimality principles [52, 66, 88]. Bifurcation laws [103] are believed to be fulfilled in real arterial trees and several target functions related to functional optimality of arterial trees have been proposed [30, 40, 52, 81].

In order to understand both diagnostic and probable effects of therapeutic interventions, studies based on hemodynamics simulations use computational models of arterial trees as their geometrical substrate. To date, these simulations can be based on any of the following classes of network models: lumped parameter models

[26, 48], anatomical models [57, 94], fractal models [16, 95] and models generated by Constrained Constructive Optimization (CCO) [72, 73].

Lumped parameter models reduce the features of complex arterial trees to several compartments, each of which represents a certain class of vessel (e.g. arteries, arterioles, capillaries, venules, veins) jointly characterized by a resistance, inertance and compliance; thereby these compartmental models ignore the details of geometric structure.

Anatomical models provide a specific and accurate representation of the coarse structure of real arterial trees. They describe severely pruned trees.

Fractal models are generated from root segment to terminal segments according to concepts of self-similarity (deterministic or stochastic). They reproduce the statistical distributions of segment radii, lengths and angles of real vascular trees. However, they do not guarantee that their segments can be arranged properly in three-dimensional space.

CCO models allow to generate realizations of arterial trees which are both detailed and realistic. Although they are based on lumped descriptions of arterial vessels, they are more detailed than lumped parameter models because of their highly complex topological structure. Furthermore, the spatial arrangement in three-dimensional space of segments is optimized instead of being predefined as in the case of fractal models. Differently from anatomical models, distributions of segment radii and lengths emerge from the generation procedure rather than being plugged in.

Without the input of anatomical information, CCO trees are grown by stepwise adding one terminal segment after the other until a given number N_{term} of terminal segments is reached. At each step of the growth process, the geometry of the newly created bifurcation is optimized according to a given target function (e.g. total intravascular volume) and a prescribed set of physiological boundary conditions and constraints.

Arterial tree models generated by CCO are able to mimic important properties of real arterial trees, such as segment radii [31, 73], branching angle statistics [74] and pressure profiles [75, 77]. Thus, these CCO trees can serve as an anatomical substrate for computational simulations of hemodynamic phenomena.

The technical approach of generating trees via CCO has previously been described in detail [31, 72, 73]. In the following, we recall the main features of the CCO method for ease of reference.

2.4.1 Model assumptions

Arterial model trees are generated by CCO based on the assumptions below [31, 73]:

- (i) the piece of tissue to be perfused is geometrically represented by a perfusion area [72, 73] or a perfusion volume [31, 80];
- (ii) the arterial tree is modeled as a binary dichotomously branching (binary) system of straight cylindrical tubes (vessel segments);

- (iii) the model of the tree starts at the root segment (feeding artery) and it is truncated in the form of terminal segments on prearteriolar level;
- (iv) the model of the tree should cover the space of the perfusion domain as homogeneously as possible without intersection of segments;
- (v) the blood is modeled as an incompressible, homogeneous, Newtonian fluid at steady state and laminar flow conditions;
- (vi) the flow resistance R of each segment in the tree is assumed to follow Poiseuille's law [22]

$$R = \left(\frac{8\eta}{\pi} \right) \frac{l}{r^4}, \quad (2.10)$$

where l and r are the length and radius of the segment, η is the constant blood viscosity ($\eta = 3.6 \times 10^{-3}$ Pa s);

- (vii) the pressure drop Δp_s along each segment is given by

$$\Delta p_s = RQ, \quad (2.11)$$

where Q is the flow through the segment.

2.4.2 Physiological boundary conditions and constraints

At each stage of the construction process, the CCO tree model satisfies a set of physiological boundary conditions and constraints:

- (i) each terminal segment supplies an individual amount of blood flow Q_{term} into the microcirculatory network;
- (ii) all terminal segments drain blood against a given, constant, terminal pressure, p_{term} ;
- (iii) the resistance of the resulting model of the tree induces a pre-specified perfusion flow Q_{perf} across the overall pressure gradient

$$\Delta p = p_{\text{perf}} - p_{\text{term}}, \quad (2.12)$$

where p_{perf} is the perfusion pressure at the inlet of the root segment;

- (iv) the model of the tree with N_{term} terminal segments has $N_{\text{tot}} = 2 \times N_{\text{term}} - 1$ segments in total due to the binary branching mode;
- (v) at bifurcations, the radii of the parent vessel (say r_0) and of the daughter segments (say r_1 and r_2) are forced to fulfill a bifurcation law

$$r_0^\gamma = r_1^\gamma + r_2^\gamma, \quad (2.13)$$

where γ is a constant exponent with ranging between 2.55 and 3, governing the shrinkage of radii across bifurcations. Minimum reflection of pulse waves at bifurcations is achieved with $\gamma = 2.55$ [3]. For uniform shear stress all over tree, γ should be set to 3 [66]. From Murray's law [81], $\gamma = 3$ is obtained as a

necessary condition for minimum energy consumption in a vascular network. In the present work, the bifurcation exponent γ was set to 3 during CCO tree construction (see Section 2.4.4).

2.4.3 Generation of 3D optimized arterial trees

The generation of 3D optimized arterial trees starts by planting the root segment with its proximal point placed over the boundary of the volume to be perfused, and its distal point located at a randomly selected position within the perfusion volume. According to the resulting length of the root segment (l_{root}) and the overall pressure drop (according to equation 2.12), the radius r_{root} is calculated such that the hydrodynamic resistance R (see expression 2.10) yields the flow Q_{term} through one terminal segment ($k_{term} = 1$) into the microcirculatory network.

Given a tree with k_{term} terminal segments, the stepwise growth of the tree is as follows. First, the location \mathbf{x}_{term} for a new terminal is selected from a pseudo-random number sequence, uniformly distributed inside the perfusion volume. In this work, we used the dSFMT (double precision SIMD-oriented Fast Mersenne Twister) free generator [68] to produce the pseudo-random number sequence for tossing.

The prospective location \mathbf{x}_{term} is accepted as a candidate for a new terminal site only if \mathbf{x}_{term} satisfies a distance criterion (see Section 2.4.3.1).

When \mathbf{x}_{term} has been accepted as a distal end of a new terminal segment, it is temporarily connected to each of the neighboring segments, one after the other. Connecting the new terminal segment to a preexisting segment violates the boundary condition regarding the terminal flows. In order to return the proper terminal flows, the flow resistance of the tree must be adjusted when performing each new temporary connection. This can only be done by rescaling the segments' radii [31] (see Section 2.4.3.2).

The bifurcation site resulting in each temporary connection is optimized in order to minimize a target function (see Section 2.4.3.3). After having investigated all possible connections in the neighborhood of \mathbf{x}_{term} , the connection that provided the lowest value for the target function (total intravascular volume in the present work) is adopted as permanent connection for the new terminal site \mathbf{x}_{term} . Thus, the tree is grown to $k_{term} + 1$ terminal segments. The process of growing the tree summarized above is repeated until $k_{term} = N_{term}$, i.e., the pre-set number of terminals N_{term} is achieved.

2.4.3.1 Distance criterion

The distance criterion consists in checking whether the minimum distance d_{crit} of a new candidate terminal location \mathbf{x}_{term} to all existing segments exceeds an adaptive threshold d_{thresh} , which is properly decreased when the number of terminal segments increases during the tree generation process, or not. In [73] a distance

criterion was formulated for the two-dimensional case. Below, the extension to the three-dimensional case is developed.

For a trial location $\mathbf{x}_{term} = (x, y, z)$, the minimum distance d_{crit} to each of the existing segments in the model tree is calculated as follows. Given $\mathbf{x}_{Dj} = (x_{Dj}, y_{Dj}, z_{Dj})$ the distal and $\mathbf{x}_{Pj} = (x_{Pj}, y_{Pj}, z_{Pj})$ the proximal ends of an existing segment j in the tree. Consider the vectors $\mathbf{u} = \mathbf{x}_{Pj} - \mathbf{x}_{Dj}$, $\mathbf{v} = \mathbf{x}_{term} - \mathbf{x}_{Dj}$, $\mathbf{w} = \mathbf{x}_{term} - \mathbf{x}_{Pj}$. Note that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. The projection of \mathbf{v} onto \mathbf{u} is calculated as

$$\text{proj}_{\mathbf{u}} \mathbf{v} = d_j(\mathbf{x}_{term}) \mathbf{u}, \quad (2.14)$$

where $d_j(\mathbf{x}_{term})$ is a scalar given by

$$d_j(\mathbf{x}_{term}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}} = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}|^2}, \quad (2.15)$$

where “ \cdot ” denotes the scalar product between vectors.

If $d_j(\mathbf{x}_{term}) \in [0, 1]$, the projection of \mathbf{v} onto \mathbf{u} lies somewhere along segment j . In this case, the orthogonal distance is considered critical and obtained as

$$d_{crit} = \frac{|\mathbf{v} \times \mathbf{w}|}{|\mathbf{u}|}, \quad (2.16)$$

where “ \times ” denotes the cross product. The numerator is simply twice the area of the triangle formed by points \mathbf{x}_{term} , \mathbf{x}_{Dj} and \mathbf{x}_{Pj} , and the denominator is the length of the base of this triangle.

If $d_j(\mathbf{x}_{term}) \notin [0, 1]$, the minimum distance between \mathbf{x}_{term} and one of the endpoints of segment j is assumed critical and calculated as

$$d_{crit} = \min \{|\mathbf{v}|, |\mathbf{w}|\}. \quad (2.17)$$

If d_{crit} obtained to each of the preexisting segments in the tree exceeds a threshold distance d_{thresh} , i.e. $d_{crit} > d_{thresh}$, then \mathbf{x}_{term} is accepted to become the location of the distal end of the next terminal segment to be added.

If $d_{crit} < d_{thresh}$, tossing is repeated up to $N_{toss} = 10$ times. In case no acceptable location has been found, the threshold distance d_{thresh} is decreased by a factor 0.9 and another loop of $N_{toss} = 10$ times is launched.

Before adding a new terminal segment, d_{thresh} is initially established proportional to the perfusion domain available per terminal site $(r_{perf}^2/k_{term})^{1/2}$ [74], where k_{term} is the current number of terminal segments and r_{perf} is a characteristic length of perfusion domain (radius).

2.4.3.2 Rescaling segments' radii

Without loss of generality, we assume that the segment *con* of the CCO model tree is chosen site for connection of the new terminal segment *term* (see Figure 2.17).

Thus, *con* is shortened and the new bifurcation segment *bif* is created, as also seen Figure 2.17.

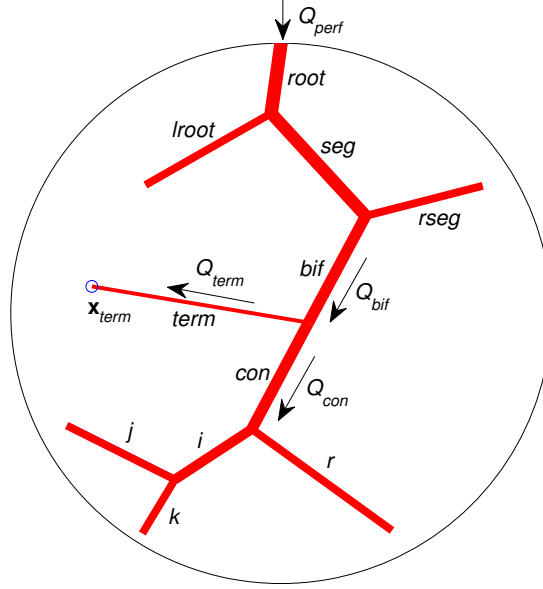


Fig. 2.17 Connecting the new terminal segment *term* to the segment *con* of the CCO model tree generates a new bifurcation segment *bif*.

The flow splitting ratio at the segment *bif* has to be adjusted so as to yield Q_{term} towards the new terminal segment *term*, and the flow Q_{con} into the subtree distal to *bif*. This can be adjusted via the ratio for the radii of the daughter segments of the segment *bif*, as given by [31]

$$\frac{r_{con}}{r_{term}} = \left(\frac{Q_{con} R_{sub,con}^*}{Q_{term} R_{term}^*} \right)^{1/4}, \quad (2.18)$$

where $R_{term}^* = 8\eta l_{term}/\pi$ is the reduced resistance of the new terminal segment *term*, and $R_{sub,con}^*$ is the reduced resistance of segment *con* including its distal continuing daughter trees (subtrees). According to [31], $R_{sub,con}^*$ can be determined by recursively traversing the subtrees of *con* in postorder mode and determining at each segment *i*

$$R_{sub,i}^* = \left(\frac{8\eta}{\pi} \right) l_i + \left[\frac{(r_{left,i}/r_i)^4}{R_{left,i}^*} + \frac{(r_{right,i}/r_i)^4}{R_{right,i}^*} \right]^{-1}, \quad (2.19)$$

where $R_{left,i}^*$ is the reduced resistance and $r_{left,i}$ is the inlet radius of the left subtree of segment i .

Using the power law (2.13) at bifurcation bif and radii ratio r_{con}/r_{term} (see equation (2.18)), the bifurcation ratios with respect to segment bif are obtained as [31]

$$\begin{aligned}\beta_{con} &= \frac{r_{con}}{r_{bif}} = \left[1 + \left(\frac{r_{con}}{r_{term}} \right)^{-\gamma} \right]^{-1/\gamma}, \\ \beta_{term} &= \frac{r_{term}}{r_{bif}} = \left[1 + \left(\frac{r_{con}}{r_{term}} \right)^{\gamma} \right]^{-1/\gamma},\end{aligned}\quad (2.20)$$

These bifurcation ratios guarantee that the flow Q_{bif} into segment bif splits correctly into the flows Q_{con} and Q_{term} . By the conservation of mass, $Q_{bif} = Q_{con} + Q_{term}$.

Since the new terminal segment $term$ is now be supplied via the segment bif (daughter of the segment seg), the respective bifurcation ratios r_{bif}/r_{seg} and r_{rseg}/r_{seg} must be recalculated by means of the equations given above. The same holds for all bifurcations upwards segment bif , towards the root segment.

The absolute value of the radius of the root segment r_{root} is given by

$$r_{root} = \left[R_{sub,root}^* \frac{Q_{tot}}{\Delta p} \right]^{1/4}, \quad (2.21)$$

where Q_{tot} is correct absolute flow into the tree, $Q_{tot} = Q_{perf}$ when a tree of $k_{term} = N_{term}$ terminal segments has been generated.

The absolute value radius of segment i is obtained as follows [31]

$$r_i = r_{root} \prod_{k=i}^{root} \beta_k, \quad (2.22)$$

where β_k are the corresponding bifurcation ratios and the product is extended over all bifurcating segments on the unique path from segment i to the root segment.

2.4.3.3 Optimization target function

CCO trees grown under different optimization target functions show differences in structure that can be quantified by appropriately chosen numerical indexes [75]. In order to quantify the optimality of the CCO tree, the total intravascular volume is chosen as the optimization target function [30, 31, 40, 52, 73, 81]. This target function reads

$$V = \pi \sum_{i=1}^{k_{tot}} l_i r_i^2, \quad (2.23)$$

where k_{tot} is the total number of segment in the tree ($k_{tot} = 2k_{term} - 1$).

Within the CCO framework, both local and global aspects of optimization are governed by same target function (2.23).

Once the location \mathbf{x}_{new} has been accepted as a new terminal site, it is temporarily connected to the midpoints of each neighboring segment, one after the other. For each temporary connection, the initial location of the resulting bifurcation is moved by a gradient optimization procedure [61] until a (local) minimum of the target function V is achieved, while the physiological boundary conditions and constraints are fulfilled by appropriate scaling of segment radii (see Section 2.4.3.2). This is called *geometric optimization* (local aspect). The minimum value V and location of the bifurcation for each of these temporary connections are stored.

Finally, the so added temporary connection with the lowest value V is adopted and that connection is made permanent. This is called *structural optimization* (global aspect). Observe that geometric optimization is nested within structural optimization.

2.4.4 Results

The CCO method described previously has been implemented. The results obtained with the CCO algorithm for different situations are here presented.

2.4.4.1 Comparison with real coronary arterial trees

For a comparison with real morphometric data corresponding to coronary arterial trees, the CCO method was employed to generate arterial trees of a portion of tissue with spherical geometry, whose perfusion volume is $V_{perf} = 100 \text{ cm}^3$ [31]. In the CCO simulations, the perfusion pressure was set to $p_{perf} = 100 \text{ mmHg}$, terminal pressure $p_{term} = 72 \text{ mmHg}$, total perfusion flow $Q_{perf} = 500 \text{ ml/min}$, and the terminal flow $Q_{term} = 2 \text{ ml/min}$.

Ten replicates of the tree with 250 terminal segments were generated for the same predefined parameters and the same optimization target function (2.23). The difference among trees lays in the sequence of pseudo-random numbers for casting the distal ends of its terminal segments.

The morphometric result of the CCO-generated trees was compared with experimental data acquired from corrosion casts from two human hearts reported in [102]. There, the authors classified each vessel segment according to the bifurcation level (number of proximal bifurcations), measured the diameter, and plotted average diameter versus bifurcation level.

Figure 2.18 displays the mean diameter and standard deviation (SD) of all vessel segments versus the corresponding bifurcation level. Close agreement with the experimental data is found for the diameter of the root segment (level 0), for levels 1-3 and levels 7-19. The result shown in Figure 2.18 is consistent with that reported in [31].

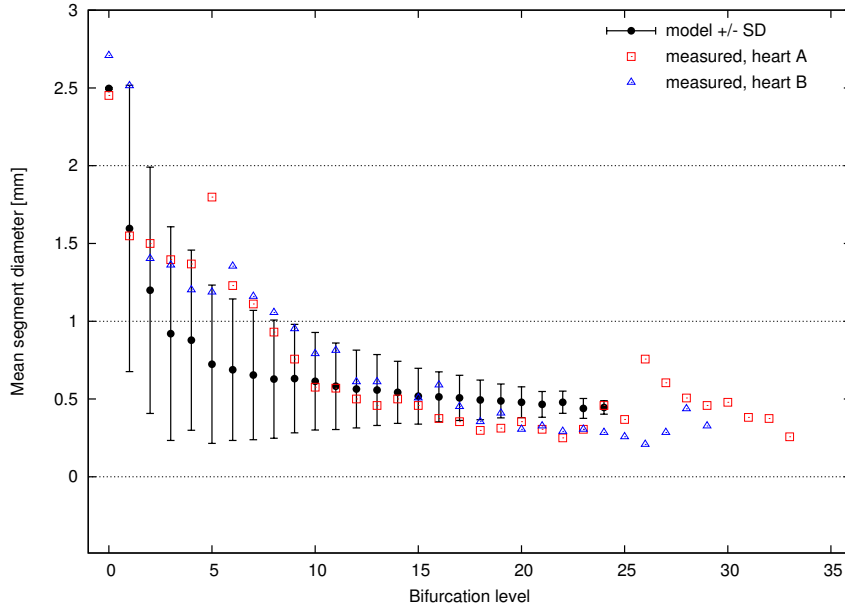


Fig. 2.18 Comparison of structural features of CCO-generated trees with real morphometric data from coronary arterial trees.

2.4.4.2 CCO-trees supplying 3D perfusion domains of different shape

In this section, the CCO method is employed to generate the arterial tree within three different perfusion domains under identical physiological boundary conditions, which are: perfusion pressure $p_{perf} = 100$ mmHg, terminal pressure $p_{term} = 60$ mmHg, total perfusion flow $Q_{perf} = 500$ ml/min and perfusion volume $V_{perf} = 100$ cm³.

Figure 2.19 illustrates the CCO-generated trees with $N_{term} = 4000$ (i.e., 7999 segments in total) optimized such that minimum intravascular volume is obtained, while maintaining equal terminal flow $Q_{term} = 0.125$ ml/min. In this figure, the image with color scale presents the vessel radii for each vessel, where it is possible to appreciate the course of the main branches within the spherical domain.

Figures 2.20 and 2.21 present a similar case regarding the physiological data to construct the tree. Note in these figures that even the shape of the 3D region to be perfused is different, the CCO method can equally be applied, generating arterial vessels with terminal points homogeneously distributed. In both figures, it can be seen that the main branches that course through the domain do it in a quite anisotropic manner in view of the high geometrical aspect ratio of the vascular domain.

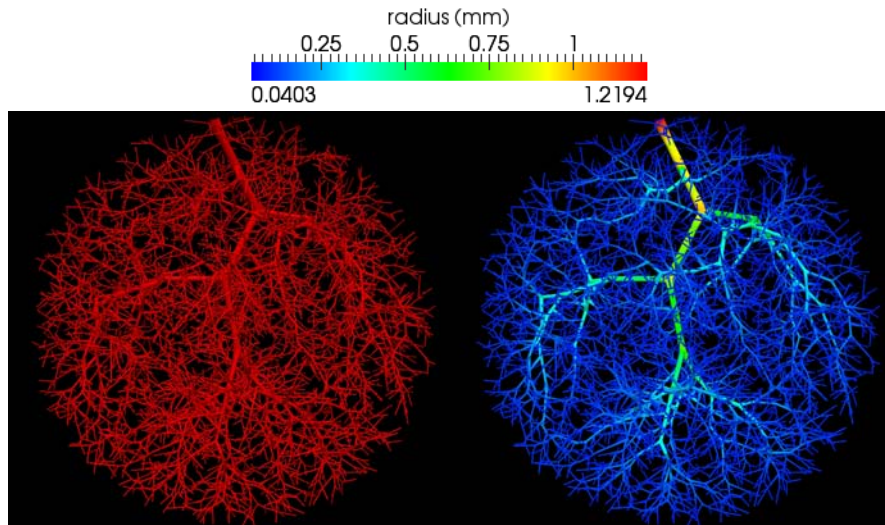


Fig. 2.19 CCO-generated arterial tree model in a spherical domain with 4000 terminal segments.

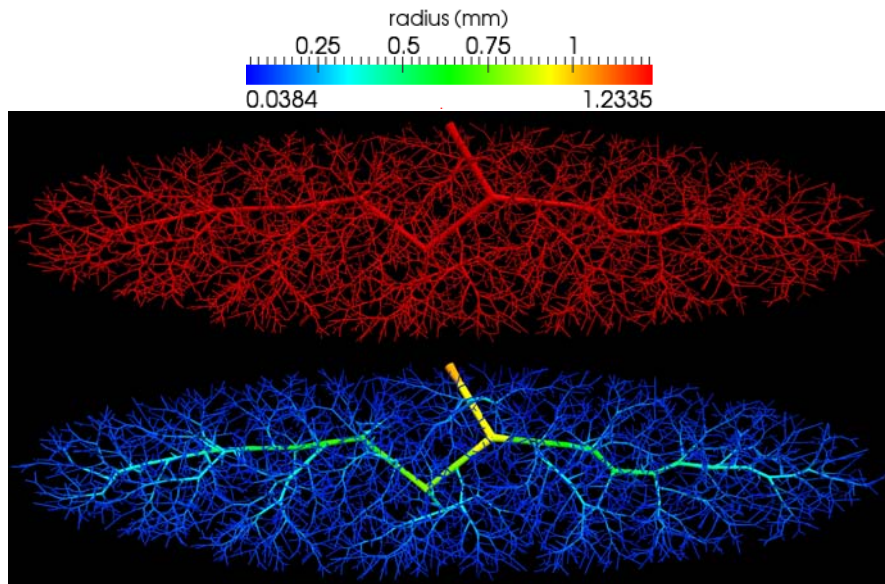


Fig. 2.20 CCO-generated arterial tree model in an elliptical domain with 4000 terminal segments.

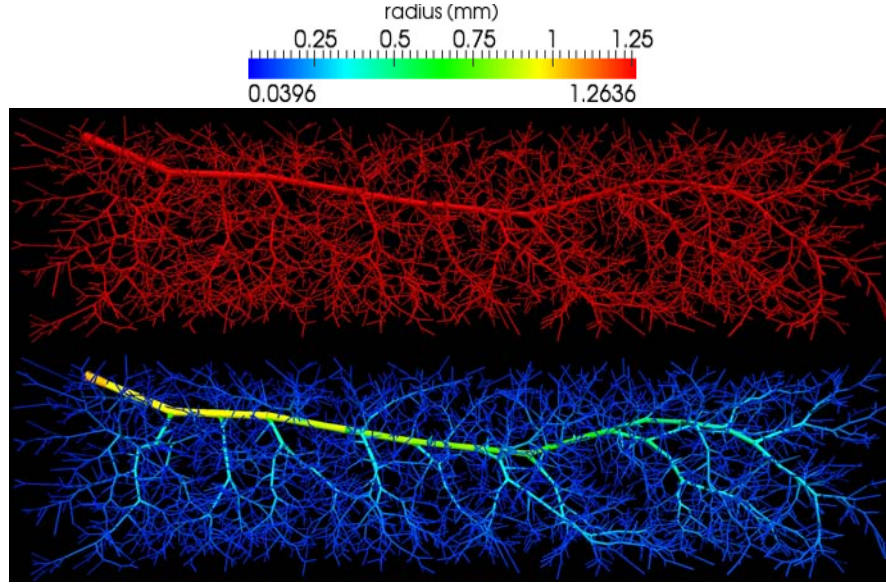


Fig. 2.21 CCO-generated arterial tree models in a parallelepiped with 4000 terminal segments.

2.4.4.3 Generation of the hepatic vascular system

The hepatic vascular system consists of three concurrent vascular networks: the arterial tree itself, the portal venous tree and the hepatic venous tree. Hepatic arteries and portal veins deliver blood and to hepatic cells, while the hepatic venous tree is responsible for blood transport from the liver back to the heart.

The algorithm behind the CCO method was modified in order to allow the generation of the entire hepatic vascular system, where the trees are assumed to be concurrent but at the same time individual CCO realizations. These CCO networks are connected together through of the capillary bed. The process of CCO multiple tree generation is as follows. Once the location \mathbf{x}_{term} randomly selected has been accepted as new terminal site in the arterial tree, it is connected to the optimum site within the three trees. The optimum site location is a concept which is local to each CCO network, so it is found in an independent manner within each tree. In addition, the local optimization procedure that perturbs the bifurcation point is carried out such that the lowest total intravascular volume of each CCO network is satisfied at each stage in the tree generation. In addition, the optimum site is chosen such that crossing among arteries (arterioles) and veins (venules) is avoided. Note that the three arterial networks are grown concurrently, that is, a set of three segments is added at each time, one for each tree.

Since liver features a non-convex shape, the intersection checkage between segment and surface volume perfusion discretized into triangle elements is performed when the location \mathbf{x}_{term} is temporarily connected to each of the neighboring seg-

ments and during the geometric optimization of the resulting bifurcation in each connection. If the intersection occurs then the candidate connection is dissolved.

As before, for the CCO simulation, the bifurcation law is set to $\gamma = 3$, the number of terminals is $N_{term} = 2000$ and the physiological parameters [38] are given in Table 2.9.

Model parameter	Hepatic arteries	Portal veins	Hepatic veins
Blood pressure at the input (mmHg)	95	25	12
Blood pressure at the output (mmHg)	20	15	5
Blood flow (ml/min)	400	1100	1500
Perfusion volume (cm ³)	1500		

Table 2.9 Physiological parameters used in the CCO simulation in order to generate the hepatic vascular system with 2000 terminal site.

In Figure 2.22 two different perspectives of the liver geometry obtained from an anatomical dataset is displayed. Also, the three locations for the different feeding arteries are provided.

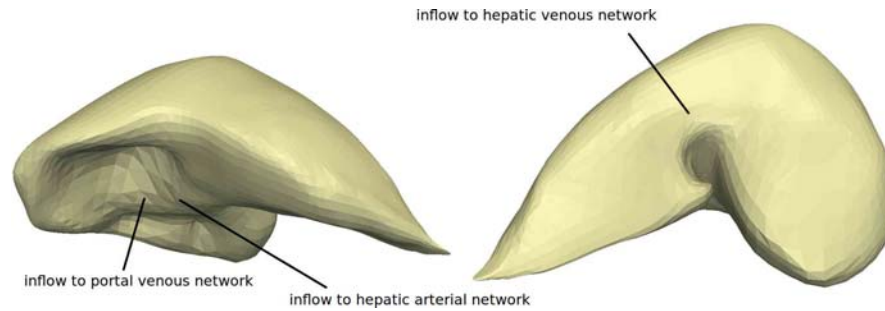


Fig. 2.22 Anatomical geometry of the liver and the three inflow locations for the concurrent vascular networks.

Figure 2.23 presents the hepatic arterial tree generated by the CCO together with a description of the lumen radii across the network.

In Figure 2.24 the generated networks are shown. Observe that the lumen radii in the vein case results much bigger than the arterial counterpart, as expected.

Finally, Figure 2.26 presents the generated vascular network for the portal venous system. Observe that the bigger lumen radii in the venous systems are because of the parameters presented in Table 2.9. Recall that the three networks are in contact with each other only at the capillary level, that is, at the very last terminal vessels.

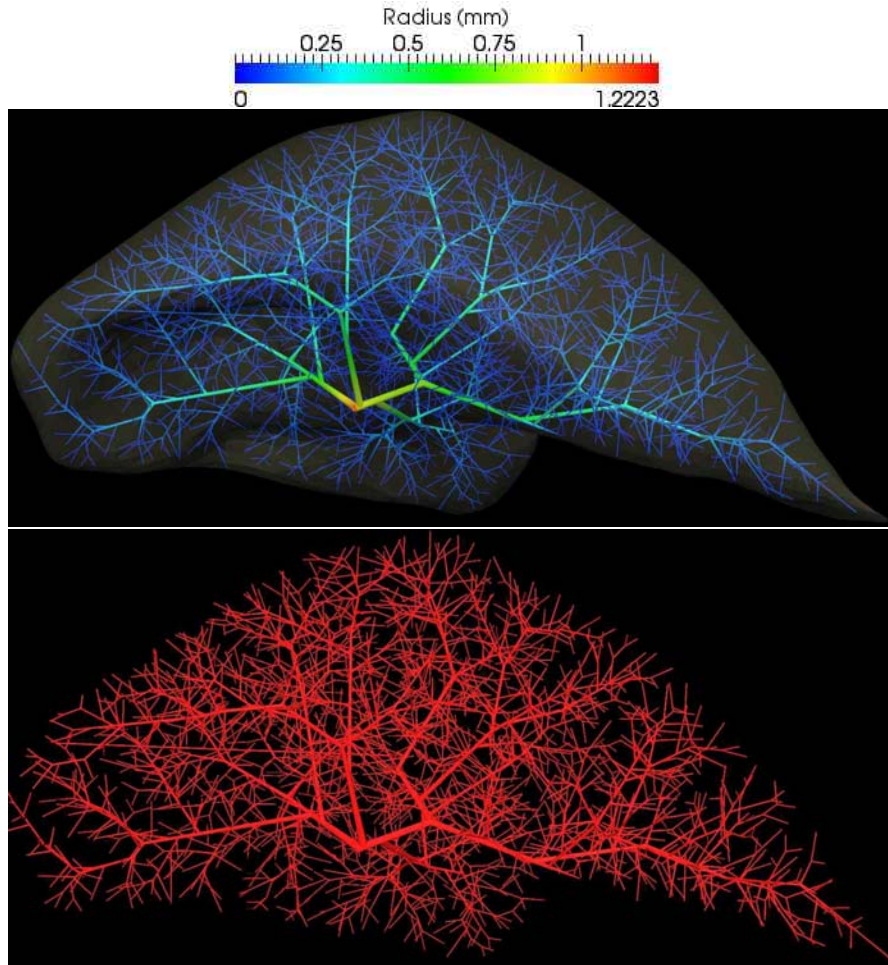


Fig. 2.23 CCO-generated hepatic arterial tree with 2000 terminal segments.

2.5 Final remarks

In the present work we have presented several developments that are characterized by being the state-of-the-art in terms of cardiovascular modeling. Specifically, the following topics have been addressed: (i) a closed-loop self-regulated model of the cardiovascular system, (ii) anatomically detailed models of the arterial network and (iii) the automatic construction of peripheral beds. In (i), the effect of the baroreflex control was analyzed in order to understand the importance of short term auto-regulation mechanisms in the conformation of the pressure pulse and blood flow distribution. In (ii) a model of the upper limb was developed following consistent and stringent physiological and anatomical considerations. This novel approach in

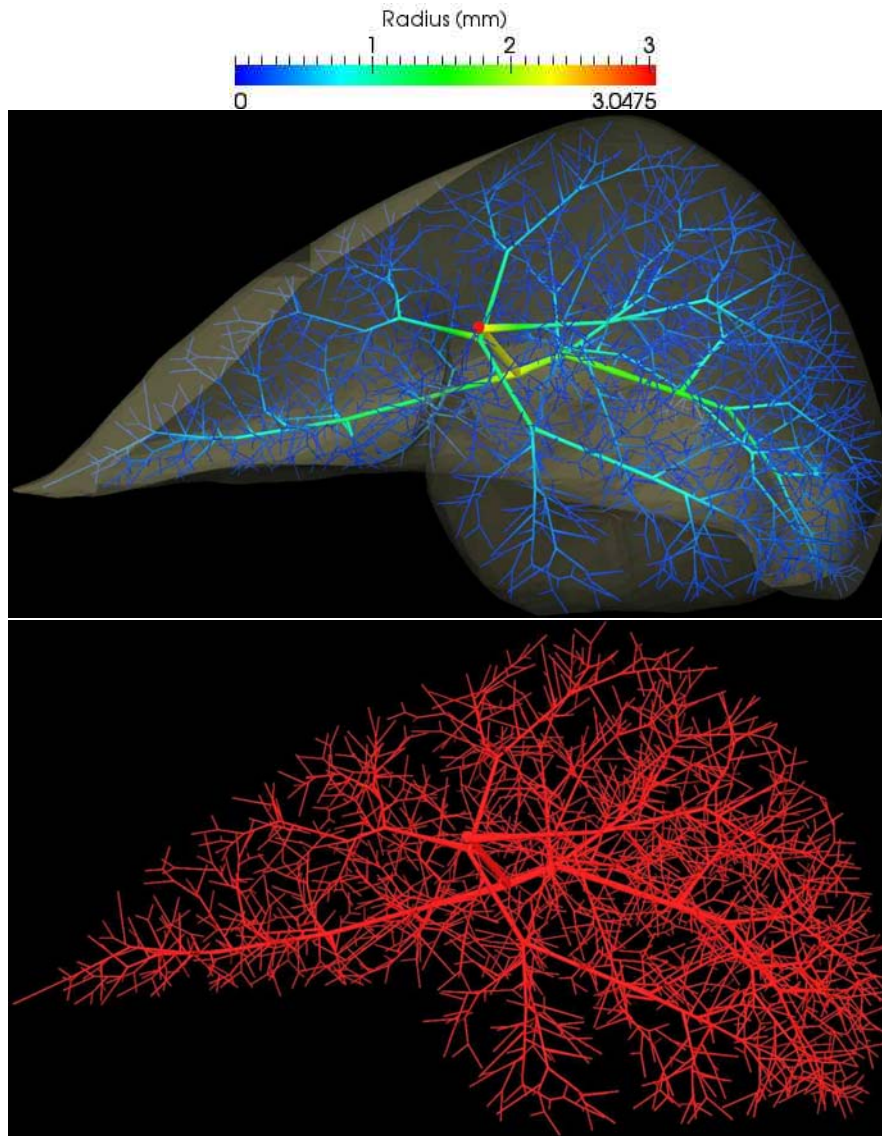


Fig. 2.24 CCO-generated hepatic venous tree with 2000 terminal segments.

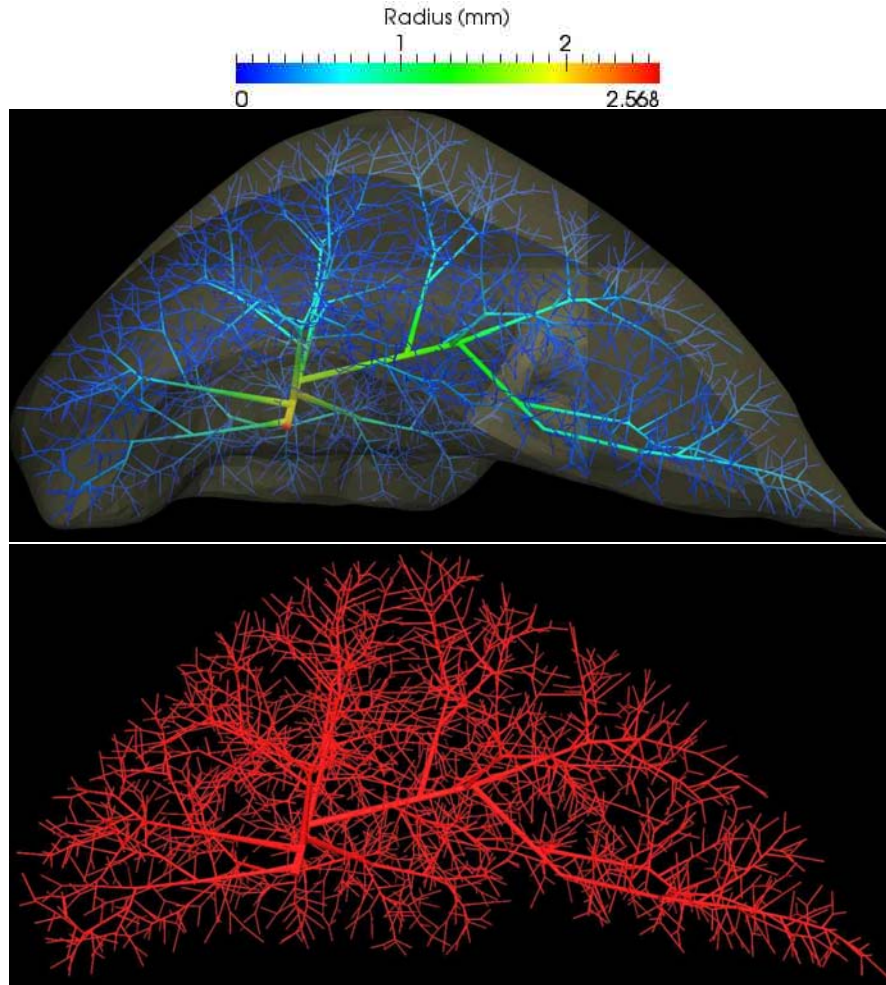


Fig. 2.25 CCO-generated portal venous tree with 2000 terminal segments.

setting cardiovascular networks is nowadays used to set up a highly detailed model of the entire cardiovascular system which will turn into the largest modeling hemodynamics *in-silico* laboratory worldwide. Finally, in (iii), automatic procedures to define the vascular network representing the peripheral beds have been developed and successfully implemented, aiming at providing the missing arterial segments, at the pre-arteriolar and arteriolar levels, in 1D models.

For each of these fields, examples of application ranging from academic to potentially-clinical have been simulated and discussed. The developments presented here, when combined altogether are intended to set up a comprehensive modeling framework to address widest range of cardiovascular scenarios observed in both

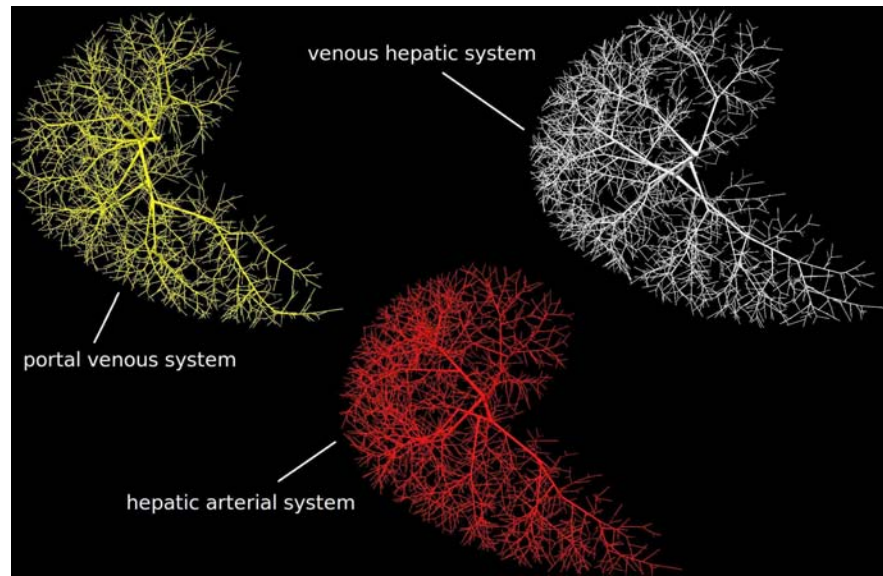


Fig. 2.26 Concurrent arterial networks with 2000 terminal segments each. The exploded visualization form helps to see the different networks.

physiological and pathophysiological conditions, the latter resulting from naturally/surgically altered conditions.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. The support of these agencies is gratefully acknowledged.

References

- [1] Alastruey, J., Parker, K., Peiró, J., Byrd, S., Sherwin, S.: Modelling the circle of Willis to assess the effects of anatomical variations and occlusions on cerebral flows. *J. Biomech.* **40**, 1794–1805 (2007).
- [2] Amundsen, B., Wisloff, U., Helgerud, J., Hoff, J., Slordahl, S.: Ultrasound recorded axillary artery blood flow during elbow-flexion exercise. *Med. Sci. Sports Exerc.* **34**, 1288–1293 (2002).
- [3] Arts, T., Kruger, R., van Gerven, W., Lambregts, J., Reneman, S.: Propagation velocity and reflection of pressure waves in the canine coronary artery. *Amer. J. Physiol.* **237**, H469–H474 (1979).
- [1] Avolio, A.: Multi-branched model of the human arterial system. *Med. Biol. Eng. Comp.* **18**, 709–718 (1980).

- [5] Bilge, O., Pinar, Y., Ozer, M., Gövsa, F.: A morphometric study on the superficial palmar arch of the hand. *Surg. Radiol. Anat.* **28**, 343–350 (2006).
- [3] Blanco, P., Feijóo, R., Urquiza, S.: A unified variational approach for coupling 3D–1D models and its blood flow applications. *Comp. Meth. Appl. Mech. Eng.* **196**, 4391–4410 (2007).
- [4] Blanco, P., Pivello, M., Urquiza, S., Feijóo, R.: On the potentialities of 3D–1D coupled models in hemodynamics simulations. *J. Biomech.* **42**, 919–930 (2009).
- [5] Blanco, P., Urquiza, S., Feijóo, R.: Assessing the influence of heart rate in local hemodynamics through coupled 3D–1D–0D models. *Int. J. Num. Meth. Biomed. Eng.* **26**, 890–903 (2010).
- [6] Blanco, P., Leiva, J., Feijóo, R., Buscaglia, G.: Black-box decomposition approach for computational hemodynamics: One-dimensional models. *Comp. Meth. Appl. Mech. Eng.* **200**, 1389–1405 (2011).
- [8] Blanco, P., Trenhago, P., Fernandes, L., Feijóo, R.: On the integration of the baroreflex control mechanism in a heterogeneous model of the cardiovascular system. *Int. J. Num. Meth. Biomed. Eng.* **28**, 412–433 (2012).
- [11] Casoli, V., Kostopoulos, E., Pélissier, P., Caix, P., Martin, D., Baudet, J.: The middle collateral artery: anatomic basis for the “extreme” lateral arm flap. *Surg. Radiol. Anat.* **26**, 172–177 (2004).
- [12] Chapleau, M., Abboud, F.: Neuro-cardiovascular regulation: From molecules to man. *Ann. New York Acad. Sci.* **940**, 13–22 (2001).
- [13] Chen, S., Xu, D., Tang, M., Ding, H., Sheng, W., Peng, T.: Measurement and analysis of the perforator arteries in upper extremity for the flap design. *Surg. Radiol. Anat.* **31**, 687–693 (2009).
- [14] Claassen, H., Schmitt, O., Werner, D., Schareck, W., Kröger, J., Wree, A.: Superficial arm arteries revisited: Brother and sister with absent radial pulse. *Ann. Anat.* **192**, 151–155 (2010).
- [15] Dauber, W.: *Pocket Atlas of Human Anatomy by Feneis*. Thieme, 5th edition (2007).
- [16] Dawant, B., Levin, M., Popel, A.: Effect of dispersion of vessel diameters and lengths in stochastic networks. I. Modeling of microcirculatory flow. *Microvasc. Res.* **31**, 203–222 (1986).
- [17] Duparc, F., Muller, J., Fréger, P.: Arterial blood supply of the proximal humeral epiphysis. *Surg. Radiol. Anat.* **23**, 185–190 (2001).
- [18] Fazan, V., Borges, C., Da Silva, J., Caetano, A., Filho, O.: Superficial palmar arch: an arterial diameter study. *J. Anat.* **204**, 307–311 (2004).
- [14] Formaggia, L., Nobile, F., Quarteroni, A., Veneziani, A.: Multiscale modelling of the circulatory system: a preliminary analysis. *Comp. Vis. Sci.* **2**, 75–83 (1999).
- [13] Formaggia, L., Gerbeau, J., Nobile, F., Quarteroni, A.: On the coupling of 3D and 1D Navier-Stokes equations for flow problems in compliant vessels. *Comp. Meth. Appl. Mech. Eng.* **191**, 561–582 (2001).
- [21] Formaggia, L., Moura, A., Nobile, F.: On the stability of the coupling of 3D and 1D fluid-structure interaction models for blood flow simulations. *ESIAM: M2AN* **41**, 743–769 (2007).
- [22] Fung, Y.: *Biomechanics: Circulation*. Springer-Verlag, New York (1984).
- [15] Grinberg, L., Anor, T., Madsen, J., Yakhot, A., Karniadakis, G.: Large-scale simulation of the human arterial tree. *Clinical and Experimental Pharmacology and Physiology* **36**, 194–205 (2009).
- [24] Guo, G., Thames, M., Abboud, F.: Differential baroreflex control of heart rate and vascular resistance in rabbits. Relative role of carotid, aortic, and cardiopulmonary baroreceptors. *Circ. Res.* **50**, 554–565 (1982).
- [25] Haerle, M., Tonagel, F., Schaller, H.: Collateral arterial pathways in the forearm. *Surg. Radiol. Anat.* **26**, 208–211 (2004).
- [26] Hoffman, J.: Maximal coronary flow and the concept of coronary vascular reserve. *Circulation* **70**, 153–159 (1984).
- [27] Holzbaur, K., Murray, W., Gold, G., Delp, S.: Upper limb muscle volumes in adult subjects. *J. Biomech.* **40**, 742–749 (2007).

- [13] Humphrey, J., Taylor, C.: Intracranial and abdominal aortic aneurysms: similarities, differences, and need for a new class of computational models. *Ann. Rev. Biomed. Eng.* **10**, 221–246 (2008).
- [14] Hunter, P., Nielsen, P.: A strategy for integrative computational physiology. *Physiol.* **20**, 316–325 (2005).
- [30] Kamiya, A., Togawa, T.: Optimal branching structure of the vascular tree. *Bull. Math. Biophys.* **34**, 431–438 (1972).
- [31] Karch, R., Neumann, F., Neumann, M., Schreiner, W.: A tree-dimensional model for arterial tree representation, generated by constrained constructive optimization. *Comput. Biol. Med.* **29**, 19–38 (1999).
- [20] Kim, H., Vignon-Clementel, I., Figueroa, C., LaDisa, J., Jansen, K., Feinstein, J., Taylor, C.: On coupling a lumped parameter heart model and a three-dimensional finite element aorta model. *Ann. Biomed. Eng.* **37**, 2153–2169 (2009).
- [33] Kim, H., Jansen, K., Taylor, C.: Incorporating autoregulatory mechanisms of the cardiovascular system in three-dimensional finite element models of arterial blood flow. *Ann. Biomed. Eng.* **38**, 2314–2330 (2010).
- [34] Kivity, Y., Collins, R.: Nonlinear wave propagation in viscoelastic tubes: application to aortic rupture. *J. Biomech.* **7**, 67–76 (1974).
- [35] Knobloch, K., Tomaszek, S., Lichtenberg, A., Karck, M., Haverich, A.: Long-term palmar microcirculation after radial artery harvesting: an observational study. *Ann. Thoracic Surg.* **81**, 1700–1707 (2006).
- [21] Korakianitis, T., Shi, Y.: Numerical simulation of cardiovascular dynamics with healthy and diseased heart valves. *J. Biomech.* **39**, 1964–1982 (2006).
- [37] Korner, P.: Integrative neural cardiovascular control. *Physiol. Rev.* **51**, 312–367 (1971).
- [38] Kretowski, M., Bezy-Wendling, J.: Computer modelling of vascular systems. *Task Quarterly* **8**, 223–229 (2004).
- [39] Kufahl, R., Clark, M.: A circle of willis simulation using distensible vessels and pulsatile flow. *J. Biomech. Eng.* **107**, 112–122 (1985).
- [40] LaBarbera, M.: Principles of design of fluid transport systems in zoology. *Science* **249**, 992–999 (1990).
- [41] Lanzarone, E., Liani, P., Baselli, G., Constantino, M.: Model of arterial tree and peripheral control for the study of physiological and assisted circulation. *Med. Eng. & Phys.* **29**, 542–555 (2007).
- [23] Leiva, J., Blanco, P., Buscaglia, G.: Partitioned analysis for dimensionally-heterogeneous hydraulic networks. *SIAM J. on Multiscale Model. Simul.* **9**, 872–903 (2011).
- [43] Leiva, J., Blanco, P., Buscaglia, G.: Iterative strong coupling of dimensionally-heterogeneous models. *Int. J. Num. Meth. Eng.* **81**, 1558–1580 (2010).
- [24] Liang, F., Takagi, S., Himeno, R., Liu, H.: Biomechanical characterization of ventricular-arterial coupling during aging: A multi-scale model study. *J. Biomech.* **42**, 692–704 (2009).
- [45] Liang, F., Liu, H.: A closed-loop lumped parameter computational model for human cardiovascular system. *JSME International Journal Series C* **48**, 484–493 (2005).
- [46] Mahabir, R., Williamson, J., Carr, N., Courtemanche, D.: Vascular resistance in human muscle flaps. *Ann. Plast. Surg.* **47**, 148–152 (2001).
- [47] Martin, B., Cirovic, S., Stergiopulos, N., Watanabe, M., Blanco, P.: Prediction of the impact of craniospinal compliance on the relative timing of arterial and cerebrospinal fluid pulsations and perivascular flow into the spinal cord. *ASME 2012 Summer Bioeng. Conf. Fajardo, Puerto Rico* (2012).
- [48] Mates, R., Klocke, F., Canty, J.: Coronary capacitance. *Prog Cardiovasc Dis.* **31**, 1–15 (1988).
- [49] Migliavacca, F., Balossino, R., Pennati, G., Dubini, G., Hsia, T., de Leval, M., Bove, E.: Multiscale modelling in biofluidynamics: Application to reconstructive paediatric cardiac surgery. *J. Biomech.* **39**, 1010–1020 (2006).
- [50] Moore, K., Dalley, A., Agur, A.: *Clinically Oriented Anatomy*. Wolters Kluwer, 6th edition (2010).

- [51] Morris, S., Tang, M., Almutari, K., Geddes, C., Yang, D.: The anatomic basis of perforator flaps. *Clin. Plast. Surg.* **37**, 553–570 (2010).
- [52] Murray, D.: The physiological principle of minimum work, I. The vascular system and the cost of blood volume. *Proc. Natl. Acad. Sci. USA* **12**, 207–214 (1926).
- [53] Netter, F.: *Atlas of Human Anatomy*. Elsevier, 5th edition (2011).
- [54] Olufsen, M.: Structured tree outflow condition for blood flow in larger systemic arteries. *Am. J. Physiol.* **276**, H257–H268 (1999).
- [55] Olufsen, M., Peskin, C., Kim, W., Pedersen, E., Nadim, A., Larsen, J.: Numerical simulation and experimental validation of blood flow in arteries with structured-tree outflow conditions. *Ann. Biomed. Eng.* **28**, 1281–1299 (2000).
- [56] Omokawa, S., Tanaka, Y., Ryu, J., Kish, V.: The anatomical basis for reverse first to fifth dorsal metacarpal arterial flaps. *J. Hand Surg.* **30**, 40–44 (2005).
- [57] Onuki, T., Nitta, S.: Computer simulation of geometry and hemodynamics of canine pulmonary arteries. *Ann. Biomed. Eng.* **21**, 107–115 (1993).
- [58] O'Rourke, M., Nichols, W.: *McDonald's Blood Flow in Arteries - Theoretical, Experimental and Clinical Principles*. Arnold, 4th edition (1998).
- [59] Ottesen, J., Olufsen, M.: Functionality of the baroreceptor nerves in heart rate regulation. *Comp. Meth. Prog. Biomed.* **101**, 208–219 (2011).
- [60] Ottesen, J., Olufsen, M., Larsen, J.: *Applied Mathematical Models in Human Physiology*. Society for Industrial and Applied Mathematics, Philadelphia (2004).
- [61] Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes in C, the Art of Scientific Computing*. Cambridge University Press (1992).
- [62] Prevel, C., Matloub, H., Ye, Z., Sanger, J., Yousif, N.: The extrinsic blood supply of the ulnar nerve at the elbow: an anatomic study. *J. Hand Surg.* **18**, 433–438 (1993).
- [63] Reichold, J., Stampanoni, M., Keller, A., Buck, A., Jenny, P., Weber, B.: Vascular graph model to simulate the cerebral blood flow in realistic vascular networks. *J. Cereb. Blood Flow Metab.* **29**, 1429–1443 (2009).
- [64] Reymond, P., Merenda, F., Perren, F., Rüfenacht, D., Stergiopoulos, N.: Validation of a one-dimensional model of the systemic arterial tree. *Am. J. Physiol. Heart Circ. Physiol.* **297**, H208–H222 (2009).
- [19] Ricotta, J., Pagan, J., Xenos, M., Alemum, Y., Einav, S., Bluestein, D.: Cardiovascular disease management: the need for better diagnostics. *Med. Biol. Eng. Comput.* **46**, 1059–1068 (2008).
- [66] Rodbard, S.: Vascular caliber. *Cardiol.* **60**, 4–49 (1975).
- [67] Royse, A., Chang, G., Nicholas, D., Royse, C.: No late ulnar artery atheroma after radial artery harvest for coronary artery bypass surgery. *Ann. Thoracic Surg.* **85**, 891–894 (2008).
- [68] Saito, M., Matsumoto, M.: *A PRNG Specialized in Double Precision Floating Point Numbers Using an Affine Transition. Monte Carlo and Quasi-Monte Carlo Methods*, Springer Berlin Heidelberg (2009).
- [69] Sauerbier, M., Unglaub, F.: Perforator flaps in the upper extremity. *Clin. Plast. Surg.* **37**, 667–676 (2010).
- [70] Schaaf, B., Abbrecht, P.: Digital computer simulation of systemic arterial pulse wave transmission: a nonlinear model. *J. Biomech. Eng.* **5**, 345–364 (1972).
- [71] Schafhalterzoppoth, I., Gray, A.: The musculocutaneous nerve: ultrasound appearance for peripheral nerve block. *Reg. Anesth. Pain Med.* **30**, 385–390 (2005).
- [72] Schreiner, W.: Computer generation of complex arterial tree models. *J. Biomed. Eng.* **15**, 148–149 (1993).
- [73] Schreiner, W., Buxbaum, P.: Computer-optimization of vascular trees. *IEEE Trans. Biomed. Eng.* **40**, 482–491 (1993).
- [74] Schreiner, W., Neumann, M., Neumann, F., Roedler, S., End, A., Buxbaum, P., Muller, M., Spieckermann, P.: The branching angles in computer-generated optimized models of arterial trees. *J. Gen. Physiol.* **103**, 975–989 (1994).

- [75] Schreiner, W., Neumann, F., Neumann, M., End, A., Roedler, S., Aharinejad, S.: The influence of optimization target selection on the structure of arterial tree models generated by constrained constructive optimization. *J. Gen. Physiol.* **106**, 583–599 (1995).
- [76] Schreiner, W., Neumann, F., Neumann, M., End, A., Müller, M.: Structural quantification and bifurcation symmetry in arterial tree models generated by constrained constructive optimization. *J. Theor. Biol.* **180**, 161–174 (1996).
- [77] Schreiner, W., Neumann, F., Neumann, M., Karch, R., End, A., Roedler, S.: Limited bifurcation asymmetry in coronary arterial tree models generated by constrained constructive optimization. *J. Gen. Physiol.* **109**, 129–140 (1997).
- [78] Schreiner, W., Neumann, F., Karch, R., Neumann, M., Roedler, S., End, A.: Shear stress distribution in arterial tree models generated by constrained constructive optimization. *J. Theor. Biol.* **198**, 27–45 (1999).
- [79] Schreiner, W., Karch, R., Neumann, M., Neumann, F., Roedler, S., Heinze, G.: Heterogeneous perfusion is a consequence of uniform shear stress in optimized arterial tree models. *J. Theor. Biol.* **220**, 285–301 (2003).
- [80] Schreiner, W., Karch, R., Neumann, M., Neumann, F., Szawlowski, P., Roedler, S.: Optimized arterial trees supplying hollow organs. *Med. Eng. & Phys.* **28**, 416–429 (2006).
- [81] Sherman, T.: On connecting large vessels to small. The meaning of Murray's law. *J. Gen. Physiol.* **78**, 431–453 (1981).
- [82] Spencer, M., Deninson, A.: The square-wave electro-magnetic flowmeter. Theory of operation and design of magnetic probes for clinical and experimental applications. *I.R.E. Trans. Med. Elect.* **6**, 220–228 (1959).
- [29] Stergiopoulos, N., Young, D., Rogge, T.: Computer simulation of arterial flow with applications to arterial and aortic stenoses. *J. Biomech.* **25**, 1477–1488 (1992).
- [84] Stettler, J., Niederer, P., Anliker, M.: Theoretical analysis of arterial hemodynamics including the influence of bifurcations, part I. *Ann. Biomed. Eng.* **9**, 145–164 (1981).
- [85] Sykora, M., Diedler, J., Turcani, P., Hacke, W., Steiner, T.: Baroreflex: A new therapeutic target in human stroke? *Stroke* **40**, e678–e682 (2009).
- [86] Taylor, G.: The angiosomes of the body and their supply to perforator flaps. *Clin. Plastic Surg.*, **30**, 331–342 (2003).
- [87] TenVoorde, B., Kingma, R.: A baroreflex model of short term blood pressure and heart rate variability. IOS Press, Amsterdam, 2000.
- [88] Thompson, D.: *On Growth and Form*, Cambridge University Press, Cambridge (1942).
- [89] Trager, S., Pignataro, M., Anderson, J., Kleinert, J.: Color flow Doppler: imaging the upper extremity. *J. Hand Surg.* **18**, 621–625 (1993).
- [31] Urquiza, S., Blanco, P., Vénere, M., Feijóo, R.: Multidimensional modelling for the carotid artery blood flow. *Comp. Meth. App. Mech. Eng.* **195**, 4002–4017 (2006).
- [91] Ursino, M.: *Mathematical Modelling in Medicine*, IOS Press (2000).
- [92] Ursino, M., Antonucci, M., Belardinelli, E.: The role of active changes in venous capacity by carotid baroreflex: analysis with a mathematical model. *Heart Circ. Physiol.* **36**, H2531–H2546 (1994).
- [93] Ursino, M., Giannessi, M.: A model of cerebrovascular reactivity including the circle of Willis and cortical anastomoses. *Ann. Biomed. Eng.* **38**, 955–974 (2010).
- [94] VanBavel, E., Spaan, J.: Branching patterns in the porcine coronary arterial tree. Estimation of flow heterogeneity. *Circ. Res.* **71**, 1200–1212 (1992).
- [95] Van Beek, J., Roger, S., Bassingthwaite, J.: Regional myocardial flow heterogeneity explained with fractal networks. *Am. J. Physiol.* **257**, H1670–H1680 (1989).
- [32] Vignon-Clementel, I., Figueiroa, C., Jansen, K., Taylor, C.: Outflow boundary conditions for three-dimensional finite element modeling of blood flow and pressure waves in arteries. *Comp. Meth. Appl. Mech. Eng.* **195**, 3776–3996 (2006).
- [97] Walther, G., Nottin, S., Dauzat, M., Obert, P.: Femoral and axillary ultrasound blood flow during exercise: a methodological study. *Med. Sci. Sports Exerc.* **38**, 1353–1361 (2006).
- [33] Wang, J., Parker, K.: Wave propagation in a model of the arterial circulation. *J. Biomech.* **37**, 457–470 (2004).

- [99] Wavreille, G., Bricout, J., Mouliade, S., Lemoine, S., Prodhomme, G., Khanchandani, P., Chantelot, C., Fontaine, C.: Anatomical bases of the free posterior brachial fascial flap. *Surg. Radiol. Anat.* **32**, 393–399 (2010).
- [100] Wavreille, G., Dos Remedios, C., Chantelot, C., Limousin, M., Fontaine, C.: Anatomic bases of vascularized elbow joint harvesting to achieve vascularized allograft. *Surg. Radiol. Anat.* **28**, 498–510 (2006).
- [101] Xiuying, M., Abboud, F., Chapleau, M.: Neurocardiovascular regulation in mice: experimental approaches and novel findings. *Clin. Exp. Pharmac. Physiol.* **30**, 885–893 (2003).
- [102] Zamir, M., Chee, H.: Segment analysis of human coronary arteries. *Blood Vessels* **24**, 76–84 (1987).
- [103] Zamir, M.: Distributing and delivering vessels of the human heart. *J. Gen. Physiol.* **91**, 725–735 (1988).
- [104] Zhang, F., Topp, S., Zhang, W., Zheng, H., Zhang, F.: Anatomic study of distally based pedicle compound flaps with nutrient vessels of the cutaneous nerves and superficial veins of the forearm. *Microsurgery* **26**, 373–385 (2006).

Chapter 3

Implementation issues of large strain formulations of hyperelastic materials for the modeling of arterial wall mechanics

Santiago A. Urquiza, Pablo J. Blanco, Gonzalo D. Ares and Raúl A. Feijóo

Abstract In this work we describe a generic framework to evaluate the mechanical equilibrium of hyperelastic materials in order to model arterial wall mechanics. This is carried out using a spatial variational formulation for the momentum equation which is linearized following a Newton-Raphson scheme. The proposed iterative algorithm is such that for a given load, the equilibrium is reached in the deformed spatial configuration. The main feature of the present approach is based on the evaluation of the second order stress tensor and of the fourth order constitutive tangent tensor using finite differences. That is, given a strain energy potential, we compute, by means of a second order centered finite difference scheme, the first (stress) and second (tangent matrix) derivatives. In this way, a generic computational implementation in the context of the finite element method is achieved, making possible to change the material behavior just changing the procedure that evaluates the elastic function and without reprogramming the entire numerical element structure. The developments are performed for compressible and incompressible materials, and some implementation issues are presented and discussed. The method is validated and then used for computing the mechanical response of fiber-reinforced materials in the context of the analysis of arterial tissues.

Santiago Adrián Urquiza

International Colaborator of the INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing and Mechanical Engineering Department, Mar del Plata National University, Av. Juan B. Justo 4302, 7600, Mar del Plata, Argentina, e-mail: `santiago.urquiza@fi.mdp.edu.ar`

Pablo Javier Blanco

Gonzalo Damián Ares

Raúl Antonino Feijóo

HeMoLab Hemodynamics Modeling Laboratory, INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing and LNCC - National Laboratory for Scientific Computing, Av. Getúlio Vargas 333, 25651-075, Petrópolis, Brazil, e-mail: `\{pjblanco, gonzalo, feij\}@lncc.br`

3.1 Introduction

The accurate representation of the arterial dynamics considering the blood flow and the arterial wall mechanical response has become increasingly important in the context of computational simulation of hemodynamic problems [5, 15, 13, 14, 15, 17, 19, 20]. Consequently, there has been growing scientific interest in the development of realistic constitutive equations representing the mechanical behavior of arterial wall tissues. Such interest is based on the fact that the constitutive response of arteries and the resulting stress state are related to the onset and development of certain pathologies, such as atherosclerosis and aneurysms. Moreover, the need for complex equations representing the arterial stress-strain relation arises naturally when modeling efforts are directed towards surgical planning of vessel repairs such as bypass grafts and also in stents placement.

Modeling arterial blood flows implies that the fluid-structure interaction (FSI) problem between the arterial walls and the blood inside them must be taken into account. This FSI problem poses some challenging issues to the corresponding computational models due to its nonlinear characteristics coming from the constitutive laws of the arterial wall, the varying geometry, the blood behavior and also due to the coupling itself between the solid and the fluid components of a given arterial segment. The authors have been working from the last decade in coupled hetero-dimensional problems discretized via the Finite Element Method (FEM) and aimed to model the arterial blood flow in compliant vessels [1, 4, 3, 31]. Up today, very simple wall models were considered (consistent with the 1D model -independent rings models-). Nevertheless, more realistic arterial wall models are needed to be implemented if we are, for example, dealing with models of aneurysms onset, development and rupture, in which assessing the stress state inside the arterial wall thickness is crucial.

When implementing a FEM computer code that takes into account appropriate constitutive equations of the arterial tissues, the developer is faced to the fact that there is a proliferation of available constitutive laws representing the material behavior [4, 9, 7, 8, 10, 11, 12, 18, 22, 23, 24]. In this sense, it is difficult to determine a priori, which ones are the best suited for computational models of hemodynamic problems. Observe that, it is mistimed to define in advance which equations should be implemented in the program code, as this would restrict the freedom of choice of the constitutive law which best suits the actual type of the problem to be solved. Furthermore, if the specific material relations are directly encoded at the *element* level calculations, the complete code associated to that *element* must be re-coded each time the constitutive law is changed. Therefore, the key aspect in order to achieve a generic code, is to split the geometric non-linearities from the constitutive non-linearities and to encapsulate the constitutive laws in specific functions embodying the material response such that they can be easily changed.

Thus, in this work, the main conceptual aspects of a FEM implementation for the mechanical simulation of generic hyperelastic materials are presented. This FEM realization is capable of solving the mechanics of solid bodies in large deformations and displacements regimes, and is coded in such a way that changes in the consti-

tutive equation do not affect the entire implementation. This is achieved in a very simple manner through the encapsulation of the constitutive relationships, characterized by the corresponding strain energy functions, in certain functions used to calculate the tangent matrix components. This brings about a code that separates the constitutive issues from the geometric ones at the same time that encapsulates and simplifies the software coding of material behavior laws. In the hyperelastic case treated here, the stresses and tangential component of the elastic tensor are obtained by a second order finite differences scheme. So, it is possible to incorporate new materials only by programming a scalar function, i.e., the hyperelastic strain energy potential. Consequently, this impacts on the simplicity of the computational implementation, facilitating and accelerating the prototyping of new constitutive laws, which can now be straightforwardly incorporated to the existing software. Additionally, these ideas become more relevant when we realize that the arterial wall is a multilayer (collagen) fiber reinforced material and that the properties vary from district to district of the arterial trees and also that models require different levels of details at different locations.

This work is organized as follows. Section 3.2 presents a review of the main ingredients from continuum mechanics which will be used throughout this work. The variational formulation of the mechanical equilibrium and the consistent Newton linearization are presented in Section 3.3. In Section 3.4 the finite-difference approximation for the calculation of the derivatives of the strain energy function is discussed, and in Section 3.5 the finite element method and additional numerical issues are addressed. The resulting iterative algorithm to reach the mechanical equilibrium is presented in Section 3.5.4. Specific numerical examples and applications are treated in Section 5.6. Finally, in Section 3.7 the final remarks are outlined.

3.2 Continuum mechanics revisited

Let \mathcal{B} be a body which for which we identify in the Euclidian space \mathcal{E} a reference, or material, configuration denoted by Ω_m , and with coordinates \mathbf{x}_m . The same body at equilibrium (see definition of this concept in the following section) occupies the current, or spatial, configuration denoted by Ω_s , and with coordinates \mathbf{x}_s . Hereafter, indexes m and s imply that the description is given in terms of the material coordinates \mathbf{x}_m and of the spatial coordinates \mathbf{x}_s , respectively. Figure 3.1 provides a schematic picture.

The mapping from the current configuration to the material one is given by the function $\phi : \Omega_m \rightarrow \Omega_s, \mathbf{x}_m \mapsto \mathbf{x}_s = \phi(\mathbf{x}_m)$ in such a way that it is explicitly given by $\mathbf{x}_s = \mathbf{x}_m + \mathbf{u}_m(\mathbf{x}_m)$, where $\mathbf{u}_m(\mathbf{x}_m)$ is the displacement field defined in the material configuration. In this manner, we introduce the deformation gradient \mathbf{F}_m that is

$$\mathbf{F}_m = \nabla_{\mathbf{x}_m} \mathbf{x}_s = \mathbf{I} + \nabla_{\mathbf{x}_m} \mathbf{u}_m. \quad (3.1)$$

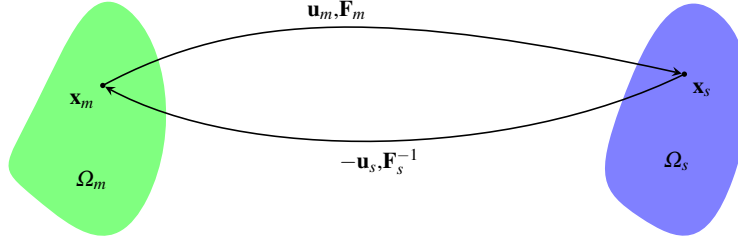


Fig. 3.1 Spatial and material configurations for the body \mathcal{B} .

Analogously, denoting by $\mathbf{u}_s(\mathbf{x}_s)$ the displacement field in the spatial configuration we have that $\mathbf{x}_m = \mathbf{x}_s - \mathbf{u}_s(\mathbf{x}_s)$, and then we have the inverse of the deformation gradient as

$$\mathbf{F}_s^{-1} = \nabla_{\mathbf{x}_s} \mathbf{x}_m = \mathbf{I} - \nabla_{\mathbf{x}_s} \mathbf{u}_s. \quad (3.2)$$

Therefore, it is possible to introduce some deformation measures in the material configuration such as the right Cauchy-Green deformation tensor $\mathbf{C}_m = \mathbf{F}_m^T \mathbf{F}_m$ and the Green-Lagrange deformation tensor

$$\begin{aligned} \mathbf{E}_m &= \frac{1}{2}(\mathbf{C}_m - \mathbf{I}) = \frac{1}{2}(\mathbf{F}_m^T \mathbf{F}_m - \mathbf{I}) = \\ &= \frac{1}{2}(\nabla_{\mathbf{x}_m} \mathbf{u}_m + (\nabla_{\mathbf{x}_m} \mathbf{u}_m)^T + (\nabla_{\mathbf{x}_m} \mathbf{u}_m)^T \nabla_{\mathbf{x}_m} \mathbf{u}_m). \end{aligned} \quad (3.3)$$

Analogously, we introduce in the spatial configuration the symmetric deformation gradient

$$\boldsymbol{\varepsilon}_s(\mathbf{u}_s) = \frac{1}{2}(\nabla_{\mathbf{x}_s} \mathbf{u}_s + (\nabla_{\mathbf{x}_s} \mathbf{u}_s)^T). \quad (3.4)$$

Also, we have the following relation

$$(\nabla_{\mathbf{x}_s} \mathbf{u}_s)_m = (\nabla_{\mathbf{x}_m} \mathbf{u}_m) \mathbf{F}_m^{-1}. \quad (3.5)$$

Regarding the description of stress states we introduce the Cauchy stress tensor denoted by $\boldsymbol{\sigma}_s$, and from here we define the first and second Piola-Kirchhoff, denoted by \mathbf{P}_m and \mathbf{S}_m , respectively, as those tensors that satisfy

$$\boldsymbol{\sigma}_m = \frac{1}{\det \mathbf{F}_m} \mathbf{P}_m \mathbf{F}_m^T \quad (3.6)$$

and

$$\boldsymbol{\sigma}_m = \frac{1}{\det \mathbf{F}_m} \mathbf{F}_m \mathbf{S}_m \mathbf{F}_m^T, \quad (3.7)$$

respectively, from which it is $\mathbf{P}_m = \mathbf{F}_m \mathbf{S}_m$.

As far as variational principles are concerned, the internal virtual power in solid mechanics is the result of some of the following duality pairings $(\boldsymbol{\sigma}_s, \hat{\boldsymbol{\varepsilon}}_s)$, $(\mathbf{S}_m, \hat{\mathbf{E}}_m)$ or $(\mathbf{P}_m, \hat{\mathbf{F}}_m)$ among others, as will be seen in the following section. Here, the hat represents an admissible action performed over the body in such a way that $\hat{\mathbf{u}}_m = \hat{\mathbf{v}}_m$ and $\hat{\mathbf{u}}_s = \hat{\mathbf{v}}_s$ represent virtual velocities in the material and spatial configurations. Here, it is clear that $\hat{\boldsymbol{\varepsilon}}_s$, $\hat{\mathbf{E}}_m$ and $\hat{\mathbf{F}}_m$ are the corresponding strain rate tensors. Finally, from the expressions above these strain rate tensors are

$$\hat{\boldsymbol{\varepsilon}}_s(\hat{\mathbf{v}}_s) = \frac{1}{2}(\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s + (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)^T), \quad (3.8)$$

$$\hat{\mathbf{E}}_m = \frac{1}{2}((\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_m + \mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)) \quad (3.9)$$

and

$$\hat{\mathbf{F}}_m = \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m. \quad (3.10)$$

3.3 Equilibrium and linearization

3.3.1 Spatial and material equilibrium expressions

In solid mechanics, the equilibrium is defined by the following variational problem: find $\mathbf{u}_s \in \mathcal{U}_s$ such that the stress state $\boldsymbol{\sigma}_s$ is such that

$$\int_{\Omega_s} \boldsymbol{\sigma}_s \cdot \hat{\boldsymbol{\varepsilon}}_s \, d\Omega_s = \int_{\Omega_s} \rho_s \mathbf{f}_s \cdot \hat{\mathbf{v}}_s \, d\Omega_s + \int_{\Gamma_{N_s}} \mathbf{g}_s \cdot \hat{\mathbf{v}}_s \, d\Gamma_s \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.11)$$

where ρ_s is the density, Γ_{N_s} is the Neumann surface in its spatial configuration whose coordinates are $\mathbf{y}_s = \mathbf{x}_s|_{\Gamma_{N_s}}$. Here, \mathbf{f}_s is a volume loading defined in the current configuration Ω_s , while \mathbf{g}_s is a surface loading acting on the boundary Γ_{N_s} . The linear manifold \mathcal{U}_s and its associated linear space $\hat{\mathcal{V}}_s$ are

$$\begin{aligned} \mathcal{U}_s &= \{\mathbf{u}_s \in \mathbf{H}^1(\Omega_s); \mathbf{u}_s = \bar{\mathbf{u}}_s \text{ on } \Gamma_{D_s}\}, \\ \hat{\mathcal{V}}_s &= \{\hat{\mathbf{v}}_s \in \mathbf{H}^1(\Omega_s); \hat{\mathbf{v}}_s = 0 \text{ on } \Gamma_{D_s}\}, \end{aligned} \quad (3.12)$$

where Γ_{D_s} is the part of the boundary where essential conditions are imposed.

Let us obtain the same equilibrium equation now in the material configuration. Then, equation (3.11) becomes

$$\begin{aligned}
\int_{\Omega_m} \boldsymbol{\sigma}_m \cdot \frac{1}{2} ((\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) \mathbf{F}_m^{-1} + \mathbf{F}_m^{-T} (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T) \det \mathbf{F}_m \, d\Omega_m = \\
\int_{\Omega_m} \rho_m \mathbf{f}_m \cdot \hat{\mathbf{v}}_m \, d\Omega_m + \int_{\Gamma_{Nm}} \mathbf{g}_m \cdot \hat{\mathbf{v}}_m \det \mathbf{F}_m |\mathbf{F}_m^{-T} \mathbf{n}_m| \, d\Gamma_{Nm} \\
\forall \hat{\mathbf{v}}_m \in \hat{\mathcal{V}}_m, \quad (3.13)
\end{aligned}$$

where ρ_m is the density in the reference configuration, and Γ_{Nm} is the surface in the material configuration with coordinates \mathbf{Y}_m and whose outward normal vector in the material configuration is \mathbf{n}_m .

Making use of the symmetry of $\boldsymbol{\sigma}_m$ and introducing the second Piola-Kirchhoff stress tensor and the Green-Lagrange rate strain tensor we have that the equilibrium in the material configuration is given by: find $\mathbf{u}_m \in \mathcal{U}_m$ such that the stress state \mathbf{S}_m is such that

$$\begin{aligned}
\int_{\Omega_m} \mathbf{S}_m \cdot \hat{\mathbf{E}}_m \, d\Omega_m = \int_{\Omega_m} \rho_m \mathbf{f}_m \cdot \hat{\mathbf{v}}_m \, d\Omega_m + \int_{\Gamma_{Nm}} \mathbf{g}_m \cdot \hat{\mathbf{v}}_m \det \mathbf{F}_m |\mathbf{F}_m^{-T} \mathbf{n}_m| \, d\Gamma_{Nm} \\
\forall \hat{\mathbf{v}}_m \in \hat{\mathcal{V}}_m. \quad (3.14)
\end{aligned}$$

Now, the linear manifold and the associated linear space in the material configurations are

$$\begin{aligned}
\mathcal{U}_m &= \{\mathbf{u}_m \in \mathbf{H}^1(\Omega_m); \mathbf{u}_m = \bar{\mathbf{u}}_m \text{ on } \Gamma_{Dm}\}, \\
\hat{\mathcal{V}}_m &= \{\hat{\mathbf{v}}_m \in \mathbf{H}^1(\Omega_m); \hat{\mathbf{v}}_m = 0 \text{ on } \Gamma_{Dm}\}.
\end{aligned} \quad (3.15)$$

From (3.11) and (3.14) we can observe the duality between $\boldsymbol{\sigma}_s$ and $\hat{\mathbf{e}}_s$ and between \mathbf{S}_m and $\hat{\mathbf{E}}_m$, as commented before. Also, straightforward manipulation of the left hand side integrand yields $\mathbf{S}_m \cdot \hat{\mathbf{E}}_m = \mathbf{P}_m \cdot \hat{\mathbf{F}}_m$.

When considering incompressible materials, the incompressibility constraint reads in the material description $\det \mathbf{F}_m = 1$, and may be incorporated in the very definition of \mathcal{U}_m and $\hat{\mathcal{V}}_m$. The spatial counterpart reads $1 = \det \mathbf{F}_s^{-1}$, or equivalently $\det \mathbf{F}_s = 1$, and it may be incorporated in the definition of \mathcal{U}_s and $\hat{\mathcal{V}}_s$ as well. Nevertheless, this constraint is usually relaxed through a Lagrange multiplier.

Both variational formulations (3.11) and (3.14) are completely equivalent at the continuum level. When introducing approximations of the functional spaces the two formulations differ. In the present work we assess the mechanical equilibrium by means of the linearized version of the spatial formulation (3.11).

3.3.2 Linearization procedure

For the linearization we make use of the Newton-Raphson method. The linearization of variational formulation (3.11) is carried out by writing it in the material configuration, linearizing there and going back to the spatial configuration. For the simplicity

in the presentation we disregard the nonlinearities arising from the nature of the loadings \mathbf{f}_s and \mathbf{g}_s .

Notice that the variational formulation (3.11) can be written in abstract form as: find $\mathbf{u}_s \in \mathcal{U}_s$ such that

$$\langle \mathcal{R}_{\Omega_s}(\mathbf{u}_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s} = 0 \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.16)$$

where in the operator \mathcal{R}_{Ω_s} the constitutive response and the loading are taken into account. Also this operator depends on the spatial coordinates \mathbf{x}_s , and that is why we use the index Ω_s to introduce this dependence. The Newton-Raphson linearization applied to this expression reads as next: find $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ such that

$$\left. \langle \mathcal{R}_{\Omega_s}(\mathbf{u}_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s} + \frac{d}{d\tau} \langle \mathcal{R}_{\Omega_s + \tau \delta \Omega_s}(\mathbf{u}_s + \tau \delta \mathbf{u}_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s + \tau \delta \Omega_s} \right|_{\tau=0} = 0 \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.17)$$

where $\delta \Omega_s$ denotes the variation of the domain due to the displacement $\delta \mathbf{u}_s$, and $\delta \mathcal{U}_s$ is the incremental affine space in which essential boundary conditions are considered in an incremental manner. Therefore, the linear problem consists in finding $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ such that

$$\langle \mathcal{D}_{\Omega_s}(\mathbf{u}_s) \delta \mathbf{u}_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = - \langle \mathcal{R}_{\Omega_s}(\mathbf{u}_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s} \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.18)$$

where $\mathcal{D}_{\Omega_s}(\mathbf{u}_s)$ is the tangent operator. In the most general situation this operator accounts for constitutive, geometrical and loads nonlinearities.

The second term in the left hand side of (3.17) can be written in the material configuration changing the index s by m (noting that Ω_m now is fixed) as next

$$\left. \frac{d}{d\tau} \langle \mathcal{R}_{\Omega_s + \tau \delta \Omega_s}(\mathbf{u}_s + \tau \delta \mathbf{u}_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s + \tau \delta \Omega_s} \right|_{\tau=0} = \left. \frac{d}{d\tau} \langle \mathcal{R}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \right|_{\tau=0}. \quad (3.19)$$

Note that

$$\mathbf{F}_{m,\tau} = \mathbf{I} + \nabla_{\mathbf{x}_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m) = \mathbf{F}_m + \tau \nabla_{\mathbf{x}_m} \delta \mathbf{u}_m, \quad (3.20)$$

such that it produces a perturbation in the Green-Lagrange tensor which is given by

$$\begin{aligned} \mathbf{E}_{m,\tau} &= \frac{1}{2} (\mathbf{F}_{m,\tau}^T \mathbf{F}_{m,\tau} - \mathbf{I}) = \\ &= \mathbf{E}_m + \frac{\tau}{2} ((\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m)^T \mathbf{F}_m + \mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m)) + \frac{\tau^2}{2} (\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m)^T (\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m). \end{aligned} \quad (3.21)$$

Then, the right hand side of equation (3.19) is the following

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{R}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \\ \frac{d}{d\tau} \left[\int_{\Omega_m} \mathbf{S}_m(\mathbf{E}_m, \tau) \cdot \frac{1}{2} (\mathbf{F}_{m,\tau}^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_{m,\tau}) d\Omega_m \right] \Big|_{\tau=0}, \end{aligned} \quad (3.22)$$

where we put in evidence the dependence of the Piola-Kirchhoff tensor with respect to the Green-Lagrange tensor, that is $\mathbf{S}_m = \mathbf{S}_m(\mathbf{E}_m)$. With the help of some classical relations from continuum mechanics (see the Appendix) we get from (3.22) the following

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{R}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \\ \int_{\Omega_m} \left[\frac{\partial \mathbf{S}_m}{\partial \mathbf{E}_m} \mathbf{F}_m^T \boldsymbol{\varepsilon}_m(\delta \mathbf{u}_s) \mathbf{F}_m \cdot \frac{1}{2} (\mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_m) \right. \\ \left. + \mathbf{S}_m \cdot \frac{1}{2} (\mathbf{F}_m^T (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m \mathbf{F}_m) \right] d\Omega_m. \end{aligned} \quad (3.23)$$

Going back to the spatial configuration using the fact that $\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m = (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)_m \mathbf{F}_m$, expression (3.23) results in

$$\begin{aligned} \langle \mathcal{D}_{\Omega_s}(\mathbf{u}_s) \delta \mathbf{u}_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = \int_{\Omega_s} \left[\frac{1}{\det \mathbf{F}_s} \mathbf{F}_s \left(\frac{\partial \mathbf{S}_m}{\partial \mathbf{E}_m} \right)_s \mathbf{F}_s^T \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \mathbf{F}_s \mathbf{F}_s^T \cdot \hat{\boldsymbol{\varepsilon}}_s(\hat{\mathbf{v}}_s) \right. \\ \left. + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \boldsymbol{\sigma}_s \cdot (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s) \right] d\Omega_s. \end{aligned} \quad (3.24)$$

Using the notation

$$\mathbf{D}_s \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) = \frac{1}{\det \mathbf{F}_s} \mathbf{F}_s \left(\frac{\partial \mathbf{S}_m}{\partial \mathbf{E}_m} \right)_s \mathbf{F}_s^T \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \mathbf{F}_s \mathbf{F}_s^T, \quad (3.25)$$

we formulate the linear problem as follows: find $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ such that

$$\begin{aligned} \int_{\Omega_s} [\mathbf{D}_s \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \cdot \hat{\boldsymbol{\varepsilon}}_s + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \boldsymbol{\sigma}_s \cdot (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)] d\Omega_s = \\ - \int_{\Omega_s} \boldsymbol{\sigma}_s \cdot \hat{\boldsymbol{\varepsilon}}_s d\Omega_s + \int_{\Omega_s} \rho_s \mathbf{f}_s \cdot \hat{\mathbf{v}}_s d\Omega_s + \int_{\Gamma_{N_s}} \mathbf{g}_s \cdot \hat{\mathbf{v}}_s d\Gamma_s \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s. \end{aligned} \quad (3.26)$$

The first term in the left hand side is derived from the constitutive response, and in components the tangent operator reads

$$[\mathbf{D}_s]_{ijkl} = \frac{1}{\det \mathbf{F}_s} [\mathbf{F}_s]_{ia} [\mathbf{F}_s]_{jb} [\mathbf{F}_s]_{kc} [\mathbf{F}_s]_{ld} \left(\left(\frac{\partial \mathbf{S}_m}{\partial \mathbf{E}_m} \right)_s \right)_{abcd}. \quad (3.27)$$

In turn, the second term in the left hand side is the contribution to the tangent matrix of the geometrical nonlinearity resulting from the fact that the configuration

is unknown. Evidently, the fourth order tensor \mathbf{D}_s and the Cauchy stress tensor σ_s depend upon the current configuration, that is $\mathbf{D}_s = \mathbf{D}_s(\mathbf{u}_s)$ and $\sigma_s = \sigma_s(\mathbf{u}_s)$.

Given the tangent matrix and the stress state, problem (3.26) constitutes a single step in the iterative algorithm by which the equilibrium is achieved. This is discussed in detail in Section 3.5.4

For ease of notation we write problem as follows: find $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ such that

$$a_s(\delta \mathbf{u}_s, \hat{\mathbf{v}}_s) = l_s(\hat{\mathbf{v}}_s) \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.28)$$

where the bilinear form $a_s(\cdot, \cdot)$ and the linear form $l_s(\cdot)$ follow straightforwardly from (3.26).

3.3.3 Strain-stress relations and tangent operators

The relation between strain and stress in hyperelastic materials like those treated in the present work is accounted for through the definition of the strain energy function. This potential is a function of the current strain level. Then, let us consider the classical energy strain function ψ depending upon the Green-Lagrange deformation tensor $\psi(\mathbf{E}_m)$. Therefore, the Piola-Kirchhoff stress tensor \mathbf{S}_m is obtained as

$$\mathbf{S}_m = \frac{\partial \psi}{\partial \mathbf{E}_m}, \quad (3.29)$$

from which it is

$$\sigma_m = \frac{1}{\det \mathbf{F}_m} \mathbf{F}_m \frac{\partial \psi}{\partial \mathbf{E}_m} \mathbf{F}_m^T. \quad (3.30)$$

In turn, the stiffness tensor (fourth order tensor denoted by \mathbf{K}_m) associated with a certain level of strain, say \mathbf{E}_m , is obtained by doing

$$\mathbf{K}_m = \frac{\partial \mathbf{S}_m}{\partial \mathbf{E}_m} = \frac{\partial^2 \psi}{\partial \mathbf{E}_m \partial \mathbf{E}_m}, \quad (3.31)$$

which, according to this notation, it is written in the material configuration.

In this way, the expressions of the tangent matrix and of the Cauchy stress in the current configuration and in components are the following

$$[\mathbf{D}_s]_{ijkl} = \frac{1}{\det \mathbf{F}_s} [\mathbf{F}_s]_{ia} [\mathbf{F}_s]_{jb} [\mathbf{F}_s]_{kc} [\mathbf{F}_s]_{ld} \left(\left(\frac{\partial^2 \psi}{\partial \mathbf{E}_m \partial \mathbf{E}_m} \right)_s \right)_{abcd}. \quad (3.32)$$

$$[\sigma_s]_{ij} = \frac{1}{\det \mathbf{F}_s} [\mathbf{F}_s]_{ik} \left(\left(\frac{\partial \psi}{\partial \mathbf{E}_m} \right)_s \right)_{kl} [\mathbf{F}_s]_{jl}. \quad (3.33)$$

Observe that, for a given configuration for which we have \mathbf{F}_m at hand (and so \mathbf{E}_m) we have to compute the tangent matrix \mathbf{K}_m and the stress tensor \mathbf{S}_m deriving ψ using

(3.31) and (3.29), respectively. Then, these quantities are consistently transformed to the current configuration using equations (3.32) and (3.33), respectively.

In view of the wide range of constitutive characterizations for arterial tissues found in the literature, the implementation of the finite strain problem becomes untractable from the point of view of the effort invested in coding tangent expressions like the ones seen above in *ad hoc* situations. It is at this point that the present work presents its major contribution. We propose computing \mathbf{D}_s and σ_s by finite differences, as will be seen in Section 3.4.

3.3.4 Incompressible materials

In the case of incompressible materials we can derive the parallel expression of (3.26). Indeed, for such case we perform the decomposition of the Cauchy stress tensor into hydrostatic and deviatoric, namely $\sigma_s = -p_s \mathbf{I} + \sigma_s^D$, where p_s is the pressure and σ_s^D is the deviatoric component, both quantities considered in the spatial configuration. Then, the analogous to (3.11) is: find $\mathbf{u}_s \in \mathcal{U}_s$ and $p_s \in \mathcal{P}_s$ such that the deviatoric stress state σ_s^D is such that

$$\begin{aligned} \int_{\Omega_s} [-p_s \operatorname{div}_{\mathbf{x}_s} \hat{\mathbf{v}}_s + \sigma_s^D \cdot \hat{\mathbf{e}}_s] d\Omega_s = \\ \int_{\Omega_s} \rho_s \mathbf{f}_s \cdot \hat{\mathbf{v}}_s d\Omega_s + \int_{\Gamma_{N_s}} \mathbf{g}_s \cdot \hat{\mathbf{v}}_s d\Gamma_s \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \end{aligned} \quad (3.34)$$

and

$$\int_{\Omega_s} (1 - (\det \mathbf{F}_s)^{-1}) \hat{q}_s d\Omega_s = 0 \quad \forall \hat{q}_s \in \mathcal{P}_s. \quad (3.35)$$

These two equations (3.34) and (3.35) stand for the momentum and mass conservation, respectively. In this case \mathcal{U}_s and $\hat{\mathcal{V}}_s$ are as in (3.12), while it is $\mathcal{P}_s = L^2(\Omega_s)$.

Now, the deviatoric part of the Cauchy stress will be related to the strain energy function through the Piola-Kirchhoff tensor.

In this case, we write the variational formulation (3.34) in abstract form as: find $\mathbf{u}_s \in \mathcal{U}_s$ and $p_s \in \mathcal{P}_s$ such that

$$\langle \mathcal{S}_{\Omega_s}(\mathbf{u}_s, p_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s} = 0 \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.36)$$

$$\langle \mathcal{T}_{\Omega_s}(\mathbf{u}_s), \hat{q}_s \rangle_{\Omega_s} = 0 \quad \forall \hat{q}_s \in \mathcal{P}_s, \quad (3.37)$$

So, the linearization of (3.34)-(3.35) renders the following problem: find $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ and $\delta p_s \in \mathcal{P}_s$ such that

$$\begin{aligned} \langle \mathcal{D}_{\Omega_s}^{11}(\mathbf{u}_s, p_s) \delta \mathbf{u}_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} + \langle \mathcal{D}_{\Omega_s}^{12}(\mathbf{u}_s) \delta p_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = \\ - \langle \mathcal{S}_{\Omega_s}(\mathbf{u}_s, p_s), \hat{\mathbf{v}}_s \rangle_{\Omega_s} \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \end{aligned} \quad (3.38)$$

$$\langle \mathcal{D}_{\Omega_s}^{21}(\mathbf{u}_s) \delta \mathbf{u}_s, \hat{q}_s \rangle_{\Omega_s} = - \langle \mathcal{T}_{\Omega_s}(\mathbf{u}_s), \hat{q}_s \rangle_{\Omega_s} \quad \forall \hat{q}_s \in \mathcal{P}_s, \quad (3.39)$$

where $\mathcal{D}_{\Omega_s}^{11}(\mathbf{u}_s, p_q)$, $\mathcal{D}_{\Omega_s}^{12}(\mathbf{u}_s)$ and $\mathcal{D}_{\Omega_s}^{21}(\mathbf{u}_s)$ are the block tangent operators that form the tangent matrix

$$\mathcal{D}_{\Omega_s}(\mathbf{u}_s, p_q) = \begin{pmatrix} \mathcal{D}_{\Omega_s}^{11}(\mathbf{u}_s, p_q) & \mathcal{D}_{\Omega_s}^{12}(\mathbf{u}_s) \\ \mathcal{D}_{\Omega_s}^{21}(\mathbf{u}_s) & 0 \end{pmatrix}. \quad (3.40)$$

Let us obtain this block tangent operator following the same guidelines of the previous section. Firstly we start with the block $\mathcal{D}_{\Omega_s}^{11}$

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{S}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m, p_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \\ \frac{d}{d\tau} \left[\int_{\Omega_m} \left(-p_m (\mathbf{F}_{m,\tau}^{-T} \cdot \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) \det \mathbf{F}_{m,\tau} \right. \right. \\ \left. \left. + \mathbf{S}_m^D(\mathbf{E}_{m,\tau}) \cdot \frac{1}{2} (\mathbf{F}_{m,\tau}^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_{m,\tau}) d\Omega_m \right) \right] \Big|_{\tau=0}, \quad (3.41) \end{aligned}$$

where \mathbf{S}_m^D is such that

$$\boldsymbol{\sigma}_m^D = \frac{1}{\det \mathbf{F}_m} \mathbf{F}_m \mathbf{S}_m^D \mathbf{F}_m^T. \quad (3.42)$$

As before, using the expressions from the Appendix we obtain

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{S}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m, p_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \\ \int_{\Omega_m} \left[p_m ((\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m^T \mathbf{F}_m^{-T} \cdot \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) \det \mathbf{F}_m \right. \\ \left. - p_m (\mathbf{F}_m^{-T} \cdot \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) (\operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s)_m \det \mathbf{F}_m \right. \\ \left. + \frac{\partial \mathbf{S}_m^D}{\partial \mathbf{E}_m} \mathbf{F}_m^T \boldsymbol{\varepsilon}_m(\delta \mathbf{u}_s) \mathbf{F}_m \cdot \frac{1}{2} (\mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_m) \right. \\ \left. + \mathbf{S}_m^D \cdot \frac{1}{2} (\mathbf{F}_m^T (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m \mathbf{F}_m) \right] d\Omega_m. \quad (3.43) \end{aligned}$$

Back to the spatial configuration yields

$$\begin{aligned} \langle \mathcal{D}_{\Omega_s}^{11}(\mathbf{u}_s, p_s) \delta \mathbf{u}_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = \int_{\Omega_s} [p_s ((\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)^T \cdot \nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s - \operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s \operatorname{div}_{\mathbf{x}_s} \hat{\mathbf{v}}_s) + \\ \mathbf{D}_s^D \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \cdot \hat{\boldsymbol{\varepsilon}}_s + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \boldsymbol{\sigma}_s^D \cdot (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)] d\Omega_s, \quad (3.44) \end{aligned}$$

where \mathbf{D}_s^D is the analogous to (3.27) when it is derived using the deviatoric component \mathbf{S}_m^D .

Now, we add and subtract in (3.44) the term $p_s (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \cdot \nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s$, and consider the operation with second order tensors given by

$$p_s \operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s \operatorname{div}_{\mathbf{x}_s} \hat{\mathbf{v}}_s = p_s (\mathbf{I} \otimes \mathbf{I}) \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \cdot \hat{\boldsymbol{\varepsilon}}_s, \quad (3.45)$$

with \mathbf{I} the identity matrix. This leads to

$$\begin{aligned} \langle \mathcal{D}_{\Omega_s}^{11}(\mathbf{u}_s, p_s) \delta \mathbf{u}_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = \\ \int_{\Omega_s} [[\mathbf{D}_s^D + p_s(2\mathbb{I} - (\mathbf{I} \otimes \mathbf{I}))] \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \cdot \hat{\boldsymbol{\varepsilon}}_s + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \boldsymbol{\sigma}_s \cdot (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)] d\Omega_s. \end{aligned} \quad (3.46)$$

In the same manner, for the block $\mathcal{D}_{\Omega_s}^{12}$ we obtain

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{S}_{\Omega_m}(\mathbf{u}_m, p_m + \tau \delta p_m), \hat{\mathbf{v}}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \\ \frac{d}{d\tau} \left[\int_{\Omega_m} \left(-(p_m + \tau \delta p_m) (\mathbf{F}_m^{-T} \cdot \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) \det \mathbf{F}_m + \right. \right. \\ \left. \left. + \mathbf{S}_m^D(\mathbf{E}_m) \cdot \frac{1}{2} (\mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) + (\nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m)^T \mathbf{F}_m) d\Omega_m \right) \right] \Big|_{\tau=0} = \\ - \int_{\Omega_m} \delta p_m (\mathbf{F}_m^{-T} \cdot \nabla_{\mathbf{x}_m} \hat{\mathbf{v}}_m) \det \mathbf{F}_m d\Omega_m, \end{aligned} \quad (3.47)$$

and in the spatial configuration it results

$$\langle \mathcal{D}_{\Omega_s}^{12}(\mathbf{u}_s) \delta p_s, \hat{\mathbf{v}}_s \rangle_{\Omega_s} = - \int_{\Omega_s} \delta p_s \operatorname{div}_{\mathbf{x}_s} \hat{\mathbf{v}}_s d\Omega_s. \quad (3.48)$$

Finally, the linearization of the mass conservation to obtain block $\mathcal{D}_{\Omega_s}^{21}$ is carried out as follows

$$\begin{aligned} \frac{d}{d\tau} \langle \mathcal{T}_{\Omega_m}(\mathbf{u}_m + \tau \delta \mathbf{u}_m), \hat{q}_m \rangle_{\Omega_m} \Big|_{\tau=0} = \frac{d}{d\tau} \left[\int_{\Omega_m} (\det \mathbf{F}_m, \tau - 1) \hat{q}_m d\Omega_m \right] \Big|_{\tau=0} = \\ \int_{\Omega_m} \det \mathbf{F}_m (\operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s)_m \hat{q}_m d\Omega_m, \end{aligned} \quad (3.49)$$

and in the spatial configuration it is

$$\langle \mathcal{D}_{\Omega_s}^{21}(\mathbf{u}_s) \delta \mathbf{u}_s, \hat{q}_s \rangle_{\Omega_s} = \int_{\Omega_s} \operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s \hat{q}_s d\Omega_s. \quad (3.50)$$

Hence, with (3.44), (3.48) and (3.50) we formulate the linear problem for incompressible materials as follows: find $\delta \mathbf{u}_s \in \mathcal{W}_s$ and $\delta p_s \in \mathcal{P}_s$ such that

$$\begin{aligned} \int_{\Omega_s} [[\mathbf{D}_s^D + p_s(2\mathbb{I} - (\mathbf{I} \otimes \mathbf{I}))] \boldsymbol{\varepsilon}_s(\delta \mathbf{u}_s) \cdot \hat{\boldsymbol{\varepsilon}}_s + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \boldsymbol{\sigma}_s \cdot (\nabla_{\mathbf{x}_s} \hat{\mathbf{v}}_s)] d\Omega_s \\ - \int_{\Omega_s} \delta p_s \operatorname{div}_{\mathbf{x}_s} \hat{\mathbf{v}}_s d\Omega_s = - \int_{\Omega_s} \boldsymbol{\sigma}_s \cdot \hat{\boldsymbol{\varepsilon}}_s d\Omega_s + \int_{\Omega_s} \rho_s \mathbf{f}_s \cdot \hat{\mathbf{v}}_s d\Omega_s + \int_{\Gamma_{Ns}} \mathbf{g}_s \cdot \hat{\mathbf{v}}_s d\Gamma_s \\ \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \end{aligned} \quad (3.51)$$

and

$$-\int_{\Omega_s} \operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s \hat{q}_s \, d\Omega_s = \int_{\Omega_s} (1 - (\det \mathbf{F}_s)^{-1}) \hat{q}_s \, d\Omega_s \quad \forall \hat{q}_s \in \mathcal{P}_s. \quad (3.52)$$

As before, in order to simplify the notation we write problem as follows: find $\delta \mathbf{u}_s \in \delta \mathcal{U}_s$ and $\delta p_s \in \mathcal{P}_s$ such that

$$c_s(\delta \mathbf{u}_s, \hat{\mathbf{v}}_s) + b_s(\delta p_s, \hat{\mathbf{v}}_s) = f_s(\hat{\mathbf{v}}_s) \quad \forall \hat{\mathbf{v}}_s \in \hat{\mathcal{V}}_s, \quad (3.53)$$

$$b_s(\delta \mathbf{u}_s, \hat{q}_s) = d_s(\hat{q}_s) \quad \forall \hat{q}_s \in \mathcal{P}_s, \quad (3.54)$$

where the bilinear forms $c_s(\cdot, \cdot)$ and $b_s(\cdot, \cdot)$ as well as the linear forms $f_s(\cdot)$ and $d_s(\cdot)$ follow from (3.51)–(3.52).

3.3.5 Constitutive equations for soft tissues

In this section some constitutive equations used to characterize hyperelastic materials are presented. As stated before, we recall that the mechanical behavior for this type of materials is defined by a scalar function ψ , called the strain energy function, which is expressed in terms of the Green-Lagrange deformation tensor \mathbf{E}_m or the Cauchy-Green deformation tensor \mathbf{C}_m .

3.3.5.1 Linear Elasticity model

The linear elastic constitutive equation characterizes the behavior of materials that present a linear relation between strains and stresses. The implemented constitutive equation is derived from the generalized Hooke's Law ($[\mathbf{S}_m]_{ij} = [\mathbf{C}_m]_{ijkl} [\mathbf{E}_m]_{kl}$), simplified to treat isotropic materials. The strain energy function presents the following form

$$\psi(I_1, I_2) = 2\mu I_2 + \frac{\bar{\lambda}}{2} I_1^2, \quad (3.55)$$

where I_1 and I_2 are invariants of the tensor \mathbf{E}_m given by

$$I_1 = \operatorname{tr}(\mathbf{E}_m) \quad I_2 = \frac{1}{2} [(\operatorname{tr} \mathbf{E})^2 - \operatorname{tr}(\mathbf{E}^2)]. \quad (3.56)$$

The parameters ν and $\bar{\lambda}$ can be related to the known parameters E (Young's elasticity modulus) and ν (Poisson's coefficient) through this relations

$$\mu = G = \frac{E}{2(1+\nu)} \quad \bar{\lambda} = \frac{E}{3(1-2\nu)} - \frac{2\mu}{3}, \quad (3.57)$$

where μ characterizes the behavior of the material when subjected to shear stress, and ν controls the compressibility of the material ($\nu = 0.5$ for incompressible materials). Additional details regarding this constitutive law can be found in [16].

3.3.5.2 Mooney-Rivlin and Neo-hookean models

The Mooney-Rivlin constitutive equation is commonly used to model the mechanical response of different kind of elastic rubbers with good results when dealing with deformation levels lower than 100%. The strain energy function is originally proposed to treat incompressible materials and is defined by the expression

$$\psi(I_1, I_2) = C_1(I_1 - 3) + C_2(I_2 - 3), \quad (3.58)$$

where I_1 and I_2 are invariants of the Cauchy-Green deformation tensor \mathbf{C}_m

$$I_1 = \text{tr}(\mathbf{C}_m) \quad I_2 = \frac{1}{2}[(\text{tr}\mathbf{C})^2 - \text{tr}(\mathbf{C}^2)]. \quad (3.59)$$

For small deformations we can recover the linear elastic model using $E = 6(C_1 + C_2)$ and $G = 2(C_1 + C_2)$. The Neo-hookean model is a particular case of this Mooney-Rivlin model, in which $C_2 = 0$. To adapt this constitutive equation to deal with compressibility, the constraint $J = \det \mathbf{F}_m = 1$ is disregarded, and a term that takes into account volumetric changes given by the invariant $I_3 = J^2 = \det \mathbf{C}_m$ is added

$$\psi(I_1, I_2, I_3) = C_1(I_1 - 3) + C_2(I_2 - 3) + \psi_h^*(I_3) \quad (3.60)$$

Since the three terms contribute to the hydrostatic component of the stress tensor, it is convenient to perform a modification using a unique term contributing to the pressure $p = -\kappa(\bar{I}_3 - 1)$. Therefore, we employ the reduced invariants

$$\bar{I}_1 = I_1 I_3^{-\frac{1}{3}} \quad \bar{I}_2 = I_2 I_3^{-\frac{2}{3}} \quad \bar{I}_3 = I_3^{\frac{1}{3}}. \quad (3.61)$$

Finally, the strain energy function reads

$$\psi(\bar{I}_1, \bar{I}_2, \bar{I}_3) = C_1(\bar{I}_1 - 3) + C_2(\bar{I}_2 - 3) + \frac{1}{2}\kappa(\bar{I}_3 - 1)^2 \quad (3.62)$$

where κ is the compressibility modulus that characterizes the response of the material due to changes in the volume. Notice that choosing a sufficiently large value can be used to enforce incompressibility via a compressible formulation.

3.3.5.3 Exponential constitutive model proposed by Delfino

This is a simplified arterial wall model which proposes an exponential and isotropic strain energy function with the goal of modeling the typical stiffening of the tissue

for high levels of deformation. This phenomenon is a consequence of the presence of collagen fibers in the arterial wall, which have a significantly higher elasticity modulus compared with the elastin matrix in which the fibers are immersed. The fibers are originally curled and they become active after reaching certain stretch. The strain energy function is expressed by means of the reduced invariant \bar{I}_1 and reads

$$\bar{\Psi} = \frac{a}{b} \left\{ \exp \left[\frac{b}{2} (\bar{I}_1 - 3) \right] - 1 \right\}, \quad (3.63)$$

where $a > 0$ and $b > 0$ (non-dimensional) are the material parameters. For more details regarding this formulation, the reader is referred to [4, 8].

3.3.5.4 Fiber-reinforced model proposed by Holzapfel

This constitutive equation proposed in [8] treats as separate entities the elastin matrix and the collagen fibers. The elastin matrix is taken into account by an isotropic contribution, similar to the first term of the Mooney-Rivlin equation, while the collagen fibers are considered by an anisotropic contribution acting in two given directions.

In the arterial wall different layers can be distinguished. We will assume that each of them has a similar mechanical behavior, the same type of strain energy function will be used, with a different set of parameters and directions in which the collagen fibers act. The strain energy function takes the form

$$\bar{\Psi}(\bar{\mathbf{C}}, \mathbf{a}_{01}, \mathbf{a}_{02}) = \bar{\Psi}_{iso}(\bar{\mathbf{C}}) + \bar{\Psi}_{aniso}(\bar{\mathbf{C}}, \mathbf{a}_{01}, \mathbf{a}_{02}), \quad (3.64)$$

where $\bar{\mathbf{C}}$ is the Cauchy deformation tensor and $\mathbf{a}_{01}, \mathbf{a}_{02}$ are the directional versors that define the orientation of the collagen fibers. The first term of the right hand side, the isotropic contribution, reads

$$\bar{\Psi}_{iso}(\bar{I}_1) = \frac{c}{2} (\bar{I}_1 - 3), \quad (3.65)$$

and the anisotropic contribution is

$$\bar{\Psi}_{aniso}(\bar{I}_4, \bar{I}_6) = \frac{k_1}{2k_2} \sum_{i=4,6} \{ \exp[k_2(\bar{I}_i - 1)^2] - 1 \}, \quad (3.66)$$

Here the invariants of the Cauchy-Green tensor $\bar{\mathbf{C}}_m$ are introduced according to the directional versors

$$\bar{I}_1 = \text{tr} \bar{\mathbf{C}}_m \quad \bar{I}_2 = \frac{1}{2} \left[(\text{tr} \bar{\mathbf{C}}_m)^2 - \text{tr}(\bar{\mathbf{C}}_m^2) \right] \quad \bar{I}_3 = \det \bar{\mathbf{C}}_m = 1 \quad (3.67)$$

$$\overline{I}_4 = \overline{\mathbf{C}}_m \cdot \overline{\mathbf{A}}_1 \quad \overline{I}_6 = \overline{\mathbf{C}}_m \cdot \overline{\mathbf{A}}_2 \quad (3.68)$$

where

$$\overline{\mathbf{A}}_1 = \mathbf{a}_{01} \otimes \mathbf{a}_{01} \quad \overline{\mathbf{A}}_2 = \mathbf{a}_{02} \otimes \mathbf{a}_{02}. \quad (3.69)$$

3.4 Calculation of the tangent matrix and the stress state

3.4.1 Approximated versus analytical calculation

As said above, for compressible or incompressible materials we need to provide the expressions for \mathbf{D}_s and σ_s , or \mathbf{D}_s^D and σ_s^D , respectively. Within the context of biomechanics modeling several potential functions are available in the literature to represent the constitutive behavior of tissues [4, 8, 16, 24]. Indeed, the tissue engineering area has flourished in the last ten years, which has been put in evidence through the proliferation of new constitutive expressions accounting for different sorts of fiber-reinforced material.

The proliferation of constitutive relations may become a major issue if the computational implementation is not adequately coded. This is because as soon as a different potential function is required for modelling purposes, the element programming code should be completely rewritten if the expressions of its derivatives are embedded and spreaded along the entire code. In this manner, it is convenient to encapsulate the expressions for the energy potential and its derivatives behind software interfaces that isolate it from the main part of the element code where the general geometric aspects are considered.

Proceeding as suggested in the previous paragraph, it is concluded that the programming structure can be further simplified only if the expression for the elastic energy is required -coded- and its first and second derivatives are evaluated by finite differences. In view of this, in this section we introduce a finite difference approach for the calculation of the tangent matrix and the stress tensor. Consequently, the constitutive behavior is characterized simply by providing the strain energy potential ψ (or ψ^D for incompressible materials), being the tangent and stress tensors computed through several evaluations of this potential.

Thus, the entire element code is re-used to evaluate the mechanical equilibrium, and the only procedure to be changed when implementing a new constitutive law is that for the few lines where the expression of the potential ψ is coded. Evidently, this *generic programming* concept makes the constitutive model prototyping a quick and easy step towards assessing and comparing different constitutive models, which in the other hand, are in continuous and increasing development as can be observed in the current biomechanics literature.

The potential disadvantage is the more expensive computations involved in the approximated evaluation of the derivatives, when compared with the *specific programming* approach in which the analytical calculation and posterior programming of the first and second derivatives is performed.

However, notice that the finite-difference evaluation of the first and second derivatives is a fully parallel step. Therefore, the expensive computations involved in this generic programming concept can be drastically reduced such that we end up with a methodology which is competitive with, and even cheaper than, the specific programming idea.

3.4.2 Finite difference approximation of derivatives

Let us consider that the constitutive behavior is given by the strain energy potential ψ , which is a function of the Green-Lagrange tensor \mathbf{E}_m . Then, the quantities to be evaluated are the following

$$\frac{\partial \psi}{\partial \mathbf{E}_m} \quad \text{and} \quad \frac{\partial^2 \psi}{\partial \mathbf{E}_m \partial \mathbf{E}_m}. \quad (3.70)$$

The finite difference approach consists in evaluating the potential ψ at different strains denoted by $\mathbf{E}_m + \Delta \mathbf{E}_m$. It is important to note that being \mathbf{E}_m a symmetric tensor the corresponding functional form of ψ must be symmetrized or, alternatively, only symmetric perturbations of \mathbf{E}_m must be considered. Consequently, to evaluate the derivatives, \mathbf{E}_m will be perturbed preserving its symmetry. Thus we define the following nine symmetric second order tensorial perturbations:

$$(\Delta^{KL})_{ij} = \frac{1}{2} (\delta_{iK} \delta_{jL} + \delta_{iL} \delta_{jK}) \varepsilon, \quad (3.71)$$

where (i, j) are the indices of the tensor Δ^{KL} , and K, L fixed indicate its perturbed components. Parameter ε is small enough to meet a certain accuracy in the approximated calculation. Observe that the above expression represents a second order tensor whose nonzero components are for the indices (k, l) and (l, k) . Then, we introduce the abbreviated notation

$$\psi^{(\pm KL)} = \psi(\mathbf{E}_m \pm \Delta^{KL}), \quad (3.72)$$

$$\psi^{(\pm KL, \pm MN)} = \psi(\mathbf{E}_m \pm \Delta^{KL} \pm \Delta^{MN}), \quad (3.73)$$

$$\psi^{(0)} = \psi(\mathbf{E}_m). \quad (3.74)$$

And, finally, we give the finite difference approximation for \mathbf{S}_m as

$$[\mathbf{S}_m]_{kl} = S_{kl} = \frac{\partial \psi}{\partial E_{kl}} \simeq \frac{\psi^{(+KL)} - \psi^{(-KL)}}{2\varepsilon}. \quad (3.75)$$

In the same manner, the *diagonal* components of \mathbf{K}_m , denoted by $(\mathbf{K}_m)_{klkl} = K_{klkl}$, $\forall k, l$, are given by

$$[\mathbf{K}_m]_{klkl} = K_{klkl} = \frac{\partial^2 \psi}{\partial E_{kl}^2} \simeq \frac{\psi^{(+KL)} - 2\psi^{(0)} + \psi^{(-KL)}}{\varepsilon^2}. \quad (3.76)$$

Analogously, the *cross* derivatives (say $k \neq m$ or $l \neq n$) are calculated as follows

$$[\mathbf{K}_m]_{klmn} = K_{klmn} = \frac{\partial^2 \psi}{\partial E_{kl} \partial E_{mn}} \simeq \frac{\psi^{(+KL, +MN)} - [\psi^{(-KL, +MN)} + \psi^{(+KL, -MN)}] + \psi^{(-KL, -MN)}}{4\varepsilon^2}. \quad (3.77)$$

It is quite obvious that the above expressions are equivalent to approximate the potential ψ as a quadratic function around \mathbf{E}_m . Hence, this calculation entails performing a moving least squares quadratic approximation with the minimum number of points required to become well defined. Embedding these ideas in a Finite Element code is a straightforward task. As the energy potentials are smooth in the components of \mathbf{E}_m , the precision in the derivative calculation is not an obstacle as far as the evaluation may be made for values of ε small enough without any numerical impediments. So, a unique code can deal with virtually any hiperelastic material by means of a “black box” procedure being called by the element code with the unique purpose of evaluating $\psi(\mathbf{E}_m)$. In consequence, if another material law is needed, this last procedure may be easily changed as required.

3.5 Finite element approximation

Hereafter, index h denotes the finite element context. Since our aim is to solve three dimensional problems using low order interpolation functions, the domain of analysis is discretized through tetrahedral simplex elements. The finite element mesh is a tetrahedralization of Ω_s , and is denoted by T_h . The number of nodes in the mesh is N and the number of elements is M .

3.5.1 Preconditioning with zero-order terms

The mechanical problem requires a preconditioning so that we are able to solve the resulting algebraic system of equations. This is accomplished by modifying the variational formulation introducing a zero-order term through a numerical parameter, that is, we consider the following bilinear forms

$$t_s(\delta \mathbf{u}_s, \hat{\mathbf{v}}_s) = \int_{\Omega_s} \gamma \delta \mathbf{u}_s \cdot \hat{\mathbf{v}}_s \, d\Omega_s. \quad (3.78)$$

Parameter γ is a problem-dependent parameter, and will be specified in due course.

3.5.2 Discrete formulation for compressible materials

Index h makes reference to the finite-dimensional setting. As said, the interpolation functions are polynomials of degree one. That means that the finite dimensional spaces used to find approximated solutions are

$$\delta\mathcal{U}_{s,h} = \{\mathbf{w}_{s,h} \in \delta\mathcal{U}_s; (\mathbf{w}_{s,h} \cdot \mathbf{e}_i)|_K \in \mathbb{P}_1 \ i = 1, 2, 3 \ \forall K \in T_h\}, \quad (3.79)$$

$$\hat{\mathcal{V}}_{s,h} = \{\mathbf{v}_{s,h} \in \mathcal{V}_s; (\mathbf{w}_{s,h} \cdot \mathbf{e}_i)|_K \in \mathbb{P}_1 \ i = 1, 2, 3 \ \forall K \in T_h\}, \quad (3.80)$$

where \mathbf{e}_i , $i = 1, 2, 3$ are the unit Cartesian vectors.

In addition, we add the contribution to the left hand side of the preconditioning term of the previous section. Therefore, the discrete counterpart to (3.28) reads: find $\delta\mathbf{u}_{s,h} \in \delta\mathcal{U}_{s,h}$ such that

$$a_{s,h}(\delta\mathbf{u}_{s,h}, \hat{\mathbf{v}}_{s,h}) + l_{s,h}(\delta\mathbf{u}_{s,h}, \hat{\mathbf{v}}_{s,h}) = l_{s,h}(\hat{\mathbf{v}}_{s,h}) \quad \forall \hat{\mathbf{v}}_{s,h} \in \hat{\mathcal{V}}_{s,h}. \quad (3.81)$$

This discrete variational formulation leads to the following system of algebraic equations

$$(A_{s,h} + T_{s,h})\delta\mathbf{u}_s = \mathbf{l}_{s,h}, \quad (3.82)$$

where the dependence of $A_{s,h}$, $T_{s,h}$ and $\mathbf{l}_{s,h}$ with respect to the current state \mathbf{u}_s have been omitted for the sake of simplicity in the notation.

3.5.3 Discrete formulation for incompressible materials

As before, index h denotes entities related to the finite element approximation. The interpolation functions for the incremental displacement field and its variations is given by (3.79) and (3.80), respectively. For the pressure field the space is also linear, that is

$$\mathcal{P}_{s,h} = \{q_{s,h} \in \mathcal{P}_s; q_{s,h}|_K \in \mathbb{P}_1 \ \forall K \in T_h\}. \quad (3.83)$$

In this incompressible case, it is well-known that this pair of equal order spaces $(\delta\mathcal{U}_{s,h}, \mathcal{P}_{s,h})$ is not a stable choice in terms of the inf-sup condition associated to the formulation. Therefore, there is need for adding stabilization terms in the formulation which somehow cure the oscillations found in the pressure field when using a raw equal order approach.

The stabilization terms added to the discrete formulation are equivalent to those encountered in computational fluid dynamics when circumventing the inf-sup condition through the use of equal order interpolation, just as in the present case. Then, the stabilization term takes the form

$$\mathbf{g}_{s,h}(\delta p_{s,h}, \hat{q}_{s,h}) = \sum_{K \in T_h} \int_{\Omega_{s,K}} \frac{h_K^2}{2\mu} \nabla_{\mathbf{x}_s} \delta p_{s,h} \cdot \nabla_{\mathbf{x}_s} \hat{q}_{s,h} \, d\Omega_{s,K}, \quad (3.84)$$

where $\Omega_{s,K}$ is the spatial configuration of element K , $\forall K \in T_h$, h_K a characteristic element length and μ a given parameter.

Using the preconditioner presented in the previous section, and the stabilization terms presented above, the discrete problem is formulated as follows: find $\delta \mathbf{u}_{s,h} \in \delta \mathcal{U}_{s,h}$ and $\delta p_{s,h} \in \mathcal{P}_{s,h}$ such that

$$c_{s,h}(\delta \mathbf{u}_{s,h}, \hat{\mathbf{v}}_{s,h}) + t_{s,h}(\delta \mathbf{u}_{s,h}, \hat{\mathbf{v}}_{s,h}) + b_{s,h}(\delta p_{s,h}, \hat{\mathbf{v}}_{s,h}) = f_{s,h}(\hat{\mathbf{v}}_{s,h}) \quad \forall \hat{\mathbf{v}}_{s,h} \in \hat{\mathcal{V}}_{s,h}, \quad (3.85)$$

$$b_{s,h}(\delta \mathbf{u}_{s,h}, \hat{q}_{s,h}) + g_{s,h}(\delta p_{s,h}, \hat{q}_{s,h}) = d_{s,h}(\hat{q}_{s,h}) \quad \forall \hat{q}_{s,h} \in \mathcal{P}_{s,h}, \quad (3.86)$$

As with the compressible case, the discrete problem is related to the following system of algebraic equations

$$\begin{pmatrix} C_{s,h} + T_{s,h} & B_{s,h} \\ B_{s,h} & G_{s,h} \end{pmatrix} \begin{pmatrix} \delta \mathbf{u}_s \\ \delta p_s \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{s,h} \\ \mathbf{d}_{s,h} \end{pmatrix}. \quad (3.87)$$

As before, the dependence of $C_{s,h}$ with respect to (\mathbf{u}_s, p_s) and of $T_{s,h}$, $B_{s,h}$, $G_{s,h}$, $\mathbf{f}_{s,h}$ and $\mathbf{d}_{s,h}$ with respect to \mathbf{u}_s have been omitted for notational simplicity.

3.5.4 Iterative scheme

In this section we make use of the linearized problem obtained in Section 3.3 to formulate the iterative approach for computing the mechanical equilibrium. The ideas are exposed in the linearized continuum context and then the algorithm is described on top of the associated discrete problems.

The current spatial configuration is denoted by Ω_s^k , whose coordinates are \mathbf{x}_s^k . The displacement field that led us to this configuration is denoted by \mathbf{u}_s^k . We seek the configuration Ω_s^{k+1} by generating a displacement $\delta \mathbf{u}_s^{k+1} = \mathbf{u}_s^{k+1} - \mathbf{u}_s^k$ with the Newton method. For the sake of brevity we will talk about linear problem (3.26). This increment in the displacement is described in the current configuration, that is, it is given in terms of the current coordinates \mathbf{x}_s^k . For incompressible materials the same ideas apply, and we have to use expressions (3.51)-(3.52) for each Newton step.

For simplicity let us consider the formulation for compressible materials, and assume also that a given ψ is given, and that we have at hand a black box code, using the approach described in Section 3.4, that computes \mathbf{D}_s and σ_s . For a given spatial configuration Ω_s^k we will denote these quantities by \mathbf{D}_s^k and σ_s^k , correspondingly.

The iterative method proposed here is derived using the Newton-Raphson linearization over the internal power in equation (3.11), as explained in Section 3.3.2, and a fixed point method for the nonlinearities arising from the external power also in equation (3.11) (for instance configuration-dependent loadings such as a follower loads).

Specifically, the algorithm at the discrete level for compressible materials reads:

Algorithm 1 Mechanical equilibrium for compressible materials

-
1. given $\Omega_{s,h}^0$ (i.e. \mathbf{x}_s^0), ψ , \mathbf{u}_s^0 and σ do
 2. for $k = 0, 1, 2, \dots$ do
 3. compute $\nabla_{\mathbf{x}_s^k} \mathbf{u}_{s,h}^k$, $(\mathbf{F}_{s,h}^k)^{-1}$ and $\mathbf{E}_{m,h}^k$
 4. compute $(\mathbf{S}_{m,h}^k)_{s,h}$ and $(\mathbf{K}_{m,h}^k)_{s,h}$
 5. compute $\sigma_{s,h}^k$ and $\mathbf{D}_{s,h}^k$
 6. assemble $\mathbf{A}_{s,h}^k$, $\mathbf{T}_{s,h}^k$ and $\mathbf{l}_{s,h}^k$
 7. solve $(\mathbf{A}_{s,h}^k + \mathbf{T}_{s,h}^k) \delta \mathbf{u}_s^{k+1} = \mathbf{l}_{s,h}^k$
 8. do $\mathbf{u}_s^{k+1} = \mathbf{u}_s^k + \delta \mathbf{u}_s^{k+1}$
 9. until $\frac{\|\delta \mathbf{u}_s^{k+1}\|}{\|\mathbf{u}_s^k\|} < \sigma$, $k = k + 1$
 10. end
-

In turn, for incompressible materials the algorithm is the following:

Algorithm 2 Mechanical equilibrium for incompressible materials

-
1. given $\Omega_{s,h}^0$ (i.e. \mathbf{x}_s^0), ψ , \mathbf{u}_s^0 , $p_s^0 = 0$ and σ do
 2. for $k = 0, 1, 2, \dots$ do
 3. compute $\nabla_{\mathbf{x}_s^k} \mathbf{u}_{s,h}^k$, $(\mathbf{F}_{s,h}^k)^{-1}$ and $\mathbf{E}_{m,h}^k$
 4. compute $(\mathbf{S}_{m,h}^{Dk})_{s,h}$ and $(\mathbf{K}_{m,h}^{Dk})_{s,h}$
 5. compute $\sigma_{s,h}^{Dk}$ and $\mathbf{D}_{s,h}^{Dk}$
 6. assemble $\mathbf{C}_{s,h}^k$, $\mathbf{T}_{s,h}^k$, $\mathbf{B}_{s,h}^k$, $\mathbf{G}_{s,h}^k$, $\mathbf{f}_{s,h}^k$ and $\mathbf{d}_{s,h}^k$
 7. solve $\begin{pmatrix} \mathbf{C}_{s,h}^k + \mathbf{T}_{s,h}^k & \mathbf{B}_{s,h}^k \\ \mathbf{B}_{s,h}^{kT} & \mathbf{G}_{s,h}^k \end{pmatrix} \begin{pmatrix} \delta \mathbf{u}_s^{k+1} \\ \delta p_s^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{s,h}^k \\ \mathbf{d}_{s,h}^k \end{pmatrix}$
 8. do $\mathbf{u}_s^{k+1} = \mathbf{u}_s^k + \delta \mathbf{u}_s^{k+1}$, $p_s^{k+1} = p_s^k + \delta p_s^{k+1}$
 9. until $\max \left\{ \frac{\|\delta \mathbf{u}_s^{k+1}\|}{\|\mathbf{u}_s^k\|}, \frac{\|\delta p_s^{k+1}\|}{\|p_s^k\|} \right\} < \sigma$, $k = k + 1$
 10. end
-

Notice that, if converges, the algorithms described above are such that $\delta \mathbf{u}_s^{k+1} \rightarrow 0$. Then, the body configuration converges to the spatial (deformed) configuration $\Omega_{s,h}$, which implies that the mechanical equilibrium is actually achieved in this spatial configuration $\Omega_{s,h}$.

3.6 Numerical examples

In this section several problems are solved whose aim is to test and validate the proposed model. Four examples are shown, the first three of them with emphasis in validating the model, comparing the obtained results with already published data,

while the fourth case is a numerical example performed in a real geometry of a cerebral aneurysm obtained from medical images.

3.6.1 Thick walled cylinder

Test description

This test is the classical problem of a thick walled cylinder loaded with an inner pressure. In order to validate our model we will compare our results with analytical results, which are well known for this problem when dealing with materials that behave according to the linear elasticity theory. Hence, we choose for this test the following values for the geometry, material properties and loadings

- Inner radius: 5 m.
- External radius: 10 m.
- Axial length: 1 m.
- $E = 1000 \frac{N}{m^2}$
- $\nu \approx 0.5$
- Applied pressure = $103.13 \frac{N}{m^2}$

For the compressible formulation we will study how close we can get ν to 0.5 while still obtaining coherent results.

Mesh

Taking advantage of the symmetry of the problem, we will reduce the domain to a quarter of the cylinder, using the boundary conditions specified in the corresponding section. The following cases are considered.

- Mesh A: tetrahedral elements with linear interpolation for displacements and pressure. The domain is discretized using 18 elements in the circumferencial direction, 1 element in the axial direction and 10 elements in the radial direction. The resulting mesh can be observed in Figure 3.2.
- Mesh B: tetrahedral elements with linear interpolation for displacements and pressure. The domain is discretized using 36 elements in the circumferencial direction, 4 element in the axial direction and 30 elements in the radial direction. The mesh can be observed in Figure 3.3.
- Mesh C: tetrahedral elements with quadratic interpolation for displacements and linear interpolation for the pressure. The domain is discretized using 36 elements in the circumferencial direction, 4 element in the axial direction and 30 elements in the radial direction. The mesh can be observed in Figure 3.4.

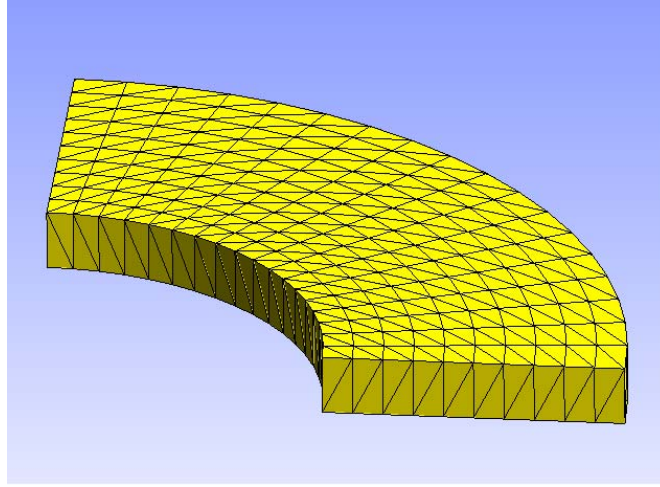


Fig. 3.2 Thick walled cylinder. Mesh A.

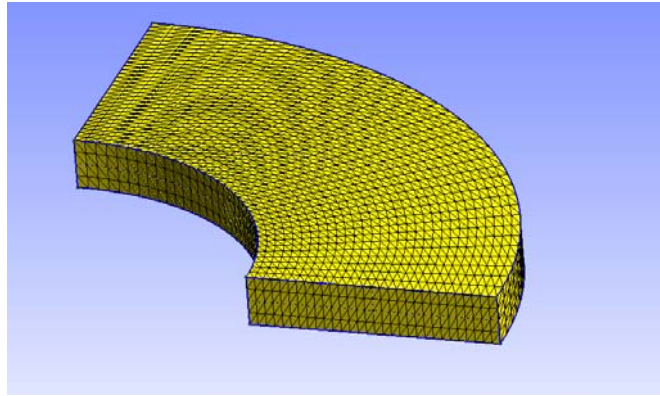


Fig. 3.3 Thick walled cylinder. Mesh B.

Boundary conditions

For this problem the boundary conditions described below are considered.

- Since we are dealing with a plane strain problem, we impose null displacements in the axial direction for the superior and inferior cross-sectional surfaces.
- Pressure applied in the inner surface of the cylinder.
- Null displacements in the normal directions of the symmetry surfaces.

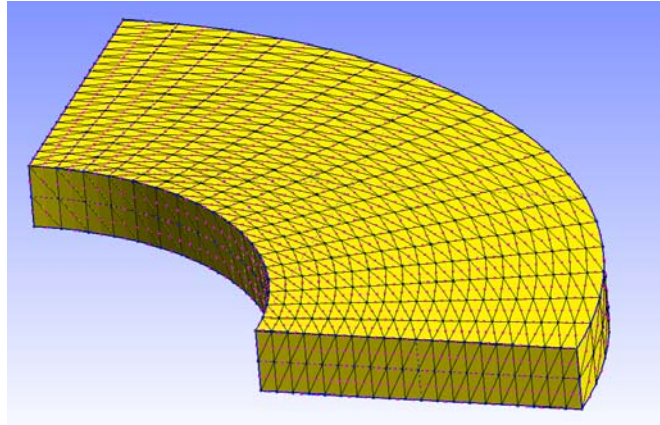


Fig. 3.4 Thick walled cylinder. Mesh C.

Results

In Table 3.1 a summarize of the obtained solutions with the compressible and incompressible formulations with linear and quadratic interpolations is presented. The radial displacement of the inner surface was selected as an indicator to perform the comparisons with the exact analytic solution for this problem. Also, some comments about each one of the tested alternatives will be exposed.

Formulation / Interpolation type	Radial displacement inner radius	Difference [%]
<i>Analytic result</i>	1.0313	-
Compressible / Linear	1.0206	1.04
Compressible / Quadratic	1.0185	1.24
Stabilized incompressible / Linear-linear	1.0310	0.03
Incompressible / Quadratic-linear	1.0302	0.11

Table 3.1 Thick walled cylinder. Results.

Compressible formulation, linear interpolation

Using this combination we are not able to obtain coherent results for the proposed value for ν . The compressible formulation with this linear interpolation is unable to represent correctly the incompressibility constraint, since it features an unstable solution. In Figures 3.5 and 3.6 the radial displacement field for different values for ν using the meshes “A” and “B”, respectively, are displayed. In both cases it can be appreciated that when ν reaches 0.5 (incompressible material) the model is not capable of producing valid results, featuring “locking” phenomena that leads to asymmetries in the solution.

When discretizing using the mesh A, the asymmetry in the displacement of the inner radius is close to 2% for $\nu = 0.47$. With mesh B better results are obtained, with adequate behavior for $\nu \leq 0.49$.

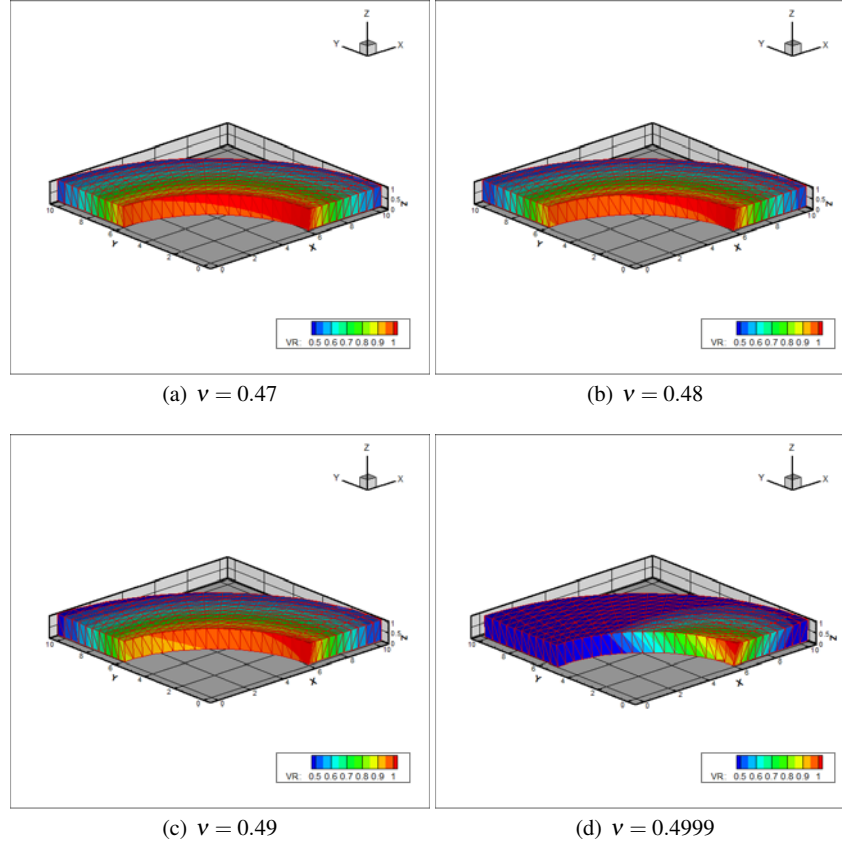


Fig. 3.5 Thick walled cylinder. Compressible formulation, linear interpolation, results with mesh A.

Compressible formulation, quadratic interpolation

Repeating the analysis performed above with linear interpolation, see Figure 3.7, the solutions for quadratic interpolation show better results when dealing with the incompressibility constraint, reaching adequate solutions even for $\nu \leq 0.499999$, which is an acceptable approximation to fulfil incompressibility in most cases.

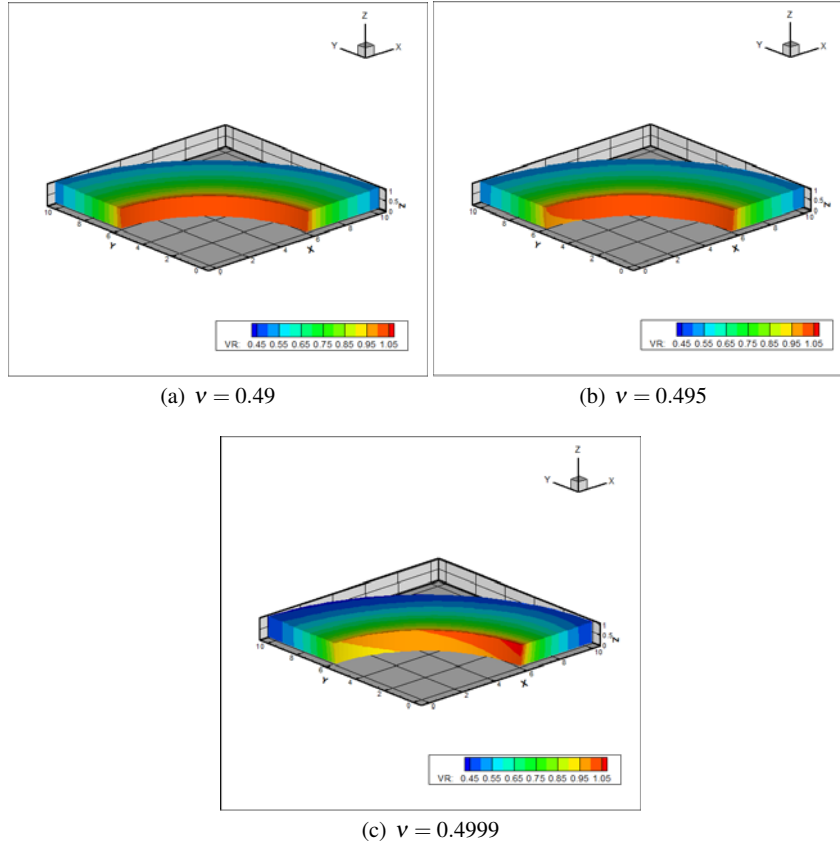


Fig. 3.6 Thick walled cylinder. Compressible formulation, linear interpolation, results with mesh B.

Incompressible formulation, quadratic/linear interpolation (displacements/pressure)

Using the incompressible formulation along with quadratic interpolation for displacements and linear interpolation for the pressure resulted in satisfactory solutions. An increase in the computational cost is however noticed, but it is negligible for this small problem. In Figure 3.8 the radial displacements field was displayed over the deformed configuration after reaching the equilibrium.

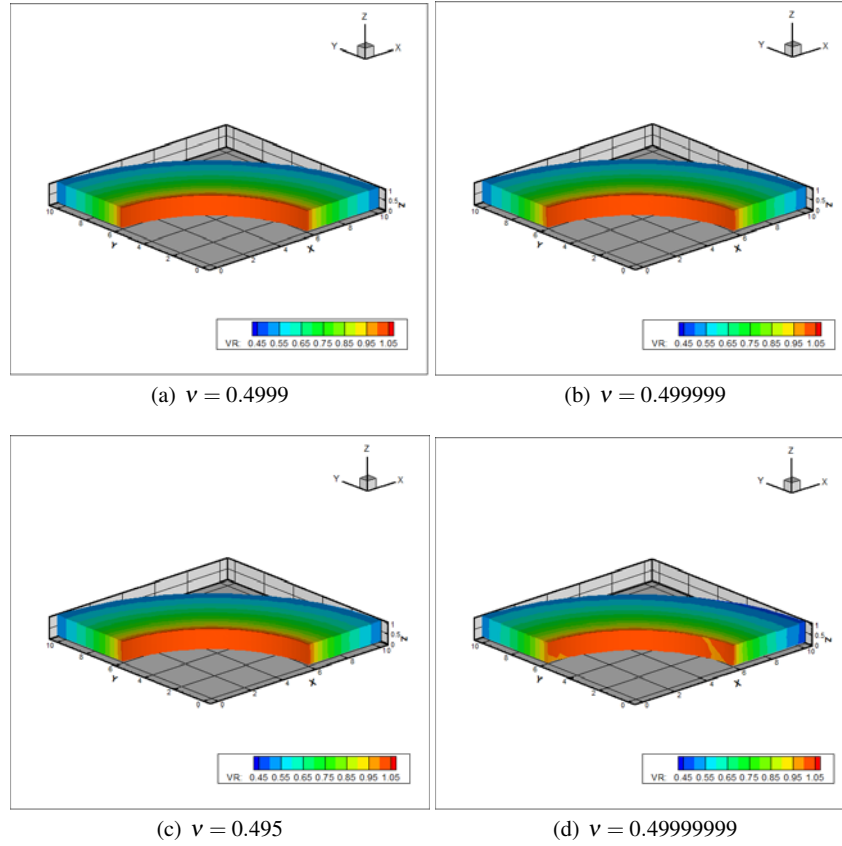


Fig. 3.7 Thick walled cylinder: compressible formulation, quadratic interpolation.

Stabilized incompressible formulation, linear interpolation (displacements/pressure)

Stabilized incompressible formulations with linear interpolation for displacement and pressure fields are required for the pure incompressible case. In Figure 3.9 the radial displacement field plotted in the deformed configuration can be observed.

Discussion

As observed, both formulations compressible and incompressible are capable of providing accurate results for this simple example. However, this test allowed us to put in evidence the problems of the compressible formulation to model materials with a behavior close to the incompressible limit. This problem is specially relevant when using linear interpolation. Also, despite being negligible for this problem, it

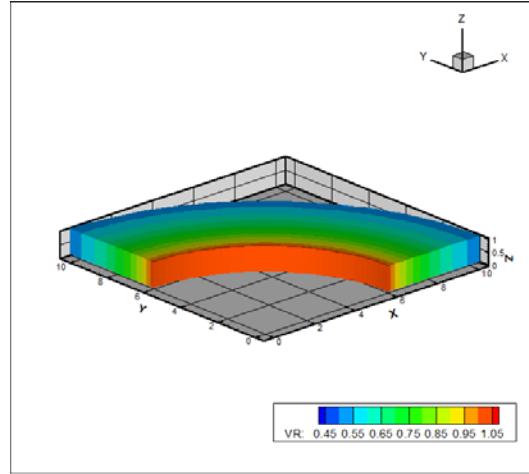


Fig. 3.8 Thick walled cylinder: incompressible formulation, quadratic interpolation for displacements and linear interpolation for pressure, results with mesh C, $\nu = 0.5$

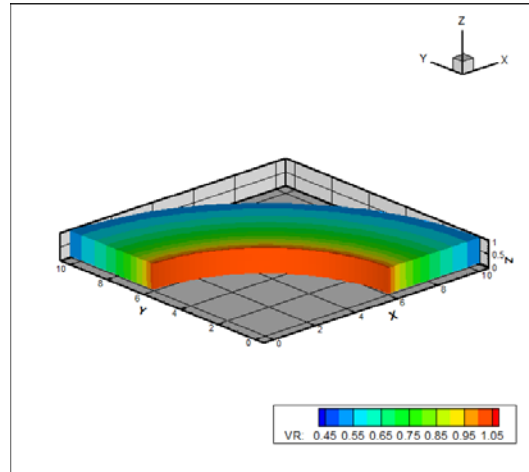


Fig. 3.9 Thick walled cylinder: stabilized incompressible formulation, linear interpolations, results with mesh B, $\nu = 0.5$

is noticeable the increment in the computational cost when using quadratic interpolation for both compressible and incompressible formulations. As result of this two factors, we choose to continue our work in the following numerical examples with the incompressible stabilized formulation combined with first order interpolation for both displacements and pressure fields.

3.6.2 Human carotid artery modeled with Delfino's constitutive law

Test description

In this test the numerical results presented in [4] are reproduced, from which we extracted the geometrical data and the values for the material parameters to model the artery using the exponential constitutive equation proposed by the same work and which was recalled. The geometry of the artery is assumed to be perfectly cylindrical, subjected to an axial pre-stress (indicated as a percentage of the axial longitude of the cylinder) and an inner pressure. The problem will be solved for a quarter of cylinder and for different levels of inner pressure. This will allow us to study the relation between inner pressure and inner radius.

- Inner radius: $R_i = 3.1$ mm
- External radius: $R_e = 4.1$ mm
- Material parameters: $a = 44.2$ kPa , $b = 16.7$

Mesh

Figure 3.10 shows the mesh used to model the arterial wall. The mesh contains 8640 linear tetrahedra, discretized with 10 elements in the radial direction, 4 in the axial direction and 36 in the circumferencial direction; resulting in 2035 nodes and 8140 degrees of freedom.

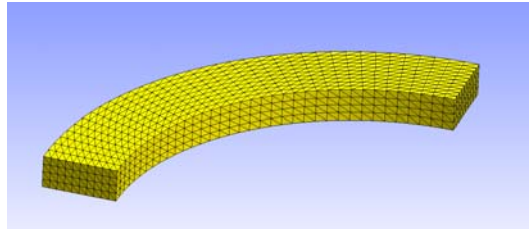


Fig. 3.10 Mesh for human carotid artery modeled with exponential constitutive equation proposed by Delfino.

Boundary condition

For this test the following boundary conditions are imposed:

- Dirichlet boundary conditions for the displacement of the nodes in the superior and inferior surfaces. In order to ensure the required pre-stretch for each case, we impose null displacements in the axial direction for the inferior surface and displacements equal to the corresponding pre-stretch for the superior surface.

- Inner pressure over the interior surface of the arterial wall.
- Null displacements in the normal direction of the symmetry surfaces.

Results

We performed simulations for different values of inner pressure between 0 and 25 kPa for each pre-stretch condition, finding the configuration of equilibrium for every set of loads proposed. The results are compared with the results published in [8] using as the data indicator the inner pressure vs. inner radius curves, shown through Figure 3.11. Furthermore, some solutions are displayed in Figure 3.12, which presents radial displacements plotted over the deformed configurations for different situations of pre-stretch and load. and in Figure 3.13, which shows examples of the pressure field.

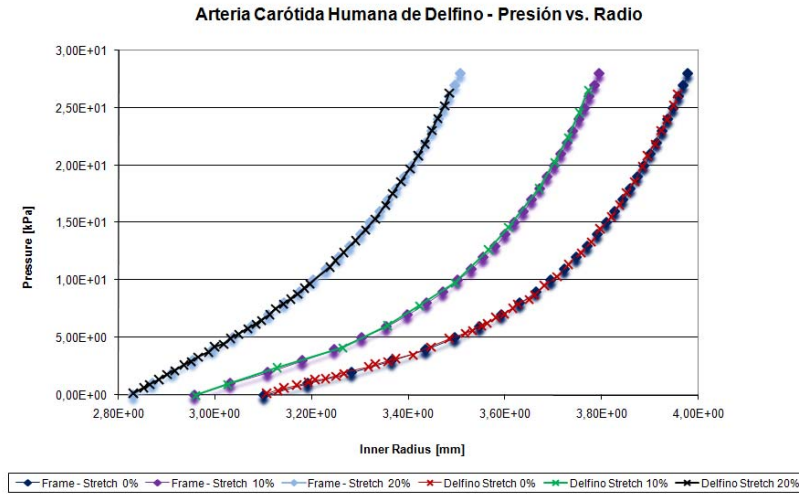


Fig. 3.11 Comparison of pressure vs. inner radius curves, human carotid artery modeled with exponential constitutive equation proposed by Delfino.

Discussion

In this case is important to remark that the proposed model is tested in an application case with similar characteristics to the hemodynamics conditions, that represent the final objective of this work. The results are reproduced with high accuracy and the pressure and displacement fields are coherent, uniform and symmetric.

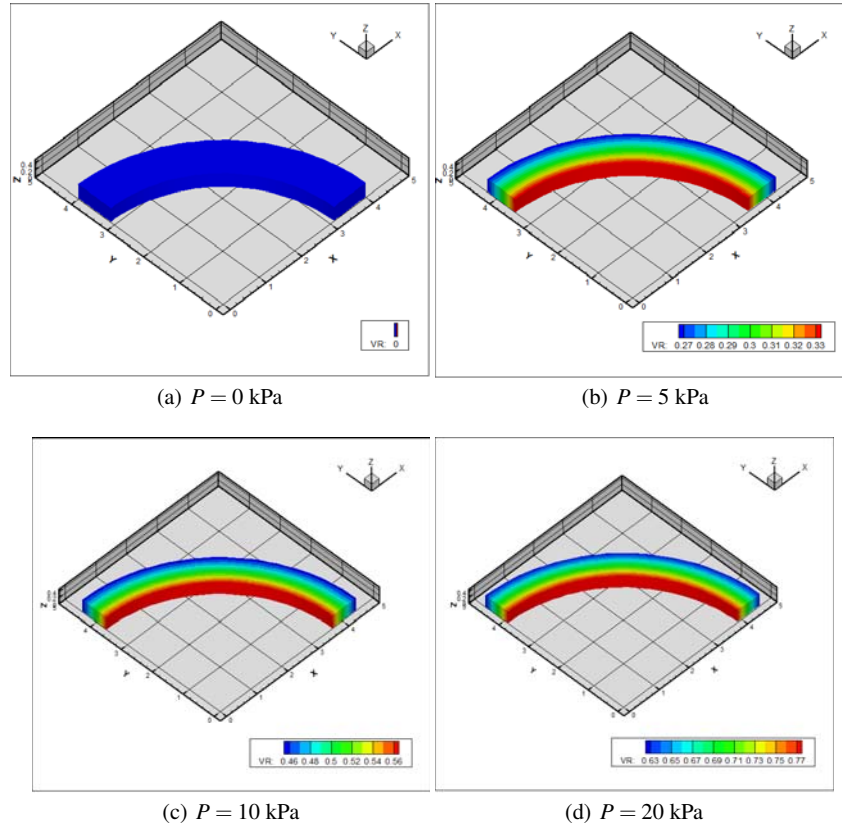


Fig. 3.12 Radial displacements, human carotid artery modeled with exponential constitutive equation proposed by Delfino.

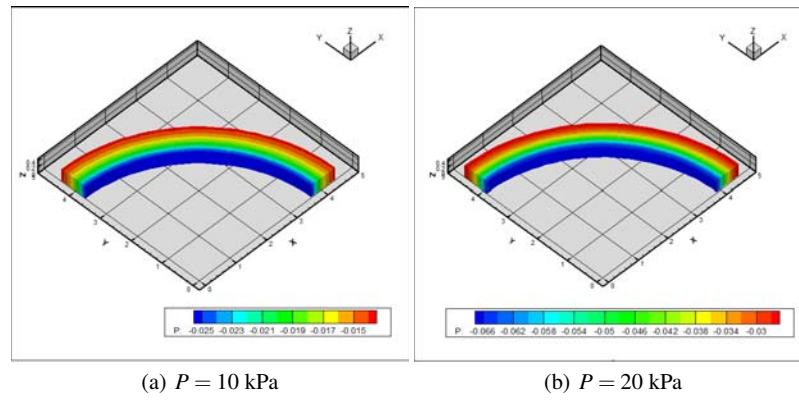


Fig. 3.13 Pressure field, human carotid artery modeled with exponential constitutive equation proposed by Delfino.

3.6.3 Rabbit carotid artery modeled with anisotropic and multi-layer constitutive equation proposed by Holzapfel

3.6.3.1 Test description

A segment of a rabbit carotid artery is modeled as described in [8]. It is assumed that the arterial wall is composed by two layers: media (interior) and adventitia (exterior). Both with a similar mechanical behaviour, as described in Section 3.3.5.4, but with a different set of material parameters. A representative scheme of the arterial wall layers is represented in Figure 3.14. The geometrical data, parameters for the strain energy functions and orientation angle of the collagen fibers are extracted from the cited publication. As done for the case of the previous section, a pre-stretch is applied and then the arterial wall segment is loaded with an inner pressure.

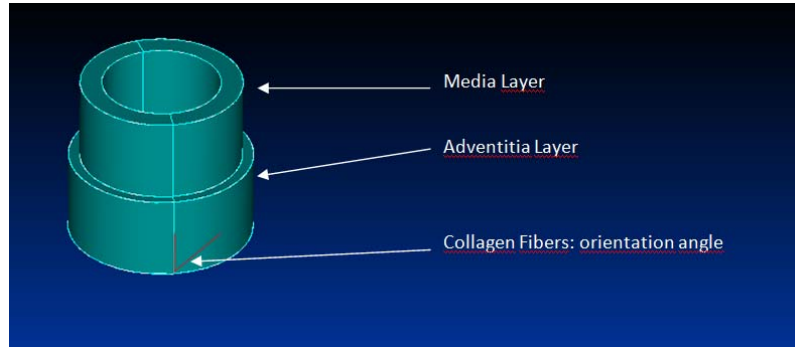


Fig. 3.14 Arterial wall scheme, media and adventitia layers, collagen fibers orientation angle

Here the geometrical data for the zero-stress state follows, while Table 3.2 shows the material parameters used for each layer.

- Inner radius: $R_i = 0.71$ mm
- Media layer thickness: $t_M = 0.26$ mm
- Adventitia layer thickness: $t_A = 0.13$ mm

	Media	Adventitia
c [kPa]	3.0000	0.3000
k_1 [kPa]	2.3632	0.5620
k_2	0.8323	0.7112
β	61	28

Table 3.2 Strain energy function parameters, rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

Mesh

We use a mesh composed by 12960 linear tetrahedra, with 2849 nodes and 11396 degrees of freedom. The volume is divided in two sectors to separate the different material layers. The discretization is performed with 5 elements in the radial direction for each layer, 36 elements in the circumferencial direction and 4 elements in the axial direction. Figure 3.15 shows the discretized domain.

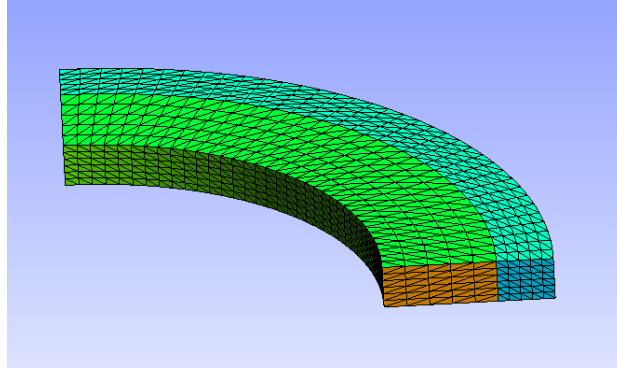


Fig. 3.15 Mesh for rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

Boundary condition

Boundary conditions are imposed in the same exact way as presented in Section 3.6.2 for the human carotid artery.

Results

The problem is solved for pressure levels between 0 and 25 kPa, for two different pre-stretch levels, 50% and 90%. It is important to remark that was necessary to adjust the parameter that control the stabilization, μ , as we increase the inner pressure applied. The highly non-linear behavior of the constitutive response leads to important variations in the material stiffness modulus, making necessary to increase the stabilization, decreasing the value of μ . This adjustment was performed manually in order to avoid instabilities in the solution, one example of this problem is shown in Figure 3.16 where oscillations in the pressure field can be noticed.

As proceeded for the human carotid, we present the pressure vs. inner radius curves obtained comparing, once again, with the results published in [8]. These results are presented in Figure 3.18. Figure 3.17 shows the radial displacements and pressure plotted over the deformed configuration when we applied only a initial pre-

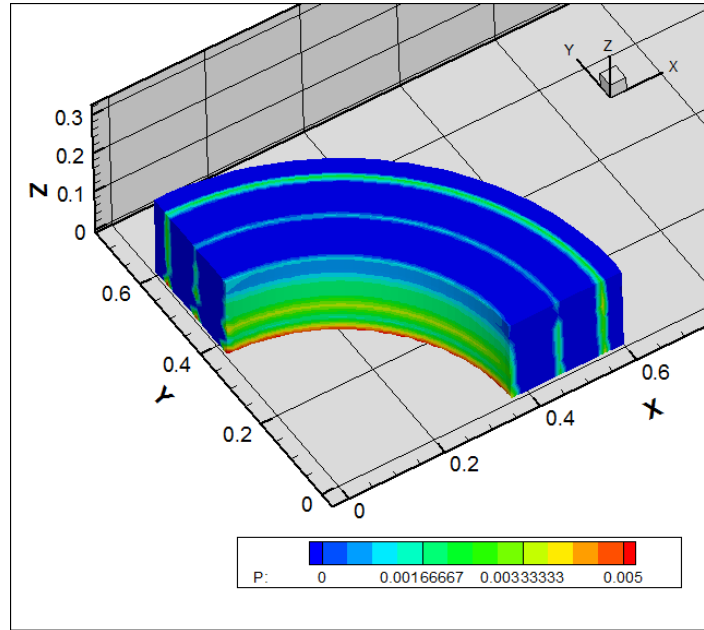


Fig. 3.16 Instability in the pressure field, inadequate stabilization in rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

stretch of 90% of the axial length of the arterial segment. Moreover, as example of the equilibrium configurations obtained in this simulations, Figures 3.19 and 3.20 display the radial displacements and pressure fields, respectively, for two different loading conditions.

Discussion

The presented results allow us to confirm that the variational formulation and the methodology employed to deal with the constitutive response are capable of reaching accurate results for this complex case. It is important to remark that besides being a multi-layer problem with constitutive and geometrical non-linearities, we are facing the additional difficulty of working with an anisotropic material with preferred directions given by the incorporation of the collagen fibers. Another aspect to be emphasized is the influence of the stabilization parameter μ , since an inadequate selection can lead to numerical oscillations (lack of stabilization) or to an excessive diffusion of the pressure field (excessive stabilization). Although in this present work we selected manually this parameter, we will work towards an adaptive method to perform this selection in an automatic way.

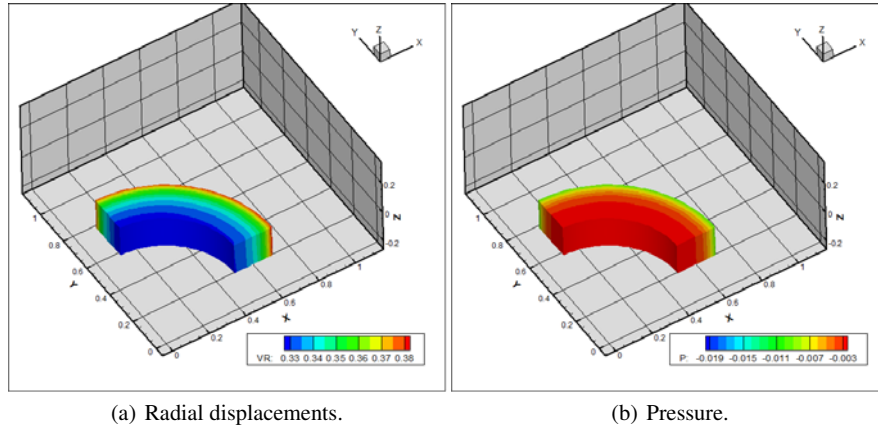


Fig. 3.17 Applied axial pre-stretch of 90% in rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

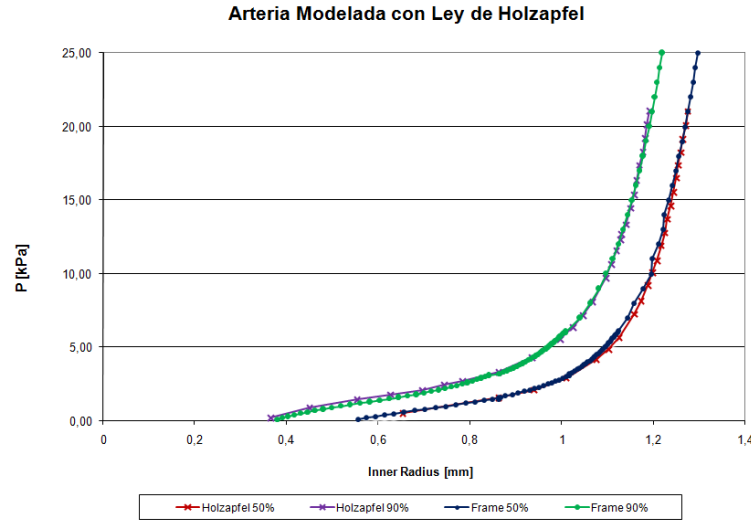


Fig. 3.18 Comparison of pressure vs. inner radius curves, inadequate stabilization in rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

3.6.4 Intracranial Aneurysm

Test description

Here we present a numerical example of a real intracranial aneurysm obtained from medical images. The thickness of the arterial wall is $t = 0.3[mm]$. The material con-

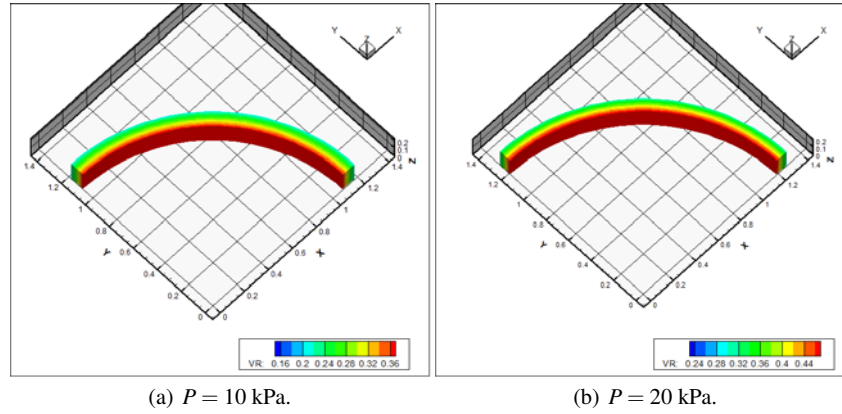


Fig. 3.19 Radial displacement field, pre-stretch 90%, inadequate stabilization in rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

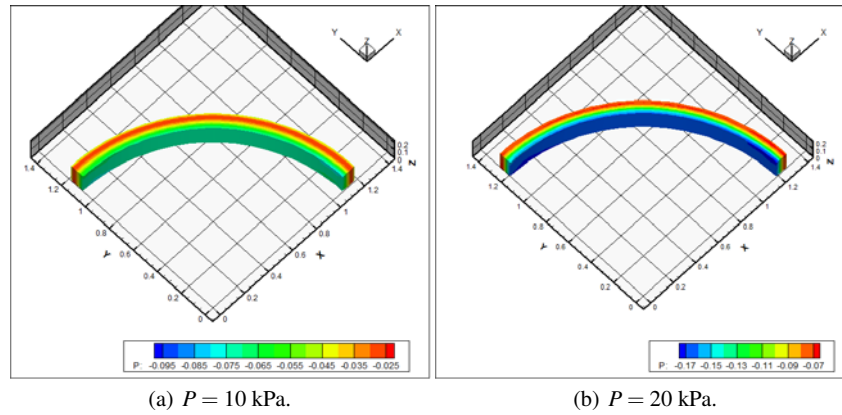


Fig. 3.20 Pressure field, pre-stretch 90%, inadequate stabilization in rabbit carotid artery modeled with constitutive equation proposed by Holzapfel.

stitutive response was modeled with the exponential equation proposed by Delfino (defined in Section 3.3.5.3). The test simply consist in applying pressure to the inner surface of the aneurysm progressively.

Mesh

Figures 3.21 show an external view of the used meshes while, Figure 3.22 presents a close up of the interior of them. Mesh sizes are:

- Mesh A: number of elements: $5 \cdot 10^4$, number of nodes: $2 \cdot 10^4$.
- Mesh B: number of elements: $2 \cdot 10^5$, number of nodes: 10^5 .

- Mesh C: number of elements: 10^6 , number of nodes: $8 \cdot 10^5$.

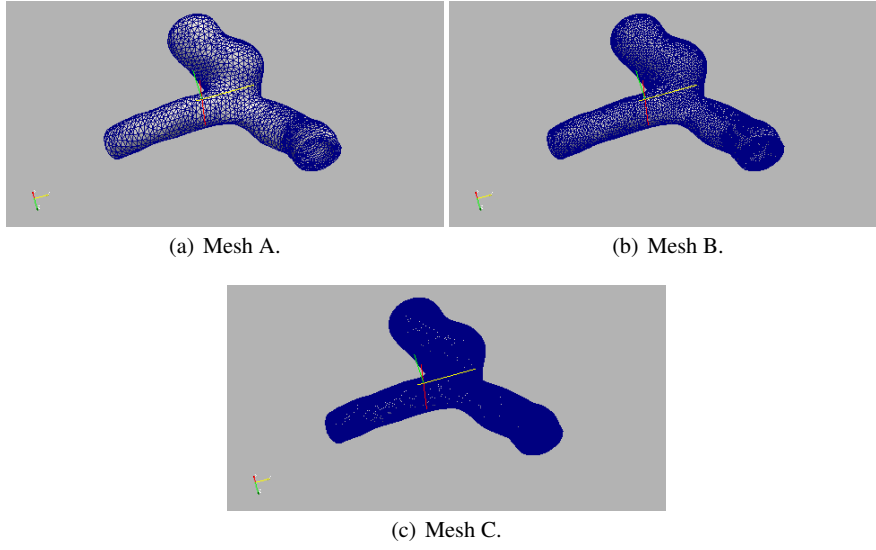


Fig. 3.21 Meshes for the intracranial aneurysm.

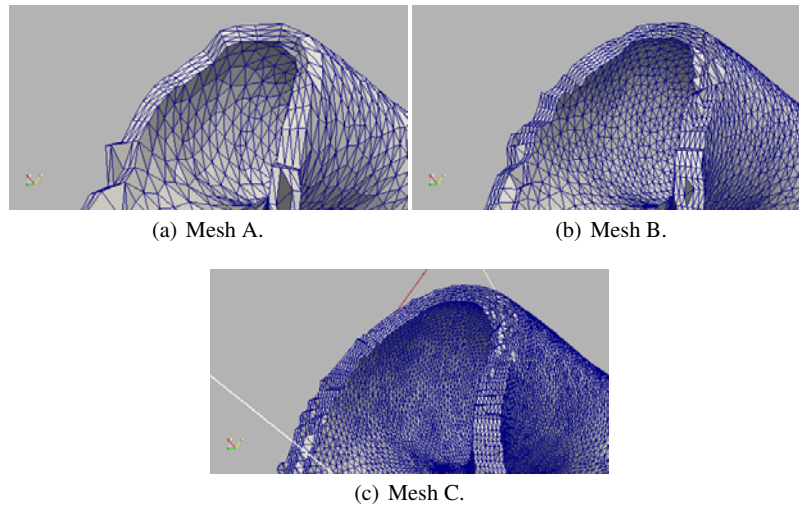


Fig. 3.22 Close up on meshes for the intracranial aneurysm.

Boundary condition

Although it is not a realistic assumption, we consider the artery clamped in both extremes for the sake of simplicity for this example. As a consequence, Dirichlet conditions that impose null displacement in all directions are set for the nodes on both extremes or the arterial wall.

Preconditioning

As result of the characteristics of the problem and the size of the meshes employed in the analysis, the system of linear equations system is ill-conditioned and presents difficulties to be solved using classical iterative methods. In order to deal with this inconvenient we add a zero-order preconditioner term as described in Section 3.5.1. We select the value for γ taking into account that higher values leads to a faster solving of the linear system but also increases the number of iterations of the Newton scheme, obtaining best results for $\gamma \approx 10\%a$, where a is the material parameter corresponding to the Delfino's constitutive equation.

Results

In this section are presented results for the arterial wall in equilibrium loaded with an inner pressure of 110 [mmHg], Figure 3.23 can be found the pressure field, while Figure 3.25 displays the magnitude of the displacements and Figure 3.24 shows a comparison between the reference and the obtained deformed configurations for each mesh.

Discussion

This example can be considered as a first step towards a realistic tridimensional model for hemodynamics problems. The obtained results are encouraging, since we reached coherent simulations under histological loads, using complex and non linear models for the arterial wall behavior on a patient-specific image based geometry. Next steps include working on fluid-structure interaction and in coupling with 1D models to provide the best possible boundary conditions for the problem.

3.7 Conclusions

In this work a generic strategy to deal with constitutive equations for hiperelastic incompressible materials in finite strains regime has been developed. This approach presents evident advantages for the rapid prototyping of existing, and eventually

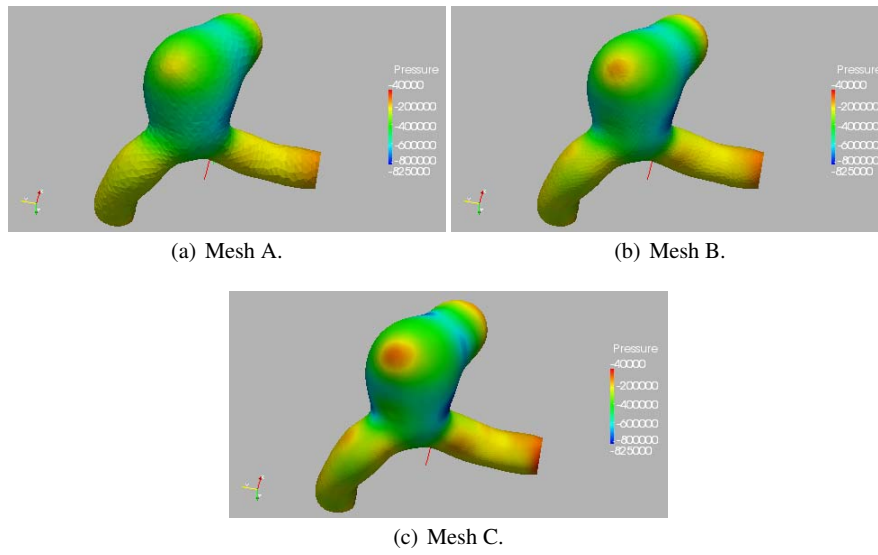


Fig. 3.23 Pressure field result for intracranial aneurysm.

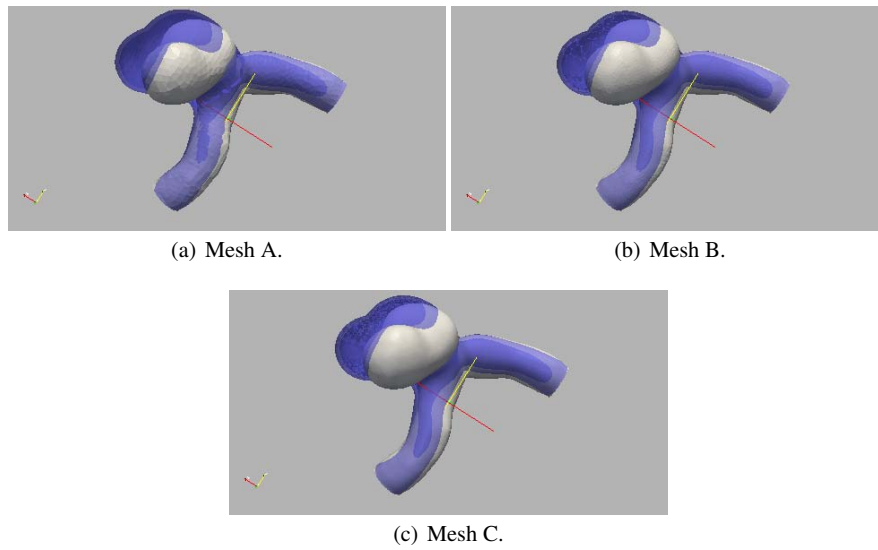


Fig. 3.24 Comparison of original and deformed configurations for intracranial aneurysm.

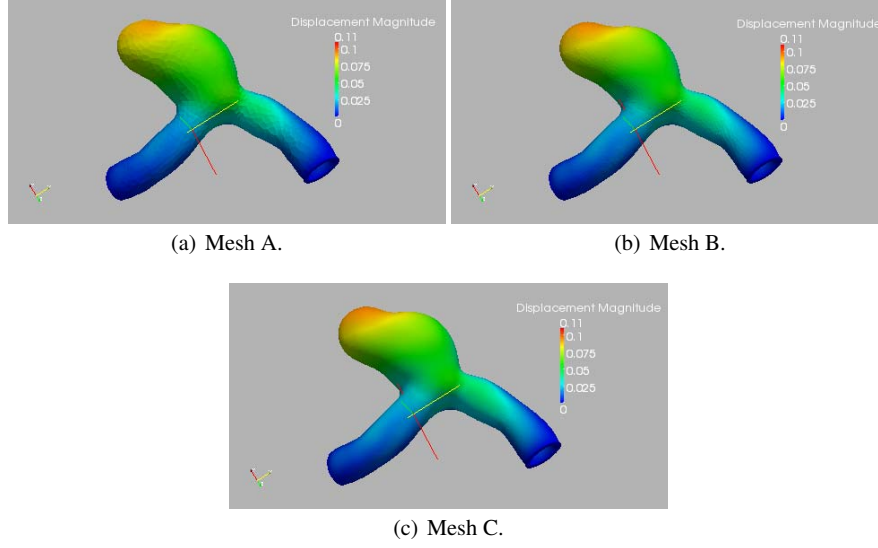


Fig. 3.25 Displacement magnitude field result for intracranial aneurysm.

new, constitutive equations for arterial wall modeling. The mechanical equilibrium is achieved in the deformed configuration through a Newton-Raphson method, while the generic approach is accomplished by means of a second-order finite difference approach to compute stresses and stiffness matrixes. Therefore, the only user-intervention within the code turns out to be the programming of the strain energy function (a scalar function) whose arguments are the invariants of the strain tensors. Several examples have been presented validating the computational implementation.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. The support of these agencies is gratefully acknowledged.

Appendix

Consider a displacement field \mathbf{u}_m which is perturbed producing $\mathbf{u}_{m,\tau} = \mathbf{u}_m + \tau \delta \mathbf{u}_m$. With this perturbation we have that the deformation gradient, originally given by $\mathbf{F}_m = \mathbf{I} + \nabla_{\mathbf{x}_m} \mathbf{u}_m$ results in $\mathbf{F}_{m,\tau} = \mathbf{I} + \nabla_{\mathbf{x}_m} \mathbf{u}_{m,\tau} = \mathbf{I} + \nabla_{\mathbf{x}_m} (\mathbf{u}_m + \tau \delta \mathbf{u}_m)$. Let us compute the expression of the derivatives of several quantities involving $\mathbf{F}_{m,\tau}$, with respect to τ . This will be employed in the linearization procedures whenever the ma-

terial configuration is known. From continuum mechanics we have

$$\left. \frac{d}{d\tau} \mathbf{F}_{m,\tau} \right|_{\tau=0} = \nabla_{\mathbf{x}_m} \delta \mathbf{u}_m = (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m \mathbf{F}_m, \quad (3.88)$$

$$\left. \frac{d}{d\tau} \mathbf{F}_m^{-1} \right|_{\tau=0} = -\mathbf{F}_m^{-1} (\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m) \mathbf{F}_m^{-1} = -\mathbf{F}_m^{-1} (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m, \quad (3.89)$$

$$\left. \frac{d}{d\tau} \det \mathbf{F}_{m,\tau} \right|_{\tau=0} = \det \mathbf{F}_m (\mathbf{F}_m^{-T} \cdot \nabla_{\mathbf{x}_m} \delta \mathbf{u}_m) = \det \mathbf{F}_m (\operatorname{div}_{\mathbf{x}_s} \delta \mathbf{u}_s)_m, \quad (3.90)$$

$$\begin{aligned} \left. \frac{d}{d\tau} \mathbf{E}_{m,\tau} \right|_{\tau=0} &= \frac{1}{2} ((\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m)^T \mathbf{F}_m + \mathbf{F}_m^T (\nabla_{\mathbf{x}_m} \delta \mathbf{u}_m)) \\ &= \frac{1}{2} \mathbf{F}_m^T ((\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m^T + (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)_m) \mathbf{F}_m \\ &= \mathbf{F}_m^T \boldsymbol{\varepsilon}_m(\delta \mathbf{u}_s) \mathbf{F}_m. \end{aligned} \quad (3.91)$$

Now, consider a displacement field \mathbf{u}_s which is perturbed $\mathbf{u}_{s,\tau} = \mathbf{u}_s + \tau \delta \mathbf{u}_s$. Then the deformation gradient, originally given by $\mathbf{F}_s^{-1} = \mathbf{I} - \nabla_{\mathbf{x}_s} \mathbf{u}_s$ results in $\mathbf{F}_{s,\tau}^{-1} = \mathbf{I} - \nabla_{\mathbf{x}_s} \mathbf{u}_{s,\tau} = \mathbf{I} - \nabla_{\mathbf{x}_s} (\mathbf{u}_s + \tau \delta \mathbf{u}_s)$. The derivatives of several quantities involving $\mathbf{F}_{s,\tau}$, with respect to τ are

$$\left. \frac{d}{d\tau} \mathbf{F}_{s,\tau}^{-1} \right|_{\tau=0} = -\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s, \quad (3.92)$$

$$\left. \frac{d}{d\tau} \mathbf{F}_{s,\tau} \right|_{\tau=0} = \mathbf{F}_s \nabla_{\mathbf{x}_s} \delta \mathbf{u}_s \mathbf{F}_s, \quad (3.93)$$

$$\begin{aligned} \left. \frac{d}{d\tau} \det \mathbf{F}_s \right|_{\tau=0} &= \det \mathbf{F}_s (\mathbf{F}_s^{-T} \cdot \nabla_{\mathbf{x}_s} \delta \mathbf{u}_s \mathbf{F}_s) \\ &= \det \mathbf{F}_{s,\tau} (\mathbf{F}_s^T \cdot \nabla_{\mathbf{x}_s} \delta \mathbf{u}_s), \end{aligned} \quad (3.94)$$

$$\begin{aligned} \left. \frac{d}{d\tau} \mathbf{E}_{s,\tau} \right|_{\tau=0} &= \frac{1}{2} (\mathbf{F}_s^T (\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)^T \mathbf{F}_s^T \mathbf{F}_s + \mathbf{F}_s^T \mathbf{F}_s \nabla_{\mathbf{x}_s} \delta \mathbf{u}_s \mathbf{F}_s) \\ &= \frac{1}{2} \mathbf{F}_s^T ((\nabla_{\mathbf{x}_s} \delta \mathbf{u}_s)^T \mathbf{F}_s^T + \mathbf{F}_s \nabla_{\mathbf{x}_s} \delta \mathbf{u}_s) \mathbf{F}_s. \end{aligned} \quad (3.95)$$

References

- [1] Blanco, P., Feijóo, R., Urquiza, S.: A unified variational approach for coupling 3D-1D models and its blood flow applications. *Comp. Meth. Appl. Mech. Eng.* **196**, 4391–4410 (2007).
- [4] Blanco, P., Pivello, M., Urquiza, S., Feijóo, R.: On the potentialities of 3D–1D coupled models in hemodynamics simulations. *J. Biomech.* **42**, 919–930 (2009).
- [3] Blanco, P., Urquiza, S., Feijóo, R.: Assessing the influence of heart rate in local hemodynamics through coupled 3D-1D-0D models. *Int. J. Num. Meth. Biomed. Eng.* **26**, 890–903 (2010).
- [4] Delfino, A., Stergiopoulos, N., Moore, J., Meister, J.: Residual strain effects on the stress field in a thick wall finite element model of the human carotid bifurcation. *J. Biomech.* **30**,

- 777–786 (1997).
- [5] Fukunari, H., Oshima, M., Torii, R., Watanabe, H., Hisada, T.: Fluid-structure interaction finite element analysis of middle cerebral artery aneurysm. *J. Biomech.* **39**, S364 (2006).
 - [15] Grinberg, L., Anor, T., Madsen, J., Yakhot, A., Karniadakis, G.: Large-scale simulation of the human arterial tree. *Clinical and Experimental Pharmacology and Physiology* **36**, 194–205 (2009).
 - [7] Holzapfel, G., Gasser, T.: A viscoelastic model for fiber-reinforced composites at finite strains: Continuum basis, computational aspects and applications. *Comp. Meth. Appl. Mech. Eng.* **190**, 4379–4403 (2001).
 - [8] Holzapfel, G., Gasser, T., Ogden, R.: A new constitutive framework for arterial wall mechanics and a comparative study of material models. *J. Elastic.* **61**, 1–48 (2000).
 - [9] Holzapfel, G., Ogden, R.: Constitutive modelling of arteries. *Proc. R. Soc. A* **466**, 1551–1597 (2010).
 - [10] Horgan, C., Saccomandi, G.: A new constitutive theory for fiber-reinforced incompressible nonlinearly elastic solids. *Journal of the Mechanics and Physics of Solids* **53**, 1985–2015 (2005).
 - [11] Humphrey, J., Canham, P.: Structure, mechanical properties, and mechanics of intracranial saccular aneurysms. *J. Elastic.* **61**, 49–81 (2000).
 - [12] Humphrey, J., Holzapfel, G.: Mechanics, mechanobiology, and modeling of human abdominal aorta and aneurysms. *J. Biomech.* **45**, 805–814 (2012).
 - [13] Humphrey, J., Taylor, C.: Intracranial and abdominal aortic aneurysms: similarities, differences, and need for a new class of computational models. *Annu. Rev. Biomed. Eng.* **10**, 221–246 (2008).
 - [14] Hunter, P., Nielsen, P.: A strategy for integrative computational physiology. *Physiology* **20**, 316–325 (2005).
 - [15] Hyun, S., Kleinstreuer, C., Archie, J.: Hemodynamics analyses of arterial expansions with implications to thrombosis and restenosis. *Med. Eng. Phys.* **22**, 13–27 (2000).
 - [16] Malvern, L.: *Introduction to the Mechanics of a Continuous Medium*. Englewood Cliffs, N J, Prentice-Hall, Inc. (1969).
 - [17] Perktold, K., Rappitsch, G.: Computer simulation of local blood flow and vessel mechanics in a compliant carotid artery bifurcation model. *J. Biomech.* **28**, 845–856 (1995).
 - [18] Peña, E., Alastrué, V., Laborda, A., Martínez, M., Doblaré, M.: A constitutive formulation of vascular tissue mechanics including viscoelasticity and softening behaviour. *J. Biomech.* **43**, 984–989 (2010).
 - [19] Ricotta, J., Pagan, J., Xenos, M., Alemum, Y., Einav, S., Bluestein, D.: Cardiovascular disease management: the need for better diagnostics. *Med. Biol. Eng. Comput.* **46**, 1059–1068 (2008).
 - [20] Torii, R., Oshima, M., Kobayashi, T., Takagi, K., Tezduyar, T.: Computer modeling of cardiovascular fluid-structure interactions with the deforming-spatial-domain/stabilized space-time formulation. *Comp. Meth. Appl. Mech. Eng.* **195**, 1885–1895 (2006).
 - [31] Urquiza, S., Blanco, P., Vénere, M., Feijóo, R.: Multidimensional modeling for the carotid blood flow. *Comp. Meth. Appl. Mech. Eng.* **195**, 4002–4017 (2006).
 - [22] Vassoler, J., Reips, L., Fancello, E.: A variational framework for fiber-reinforced viscoelastic soft tissues. *Int. J. Num. Meth. Biomed. Eng.* **bf 89**, 1691–1706 (2012).
 - [23] Volokh, K.: Hyperelasticity with softening for modeling materials failure. *Journal of the Mechanics and Physics of Solids* **55**, 2237–2264 (2007).
 - [24] Zulliger, M., Fridez, P., Hayashi, K., Stergiopoulos, N.: A strain energy function for arteries accounting for wall composition and structure. *J. Biomech.* **37**, 989–1000 (2004).

Chapter 4

Black-box decomposition methods to model blood flow in the cardiovascular system using dimensionally-heterogeneous models

Pablo J. Blanco, Jorge S. Leiva and Gustavo C. Buscaglia

Abstract In this work, an efficient black-box approach for the strong iterative coupling of dimensionally-heterogeneous hemodynamics models is presented. A heterogeneous model of the cardiovascular system is formed by several components whose aims are to simulate the behavior of the different vascular entities. These components are ultimately connected through coupling equations, which constitute a system of non-linear equations. The assembling of the system of equations is presented within the context of coupling complex and simple components. To solve this system of equations a Broyden-based algorithm is proposed. This algorithm is employed to split a 3D-1D-0D closed-loop model of the cardiovascular system into corresponding black-box components standing for the 3D (specific vessels), 1D (systemic arteries/peripheral vessels) and 0D (venous/cardiac/pulmonary circulation) components. Several examples of application in computational hemodynamics are presented showing the robustness and suitability of this novel approach.

Pablo Javier Blanco

HeMoLab - Hemodynamics Modeling Laboratory, INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing and LNCC - National Laboratory for Scientific Computing, Av. Getúlio Vargas 333, 25651-075, Petrópolis, Brazil, e-mail: pjblanco@lncc.br

Jorge Sebastián Leiva

Balseiro Institute, and Av. Bustillo Km. 9.5, 8400, San Carlos de Bariloche, Argentina, e-mail: leivaj@ib.cnea.gov.ar

Gustavo Carlos Buscaglia

Institute of Mathematical and Computational Sciences, São Paulo University and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Av. do Trabalhador Sãocarlense 400, 13560-970, São Carlos, Brazil, e-mail: gustavo.buscaglia@icmc.usp.br

4.1 Introduction

The coupling of models of different dimensionality in fluid dynamics was firstly addressed in [27]. Alternative formulations and applications have been presented subsequently, proving the usefulness of such approach to model complex systems. Generally speaking, such situation arises when a complex system, such as a hydraulic network, is split into different subsystems with different characteristic geometrical scales. Complex components of the system must be modeled using the full Navier–Stokes equations whereas some approximation is acceptable in other, simpler components, such as long pipes or valves, for example. In general, the global dynamics results from the interaction among the parts of the whole network, requiring the use of coupled dimensionally-heterogeneous models.

In turn, recent years have witnessed the efforts of scientific community directed towards the integration of different mathematical models of the cardiovascular system (CVS) [3, 13, 31, 32]. Particularly, the use of dimensionally-heterogeneous models has become customary in hemodynamics applications [4, 5, 6, 8, 9, 14, 15, 16, 20, 25].

Unlike pure 3D models, dimensionally-heterogeneous models allow to incorporate in a natural manner the complex effects of the vasculature present upstream and downstream of an artificially isolated 3D region. This means that we are able to consider the hemodynamics environment posed by the global state of the CVS. Therefore, non-stationary conditions as those encountered when considering regulation mechanisms [8], or analyses in multiple cardiovascular scenarios [5] can be naturally accounted for through the use of dimensionally-heterogeneous models. In these examples, standalone 3D models would require boundary data which has to be somehow provided (i) for all the artificial interfaces generated by the isolation of the 3D geometry, and (ii) for the different varying conditions. Either in cases with several 3D artificial inlets/outlets (consider, for instance, the full aorta artery, or multiple 3D models in the same analysis) or in the non-periodic scenarios mentioned above, this is nowadays unfeasible.

The CVS is understood as being formed by the coupling of components, each of which accounts for different vascular entities, e.g. arterial vessels, peripheral beds, cardiac chambers and so on. In this context, these components are associated with black-box codes specifically developed to deal with their own mathematical characteristics.

A black-box approach for the modeling of the CVS presents several advantages in contrast to a monolithic approach:

- i. there is no need to have full access to the computational codes;
- ii. it is based on simple input/output relations of well-validated codes;
- iii. the impact of non-linearities of each component are bounded to each specific component;
- iv. the time-discretization is carried out according to the specific requirements of each component.

Let us exemplify the last point listed above. Consider that the most expensive component is non-linear and takes 3-4 non-linear iterations to converge in a standard simulation (e.g., the Navier-Stokes equations in the present context). Consider also that this component is coupled to a highly non-linear component, but far cheaper to be solved, which takes 10-15 iterations to converge (e.g., the 0D model for the cardiac valves in the present context). In a monolithic approach, the cheaper component forces the expensive one to be solved more times than needed, making the computation more expensive than necessary. Similarly, in a monolithic setting the time-step is necessarily the same for all the components. This fact may signify unnecessary expensive computations if the required time step for a 3D model to advance in time is larger than the time-step needed by the rest of the -cheaper- components. In contrast to the monolithic approach, in a black-box setting these restrictions on the time-step, which depend upon the time-advancing schemes chosen for each component, do not affect the time-step of the coupled problem.

The first works to address the iterative solution of heterogeneous models made use of explicit methods as well as of implicit Gauss-Seidel-like methods with sub-relaxation [3, 13, 25, 26]. The common feature of those approaches is that it is necessary to define a hierarchy in the network of components regarding boundary data. Assume that we are coupling a 1D with a 3D model and that we use a Gauss-Seidel method (the same holds for Jacobi, or for explicit methods) to solve the problem in an iterative manner. Suppose that we decide to impose the pressure as boundary data to the 1D model. This model will give back a certain flow rate that is to be imposed as boundary data on the 3D model, from which a certain pressure is going to be retrieved to be subsequently applied to the 1D model, and so on. This procedure poses restrictions on the choice of boundary data to be imposed on the different components, something which is known as subdomain coloring. Moreover, Gauss-Seidel-like methods have poor convergence properties.

Recent developments for the strong coupling of heterogeneous components have been presented in [22]. In [23], more sophisticated iterative strategies for the decomposition of hydraulic networks in the context of incompressible fluids and rigid pipes has been presented. Comparisons between Gauss-Seidel, Newton-GMRES and Broyden methods are presented in the aforementioned work, from which it is concluded that the Broyden method turns out to be the best choice in terms of computational cost to solve dimensionally-heterogeneous flow models. Similar methodologies were used in [6] to carry out the decomposition of 1D networks for the solution of the arterial blood flow in the entire arterial system.

In the present work we present a detailed description of the development of a coupling strategy to solve in an iterative manner the interaction between dimensionally-heterogeneous models. This methodology is employed to address the solution of a dimensionally-heterogeneous closed-loop model of the CVS through the iterative solution of standalone components. Specifically, we propose an algorithm based on the Broyden method and decompose the CVS into components which stand for: (i) blood flow in specific vessels (3D models), (ii) wave propagation phenomena in systemic arteries (1D models) and peripheral beds, and (iii) venous, cardiac and

pulmonary circulations (0D models), forming thus a 3D-1D-0D closed-loop model of the CVS.

Components used in the description of the CVS are connected through coupling - interface- equations; the full problem is rewritten in terms of interface unknowns; an input/output reasoning is applied to define the character of the interface unknowns for each black-box component, and the Broyden-based iterative method is used to solve the associated system of non-linear equations. The performance and robustness of the proposed methodology are assessed through several examples of application.

The work is organized as follows. Section 4.2 the basic concepts are introduced and put to work in a simple problem. In Section 4.3 the setting of the dimensionally-heterogeneous problem using flow models is given, and in Section 4.4 the concepts behind the partitioning procedure are presented. In Section 4.5 the mathematical models and the black-box concept are described in the hemodynamics context, while in Section 4.6 the algorithm for the iterative solution is presented. The application to practical situations is given in Section 5.6, and the final remarks are given in Section 4.8.

4.2 Underlying concepts for the partitioning strategy

Let us exemplify the core ideas in a simple example. Consider the diffusion problem defined in the 1D domain of length L shown in Figure 4.1; i.e.,

$$\begin{cases} -\frac{d}{dx}\left(k\frac{dU}{dx}\right) = f & \text{in } (0,L), \\ U(0) = 0, \\ U(L) = 0. \end{cases} \quad (4.1)$$

To keep the exposition as simple as possible, let us assume that k and f are constants, and that homogeneous Dirichlet boundary conditions are applied at the boundaries.

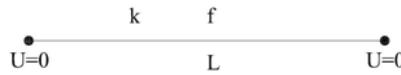


Fig. 4.1 Diffusion problem in a 1D domain.

We decompose the domain into two non-overlapping regions $\Omega_1 = (0, c)$, $\Omega_2 = (c, L)$. Our aim, as in all domain decomposition methods, is to solve problem (4.1) by solving separate problems in Ω_1 and Ω_2 . For the partitioned solution (U_1, U_2) to coincide with the exact one, in the sense that $U_1 = U|_{\Omega_1}$ and $U_2 = U|_{\Omega_2}$, it is necessary and sufficient that both functions, together with their fluxes, match at point c . In other words, defining the (unknown) *coupling variables* u_c and q_c , the

following must hold

$$\begin{aligned} U_1(c) &= U_2(c) = u_c, \\ -k \frac{dU_1}{dx}(c) &= -k \frac{dU_2}{dx}(c) = q_c. \end{aligned} \quad (4.2)$$

In a more general case, we would have the situation where the sub-domains have different physical parameters, that is $k_1 = k(c^-) \neq k(c^+) = k_2$. We take the case $k_1 = k_2 = k$ for the sake of simplicity. We focus first on Ω_1 , in which we will solve the original equation (4.1)₁ with boundary condition (4.1)₂, but we are missing a boundary condition at $x = c$. Let us decide, arbitrarily, that $x = c$ will be a *Dirichlet* boundary for Ω_1 . If we impose a value γ , the solution w_γ of

$$\begin{cases} -\frac{d}{dx} \left(k \frac{dw_\gamma}{dx} \right) = f & \text{in } (0, c), \\ w_\gamma(0) = 0, \\ w_\gamma(c) = \gamma, \end{cases} \quad (4.3)$$

will depend on γ , as implied by the subscript. Let us denote by $\mathcal{Q}_c^1(\gamma)$ the flux at $x = c$ corresponding to the solution w_γ , i.e.,

$$\mathcal{Q}_c^1(\gamma) = -k \frac{dw_\gamma}{dx}(c). \quad (4.4)$$

It is then clear that, for (q_c, u_c) to satisfy (4.2), they must satisfy the equation

$$-q_c + \mathcal{Q}_c^1(u_c) = 0. \quad (4.5)$$

Let us now turn to Ω_2 , and arbitrarily decide that we want $x = c$ to be a *Neumann* boundary for it. Proceeding analogously as above, let z_δ be the solution of

$$\begin{cases} -\frac{d}{dx} \left(k \frac{dz_\delta}{dx} \right) = f & \text{in } (c, L), \\ z_\delta(L) = 0, \\ -k \frac{dz_\delta}{dx}(c) = \delta, \end{cases} \quad (4.6)$$

and let us denote by $\mathcal{U}_c^2(\delta)$ the value of z_δ at $x = c$, i.e.,

$$\mathcal{U}_c^2(\delta) = z_\delta(c). \quad (4.7)$$

With this notation, it is clear that (q_c, u_c) must satisfy

$$-\mathcal{U}_c^2(q_c) + u_c = 0. \quad (4.8)$$

In our simple example \mathcal{Q}_c^1 and \mathcal{U}_c^2 can be calculated explicitly. They are given by the affine functions

$$\mathcal{Q}_c^1(\gamma) = -\frac{k}{c}\gamma + \frac{cf}{2}, \quad (4.9)$$

$$\mathcal{U}_c^2(\delta) = \frac{L-c}{k}\delta + \frac{f(L-c)^2}{2k}. \quad (4.10)$$

We thus arrive at a system of two equations in the two unknowns q_c and u_c

$$\begin{aligned} r_1(q_c, u_c) &:= -q_c + \mathcal{Q}_c^1(u_c) = 0, \\ r_2(q_c, u_c) &:= -\mathcal{U}_c^2(q_c) + u_c = 0. \end{aligned} \quad (4.11)$$

We group the unknowns into the vector $X = (q_c, u_c)$ and the residuals into the vector $r(X) = (r_1(X), r_2(X))$. Whether these are row or column vectors will be obvious from the context.

Up to now we have made two arbitrary choices, namely to take $x = c$ as Dirichlet boundary for Ω_1 and as Neumann boundary for Ω_2 . Let us make a third arbitrary choice and solve the system above by the *Gauss-Seidel* method starting from $X_0 = (q_c^0, u_c^0)$. This leads to the iterative method

$$\begin{aligned} q_c^{k+1} &= \mathcal{Q}_c^1(u_c^k), \\ u_c^{k+1} &= \mathcal{U}_c^2(q_c^{k+1}), \end{aligned} \quad (4.12)$$

which is nothing but the classical *Dirichlet-to-Neumann* (D-to-N) method of domain decomposition.

Notice that, in the simple case in which k and f are constants, eliminating q_c^{k+1} from the previous equations (using (4.12) along with (4.9) and (4.10)) leads to

$$u_c^{k+1} = -\frac{L-c}{c}u_c^k + \frac{fL(L-c)}{2k}, \quad (4.13)$$

implying, since $u_c = \frac{fc}{2k}(L-c)$, that

$$u_c^{k+1} - u_c = -\frac{L-c}{c}(u_c^k - u_c),$$

so that if $c > L/2$ the method will converge linearly, with constant $(L-c)/c$, and if $c \leq L/2$ it will not converge. Taking, for example, $L = 3$, $c = f = k = 1$, $q_c^0 = u_c^0 = 0$, leads to $q_c^1 = 1/2$, $u_c^1 = 3$, $q_c^2 = -5/2$, $u_c^2 = -3$, etc., diverging from the exact solution $q_c = -1/2$, $u_c = 1$.

The reinterpretation of the D-to-N method introduced above allows us to study many alternatives simply by changing the choices made along the derivation. Let us work out some of these alternatives in detail.

Consider for example changing the Gauss-Seidel solution of (4.11) by the following matrix-free GMRES technique:

1. Given X_0 , compute $r(X_0)$, set $Z = \|r(X_0)\|_2$, and $v_1 = r(X_0)/Z$.
2. For $j = 1, \dots, m$, do
3. Compute $w_j = r(X_0) - r(X_0 + v_j)$

4. For $i = 1, \dots, j$, do
5. $H_{ij} = (w_j, v_i)$
6. $w_j = w_j - H_{ij}v_i$
7. Enddo
8. $H_{j+1,j} = \|w_j\|_2$. If $H_{j+1,j} = 0$ set $m = j$ and goto 11
9. $v_{j+1} = w_j / H_{j+1,j}$
10. Enddo
11. Define the $(m+1) \times m$ Hessenberg matrix $H_m = \{H_{ij}\}_{1 \leq i \leq m+1, 1 \leq j \leq m}$
12. Compute y_m , the minimizer of $\|Ze_1 - H_m y\|_2$
13. Do $X_m = X_0 + V_m y_m$, where the columns of V_m are the vectors v_j computed through the iterations.

With the same data as before, and in particular $X_0 = (q_c^0, u_c^0) = (0, 0)$, one obtains

$$r(X_0) = \begin{pmatrix} \frac{1}{2} \\ -2 \end{pmatrix}, \quad v_1 = \frac{1}{\sqrt{17}} \begin{pmatrix} 1 \\ -4 \end{pmatrix},$$

to calculate $r(X_0)$ we solved on Ω_1 with Dirichlet boundary condition equal to u_c^0 and on Ω_2 with Neumann boundary condition equal to q_c^0 . We now need to calculate $r(X_0 + v_1)$. For this purpose we solve on Ω_1 with Dirichlet condition equal to $(X_0 + v_1)_2 = -\frac{4}{\sqrt{17}}$ and on Ω_2 with Neumann condition equal to $(X_0 + v_1)_1 = \frac{1}{\sqrt{17}}$, this leads to

$$\begin{aligned} w_1 &= r(X_0) - r(X_0 + v_1) = \frac{1}{\sqrt{17}} \begin{pmatrix} -3 \\ 6 \end{pmatrix}, & H_{11} &= (w_1, v_1) = -\frac{27}{17}, \\ w_1 &= w_1 - H_{11}v_1 = -\frac{6}{17^{\frac{3}{2}}} \begin{pmatrix} 4 \\ 1 \end{pmatrix}, & H_{21} &= \frac{6}{17}, \\ v_2 &= -\frac{1}{\sqrt{17}} \begin{pmatrix} 4 \\ 1 \end{pmatrix}. \end{aligned}$$

We now calculate $r(X_0 + v_2)$ imposing $(X_0 + v_2)_2$ as Dirichlet condition for Ω_1 and $(X_0 + v_2)_1$ as Neumann condition for Ω_2 . Notice that the GMRES algorithm does not impose to the Neumann domain the flux obtained from the Dirichlet domain or viceversa. The consequence is that, once the orthogonalization steps 4–7 are completed, we end up with $w_2 = 0$ (which was expected, since the space is two-dimensional). As a result, step 12 yields

$$\begin{pmatrix} -\frac{27}{17} & \frac{23}{17} \\ \frac{6}{17} & \frac{27}{17} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{17}}{2} \\ 0 \end{pmatrix} \Rightarrow y = \frac{1}{2\sqrt{17}} \begin{pmatrix} -9 \\ 2 \end{pmatrix},$$

leading to

$$X_2 = X_0 + y_1 v_1 + y_2 v_2 = \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} q_c \\ u_c \end{pmatrix},$$

the exact result.

The point we have made in the simple exercise above is that, keeping the same choices of boundary conditions as in the D-to-N method (Dirichlet for Ω_1 , Neumann for Ω_2), by changing the solver to GMRES we obtain convergence in two iterations for any choice of parameters and of initial guess.

Let us now keep the GMRES solver and change the boundary condition to Dirichlet for both subdomains. The system now becomes

$$\begin{aligned} r_1(q_c, u_c) &:= -q_c + \mathcal{Q}_c^1(u_c) = 0, \\ r_2(q_c, u_c) &:= -q_c + \mathcal{Q}_c^2(u_c) = 0, \end{aligned} \quad (4.14)$$

where $\mathcal{Q}_c^2(u_c)$ is the flux obtained at $x = c$ when the equation is solved in Ω_2 with Dirichlet condition (at $x = c$) with value u_c . In this very simple case we have

$$\mathcal{Q}_c^2(\gamma) = \frac{k}{L-c} \gamma - \frac{f(L-c)}{2}. \quad (4.15)$$

Taking the same values as before, we get $r(X_0) = (\frac{1}{2}, -1)$. Performing all the computations we get, after just one iteration (due to the particular choice of initial conditions) $X_1 = (-\frac{1}{2}, 1)$, which is again exact.

Classical approaches in the domain decomposition literature amount to solving the following equations (without having considered preconditioning for the sake of simplicity)

$$Su_c = \chi \quad \text{Steklov-Poincaré equation,} \quad (4.16)$$

and

$$Rq_c = \xi \quad \text{FETI equation.} \quad (4.17)$$

Such interface problems are particular cases of the approach introduced by (4.11). Indeed, from system (4.14) (equivalent to system (4.11)) one obtains

$$\mathcal{Q}_c^1(u_c) - \mathcal{Q}_c^2(u_c) = 0, \quad (4.18)$$

which is nothing but the Steklov-Poincaré equation (4.16). The FETI equation is also easily obtained from the proposed methodology if Neumann conditions are applied to both subdomains and the unknown u_c is eliminated. If one wants the choice of boundary conditions to remain arbitrary, however, both unknowns must be kept.

The added flexibility comes with an additional cost: The number of unknowns is doubled and the system is always non-symmetric, even if the underlying problem is symmetric. It is for this reason that we find the proposed method especially well suited for dimensionally-heterogeneous coupling, since in this area the number of interfaces is typically small. Also notice that we do not discuss preconditioning of the system, which we have not investigated and expect not to be necessary in most cases due again to the small size of the system.

4.3 Mathematical setting for dimensionally-heterogeneous models

This section extends the ingredients introduced in Section 4.2 for the coupling of dimensionally-heterogeneous flow models. In this case we address a simplified 3D-0D setting. Respectively, the 3D and the 0D components are denoted complex and simple components.

For the flow models presented in this section we assume incompressibility of the fluid and also rigidity of the pipes. This is by no means restrictive from the point of view of the domain decomposition approach. Indeed, compliant vessels are explored in the cardiovascular setting. This is why the formulation is not oversimplified in some parts where it could be, but it is presented in a more general framework so as to embrace more general situations like the one encountered in hemodynamics.

In order to introduce the model we resort to the simplest situation shown in Figure 4.2 where the interaction between two systems with different leading geometrical scales is shown. Also, we employ graph notation, denoting by \mathcal{N} the nodes (or complex models), by \mathcal{C} the connections (or simple models) and by \mathcal{I} the coupling interfaces. The *Times* font is used for the elements described with the calligraphic font within the text.

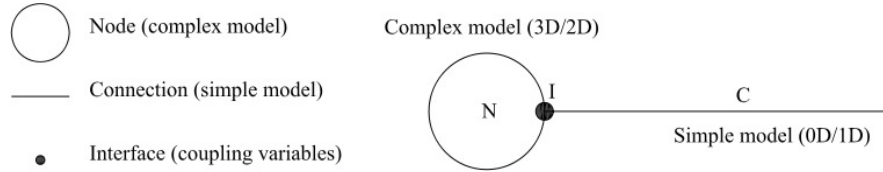


Fig. 4.2 Coupled system consisting of two geometrical scales.

System \mathcal{N} has complex dynamics and is modeled via the full Navier–Stokes equations (2D or 3D) and thus \mathcal{N} denotes a spatial domain $\Omega \in \mathbb{R}^n$ ($n = 2, 3$). In order to have existence and uniqueness of the solution we will always assume that the data is sufficiently small. In turn, system \mathcal{C} has simple dynamics and is represented through the condensed Navier–Stokes equations (particularly it can be 0D models when no change of area is considered or 1D models when the pipes are deformable).

Over the coupling interface \mathcal{I} , which denotes a given portion of the boundary $\partial\Omega$ of Ω , continuity of mean velocity (or equivalently mass flow) and of normal traction are assumed to hold. The governing equations are (here we consider 0D models for the simple components):

$$\begin{cases}
\rho \frac{\partial \mathbf{u}}{\partial t} + \rho (\nabla \mathbf{u}) \mathbf{u} + \nabla p - \mu \Delta \mathbf{u} = 0 & \text{in } \mathcal{N}, \\
\operatorname{div} \mathbf{u} = 0 & \text{in } \mathcal{N}, \\
\mathbf{u} = 0 & \text{on } \partial \Omega \setminus \Gamma, \\
Q_1 = - \int_{\Gamma} \mathbf{u} \cdot \mathbf{n} d\Gamma & \text{in } \mathcal{I}, \\
-P_1 \mathbf{n} = \frac{1}{|\Gamma|} \int_{\Gamma} (-p \mathbf{I} + \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)) \mathbf{n} d\Gamma & \text{in } \mathcal{I}, \\
M \frac{dQ_1}{dt} + K Q_1 - (P_2 - P_1) - B = 0 & \text{in } \mathcal{C}, \\
Q_1 + Q_2 = 0 & \text{in } \mathcal{C}, \\
P_2 = P_{\text{ref}} & \text{in } \mathcal{O},
\end{cases} \quad (4.19)$$

where Γ is the coupling interface corresponding to \mathcal{I} with unit outward normal \mathbf{n} (seen from system \mathcal{N}), (\mathbf{u}, p) is the pair velocity-pressure in \mathcal{N} , ρ and μ are the density and dynamic viscosity. Note that a Dirichlet boundary condition is imposed over the complementary boundary to the coupling interface $\partial \Omega \setminus \Gamma$. Also, Q_1 and Q_2 are the flow rates at both boundaries (ends) of the \mathcal{C} component. Here note that just one flow rate should be necessary, however, as said before, we keep both quantities so the ideas can be straightforwardly applied to more general contexts. The drop pressure is given by $\Delta P = P_2 - P_1$, noting that here we imposed a pressure boundary condition to this component given by P_{ref} at the outlet boundary of the network, denoted by \mathcal{O} . Finally, B is a source term that represents a pump acting in such connection, and M and K are the inertial and friction coefficients, which read as follows

$$\begin{aligned}
M &= \frac{\rho L}{A}, \\
K &= \frac{8\pi\mu L}{A^2} && \text{circular pipe,} \\
K &= \frac{12\mu L}{a^3 H} && \text{parallel plates,}
\end{aligned} \quad (4.20)$$

where L is the length of the pipe, A is the sectional area of the pipe (in the case of a circular pipe), a is the separation between plates with depth equal to H (in the case of parallel plates). Over Γ continuity of flow rate and continuity of mean traction are considered.

The problem stated by equation (4.19) is however not well-posed. We thus make the choice

$$\frac{1}{|\Gamma|} \int_{\Gamma} (-p \mathbf{I} + \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)) \mathbf{n} d\Gamma = (-p \mathbf{I} + \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)) \mathbf{n} \quad \text{in } \mathcal{I}, \quad (4.21)$$

which implies that the quantity $(-p \mathbf{I} + \mu (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)) \mathbf{n}$ is constant across Γ . With this choice we are making use of the do-nothing approach and therefore we recover uniqueness (see [17]), so that we end up with the following problem for the coupled, dimensionally-heterogeneous system:

$$\left\{ \begin{array}{ll} \rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\nabla \mathbf{u})\mathbf{u} + \nabla p - \mu \Delta \mathbf{u} = 0 & \text{in } \mathcal{N}, \\ \operatorname{div} \mathbf{u} = 0 & \text{in } \mathcal{N}, \\ \mathbf{u} = 0 & \text{on } \partial\Omega \setminus \Gamma, \\ Q_1 = - \int_{\Gamma} \mathbf{u} \cdot \mathbf{n} d\Gamma & \text{in } \mathcal{I}, \\ -P_1 \mathbf{n} = (-p\mathbf{I} + \mu(\nabla \mathbf{u} + (\nabla \mathbf{u})^T))\mathbf{n} & \text{in } \mathcal{I}, \\ M \frac{dQ_1}{dt} + KQ_1 - (P_2 - P_1) - B = 0 & \text{in } \mathcal{C}, \\ Q_1 + Q_2 = 0 & \text{in } \mathcal{C}, \\ P_2 = P_{\text{ref}} & \text{in } \mathcal{O}. \end{array} \right. \quad (4.22)$$

In order to treat the different components following a black-box approach we choose now the interface variables as being (Q^1, P^1) , that is, the flow rate and pressure at the coupling interface. With this, the problem is augmented and reads

$$\left\{ \begin{array}{ll} \rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\nabla \mathbf{u})\mathbf{u} + \nabla p - \mu \Delta \mathbf{u} = 0 & \text{in } \mathcal{N}, \\ \operatorname{div} \mathbf{u} = 0 & \text{in } \mathcal{N}, \\ \mathbf{u} = 0 & \text{on } \partial\Omega \setminus \Gamma, \\ Q_{\mathcal{N},1} = - \int_{\Gamma} \mathbf{u} \cdot \mathbf{n} d\Gamma & \text{in } \mathcal{I}, \\ -P_{\mathcal{N},1} \mathbf{n} = (-p\mathbf{I} + \mu(\nabla \mathbf{u} + (\nabla \mathbf{u})^T))\mathbf{n} & \text{in } \mathcal{I}, \\ Q^1 + Q_{\mathcal{N},1} = 0 & \text{in } \mathcal{I}, \\ P^1 - P_{\mathcal{N},1} = 0 & \text{in } \mathcal{I}, \\ Q^1 - Q_{\mathcal{C},1} = 0 & \text{in } \mathcal{I}, \\ P^1 - P_{\mathcal{C},1} = 0 & \text{in } \mathcal{I}, \\ M \frac{dQ_{\mathcal{C},1}}{dt} + KQ_{\mathcal{C},1} - (P_{\mathcal{C},2} - P_{\mathcal{C},1}) - B = 0 & \text{in } \mathcal{C}, \\ Q_{\mathcal{C},1} + Q_{\mathcal{C},2} = 0 & \text{in } \mathcal{C}, \\ P_{\mathcal{C},2} = P_{\text{ref}} & \text{in } \mathcal{O}, \end{array} \right. \quad (4.23)$$

where we used $(Q_{\mathcal{N},1}, P_{\mathcal{N},1})$ and $(Q_{\mathcal{C},1}, P_{\mathcal{C},1})$ to denote the variables from the complex and simple models that must match the interface state variables (Q^1, P^1) .

This system can be extended to a generic graph-like network as the one shown in Figure 4.3. Let us say that it consists of n_T nodes (complex models), c_T connections (simple models), i_T interfaces (coupling interfaces) and o_T network outlets.

Therefore, as happened with equation (4.23), in this case we have

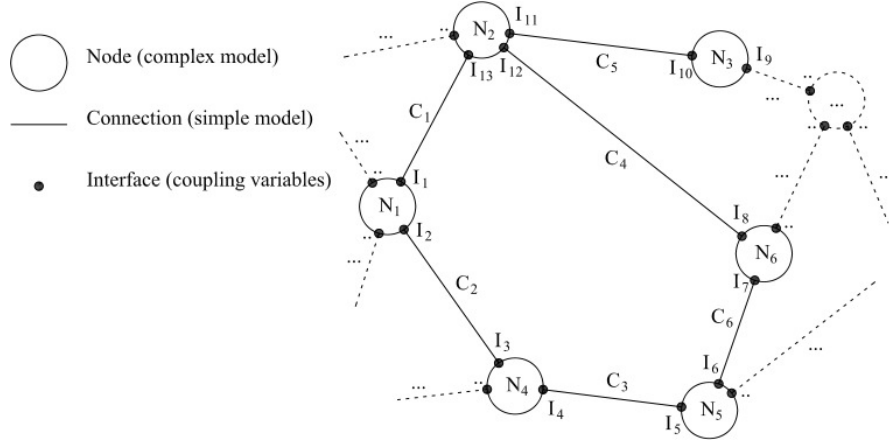


Fig. 4.3 Scheme of a generic coupled dimensionally-heterogeneous system.

$$\begin{cases}
 \rho \frac{\partial \mathbf{u}_n}{\partial t} + \rho (\nabla \mathbf{u}_n) \mathbf{u}_n + \nabla p_n - \mu \Delta \mathbf{u}_n = 0 & \text{in } \mathcal{N}_n, n = 1, \dots, n_T, \\
 \operatorname{div} \mathbf{u}_n = 0 & \text{in } \mathcal{N}_n, n = 1, \dots, n_T, \\
 \mathbf{u}_n = 0 & \text{on } \partial \Omega_n \setminus \Gamma_n, n = 1, \dots, n_T, \\
 Q_{\mathcal{N}_n, l} = - \int_{\Gamma_l} \mathbf{u}_n \cdot \mathbf{n}_l d\Gamma & \text{on } \Gamma_l, l = 1, \dots, N_n, n = 1, \dots, n_T, \\
 -P_{\mathcal{N}_n, l} \mathbf{I} = (-p_n \mathbf{I} + \mu (\nabla \mathbf{u}_n + (\nabla \mathbf{u}_n)^T)) \mathbf{n}_l & \text{on } \Gamma_l, l = 1, \dots, N_n, n = 1, \dots, n_T, \\
 Q^i + Q_{\mathcal{N}_n, l} |_{\mathcal{J}_i} = 0 & \text{in } \mathcal{J}_i, i = 1, \dots, i_T, \\
 P^i - P_{\mathcal{N}_n, l} |_{\mathcal{J}_i} = 0 & \text{in } \mathcal{J}_i, i = 1, \dots, i_T, \\
 Q^i - Q_{\mathcal{C}_c, k} |_{\mathcal{J}_i} = 0 & \text{in } \mathcal{J}_i, i = 1, \dots, i_T, \\
 P^i - P_{\mathcal{C}_c, k} |_{\mathcal{J}_i} = 0 & \text{in } \mathcal{J}_i, i = 1, \dots, i_T, \\
 M_c \frac{dQ_{\mathcal{C}_c, 1}}{dt} + K_c Q_{\mathcal{C}_c, 1} - (P_{\mathcal{C}_c, 2} - P_{\mathcal{C}_c, 1}) - B_c = 0 & \text{in } \mathcal{C}_c, c = 1, \dots, c_T, \\
 Q_{\mathcal{C}_c, 1} + Q_{\mathcal{C}_c, 2} = 0 & \text{in } \mathcal{C}_c, c = 1, \dots, c_T, \\
 P_{\mathcal{C}_c, k} |_{\mathcal{O}_o} = P_{\text{ref}}^o & \text{in } \mathcal{O}_o, o = 1, \dots, o_T.
 \end{cases} \quad (4.24)$$

Here $Q_{\mathcal{C}_c, k} |_{\mathcal{J}_i}$ and $P_{\mathcal{C}_c, k} |_{\mathcal{J}_i}$ are the flow rate and pressure in the simplified model \mathcal{C}_c over the coupling interface \mathcal{J}_i (with $k = 1, 2$ depending whether \mathcal{J}_i is an inlet or an outlet for the \mathcal{C}_c component). Analogously, $P_{\mathcal{C}_c, k} |_{\mathcal{O}_o}$ is the pressure in the simplified model over the network outlet \mathcal{O}_o (with $k = 1, 2$ depending whether \mathcal{O}_o is an inlet or an outlet for the \mathcal{C}_c component), which equals the data P_{ref}^o . Likewise, $Q_{\mathcal{N}_n, l} |_{\mathcal{J}_i}$ and $P_{\mathcal{N}_n, l} |_{\mathcal{J}_i}$ are the flow rate and pressure in the complex model \mathcal{N}_n over the local boundary Γ_l , and that corresponds to the coupling interface \mathcal{J}_i . We denote by N_n the number of interfaces owned by the complex model \mathcal{N}_n , while the number of interfaces owned by the simple model \mathcal{C}_c is trivially $N_c = 2$ (however, more general situations can be considered). Notice that subscripts denote indexes referring to lo-

cal numeration (component-specific) whereas superscripts refer to indexes in global numeration (interfaces and network outlets).

Problem (4.24) could be simplified by eliminating, for example, the unknowns $Q_{\mathcal{N}_n, l}|_{\mathcal{I}_i}$ and $P_{\mathcal{N}_n, l}|_{\mathcal{I}_i}$. Nonetheless, we keep them in order to put in evidence, when selected, the variables that are going to play the role of boundary conditions in the domain decomposition perspective that will be discussed in Section 4.4.

4.4 Partitioned strategy

Firstly, in this section the flow problem presented in the previous section is recast in terms of coupling (or interface) variables, as done in Section 4.2 for linear problems.

4.4.1 Interface variables problem

Generally speaking, the choice of the interface variables depends upon the underlying models of a given coupling interface, specifically upon the physical quantities for which some sense of conservation or continuity exists. When dealing with heterogeneous, or non-conforming media, besides the physical considerations other numerical-mathematical conditions have to be taken into account. Particularly, the trace of the approximation spaces employed at each side of the coupling interface. It is worth noting that there exist inf-sup-like conditions that pose the necessary requirements, regarding what was said above, to ensure existence and uniqueness of the solutions of the global system. Nevertheless, we will not dwell into these mathematical aspects, which are partially treated in [2, 10].

The simplest choice consists in taking as interface unknowns the variables corresponding to the simplest model among those who share a given coupling interface. In the cases developed in the present work such choice corresponds to flow rate Q^i and pressure P^i , $i = 1, \dots, i_T$. Hence, when talking about Neumann boundary conditions we understand the imposition of the pressure P^i , whilst by Dirichlet boundary conditions it is understood the imposition of the flow rate Q^i .

Rewriting problem (4.24) in terms of the interface variables $(\{P^i\}_{i=1}^{i_T}, \{Q^i\}_{i=1}^{i_T})$ requires the elimination of the internal degrees of freedom of the system. In order to do so, it is necessary to define the mappings that relate the interface variables among them, in the spirit of the classical Steklov–Poincaré operator found in linear problems. Let us isolate the treatment of each sub-model (complex and simple ones). In what follows the numeration of the quantities is given locally, that is a local numeration for each sub-model, without relation with the numeration given in the whole problem (4.24).

Firstly, for each complex model \mathcal{N}_n we have N_n coupling boundaries. It is assumed that there are N_n^Q coupling boundaries where the flow rate is imposed (the data at hand is $(\{Q_{\mathcal{N}_n, j}\}_{j \in S_n^Q})$, being S_n^Q the corresponding set of indexes and N_n^Q its

cardinality), whereas N_n^P is the number of coupling boundaries where a Neumann boundary condition, via the value of the pressure, is considered (the data at hand is $(\{P_{\mathcal{N}_n,j}\}_{j \in S_n^P})$, being S_n^P the corresponding set of indexes and N_n^P its cardinality). Then it is $N_n = N_n^Q + N_n^P$. Thus, for each complex model \mathcal{N}_n , from problem (4.24), we can write the following abstract functional relation among the unknowns $(\{P_{\mathcal{N}_n,j}\}_{j \in S_n^Q}, \{Q_{\mathcal{N}_n,j}\}_{j \in S_n^P})$ and the data $(\{P_{\mathcal{N}_n,j}\}_{j \in S_n^P}, \{Q_{\mathcal{N}_n,j}\}_{j \in S_n^Q})$ as follows

$$\begin{bmatrix} \{P_{\mathcal{N}_n,j}\}_{j \in S_n^Q} \\ \{Q_{\mathcal{N}_n,j}\}_{j \in S_n^P} \end{bmatrix} = \mathcal{G}_{\mathcal{N}_n} \left(\begin{bmatrix} \{Q_{\mathcal{N}_n,j}\}_{j \in S_n^Q} \\ \{P_{\mathcal{N}_n,j}\}_{j \in S_n^P} \end{bmatrix} \right). \quad (4.25)$$

Note that there is a time dependence in the definition of the operator $\mathcal{G}_{\mathcal{N}_n}$. This has been disregarded for notational simplicity. The functional relation given by this operator maps, at any time, $\mathbb{R}^{N_n^Q + N_n^P}$ onto itself, that is

$$\mathcal{G}_{\mathcal{N}_n} : \underbrace{(\mathbb{R} \times \dots \times \mathbb{R})}_{N_n^Q \text{ times}} \times \underbrace{(\mathbb{R} \times \dots \times \mathbb{R})}_{N_n^P \text{ times}} \rightarrow \underbrace{(\mathbb{R} \times \dots \times \mathbb{R})}_{N_n^Q \text{ times}} \times \underbrace{(\mathbb{R} \times \dots \times \mathbb{R})}_{N_n^P \text{ times}} \quad \forall t, \quad (4.26)$$

and amounts to the solution of the following boundary value problem (for a given complex model \mathcal{N}_n)

$$\begin{cases} \rho \frac{\partial \mathbf{u}_n}{\partial t} + \rho(\nabla \mathbf{u}_n) \mathbf{u}_n + \nabla p_n - \mu \Delta \mathbf{u}_n = 0 & \text{in } \mathcal{N}_n, \\ \operatorname{div} \mathbf{u}_n = 0 & \text{in } \mathcal{N}_n, \\ \mathbf{u}_n = 0 & \text{on } \partial \Omega_n \setminus \Gamma_n, \\ - \int_{\Gamma_j} \mathbf{u}_n \cdot \mathbf{n}_j d\Gamma = Q_{\mathcal{N}_n,j} & \text{on } \Gamma_j, j \in \mathcal{N}_n^Q, \\ -P_{\mathcal{N}_n,j} \mathbf{n}_j = (-p_n \mathbf{I} + \mu(\nabla \mathbf{u}_n + (\nabla \mathbf{u}_n)^T)) \mathbf{n}_j & \text{on } \Gamma_j, j \in \mathcal{N}_n^Q, \\ Q_{\mathcal{N}_n,j} = - \int_{\Gamma_j} \mathbf{u}_n \cdot \mathbf{n}_j d\Gamma & \text{on } \Gamma_j, j \in \mathcal{N}_n^P, \\ (-p_n \mathbf{I} + \mu(\nabla \mathbf{u}_n + (\nabla \mathbf{u}_n)^T)) \mathbf{n}_j = -P_{\mathcal{N}_n,j} \mathbf{n}_j & \text{in } \Gamma_j, j \in \mathcal{N}_n^P. \end{cases} \quad (4.27)$$

The mapping $\mathcal{G}_{\mathcal{N}_n}$ relates the imposed boundary conditions with the corresponding unknowns (dual variables). In vector notation expression (4.25) is

$$\begin{bmatrix} \mathbf{P}_{\mathcal{N}_n^Q} \\ \mathbf{Q}_{\mathcal{N}_n^P} \end{bmatrix} = \mathcal{G}_{\mathcal{N}_n} \left(\begin{bmatrix} \mathbf{Q}_{\mathcal{N}_n^Q} \\ \mathbf{P}_{\mathcal{N}_n^P} \end{bmatrix} \right), \quad (4.28)$$

where each component of the vector represents a local interface of the complex model \mathcal{N}_n .

Suppose $N_n^P = 0$ ($N_n = N_n^Q$), and that the problem is linear and homogeneous. Then, expression (4.28) becomes

$$\mathbf{P}_{N_n} = \mathcal{G}_{\mathcal{N}_n}^{SP}(\mathbf{Q}_{N_n}), \quad (4.29)$$

which is the classical Steklov–Poincaré operator that, for a given primal quantity returns the dual variable. Likewise, if $N_n^Q = 0$ ($N_n = N_n^P$) and under the same assumptions expression (4.28) becomes

$$\mathbf{Q}_{N_n} = [\mathcal{G}_{N_n}^{SP}]^{-1}(\mathbf{P}_{N_n}). \quad (4.30)$$

Depending upon the model under study, it is not always possible to find an explicit relation like (4.28), but it is possible to write an implicit relation as follows

$$\overline{\mathcal{F}}_{N_n} \left(\begin{bmatrix} \mathbf{P}_{N_n^Q} \\ \mathbf{Q}_{N_n^P} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{N_n^Q} \\ \mathbf{P}_{N_n^P} \end{bmatrix} \right) = 0. \quad (4.31)$$

With the purpose of simplifying the notation we reorder the variables as follows

$$\mathcal{F}_{N_n} \left(\begin{bmatrix} \mathbf{P}_{N_n^Q} \\ \mathbf{P}_{N_n^P} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{N_n^Q} \\ \mathbf{Q}_{N_n^P} \end{bmatrix} \right) = \mathcal{F}_{N_n}(\mathbf{P}_{N_n}, \mathbf{Q}_{N_n}) = 0. \quad (4.32)$$

Here we have analyzed the complex models. Analogously, let us turn the attention to the simple models (the connections). In the class of models treated in the present work we always have $N_c = 2$ for any simple model \mathcal{C}_c (a generalization could be in order, but we stick to the simplest case for the sake of clarity). Then we have three possibilities $(N_c^P, N_c^Q) = (1, 1)$, $(N_c^P, N_c^Q) = (2, 0)$ and $(N_c^P, N_c^Q) = (0, 2)$. Particularly, in an iterative setting the latter case suffers from the same ill-posedness than the complex model with $N_n^P = 0$. For a simple model \mathcal{C}_c we can write, from problem (4.24), the following relation (for the case where $(N_c^P, N_c^Q) = (1, 1)$)

$$\begin{bmatrix} P_{\mathcal{C}_c,1} \\ Q_{\mathcal{C}_c,2} \end{bmatrix} = \mathcal{G}_{\mathcal{C}_c} \left(\begin{bmatrix} Q_{\mathcal{C}_c,1} \\ P_{\mathcal{C}_c,2} \end{bmatrix} \right), \quad (4.33)$$

where, as with expression (4.25), there is a time dependence that was not given explicitly for ease of notation. Hence, the functional relation maps, at any time, \mathbb{R}^2 onto itself, that is

$$\mathcal{G}_{\mathcal{C}_c} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \times \mathbb{R} \quad \forall t, \quad (4.34)$$

and represents the following problem (assuming that none of the two boundaries of this simple model are network outlets)

$$\begin{cases} M_c \frac{dQ_{\mathcal{C}_c,1}}{dt} + K_c Q_{\mathcal{C}_c,1} - (P_{\mathcal{C}_c,2} - P_{\mathcal{C}_c,1}) - B_c = 0 & \text{in } \mathcal{C}_c, \\ Q_{\mathcal{C}_c,1} + Q_{\mathcal{C}_c,2} = 0 & \text{in } \mathcal{C}_c. \end{cases} \quad (4.35)$$

Proceeding as before we can rearrange the equation to recast it through an implicit operator $\mathcal{F}_{\mathcal{C}_c}$ as follows

$$\mathcal{F}_{\mathcal{C}_c} \left(\begin{bmatrix} P_{\mathcal{C}_c,1} \\ P_{\mathcal{C}_c,2} \end{bmatrix}, \begin{bmatrix} Q_{\mathcal{C}_c,1} \\ Q_{\mathcal{C}_c,2} \end{bmatrix} \right) = \mathcal{F}_{\mathcal{C}_c}(\mathbf{P}_{\mathcal{C}_c}, \mathbf{Q}_{\mathcal{C}_c}) = 0. \quad (4.36)$$

4.4.2 Multi-component treatment

The way in which we treat separately a given complex model or a given simple model has already been explained. In this section we describe the formulation involving the interaction of many components, which corresponds to the coupled problem (4.24).

As a first step we group all the interface variables $\{P^i\}_{i=1}^{i_T}$ and $\{Q^i\}_{i=1}^{i_T}$, given in global numeration, in the vector quantities \mathbf{P} and \mathbf{Q} , respectively, that is

$$\mathbf{P} = \begin{bmatrix} P^1 \\ P^2 \\ \vdots \\ P^{i_T} \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} Q^1 \\ Q^2 \\ \vdots \\ Q^{i_T} \end{bmatrix}. \quad (4.37)$$

Such global quantities are related to the local quantities of each complex model through a mapping \mathbf{R}_{N_n} as follows

$$\mathbf{P}_{N_n} = \mathbf{R}_{N_n} \mathbf{P} \quad \mathbf{Q}_{N_n} = \mathbf{R}_{N_n} \mathbf{Q}, \quad (4.38)$$

where

$$\mathbf{P}_{N_n} = \begin{bmatrix} P_{\mathcal{N}_n,1} \\ P_{\mathcal{N}_n,2} \\ \vdots \\ P_{\mathcal{N}_n,N_n} \end{bmatrix} \quad \mathbf{Q}_{N_n} = \begin{bmatrix} Q_{\mathcal{N}_n,1} \\ Q_{\mathcal{N}_n,2} \\ \vdots \\ Q_{\mathcal{N}_n,N_n} \end{bmatrix}. \quad (4.39)$$

Likewise, we give the counterpart relation, denoted by \mathbf{R}_{C_c} , with the local quantities of each simple model

$$\mathbf{P}_{N_c} = \mathbf{R}_{C_c} \mathbf{P} \quad \mathbf{Q}_{N_c} = \mathbf{R}_{C_c} \mathbf{Q}, \quad (4.40)$$

where

$$\mathbf{P}_{N_c} = \begin{bmatrix} P_{\mathcal{C}_c,1} \\ P_{\mathcal{C}_c,2} \end{bmatrix} \quad \mathbf{Q}_{N_c} = \begin{bmatrix} Q_{\mathcal{C}_c,1} \\ Q_{\mathcal{C}_c,2} \end{bmatrix}. \quad (4.41)$$

These mappings also satisfy

$$\mathbf{P} = \sum_{n=1}^{n_T} \mathbf{R}_{N_n}^T \mathbf{P}_{N_n} \quad \mathbf{Q} = \sum_{n=1}^{n_T} \mathbf{R}_{N_n}^T \mathbf{Q}_{N_n}, \quad (4.42)$$

$$\mathbf{P} = \sum_{c=1}^{c_T} \mathbf{R}_{C_c}^T \mathbf{P}_{N_c} \quad \mathbf{Q} = \sum_{c=1}^{c_T} \mathbf{R}_{C_c}^T \mathbf{Q}_{N_c}, \quad (4.43)$$

establishing a mapping between local and global numerations.

Notice that the linear mappings $\mathbf{R}_{N_n} : \mathbb{R}^{i_T} \rightarrow \mathbb{R}^{N_n}$ and $\mathbf{R}_{C_c} : \mathbb{R}^{i_T} \rightarrow \mathbb{R}^{N_c}$ (recall that $N_c = 2$), are, algebraically speaking, rectangular matrixes whose entries are zeros and ones.

For each coupling interface we have two unknowns P^i and Q^i , and the equations are provided by the corresponding implicit operators (4.32) (complex models) and (4.36) (simple models), that is

$$\begin{aligned}\mathcal{F}_{\mathcal{N}_n}(\mathbf{P}_{N_n}, \mathbf{Q}_{N_n}) &= 0 & n &= 1, \dots, n_T, \\ \mathcal{F}_{\mathcal{C}_c}(\mathbf{P}_{C_c}, \mathbf{Q}_{C_c}) &= 0 & c &= 1, \dots, c_T.\end{aligned}\quad (4.44)$$

Changing now to a global numeration of the variables we get

$$\begin{aligned}\mathcal{F}_{\mathcal{N}_n}(\mathbf{R}_{N_n} \mathbf{P}, \mathbf{R}_{N_n} \mathbf{Q}) &= 0 & n &= 1, \dots, n_T, \\ \mathcal{F}_{\mathcal{C}_c}(\mathbf{R}_{C_c} \mathbf{P}, \mathbf{R}_{C_c} \mathbf{Q}) &= 0 & c &= 1, \dots, c_T.\end{aligned}\quad (4.45)$$

The next step consists in renummerating the interface equations using the transpose of the mappings \mathbf{R}_{N_n} and \mathbf{R}_{C_c} , which implies

$$\begin{aligned}\mathbf{R}_{N_n}^T \mathcal{F}_{\mathcal{N}_n}(\mathbf{R}_{N_n} \mathbf{P}, \mathbf{R}_{N_n} \mathbf{Q}) &= 0 & n &= 1, \dots, n_T, \\ \mathbf{R}_{C_c}^T \mathcal{F}_{\mathcal{C}_c}(\mathbf{R}_{C_c} \mathbf{P}, \mathbf{R}_{C_c} \mathbf{Q}) &= 0 & c &= 1, \dots, c_T.\end{aligned}\quad (4.46)$$

With these sets of equations we assemble the global problem consisting of $2i_T$ interface unknowns

$$\begin{aligned}\mathcal{F}_{\mathcal{N}}(\mathbf{P}, \mathbf{Q}) &= \sum_{n=1}^{n_T} \mathbf{R}_{N_n}^T \mathcal{F}_{\mathcal{N}_n}(\mathbf{R}_{N_n} \mathbf{P}, \mathbf{R}_{N_n} \mathbf{Q}) = 0, \\ \mathcal{F}_{\mathcal{C}}(\mathbf{P}, \mathbf{Q}) &= \sum_{c=1}^{c_T} \mathbf{R}_{C_c}^T \mathcal{F}_{\mathcal{C}_c}(\mathbf{R}_{C_c} \mathbf{P}, \mathbf{R}_{C_c} \mathbf{Q}) = 0.\end{aligned}\quad (4.47)$$

This is the system of non-linear interface equations associated to the coupled problem (4.24).

In some particular cases, from (4.47) it is possible to obtain the explicit relation between \mathbf{P} and \mathbf{Q} for components \mathcal{N} and \mathcal{C} yielding

$$\mathbf{P} = \mathcal{G}_{\mathcal{N}}^{SP}(\mathbf{Q}) \quad \text{or} \quad \mathbf{Q} = [\mathcal{G}_{\mathcal{N}}^{SP}]^{-1}(\mathbf{P}), \quad (4.48)$$

$$\mathbf{P} = \mathcal{G}_{\mathcal{C}}^{SP}(\mathbf{Q}) \quad \text{or} \quad \mathbf{Q} = [\mathcal{G}_{\mathcal{C}}^{SP}]^{-1}(\mathbf{P}), \quad (4.49)$$

4.4.3 An example to clarify notation

In order to exemplify the notation introduced so far, let us consider a specific example like the one shown in Figure 4.4. It is a closed network (the set \mathcal{O}_o is empty), and we will consider that we impose the traction (Neumann data) to all coupling points in all complex and simple components. Then, for the problem to be well-posed the pressure has to be fixed at any point in the network, for which we assume a reference pressure at interface \mathcal{I}_1 . This example resembles the cardiovascular a network

that will be presented in Section 4.5. The aim of this example is to provide all the elements introduced in the previous section to this specific scenario.

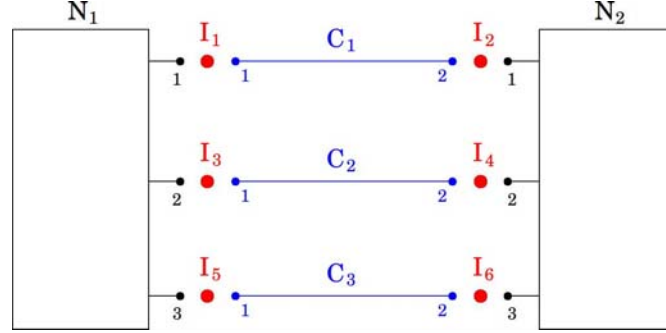


Fig. 4.4 Specific network with $i_T = 6$, $n_T = 2$, $c_T = 3$ and $o_T = 0$.

The matrixes \mathbf{R}_{N_n} , $n = 1, 2$ and \mathbf{R}_{C_c} , $c = 1, 2, 3$ are given by

$$\begin{aligned} \mathbf{R}_{N_1} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} & \mathbf{R}_{C_1} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \\ \mathbf{R}_{N_2} &= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} & \mathbf{R}_{C_2} &= \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \\ & & \mathbf{R}_{C_3} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Let us take for simplicity just the complex model \mathcal{N}_1 . In this case, equation (4.44) for $n = 1$ reads

$$\mathcal{F}_{\mathcal{N}_1}(\mathbf{P}_{N_1}, \mathbf{Q}_{N_1}) = \begin{pmatrix} Q_{\mathcal{N}_1,1} - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_1^{-1} (P_{\mathcal{N}_1,1}, P_{\mathcal{N}_1,2}, P_{\mathcal{N}_1,3}) \\ Q_{\mathcal{N}_1,2} - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_2^{-1} (P_{\mathcal{N}_1,1}, P_{\mathcal{N}_1,2}, P_{\mathcal{N}_1,3}) \\ Q_{\mathcal{N}_1,3} - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_3^{-1} (P_{\mathcal{N}_1,1}, P_{\mathcal{N}_1,2}, P_{\mathcal{N}_1,3}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (4.50)$$

where

$$[\mathcal{G}_{\mathcal{N}_1}^{SP}]^{-1} = \begin{pmatrix} [\mathcal{G}_{\mathcal{N}_1}^{SP}]_1^{-1} \\ [\mathcal{G}_{\mathcal{N}_1}^{SP}]_2^{-1} \\ [\mathcal{G}_{\mathcal{N}_1}^{SP}]_3^{-1} \end{pmatrix}. \quad (4.51)$$

Then, expression (4.45) is

$$\mathcal{F}_{\mathcal{N}_1}(\mathbf{R}_{N_1} \mathbf{P}, \mathbf{R}_{N_1} \mathbf{Q}) = \begin{pmatrix} Q^1 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_1^{-1} (P^1, P^3, P^5) \\ Q^3 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_2^{-1} (P^1, P^3, P^5) \\ Q^5 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_3^{-1} (P^1, P^3, P^5) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (4.52)$$

The application of the mapping \mathbf{R}_{N_1} to $\mathcal{F}_{\mathcal{N}_1}$ yields the specific counterpart to (4.46), that is

$$\mathbf{R}_{N_1}^T \mathcal{F}_{\mathcal{N}_1}(\mathbf{R}_{N_1} \mathbf{P}, \mathbf{R}_{N_1} \mathbf{Q}) = \begin{pmatrix} Q^1 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_1^{-1} (P^1, P^3, P^5) \\ 0 \\ Q^3 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_2^{-1} (P^1, P^3, P^5) \\ 0 \\ Q^5 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_3^{-1} (P^1, P^3, P^5) \\ 0 \end{pmatrix} = 0. \quad (4.53)$$

Performing the same steps for \mathcal{N}_2 we assemble the system of equations provided by the two complex models (see (4.47)) as follows

$$\mathcal{F}_{\mathcal{N}}(\mathbf{P}, \mathbf{Q}) = \begin{pmatrix} Q^1 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_1^{-1} (P^1, P^3, P^5) \\ Q^2 - [\mathcal{G}_{\mathcal{N}_2}^{SP}]_1^{-1} (P^2, P^4, P^6) \\ Q^3 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_2^{-1} (P^1, P^3, P^5) \\ Q^4 - [\mathcal{G}_{\mathcal{N}_2}^{SP}]_2^{-1} (P^2, P^4, P^6) \\ Q^5 - [\mathcal{G}_{\mathcal{N}_1}^{SP}]_3^{-1} (P^1, P^3, P^5) \\ Q^6 - [\mathcal{G}_{\mathcal{N}_2}^{SP}]_3^{-1} (P^2, P^4, P^6) \end{pmatrix} = 0. \quad (4.54)$$

Analogously for the simple components after the assembling process we have

$$\mathcal{F}_{\mathcal{C}}(\mathbf{P}, \mathbf{Q}) = \begin{pmatrix} Q^1 - [\mathcal{G}_{\mathcal{C}_1}^{SP}]_1^{-1} (P^1, P^2) \\ Q^2 - [\mathcal{G}_{\mathcal{C}_1}^{SP}]_2^{-1} (P^1, P^2) \\ Q^3 - [\mathcal{G}_{\mathcal{C}_2}^{SP}]_1^{-1} (P^3, P^4) \\ Q^4 - [\mathcal{G}_{\mathcal{C}_2}^{SP}]_2^{-1} (P^3, P^4) \\ Q^5 - [\mathcal{G}_{\mathcal{C}_3}^{SP}]_1^{-1} (P^5, P^6) \\ Q^6 - [\mathcal{G}_{\mathcal{C}_3}^{SP}]_2^{-1} (P^5, P^6) \end{pmatrix} = 0, \quad (4.55)$$

where, for $c = 1, 2$, it is

$$[\mathcal{G}_{\mathcal{C}_c}^{SP}]^{-1} = \begin{pmatrix} [\mathcal{G}_{\mathcal{C}_c}^{SP}]_1^{-1} \\ [\mathcal{G}_{\mathcal{C}_c}^{SP}]_2^{-1} \end{pmatrix}. \quad (4.56)$$

Therefore, to obtain the solution (\mathbf{P}, \mathbf{Q}) we have to solve the system of nonlinear equations composed by (4.54) and (4.55) using any iterative method.

4.5 Dimensionally-heterogeneous blood flow components

From this section onwards the notation slightly differs from the previous sections, since it is now specified and simplified for the cardiovascular setting.

The basic ingredients of the iterative algorithm responsible for solving the dimensionally-heterogeneous model of the CVS are the same introduced in Section 4.2 and in Section 4.4.

4.5.1 Hemodynamics components

The closed-loop circulation is described in terms of three classes of components:

High-pressure component This component is shortly denoted by HP , and comprises the arterial network and peripheral beds. These vascular entities are modeled using 1D and 0D (lumped) representations, respectively. The systemic circulation is modeled as the flow of an incompressible fluid in compliant vessels [18, 29, 33]. At each terminal vessel a Windkessel element is considered to model the peripheral circulation [29]. The topological configuration and data set (based on [1]) are taken from our previous works [3, 4, 5, 6].

Low-pressure component This component is shortly denoted by LP , and is responsible for closing the cardiovascular circuit incorporating venules, veins, cavas, atria, ventricles and valves, modeled through 0D models. The venous and pulmonary circulations are modeled using standard analog electric circuit models as in [24]. The four cardiac chambers are introduced following also [24]. The models for the four valves are the ones presented in [21]. The data for the different vascular regions were taken from [24] and [21], correspondingly.

Specific vessel component This component is shortly denoted by SV , and stands for specific vessels of interest in which blood flow is to be simulated in detail such as bifurcations and aneurysms, among others. The blood flow is modeled using the Navier-Stokes equations in deforming domains [11, 30]. The structural behavior is considered through independent ring models [31]. Standard as well as patient-specific arterial vessels can be considered in the model.

4.5.2 Full closed-loop model of the CVS

This three-component closed-loop description of the CVS is schematically shown in Figure 4.5. Hereafter, for notational simplicity we assume that we have just one SV component in the model. The closed-loop circuit is formed by the coupling between the LP and the HP components. In the example of the figure the SV component is embedded in the HP one.

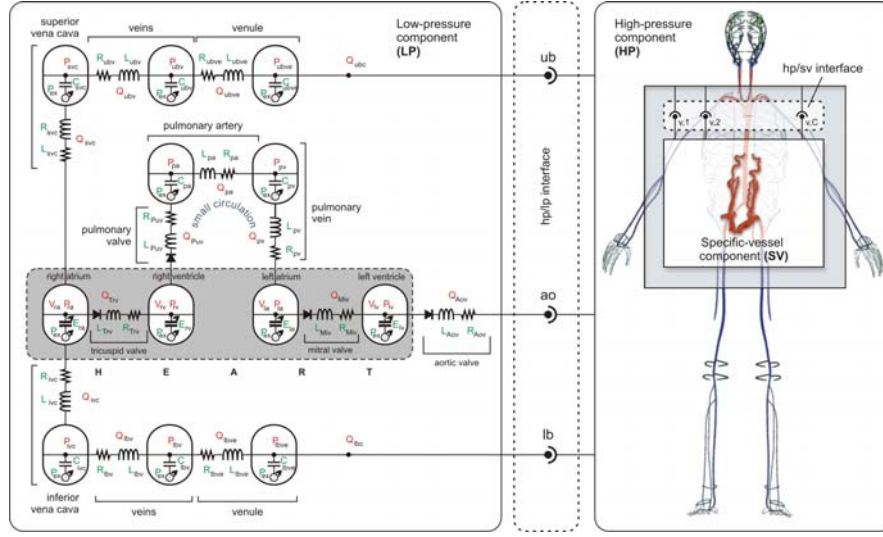


Fig. 4.5 Representation of the CVS through heterogeneous components.

There are three coupling interfaces between the LP and the HP components. The first interface is placed at the aortic root right after the aortic valve, and is named as ao interface. The second and third interfaces are placed over the upper and lower body regions, and are named ub and lb , respectively. These upper and lower connections collect all the corresponding Windkessel terminals from the arterial network and derive the blood flow into the upper and lower venous networks. In the SV component we consider C coupling interfaces which are shared with the HP component, and that correspond to the artificial inlets/outlets when the vessel is isolated from the rest of the arterial system. Therefore, the total number of coupling interfaces in the network is $N = 3 + C$.

4.5.3 Black-box paradigm

Over each coupling interface we define two degrees of freedom: flow rate Q and pressure P . As usual, mass conservation and force equilibrium are assumed to be satisfied at these points, then Q and P are continuous quantities across these interfaces. The system is fully described by the vector of interface unknowns $\mathbf{X} = (\mathbf{Q}, \mathbf{P})$, where $\mathbf{Q} = (Q_{ao}, Q_{ub}, Q_{lb}, \{Q_{v,i}\}_{i=1}^C)$ and $\mathbf{P} = (P_{ao}, P_{ub}, P_{lb}, \{P_{v,i}\}_{i=1}^C)$. These are global unknowns, which have a counterpart at the level of the different components. Hence, component-wise, we map the global unknowns as follows

$$\begin{aligned}
\mathbf{Q}_{\text{HP}} &= (Q_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}}, \{Q_{v,i}\}_{i=1}^C), & \mathbf{P}_{\text{HP}} &= (P_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}}, \{P_{v,i}\}_{i=1}^C), \\
\mathbf{Q}_{\text{LP}} &= (Q_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}}), & \mathbf{P}_{\text{LP}} &= (P_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}}), \\
\mathbf{Q}_{\text{SV}} &= (\{Q_{v,i}\}_{i=1}^C) & \mathbf{P}_{\text{SV}} &= (\{P_{v,i}\}_{i=1}^C).
\end{aligned}$$

From the component interface unknowns vector $\mathbf{X}_X = (\mathbf{Q}_X, \mathbf{P}_X)$ we define those quantities considered as input data \mathbf{I}_X and those considered output data \mathbf{O}_X , $X = \text{HP}, \text{LP}, \text{SV}$.

For the SV component the most natural choice for the input data in the present setting is to take Neumann boundary conditions on every inlet/outlet of the domain. For the HP component, our previous experience dictates that the most convenient choice from the performance point of view is to take the pressure as input data (see [6]). Finally, for the LP component the natural choice according to the morphology of the component (see Figure 4.5) is to consider as input data the flow rate at the ub and lb connecting points, and the pressure at the ao connecting point. These choices is summarized in Table 4.1.

Component	Input data	Output data
HP	$\mathbf{I}_{\text{HP}} = (P_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}}, \{P_{v,i}\}_{i=1}^C)$	$\mathbf{O}_{\text{HP}} = (Q_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}}, \{Q_{v,i}\}_{i=1}^C)$
LP	$\mathbf{I}_{\text{LP}} = (P_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}})$	$\mathbf{O}_{\text{LP}} = (Q_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}})$
SV	$\mathbf{I}_{\text{SV}} = (P_{v,1}, P_{v,2}, \dots, P_{v,C})$	$\mathbf{O}_{\text{SV}} = (Q_{v,1}, Q_{v,2}, \dots, Q_{v,C})$

Table 4.1 Input and output data for the different components (see Figure 4.5 for notation).

As seen in Section 4.4, the governing equation can be written in abstract operator form as follows

$$\mathcal{R}_X^t(\mathbf{X}_X) = \mathcal{R}_X^t(\mathbf{Q}_X, \mathbf{P}_X) = 0 \quad X = \text{HP}, \text{LP}, \text{SV}. \quad (4.57)$$

Reordering the degrees of freedom we write

$$\mathcal{F}_X^t(\mathbf{O}_X, \mathbf{I}_X) = 0 \quad X = \text{HP}, \text{LP}, \text{SV}, \quad (4.58)$$

while in explicit form each equation reads

$$\mathbf{O}_X = \mathcal{G}_X^t(\mathbf{I}_X) \quad X = \text{HP}, \text{LP}, \text{SV}. \quad (4.59)$$

Hence, solving component X implies obtaining the output \mathbf{O}_X from given input data \mathbf{I}_X . The full system of equations to be solved can then be written as

$$\begin{aligned}
(Q_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}}, \{Q_{v,i}\}_{i=1}^C) - \mathcal{G}_{\text{HP}}^t(P_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}}, \{P_{v,i}\}_{i=1}^C) &= 0, \\
(Q_{\text{ao}}, P_{\text{ub}}, P_{\text{lb}}) - \mathcal{G}_{\text{LP}}^t(P_{\text{ao}}, Q_{\text{ub}}, Q_{\text{lb}}) &= 0, \\
(Q_{v,1}, Q_{v,2}, \dots, Q_{v,C}) - \mathcal{G}_{\text{SV}}^t(P_{v,1}, P_{v,2}, \dots, P_{v,C}) &= 0.
\end{aligned} \quad (4.60)$$

The number of coupling equations in (4.60) is $2N$, being N the number of coupling interfaces. Recall that the number of interface unknowns is also $2N$. This is due to the fact that each component provides one equation per interface correspond-

ing to the output data defined at such interface. So, two equations are provided in total per interface. Thus, by construction the system is always closed.

Operators \mathcal{G}_X^t , $X = \text{HP, LP, SV}$, in (4.60), are time-dependent and non-linear, and entail dealing with ordinary differential equations, 1D and 3D partial differential equations, correspondingly. Thus, in order to solve each single component it is necessary to apply both time discretization and linearization procedures for each component. In the HP and SV components a spatial discretization is also needed. The interested reader is referred to [3, 9, 31] for details of time and spatial discretization of the different components.

4.6 Component decomposition strategy

4.6.1 Broyden method

As said, the present approach make use of the Broyden method in order to construct a robust algorithm for the solution of system (4.60). For a generic non-linear equation $\mathcal{R}(x) = 0$, the Broyden method reads [12, 19]

Algorithm 1 Broyden's algorithm

1. Given x^0 and B^0 do
 2. Compute $r^0 = \mathcal{R}(x^0)$
 3. For $j = 0, 1, \dots$ until $\frac{\|r^j\|}{\|r^0\|} \leq \varepsilon$, or $\|r^j\| \leq \varepsilon^a$, do
 4. $w^j = -(B^j)^{-1} r^j$
 5. $x^{j+1} = x^j + w^j$
 6. $r^{j+1} = \mathcal{R}(x^{j+1})$
 7. $B^{j+1} = B^j + \frac{r^{j+1}(w^j)^T}{(w^j)^T w^j}$
 8. Enddo
-

In our context the unknown x^j is the vector of interface unknowns $\mathbf{X}^j = (\mathbf{Q}^j, \mathbf{P}^j)$. It is worthwhile to note that the algorithm requires an approximation of the Jacobian matrix as initial matrix, which is called B^0 , after which a rank-one update is performed at each iteration (see Step 7). The convergence criterion depends on the relative residual (tolerance ε) and on the absolute residual (tolerance ε^a). Step 6 (and also the initial Step 2) requires the evaluation of the residual r^j , which is

$$r^j = \begin{pmatrix} r_{\text{HP}}^j \\ r_{\text{LP}}^j \\ r_{\text{SV}}^j \end{pmatrix} = \begin{pmatrix} \mathcal{R}_{\text{HP}}^t(\mathbf{X}^j) \\ \mathcal{R}_{\text{LP}}^t(\mathbf{X}^j) \\ \mathcal{R}_{\text{SV}}^t(\mathbf{X}^j) \end{pmatrix} = \begin{pmatrix} \mathbf{O}_{\text{HP}}^j - \mathcal{G}_{\text{HP}}^t(\mathbf{I}_{\text{HP}}^j) \\ \mathbf{O}_{\text{LP}}^j - \mathcal{G}_{\text{LP}}^t(\mathbf{I}_{\text{LP}}^j) \\ \mathbf{O}_{\text{SV}}^j - \mathcal{G}_{\text{SV}}^t(\mathbf{I}_{\text{SV}}^j) \end{pmatrix}. \quad (4.61)$$

Recall that the problem is time-dependent, so we need to go through this algorithm at each time step. That means that we need to provide the matrix B^0 at each new time step $n + 1$, say $B^{0,n+1}$. This is called the initialization of the Broyden's algorithm. The strategy followed in the present work is to take $B^{0,n+1} = B^n$, which is equivalent to using a preconditioner for the non-linear problem [19]. In the present work, at the first time step the matrix B^0 is initialized with the Jacobian of the system computed by finite differences, that is

$$B_{kl}^0 = \frac{\mathcal{R}_k(x_l^0 + \delta x_l^0) - \mathcal{R}_k(x_l^0)}{\delta x_l^0} \quad k, l = 1, \dots, 2N. \quad (4.62)$$

where \mathcal{R}_k is the k th component of the vector operator \mathcal{R} and $\delta x_l^0 = \varepsilon x_l^0$, $l = 1, \dots, 2N$, with ε small enough.

4.6.2 Inner/outer iterations and local/global time steps

To reach the final form of the coupling algorithm we need to consider two additional features which are present when coupling heterogeneous components in blood flow simulations.

The first is related to the non-linearities in the components. In this way, we will refer to the iterations performed by the Broyden algorithm as *outer iterations* (or Broyden iterations), and whenever a component is non-linear we will refer to the iterations performed in the specific components as *inner iterations*. The former are responsible for gluing the solution at the coupling interfaces, and the latter are mandatory in order to evaluate the residual in each specific non-linear component properly, something which has to be done at each outer iteration. For example, given the data at iteration j , \mathbf{I}_{LP}^j , and the output at the same iteration j , \mathbf{O}_{LP}^j , we have to iterate until we reach the corresponding fixed point in the LP component from which we retrieve the corresponding output $\mathcal{G}_{LP}^i(\mathbf{I}_{LP}^j)$, and then we measure the residual as in the second line of (4.61). Observe that the evaluation of the residual in a certain component is a local operation, being independent from the rest of the components.

The second feature is related to the time discretization. Here we discern between two levels of time-stepping. We introduce the *global time-step* Δt_G which determines the time instants in the simulation at which the system of equations (4.60) is satisfied. Independently from this, we introduce in each component the *local time step* $\Delta t_{L,X}$, which is the result of time discretization in component $X = \text{HP}, \text{LP}, \text{SV}$. This local time step may depend upon stability requirements of each specific component. For instance, we can weaken the coupling equations by enforcing them at larger time steps, say $\Delta t_G > \Delta t_{L,LP} = \Delta t_{L,HP} = \Delta t_{L,SV}$, as well as we can have a different time step for different components, say $\Delta t_G = \Delta t_{L,SV} > \Delta t_{L,HP} = \Delta t_{L,LP}$. As we will see, the latter strategy is the best choice when aiming at reducing computational cost.

4.6.3 Broyden-based iterative algorithm

According to the additional features introduced in the previous section, we now present the full version of the decomposition algorithm discriminating outer and inner iterations as well as global and local time-steps.

Algorithm 2 Broyden-based algorithm for coupling heterogeneous models

1. Given x^0 and B^0 do
 2. For $n = 0, 1, \dots, n_{\max}$ do (time step: $\Delta t_G = t^{n+1} - t^n$)
 3. Set $x^{0,n+1} = x^n$ and $B^{0,n+1} = B^n$
 4. Evaluate $r^{0,n+1}(x^{0,n+1})$ through Algorithm 3
 5. For $j = 0, 1, \dots$ until $\frac{\|r^{j,n+1}\|}{\|r^{0,n+1}\|} \leq \varepsilon_O$, or $\|r^{j,n+1}\| \leq \varepsilon^a$, do
 6. $w^j = -(B^{j,n+1})^{-1} r^{j,n+1}$
 7. $x^{j+1,n+1} = x^{j,n+1} + w^j$
 8. Evaluate $r^{j+1,n+1}(x^{j+1,n+1})$ through Algorithm 3
 9. $B^{j+1,n+1} = B^{j,n+1} + \frac{r^{j+1,n+1}(w^j)^T}{(w^j)^T w^j}$
 10. Enddo
 11. Set $x^{n+1} = x^{j+1,n+1}$ and $B^{n+1} = B^{j+1,n+1}$
 12. Enddo
-

This algorithm is complemented by the following procedure for the evaluation of the residual using substepping and inner non-linear iterations. Thus, the full assembling of the system of equations is performed through the assembling of the equations corresponding to the local components.

Algorithm 3 Residual evaluation and system assembling

1. Given $x^{j,n+1}$ compute $r^{j,n+1}(x^{j,n+1})$ doing
 2. For $X \in \{\text{HP}, \text{LP}, \text{SV}\}$ do
 3. Set $a_X = \frac{\Delta t_G}{\Delta t_{L,X}}$
 4. For $m_X = \frac{1}{a_X}, \frac{2}{a_X}, \dots, 1$ do (time step: $\Delta t_{L,X} = t^{n+m_X} - t^{n+m_X - \frac{1}{a_X}}$)
 5. Compute $r_X^{j,n+m_X,0} = \mathcal{R}_X^{j,n+m_X,0}(x^{j,n+1})$
 6. For $k_X = 0, 1, \dots$ until $\frac{\|r_X^{j,n+m_X,k_X}\|}{\|r_X^{j,n+m_X,0}\|} \leq \varepsilon_{I,X}$, or $\|r_X^{j,n+m_X,k_X}\| \leq \varepsilon^a$, do
 7. $r_X^{j,n+m_X,k_X+1} = \mathcal{R}_X^{j,n+m_X,k_X+1}(x^{j,n+1})$
 8. Enddo
 9. $r_X^{j,n+m_X} = r_X^{j,n+m_X,k_X+1}$
 10. Enddo
 11. Assemble the contribution of $r_X^{j,n+1}$ to $r^{j,n+1}$ (see (4.61))
 12. Enddo
-

For outer iterations the convergence tolerance is ε_O , and for inner iterations the convergence tolerance is $\varepsilon_{I,X}$, $X = \text{HP}, \text{LP}, \text{SV}$.

Figure 4.6 displays the flow chart for the iterative algorithm featuring the global -coupling- time step, the local -component- time steps, the outer -coupling- iterations and the inner -component- iterations.

4.7 Hemodynamics simulations

Unless stated otherwise four cardiac cycles were simulated. The convergence tolerances are $\varepsilon_O = 1 \cdot 10^{-6}$, $\varepsilon_{I,X} = 1 \cdot 10^{-7}$, $X = \text{HP}, \text{LP}, \text{SV}$ and $\varepsilon^a = 1 \cdot 10^{-14}$. Simulations are characterized by the global time step Δt_G and the local ratios α_X , $X = \text{HP}, \text{LP}, \text{SV}$. In all cases the cardiac period is $T = 1$ s. The number of degrees of freedom in the HP component is 16,776, and in the LP component is 46. For the HP and LP components, a direct solver is used to solve the algebraic system of equations, and for the SV component a GMRES algorithm is employed. Recall that the focus of the following applications is the analysis of the performance of Algorithm 2.

In all the examples, we fix the spatial discretization of the components (specifically of the HP and SV components) and perform sensitivity analyses with respect to the global and local time steps. Notice that, unlike classical domain decomposition methods, the number of interface unknowns in the system of equations (4.61) does not depend on the spatial discretization in the components. As a consequence, the proposed iterative method is insensitive to the number of internal degrees of freedom of each component. A mathematical proof of this result for linear problems is given in [7].

In the present work, we make use of implicit numerical methods for solving the time advancing equations in the different components. Therefore, there is no CFL-like constraints over the local time steps.

4.7.1 Blood flow in the closed-loop cardiovascular model

The first example is used to explore the iterative strategy for the closed-loop HP-LP system (no SV component is included in the present analysis).

In Tables 4.2-4.3 the results of the simulations are summarized when combining Algorithms 2 and 3. We have included the case in which we solve the HP-LP system in a monolithic manner (considering $\Delta t = \Delta t_G$ and tolerance $\varepsilon = \varepsilon_O$). Notation $(\cdot)^{(L)}$ indicates that a single Picard inner iteration was performed. This is equivalent to considering the component as a linear one. Observe that this is related to the inner iterations, which are independent from the outer iterations that drive the convergence of the entire system (4.60). Running the HP component as a linear component is possible, on the contrary, running the LP component as a linear one makes the iterative method to diverge in all tested cases. This phenomenon is caused by

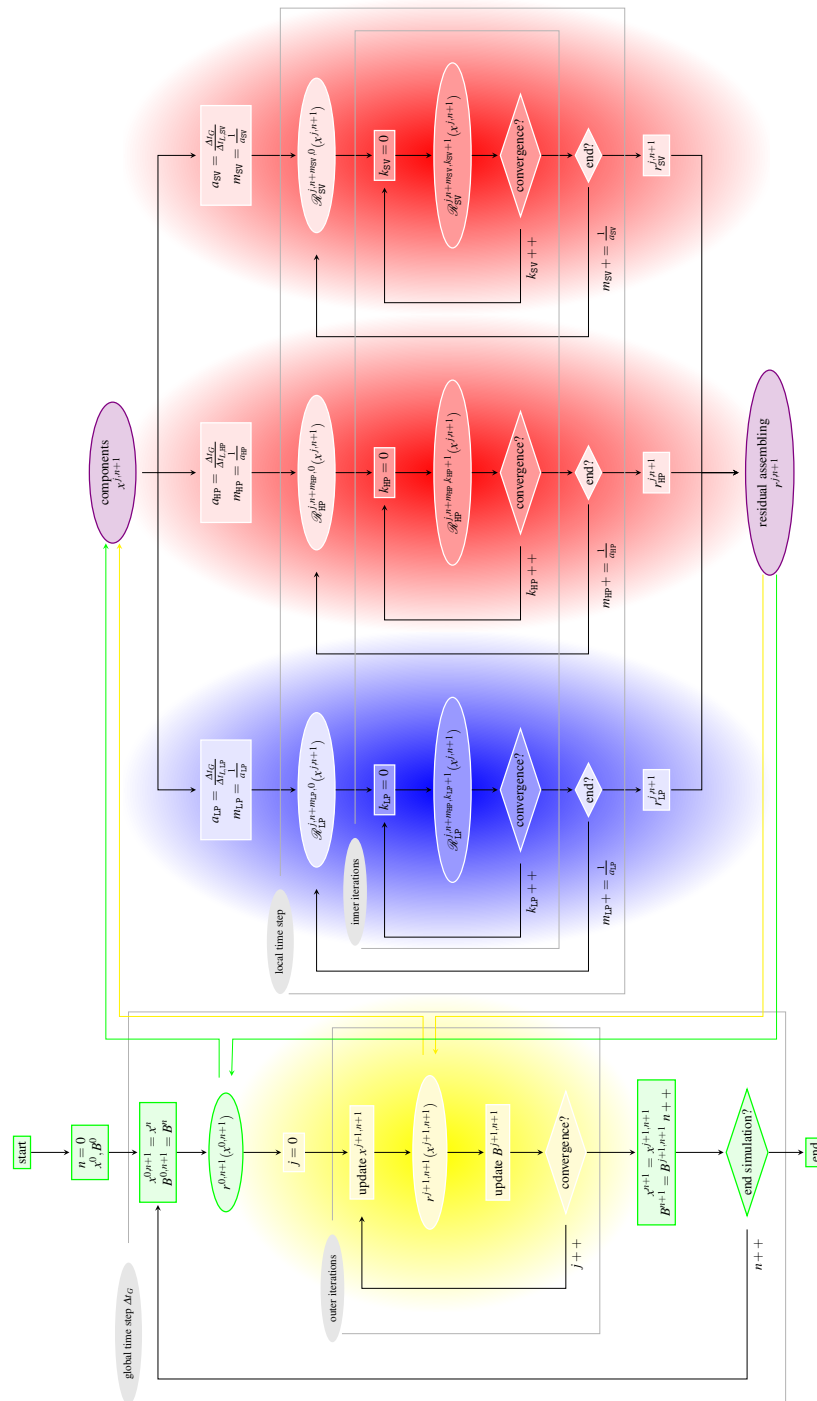


Fig. 4.6 Iterative algorithm featuring global/local time steps and outer/inner iterations.

the nondifferentiable behavior of cardiac valves around the opening/closing phases. Lack of convergence of the fixed point iterations in valve dynamics turns the entire LP unstable, from which the iterative method for the coupled problem fails to converge.

Per cardiac beat we have: N_{ci} the total number of coupling iterations between the two components, $N_{ss,X}$ the number of times component X is solved and $N_{ss,HP-LP}$ the number of times which the monolithic HP-LP system is solved. A crashed simulation is indicated by “–” (LP inner iterations did not converge in all these cases). Analogously, notation $-(L)$ indicates that the simulation crashed as a consequence of the lack of convergence of the LP component, also in the case of running the HP component as a linear one.

Time stepping			Iterative solution			Monolithic solution
Δt_G [s]	a_{HP}	a_{LP}	$N_{ci} (\frac{N_{ci}}{N_{\Delta t_G}})$	$N_{ss,HP}$	$N_{ss,LP}$	$N_{ss,HP-LP}$
0.0005	1	1	8238 (4.12)	21313	25707	11564
	1	1	10414 (5.21)	10414 ^(L)	30581	
0.0010	1	1	4160 (4.16)	11779	14082	6519
	1	1	5912 (5.91)	5912 ^(L)	18271	
	1	2	4071 (4.07)	11336	46952	
	1	2	5333 (5.33)	5333 ^(L)	62011	
	2	2	3766 (3.77)	30954	43829	
	2	2	–	–	–	
0.0020	1	1	2307 (4.61)	7368	7987	3750
	1	1	3034 (6.07)	3034 ^(L)	9585	
	1	2	2409 (4.82)	7324	31194	
	1	2	3052 (6.10)	3052 ^(L)	39755	
	2	2	2057 (4.11)	18217	26689	
	1	4	2338 (4.68)	7186	53960	
	1	4	2985 (5.97)	2985 ^(L)	69337	
	2	4	2062 (4.12)	18268	47788	
	4	4	2063 (4.13)	33928	47830	

Table 4.2 Iterative and monolithic solution of the HP-LP system for several combinations of global and local time steps (Δt_G , a_{HP} and a_{LP} specified in each case; in the monolithic case $\Delta t = \Delta t_G$). Per cardiac beat, N_{ci} denotes the total number of coupling iterations among components, $N_{\Delta t_G}$ is the number of global time steps ($\frac{T}{\Delta t_G}$), $N_{ss,X}$ is the number of times component X was solved and $N_{ss,HP-LP}$ is the number of times the coupled HP-LP system was solved. Notation $(\cdot)^{(L)}$ implies that the HP component was run as a linear component regarding the inner iterations (in such case $a_{HP} = 1$ always). The symbol “–” denotes lack of convergence, while notation $(\cdot)^{(*)}$ indicates that the solution is far from the actual solution (due to the time stepping strategy).

The main outcomes from Tables 4.2-4.3, are summarized here:

- The coupling iterations per global time step, that is $\frac{N_{ci}}{N_{\Delta t_G}}$, are almost insensitive to Δt_G , while are sensitive to the nonlinear/linear character of the HP component.
- There is no relation between the number of times the HP and the LP component are solved. The nonlinearities of the latter component are bounded to that

Time stepping			Iterative solution			Monolithic solution
Δt_G [s]	a_{HP}	a_{LP}	$N_{ci} (\frac{N_{ci}}{N_{\Delta t_G}})$	$N_{ss,HP}$	$N_{ss,LP}$	$N_{ss,HP-LP}$
0.0040	1	1	—	—	—	—
	1	2	1398 (5.59)	5154	20528	
	1	2	1802 (7.21)	1802 ^(L)	26679	
	2	2	1104 (4.41)	12154	16358	
	1	4	1418 (5.67)	5215	36669	
	1	4	—	— ^(L)	—	
	2	4	1080 (4.32)	11886	28145	
	4	4	1094 (4.38)	19507	28611	
	1	8	1415 ^(*) (5.66)	5174	65301	
	1	8	1783 ^(*) (7.13)	1783 ^(L)	82664	
	2	8	1078 (4.31)	11873	50257	
	4	8	1083 (4.33)	19295	50475	
	8	8	1093 (4.37)	36017	51025	
0.0080	1	1	—	—	—	—
	2	2	—	—	—	
	1	4	788 (6.30)	3900	23210	
	1	4	1156 (9.25)	1156 ^(L)	34078	
	2	4	594 (4.75)	8132	17769	
	4	4	590 (4.72)	13008	17640	
	1	8	794 (6.35)	3893	41270	
	1	8	1157 ^(*) (9.26)	1157 ^(L)	60083	
	2	8	594 (4.75)	8150	31377	
	4	8	595 (4.76)	13129	31447	
	8	8	595 (4.76)	21234	31352	

Table 4.3 (Continuation of Table 4.2) Iterative and monolithic solution of the HP-LP system for several combinations of global and local time steps (Δt_G , a_{HP} and a_{LP} specified in each case; in the monolithic case $\Delta t = \Delta t_G$). See notation in caption of Table 4.2.

component. Furthermore, it is possible to run the HP component as a linear component and reduce computational cost.

- The nonlinearities of the LP are responsible for lack of convergence of the monolithic system at large time steps. Substepping allows to circumvent this issue by setting adequate time steps according to each component requirements.
- Cheaper simulations are obtained using the largest possible time step for the most expensive component (in this case the HP component), while the time step should be reduced in the rest of the components (in this case the LP component).
- A measure of the computational cost for the solution of the coupled system is given, in this case, by the number of iterations $N_{ss,HP}$. The physics captured by the different simulations are, roughly speaking, equivalent and independently of the time step. The lack of convergence of the monolithic strategy, which limits the time step to 0.0020 s, is thus just a numerical artifact. The proposed iterative method already reduces the cost from $N_{ss,HP} = 3750$ to $N_{ss,HP} = 2985$ for $\Delta t_G = 0.0020$ s (20% reduction), but by exhibiting convergence up to $\Delta t_G = 0.0080$ s it allows for further reduction of the cost to $N_{ss,HP} = 1156$ (69% reduction).

Finally, Figure 4.7 displays the solution, in terms of flow rate and pressure at several points in the HP-LP closed-loop cardiovascular system. As well, in Figure 4.8 the cardiac circulation is presented by plotting the dynamics of the four cardiac valves and the corresponding flow through each of them.

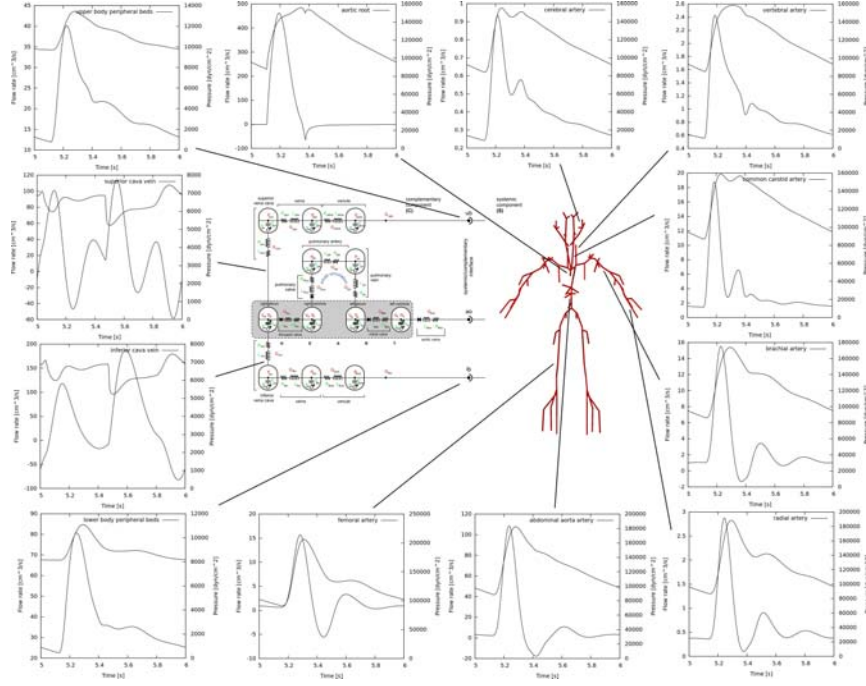


Fig. 4.7 Flow rate and pressure at several points of the HP-LP closed-loop cardiovascular system.

4.7.2 Blood flow in a cerebral aneurysm

In this case we use the same HP-LP system as in the previous section and couple a SV component representing a patient-specific cerebral aneurysm located at the middle cerebral artery. Clearly, the computational cost is driven by the SV component for which an adequate time step is $\Delta t_{L,SV} \leq 0.002$. For this application we limit ourselves to a few number of simulations to compare the performance, since the goal here is to show the applicability in cases of potential medical interest.

The geometry of the aneurysm is reconstructed from a set of DICOM (*Digital Imaging and Communication in Medicine*) images following standard segmentation procedures. The number of degrees of freedom in the SV component is 165,046.

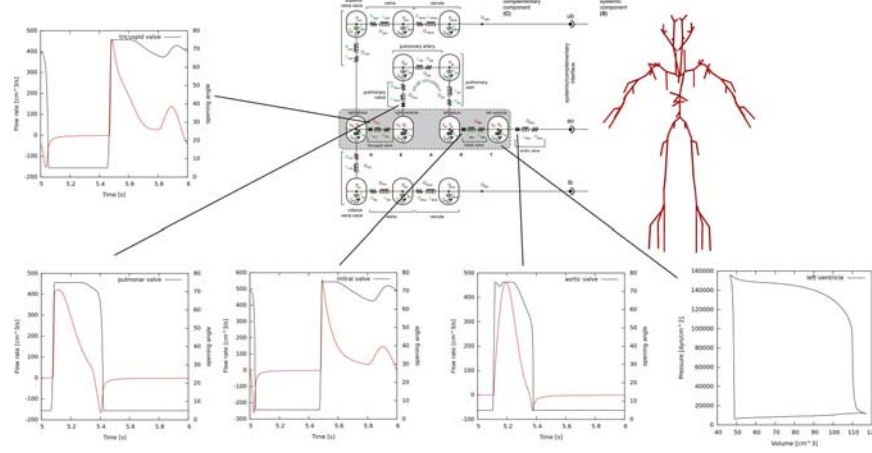


Fig. 4.8 Valve dynamics in the cardiac circulation present in the LP component.

Table 4.4 summarizes the simulations performed in this case. The cheapest simulation, given that $a_{SV} = 1$, is obtained when $a_{HP} = 2$ and $a_{LP} = 2$ for both time steps ($\Delta t = 0.001$ and $\Delta t = 0.002$), which is consistent with the conclusions obtained from the previous section.

Time stepping				Iterative solution			
Δt_G [s]	a_{HP}	a_{LP}	a_{SV}	$N_{ci} (\frac{N_{ci}}{N_{\Delta t_G}})$	$N_{ss,HP}$	$N_{ss,LP}$	$N_{ss,SV}$
0.0010	1	1	1	4239 (4.24)	14044	13912	11130
	1	2	1	4289 (4.29)	14089	49612	11135
	2	2	1	4237 (4.24)	34845	49024	11110
0.0020	1	1	1	2514 (5.03)	9719	8287	7517
	1	2	1	2549 (5.10)	9854	33019	7610
	2	2	1	2475 (4.95)	22180	32085	7432

Table 4.4 Iterative solution of the HP-LP-SV system for some combinations of global and local time steps. See notation introduced in Tables 4.2-4.3.

The number of iterations showed to be insensitive to the multi-stepping strategy employed (once Δt_G has been fixed), as concluded from Table 4.4. Moreover, the convergence history is approximately constant throughout the entire cardiac cycle. For $\Delta t_G = 0.001$ the number of iterations is smaller than for $\Delta t_G = 0.002$, but the computational cost turns to be less in the latter case because the reduction in $\frac{N_{ci}}{\Delta t_G}$ does not make up for the increase in the number of times per cardiac cycle the SV component has to be solved (see last column in Table 4.4). Figure 4.9 features the iterations performed to reach convergence at each time step in the four cardiac cycles. As seen in Table 4.4 it is possible to appreciate that the number of iterations is insensitive to the global/local time-stepping strategy.

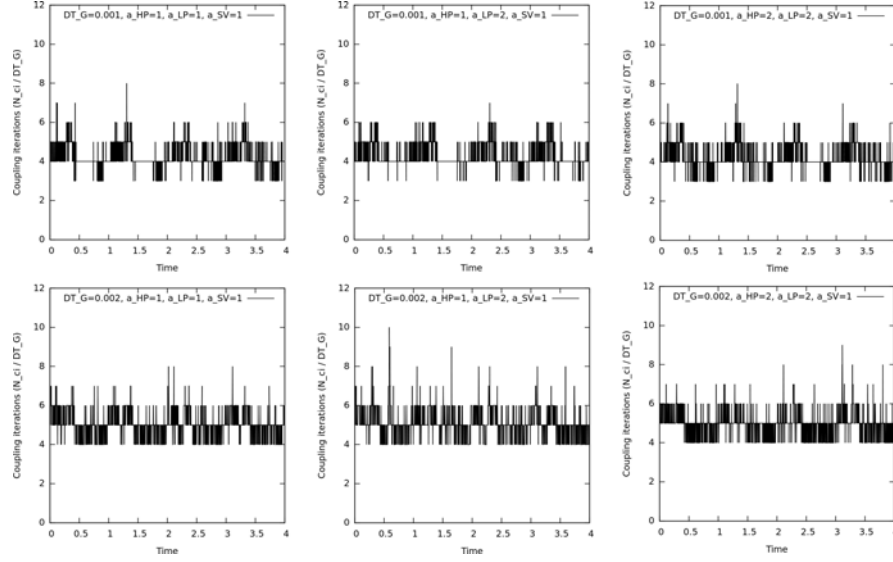


Fig. 4.9 Iterations per time step throughout the four cardiac cycles for the different configurations of global/local time-steps.

In this case, the measure of the computational cost for the solution of the coupled system is given by the number of iterations of the most expensive component, that is $N_{ss,sv}$.

Figure 4.10 shows the geometry of the patient-specific vessel and the results at all the coupling interfaces. These results correspond to the case $a_{HP} = 2$, $a_{LP} = 2$ and $a_{SV} = 1$ and to the last cardiac cycle.

Finally, in Figure 4.11 the local structure of blood flow is presented at different time instants corresponding to the last cardiac cycle through the computation of the streamlines.

4.7.3 Blood flow in the arm

This example is used to show an application in which several SV components are considered within the HP-LP system seen before. These components correspond to the five arterial branchings taking place at one of the arms. Each bifurcation is constructed such that it matches the lumen radii and arterial wall parameters with the corresponding artery in the HP component. The aim here is to show the robustness in the solution process even for a large number of SV components (see Figure 4.12 for a schematic placement of the bifurcations in the region of interest). In this case, the computational cost should be carefully evaluated because the number of degrees of freedom in the SV components ranges between 85,365 and 136,941, and this

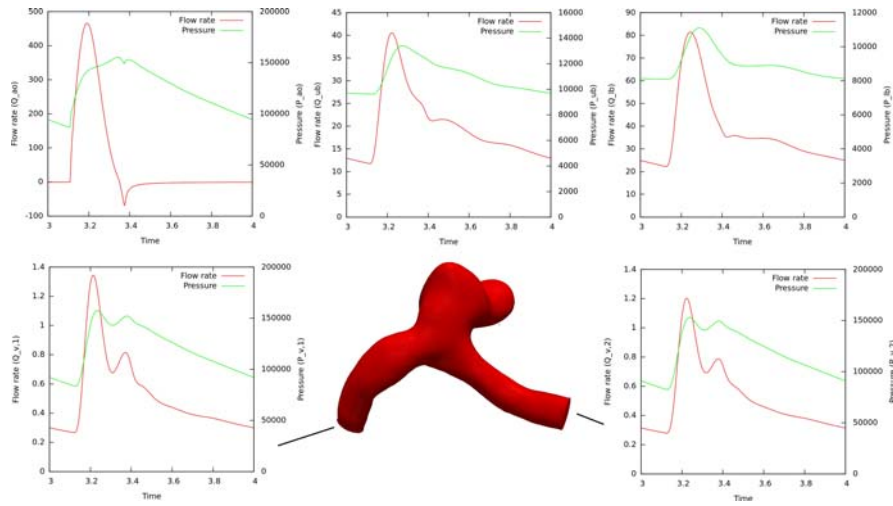


Fig. 4.10 Aneurysm geometry and results (pressure and flow rate) at coupling interfaces. Units are dyn/cm^2 for pressure and cm^3/s for flow rate.

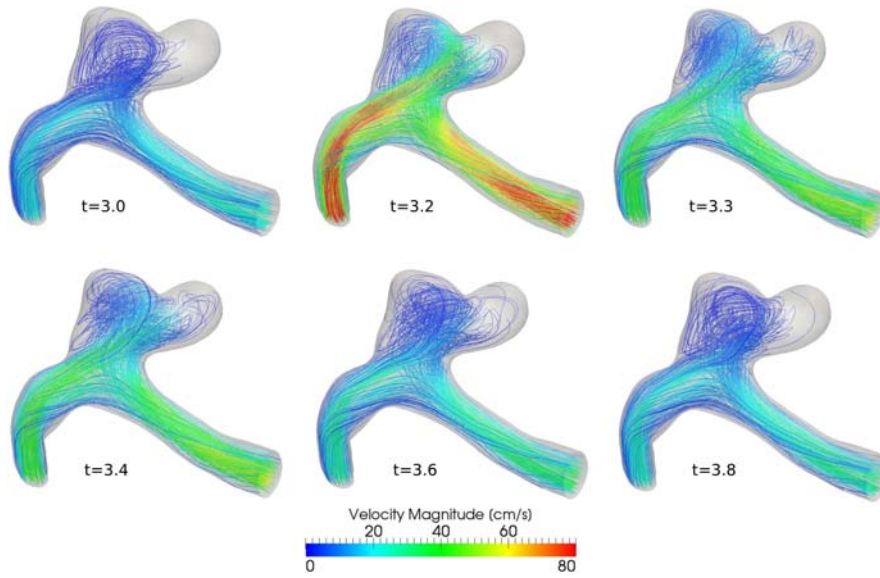


Fig. 4.11 Local blood flow in the aneurysm visualized through streamlines at different time instants within the cardiac cycle.

can generate problems in the balance of load. This is currently a matter of current research.

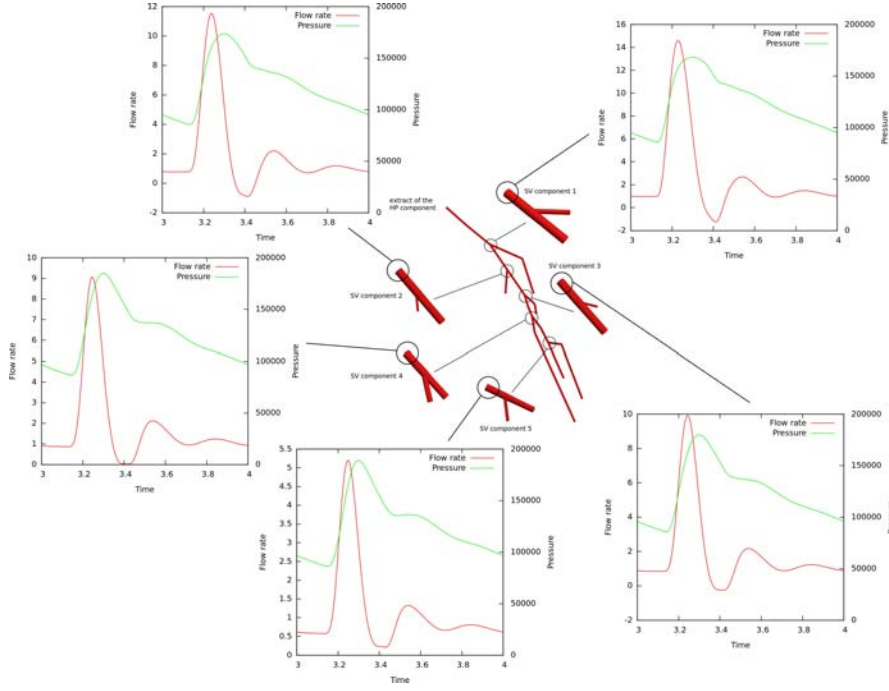


Fig. 4.12 Embedding of five SV components to replace arterial branches of the arm in the HP component. Results, pressure and flow rate, at coupling interfaces are displayed. Units are dyn/cm^2 for pressure and cm^3/s for flow rate.

We consider $\Delta t_G = 0.001$ and $a_{\text{HP}} = a_{\text{LP}} = a_{\text{SV}} = 1$. In Figure 4.12 the results in terms of pressure and flow rate at the inlets of the SV components is shown. In turn, Figure 4.13 shows the convergence history in the last cardiac cycle and the change in the pressure wave at the inlets of the SV components as we move forward in the distal direction. Even though the system of nonlinear equations comprises 36 interface unknowns, the iterative algorithm takes just 4 to 6 iterations (ocasionally 7 and 8) to converge at each time step.

Finally, Figure 4.14 presents the pressure field and the velocity magnitude inside each bifurcation at some time instants within the last cardiac cycle. Notice that each SV component is embedded in its corresponding hemodynamics environment, obtained from the closed-loop interaction provided by the HP-LP components. Clearly, a Womersley-like velocity arises after systole as a result of the inversion of the pressure gradient, and then the flow stabilizes its direction during diastole (see also Figure 4.12). In the fifth SV component this phenomenon is less evident, since back flow is not present after systole as in the rest of the SV components (see Figure 4.13).

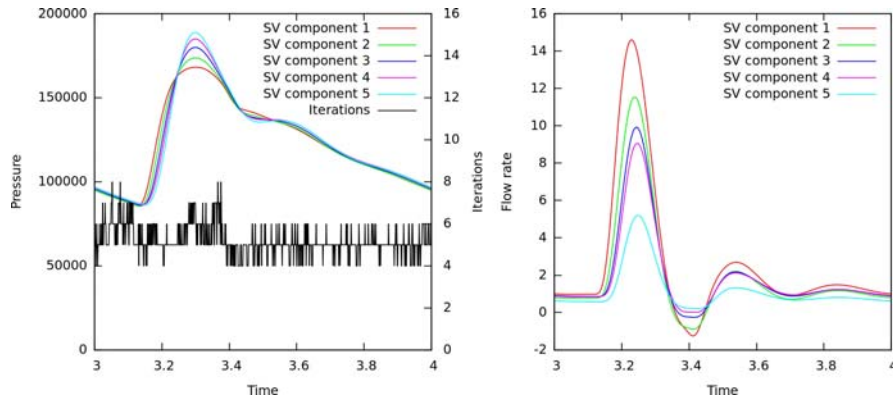


Fig. 4.13 Convergence history for the blood flow simulation in the arm and results at the inlets for all the SV components. Pressure units are dyn/cm^2 and flow rate units are cm^3/s .

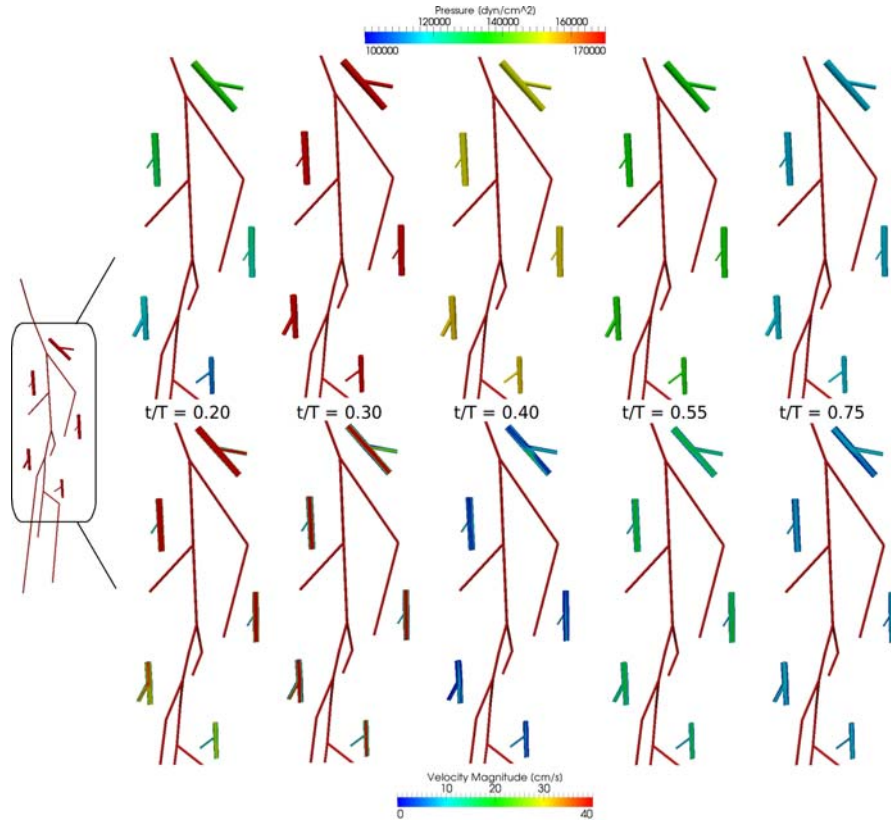


Fig. 4.14 Slices showing the arterial pressure and the blood flow in the SV components at different time instants.

This kind of application motivates the use of heterogeneous models in order to account for the effect of bifurcations (SV components) in the flow rate-pressure relation, for which simple and reliable models based on correlations are not available.

4.7.4 Blood flow in the entire aorta

Finally, in this last application we evaluate the performance of the iterative algorithm to couple a SV component with many coupling interfaces with the HP-LP system used in the previous sections. In this case such component is the largest arterial vessel in the human body, the aorta artery with its main branches as seen in Figure 4.15. The geometry of the aorta was obtained from a patient-specific computed tomography using standard image segmentation procedures. Notice that in such figure we omit the LP component for the sake of simplicity. As in the previous sections, the computational cost is up to the SV component, for which the number of degrees of freedom is 10,974,215. Altogether, the system of non-linear interface equations consists of 30 unknowns.

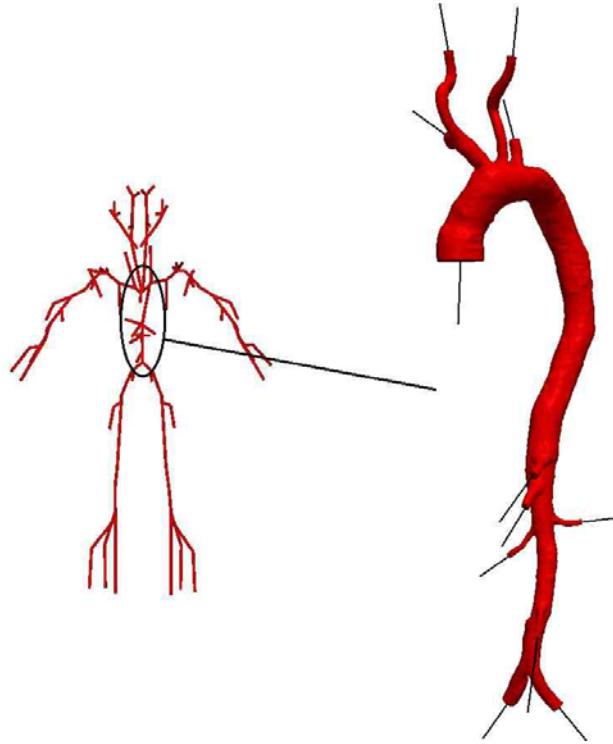


Fig. 4.15 Coupling of the entire aorta artery (SV component) to the closed-loop cardiovascular system formed by the HP and LP components.

For the simulation we take $\Delta t_G = 0.001$ and $a_{HP} = a_{LP} = a_{SV} = 1$, and two cardiac cycles are simulated. The number of iterations at each time step that the algorithm took to converge is shown in Figure 4.16. As happened with many *SV* components, in this case of several coupling unknowns the performance of the algorithm is satisfactory, taking between 3 and 11 iterations per time step to reach convergence, with an average, over a cardiac cycle, of 5.21 iterations per time step.

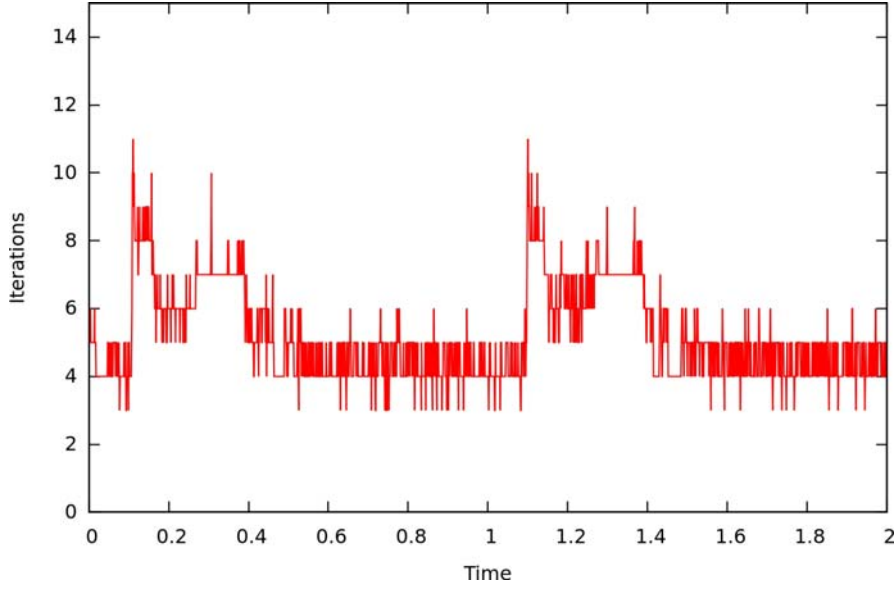


Fig. 4.16 Broyden iterations to reach convergence throughout two cardiac cycles.

In Figure 4.17 the results in terms of pressure and flow rate at the coupling interfaces between the *SV* and the *HP* components are shown for the last cardiac beat. Notice that well-known features of the cardiovascular system such as the distal smoothing of the diastolic notch, and the distal impingement pressure are reproduced by the model.

As stated in the introduction, three dimensional modeling is able to provide a full insight of the hemodynamics phenomena. For instance, the pressure distribution along the second cardiac cycle is featured in the sequence of images presented in Figure 4.18. After a sudden growth of the arterial pressure at the aortic root (Figure 18(b)), the development of the systolic flow occurs (Figure 18(c)) and an inversion of the pressure gradient is observed during diastole (Figures 18(d) and 18(e)) triggering complex blood flow patterns, as seen in Figure 4.19.

In Figure 4.19 the structure of the velocity field is presented through the computation of an isosurface of the velocity magnitude. The isosurface plotted corresponds to a magnitude equal to 8 cm/s. This isosurface shows clearly the systolic flow (Figures 19(a), 19(b) and 19(c)) and the subsequent development of backward flow and

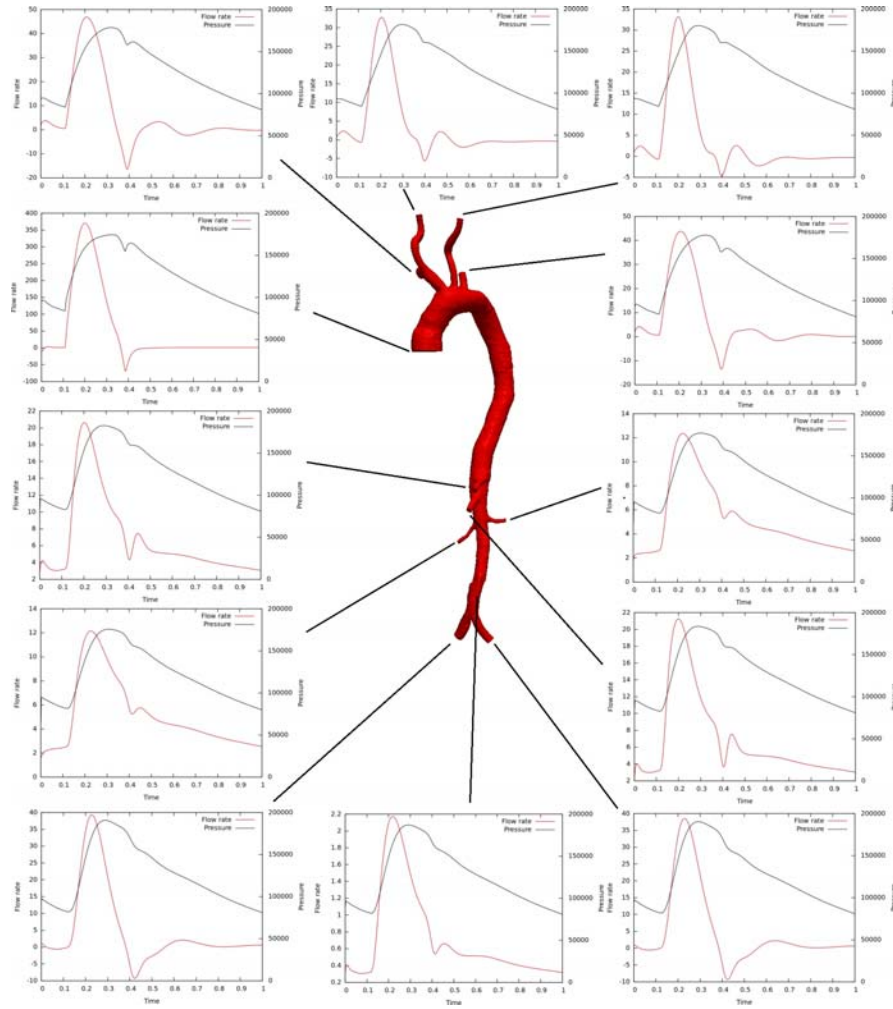


Fig. 4.17 Pressure and flow rate at coupling interfaces in the aorta artery (SV component). Units are dyn/cm^2 and flow rate units are cm^3/s .

complex vortical structures during diastole as a result of the rapid desacceleration and inverted pressure gradient (Figures 19(d), 19(e), 19(f) and 19(g)).

With this kind of analysis it is possible to study the complexity of the blood flow pattern throughout the entire aorta artery under the hemodynamics environment imposed by the surrounding HP-LP model. It is well-known that there are features of the three dimensional structure of the blood flow which are related to the onset and progress of some cardiovascular diseases. Recently, it has been postulated a correlation between aortic back-flow (due to aortic valve insufficiency) and atherosclerosis at the level of the thoracic aorta [28]. Using the present approach it is plausible to

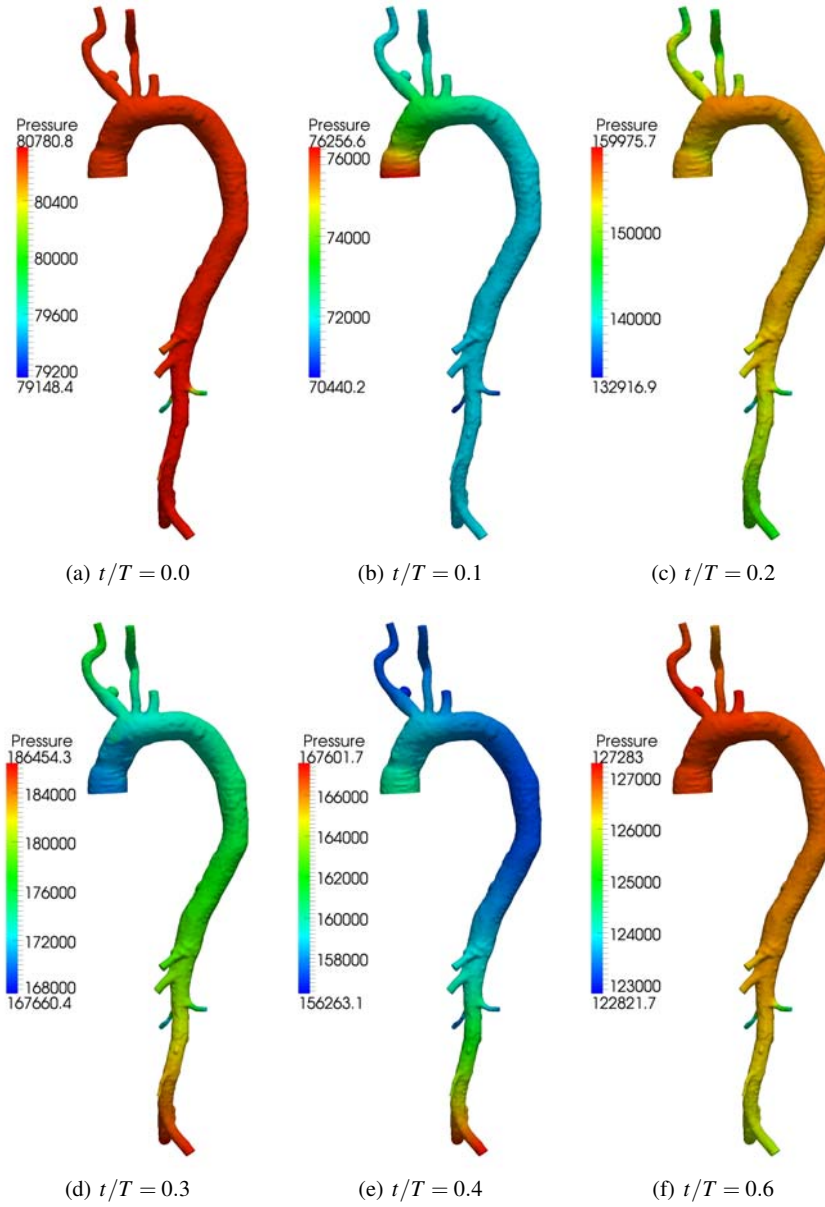


Fig. 4.18 Pressure in the aorta artery for different time instants in the second cardiac beat. Units are dyn/cm^2 .

provide a quantification of the stress exerted by the fluid over the arterial walls as a function of the severity of the pathological condition of the regurgitating valve. This is being matter of current research in the applications field.

4.8 Conclusions

In this article a black-box decomposition strategy for dealing with dimensionally-heterogeneous hemodynamics components has been presented. Its applicability to closed-loop 3D-1D-0D modeling of the cardiovascular system has been demonstrated through several detailed examples.

Let us now explain the idea in more general terms, so as to elucidate its wide applicability. Suppose you have a collection of systems (mechanical, thermal, electric, whatsoever), each of which you model with a black-box code. Assume, further, that you control this code by a set of inputs at its boundaries, and that each input of the set has some “conjugate”, or “associated” output. This conjugacy between an input (**I**) and an output (**O**) is to be understood in the following sense: At each boundary, the black-box code allows you to impose the value either of **I** or of **O**, but not both. Also, having imposed the value of **I** (temperature, displacement, pressure, voltage, ...) the code provides, as an output, the value of **O** (heat flux, force, flow rate, current, ...); and viceversa. For each model you have a preferred set of inputs which are easier to impose, or for which you have tuned some numerical parameters.

Assume now that you decide to *couple* these systems together into a larger system of which the aforementioned black-box models are *components*. In formal terms, to couple a set of components means that *the input and output values of component X at some boundary are required to equal some corresponding input/output values of another component Y*, this happening simultaneously for all of the coupling interfaces in the system. *Solving the coupled system* amounts to finding the value of each input of each black-box component, such that imposing these values and running all component codes the output of each code at each interface matches the corresponding input/output variable of the neighbor component with which that interface is shared.

At first, you would probably try to find the solution heuristically. You would choose that each output of each code corresponds to one input of the codes of the neighboring components, and iterate starting from some initial guess of values. After each component code is run, its outputs are retrieved and imposed as inputs to the neighboring components, and so on until, hopefully, a fixed point is reached, which would be the sought-for solution. This heuristic algorithm, however, is neither flexible nor effective. It is not flexible because you would have to resign your preferred set of inputs for each component, favoring inputs that coincide with outputs of the component's neighbors. And it is not effective because, in general, the convergence to the fixed point is either very slow or simply does not take place.

The algorithm described in this article provides a systematic way out of that conundrum. The essential point is to forget about considering the inputs as data and

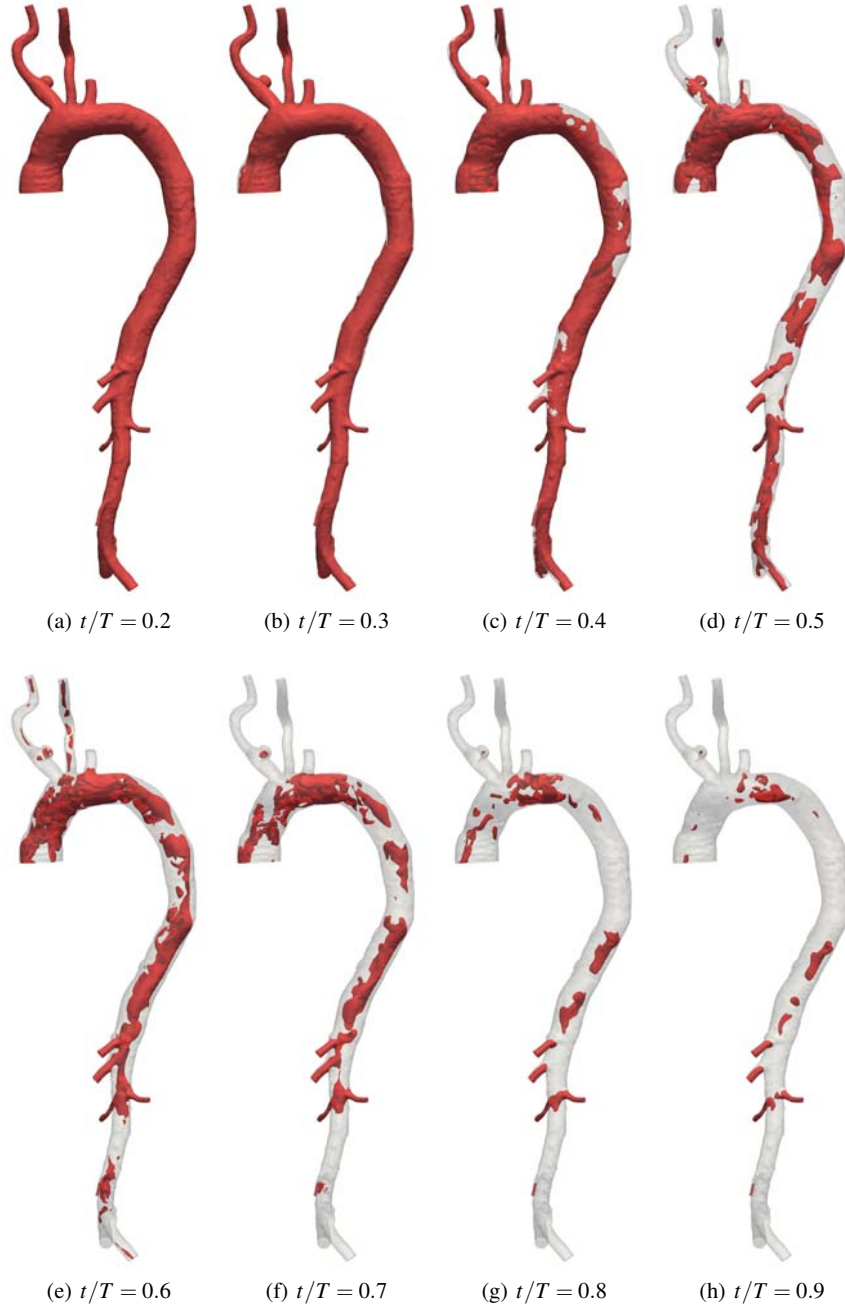


Fig. 4.19 Isosurface of velocity magnitude 8 cm/s for different time instans of the second cardiac cycle.

the outputs as results, and instead to consider as *unknowns* the vector consisting of *all* the coupling variables of *all* interfaces. The vector $\mathbf{X} = (\mathbf{Q}, \mathbf{P})$ of Section 4.5.3, consisting of flow rates *and* pressures at all the interfaces, is a specific example. Both conjugate variables of each interface (be it flow rate and pressure, displacement and force, etc.) are unknowns with equal standing, which allows us to write the set of equations (4.60) in terms of inputs and outputs of each code, both inputs *and* outputs being unknowns of the system. Notice in Table 4.1, for example, that P_{ao} is input for both HP and LP, and Q_{ao} is, consequently, output for both. Since both P_{ao} and Q_{ao} are unknowns with no hierarchy, we can choose which one is input and which is output at our will, independently for each of the two components (HP and LP) that share the aortic interface.

Let us go back to the system of equations (4.60) and elaborate on a more intuitive interpretation. Assume we have a candidate solution vector \mathbf{X} . Some of the entries in \mathbf{X} correspond to the input variables of component HP, for example. The rather abstract operator \mathcal{G}_{HP}^I in the first equation of (4.60) tells us, simply, “run the code of component HP and compute its outputs”. These outputs also correspond to specific entries in \mathbf{X} , and that equation tells us to compare the values in \mathbf{X} with those yielded by the HP code. If they are the same, the first equation of (4.60) is satisfied, and if this happens for all the components, the candidate vector \mathbf{X} is indeed a solution of the *coupled system*. If not, a *residual* will appear from the set of equations, by simply subtracting values. This residual, or better the procedure for building it so described, corresponds to Step 8 in Algorithm 2 (i.e., Algorithm 3), the only system-dependent step of the procedure proposed here. The rest is numerical technology for solving systems of equations having access just to the residuals, and Broyden’s algorithm is not the only choice but instead the one that has proved to be more robust and effective in the (many) tests that we have conducted.

Our whole contribution, thus, is already contained in equation (4.60). In a nutshell, it amounts to separating the “input versus output” reasoning, which we use to build the equations of each component, from the “datum versus result” reasoning, which is overruled once the system is coupled, both inputs and outputs being, simply, *unknowns*.

With some imagination, the proposed approach can be applied to a large variety of biomechanical systems, though not to all. Mathematically speaking, the components have to behave elliptically. In other words, they must *need* a boundary condition at each interface and they must yield a conjugate variable there. Subsonic, rigid-wall fluidic systems behave in this way, as do flexible-wall fluidic if the fluid velocity is smaller than that of the deformation waves. This is the case of practically all hemodynamic flows, justifying the specific applications discussed above. If the flow had *supercritical* regions, that is, regions in which the local velocity exceeded that of the deformation waves, the straightforward algorithm proposed here would not be adequate. Further work is needed to generalize the methodology so as to encompass non-elliptic-like components.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. The support of these agencies is gratefully acknowledged.

References

- [1] Avolio, A.: Multi-branched model of the human arterial system. *Med. Biol. Engrg. Comp.* **18**, 709–718 (1980).
- [2] Bertoluzza, S.: Substructuring preconditioners for the three fields domain decomposition method. *Math. Comp.* **73**, 659–689 (2003).
- [3] Blanco, P., Feijóo, R., Urquiza, S.: A unified variational approach for coupling 3D–1D models and its blood flow applications. *Comp. Meth. Appl. Mech. Engrg.* **196**, 4391–4410 (2007).
- [4] Blanco, P., Pivello, M., Urquiza, S., Feijóo, R.: On the potentialities of 3D–1D coupled models in hemodynamics simulations. *J. Biomech.* **42**, 919–930 (2009).
- [5] Blanco, P., Urquiza, S., Feijóo, R.: Assessing the influence of heart rate in local hemodynamics through coupled 3D-1D-0D models. *Int. J. Num. Meth. Biomed. Engrg.* **26**, 890–903 (2010).
- [6] Blanco, P., Leiva, J., Feijóo, R., Buscaglia, G.: Black-box decomposition approach for computational hemodynamics: One-dimensional models. *Comp. Meth. Appl. Mech. Engrg.* **200**, 1389–1405 (2011).
- [7] Blanco, P., Discacciati, M., Quarteroni, A.: Modeling dimensionally-heterogeneous problems: analysis, approximation and applications. *Numerische Mathematik* **119**, 299–335 (2011).
- [8] Blanco, P., Trenhago, P., Fernandes, L., Feijóo, R.: On the integration of the baroreflex control mechanism in a heterogeneous model of the cardiovascular system. *Int. J. Num. Meth. Biomed. Engrg.* **28**, 412–433 (2012).
- [9] Blanco, P., Feijóo, R.: The role of the variational formulation in the dimensionally-heterogeneous modelling of the human cardiovascular system. In: *Modeling of Physiological Flows*, D. Ambrosi, A. Quarteroni and G. Rozza eds., Springer, Italy, 251–288 (2012).
- [10] Brezzi, F., Marini, D.: A three-field domain decomposition method. In: *Domain Decomposition Methods in Science and Engineering*, Quarteroni et al. eds., American Mathematical Society, **157** 27–34 (1994).
- [11] Cebal, J., Yim, P., Löhner, R., Soto, O., Choyke, P.: Blood flow modeling in carotid arteries with computational fluid dynamics and MR Imaging. *Acad. Radiol.* **9**, 1286–1299 (2002).
- [12] Dennis, J., Schnabel, R.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Classics in applied mathematics, SIAM (1996).
- [13] Formaggia, L., Gerbeau, J., Nobile, F., Quarteroni, A.: On the coupling of 3D and 1D Navier–Stokes equations for flow problems in compliant vessels. *Comp. Meth. Appl. Mech. Engrg.* **191**, 561–582 (2001).
- [14] Formaggia, L., Nobile, F., Quarteroni, A., Veneziani, A.: Multiscale modelling of the circulatory system: a preliminary analysis. *Comp. Vis. Science* **2**, 75–83 (1999).
- [15] Grinberg, L., Anor, T., Madsen, J., Yakhot, A., Karniadakis, G.: Large-scale simulation of the human arterial tree. *Clinical and Experimental Pharmacology and Physiology* **36**, 194–205 (2009).
- [16] Grinberg, L., Cheever, E., Anor, T., Madsen, J., Karniadakis, G.: Modeling blood flow circulation in intracranial arterial networks: A comparative 3D/1D simulation study. *Annals of Biomedical Engineering* **39**, 297–309 (2011).

- [17] Heywood, J.G., Rannacher, R., Turek, S.: Artificial boundaries and flux and pressure conditions for the incompressible Navier–Stokes equations. *Int. J. Num. Meth. Fluids* **22**, 325–352 (1996).
- [18] Hughes, T.: On the one-dimensional theory of blood flow in the larger vessels. *Mathematical Biosciences* **18**, 161–170 (1973).
- [19] Kelley, C.: *Iterative Methods for Linear and Nonlinear Equations*. Frontiers in applied mathematics, SIAM (1995).
- [20] Kim, H., Vignon-Clementel, I., Figueroa, C., LaDisa, J., Jansen, K., Feinstein, J., Taylor, C.: On coupling a lumped parameter heart model and a three-dimensional finite element aorta model. *Annals of Biomedical Engineering* **37**, 2153–2169 (2009).
- [21] Korakianitis, T., Shi, Y.: Numerical simulation of cardiovascular dynamics with healthy and diseased heart valves. *J. Biomech.* **39**, 1964–1982 (2006).
- [22] Leiva, J., Blanco, P., Buscaglia, G.: Iterative strong coupling of dimensionally-heterogeneous models. *Int. J. Num. Meth. Engng.* **81**, 1558–1580 (2010).
- [23] Leiva, J., Blanco, P., Buscaglia, G.: Partitioned analysis for dimensionally-heterogeneous hydraulic networks. *SIAM Multiscale Modeling and Simulation* **9**, 872–903 (2011).
- [24] Liang, F., Takagi, S., Himeno, R., Liu, H.: Multi-scale modeling of the human cardiovascular system with applications to aortic valvular and arterial stenoses. *Med. Biol. Eng. Comput.* **47**, 743–755 (2009).
- [25] Quarteroni, A., Veneziani, A.: Analysis of a geometrical multiscale model based on the coupling of PDE's and ODE's for blood flow simulations. *SIAM J. on Multiscale Model. Simul.* **1**, 173–195 (2003).
- [26] Quarteroni, A., Ragni, S., Veneziani, A.: Coupling between lumped and distributed models for blood flow problems. *Comp. Vis. Science* **4**, 111–124 (2001).
- [27] Quarteroni, A., Veneziani, A.: Coupling between lumped and distributed models for blood flow problems. In: *Computational Science for 21st Century*, Lions et al. eds.. John Wiley and Sons, 339–350 (1997).
- [28] Shimoni, S., Zilberman, L., Edri, O., Bar, I., Golland, S., Gendelman, G., Swissa, M., Livshitz, S., Paz, O., Ayzenberg, O., George, J.: Thoracic aorta atherosclerosis in patients with aortic regurgitation. *Atherosclerosis* **218**, 107–109 (2011).
- [29] Stergiopoulos, N., Young, D., Rogge, T.: Computer simulation of arterial flow with applications to arterial and aortic stenoses. *J. Biomech.* **25**, 1477–1488 (1992).
- [30] Taylor, C., Hughes, T., Zarins, C.: Finite element modeling of blood flow in arteries. *Comp. Meth. Appl. Mech. Engng.* **158**, 155–196 (1998).
- [31] Urquiza, S., Blanco, P., Vénere, M., Feijóo, R.: Multidimensional modelling for the carotid artery blood flow. *Comp. Meth. Appl. Mech. Engng.* **195**, 4002–4017 (2006).
- [32] Vignon-Clementel, I., Figueiroa, C., Jansen, K., Taylor, C.: Outflow boundary conditions for three-dimensional finite element modeling of blood flow and pressure waves in arteries. *Comp. Meth. Appl. Mech. Engng.* **195**, 3776–3996 (2006).
- [33] Wang, J., Parker, K.: Wave propagation in a model of the arterial circulation. *J. Biomech.* **37**, 457–470 (2004).

Chapter 5

Mathematical and Statistical Models for Medical Image Processing

Gilson A. Giraldi, Paulo S. S. Rodrigues, Denise Guliato and Antonio Lopes Apolinário Jr.

Abstract This chapter is focused on elements of functional analysis, statistical learning and information theory for image processing. Theoretical models will be discussed using entropy for compression and segmentation, neural networks for image information processing and fractals/wavelets for data representation and dimensionality reduction. A review will be made, presenting some aspects of wavelet theory and fractal geometry using elegant functional frameworks. Subsequently, the nonextensive Tsallis information theory and multilayer perceptron will be considered. Afterward, two applications of the mentioned models will be presented: 2D shape characterization pipeline and image segmentation. Finally, some perspectives for wavelets and Tsallis information theory in image processing will be discussed.

Gilson A. Giraldi

LVCVRV - Scientific Visualization and Virtual Reality Laboratory an Associated Laboratory of INCT-MACC and LNCC - National Laboratory for Scientific Computing, Av. Getúlio Vargas 333, Quitandinha, 25651-075, Petrópolis, RJ, Brazil, e-mail: gilson@lncc.br

Paulo S. S. Rodrigues

LVCVRV - Scientific Visualization and Virtual Reality Laboratory an Associated Laboratory of INCT-MACC and FEI University Center, São Bernardo do Campo, SP, Brazil, e-mail: psergio@fei.edu.br

Denise Guliato

LBD - Database Laboratory an Associated Laboratory of INCT-MACC and Federal University of Uberlândia, MG, Brazil, e-mail: dguliato@gmail.com

Antonio Lopes Apolinário Jr.

LVCVRV - Scientific Visualization and Virtual Reality Laboratory an Associated Laboratory of INCT-MACC and Federal University of Bahia, Salvador, Bahia, Brazil, e-mail: apolinario@dcc.ufba.br

5.1 Introduction

In the last decades, medical imaging techniques have been applied to scan human body for clinical purposes (diagnoses and surgical planning) and medical science studies like anatomy and physiology. These areas incorporate imaging technology based on radiography (projection or fluoroscopy), nuclear medicine, endoscopy, microscopy (for human pathological investigations), tomography, magnetic resonance imaging, ultrasound, among others [28]. From the viewpoint of computer science, each one of these modalities needs specific algorithms and mathematical methods for automating the process of image analysis [30].

In this sense, the area of medical image processing is a very active one that always requires innovative techniques. Image processing in general, and medical image processing in particular, can be roughly divided in: reconstruction, representation, image enhancement, restoration, filtering, segmentation, image analysis and compression [14, 9].

The reconstruction area, the background for computerized tomography, has the mathematical model based on the Radon Transform [19]. The techniques for digital image representation compose a class of methods used to characterize the information therein [9, 14]. A classical approach for image representation is based on unitary matrices that define the so called *image transforms*. The Discrete Fourier Transform (**DFT**) is the most known example in this class [14]. Multiresolution representations, such Pyramid [15] and Wavelets [32], compose another important class in this field.

Image enhancement refers to operations used to highlight image features such as edges, boundaries or contrast, in order to make a digital image more useful for display and/or analysis. Image restoration is concerned with processing the acquired image in order to reduce the effects of degradations due to sensor noise, blur due to camera focus problems, relative object-camera motion, random factors related to atmospheric events, and so on [14].

The enhancement and restoration techniques depend on some kind of filtering process which can be formulated in discrete or continuous spaces depending on the corresponding signal representation [27, 14].

Segmentation is a fundamental step in image analysis. From a practical point of view, segmentation is the partition of an image into multiple regions (sets of pixels) according to some criterion of homogeneity of features such as color, shape, texture and spatial relationship. These fundamental regions are disjoint sets of pixels and their union compose the original whole scene [9]. Once completed the segmentation, the image can be considered as a composition of several objects. So, the image analysis can be started, which is the extraction of meaningful information from digital images. This field combines techniques for features extraction and pattern recognition [29].

Image data compression is concerned with minimizing the number of bits required to represent an image. Information theory, which deals with measurement and transmission of information through a channel, is an important tool for this area (see [20], Chapter 11).

This chapter is focused on mathematical and statistical models for image processing. The models are based on elements of functional analysis, statistical learning and information theory. Specifically, a review of some aspects of wavelet theory and fractal geometry using two elegant functional frameworks will be made in sections 5.2 and 5.3, respectively. Then, the nonextensive Tsallis information theory (section 5.4) and multilayer perceptron (section 5.5) are considered with two goals in mind: information quantification and processing.

Next, in section 5.6, we describe two applications of the considered models; the first one is a 2D shape characterization pipeline and the second one in segmentation tasks. The first application focus on the segmentation of ultrasound images of breast lesions through the non-extensive entropy segmentation technique described in [25]. In the second application the aim is to characterize 2D and 3D shapes of breast lesions as malignant tumor or benign mass. In this way, it was proposed a new shape descriptor based on the Hilbert space-filling curve, a kind of fractal structure, discrete wavelet transform and a neural network.

Following, section 5.7 presenting perspectives for solving PDE problems in image processing using wavelets and in nonextensive information theory. Finally, section 5.8 presents the final comments.

5.2 Wavelet Theory

In this section, some aspects of discrete wavelet theory will be reviewed. The basic elements of orthonormal wavelet basis and multiscale analysis will be presented. The section ends with the discrete wavelet transform presentation. The text is organized based on the references [18, 23, 4, 6]. There are many sites that can be consulted to help the beginners in wavelet theory [21]. Besides, toolboxes can be found for the MatLab package [13].

A multiscale analysis (MSA) of $L^2(\mathfrak{R})$ is an increasing sequence of closed subspaces, called **scale spaces**, $V_m \subset L^2(\mathfrak{R})$:

$$\{0\} \subset \cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots \subset L^2(\mathfrak{R}) \quad (5.1)$$

such that the following are true [18]:

$$\overline{\bigcup_{m \in \mathbb{Z}} V_m} = L^2(\mathfrak{R}), \quad (5.2)$$

$$\bigcap_{m \in \mathbb{Z}} V_m = \{0\}, \quad (5.3)$$

$$f(x) \in V_m \iff f(2^m x) \in V_0 \quad (5.4)$$

There is a function $\varphi \in L^2(\mathfrak{R})$, named **scaling function**, whose integer translates generate a Riesz basis of V_0 (that mean, V_0 is the closure of the set $\text{span} \{ \varphi_{m,k} \mid k \in \mathbb{Z} \}$):

$$V_0 = \overline{\text{span}\{\varphi(x-k) \mid k \in \mathbb{Z}\}}$$

and

$$A \sum_{k \in \mathbb{Z}} c_k^2 \leq \left\| \sum_{k \in \mathbb{Z}} c_k \varphi(\cdot - k) \right\|_{L^2}^2 \leq B \sum_{k \in \mathbb{Z}} c_k^2 \quad (5.5)$$

for all $\{c_k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$, where A and B are positive constants.

A remarkable consequence of the above properties is the fact that the space V_m is spanned by the functions:

$$\varphi_{m,k}(x) := 2^{-m/2} \varphi(2^{-m}x - k), \quad (5.6)$$

$$V_m = \overline{\text{span}\{\varphi_{m,k} \mid k \in \mathbb{Z}\}}. \quad (5.7)$$

This is based on (5.4) and (5.5). The functions in (5.6) all have the same L^2 -norm $\|\varphi_{m,k}\|_{L^2} = \|\varphi\|_{L^2}$.

The scaling function φ satisfies a *scaling equation*, i.e. there is a sequence $\{h_k\}_{k \in \mathbb{Z}}$ of real numbers such that:

$$\varphi(x) = \sqrt[2]{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k). \quad (5.8)$$

This follows from the fact that $\varphi \in V_0 \subset V_{-1} = \overline{\text{span}\{\sqrt[2]{2}\varphi(2x - k) \mid k \in \mathbb{Z}\}}$.

5.2.1 Wavelet Spaces

The key for the construction of both orthogonal wavelet bases and fast algorithms lies in equation (5.8). In this way, W_m can be denoted as the orthogonal complement of V_m in V_{m-1} , that means:

$$V_{m-1} = W_m \oplus V_m, \quad V_m \perp W_m. \quad (5.9)$$

If Q_m is the orthogonal projector of $L^2(\mathfrak{R})$ in W_m and P_{m-1} denotes the orthogonal projector onto V_{m-1} then the decomposition (5.9) means:

$$P_{m-1} = Q_m + P_m. \quad (5.10)$$

From (5.9) it follows that:

$$V_{m-1} = W_m \oplus (W_{m+1} \oplus V_{m+1}), \quad V_{m+1} \perp W_{m+1}, \quad (5.11)$$

and so on. Therefore, using the MSA definition it follows that:

$$V_{m-1} = \bigoplus_{j \geq m} W_j \quad (5.12)$$

and so:

$$L^2(\mathfrak{R}) = \bigoplus_{j \in \mathbb{Z}} W_j. \quad (5.13)$$

The spaces W_m , with $m \in \mathbb{Z}$, are named **wavelet spaces**. Let $\{V_m\}_{m \in \mathbb{Z}}$ be an MSA generated by the orthogonal scaling function $\varphi \in V_0$ and the function $\psi \in V_{-1}$, defined by

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2x - k) = \sum_{k \in \mathbb{Z}} g_k \varphi_{-1,k}(x), \quad (5.14)$$

$$g_k = (-1)^k h_{1-k}, \quad (5.15)$$

where $\{h_k\}_{k \in \mathbb{Z}}$ are the coefficients of the scaling equation (5.8).

Theorem 5.1. *The function ψ defined by expressions (5.14)-(5.15) has the following properties [18]:*

- i. $\{\psi_{m,k}(x) = 2^{-m/2} \varphi(2^{-m}x - k) \mid k \in \mathbb{Z}\}$ is an orthonormal basis for W_m ,
- ii. $\{\psi_{m,k} \mid m, k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\mathfrak{R})$,
- iii. ψ is a wavelet; that means, $\psi \in L^2$ and satisfies the admissibility condition

$$0 < c_\psi := 2\pi \int_{\mathfrak{R}} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty.$$

A known example of MSA is given by the Haar Wavelets. Let the Haar scaling function given by:

$$\varphi(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases} \quad (5.16)$$

The corresponding space V_m is generated by expressions (5.6)-(5.7) and the wavelet space W_m is given by Theorem 5.1. In [6] it is presented a study of the MSA generated through the Haar scaling function.

5.2.2 Discrete Wavelet Transform

In this section, the basic algorithms for the fast computation of the discrete wavelet transform will be introduced. It is performed in the context of multiscale analysis and by applying the scaling equation.

Consider a function f in V_0 , the basic space of a multiscale analysis, as well as an orthogonal scaling function $\varphi \in V_0$. From expression (5.7) we know f has the expansion:

$$f(x) = \sum_{k \in \mathbb{Z}} c_k^0 \varphi(x - k). \quad (5.17)$$

As previously, ψ denote the orthogonal wavelet corresponding to the scaling function φ , which generates an orthonormal basis:

$$\left\{ \psi_{m,k} = 2^{-m/2} \varphi(2^{-m} \cdot -k) \mid m, k \in \mathbb{Z} \right\},$$

of $L^2(\mathbb{R})$, according to the Theorem 5.1.

Now, the discrete wavelet transform can be computed, which is given by the scalar products:

$$\sqrt{c_\psi} L_\psi f(2^m, 2^m k) = \langle f, \psi_{m,k} \rangle_{L^2}, \quad m, k \in \mathbb{Z}.$$

The following notation will be used:

$$d_k^m = \langle f, \psi_{m,k} \rangle_{L^2}, \quad d^m = \{d_k^m \mid k \in \mathbb{Z}\} \in l^2(\mathbb{Z}),$$

$$c_k^m = \langle f, \varphi_{m,k} \rangle_{L^2}, \quad c^m = \{c_k^m \mid k \in \mathbb{Z}\} \in l^2(\mathbb{Z}).$$

Applying the scaling equations (5.8) and (5.14), it can be derived:

$$d_k^m = \langle f, \psi_{m,k} \rangle_{L^2} = \sum_{l \in \mathbb{Z}} g_l \langle f, \varphi_{m-1, 2k+l} \rangle_{L^2} = \sum_{l \in \mathbb{Z}} g_{l-2k} c_l^{m-1},$$

$$c_k^m = \langle f, \varphi_{m,k} \rangle_{L^2} = \sum_{l \in \mathbb{Z}} h_l \langle f, \varphi_{m-1, 2k+l} \rangle_{L^2} = \sum_{l \in \mathbb{Z}} h_{l-2k} c_l^{m-1}.$$

Thus the decomposition algorithm is already completed: starting with the sequence c^0 the discrete wavelet decomposition can be computed recursively, using discrete convolutions. It is important to observe that all operations are carried out discretely on the sequences of coefficients c^m and d^m .

This process can be better expressed with the help of the decomposition operators H and G , defined by:

$$H : l^2(\mathbb{Z}) \longrightarrow l^2(\mathbb{Z})$$

$$c \longmapsto Hc = c *_2 h = \left\{ (Hc)_k = \sum_{l \in \mathbb{Z}} h_{l-2k} c_l \right\}, \quad (5.18)$$

$$G : l^2(\mathbb{Z}) \longrightarrow l^2(\mathbb{Z})$$

$$c \longmapsto Gc = c *_2 g = \left\{ (Gc)_k = \sum_{l \in \mathbb{Z}} g_{l-2k} c_l \right\}, \quad (5.19)$$

where $h = \{h_k \mid k \in \mathbb{Z}\}$ and $g = \{g_k \mid k \in \mathbb{Z}\}$ are, respectively, the sequences of scaling and wavelet coefficients.

It shall be observed that in these expression it was not used the standard discrete convolution - in the convolutions above it enters only the second index. That is way the symbol $*_2$ for this operation should be used. The decomposition of the initial sequence c^0 in the sequences of coefficients $\{c^m, d^m \mid m = 1, \dots, M\}$, can be summarized in the Algorithm 1.

Observe that, by keeping c^M and $d^m, m = 1, \dots, M$, it is possible to recursively reconstruct the sequence c^0 . In fact, fast wavelet reconstruction can be done [4].

Algorithm 1 Fast Wavelet Transform

Input: $c^0 = \{c_k \mid k \in Z\}$
 M levels of decomposition (number of scales)
for $m = 1, \dots, M$ **do**
 $d^m = Gc^{m-1}$

$$c^m = Hc^{m-1}$$

end for

Output: c^M
 $d^m, m = 1, \dots, M$

$$\begin{array}{ccccccc} c^0 & \xrightarrow{H} & c^1 & \xrightarrow{H} & c^2 & \dots & c^{M-1} & \xrightarrow{H} & c^M \\ & \searrow & & \searrow & & & & \searrow & \\ & d^1 & & d^2 & & \dots & & d^M \end{array}$$

5.3 Fractal Geometry

This section follows the presentation of [2]. All the definitions and proofs can be found in that reference.

Let (X, d) be a complete metric space and $\mathbf{H}(X)$ the space whose points are the compact subsets of X , other than the empty set. Let $x \in X$ and $A, B \in \mathbf{H}(X)$. Then, the distance from the point x to the set B and the distance from set A to set B can be defined, respectively, as:

$$d(x, B) = \text{Min} \{d(x, y); \quad y \in B\}.$$

$$d(A, B) = \text{Max} \{d(x, B); \quad x \in A\}.$$

Besides, the **Hausdorff distance** between points (sets) $A, B \in \mathbf{H}(X)$ is defined by:

$$h(A, B) = \text{Max} \{d(A, B), d(B, A)\}.$$

It can be prove that the metric space $(\mathbf{H}(X), h)$ is also complete and it is called the **space of fractals** [2].

Definition 5.1. A transformation, or a function, $f : X \rightarrow X$ on a metric space (X, d) is called contractive or a **contraction mapping** if there is a constant $0 \leq s < 1$ such that:

$$d(f(x), f(y)) \leq s \cdot d(x, y), \quad \forall x, y \in X.$$

Any such number s is called the **contractivity factor** for f . A point $x_f \in X$ such that $f(x_f) = x_f$ is called a **fixed point** of the transformation.

Theorem 5.2. Let $f : X \rightarrow X$ be a contraction mapping on a complete metric space (X, d) . Then f possesses exactly one fixed point $x_f \in X$ and moreover for any point

$x \in X$, the sequence $\{f^{(n)}(x); n = 0, 1, 2, \dots\}$ converges to x_f ($f^{(n)} = f \circ f \circ \dots \circ f$, n times). That is::

$$\lim_{n \rightarrow \infty} f^{(n)}(x) = x_f, \quad \forall x \in X.$$

A beautiful result is that a **Deterministic Fractal** can be defined as a fixed point of a contraction mapping on $(\mathbf{H}(X), h)$. The following theorems characterizes what kind of contraction transformations the fractal theory deals with.

Theorem 5.3. Let $w : X \rightarrow X$ be a contraction mapping on the metric space (X, d) with contractivity factor s . Then $w : \mathbf{H}(X) \rightarrow \mathbf{H}(X)$ defined by:

$$w(B) = \{w(x); x \in B\},$$

for any $B \in \mathbf{H}(X)$ is a contraction mapping on $(\mathbf{H}(X), h)$ with contractive factor s .

Theorem 5.4. Let (X, d) be a metric space and $\{w_n : \mathbf{H}(X) \rightarrow \mathbf{H}(X), n = 1, 2, \dots, N\}$ a set of contraction mapping in the metric space $(\mathbf{H}(X), h)$. Let the contractivity factor for w_n be denoted by s_n for each n . Define $W : \mathbf{H}(X) \rightarrow \mathbf{H}(X)$ by:

$$W(B) = \cup_{n=1}^N w_n(B),$$

for each $B \in \mathbf{H}(X)$. Then, W is a contraction mapping on $(\mathbf{H}(X), h)$ with contractive factor $s = \max\{s_n; n = 1, 2, \dots, N\}$.

Definition 5.2. A (hyperbolic) iterated function system (**IFS**) consists of a complete metric space (X, d) together with a finite set of contraction mappings $w_n : X \rightarrow X$, with respective contractivity factors s_n , for $n = 1, \dots, N$.

Theorem 5.5. Let a hyperbolic iterated function system

$$\{w_n : X \rightarrow X; w_n, n = 1, 2, \dots, N\}$$

with contractive factor $s = \max\{s_n; n = 1, 2, \dots, N\}$. Then the transformation

$$W : \mathbf{H}(X) \rightarrow \mathbf{H}(X)$$

defined by:

$$W(B) = \cup_{n=1}^N w_n(B),$$

for all $B \in \mathbf{H}(X)$, is a contraction mapping on the complete metric space $(\mathbf{H}(X), h)$ with contractivity factor s .

Therefore:

$$h(W(B), W(C)) \leq s \cdot h(B, C),$$

for all $B, C \in \mathbf{H}(X)$. Its unique fixed point, $A \in \mathbf{H}(X)$ obeys:

$$A = W(A) = \cup_{n=1}^N w_n(A),$$

and is given by

$$A = \lim_{n \rightarrow \infty} W^{(n)}(B),$$

for any $B \in \mathbf{H}(X)$. This fixed point is called the attractor of the IFS.

A **deterministic fractal is an attractor of an IFS**. However, careful must be taken because there are attractors that are not fractals. There are lots of interesting and beautiful examples of deterministic fractals. Among them, the Hilbert curve, also known as Hilbert space-filling curve, is a continuous path that passes through every point of an n -dimensional grid and never crosses itself [22]. The one-dimensional representation of a volume scanned following the path defined by the Hilbert curve keeps the spatial correlation of the voxels, as close as possible.

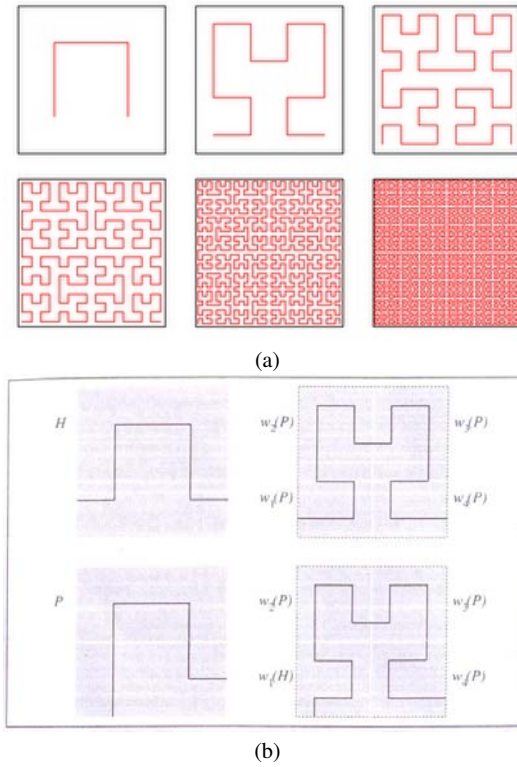


Fig. 5.1 (a) Six stages of Hilbert curve generation (source: fortygrins.blogspot.com). (b) IFS for Hilbert curve (source: [22]).

In fact, the Hilbert curve is a deterministic fractal that can be computed as an attractor of an IFS. The Figure 5.1.(a) shows some stages of that fractal generation. To see this, a suitable hyperbolic iterated function system must be defined. So, consider the contraction mappings $w_n : \mathbb{R}^2 \rightarrow \mathbb{R}^2, n = 1, 2, 3, 4$, with respective contractivity factors $s = 1/2$:

$$w_1(x, y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}, \quad (5.20)$$

$$w_2(x, y) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}, \quad (5.21)$$

$$w_3(x, y) = \begin{pmatrix} -1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \quad (5.22)$$

$$w_4(x, y) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}, \quad (5.23)$$

$$(5.24)$$

These mappings are composed by simple scale, rotations and translations. For instance, the w_4 is composed by a scale, a 90° rotation followed by a translation. In [22] it is claimed that the following (generalized) IFS scheme generates the Hilbert curve construction sequence of Figure 5.1.(a), provided that H_0 is the curve in the upper left corner of Figure 5.1.(b) and P_0 is the curve pictured in the lower left corner of Figure 5.1.(b).

5.4 Information Theory and Compression Problem

A fundamental work in this area is the Shannon's Information Theory (see [20], Chapter 11), which provides many useful tools that are based on measuring information in terms of the complexity of structures needed to encode a given piece of information.

Shannon's theory solves a central problems for classical information: How much can a message be compressed; i.e., how redundant is the information? (*The noiseless coding theorem*).

In this theory, the information and the transmission channel are formulated in a probabilistic point of view. In particular, the concept of entropy appears as a measure of the information of a system. In this context, a message is defined as a string of n letters chosen from an alphabet of W letters:

$$A = \{a_1, a_2, \dots, a_W\}$$

Let us suppose a priori probability distribution p :

$$p(a_i) = p_i, \quad \text{such that} \quad \sum_{i=1}^W p(a_i) = 1. \quad (5.25)$$

For example, the simplest case is for a binary alphabet where $p(1) = p$ and $p(0) = 1 - p$; $0 \leq p \leq 1$.

For n very large, the law of large numbers tells us that typical strings will contain (in the binary case) about $n(1 - p)$ 0's and about np 1's. The number of distinct

strings of this form is given by the binomial coefficient $\binom{n}{np}$. From Stirling approximation [20] it is known that $\log(n!) = n \log n - n + O(\log n)$. Thus, the binomial coefficient can be approximated by:

$$\log \binom{n}{np} \simeq nH(p), \quad (5.26)$$

where:

$$H(p) = -p \log p - (1-p) \log (1-p) \quad (5.27)$$

is the entropy function, called Shannon's entropy (observe that log's have base 2).

Thus, from equation (5.26) it can be seen that the number of typical strings is of order $2^{nH(p)}$.

Furthermore, the entropy H has the following properties:

- a. $0 \leq H(p) \leq 1$, if $1 \leq p \leq 1$;
- b. $H(p) = 1$ only if $p = \frac{1}{2}$.

Thus, from property (a):

$$2^{nH(p)} < 2^n \quad \text{if} \quad p \neq \frac{1}{2} \quad (5.28)$$

that is, there is no need for a codeword for every n -letter sequence, but only for the typical ones. In another way, we can compress the information in a shorter string.

This can be generalized:

$$\frac{n!}{\prod_x (np(x))!} \simeq 2^{nH(X)} \quad (5.29)$$

$$H(X) = \sum_x p(x) (-\log p(x)) \quad (5.30)$$

where $X : A \rightarrow \mathfrak{R}$ is a random variable with probability distribution $p(x)$.

More recently, Tsallis has proposed the following generalization for Shannon's entropy [31] :

$$S_q = k \frac{1 - \sum_{i=1}^W p_i^q}{q-1}, \quad (5.31)$$

where k is a normalization constant. By L'Hopital's rule, it can be shown that:

$$-\log p_i = \lim_{q \rightarrow 1} \left(\frac{1 - p_i^{q-1}}{q-1} \right) \quad (5.32)$$

$$\lim_{q \rightarrow 1} S_q = S_1 = H(X), \quad (5.33)$$

where X is the random variable such that $p(X = x_i) = p_i$.

Tsallis entropy is not additive because, given two independent random variables X and Y ; that means, the probability $p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$, it can be proved that [31]:

$$S_q(X + Y) = S_q(X) + S_q(Y) + (1 - q) \cdot S_q(X) \cdot S_q(Y), \quad (5.34)$$

where the symbol $X + Y$ means joint distribution.

The result (5.29) points out to a compression scheme. Suppose that $X_1, X_2, X_3, \dots, X_n$ is a *independent and identically distributed* classical information source over some finite alphabet; that is, the expectations and variances are such that $E(X_1) = E(X_2) = \dots = E(X_n) \equiv E(X)$ and $D(X_1) = D(X_2) = \dots = D(X_n) \equiv D(X)$, where X represents any of the random variables, and expression (5.35) holds.

A *Compression Scheme of Rate R* , denoted by $C^n(x)$, maps possible sequences $x = (x_1, x_2, \dots, x_n)$ to a bit string of length nR . The matching *decompression scheme* D^n takes the nR compressed bits and maps them back to a string of n letters. This operation is denoted by $D^n(C^n(x))$. A compression-decompression scheme is said to be reliable if the probability that $D^n(C^n(x)) = x$ approaches to one as $n \rightarrow \infty$. A fundamental result in this theory is the *Shannon's Noiseless Channel Coding Theorem*:

Theorem 5.6. *Suppose that $\{X_i\}$ are independent and identically distributed random variables that define an information source with entropy $H(X)$. Suppose $R > H(X)$. Then there exists a reliable compression scheme of rate R for the source. Conversely, if $R < H(X)$ then any compression scheme will not be reliable.*

In [20], pp. 537, this theorem is demonstrated following the development given next. Let a particular n -message: x_1, x_2, \dots, x_n . So, by the assumption of statistically independent random variables:

$$P(x_1, x_2, \dots, x_n) = p(x_1) \cdot p(x_2) \cdot \dots \cdot p(x_n) \quad (5.35)$$

Thus, typically, it is expect that:

$$P(x_1, x_2, \dots, x_n) \approx p^{np} (1 - p)^{(1-p)n}. \quad (5.36)$$

So:

$$\frac{-1}{n} \log P(x_1, x_2, \dots, x_n) \approx \langle -\log(p(x)) \rangle \equiv H(X), \quad (5.37)$$

in the sense that, for any $\varepsilon > 0$ and for n large enough:

$$H(X) - \varepsilon \leq \frac{-1}{n} \log P(x_1, x_2, \dots, x_n) \leq H(X) + \varepsilon \quad (5.38)$$

Thus:

$$2^{-n(H(X) + \varepsilon)} \geq P(x_1, x_2, \dots, x_n) \geq 2^{-n(H(X) - \varepsilon)} \quad (5.39)$$

A useful equivalent reformulation of this expression is:

$$\left| \frac{-1}{n} \log P(x_1, x_2, \dots, x_n) - H(X) \right| \leq \varepsilon. \quad (5.40)$$

A sequence that satisfies this property is called ε -*typical*. The following lemma states an important property about ε -*typical* sequences [20].

Lemma 5.1. *Fix $\varepsilon > 0$. Then, for any $\delta > 0$, for sufficiently large n , the probability that a sequence is ε -typical is at least $1 - \delta$.*

Proof. See [20], pages 537 – 539.

5.5 Multilevel Neural Network

Artificial Neural Networks (NN) are dynamic systems of information processing inspired by the human cognitive system [3]. It is widely employed in several fields, such as robotics, signal processing and artificial intelligence [8]. The NN has attracted high interest due to its robustness, strong potential for parallelism and the capability of generalization and of learning. A NN consists of processing elements, called neurons, and of weighted connections between those neurons. In fact, the neurons in a NN are logical models that mimics some aspects of their biological counterpart [3].

Such *logical neuron* was firstly developed by W. S. McCulloch and W.A. Pitts in 1943 [3]. It describes the fundamental functions and structures of a neural cell reporting that a neuron will fire an impulse only if a threshold value is exceeded.

Figure 5.2.(a) shows the basic elements of McCulloch-Pitts model: x is the input vector, w are weights input associated, y is output, R is number of elements in input and f is the *activation function* that determines the value in output. A simple choice for f is the signal function $sgn(\cdot)$. In this case, if the sum, across all the inputs with its respective weights exceeds the threshold b the output is 1 else the value of y is -1 , that is:

$$y = sgn\left(\sum_{i=1}^R w_i x_i - b\right). \quad (5.41)$$

But the McCulloch-Pitts neuron did not have a mechanisms for *learning*. Based on biological evidences, D.O. Hebb suggested a rule to adapt the weights input, which is interpreted as a learning rule [3]. This biological inspired procedure can be expressed in the following manner:

$$w_i^{new} = w_i^{old} + \Delta w_i; \quad \Delta w_i = \eta (y^{desired} - y) x_i, \quad (5.42)$$

where w^{new} and w^{old} are adapted weights and initials weights respectively, η is a real parameter to control the rate of learning and $y^{desired}$ is the desired (know) output. This *learning rule* plus the elements of Figure 5.2 is called the perceptron model for a neuron.

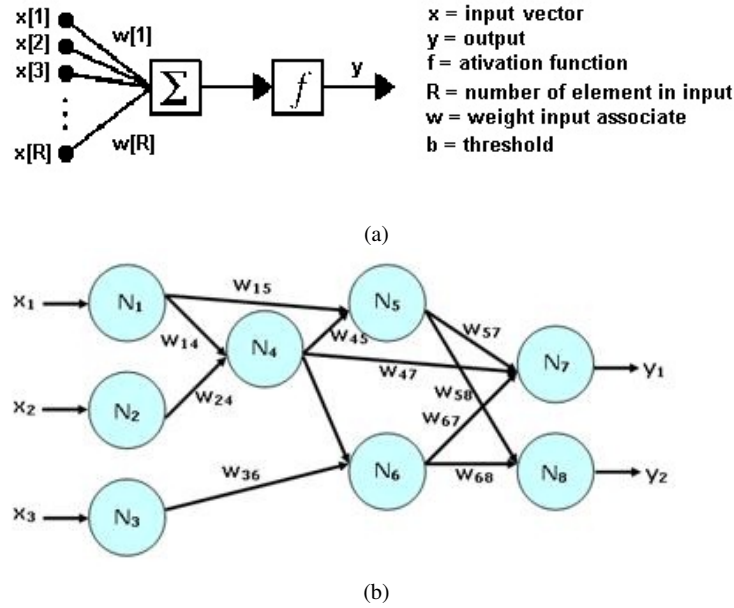


Fig. 5.2 (a) McCulloch-Pitts neuron model. (b) Artificial neural network scheme (source: www.emilstefanov.net/).

Then, the learning typically occurs through training, or exposure to a know set of input/output data. The training algorithm iteratively adjusts the connection weights $\{w_i\}$ analogous to synapses in biological nervous. These connection weights store the knowledge necessary to solve specific problems.

In a NN, the neurons are disposed in layers so that neurons of one layer are connected with neurons of an adjacent layer, as shown in Figure 5.2.(b). This figure shows the input layer, composed in this case by neurons N_1, N_2 and N_3 , the hidden layer (neurons N_4, N_5, N_6) and the output layer composed by the neurons N_7, N_8 . The input layer receives the values from the application; in this case, x_1, x_2, x_3 and the output layer sends the results (y_1, y_2) to the application. The hidden layers (one or more) are those in between input and output ones.

In the case of multilayer perceptron neural networks, each neuron is a McCulloch-Pitts one (Figure 5.2.(a)) with a learning rule to update the weights $w_{i,j}$ in order to reduce the error between the obtained output and the target one. Let us present its mathematical formulation following the reference [3]. So, let E_p the error function for pattern p , t_{pj} represents the target output for pattern p on node j , whilst o_{pj} represents the actual output at that node and w_{ij} is the weight from node i to node j .

The error function is defined as:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2. \quad (5.43)$$

The activation of each unit j , for pattern p , can be written simply as the weighted sum, as in the perceptron:

$$net_{pj} = \sum_i w_{ij} o_{pi}. \quad (5.44)$$

The output from each unit j is the threshold function f_j acting on the weighted sum. In the perceptron it was the step function. In the multilayer perceptron, it usually is a sigmoid function:

$$f(x) = \frac{1}{1 + \exp(-k \cdot net)}, \quad (5.45)$$

but any continuously differentiable monotonic function can be used in order to define the output:

$$o_{pj} = f_j(net_{pj}). \quad (5.46)$$

By the chain rule we can write:

$$\frac{\partial E_p}{\partial w_{ij}} = \frac{\partial E_p}{\partial net_{pj}} \frac{\partial net_{pj}}{\partial w_{ij}} = \frac{\partial E_p}{\partial net_{pj}} o_{pi}, \quad (5.47)$$

due to the definition of net_{pj} in expression (5.44).

Now, defining:

$$-\frac{\partial E_p}{\partial net_{pj}} = \delta_{pj}, \quad (5.48)$$

the expression (5.47) can be rewritten as:

$$-\frac{\partial E_p}{\partial w_{ij}} = \delta_{pj} o_{pi}. \quad (5.49)$$

So, to decrease E_p the weight must change proportional to $\delta_{pj} o_{pi}$:

$$\Delta_p w_{ij} = \eta \delta_{pj} o_{pi}. \quad (5.50)$$

Now, if it is known what δ_{pj} is for each unit then the error E_p can be decreased. So, using expression (5.48) and the chain rule again:

$$\delta_{pj} = -\frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial net_{pj}}. \quad (5.51)$$

Using expressions (5.43) and (5.46) we can write:

$$\delta_{pj} = f'_j(net_{pj})(t_{pj} - o_{pj}). \quad (5.52)$$

This expression is useful for the output units, since the target and output are both available. For hidden units we must observe that, if unit j is not an output unit, we can write:

$$\frac{\partial E_p}{\partial o_{pj}} = \sum_k \frac{\partial E_p}{\partial net_{pk}} \frac{\partial net_{pk}}{\partial o_{pj}}.$$

By inserting the expressions (5.44)-(5.48) in this equations it directly follows that:

$$\frac{\partial E_p}{\partial o_{pj}} = - \sum_k \delta_{pk} w_{jk}. \quad (5.53)$$

Now, by substituting (5.53) in (5.51) we finally get :

$$\delta_{pj} = f'_j(net_{pj}) \sum_k \delta_{pk} w_{jk}. \quad (5.54)$$

This equation provides a method for changing the error function so as to be sure of reducing it. The error has to be calculated in the output units first, through expression (5.52) and then passed back through the net to the earlier units to allow them to alter their connection weights. This process of passing back of this error value that leads to the network been refered to as back-propagation network.

Algorithm 2 Multilayer Perceptron Algorithm (Beale and Jackson [3])

Initialize weights and thresholds: Set all weights and thresholds to small random values.

Present input $X_p = (x_0, x_1, \dots, x_{n-1})$ and desired output: $T_p = (t_0, t_1, \dots, t_{m-1})$ where n is the number of input nodes and m is the number of output nodes.

Compute actual output: Each layer calculates:

$$y_{pj} = f \left(\sum_{i=0}^{n-1} w_{ij} x_i \right). \quad (5.55)$$

and passes that as input to the next layer. The final layer outputs values o_{pj}

Update weights: Stating from the output layer work backwards:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj}, \quad (5.56)$$

where η is a gain term. For output units:

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj}) \quad (5.57)$$

For hidden units:

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) \sum_k \delta_{pk} w_{jk}, \quad (5.58)$$

where the sum is over the k nodes in the layer next node j .

In this work the term Multilevel Neural Network (MNN) is used to describe a system composed of two or more sets of individual Multilayer Perceptron Neural Networks in two or more levels. The individual neural nets are independents in the same level. The outputs in the level i are the inputs of the level $i+1$.

5.6 Applications in Breast Cancer Imaging

Aside from non-melanoma skin cancer, breast cancer is the most common cancer among women in most countries, including Brazil. It is also one of the leading causes of cancer death among women of all races. The mammography is the most common exam for breast cancer diagnosis. However, mammograms do not detect all breast cancers.

Exams such as ultrasound, positron emission mammography (PEM), and magnetic resonance image (MRI) are adjuncts to mammography. In mammograms, the shape of breast lesion is represented in the two dimensional space while in MRI and PEM, the lesion shape is represented in the 3D space. In both cases, the characteristic of the lesion shape represents an important play for the breast cancer diagnosis. The section 5.6.1 shows an application of Tsallis entropy for breast cancer segmentation while section 5.6.2 applies discrete wavelet, Hilbert curve and multilevel neural network for the lesion shape description.

5.6.1 Non-Extensive Segmentation of Breast Lesions

Among the segmentation methods, the thresholding techniques are simple but effective tools to separate objects from their backgrounds [26]. An approach to implement a thresholding technique is to search for a threshold value constrained to the maximization of some information measure or entropy [1].

So, let us suppose an image with k gray-levels with probability distribution $P = \{p_1, p_2, \dots, p_k\}$. Given a luminance level t , the image histogram can be partitioned into two regions (background and foreground) which distinguishes two random variables, say A and B , with probability distributions:

$$P_A = \left\{ \frac{p_1}{p_A}, \frac{p_2}{p_A}, \dots, \frac{p_t}{p_A} \right\}, \quad P_B = \left\{ \frac{p_{t+1}}{p_B}, \frac{p_{t+2}}{p_B}, \dots, \frac{p_k}{p_B} \right\} \quad (5.59)$$

where $p_A = \sum_{i=1}^t p_i$ and $p_B = \sum_{i=t+1}^k p_i$.

Now, following the Equation (5.31), the Tsallis entropy can be calculated for each distribution:

$$S_A(t, q) = \frac{1 - \sum_{i=1}^t \left(\frac{p_i}{p_A} \right)^q}{q - 1}, \quad (5.60)$$

$$S_B(t, q) = \frac{1 - \sum_{i=t+1}^k \left(\frac{p_i}{p_B} \right)^q}{q - 1}. \quad (5.61)$$

Observe that the Tsallis entropy computed by Equations (5.60) and (5.61), directly depends on the threshold t . The Tsallis entropy for the random variable $A + B$ may be computed through expression (5.34):

$$S_{A+B}(t, q) = S_A(t, q) + S_B(t, q) + (1 - q)S_A(t, q)S_B(t, q), \quad (5.62)$$

where $S_A(t, q)$, $S_B(t, q)$ are given by expressions (5.60) and (5.61), respectively.

In [1] the segmentation task is accomplished by seeking for the luminance level $t_{opt}(q)$ that maximizes the non-extensive information measure $S_{A+B}(t, q)$:

$$t_{opt}(q) = \underset{t}{\operatorname{argmax}}[S_{A+B}(t, q)]. \quad (5.63)$$

The result depends on the choice for the entropic parameter q . In [24] it is described a systematic procedure to compute the optimum value of q for a specific database. Following the above definitions and formulations, each distribution P_A and P_B can be taken and subdivided into two news distribution, $P_{A_1}, P_{A_2}, P_{B_1}$ and P_{B_2} as follows:

$$P_{A_1} = \left\{ \frac{p_1}{p_{A_1}}, \frac{p_2}{p_{A_1}}, \dots, \frac{p_{t_A}}{p_{A_1}} \right\}, \quad P_{A_2} = \left\{ \frac{p_{t_A+1}}{p_{A_2}}, \frac{p_{t_A+2}}{p_{A_2}}, \dots, \frac{p_t}{p_{A_2}} \right\}, \quad (5.64)$$

$$P_{B_1} = \left\{ \frac{p_{t+1}}{p_{B_1}}, \frac{p_{t+2}}{p_{B_1}}, \dots, \frac{p_{t_B}}{p_{B_1}} \right\}, \quad P_{B_2} = \left\{ \frac{p_{t_B+1}}{p_{B_2}}, \frac{p_{t_B+2}}{p_{B_2}}, \dots, \frac{p_k}{p_{B_2}} \right\}, \quad (5.65)$$

where, $p_{A_1} = \sum_{i=1}^{t_A} p_i$, $p_{A_2} = \sum_{i=t_A+1}^t p_i$, $p_{B_1} = \sum_{i=t+1}^{t_B} p_i$, $p_{B_2} = \sum_{i=t_B+1}^k p_i$. For each one of these four distributions its respective non-extensive entropy can be compute as, $S_{A_1}(t_A, q)$, $S_{A_2}(t_A, q)$, $S_{B_1}(t_B, q)$, $S_{B_2}(t_B, q)$, likewise in expressions (5.60)-(5.61). Besides, the entropy of the joint distributions $(A_1 + A_2)$ and $(B_1 + B_2)$ can be estimated by:

$$\begin{aligned} S_{B_1+B_2}^{A_1+A_2}(t_A, t_B, q) &= (S_{A_1} + S_{A_2} + (1 - q) \cdot S_{A_1} \cdot S_{A_2}) + \\ &\quad (S_{B_1} + S_{B_2} + (1 - q) \cdot S_{B_1} \cdot S_{B_2}) + \\ &\quad (1 - q) \cdot (S_{A_1} + S_{A_2} + (1 - q) \cdot S_{A_1} \cdot S_{A_2}) \cdot \\ &\quad (S_{B_1} + S_{B_2} + (1 - q) \cdot S_{B_1} \cdot S_{B_2}). \end{aligned} \quad (5.66)$$

In this case, two optimal luminance levels must be found: t_A and t_B (remember that $t = t_{opt}(q)$ was already computed in equation (5.63)). These values are computed by solving the optimization problem:

$$(t_A^{opt}, t_B^{opt})(q) = \underset{(t_A, t_B)}{\operatorname{argmax}}[S_{B_1+B_2}^{A_1+A_2}(t_A, t_B, q)]. \quad (5.67)$$

Expressions (5.63)-(5.67) compose the first and second levels of a recursive scheme, named Non-Extensive Segmentation Recursive Algorithm - NESRA [25], that searches for an optimum threshold for each sub-region found in the present level. However, the experimental results show that it is not necessary more than two or three recursions in order to obtain results which are equals or better than the traditional methods. Second, the growing of the number of recursion does not enlarge the algorithm complexity or computation, since this growing accompanies a drop-

ping of states to be computed at each recursion, like in the algorithms for binary partitions, which have an $O(n \log_2 n)$ complexity. At each iteration the NESRA generates 2^{r+1} regions, where r is the number of recursions. Note that for $r = 0$, the NESRA is the simple image binarization algorithm proposed in [1]. The recursive algorithm for the previous presentation is the following.

Algorithm 3 Non-Extensive Segmentation Recursive Algorithm

NESRA (H, i, k)

Input: H = image histogram, i = first histogram bin, k = last histogram bin

if Histogram H is homogeneous **then**

return

end if

for all $t = i$ until k **do**

 compute normalization for background

 compute normalization for foreground

 compute q-entropy for background according to Equation (5.60)

 compute q-entropy for foreground according to Equation (5.61)

 compute composed q-entropy according to Equations (5.62) and (5.63)

end for

$topt = \text{argmax of the composed q-entropy}$

 Call NESRA($H, i, topt$)

 Call NESRA($H, topt + 1, k$)

In order to test the NESRA method a data base composed by 250 breast ultrasound images (100 benign and 150 malignant) was used. The Figure 5.3 shows an original benign image example (left) and the NESRA result (right).

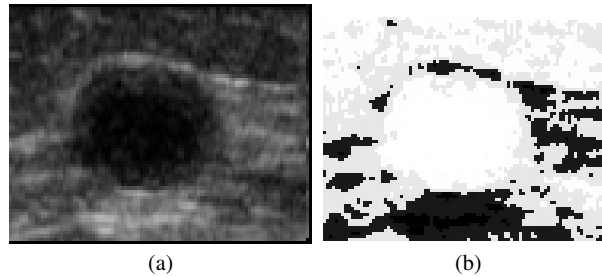


Fig. 5.3 Original ultrasound benign image (left) and the NESRA results (right) with two recursions and $q = 0.5$.

This result was obtained with one NESRA recursion only, which means that the gray scale distribution was partitioned into two new distributions and each one was partitioned into two others, generating four distinct regions. Although we wish to split the image into two regions only the recursion in this case helps to delimit a tumor's nucleus and also a boundary as a narrow region around the nucleus. Note in the Figure 5.3 the white region around the image center (tumor) and the interme-

diate gray-level region around the tumor's region (the transition region between the tumor and background). This transition region can be isolated and help to delimit accurately the lesion boundary. For the result of Figure 5.3-b, $q = 0.5$ was used. In this case, the ROI is the tumor's nucleus and its transition region.

The region obtained by the NESRA algorithm is a coarse representation of the lesion region which may be improved by traditional morphological operations, as performed in [25] through the following rule. Considering the binary image generated by NESRA (e.g Figure 5.3-b), let α and β be the total ROI's area and the total image area, respectively. If $\alpha \geq \xi\beta$ an erosion is carried out; and if $\alpha \leq \delta\beta$ a dilation is carried out. After, assuming that the ROI has a geometric center near to the image center, a region growing algorithm which defines the final ROI's boundary is applied (see [25] for details about these operations). In the experiments described bellow, the values $\xi = 0.75$ and $\delta = 0.25$ were fixed in order to correctly extract most of all ROIs. The result of this morphological rule applied in the image of Figure 5.3-b is shown in Figure 5.4.

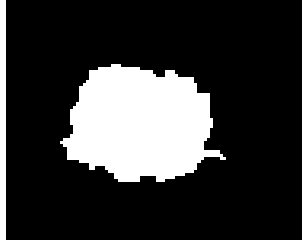


Fig. 5.4 ROI after morphological chain application.

5.6.2 High level 2D and 3D Shape Descriptors

With the aim of characterizing the 2D and 3D shape of breast lesions as malignant tumor or benign mass, it was proposed in [10, 5] two new shape descriptors based on the Hilbert space-filling curve named $2HqRF$ and $3HqRF$. The latter approach is a direct extension of the first one. In order to obtain the features, the 2D contour or 3D surface of the lesion must be mapped to a 2D or 3D silhouette, respectively. The Hilbert curve (see section 5.3) is scanned through the silhouette and an one-dimensional vector is obtained. Because this vector possesses a large number of elements, it is sub-sampled and the value of the resulting elements is normalized to the interval $[0,1]$. To obtain a high level description of the shape, the feature is finally obtained by a multi-level neural network.

The evaluation of the results was carried out using a breast cancer database from University of Calgary, Canada and a synthetic 3D database with 3D-shapes that simulate 3D breast lesions.

Let us present the $2HqRF$ approach which was inspired in the work proposed by Ebrahim et al. [7] in the sense that both the methods derive the shape feature by scanning the Hilbert curve through the silhouette of the shape. The $2HqRF$ method differs from the method proposed by Ebrahim in the sense that it can be processed in parallel and yields a descriptor that preserves relevant information from the original shape thanks to the use of a neural network. A general flowchart to obtain the $2HqRF$ method is illustrated in Figure 5.5. The procedure to obtain the $3HqRF$ is a direct extension to the 3D voxel space. The main steps are described following:

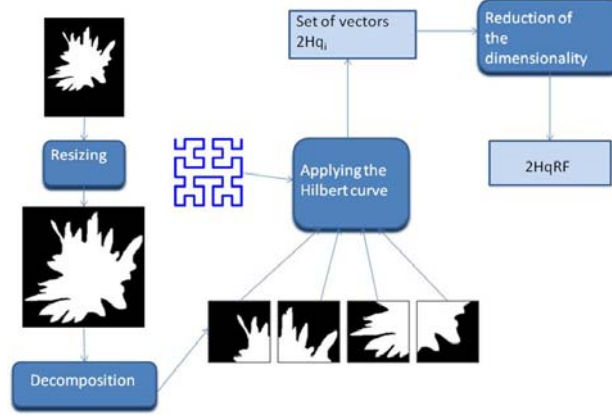


Fig. 5.5 A flowchart to derive the $2HqRF$.

- Resizing module:** in this module, the 2D-shape is resized to $2^N \times 2^N$ dimension, where N is a positive integer. The resizing procedure consists of three steps: i) the 2D shape of an arbitrary size is cropped into the smaller convex hull that contains the region of interest (ROI). ii) the region, now with $P \times Q$ size, has its background extended to a $S \times S$ size, where $S = \text{MAX}\{P, Q\}$; iii) the result is re-scaled to a region with $2^N \times 2^N$ size. This region we call $2D\text{-silhouette}$. The resizing module guaranties that all pixels of the silhouette are visited by the Hilbert curve scan.
- Decomposition module:** in this step, the 2D-silhouette is equally divided along each dimension resulting in four sub-regions $r_i, i = 1, 2, 3, 4$, with the same size. Each sub-region r_i may be successively divided into another four sub-regions. The number L of successive divisions is referred as levels of decomposition.
- Applying the Hilbert curve module:** the 2D Hilbert curve, a fractal structure presented in section 5.3, is applied to each sub-region r_i obtained in the previous step, resulting in a 1D representation vector for each one, termed $2Hq_i$. Note that the number of elements in the $2Hq_i$ is equal to the number of the pixels in the correspondent sub-region r_i , what is too large to be used in a image retrieval task.

- d. **Reduction of the dimensionality:** In order to reduce the amount of data while preserving relevant information, we reduce the dimensionality of each $2Hq_i$ using the coefficients of approximation of the discrete wavelets transform (Algorithm 1) to a predefined size q . Then, the domain of the data are normalized into the interval $[0, 1]$. At this point each $2Hq_i$ possesses q elements, then the 2D-shape is represented by $q \times 4^L$ elements, what still is a large amount of data for image retrieval process. An important contribution of this work is to use an artificial neural network to further reduce the dimensionality of the 2D-shape representation yielding the high level representation $2HqRF$. The multilevel neural network (MNN), described on section 5.5, was designed using a multi-layer perceptron network with back propagation with 4^L inputs, related to each $2Hq_i$, and c outputs, where c is the number of categories of a training dataset. The architecture for the MNN is presented in Figure 5.6.

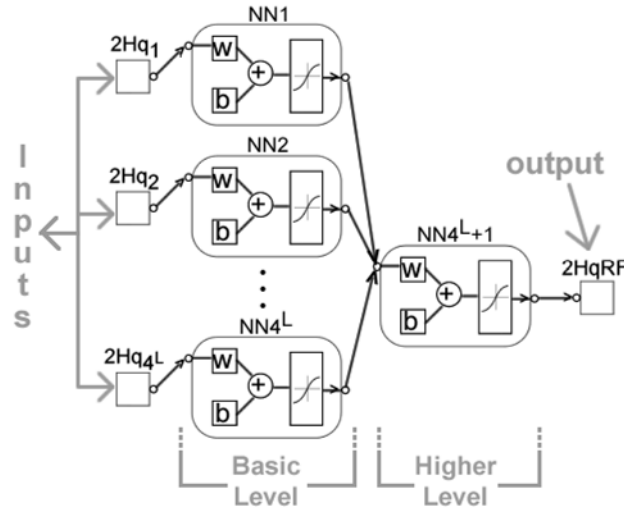


Fig. 5.6 The architecture of the MNN for the high level dimensionality reduction.

The Dataset of Breast Lesions: The images that comprise the database used in this study are from two different sources. The first set comprises twenty mammograms obtained from Screen Test: the Alberta Program for the Early Detection of Breast Cancer. This mammograms were digitized using a Lumiscan 85 scanner with resolution of $50\mu m$ at 12 bits per pixel. The set has 57 regions of interest (ROIs), of which 37 correspond to benign masses and 20 correspond to malignant tumors. The second set of images was obtained from the Mammographic Image Analysis Society (MIAS, UK) database and the teaching library of the Foothills Hospital (Calgary). The MIAS images were digitized at a resolution of $50\mu m$. The Foothills Hospital images were digitized at $62\mu m$. This set includes 28 benign masses and 26 malignant tumors. The contour of each mass was manually drawn by a radiologist

specialized in mammography. The combined dataset has 111 contours, including both typical and atypical shapes of benign masses (65) and malignant tumors (46). The diagnostic classification was based upon biopsy. Figure 5.7 shows a set of 2D-silhouette from the database named BC-111.

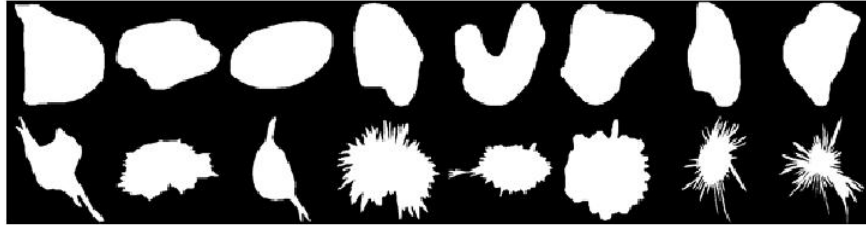


Fig. 5.7 A set of benign masses and malignant tumors from the database BC111 used in the experiments.wim2011

In order to evaluate the proposed $3HqRF$ shape descriptor, we created the synthetic 3D breast lesion, named W3BC111, derived from the contours in the dataset BC-111. In this case, the 3D lesions are classified as round and spiculate. An example is shown in Figure 5.8.

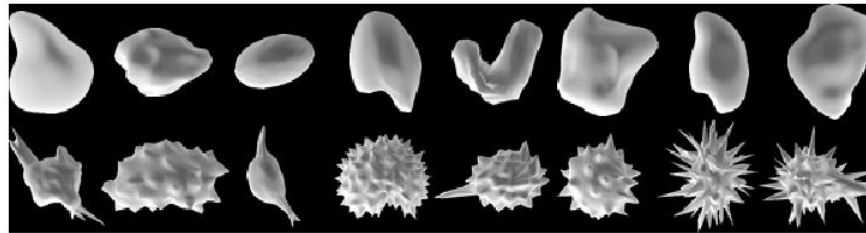


Fig. 5.8 A set of round and spiculate synthetic shapes derived from the database BC111.

Table 5.1 Parameters used in the tests.

Parameters	$2HqRF$	$3HqRF$
Decomposition Level (L)	3	1
Wavelet Family	Haar	Haar
Value of q	256	256
Size of the shape factor	2	2

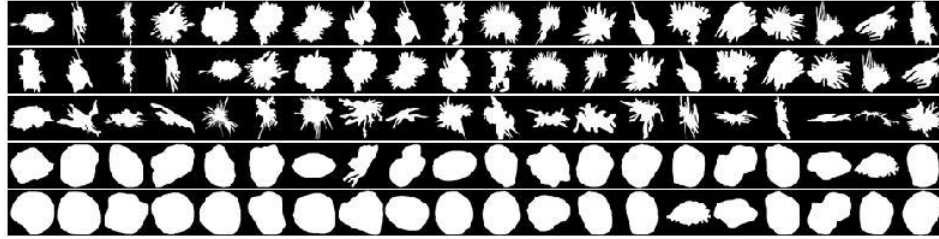
Parameters used in the experiments: The shape descriptors based on the Hilbert curve were derived for the dataset BC-111 and W3BC111 according to the parameters shown in Table 5.1. The Table 5.2 presents the parameters used to con-

Table 5.2 Parameters for the multilevel neural network.

Parameters	
No. of neurons for each layer	20
Transfer function	Hyperbolic tangent sigmoid
Back-propagation function	Levenberg-Marguardt
Validation strategy	10-fold cross validation

figure the multilevel Neural Network to derive both $2HqRF$ and $3HqRF$ shape descriptors.

Discussion of the Results: In order to evaluate the quality of the results obtained with the $2HqRF$ shape descriptors, we compared its average of the normalized discount cumulative gain (NDCG) with those obtained from our previous works using the dataset BC-111: spiculate index SI_{TF} [12], convex index CX_{TA} and Fractal index FD_{TA} [11]. The experiments were run using the 10-fold cross validation. The similarity between two shape descriptor is determined by the Euclidean distance. The analysis of the results are based on the diagnosis of the lesion. Figure 5.9 illustrates the top-10 images resulting from five different similarity searchings. Table 5.3 presents the NDCG values for all the experiments. It is worth noting that the $2HqRF$ is obtained directly from the original 2D-silhouette, different from the competing works, which need to derive the indexes from the turning angle signature function. In these experiments, the $2HqRF$ presented the best results in all the cases.

**Fig. 5.9** The top-10 images resulting of the similarity searching. The first object in each line is the query object and the remainder objects in the same line are the target.**Table 5.3** Comparison of the values of NDCG for different shape descriptors.

shape descriptor	NDCG
$2HqRF$	1.0
SI_{TF}	0.88
CX_{TA}	0.96
FD_{TA}	0.95

To evaluate the efficacy of the $3HqRF$, we run the experiments using the dataset W3BC111. Figure 5.10 presents the top-20 images resulting from five different similarity searchings. Because the W3BC111 is a synthetic dataset, the shapes was classified as round or spiculated. The idea behind that is show that the $3HqRF$ is able to discriminate between shapes, what is a relevant information for breast cancer classification. The proposed method reached a NDCG equal to 0.99.

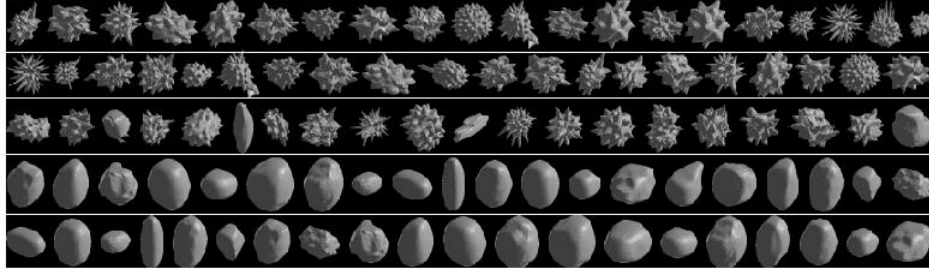


Fig. 5.10 The top-20 images resulting of the similarity searching. The first object in each line is the query object and the remainder objects in the same line are the target.

5.7 Perspectives in Wavelets Applications and Information Theory

In the last decades wavelet methods have been applied for solving partial differential equations (PDEs) and numerous papers about this topics have been published [17]. Among all the wavelet families (Haar, Daubechies, Gaussian, Mexican Hat, Morlet, etc.) the Haar wavelets are of special interest for PDE solutions because they are simply piecewise constant functions that can be analytically integrated in arbitrary times. However, Haar wavelets are not continuous functions and their derivatives do not exist in the breaking points. So, it is not possible to directly apply these wavelets for solving PDEs. Section 5.7.1 presents a simple and elegant way to address this issue without losing the simplicity of Haar wavelet framework.

For system that can not be described through the classical information framework, is that a bound for message compression, like in Shannon's theory? Among the possibilities to address this problem [16], in Section 5.7.2 we consider this question in the context of ε – *typical* sequences (see Section 5.4).

5.7.1 Haar Wavelets for Solving PDEs

Consider the linear PDE:

$$\sum_{\gamma=0}^{\Gamma} \sum_{\lambda=0}^{\Lambda} D_{\gamma\lambda} \frac{\partial^{(\gamma+\lambda)} u}{\partial x^{\gamma} \partial y^{\lambda}} = f(x, y), \quad (5.68)$$

where Γ, Λ are natural numbers, $D_{\gamma\lambda}$ are real constants and $f : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ is a known function. For simplicity, the domain is a rectangular region $\Sigma = \{(x, y); \ x \in [0, L_1] \text{ and } y \in [0, L_2]\}$. Besides, the solution $u(x, y)$ must be constrained to proper boundary conditions.

The method described in [17] uses Haar wavelets (see section 5.2.1), properly scaled and dilated, and divides the intervals $[0, L_1]$, $[0, L_2]$ into $2M_1$ and $2M_2$ parts of equal length, respectively. Based on that, it is possible to seek for a solution $u(x, y)$ that satisfies:

$$\frac{\partial^{(\Gamma+\Lambda)} u}{\partial x^{\Gamma} \partial y^{\Lambda}} = \sum_{i=1}^{2M_1} \sum_{l=1}^{2M_2} a_{il} h_i(x) h_l(y), \quad (5.69)$$

where a_{il} are the wavelets coefficients and $h_i(x)$, $h_l(y)$ are Haar functions. By multiply integrating the expression (5.69) lower order derivatives of u can be calculated. This fact, together with the boundary conditions and a set of $2M_1 \cdot 2M_2$ collocation points $(x_r, y_s) \in \Sigma$ allows to calculate a system with the form:

$$\sum_{i=1}^{2M_1} \sum_{l=1}^{2M_2} R_{ilrs} = f(x_r, y_s), \quad (5.70)$$

where terms R_{ilrs} linearly depends on a_{il} . So, by solving the system (5.70) the wavelets coefficients can be found and then a solution can be obtained by multiply integrating expression (5.69) or some lower order derivative. An example can be helpful to understand these ideas.

The diffusion equation should be solved:

$$\frac{\partial u}{\partial t} = A \frac{\partial^2 u}{\partial x^2}, \quad (x, t) \in [0, 1], \quad (5.71)$$

subject to:

$$u(x, 0) = g(x), \quad u(0, t) = u(1, t) = 0. \quad (5.72)$$

For simplicity, consider $M_1 = M_2 = M$. In this case, expression (5.69) renders:

$$\frac{\partial^3 u}{\partial t \partial x^2} = \sum_{i=1}^{2M} \sum_{l=1}^{2M} a_{il} h_i(x) h_l(t). \quad (5.73)$$

By integrating expression (5.73) respect to t and multiply integrating respect to x and considering the initial condition in expression (5.72) we obtain:

$$\frac{\partial^2 u}{\partial x^2} = \sum_{i=1}^{2M} \sum_{l=1}^{2M} a_{il} h_i(x) p_{1l}(t) + \frac{\partial^2 g}{\partial x^2}, \quad (5.74)$$

$$\frac{\partial u}{\partial t} = \sum_{i=1}^{2M} \sum_{l=1}^{2M} a_{il} p_{2i}(x) h_l(t) + x \frac{\partial^2 u}{\partial t \partial x} \Big|_{x=0} + \frac{\partial u}{\partial t} \Big|_{x=0}. \quad (5.75)$$

The functions $p_{1l}(t)$ and $p_{2i}(x)$ are analytically calculated in [17] using the Haar wavelets expressions. Now, the boundary conditions can be inserted in expression (5.75) and substitute both equation (5.74)-(5.75) in expression (5.71). Finally, the obtained expression can be evaluated in the collocation points (x_r, y_s) to obtain a system in the form (5.70).

Partial differential equations have been applied in image processing for inpainting, noise reduction and image modeling. For instance, the 2D diffusion problem:

$$\frac{\partial I}{\partial t} = A \left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \right), \quad (x, y) \in \Sigma \quad \text{and} \quad t \in \mathfrak{R}^+, \quad (5.76)$$

subject to:

$$I(x, y, 0) = I_0(x, y), \quad (5.77)$$

where $I_0(x, y)$ is an image intensity field, defines linear diffusion which solution may be a family $I(x, y, t)$ of functions that are smoothed versions of the $I_0(x, y)$. Although this is well known result in linear filtering, the above technique for solving PDE allows to explore the multiscale analysis of wavelet spaces. This characteristic shall be considered in further works.

5.7.2 Data Compression and Tsallis Entropy

An important result in the nonextensive information theory is the generalization of *The Noiseless Channel Coding Theorem* for the nonextensive entropic form of Tsallis, given by expression (5.31).

So, the following interesting property can be demonstrated, similar to Lemma 5.1, but now considering the entropy given by expression (5.31):

Lemma 5.2. Fix $\varepsilon > 0$. Then, for any $\delta > 0$, for sufficiently large n , it can be shown that:

$$P \left(\left| \frac{1}{n} \sum_{i=1}^n \left(\frac{1 - p_i^{q-1}}{q-1} \right) - S_q \right| \leq \varepsilon \right) \geq 1 - \delta.$$

Proof. By using equation (5.32) the left-hand side of expression (5.40) can be rewritten as:

$$\begin{aligned}
& \left| -\frac{1}{n} \log P(x_1, x_2, \dots, x_n) - H(X) \right| = \\
& \left| -\frac{1}{n} \log P(x_1, x_2, \dots, x_n) - S_1 \right| = \\
& \lim_{q \rightarrow 1} \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{1 - p_i^{q-1}}{q-1} \right) - \left(\frac{1 - \sum_{i=1}^W p_i^q}{q-1} \right) \right|
\end{aligned} \tag{5.78}$$

where $k = 1$.

Let us define the following random variables:

$$\xi_i = \frac{1 - p_i^{q-1}}{q-1}, \quad i = 1, \dots, n. \tag{5.79}$$

Thus:

$$E(\xi_i) = \sum_{j=1}^W \left(\frac{1 - p_j^{q-1}}{q-1} \right) p_j = \tag{5.80}$$

$$\frac{\sum_{j=1}^W p_j - \sum_{j=1}^W p_j^q}{q-1} = \tag{5.81}$$

$$\frac{1 - \sum_{j=1}^W p_j^q}{q-1} = S_q. \tag{5.82}$$

Thus, the random variables $\xi_i, i = 1, \dots, W$, are such that $E(\xi_i) = S_q, i = 1, \dots, W$. Now, let us define:

$$y_i = \frac{\xi_i - S_q}{n}, \quad i = 1, \dots, n \tag{5.83}$$

$$\xi_S = \sum_{i=1}^n y_i, \tag{5.84}$$

It can be observed that:

$$\xi_S = \sum_{i=1}^n \left(\frac{\xi_i}{n} \right) - S_q, \tag{5.85}$$

From expression (5.83), it can be stated:

$$E(y_i) = 0, \tag{5.86}$$

$$D(y_i) = D\left(\frac{\xi_i - S_q}{n}\right) = \frac{1}{n^2} D(\xi_i - S_q) = \frac{1}{n^2} D(\xi_i).$$

But, $D(\xi_1) = D(\xi_2) = \dots = D(\xi_n)$. Henceforth $D(y_1) = D(y_2) = \dots = D(y_n)$. Thus, the random variables y_i are identically distributed random variables. Besides,

they are also statistically independent due to definition (5.83). Thus, we can apply the Central Limit Theorem to obtain:

$$P(\xi_s = l, n) = \frac{1}{\sqrt{2\pi \frac{D(\xi)}{n}}} \exp \left\{ \frac{-[l-0]^2}{2 \frac{D(\xi)}{n}} \right\}. \quad (5.87)$$

So, when $n \rightarrow \infty$ the random variable ξ_s tends to a gaussian distributed random variable with zero mean and variance $\frac{D(\xi)}{n} \rightarrow 0$. The Lemma 5.2 is a straightforward consequence of this result.

This fact emphasizes the belief that we could obtain a nonextensive generalization of the Noiseless Channel Coding Theorem in the context of ε – *typical* sequences and their properties (see section 5.4).

5.8 Conclusions

This chapter shows elegant frameworks in functional analysis to develop wavelet and fractal theory. The discrete wavelet transform was presented and the Hilbert curve constructed in the context of deterministic fractal theory. These elements were applied for shape description in medical imaging. Theoretical models in information theory and neural networks were discussed and applied for compression, image segmentation and shape descriptors processing.

The two-dimensional Haar wavelets approach for solving PDEs is mathematically simple and fast with promising consequences for multiscale approaches.

This chapter also reports a preliminary result in the extension of Shannons Information Theory to nonextensive approaches based on properties of ε – *typical* sequences. The Tsallis entropy was considered to accomplish this goal.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq, CAPES, FAPERJ and FAPEMIG. The support of these agencies is gratefully acknowledged.

References

- [1] M. P. Albuquerque, M. P. Albuquerque, I.A. Esquef, and A.R.G. Mello. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25:1059–1065, 2004.
- [2] M. Barnsley. *Fractals Everywhere*. Academic Press INC., 1988.
- [3] R. Beale and T. Jackson. *Neural Computing*. MIT Press, 1994.
- [4] C.K. Chui. *An Introduction to Wavelets*. New York Academic Press, 1992.

- [5] W.A.A. de Oliveira and D. Guliato. Hqrf: Novos descritores de forma para auxiliar no diagnóstico de câncer de mama. In *Congresso da Sociedade Brasileira de Computação - Proc. do Workshop de Informática Médica*, Natal, Brazil, 2011.
- [6] D. do Amaral and G. A. Giraldi. Continuous and discrete wavelet transform and multiscale analysis. Technical Report 21, Department of Computer Science, National Laboratory for Scientific Computing, 2009.
- [7] Y. Ebrahim, M. Ahmed, W. Abdelsalam, and S.C. Chau. Shape representation and description using the hilbert curve. *Pattern Recognition Letters*, 30(4):348 – 358, 2009.
- [8] L. Fausett. *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*. Prentice Hall, 1993.
- [9] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [10] D. Guliato, W.A.A. de Oliveira, and C. Traina Jr. A new feature descriptor derived from Hilbert space-filling curve to assist breast cancer classification. In *Proceedings of the 23rd IEEE International Symposium on Computer-Based Medical Systems*, volume 1, pages 303–308, 2010.
- [11] D. Guliato, R.M. Rangayyan, Carvalho, and S.A. Santiago. Polygonal modeling of contours of breast tumors with the preservation of spicules. *IEEE Trans Biomed Eng.*, 2008.
- [12] D. Guliato, R.M. Rangayyan, J.D. de Carvalho, and S.A. Santiago. Feature extraction from the turning angle function for the classification of contours of breast tumors. *J. Digital Imaging*, 2008.
- [13] The MathWorks Inc. Wavelet toolbox for matlab. <http://www.mathworks.com/products/wavelet/>, 2009.
- [14] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., 1989.
- [15] J. M. Jolion and A. Montanvert. The adaptive pyramid: A framework for 2d image analysis. *CVGIP: Image Understanding*, 55(3):339–348, May 1992.
- [16] Satish Kumar and Rajesh Kumar. Some noiseless coding theorem connected with havrda and charvat and tsallis's entropy. *Kragujevac Journal of Mathematics*, 35(1):111–117, 2011.
- [17] Ülo Lepik. Solving pdes with the aid of two-dimensional haar wavelets. *Computers & Mathematics with Applications*, 61(7):1873–1879, 2011.
- [18] A.K. Louis, P. Maab, and A. Rieder. *Wavelet Analysis: Theory and Applications*. John Wiley & Sons, 1997.
- [19] F. Natterer and F. Wubbeling. *Mathematical Methods in Image Reconstruction*. Soc. for Ind. and Appl. Math. (SIAM), 2001.
- [20] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press., December 2000.
- [21] Swiss Federal Institute of Technology Lausanne. Services for exchanging knowledge in wavelets. <http://www.wavelet.org/>, 2009.
- [22] Heinz-Otto Peitgen, H. Jurgens, and D. Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer-Verlag, New York, Inc., 1992.
- [23] H.L. Resnikoff and Jr. R.O. Wells. *Wavelet Analysis: The scalable structure of information*. Springer, 1998.
- [24] P. S. Rodrigues and G. A. Giraldi. Computing the q-index for tsallis nonextensive image segmentation. In *Proc. of the 2009 XXII Brazilian Symp. on Comp. Graph. and IP, SIBGRAPI '09*, pages 232–237, 2009.
- [25] P. S. Rodrigues, G. A. Giraldi, R.-F. Chang, and J. S. Suri. Non-extensive entropy for cad systems of breast cancer images. In *In Proc. of International Symposium on Computer Graphics, Image Processing and Vision - SIBGRAPI'06*, Manaus, Amazonas, Brazil, 2006.
- [26] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165, January 2004.
- [27] G. Strang and T. Nguyen. *wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
- [28] P. Suetens. *Fundamentals of Medical Imaging*. Cambridge University Press, New York., 2002.

- [29] J. S. Suri and D. L. Wilson S. Laxminarayan., editors. *Hand Book of Biomedical Analysis*, volume I, II and III. Kluwer Academic/Plenum, New York, Boston, Dordrecht, London, Moscow, 2005.
- [30] J.S. Suri, D. Wilson, and S. Laxminarayan, editors. *International Topics in Biomedical Engineering - Handbook of Biomedical Image Analysis: Volume 3: Registration Models*. Springer-Verlag New York, Inc, 2005.
- [31] C. Tsallis. Nonextensive statistics: Theoretical, experimental and computational evidences and connections. *Brazilian Journal of Physics*, 29(1), March 1999.
- [32] Y. Zeevi and R. Coifman. *Signal and Image Representation in Combined Spaces (Wavelet Analysis and Its Applications)*. Academic Press, 1998.

Chapter 6

Computational methods for medical image quantification in SPECT

Marco A. Gutierrez, Ramon A. Moreno and Marina S. Rebelo

Abstract Single Photon Emission Computed Tomography (SPECT) is a noninvasive technique used in the visualization and analysis of the distribution of radioactive counts, for instance within the myocardium and surrounding structures. In the last two decades a number of techniques have been developed to aid in the classification of the SPECT images. These tools were continuously expanded providing LV perfusion and function parameters in 2D, 3D and 4D. In this chapter we present a set of techniques to quantify and visualize myocardial function in SPECT studies in both spatial and frequency domains. The quantification of the myocardial motion by computer methods provides reproducible measures and parameters, thus improving the quality of the diagnostic and providing more reliability and reproducibility of the data interpretation. However, the search for robust methods to analyze automatically the complete series of images is yet an open study area.

Marco A. Gutierrez

Hear Institute University of São Paulo and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Av. Dr. Enéas de Carvalho Aguiar, 44, CEP 05403-000, São Paulo, SP, Brazil, e-mail: marco.gutierrez@incor.usp.br

Ramon A. Moreno

Hear Institute University of São Paulo and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Av. Dr. Enéas de Carvalho Aguiar, 44, CEP 05403-000, São Paulo, SP, Brazil, e-mail: ramon.moreno@incor.usp.br

Marina S. Rebelo

Hear Institute University of São Paulo and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Av. Dr. Enéas de Carvalho Aguiar, 44, CEP 05403-000, São Paulo, SP, Brazil, e-mail: marina.rebelo@incor.usp.br

6.1 Introduction

Left ventricular contractile abnormalities can be an important manifestation of coronary artery disease. Wall motion changes may represent ischemia or infarction of the myocardium [92]. Quantifying the extent of regional wall motion abnormality may aid in determining the myocardial effects of coronary artery disease. It would also simplify the analysis of wall motion changes after diagnostic and therapeutic interventions and permit comparison of different imaging techniques to assess their diagnostic accuracy. For this reason, the proper tracking of left ventricle wall can contribute to isolate the location and extent of ischemic or infarcted myocardium and constitutes a fundamental goal of medical image modalities, such as Nuclear Medicine.

Radionuclide imaging is a common noninvasive technique used in the evaluation of cardiac function and disease. The process to obtain images in this modality involves the detection of the radiation emitted from a patient's organ or region after the injection of a radiopharmaceutical such as thallium-201 (^{201}Tl) and technetium-99m sestamibi ($^{99}\text{Tc}^m\text{-MIBI}$). Using a gamma camera part of the emitted radiation can be used to produce an image indicating the distribution of the radionuclide in the myocardium [48,92].

Temporal changes in the spatial distribution of radiopharmaceuticals can be obtained by taking multiple images over periods of time that may vary from milliseconds to hundreds of seconds. Since the resulting image, known as planar image, contains information of the projection of a radioactive volume over the detector's face, it is often difficult to determine clearly the function of tissue deep in the body. Tomographic studies obtained by taking multi-view acquisitions of the object overcome most problems caused by superposition of information in a single planar view. The technique of Single-Photon Emission Computerized Tomography (SPECT) provides the clinician a set of images that permits the visualization of the distribution of radioactive counts within the myocardium and surrounding structures. Defects on the distribution of some radionuclides in the myocardium indicate a muscle hypoperfusion due to obstruction of the coronary arteries (Figure 6.1).

Due to the capability to acquire myocardial perfusion SPECT studies gated to the electrocardiogram signal (ECG) provide the potential to extract additional quantitative information about cardiac function from the time-dependent motion of the myocardium. While gated SPECT studies are usually interpreted as series of two dimensional (2D) images, there is an increasing interest in the presentation of the data in a more realistic three dimensional (3D) form. These new displays provide an overview of the many slices in a single image, and enhance the perception of spatial relationships. On the other hand, the visualization of the gated SPECT studies is complicated by the fact that 3D density images cannot be directly presented using common display devices that produce only 2D pictures. In general, some kind of transformation must be performed to the image sequence, which introduces reduction of information.

Typically, there are two approaches: 2D sequence images or 3D surface rendering. The first approach preserves most of the intensity information. However, 3D

relations between each slice are only implicit. Moreover, a typical 3D gated SPECT study produces hundreds of images, which are usually analyzed in a time-consuming procedure. The second alternative makes 3D information explicit (Figure 6.2), but density is represented indirectly through the shape of the surface.

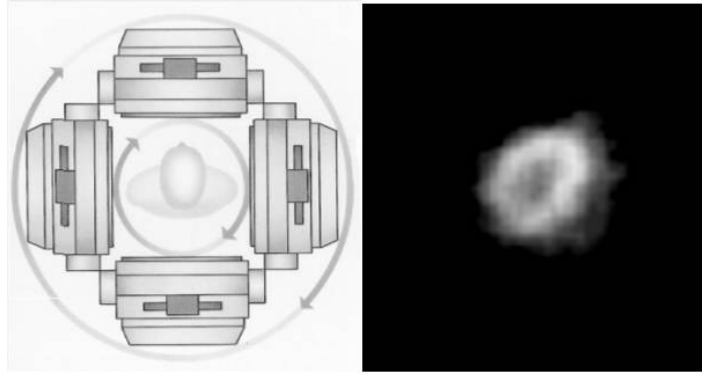


Fig. 6.1 A rotating gamma camera for emission tomography (left). By using reconstruction algorithms it is possible to estimate the Left-Ventricle (LV) transaxial image.

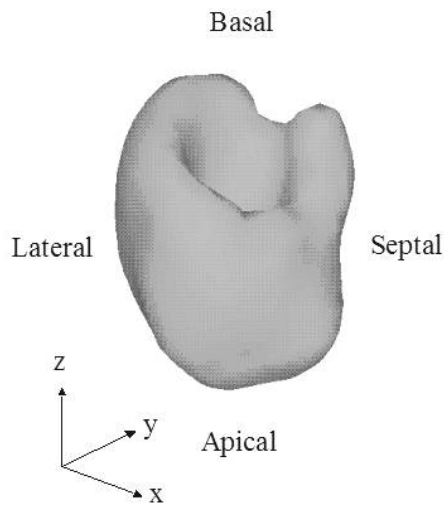


Fig. 6.2 Three dimensional rendering of a normal cardiac left ventricle.

6.2 Automatic Analysis of Cardiac Nuclear Images

A number of techniques, most of them concerned with visualization, have been developed to aid in the classification of the images. Perfusion abnormalities or defects were visually interpreted with the knowledge that normal myocardium displayed different radioactive distributions than those of abnormal myocardium [13, 60, 79, 98, 112]. However, it has been shown that interpretation of images by strictly visual techniques is subject to errors and inconsistencies due to variation in the size and orientation of the heart, the low contrast of planar imaging, photon absorption and scatter. For these reasons, assistance in diagnosis can be improved only through the development of automatic or semi-automatic methods to analyze and to quantify medical image parameters with the expectation of a significant reduction in intra-observer variability [14, 35].

6.2.1 Planar Imaging Analysis

To solve the problem of interobserver variability, several methods of quantitative analysis of the radiotracer distribution have been described [12, 48, 77, 115, 116]. Watson [12, 116] co-registered the serial images and reduced the 2D images to a series of linear radiotracer intensity profiles. Relative perfusions were then extracted from an analysis of these profiles as a function of time. Garcia [48, 77] replaced the co-registration step by using a circumferential profile that reduced the 2D image into a line profile. The construction of the circumferential profile began with the manual definition of the center of the heart in the image. From that definition, the algorithm [48, 77] sampled the myocardial tracer uptake along rays emanating from the user defined center. At 6 degree steps circling clockwise about the left ventricle, myocardial intensity profiles were recorded. From the myocardial intensity profiles along each ray, the maximum myocardial intensity was recorded as a function of angle, reducing the 2D images to a line profile.

The advances in quantification methods by imaging patients with low likelihood for coronary artery disease allowed the construction of normal databases and the definition of mean and variance limits for normal patients. With the definition of mean and variance limits for normal perfusion, objective abnormality thresholds were defined. The use of these quantification tools and normal databases improved the diagnostic sensitivity for detecting disease in individual coronary arteries and significantly reduced the intraobserver variability for the patient's classification [115].

6.2.2 SPECT Imaging Analysis

The advances in camera technology in the late 80s allowed single photon emission computed tomography (SPECT) to be used routinely. With its inherently better image contrast, SPECT imaging gradually replaced several exams usually carried out by conventional planar imaging. Typically, a gamma-camera SPECT acquisition may consist of 64 planar views, or projections, each containing 64×64 image pixels and acquired at 64 discrete angles covering 360 degrees around the patient. By the use of reconstruction algorithms it is possible to estimate transaxial, sagittal and coronal sectional images, creating 3D volumes or stacks of short-axis images that served as input to the quantitative software.

Although the objective of defining regional myocardial perfusion abnormalities with SPECT imaging did not change from that of planar imaging, the complexity of the problem increased significantly with the myocardium represented as a 3D volume rather than planar images.

The first effort to address this problem was carried out by Tamaki et al [109]. This quantitative tomographic analysis was assessed by the circumferential profile curves of the three short-axis sections (apical, middle and basal) of the left ventricle. To assess the apical region, the most central long-axis section cutting through the apex was also selected. From these images, circumferential maximal count profiles of myocardial distribution were obtained in a manner similar to that of Meade [79] and Burow [14]. The center of the left ventricle cavity was manually determined and each tomographic image was divided into 36 radial segments at 10 degrees intervals. The distribution of the radionuclide was determined by calculating the highest activity per pixel along each radial segment, normalizing the data to the segment with the highest counts to display a circumferential profile curve in each image. Perfusion defects were classified by use of uptake and washout distributions similar to the methodology used to quantify planar images. The results from these studies showed better sensitivity for detection of diseased coronary vessels than qualitative analysis.

Garcia et al [28, 46, 49] extended the method proposed by Tamaki [109] to the entire SPECT volume by introducing the polar map, or bull's-eye displays, for viewing circumferential profiles. The polar maps allow a quick and comprehensive overview of the circumferential samples from all slices by combining them into a single color-coded image. The points of each circumferential profile are assigned a color based on normalized count values, and the colored profiles are shaped into concentric rings. The most apical slice processed with circumferential profiles defines the center of the polar map, and each successive profile from each successive short-axis slice is displayed as a new ring surrounding the previous. The most basal slice of the Left-Ventricle (LV) makes up the outermost ring of the polar map. Additional maps such as standard deviation map that shows the number of standard deviations below normal of each point in each circumferential profile can aid in evaluation of the study by indicating the severity of any abnormality.

In addition to tabulated defect extent values, various bull's-eye displays were developed to aid in the interpretation of the tomographic slices. The blackout polar

map set those sectors in the polar map that were abnormal to the color black while maintaining the intensity of normal sectors. The blackout map provided a visual representation of the defect location and size. A second map, the defect severity map, mapped the defect abnormality in units of standard deviation below the normal mean. Both maps in combination with the vascular overlay template provided extent and severity values in each vascular region.

To account for differences among patients in the axial dimension of the heart, the slice thickness or polar map ring width was normalized so that each patient's polar map consisted of the same number of sampling points. With this technique, the polar map became the standard template for representing the 3D volume of the LV (Figure 6.3). By eliminating the anatomic variations of the LV, the polar map provided the construction of normal databases as a pattern to compare 3D perfusion distribution from patients under study.

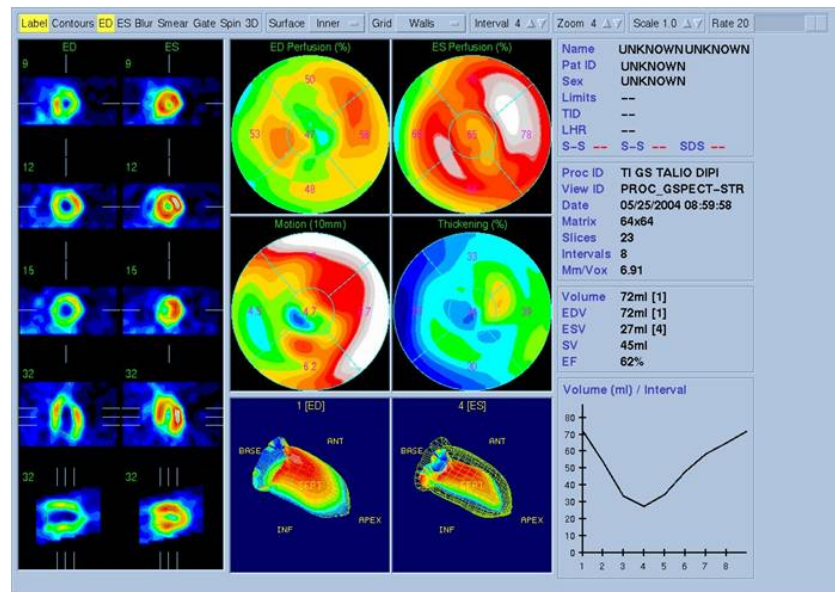


Fig. 6.3 A typical report of a cardiac SPECT study. Some slices are shown on the left part of the screen. The polar maps (bull's eye) are shown on the upper middle and below, the three dimensional surfaces in diastole and systole. (Screenshot taken from QGS – Guido Germano software package).

6.2.3 Gated SPECT Imaging Analysis

Commercial SPECT systems nowadays have the capability of acquiring and reconstructing electrocardiogram (ECG) gated myocardial perfusion studies. This feature

requires ECG triggering, where the R-wave signal (Figure 6.4) is used to align the acquisition data from different cardiac cycles. A gated SPECT study will supply information about the cardiac function averaged over many heart beats during a period when the radiopharmaceutical (^{99}Tc m and ^{201}Tl) is uniformly distributed.

This feature has promoted the use of SPECT for the total automatic assessment of myocardial perfusion and for global and regional function, including the assessment of LV ejection fraction, LV end-diastolic (ED) and end-systolic (ES) volumes and LV myocardial wall motion and thickening.

However, the quantification of global and regional LV parameters requires knowledge of the endocardial and, depending on the parameters being determined, the epicardial surface throughout the cardiac cycle. These surfaces determination may be manually assigned or automatically detected. Manual methods are subjective and time consuming. The automated methods use either boundary detection or geometric modeling.

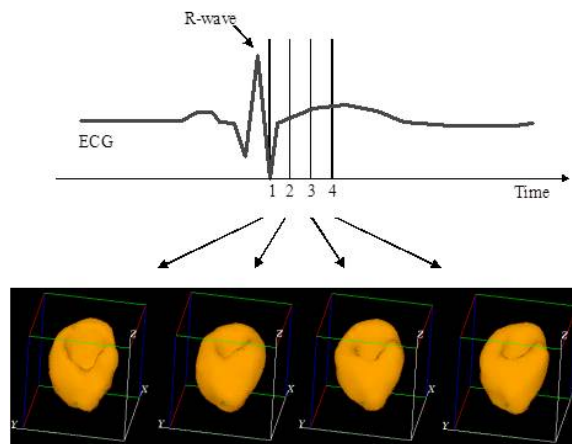


Fig. 6.4 The use of an ECG signal to provide gating for the SPECT study.

A large effort has been devoted to the analysis and segmentation of cardiac images by methods guided by prior geometric knowledge. In general they can provide, with a limited number of global parameters, a rough shape approximation.

Faber et al. [33] developed a general model-based surface detector for finding the four dimensional (three spatial dimensions plus time) endocardial and epicardial left ventricular boundaries for both perfusion and blood-pool SPECT studies. The model encoded LV shape, smoothness, and connectivity into the compatibility coefficients of a relaxation labeling algorithm. From the surface estimates, volumetric data were extracted to estimate the LV ejection fraction and regional wall motion was determined using a modified version of the centerline method. The final surfaces were visualized in a 3D rendering in which the displayed perfusion information could be presented with regional wall motion.

Cauvin et al. [15] modeled the LV as a truncated bullet, a combination of an ellipsoid and a cylinder, that is fitted to the morphological skeleton of the LV. Faber et al. [32] used a combination of cylindrical and spherical coordinate systems to build a discrete model of the LV in SPECT perfusion images. A radius function defined in a discrete space of longitudinal and circumferential coordinates describes the LV. For each orientation, the radius is determined by finding the position of maximal perfusion. After low-pass filtering to remove outlier radii, the radius function is mapped back to Cartesian space where the surface is represented using triangular meshes.

Germano et al. [51, 52, 53, 54, 55] introduced a fully automated algorithm that determined the LV surfaces from gated perfusion SPECT. In this approach, automatic border definition is done using thresholds of a Gaussian function pre-calibrated to a phantom. Although the spatial resolution of SPECT is too low to actually measure the edges of the endocardial surface, the pre-calibration to a phantom and the high tolerance of the ejection fraction calculation provided highly reproducible volumetric estimates that correlated well with LV ejection fraction from other imaging modalities. This method has become popular and commercially available because it was the first to offer a fully automatic processing for gated perfusion. Figure 6.3 shows the patient's myocardial endocardial surface at end-diastole and end-systole depicted on an arbitrary orientation calculated by this method. This display may be viewed in a dynamic fashion to visually assess wall motion showing excursion and thickening of the wall as changes in color. The color is a function of both myocardial perfusion and myocardial thickness. A myocardial segment that increases in color brightness indicates thickening, and thus viable myocardium.

Garcia et al. [31] proposed an expert system called PERFEX as a tool for computer-assisted diagnosis of stress/rest myocardial perfusion SPECT studies. This approach has the potential for standardizing the image interpretation process. After reviewing hundreds of studies from patients with coronary artery disease, heuristic rules were derived that best correlated the location of perfusion defects in SPECT studies. Two sets of gold standards were used to define the rules: the interpretation by human experts and the results of coronary angiography studies. These rules operate on data that is input to the PERFEX from the SPECT quantification process, which identifies defects as portions of the myocardium where normalized perfusion falls below a predetermined number of standard deviation when compared to a normal file.

6.2.4 Motion Analysis in SPECT

A number of techniques have been used in order to describe and quantify the motion of the cardiac structures. They can be divided into three main categories: (1) detecting endocardial motion by observing image intensity changes; (2) determining the boundary wall of the left ventricle and tracking it; (3) attempting to track anatomical implanted or induced myocardial landmarks. There are few problems involved with each of these techniques [39]. Assumptions must be made about the motion model

in the first two categories in order to obtain a unique point-wise correspondence between frames. To this end, optical flow methods [26, 27, 29, 56, 81, 92, 107, 108] have been applied for (1) and curvature-based matching [4, 11, 40, 67, 84] has been applied to find point correspondence in (2). Landmark-based methods [3, 5, 27, 36, 68, 72, 121, 122] provide information on material point correspondence. However, this information is mostly sparse and assumptions on the type of motion have to be made in order to regularize the problem of finding a dense displacement field. The use of implanted markers adds the extra complication of being invasive. Although implanted markers are usually regarded as the gold standard, there are some concerns about their influence on both image quality and modifications of motion patterns.

In gated SPECT, Bardin et al. [7, 8] extended a geometric model based on superquadrics [9] with the use of a free-form deformation (FFD), a technique introduced in computer graphics by Sederberg and Parry [104], to estimate LV wall motion. This is accomplished by deforming the full model (superquadric and FFD) in the first frame, and modifying only the FFD in the subsequent frames. By tracking points with the same parametric coordinates along the cardiac cycle, a number of dynamic parameters like wall thickening and twisting motion can be computed.

Declercq et al. [23] have introduced a spatio-temporal model to segment the LV and to analyze motion from gated SPECT sequences. The model relies on a planispheric transformation that maps endocardial points in one time frame to the corresponding material points in any other frame. First, endocardial edge points are detected in all frames using a Canny-Deriche edge detector [86] in spherical coordinates. Selected points in subsequent frames are matched to the current frame using a modification of the Iterative Closest Point algorithm [24]. Based on corresponding point pairs, the parameters of a planispheric transformation are retrieved by least square approximation. This transformation allows the description of motion parameters such as radial motion, twisting motion and long-axis shortening.

Gutierrez et al. [57, 58] proposed the combination of optical method with a multi-resolution technique to estimate the myocardial Kinetic Energy (KE) obtained directly from gated SPECT images. Specifically, the method quantifies the 3D LV motion by a series of 3D velocity vector fields computed automatically for each voxel on the sequence of cardiac volumes. The velocity estimation for each voxel is used to compute the corresponding kinetic energy and the results are presented in a compact 2D form. The method was validated with a synthetic phantom and applied to groups of 30 volunteers and 29 patients. Using the method, the cardiac condition of each subject can be studied taking the relation between the maximum and minimum values of kinetic energy observed during the cardiac cycle.

Meyering et al. [82] proposed a method to automatically estimate the velocity vector field based on the study of variation in frequency content in a time series of 2D gated SPECT images. The frequency analysis is performed by computing the Wigner-Ville and the Choi-Williams distributions [20, 21] for each image pixel, yielding the corresponding 3D-frequency spectrum. From this 3D spectrum the local velocity of each pixel is calculated by employing a multiple linear regression

model. Experimental validation was carried out using synthetic phantoms that simulate translation and rotation between successive frames.

The tracking of structures in time series images has been studied by the computer vision community, especially in the areas of non-rigid motion, segmentation and surfacing mapping [37]. The goal is to obtain a displacement field that establishes a correspondence between certain points in the structure at time t and time $t+1$.

Generally speaking, the common methods to obtain velocity vector fields lie within feature matching, gradient, frequency and spatio-temporal based techniques. The matching technique is very sensitive to ambiguity among the structures to be matched. This is particularly true in the considerations of deformable structures where the geometrical relationships between object points may be distorted. Moreover, the correspondence method can be beset with combinatorial explosions of matching possibilities between two time-varying images.

6.2.4.1 Gradient based approach

The gradient based method or the differential method was introduced by Fennema and Thompson [10, 34, 73] and developed by Horn as the Optical Flow (OF) equation [63]. This method is based on the assumption that the intensity of image elements is conserved between the images. The equation formulated in the continuum is the well-known motion constraint equation:

$$E_x u + E_y v + E_t = 0 \quad (6.1)$$

where, E_x, E_y and E_t are the image derivatives in x, y and t directions, u and v are the components of the local velocity vector \mathbf{v} along the directions x and y , respectively.

In cardiac SPECT images using a radionuclide such as ^{201}TI and $^{99}\text{Tc}^m\text{-MIBI}$, the principal structure is the myocardium, therefore it is a reasonable assumption that the total intensity of voxels within the ventricle is conserved between successive frames, and that the motion of connected tissue within the myocardium should be smooth [22, 69].

Eq. (6.1) is usually called the OF constraint (OFC) and to evaluate the flow velocity, Horn and Schunck [63] have introduced a “smoothness” constraint in addition to the fundamental constraint. Using the Horn and Schunck method and the extension to three dimensions, the velocity flow field can be obtained by minimizing the functional:

$$\min \iiint \left[(E_x u + E_y v + E_z w + E_t)^2 + \alpha^2 (u_x^2 + u_y^2 + u_z^2 + v_x^2 + v_y^2 + v_z^2 + w_x^2 + w_y^2 + w_z^2) \right] dx dy dz \quad (6.2)$$

where the first term is the OFC, the second is a measure of the Optical Flow field smoothness, and α is a weighting factor that controls the influence of the smoothness constraint. The functionally Eq. (6.2) is minimized by using the calculus of variations [102], which leads to a system of three coupled differential equations from the Euler-Lagrange equations:

$$\begin{aligned}
\nabla^2 u &= \frac{E_x}{\alpha^2} (E_x u + E_y v + E_z w + E_t) \\
\nabla^2 v &= \frac{E_y}{\alpha^2} (E_x u + E_y v + E_z w + E_t) \\
\nabla^2 w &= \frac{E_z}{\alpha^2} (E_x u + E_y v + E_z w + E_t)
\end{aligned} \tag{6.3}$$

These equations can be easily decoupled, and an interactive solution can be defined using discrete approximation of the Laplacian operator with a finite difference method. Therefore, the following set of equations is used to estimate the 3D OF components at each time instant

$$\begin{cases}
u^{n+1} = \bar{u}^n - \frac{E_x(E_x \bar{u}^n + E_y \bar{v}^n + E_z \bar{w}^n + E_t)}{\alpha^2 + E_x^2 + E_y^2 + E_z^2} \\
v^{n+1} = \bar{v}^n - \frac{E_y(E_x \bar{u}^n + E_y \bar{v}^n + E_z \bar{w}^n + E_t)}{\alpha^2 + E_x^2 + E_y^2 + E_z^2} \\
w^{n+1} = \bar{w}^n - \frac{E_z(E_x \bar{u}^n + E_y \bar{v}^n + E_z \bar{w}^n + E_t)}{\alpha^2 + E_x^2 + E_y^2 + E_z^2}
\end{cases} \tag{6.4}$$

where n is the iteration index; E_x, E_y, E_z are the partial derivatives of the image intensity in the directions x, y and z , respectively; E_t is the partial derivative in time; \bar{u}, \bar{v} and \bar{w} are the mean velocities in each direction, for the voxels in a neighborhood of a given voxel; and α is a weighting factor. Velocity components in the x, y and z directions for each voxel are computed as the solution of a linear algebraic system of equations whose coefficients are determined by the spatial and temporal derivatives given by Eq. (6.4). The linear system described in Eq. (6.4) can be solved by methods like Conjugate Gradient methods [6] or Algebraic Reconstruction techniques [62].

As can be noted, the asymptotical complexity of the method on a sequential machine is a function of the number of iterations, n , that are needed to obtain the final solution and of the image dimensions, which is $O(n^3 D_x D_y D_z)$. It should be noted that, in the iterative solution presented, the initial value for OF estimation is set to zero (i.e., $\bar{u}^{n=0} = \bar{v}^{n=0} = \bar{w}^{n=0} = 0$) and the guessed value at time t is obtained from the previous time-step (i.e., $\bar{u}^n = u^{n-1}$; $\bar{v}^n = v^{n-1}$; $\bar{w}^n = w^{n-1}$, where $n-1$ is the number of iterations executed at the previous time-step). This reduces the number of iterations needed to obtain the OF estimation, as well as the computational effort.

The estimation of velocity components using OF approach depends on derivative approximations from discrete data. Numerical derivative is ill-posed in the sense of Hadamard theory [59]. A problem is well-posed in the sense of Hadamard if its solution: **(1)** exists, **(2)** is unique, and **(3)** depends continuously on the initial data. Digital images are always “dirty” in the differential geometric sense. In theory, $E(\mathbf{x})$ may be everywhere discontinuous. Some algorithms for differentiation try to avoid this problem by regularizing the data, for instance, by blurring the original data. Nevertheless data should not be modified, and a regularization of the operand must be found.

Ill-posedness of numerical differentiation can be modified into a well-posed problem by using filters in the form of scaled convolution operators [71]. It has been established that for front-end vision systems (in which no knowledge about the observed scene is present) the unique linear scale space constructor is the Gaussian function:

$$G(\mathbf{x}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^D} e^{-\frac{\|\mathbf{x}\|^2}{2\sigma^2}} \quad (6.5)$$

where D denotes the dimension of the input domain. A blurred replica of the original image is obtained by convolution with $G(\mathbf{x}; \sigma)$ for a specific σ . The stack of images as a function of increasing scale parameter σ is coined a linear scale-space.

Equivalently, this set can be generated by the solution to the heat diffusion equation:

$$\frac{\partial L(\mathbf{x}, s)}{\partial s} \equiv \sum_{i=1}^D \frac{\partial^2 L(\mathbf{x}, s)}{\partial x_i^2} \quad (6.6)$$

where L is the pixel intensity located at \mathbf{x} , s is the continuous scale parameter ($\sigma^2 = 2s$) and the original image $L(\mathbf{x}; 0) \equiv E(\mathbf{x})$ is the initial condition for the linear second order differential Eq. (6.6).

One of the most useful results of scale space theory is that the derivatives of the Gaussian function are also solutions to the diffusion equations, and together with the 0^{th} order Gaussian they form a complete family of differential operators. The 0^{th} order kernel is a scalar, all higher order kernels are tensorial quantities as, for example, the first order kernel, the gradient, is a vector.

A scaled image $E(\mathbf{x}; \sigma)$ is a smooth function for all $\sigma > 0$ making scaled differentiation well-posed by nature. Taking the following identity:

$$\partial \{E(\mathbf{x}) \otimes G(\mathbf{x}, \sigma)\} = \partial E(\mathbf{x}) \otimes G(\mathbf{x}, \sigma) = E(\mathbf{x}) \otimes \partial G(\mathbf{x}, \sigma) \quad (6.7)$$

It follows that to obtain well-posed Cartesian partial derivatives of order n of a re-scaled image $L(\mathbf{x}; \sigma)$ it is only necessary to convolve the original image $E(\mathbf{x})$ with the corresponding partial derivatives of the 0^{th} order Gaussian $G(\mathbf{x}; \sigma)$. This result enables the extraction of any order of derivative for discrete data. There is, however, a close relation between the order of differentiation, the scale of the operator and the accuracy of the result. This leads to the important concern about what should be the best scale to a specific problem. The issue can be addressed theoretically and a criterium for best scale selection was suggested by Lindeberg [75]. However, in practice we are interested in some specific objects and, in this case, some scales are better suited than others and can be determined experimentally.

6.2.4.2 Frequency based approach

This method is most appropriate for determining the motion of a single object moving across a uniform background. The method takes advantage of the shift property of the Fourier transform.

Assume one is given two images, denoted $f(x, y)$ and $g(x, y)$, that are related by a pure translation $(\Delta x, \Delta y)$. Since the phase functions of their Fourier transforms $F(u, v)$ and $G(u, v)$ respectively, are given by:

$$\begin{aligned}\phi_F(u, v) &= \arg \{F(u, v)\} \\ \text{and} \\ \phi_G(u, v) &= \arg \{G(u, v)\} = \arg \{F(u, v)\} - (u\Delta x + v\Delta y) \\ \text{and} \\ \phi_G - \phi_F &= [\arg \{F(u, v)\} - (u\Delta x + v\Delta y)] - \arg \{F(u, v)\}\end{aligned}\quad (6.8)$$

If ϕ_G and ϕ_F are known at two distinct frequencies (u_0, v_0) and (u_1, v_1) then Eq. (6.8) defines a system of two equations in two unknowns which can be solved for $(\Delta x, \Delta y)$. This method is clearly more restrictive than the gradient approach since only a single non-localized velocity vector is obtained for each image frame.

6.2.4.3 Spatio-temporal frequency based approach

The spatio-temporal frequency (STF) based approach to optical flow derivation encompasses all methods which are based upon some underlying spatio-temporal frequency image representation. The major motivation for considering the use of the STF image representation as a basis for computing optical flow comes from the literature on mammalian vision. In particular, recent investigations have demonstrated that many neurons in various visual cortical areas of the brain behave as spatio-temporal frequency bandpass filters [2, 42, 43].

In the field of non-stationary signal analysis, the Wigner-Ville Distribution (WVD) has been used for the representation of speech and image. Jacobson and Wechsler [64, 65] first suggested the use of the WVD for the representation of shape and texture information. In particular, they formulated a theory for invariant visual pattern recognition in which the WVD plays a central role. Meyering *et al.* [82] extended this concept to analyze motion in gated SPECT images.

Given a time-varying image $f(x, y, t)$, its WVD is a 6-dimensional function defined as:

$$W_f(x, y, t, w_x, w_y, w_t) = \int_{-\infty}^{+\infty} \int \int R_f(x, y, t, \alpha, \beta, \tau) e^{-j(\alpha w_x + \beta w_y + \tau w_t)} d\alpha d\beta d\tau \quad (6.9)$$

where,

$$R_f(x, y, t, \alpha, \beta, \tau) = f(x + \alpha, y + \beta, t + \tau) \cdot f^*(x - \alpha, y - \beta, t - \tau) \quad (6.10)$$

is the pseudo-correlation function, where $*$ denotes complex conjugation.

For the special case where a time-varying image is uniformly translating at some constant velocity (u, v) , the image sequence can be expressed as a convolution between a static image and a translating delta function

$$f(x, y, t) = f(x, y) * \delta(x - ut, y - vt) \quad (6.11)$$

Using the convolution and windowing properties of the WVD, we obtain:

$$W_f(x, y, t, w_x, w_y, w_t) = \delta(uw_x + vw_y + w_t) W_f(x - ut, y - vt, w_x, w_y). \quad (6.12)$$

The WVD of a linearly translating image with constant velocity is everywhere zero except in the plane defined by $\{(x, y, t, w_x, w_y, w_t) : uw_x + vw_y + w_t = 0\}$, for fixed (u, v) . Equivalently, for an arbitrary pixel at x, y and t , each local spatial and temporal frequency spectrum of the WVD is zero everywhere except on the plane defined by $\{(w_x, w_y, w_t) : uw_x + vw_y + w_t = 0\}$.

From Eq. (6.9), the WVD assigns a three-dimensional spatio-temporal frequency spectrum to each pixel over which the image is defined. However, the WVD assigns a 3D spectrum with interference due to cross correlation when more than one frequency is present [21].

In order to smooth the spectrum of WVD a filter must be introduced. Meyering *et al.* [82] adopted a Hanning filter to smooth the spectrum

$$h = 0,5 * \left[1 - \cos \left(\frac{2\pi n}{N} \right) \right] \text{ for } 0 \leq n \leq N - 1 \quad (6.13)$$

They also used the Choi-Williams distribution (CWD) to reduce the effects of the cross-terms. The CWD was introduced with the aim of controlling the cross-terms encountered in Wigner-Ville distribution. The exponential kernel introduced by Choi and Williams is defined as:

$$cw = \frac{1}{\sqrt{4\pi\tau^2\sigma}} \exp \left(\frac{-(\mu - t)^2}{4\tau^2\sigma} \right) \quad (6.14)$$

From the Eq. (6.14), if a small σ is chosen, the Choi-Williams distribution approaches the Wigner-Ville distribution, since the kernel approaches to one. On the other hand, for large σ more cross-terms are suppressed and auto-terms are affected.

The spatial orientation of the smoothed 3D frequency spectrum is completely governed by the pixel velocity, whose components can be obtained through a simple multiple linear regression model [17, 70]:

$$w_t = b + uw_x + vw_y + e \quad (6.15)$$

Eq. (6.15) is a linear regression extension where w_t is a linear function of two independent variables w_x and w_y . The values of the coefficients b , u and v are achieved by solving the following linear system:

$$\begin{bmatrix} n & \sum_{k=1}^n w_{x_k} & \sum_{k=1}^n w_{y_k} \\ \sum_{k=1}^n w_{x_k} & \sum_{k=1}^n w_{x_k}^2 & \sum_{k=1}^n w_{x_k} w_{y_k} \\ \sum_{k=1}^n w_{y_k} & \sum_{k=1}^n w_{x_k} w_{y_k} & \sum_{k=1}^n w_{y_k}^2 \end{bmatrix} \cdot \begin{bmatrix} b \\ u \\ v \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n w_{t_k} \\ \sum_{k=1}^n w_{x_k} w_{t_k} \\ \sum_{k=1}^n w_{y_k} w_{t_k} \end{bmatrix} \quad (6.16)$$

where n is the number of pixels, w_{x_k} , w_{y_k} and w_{t_k} are the frequency components in each direction and u and v are the velocity components on x and y directions, respectively.

Figure 6.5 shows the LV 3D velocity vector field estimated by the gradient method. In this figure the complex deformation observed in the LV, during the systole can be explained by observing the principal fiber paths in a dissected LV as reported by Torrent-Gausp [110]. Figure 6.6 shows a SPECT slice at systole and diastole, superimposed by their velocity vector fields calculated employing the spatio-temporal frequency based approach using WVD and CWD.

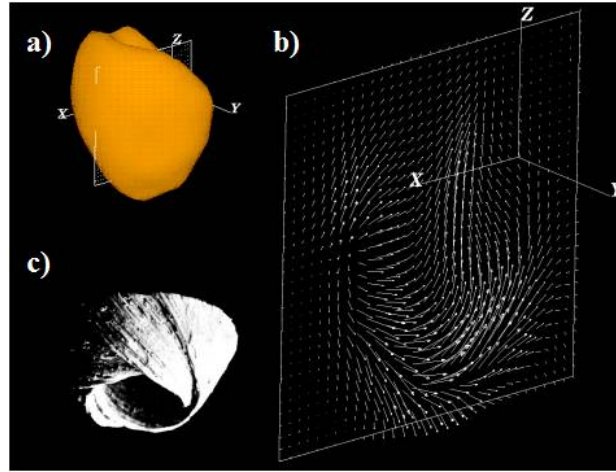


Fig. 6.5 a) a view of the 3D left ventricle image of a normal heart; b) the corresponding velocity vector field in a coronal slice at systole estimated by the gradient-based approach; c) the principal fiber pathways of a cow's LV (Torrent-Gausp¹¹⁰).

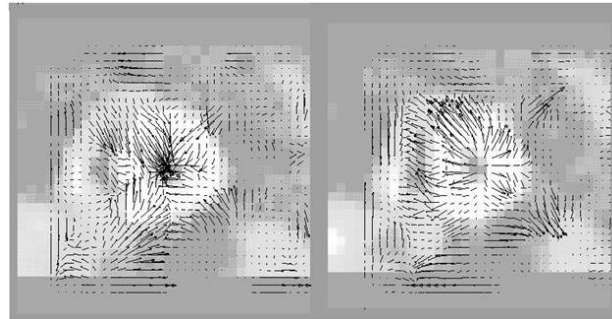


Fig. 6.6 Oblique SPECT slice at systole (left) and diastole (right), superimposed by their velocity vector fields estimated employing the frequency-based approach using WVD.

6.3 Visualization of myocardial movement

Any data analysis task performed by humans can be improved by the use of meaningful visualization techniques [111]. Therefore, as in many other fields, the visualization step plays an important role in the process of medical image analysis. The most common method to analyze the left ventricle movement is by visually inspecting sequential time frames of a single slice. The so-called cine mode provides an intuitive visualization way for doctors to analyze the movement of both the cardiac walls and the blood pool. To achieve a more comprehensive understanding of the movement of a particular cardiac acquisition, it is important to perform this inspection in more than one slice. The whole process usually takes a lot of time and due to its highly subjective nature such visual analyses lack accuracy and reproducibility [76, 106]. Besides they may introduce considerable inter-observer and intra-observer variability.

The quantification of the myocardial motion by computer methods provides reproducible measures and parameters, thus improving the quality of the diagnostic and providing more reliability and reproducibility of the data interpretation [93]. Motion analysis methods, discussed in item 2, usually produce velocity or displacement fields as output data. Intra-myocardial motion can also be object of study by dividing the myocardium into different layers [103]. Derived parameters from such fields, such as wall stress, strain (that provides information about the myocardial deformation) and strain rate (temporal derivative of the deformation) [117, 119] can add important information about the deformation of the myocardium throughout the cardiac cycle [39]. Usually there is a lot of information to be presented simultaneously making the chosen visualization method a key point in the interpretation of the results.

A large number of methods to present the results of the quantification techniques in friendly and easy ways have been proposed in the last decade. In the following items we are going to present a short review of the techniques described in the literature. Since the difficulties faced by methods of visualization of vector and displacement fields are shared with other knowledge areas, a description of general methods to visualize vector fields is provided in the first item. Then, specific techniques used for visualization of movement in medical images are presented. Finally we present some visualization schemes developed by our research group for myocardial movement analysis.

6.3.1 Visualization of velocity vector fields

The visualization of a vector field and a flow can be considered equivalent in many applications since the velocity of a flow is represented by a vector field and a vector field can define a flow [78]. The straightforward method of visualization of a vector field is to display all vectors as arrows with their tails at each point of the field grid. However, this method can generate output images with complex and sometimes

unintelligible information. Figure 6.5b presented in the previous section is a good example of this effect. To make this bunch of information useful for diagnostic purposes, it is necessary to find compact and friendly representations for it.

Since flow visualization is a relevant topic for many knowledge areas, it has been object of many research works that have produced elegant and complex visualization methods and tools for scientific data. Fuchs and Houser [41] present a comprehensive survey on visualization of complex data in many knowledge fields. Max [78] presents a concise review of methods for scientific visualization and a state of the art report on flow visualization is presented by Laramée et al [74]. The main techniques for flow visualization presented by the three reviews can be summarized as: (1) drawing arrows; (2) color coding the velocity field; (3) use of texture to generate a dense representation of the flow; (4) use of geometric objects as a basis for flow visualization, for example streamlines, streak lines, path lines and glyphs; (5) Layering, where multiple layers of information are put on top of each other; (6) topology that locates critical points where the velocity is zero and analyzes the derivatives of the velocity in those points. The authors also emphasize the fact that sometimes the visualization of the flux of the vector field itself may not be useful to understand the phenomenon under study. In such cases, the visualization of derived quantities or parameters may be the best choice.

Although very elegant solutions to the problem of visualization of vector fields are provided by many of the method proposed for general purposes, their usefulness can be seriously challenged by the difficulty the doctors would have to interpret the sometimes complex display of information. Since time is a critical issue when dealing with medical information, the ease of use and understanding are important requisites for any tool deployed for medical visualization applications [99].

6.3.2 Visualization of velocity vector fields in medical images

Although this chapter is dedicated to Nuclear Medicine quantification methods, there are many visualization techniques described for other for other medical imaging modalities that present useful solutions that can be adopted in Nuclear Medicine.

The solutions for displaying velocity fields in medical images of all modalities are usually based on some of the simpler methods described for flow visualization in the former item. In cardiac images most common visualization methods of vector fields are drawing arrows and color coding [31, 61, 80, 89, 101, 120]. Early papers [61] presented plots of the vector fields as arrows. Xu et al [120] display information about the velocity field from magnetic resonance (MR) images using both arrows and color coding superposed to the 2D slices and 3D rendered surfaces. Color coding has been commonly used to quantify blood flow in echocardiographic Doppler images since its description in 1987 [80]. New applications for the technique are described frequently [44, 87, 113, 114]. In MRI arrows indicating the velocity field are displayed at each pixel of 2D long axis myocardial view by Wen et al [118]. Jung et al [66] present velocity information using high temporal resolution Tissue

Phase Mapping in 2D slices both by color coding the velocity values and by arrows representing the velocity vectors. Large arrows representing the mean values over areas of the Left Ventricle (LV) are superimposed to the arrows representing each velocity vector in the grid. This representation allows a better visualization of the overall behavior of LV regions. Osman and Prince [95] present methods for visualization of myocardial function using Harmonic phase MR imaging (HARP MRI). Pathways and velocity field arrows are used for displaying movement information in 2D slices. Eusemann et al [31] use a color map distribution to present the regional maximum speed superposed to rendered left ventricle planes in Computed Tomography (CT) images.

In recent years, glyph based visualization methods have been studied in medical applications [20, 83, 89, 99]. Glyphs are geometric objects described by a number of properties. By changing these properties one also alters the visual representation of the glyph. The visualization of multiple information - or parameters - is achieved by associating each parameter with one or more properties of a glyph [99]. Shape, color, transparency and placement of a glyph are the main characteristics used in glyph-based visualization techniques. Meyer-Spradow et al [83] presented a technique for movement analysis in SPETC images where they present parameters (with systolic and diastolic information) simultaneously by exploring different glyph properties. Pathlines were used to visualize myocardial movement in MR images in a paper by Abd-Elmoniem et al [1] and in CT images by Eusemann et al [31].

In SPECT, a widely used visualization tool for visualization of the myocardial movement was introduced by Germano et al [50]. The tool – QGS/QPS/AutoQUANT (Cedars-Sinai) – depicts the endocardium as a rendered surface and the epicardium as a 3D grid at both end diastole and end systole. Like AutoQUANT most tools for movement analysis are restricted to maximum diastole and systole frames. However, since cardiac SPECT acquisition is performed over 8 or 16 time frames, the ability to visualize information about the myocardial movement throughout the cardiac cycle is a desired feature of the analysis tool.

Derived parameters are mainly represented by curves and color coding images. Strain information from MR images is presented in 2D slices using color coding by Abd-Elmoniem et al [1] and Wen et al [118]. Color coding has also been used to display strain and strain rate information in echocardiographic studies [30, 87, 92].

6.3.3 Myocardial movement visualization schemes

6.3.3.1 2D visualization

Generally speaking, the heart can be described as a non-rigid object that deforms throughout the cardiac cycle and has very complex mechanical properties [58]. To simplify the analysis of the left ventricular movement, it can be described in terms of contraction/expansion and torsion. In order to evaluate the movement of the LV three movement directions were defined, each with two possible orientations: (1)

radial movement is described as a contraction towards the center of the LV during systole and as an expansion from the center during diastole; (2) horizontal rotation represents the clockwise and counter-clockwise movement of the cardiac walls; (3) vertical rotation represents the movement towards the base (upwards) during systole and towards the apex during diastole (downwards). These directions are depicted in Figure 6.7.

The calculation of the velocity fields by solving the linear system given by Equation 6.4 gives the values of the velocity components for each voxel of the cardiac volume in Cartesian coordinates. However, the spherical coordinate system is more suited for the description of the movements presented in the former paragraph. A transformation of the Cartesian component velocities to the unit vectors in the directions r , θ , ϕ in the spherical coordinate system is performed for the presentation of the described directions. The radial movement is described by the unit vector for the r component, the horizontal rotation by the unit vector for the θ component and the vertical movement by the unit vector for the ϕ component.

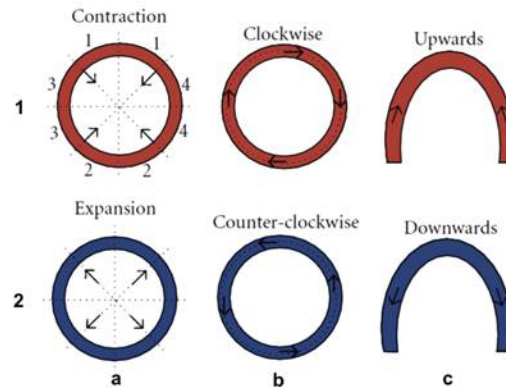


Fig. 6.7 Motion directions defined in the cardiac cycle: (a) radial movement in terms of contraction (Line 1) and expansion (Line 2); (b) horizontal rotation is either clockwise or counterclockwise; and (c) vertical rotation is upwards or downwards. The two orientations for each direction are colored by the defined coding scheme. Movements are depicted using the short axis view in (a) and (b) and the horizontal long axis view in (c). The LV walls are depicted in line 1(a): region 1 is the anterior wall; region 2 is the inferior wall; region 3 is the septal wall; and region 4 is the lateral wall. (The nomenclatures of cardiac planes and wall segments used in this work follow the recommendations of the American Heart Association, as described in Cerqueira et al. (2002)).

In order to indicate velocity intensity, a discrete lookup table is used. In this table, the absence of motion is depicted as white, positive values are depicted as blue, and negative values are depicted as red. The positive and negative colors are divided into 128 steps by changing their saturation, such that fast movement is represented by a high intensity color, and a slower movement has a lighter color (Figure 6.8).

By analyzing left ventricle images, the resulting visualization of the motion patterns can be compared with the motion expected from the heart physiology (depicted schematically in Figure 6.7). As an example, Figure 6.8 shows the results obtained

for one normal subject and one patient¹ using the velocity color coding scheme. In this Figure, the velocity images for the radial directions in a slice from the midcavity portion at both systole (line 1) and diastole (line 2) are schematically presented. During the systole the myocardium contracts, starting at the apex and moving upwards to the base. Simultaneously, the septal and lateral walls move towards the center of the left ventricle. Therefore the expected result in systole is the contraction, as is seen in Figure 6.8, line 1 of Normal subject.

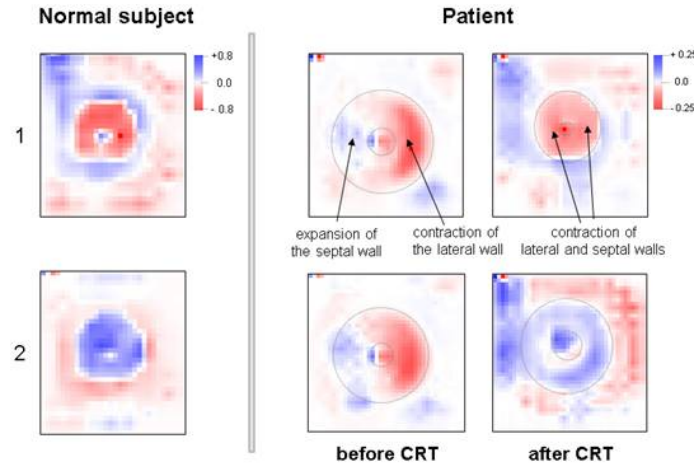


Fig. 6.8 Radial velocity images of a normal subject (left) and a patient (right) in a slice from the midcavity portion of the LV. Line 1 presents systolic images and line 2 presents diastolic images. The color scales for normal subject and patient are shown at the top right of each corresponding figure.

The contraction movement is represented by different tones of red, indicating the contraction with varying intensities. After ejection the heart enters the diastole, where the overall motion is of expansion. The expected colors are therefore also the opposite of the ones observed in systole, as can be observed in Figure 6.8, line 2 of Normal subject, where the expansion movement is depicted as different tones of blue. The Patient image presents the results for a patient diagnosed with a Ventricular Dyssynchrony [18] condition and that underwent to cardiac resynchronization therapy (CRT) [97]. The figure presents the effect of a successful application of the CRT. The arrows in systole images point to septal and lateral walls. Before CRT pix-

¹ The real images presented in this chapter were acquired at the Nuclear Medicine Department of the Heart Institute (InCor) – University of Sao Paulo Medical School. All acquisitions were performed by a dual-head rotating gamma camera (ADAC VertexPlus with a LEAP Collimator). The acquisition process was synchronized with the electrocardiogram and the cardiac cycle was divided into 8 or 16 frames/cycle. A total of 64 projections were obtained over a semicircular 180-deg orbit. All projection images were stored using a 64x64, 16-bit matrix. The transverse tomograms were reconstructed with a thickness of 1 pixel/slice (6.47 mm).

els from the septal wall present positive values (blue color) while in the lateral wall they present negative values (red color), indicating a dyssynchrony in the movement of the opposite walls. The expected normal movement would be an overall contraction of the walls, represented by red color. Such movement can be observed in systole after CRT. The arrows in the figure indicate the global contraction movement depicted in tones of reds. The overall intensity of the movement in patient images is weaker when compared to the normal subjects (see color legend values for normal and patient figures).²

6.3.3.2 Speed polar map

The application of the coding scheme to visualization in 2D planes presented in the previous item is a cumbersome task, since it involves the inspection of many slices in several phases. Besides, a quantitative comparison of heart images from different subjects is difficult, and is only meaningful if a previous normalization for their size is performed. A way of overcoming this problem is to present the coded velocity components as a functional polar map. Intensity polar maps or bulls eye presents the 3D information as 2D maps and are widely used in Nuclear Medicine for quantification of SPECT intensity images.

The bull's eye projection is a 2D map of a 3D left ventricle, whose display is a polar projection with the apex in the center, the mid-cavity in the middle, and the base in the periphery. The standard representation for bull's eye set up by the American Heart Association (AHA) is independent of size and orientation of the heart [16]. The left ventricle is divided in three thick slices with a distribution of 35%, 35%, and 30% for the basal, mid -cavity, and apical thirds of the left ventricle. These slices are further divided into segments, giving a total of 17 segments as it is shown in Figure 6.9a. The 17 segments are then arranged in a polar projection as shown in Figure 6.9b.

In the proposed scheme AHA bull's eye representation is used to present the velocity in each movement direction, thus allowing the analysis of the myocardial movement in one time frame by the inspection of the three maps. The information presented in the polar map is the mean or maximum speed - the modulus of the velocity vector - of each segment. As an example, Figure 6.10 depicts the mean speed bull's eyes of a normal subject for the three defined movement directions for two frames, in diastole and systole. The information presented is the mean speed from the voxels belonging to each segment.

Bull's eye have been traditionally used for inspecting parameters at a specific frame of time. Some articles describe methods to present more than one parameter in the bulls eye map. Oeltze et al [89] described a refined Bull's eye map for SPECT and PET images where each segment has been bisected, with resulting outer and inner rings presenting the stress and the rest perfusion maps. Föll et al [38]

² A comprehensive description of the experiments and results obtained with the application of the proposed scheme to patients with Ventricular Dyssynchrony has been published by Rebelo et al (2010) [96].

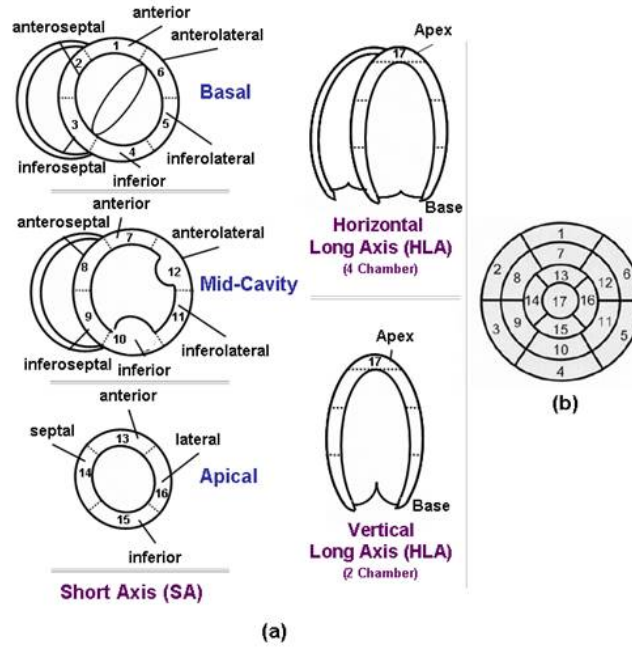


Fig. 6.9 (a) Segmentation of the left ventricle in 17 segments; (b) The segments displayed on a circumferential polar plot (Drawing based on a figure from Cerqueira et al [16]).

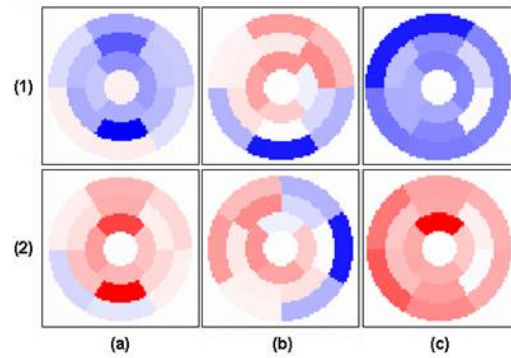


Fig. 6.10 Mean speed polar maps of a normal subject. Line 1 depicts the maps at the maximum relaxation frame in diastole and line 2 depicts the maps at the maximum contraction frame in systole. Column (a) depicts the maps of the radial component, column (b) the maps of the horizontal rotation component and column (c) the maps of the vertical component.

present velocity information in MR Tissue Phase Mapping images from two different planes in a single polar map. The short axis plane component is depicted as vectors centered in the corresponding segment and the long axis plane component is color coded. Each polar map contains information about one time frame. Changes in the movement throughout the cardiac cycle are visualized by displaying five polar maps calculated in different points of the cardiac cycle.

6.3.3.3 Dynamic polar map

We propose a novel scheme with the goal of creating an easily understandable visualization that preserves the temporal and regional information of any temporal parameter distribution. The time varying parameter is displayed in one polar map presenting the medium or maximum parameter value extracted from all time frames of any temporal parameter distribution combined with a graphical representation of the mean parameter value over time. The visualization is able to provide information about the regional and temporal myocardial condition, which may prove viable for use within a clinical setting.

As an example application, Figure 6.11 presents the proposed dynamic bulls eye visualization for the parameter kinetic energy (described in section 2.2.4), calculated for a set of images from the phantom XCAT [105].

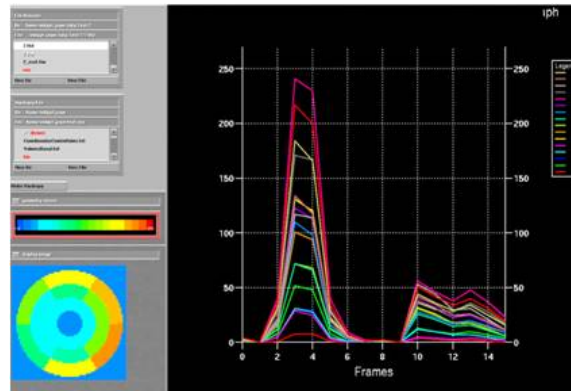


Fig. 6.11 Screenshot of dynamic bulls eye interface showing the polar map and curve graph representation of the kinetic energy distribution (kinetic energy data obtained from a simulated XCAT phantom data set).

6.4 Conclusions and Outlook

In this Chapter we have reviewed the techniques to analyze, quantify and visualize LV parameters by means of SPECT proposed in the last two decades. The strength and availability of these quantitative tools have in many ways provided a competitive advantage to Nuclear Cardiology compared with other higher-resolution non-invasive imaging modalities for the detection of coronary artery disease. During the 80s, the planar imaging of patients with low probability of coronary artery disease allowed the construction of normality databases and the definition of mean and variance limits for normal patients. With the definition of mean and variance limits for normal perfusion, objective abnormality thresholds were defined. These tools were expanded for gated SPECT, providing LV perfusion and function parameters in 2D, 3D and 4D. Although a number of techniques have been developed to aid the classification of the SPECT images, the search for robust methods to analyze automatically the complete series of images is yet an open study area. Moreover, the comparison of the different techniques is a difficult task due to the diversity of approaches, image acquisition protocol and the lack of a “gold standard” to compare the results.

The methodology of two distinct approaches to quantify the myocardial motion in SPECT images were also described in deep. The first method adopted the gradient based approach to compute OF in 3D. The derivatives estimation, a fundamental step in any OF implementation, was improved by a multi-resolution technique. In the second method a spatio-temporal frequency based approach was used to OF derivation. The majors motivations for considering this kind of motion representation were: (1) in non-rigid motion, the image frequency components change with time; (2) some investigations on mammalian vision have demonstrated that many neurons in various visual cortical areas of the brain behave as spatio-temporal frequency band-pass filters.

However, an efficient method to present the myocardial velocity information in order to make it useful for diagnostic purposes is still an open question. In this Chapter, we have reviewed some visualization strategies for displaying myocardial motion and parameters. We presented a color coded visualization scheme for displaying the wall movements that comprises the separation of the global wall motion into three individual directions, which are displayed using a compression color coding that integrates orientation and intensity of the velocity vector at each voxel. The velocity coded information of the volumes can be both presented as 2D slices or polar map. Moreover, a novel scheme with the goal of creating an easily understandable visualization that preserves the temporal and regional information of any temporal parameter distribution was also presented. The time varying parameter is displayed in one polar map presenting the medium or maximum parameter value extracted from all time frames of any temporal parameter distribution combined with a graphical representation of the mean parameter value over time. The visualization goal is to provide information about the regional and temporal myocardial condition, which may prove viable for use within a clinical setting.

Acknowledgements

This work was supported in part by National Council for Scientific and Technological Development (CNPq) and Zerbini Foundation.

References

- [1] Abd-Elmoniem KZ, Stuber M, Prince JL. Direct three-dimensional myocardial strain tensor quantification and tracking using zHARP. *Medical Image Analysis*, **12**: 778–786, 2008.
- [2] Adelson E.H., Bergen J.R.: Spatiotemporal energy models for the perception of motion, *J Opt Soc Am A*, **2**(2): 284-299, 1985.
- [3] Amartur S.C., Vessele H.J.: A new approach to study cardiac motion: The optical flow of cine MR images, *Magn Res Med*, **29**(1): 59-67, 1993.
- [4] Amini A.A., Duncan J.S.: Bending and stretching models for LV wall motion analysis from curves and surfaces, *Imag Vis Comp*, **10**(6): 418-430, 1992.
- [5] Axel L., Dougherty L.: Heart wall motion: improved method for spatial modulation of magnetization for MR imaging, *Radiology*, **172**(2): 349-350, 1989.
- [6] Axelsson O.. In: *Iterative solution methods*, Cambridge University Press, Cambridge, 1996.
- [7] Bardinet E., Cohen L.D., Ayache N.: A parametric deformable model to fit unstructured 3D data, *Comp Vis Image Underst*, **71**(1): 39-54, 1998.
- [8] Bardinet E., Cohen L.D., Ayache N.: Tracking and motion analysis of the left ventricle with deformable superquadrics, *Med Image Anal*, **1**(2): 129-150, 1996.
- [9] Barr A.H.: Superquadrics and Angle-Preserving Transformations, *IEEE Comp Graph & Appl*, **1**(1): 11-23, 1981.
- [10] Beauchemin S.S., Barron J.L.: The Computation of Optical Flow, *ACM Comput Surveys*, **27**(3): 427-433, 1995.
- [11] Benayoun S., Ayache N.: Dense Non-Rigid Motion Estimation in Sequences of Medical Images Using Differential Constraints, *Int J Comput Vision*, **26**(1): 25-40, 1998.
- [12] Berger B.C., Watson D.D., Taylor G.J., et al.: Quantitative thallium-201 exercise scintigraphy for detection of coronary artery disease, *J Nucl Med*, **22**: 585-593, 1981.
- [13] Botvinick E.H., Taradash M.R., Shames D.M., Parmley W.W.: Thallium-201 myocardial perfusion scintigraphy for the clinical clarification of normal, abnormal and equivocal stress tests, *Am J Cardiol*, **41**: 43-59, 1978.
- [14] Burow R.D., Pond M., Schafer A.W., Becker L.: "Circumferential profiles": A new method for computer analysis of thallium-201 myocardial perfusion images, *J Nucl Med*, **20**: 771-781, 1979.
- [15] Cauvin J.C., Boire J.Y., Zanca M., Bonny J.M., Maublant J., Veyre A.: 3D modeling in myocardial 201TL SPECT, *Comp Med Im Graph*, **17**(4): 345-350, 1993.
- [16] Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK. Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart. A Statement for Healthcare Professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association, *Circulation*, **105**(4): 539-542, 2002.
- [17] Chapra S.C., Canale R.P. In: *Numerical methods for engineers with programming and software applications*, McGraw-Hill, New York, 1998.
- [18] Chen J, Garcia EV, Folks RD et al., Onset of left ventricular mechanical contraction as determined by phase analysis of ECG-gated myocardial perfusion SPECT imaging: development of a diagnostic tool for assessment of cardiac mechanical dyssynchrony. *Journal of Nuclear Cardiology*, **12**(6): 687–695, 2005.

- [19] Choi H.I., Williams W.J.: Improved time-frequency representation of multicomponent signals using exponential kernels, *IEEE Trans Acous Speech Sig Proc*, **37**(6): 862-871, 1989.
- [20] Choi S-M, Lee D-S, Yoo S-J, Kim M-H. Interactive Visualization of Diagnostic Data from Cardiac Images Using 3D Glyphs. *Medical Data Analysis Lecture Notes in Computer Science*. **2868**: 83-90, 2003.
- [21] Cohen L., 1995, In: *Time-frequency analysis*, Prentice Hall PTR, New Jersey, 1995.
- [22] Debreuve E., Barlaud M., Aubert G., Laurette I., Darcourt J.: Space-time segmentation using level set active contours applied to myocardial gated SPECT, *IEEE Trans Med Imag*, **20**(7):643-659, 2001.
- [23] Declerck J., Feldmar J., Ayache N.: Definition of a four-dimensional continuous planispheric transformation for the tracking and the analysis of left-ventricle motion, *Med Image Anal*, **2**(2): 197-213, 1998.
- [24] Declerck J., Feldmar J., Goris M.L., Betting F.: Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images, *IEEE Trans Med Imag*, **16**(6): 727-737, 1997.
- [25] Denney T.S., Prince J.L.: Frequency domain performance analysis of Horn and Schunck's optical flow algorithm for deformable motion, *IEEE Trans Imag Proc*, **4**(9): 1324-1327, 1995.
- [26] Denney T.S., Prince J.L.: Optimal brightness functions for optical flow estimation of deformable motion, *IEEE Trans Imag Proc*, **3**(2): 178-191, 1994.
- [27] Denney T.S., Prince J.L.: Reconstruction of 3-D left ventricular motion from planar tagged cardiac MR images: an estimation theoretic approach, *IEEE Trans Med Imag*, **14**(4): 625-635, 1995.
- [28] Depasquale E., Nody A., DePuey G., Garcia E.V., Pilcher G.: Quantitative rotational thallium-201 tomography for identifying and localizing coronary artery disease, *Circulation*, **77**: 316-327, 1988.
- [29] Dougherty L., Asmuth J.C., Blom A.S., Axel L., Kumar R.: Validation of an optical flow method for tag displacement estimation, *IEEE Trans Med Imag*, **18**(4): 359-363, 1999.
- [30] Edvardsen T, Gerber BL, Garot J, Bluemke DA, Lima JAC, Smiseth OA. Quantitative Assessment of Intrinsic Regional Myocardial Deformation by Doppler Strain Rate Echocardiography in Humans. *Circulation*, **106**: 50-56, 2002.
- [31] Eusemann CD, Ritman EL, Bellemann ME, Robb RA. Parametric display of myocardial function. *Computerized Medical Imaging and Graphics*, **25**: 485-493, 2001.
- [32] Faber T.L., Cooke C.D., Folks R.D., Vansant J.P., Nichols K.J., DePuey E.G., Pettigrew R.I., Garcia E.V.: Left ventricular function and perfusion images: an integrated method, *J Nucl Med*, **40**: 650-659, 1999.
- [33] Faber T.L., Stokely E.M., Peshock R.M., Corbett J.R.: A Model-Based Four-Dimensional Left Ventricular Surface Detector, *IEEE Trans Med Imag*, **10**: 321-329, 1991.
- [34] Fenema C., Thompson W.: Velocity determination in scenes containing several moving images, *Comp Vis Graph & Im Proc*, **9**: 301-315, 1979.
- [35] Ficaro E.P., Corbett J.R.: Advances in quantitative perfusion SPECT imaging, *J Nucl Cardiol*, **11**(1): 62-70, 2004.
- [36] Fischer S.E., McKinnon G.C., Scheidegger M.B., Prins W., Meier D., Boesiger P.: True myocardial motion tracking, *Magn Res Med*, **31**(4): 401-413, 1994.
- [37] Fitzpatrick J.M.: The existence of geometrical density-image transformations corresponding to object motion, *Comp Vis Graph & Im Proc*, **44**: 155-174, 1988.
- [38] Fll D, Jung B, Staehle F, Schilli E, Bode C, Hennig J, Markl M. Visualization of Multidirectional Regional Left Ventricular Dynamics by High-Temporal-Resolution Tissue Phase Mapping. *Journal of Magnetic Resonance Imaging*, **29**: 1043-1052, 2009.
- [39] Frangi A.F., Niessen W.J., Viergever M.A.: Three-Dimensional Modeling for Functional Analysis of Cardiac Images: A Review, *IEEE Trans Med Imag*, **20**(1): 2-5, 2001.
- [40] Friboulet D., Magnin I.E., Mathieu C., Pommert A., Hoehne K.H.: Assessment and visualization of the curvature of the leftventricle from 3D medical images, *Comp Med Im Graph*, **17**(4): 257-262, 1993.

- [41] Fuchs R, Hauser H. Visualization of Multi-Variate Scientific Data. *Computer Graphics Forum*, **28**(6): 1670–1690, 2009.
- [42] Gafni H., Zeevi Y.Y.: A model for processing of movement in the visual system, *Biol Cybern*, **32**(3): 165-173, 1979.
- [43] Gafni H., Zeevi Y.Y.: A Model for Separation of Spatial and Temporal Information, *Biol Cybern*, **28**(2): 73-82, 1977.
- [44] Garcia D, del Iamo JC, Tann D, Yotti R, Cortina C, Bertrand , Antoranz JC, Prez-David E, Rieu R, Fernandez-Aviles F, Bermejo J. Two-Dimensional Intraventricular Flow Mapping by Digital Processing Conventional Color-Doppler Echocardiography Images, *IEEE Transactions on Medical Imaging*, **29**(10): 1701-1713, 2010.
- [45] Garcia E.V., Cooke C.D., Folks R.D., Santana C.A., Krawczynska E.G., De Braal L., Ezquerria N.F.: Diagnostic performance of an expert system for the interpretation of myocardial perfusion SPECT studies, *J Nucl Med*, **42**(8): 1185-1191, 2001.
- [46] Garcia E.V., Cooke C.D., Van Train K.F., Folks R.D., Peifer J.W.: Technical aspects of myocardial SPECT imaging with technetium- 99m sestamibi, *Am J Cardiol*, **66**: 23-31, 1990.
- [47] Garcia E.V., Faber T.L., Galt J.R., et al.: Advances in nuclear emission PET and SPECT imaging. *IEEE Eng Med Bio*, **5**: 21-33, 2000.
- [48] Garcia E.V., Maddahi J., Berman D.S., Waxman A.: Space-time quantitation of 201 Tl myocardial scintigraphy. *J Nucl Med*, **22**: 309-317, 1981.
- [49] Garcia E.V., Van Train K. F., Maddahi J., Prigent F., Fredman J., Areeda J., Waxman A. and Berman D.S.: Quantification of rotational thallium-201 myocardial tomography, *J Nucl Med*, **26**: 17-26, 1985.
- [50] Germano G, Kavanagh PB, Slomka PJ, Van Kriekinge SD, Pollard G, Berman DS. Quantitation in gated perfusion SPECT imaging: The Cedars-Sinai approach. *J Nucl Cardiol*, **14**: 433-54, 2007.
- [51] Germano G., Berman D.S.: Acquisition and processing for gated perfusion SPECT: Technical aspects. In *Clinical gated cardiac SPECT*, Futura Publishing Co, Armonk, NY, 1999.
- [52] Germano G., Erel J., Kiat H., Kavanagh P.B., Berman D.S.: Quantitative LVEF and qualitative regional function from gated thallium-201 perfusion SPECT, *J Nucl Med*, **38**(5): 749-754, 1997.
- [53] Germano G., Kavanagh P.B., Chen J., Berman D.S.: An automatic approach to the analysis, quantitation and review of perfusion and function from myocardial perfusion SPECT images, *Int J Card Imag*, **13**(4): 337-346, 1997.
- [54] Germano G., Kavanagh P.B., Chen J., Waechter P., Su H-T., Kiat H., Berman D.S.: Operatorless processing of myocardial perfusion SPECT studies, *J Nucl Med*, **36**(6): 2127-2132, 1995.
- [55] Germano G., Kiat H., Kavanagh P.B., Moriel M., Mazzanti M., et al.: Automatic quantification of ejection fraction from gated myocardial perfusion SPECT, *J Nucl Med*, **36**(11): 2138-2147, 1995.
- [56] Gorce J.M., Friboulet D., Magnin I.E.: Estimation of three-dimensional cardiac velocity fields: assessment of a differential method and application to three-dimensional CT data, *Med Imag Anal*, **1**(3): 245-261, 1996.
- [57] Gutierrez M.A., Furuie S.S., Moura L.A., Melo C.P., Meneghetti J.C., Alens N.: Estimation of myocardial kinetic energy in Gated SPECT images, *Comput. Cardiol.*, **22**: 75-78, 1995.
- [58] Gutierrez M.A., Rebelo M.S., Furuie S.S., Meneghetti J.C.: Automatic quantification of three-dimensional kinetic energy in gated myocardial perfusion single-photon-emission computerized tomography improved by a multiresolution technique, *J Elec Imag*, **12**: 118-123, 2003.
- [59] Hadamard J.: Sur les problemes aux drives partielles et leur signification physique, *Princeton University Bulletin*, **13**: 49-52, 1902.
- [60] Hamilton G.W., Trobaugh G.B., Ritchie J.L., Williams D.L., Weaver W.D., Gould K.L.: Myocardial imaging with intravenously injected thallium-201 in patients with suspected coronary artery disease, *Am J Cardiol*, **39**: 347-354, 1977.

- [61] Hennig J, Markl M, Schneider B, Peschl. Regional myocardial function with tissue phase mapping. *Magnetic Resonance Materials in Physics, Biology and Medicine*, **6**: 145-146, 1998.
- [62] Herman G.T.. In: *Medical imaging systems techniques and applications: computational techniques*, (C.T. Leondes, Ed.), Gordon and Breach Science Publishers, Amsterdam, 1998.
- [63] Horn B.K.P., Schunck B.G.: Determining optical flow, *Artificial Intelligence*, **17**: 185-203, 1981.
- [64] Jacobson L.D., Wechsler H.: Derivation of optical flow using a spatiotemporal-Frequency approach, *Comp Vis Graph Im Proc*, **38**(1): 29-65, 1987.
- [65] Jacobson L.D., Wechsler H.: Joint spatial/spatial-frequency representation, *Signal Proc*, **14**(1): 37-68, 1988.
- [66] Jung B, Markl M, Fll D, Hennig J. Investigating myocardial motion by MRI using tissue phase mapping. *European Journal of Cardio-thoracic Surgery*, **29S**: S150—S157, 2006.
- [67] Kambhamettu C., Goldgof D.B.: Curvature-Based Approach to Point Correspondence Recovery in Conformal Non-Rigid Motion, *Comp Vis Graph Im Proc*, **60**(1): 26-43, 1994.
- [68] Kerwin W.S., Prince J.L.: Cardiac material markers from tagged MR images, *Med Imag Anal*, **2**(4): 339-353, 1998.
- [69] Klein G.J., Reutter B.W., Heusman R.H.: Non-rigid summing of gated PET via optical flow, *IEEE Trans Nucl Sci*, **44**(4), 1509-1512, 1997.
- [70] Kleinbaum D.G., Kupper L.L., Muller K.E.. In: *Applied regression analysis and other multivariable methods*, PWS-KENT Publishing Company, Boston, 1988.
- [71] Koenderink J.J.: The structure of images. *Biological Cybernetics*, **50**: 363-370, 1984.
- [72] Kumar S., Goldgof D.: Automatic tracking of SPAMM grid and the estimation of deformation parameters from cardiac MR images, *IEEE Trans Med Imag*, **13**(1): 122-132, 1994.
- [73] Laplante P.A., Stoyenko A.D.. In *Real-time imaging: theory, techniques and applications*, IEEE Press, New York, 1996.
- [74] Laramée RS, Hauser H, Doleisch H, Vrolijk B, Post FH, Weiskopf D. The State of the Art in Flow Visualization: Dense and Texture-Based Techniques. *Computer Graphics Forum*, **23**(2): 203–221, 2004.
- [75] Lindeberg T.: Discrete Derivative Approximations with Scale-Space Properties: A Basis for Low-Level Feature Extraction, *J Math Imag Vis*, **3**(4): 349-376, 1993.
- [76] Liu YH, Sinusa A.J, Khaimov D, Gebuza BI, Wackers F. New hybrid count- and geometry-based method for quantification of left ventricular volumes and ejection fraction from ECG-gated SPECT: Methodology and validation. *Journal of Nuclear Cardiology*, **12**(1): 55–65, 2005.
- [77] Maddahi J., Garcia E.V., Berman D.S., Waxman A., Swan H.J.C., Forrester J.: Improved noninvasive assessment of coronary artery disease by quantitative analysis of regional stress myocardial distribution and washout of thallium-201, *Circulation*, **64**(5): 924-935, 1981.
- [78] Max N. Progress in scientific visualization. *Visual Computing*, **21**: 979–984, 2005.
- [79] Meade R.C., Bamrah V.S., Horgan J.D., Ruetz P.P., Kronenwetter C., Yeh E.L.: Quantitative methods in the evaluation of thallium-201 perfusion images, *J Nucl Med*, **19**: 1175-1178, 1978.
- [80] Merrit CRB. Doppler Color Flow Imaging. *J Clin Ultrasound*, **15**: 591-597, 1987.
- [81] Meyer F.G., Constable R.T., Sinusas A.J., Duncan J.S.: Tracking Myocardial Deformation Using Phase Contrast MR Velocity Fields: A Stochastic Approach, *IEEE Trans Med Imag*, **15**(4): 453-465, 1996.
- [82] Meyering W.I., Gutierrez M.A., Robilotta C.C., Furuie S.S.: Determining velocity displacement field from cardiac image sequence, *SPIE Medical Imaging*, **4683**: 176-185, 2002.
- [83] Meyer-Spradow J, Stegger L, Dring C, Ropinski T, Hinrichs K. Glyph-Based SPECT Visualization for the Diagnosis of Coronary Artery Disease. *IEEE Transactions on Visualization and Computer Graphics*, **14**(6): 1499-1506, 2008.
- [84] Mishra S.K., Goldgof D.B., Huang T.S., Kambhamettu C.: Curvature-Based Non-Rigid Motion Analysis from 3D Point Correspondences, *Int J Image Syst Tech*, **4**: 214-225, 1992.

- [85] Miller TM. Human Factors in Visualization Research. *IEEE Transactions on Visualization and Computer Graphics*, **10**(1): 72-84, 2004.
- [86] Monga O., Deriche R., Rocchisani J-M.: 3D edge detection using recursive filtering: Application to scanner images, *Comp Vis Graph & Im Proc: Imag Underst*, **53**(1): 76-87, 1991.
- [87] Mor-Avi V, Lang RM, Badano LP, Belohlavek M, Cardim NM, Derumeaux G, Galderisi M, Marwick T, Nagueh SF, Sengupta PP, Sicari R, Smiseth OA, Smulevitz B, Takeuchi M, Thomas JD, Vannan M, Voigt J-U, Zamorano JL. Current and Evolving Echocardiographic Techniques for the Quantitative Evaluation of Cardiac Mechanics: ASE/EAE Consensus Statement on Methodology and Indications. *Journal of American Society of Echocardiography*, **24**: 277-313, 2011.
- [88] Oeltze S, Grothues F, Hennemuth A., Ku A., Preim B, Integrated Visualization of Morphologic and Perfusion Data for the Analysis of Coronary Artery Disease. *Proc. Eurographics/IEEE VGTC Symp. Visualization '06*, 131-138, 2006.
- [89] Oeltze S, Hennemuth A., Glaer S, Khnel C, Preim B. Glyph-Based Visualization of Myocardial Perfusion Data and Enhancement with Contractility and Viability Information. *Eurographics Workshop on Visual Computing for Biomedicine*, 2008.
- [90] Osman NF, Prince JL. Visualizing myocardial function using HARP MRI. *Phys. Med. Biol.*, **45**: 1665-1682, 2000.
- [91] Ott R.J., Flower M.A., Babich J.W., Marsden P.K.: The physics of radioisotope imaging. In: Webb S (ed.) *The Physics of Medical Imaging*. Bristol, Chap., **6**: 214-226, 1988.
- [92] Perk G, Tunick PA, Kronzon I. Non-Doppler Two-dimensional Strain Imaging by Echocardiography-From Technical Considerations to Clinical Applications. *Journal of the American Society of Echocardiography*, **20**(3): 234-243, 2007.
- [93] Petretta M, Cuocolo R, Acampa W, Cuocolo A.. Quantification of Myocardial Perfusion: SPECT. *Cardiac Nuclear Imaging* (RJ Gropler, Section Editor), Published online, DOI: 10.1007/s12410-012-9131-y, 2012.
- [94] Preim B, Oeltze S, Mlejnek M, Groller E, Hennemuth A., Behrens S. Survey of the Visual Exploration and Analysis of Perfusion Data. *IEEE Transactions on Visualization and Computer Graphics*, **15**(2): 205-220, 2009.
- [95] Prince J.L., E.R. McVeigh: Motion estimation from tagged MR image sequences, *IEEE Trans Med Imag*, **11**(2): 238-249, 1992.
- [96] Rebelo MS, Aarre AKH, Clemmesen K-L, Brando SCS, Giorgi MC, Meneghetti JC, Gutierrez MA. Determination of Three-Dimensional Left Ventricle Motion to Analyze Ventricular Dyssynchrony in SPECT Images. *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 290695, 2010.
- [97] Rioual K, Unanua E, Laguitton S. et al. MSCT labelling for pre-operative planning in cardiac resynchronization therapy. *Computerized Medical Imaging and Graphics*, **29**(6): 431-439, 2005.
- [98] Ritchie J.L., Trobaugh G.B., Hamilton G.W., Gould K.L., Narahara KA, Murray J.A., Williams D.L.: Myocardial imaging with thallium-201 at rest and during exercise: Comparison with coronary angiography and resting and stress electrocardiography, *Circulation*, **56**: 66-71, 1977.
- [99] Ropinski T, Oeltze S, Preim B. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers & Graphics* **35**(2): 392-401, 2011.
- [100] Ropinski T, Specht M, Meyer-Spradow J, Hinrichs K, Preim B.. Surface Glyphs for Visualizing Multimodal Volume Data. In: *Proceedings of Vision, Modelling and Visualization*, November 7th- 9th, Saarbrücken, Germany, 2007.
- [101] Rutz AK, Ryf S, Plein S, Boesiger P, Kozerk S. Accelerated Whole-Heart 3D CSPAMM for Myocardial Motion Quantification. *Magnetic Resonance in Medicine*, **59**: 755-763, 2008.
- [102] Sagan H.. In: *Introduction to the calculus of variations*, Dover Publications, New York, 1969.
- [103] Santos A., Ledesma-Carbayo MJ. *Cardiac Imaging*. Wiley Encyclopedia of Biomedical Engineering, 2006.

- [104] Sederberg T.W., Parry S.R.: Free-Form Deformation of Solid Geometric Models, SIGGRAPH'86, **20**: 151-160, 1986.
- [105] Segars, W. P., Sturgeon, G., Mendonca, S., Grimes, J., & Tsui, B. M. W.. 4d xcat phantom for multimodality imaging research. *Med Phys*, **37**(9): 4902-4915, 2010.
- [106] Shen MYH, Liu YH, Sinusas A.J, Bruni W, Drozhinin O, Zaret BL, Wackers FJ. Quantification of regional myocardial wall thickening on electrocardiogram-gated SPECT imaging. *Journal of Nuclear Cardiology*, **6**(6): 583–595, 1999.
- [107] Song S.M., Leahy R.M., Boyd D.P., Brundage B.H., Napel S.: Determining cardiac velocity fields and intraventricular pressure distribution from a sequence of ultrafast CT cardiac images, *IEEE Trans Med Imag*, **14**(2): 386-397, 1994.
- [108] Song S.M., Leahy R.M.: Computation of 3-D velocity fields from 3-D cine CT images of a human heart, *IEEE Trans Med Imag*, **10**(3): 295-306, 1991.
- [109] Tamaki N., Yonekura Y., Mukai T., Kodama S., Kadota K., Kambara H., Kawai C., Torizuka K.: Stress thallium-201 transaxial emission computed tomography: quantitative versus qualitative analysis for evaluation of coronary artery disease, *J Am Coll Cardiol*, **4**: 1213-1221, 1984.
- [110] Torrent-Gausp F. In: *The cardiac muscle*, Fundacion Juan March, Madrid, 1942.
- [111] Tory M, Miller T. Human Factors in Visualization Research. *IEEE Transactions on Visualization and Computer Graphics*, **10**(1): 72-84, 2004.
- [112] Trobaugh G.B., Wackers F.J., Sokole E.B., DeRouen T.A., Ritchie J.L., Hamilton G.W.: Thallium-201 myocardial imaging: an interinstitutional study of observer variability, *J Nucl Med*, **19**: 359-363, 1978.
- [113] Tsao T-F, Wu Y-L, Yu J-M, Kang R-J, Tseng Y-H, H-H Huang HH, S-W Hung SW, M-K Gueng MK, Y-C Lin YC, Y-S Tyan YS, C-H Su CH. Color Doppler Twinkling Artifact of Calcified Cardiac Valves in Vitro: A Not Well Known Phenomenon in Echocardiography. *Ultrasound in Med. & Biol.*, **37**(3): 386–392, 2011.
- [114] Turan S, Turan OM, TyY-Torredes K, Harman CR, Baschat A.A.. Standardization of the first-trimester fetal cardiac examination using spatiotemporal image correlation with tomographic ultrasound and color Doppler imaging, *Ultrasound Obstet Gynecol*, **33**(6): 652–656, 2009.
- [115] Van Train K.F., Berman D.S., Garcia E.V., Berger H.J., Sands M.J., Friedman J.D., Freeman M.R., Pryzlak M., Ashburn W.L., Norris S.L., Green A.M., Maddahi J.: Quantitative analysis of stress thallium-201 myocardial scintigrams: A multicenter trial, *J Nucl Med*, **27**: 17-25, 1986.
- [116] Watson D.D., Campbell N.P., Read E.K., Gibson R.S., Teates C.D., Beller G.A.: Spatial and temporal quantitation of plane thallium myocardial images, *J Nucl Med*, **22**: 577-584, 1981.
- [117] Weidemann F, Jamal F, Sutherland GR, Claus P, Kowalski M, Hatle L, De Scheerder I, Bijnens B., Rademakers FE. Myocardial function defined by strain rate and strain during alterations in inotropic states and heart rate. *Am J Physiol Heart Circ Physiol*, **283**: H792–H799, 2002.
- [118] Wen H, Bennett E, Epstein N, Plehn J. Magnetic Resonance Imaging Assessment of Myocardial Elastic Modulus and Viscosity Using Displacement Imaging and Phase-Contrast Velocity Mapping. *Magnetic Resonance in Medicine*, **54**: 538–548, 2005.
- [119] Wnsche BC, Lobb R. The Visualization of Myocardial Strain for the Improved Analysis of Cardiac Mechanics. *Proceeding GRAPHITE '04 Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, 2004.
- [120] Xu C, Pilla JJ, Isaac G, Gorman JH, Blom A.S, Gorman RC, Ling Z, Dougherty L. Deformation analysis of 3D tagged cardiac images using an optical flow method. *Journal of Cardiovascular Magnetic Resonance*, **12**(19), 2010.
- [121] Young A.A., Axel L.: Three-dimensional motion and deformation of the heart wall: estimation with spatial modulation of magnetization—a model-based approach, *Radiology*, **185**: 241-247, 1992.

- [122] Young A.A., Kraitchman D.L., Dougherty L., Axel L.: Tracking and finite element analysis of stripe deformation in magnetic resonance tagging, *IEEE Trans Med Imag*, **14**(3): 413-421, 1995.

Chapter 7

MRI volumetric measurements of Amygdala and Hippocampus: evaluation of automatic image segmentation methods

Flávio L. Seixas, Andrea S. de Souza, Fernanda Tovar-Moll, Aura Conci, and Débora C. Muchaluat-Saade

Abstract Alzheimer’s disease (AD) is the most common cause of dementia among elderly. The hippocampus and the amygdala atrophy detected by Magnetic Resonance Image (MRI) volume evaluation is a sensitive feature of early AD diagnosis. This chapter discusses automatic image segmentation methods, comparing three methods available as public-domain medical image analyzing libraries: Freesurfer, IBASPM and FSL. The objective is to verify if there is a consensus on hippocampus and amygdala volumes obtained automatically by those image analysis packages. The MRI set used in this work is composed by exams from different normal controls individuals. The comparison evaluation among automatic segmentation methods is performed using the volume overlay index. The results showed a low agreement rate considering the three automatic image segmentation methods. FSL and Freesurfer volumes were more similar to each other considering the volume overlay index. Fi-

Flávio Luiz Seixas

UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: fseixas@ic.uff.br

Andrea Silveira de Souza

UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, LABS D’Or Network, Radiology Department, Rua Diniz Cordeiro, 39, 22281-100, Rio de Janeiro, RJ, Brazil, e-mail: andrea.silveiradesouza@gmail.com

Fernanda Tovar-Moll

LABS D’Or Network, Radiology Department, Rua Diniz Cordeiro, 39, 22281-100, Rio de Janeiro, RJ, Brazil

Aura Conci

UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: aconci@ic.uff.br

Débora C. Muchaluat-Saade

MidiaCom Lab/UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: debora@ic.uff.br

nally, segmentation results obtained by the three automatic segmentation methods were qualitatively compared to the correspondent structure anatomical definition.

7.1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative condition associated with pathological accumulation of amyloid plaques and neurofibrillary tangles in the brain, first affecting memory systems, progressing to language, and executive dysfunction and behavior compromise. AD is the most common cause of dementia among elderly, ranging from 49.9% to 84.5% in Latin America [20]. Dementia prevalence reached 6.8% in a community of São Paulo city, Brazil, considering subjects aged 60 years old and over, mirroring the rates of developed countries [3]. Dementia is becoming a major public health problem where the population is progressively growing older, according to last reports published by the Brazilian Institute of Geography and Statistics (IBGE). In 2020, subjects aged 60 years old and over will be beyond 32 million, representing more than 13.8% of the Brazilian population.

The hippocampus and the amygdala are among the first structures affected in AD [4]. Hippocampus volumetric measurements using magnetic resonance images (MRI) provide one useful biomarker for AD [13, 15, 23, 9]. Studies have proposed volumetric measurements of the hippocampus to differentiate normal aging from AD, and from mild cognitive impairment (MCI) [23]. However, the hippocampus volumetric measurement mostly relies on highly time-consuming manual segmentation, which is a partial subjective process and is not feasible in clinical routine. An expert may require 30 minutes to trace a single structure such as the hippocampus [10]. Fortunately, automatic methods using computational algorithms have provided consistent results comparable to manual segmentation [16].

Powell *et al.* [24] compared four fully automatic brain structure segmentation methods: (1) template-based, (2) probabilistic-based, (3) artificial neural network-based (ANN) and (4) support vector machine-based (SVM) segmentation. The image base was composed by 25 normal control subjects. Segmentation results were compared to manually defined regions from the sample. The comparison evaluation between automatic and manual segmentation was performed using three overlapping metrics: relative, similarity and spatial. The relative overlap for the hippocampus was lower than 59%. The comparison evaluation among automatic segmentation was performed using t-test p-values of relative overlap measured among the segmentation methods. The results showed a little disparity between ANN and SVM. In contrast to that, this chapter evaluates the relative overlap volumes considering just automatic methods, disregarding manual segmentation volumes.

Carmichael *et al.* [6] compared the atlas-based segmentation method performance with four different registration modules: (1) AIR, (2) SPM, (3) FLIRT and (4) deformable method proposed by Chen [7]. The atlas-based segmentation method first co-registers a subject image and a reference image called atlas image on which structures of interest, including the hippocampus, have been manually traced by

experts. The result was a transformation model aligning both images. The inverse transformation was then used to map the voxel label from the atlas image into the subject image space. A total of 58 subject images were collected, including 2 with AD and 2 with MCI. The left and right hippocampi were traced on all subject images. Manual and automatic segmentation using three atlases were compared: (1) MNI, (2) Harvard and (3) cohort atlas. Cohort atlas images were randomly selected from the subject population. The results were evaluated in terms of overlap ratio. Chen fully deformable registration method produced the best overlap ratio. Also, the overlap ratio for cohort-atlas-based and standard-atlas-based segmentation was about 0.5, concluding that the tracing protocols from the atlas institution did not coincide with those of the investigators.

Morey *et al.* [19] compared the performance of two popular and fully automatic segmentation tools: (1) FSL/FIRST and (2) Freesurfer. The hippocampus and amygdala volumes derived from each automatic segmentation measurement were compared to expert hand tracing for percent volume overlap, also called Dice's coefficient, and across-sample correlation. A total of 20 normal control subjects were collected. When comparing automatic segmentation measures to manual tracing, the left-right hippocampus segmentation volume overlap produced by Freesurfer (83% and 82% respectively), was superior to FIRST (79% and 80% respectively). In contrast to that, when considering the left-right amygdala, FIRST produced greater volume overlap than Freesurfer. The conclusions were: (1) the automatic methods systematically generated larger volumes than manual tracing, (2) Freesurfer and FIRST are not equal when compared to manual tracing and (3) Freesurfer was superior for segmenting the hippocampus by all the objective measurements performed and FIRST was superior for segmenting the amygdala.

Tae *et al.* [26] used T1-weighted MRI of 21 patients with chronic major depressive disorder (MDD) and 20 normal controls. The objective was to validate the usefulness of the volumetric methods, comparing the hippocampus volumes got from two fully automatic segmentation tools to manual segmentations. The automatic segmentation methods were: (1) surface-based parcellation using Freesurfer and (2) individual-atlas volumetry using IBASPM. The intra-class correlation coefficient (ICC) was used for comparison evaluation. The test of agreement between the volumetric methods got an ICC right 0.84 and left 0.85 when comparing Freesurfer to manual, and ICC right 0.65 and left 0.71 when comparing IBASPM to manual. The conclusions were: (1) the automatic hippocampal volumetric segmentation methods showed good agreement with manual hippocampal volumetry, (2) the absolute volume measured using Freesurfer was 35% larger than manual, (3) the agreement was questionable with IBASPM. In contrast to this chapter, that paper did not compare both automatic segmentation methods: it compared each one with manual segmentation classified as gold-standard.

Those cited works compared automatic segmentation methods with manual segmentation. On the other hand, this chapter discusses fully automatic segmentation methods and compares the obtained results. Three fully automatic methods available as public-domain image packages are compared: FIRST/FSL, IBASPM and Freesurfer. FIRST/FSL (www.fmrib.ox.ac.uk/fsl) is a model-based seg-

mentation and registration tool applied to get subcortical brain segmentation using Bayesian shape and appearance models, developed by FMRIB group from Oxford University [21]. IBASPM (www.thomaskoenig.ch/Lester/ibaspm.htm) is a toolbox for brain segmentation of structural MRI, developed by Cuban-Neuroscience Center [1]. Freesurfer (surfer.nmr.mgh.harvard.edu/) is a set of software tools for the study of cortical and subcortical anatomy developed by members of Athinoula A. Martinos Center for Biomedical Imaging [11].

The main goals of this chapter are: (1) presenting a synthesized medical image analysis pipeline for each of the three methods previously cited; (2) comparing the hippocampus and amygdala volumes obtained by each method, quantifying the volume agreement among them; (3) discussing the quality of the volumetric measurements obtained with each method.

This rest of this chapter is organized in four sections. Section 7.2 explains magnetic resonance imaging concepts and presents the amygdala and hippocampus anatomical definitions, describing the image data set used for comparison among automatic segmentation methods. Section 7.3 presents the automatic segmentation methods to be Freesurfer, FIRST/FSL and IBASPM, synthesizing their image analysis pipelines. Section 7.4 discusses practical experiment results and comparison among the three methods. Section 7.5 presents the final remarks.

7.2 MRI concepts and Amygdala and Hippocampus anatomical definitions

This section is organized into three subsections. Section 7.2.1 presents magnetic resonance image orientation definitions. Section 7.2.2 presents the amygdala and hippocampus anatomical definition. Section 7.2.3 describes the images used for comparing the segmentation methods.

7.2.1 Magnetic resonance image orientation

Magnetic resonance scanner produces tridimensional images based on a matrix of numbers representing the physical property of the tissue voxel. A voxel is a volume element, representing a value on a regular grid in three dimensions. The numerical value assigned to each voxel is proportional to the intensity of the radio wave signal emanating from the tissue in which nuclei (hydrogen) have been perturbed by a characteristic radio frequency energy input. Some advantages of magnetic resonance imaging (MRI) are: (1) no exposure to ionizing radiation, (2) non-invasive clinical technique, (3) superior contrast resolution and (4) superior tissue characterization [25].

With the emergence of MRI studies of the human brain, there have been numerous structural [14] and functional [17] studies of the human hippocampal forma-

tion. Studies have used both manual tracing techniques as well as algorithms for automatic segmentation. Recently development in MRI data acquisition technology offers higher resolution and better signal-to-noise ratio, allowing higher differentiation of hippocampal formation components.

Medical scanners create regular rectangular arrays of voxels. The order of arrays and geometric location of each voxel depends on the coordinate system adopted. In medical images, the terms that represent their axes are summarized in Figure 7.1a: (1) right-left (R-L), (2) anterior-posterior (A-P) and (3) inferior-superior (I-S). Regarding positive axes directions, there are two conventions. (1) In neurological convention, or right-hand coordinate system (RAS), the images are viewed as though looking from the top of the head downward (Figure 7.1b). So, patient-right appears on screen-right. (2) In radiological convention for axes (LAS), or left-hand coordinate system (LAS), the images are viewed as though looking upward from the feet of the subject (Figure 7.1c). So, patient-right appears on screen-left.

The most common axes system in computational neurology is the right-hand coordinate system (RAS). The abbreviation RAS means that voxels are ordered from right to left and grouped in a row. Rows are ordered from anterior to posterior and grouped in a slice. Slices are stored from superior to inferior and grouped in a volume. There are three possible slice planes: (1) axial or transverse (R-L x A-P planes), (2) coronal (R-L x S-I planes) and (3) sagittal (A-P x S-I planes).

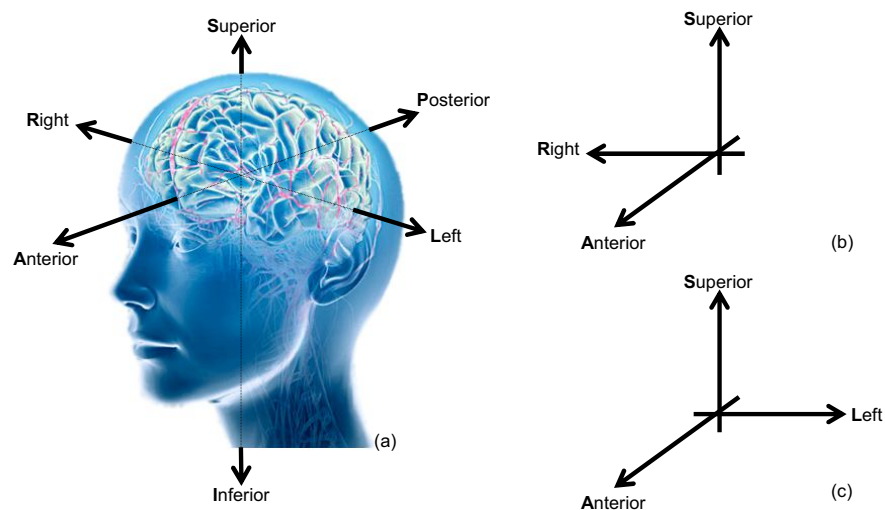


Fig. 7.1 (a) Medical image axes terms. Correspondent axes directions: (b) neurological convention or right-hand coordinate system; (c) radiology convention or left hand coordinate system.

7.2.2 *Amygdala and Hippocampus anatomical definitions*

Hippocampus is a curved structure on the medial aspect of the temporal lobe that bulges into the floor of the temporal horn of the lateral ventricle. The gray matter of the hippocampus is an extension of the subiculum of the parahippocampal gyrus. The hippocampus itself consists of two interlocking C-shaped structures: the cornu of Ammonis and the dentate gyrus [2]. Amygdala is a large complex of gray nuclei, medial to uncus, just in front of the temporal horn of the lateral ventricle. Amygdala is separated from hippocampus by uncus recess of the temporal horn or alveus of hippocampus [29].

Table 11.1 summarizes the hippocampus and amygdala boundaries according to anatomical definitions.

7.2.3 *Image dataset description*

The Magnetic Resonance (MR) images used in this work were provided by LABS DOr Network Research Group, Rio de Janeiro, Brazil. There were nineteen exams from distinct cognitively intact patients. The images were grouped in two classes: Group I was scanned in a 1.5 Tesla MR unit, using images with 256x256 resolution and a total of 160 slices; Group II was scanned in a 3 Tesla MR unit, using images with 512x512 resolution and a total of 360 slices (both from Philips, The Netherlands). Volumetric T1-weighted images were acquired in the sagittal plane. The subjects were aged from 21 to 57 years old, with 5 men and 14 women, as summarized in Table 11.3.

The next section compares the image analysis pipelines from the three image segmentation packages considered. High-level image processing concepts are defined and assigned to each of the image analyzing stages from each package.

7.3 Medical image analysis methods

The three methods here compared are well-known public-domain medical image analyzing tools, applied to segment anatomical brain structures automatically, using tridimensional image acquisition by magnetic resonance scanners. The image analysis pipeline for each medical image analyzing package is synthesized below. In order to facilitate comparison, we defined high-level image analysis processing stages and assigned it to each pipeline stage from a particular package. Figure 7.2 shows graphically the image analysis pipeline for each studied package, Freesurfer, FIRST/FSL and IBASPM.

- (1) *Preprocessing*. Preprocessing involves preparing the images for feature selection and correspondence. Thus, image enhancements are performed, such as

Table 7.1 Amygdala and Hippocampus anatomical boundaries.

Boundaries	Hippocampus	Amygdala
Lateral	Hippocampus is bordered laterally by the temporal horn of the lateral ventricle and medially by the ambient cistern.	The lateral border of the amygdala is marked by the thick, central white matter tract of the temporal lobe.
Anterior-Posterior	Hippocampus most anterior part (hippocampus head) is recognized by digitations on its superior surface. The hippocampus loses digitations in its posterior part (hippocampus body). The most posterior part (hippocampus tail) is the narrowest portion. It narrows the curves around the splenium to form indusium griseum above corpus callosum.	Amygdala lies anterior and superior to the hippocampus, at the medial aspect of the temporal lobe, lateral to the uncus. The transition slice is defined as the slice where both the amygdala and the hippocampus are present. In those slices, the amygdala appears as a small diffuse ball, while the hippocampus have crisp borders with white matter and appears as a line of gray that curves superiorly as it moves medially. Sometimes there is a band of white matter or cerebrospinal fluid between those two structures.
Inferior-Superior	Superior border of the hippocampus is marked by white matter. The most anterior portion of the hippocampus is limited by the choroid plexus of the lateral ventricle. Moving posteriorly through slices, the hippocampus moves in a superior direction in the coronal plane. The hippocampus superior border will come in contact with the thalamus. Moving further posteriorly through slices, the formation of the fornix should be viewed. The fornix appears as a white matter tract stemming from the body of the lateral ventricle. The hippocampus superior-lateral borders are marked by the fornix lateral border.	The superior border of the amygdala should be marked by a light white matter tract or more likely, a cerebrospinal-fluid boundary. The superior border is marked by the end of entorhinal sulcus. Many times the superior border of the amygdala comes in close proximity to other gray matter nuclei. The inferior border is the cortical white matter.

Table 7.2 MR image subjects grouped by scanner unit, age and gender.

Scanner Unit	Male		Female	
	Total	Age (years old)	Total	Age (years old)
1.5 Tesla	4	37 to 49 (43.5) ^a	6	46 to 57 (51) ^a
3.0 Tesla	1	28	8	21 to 54 (34.2) ^a

a. average age

image contrast improvement, image noise removal, local field inhomogeneity correction and slight head movement correction. Also, the preprocessing phase can perform image reorientation. The image orientation parameters are typically registered in the image header. However, when using different image file formats, one should make sure the image orientation read by the image application is correct. If not, it is necessary to correct the image orientation, correcting the image header or resampling the whole image.

- (2) *Registration*. Image registration is the process of aligning images so that corresponding features can easily be related [12]. The term is also used to mean aligning images with a computer model or aligning features in an image with locations in physical space or standard coordinate system. In neuroimage, there are two well-known and public-domain sets of standard coordinate systems: (1) Talairach coordinate system [27] and (2) MNI coordinate system [8].
- (3) *Segmentation*. Image segmentation methods are grouped into two classes: (1) in voxel-based methods, the anatomical brain structure is segmented using their voxel features. Some voxel features are used to represent the structure of interest: signal intensity, x, y and z positions, neighborhood-based features using gradient filters, mean filters, standard deviation filters and Haar filters, applying different sizes. The tridimensional positions can be determined using stereotaxic coordinates after spatial normalization to the standard spaces, or using spherical coordinates. (2) In the vertex-based methods, the anatomical structure of interest is represented by its contour. A contour is modeled by a series of connected vertices along the edge of a structure. A deformable contour is an iterative procedure that aims at finding out the contour of the structure of interest, i.e. the vertices take small steps towards the anatomical structure boundary per iteration. Each vertex is driven by the image intensities, typically being attracted to voxels with high intensity gradients. Physical parameters can be used to constraint the deformable contour and improve robustness, such as tension, rigidity and curvature [8, 18].
- (4) *Labeling*. Image labeling is the process of assigning a label to every voxel in an image such that voxels with the same label share certain visual characteristics and belong to the same anatomical structure.

Table 7.3 summarizes the main differences among the three studied medical image analyzing tools.

7.4 Comparison and results

Table 11.4 shows the amygdala and hippocampus average absolute volumes obtained by the three medical image analyzing packages: Freesurfer (FS), IBASPM and FSL. The volumes are grouped as described in Section 7.2.3. Figure 7.3 and Figure 7.4 show the scatter plot considering subject age and brain structure absolute volume as x, y dimensions, respectively.

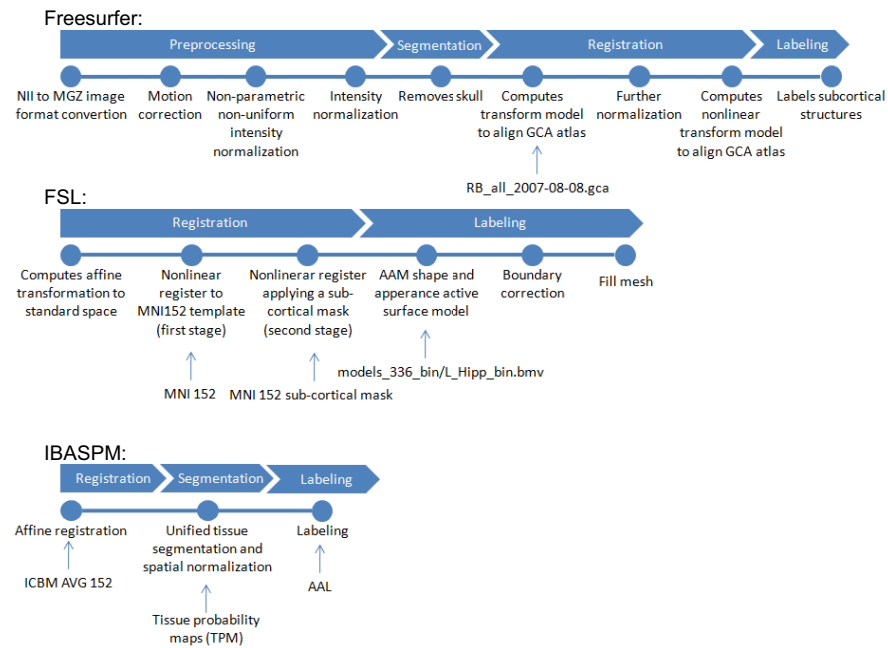


Fig. 7.2 Freesurfer, FSL, and IBASPM image analysis pipelines.

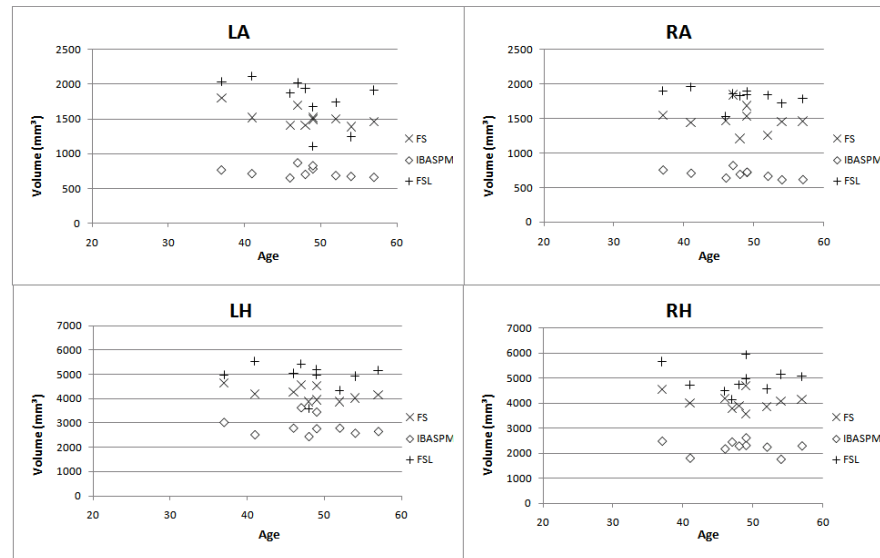


Fig. 7.3 Amygdala and Hippocampus absolute volumes obtained from Group I (1.5 Tesla MR unit).

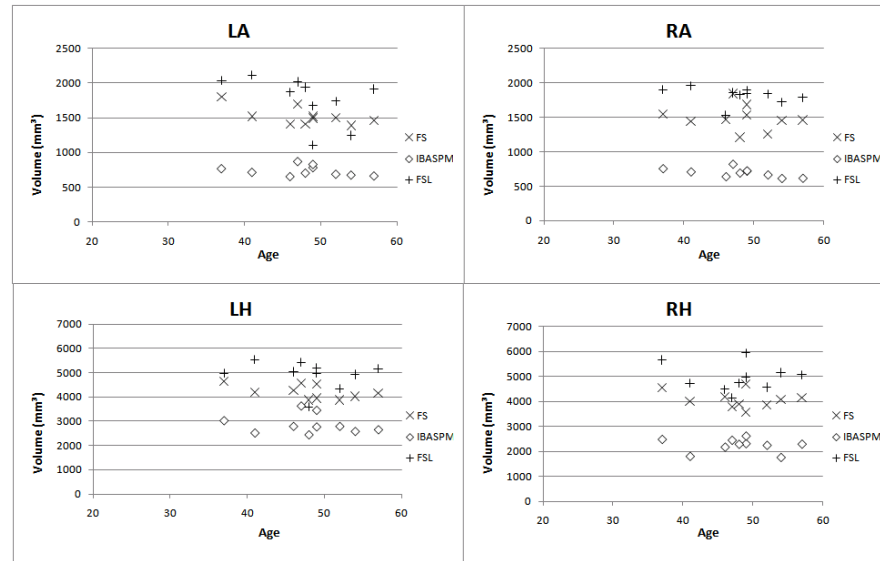
Table 7.3 Comparison among the three studied medical image analyzing tools.

Characteristic	Freesurfer	FSL	IBASPM
Native image file format	MGZ	NII	NII
Preprocessing	Non-parametric non-uniform intensity normalization driven by N3 protocol	Not available	Not available
Image registration	Affine registration driven by 9 degrees of freedom	Performed in two stages. The first makes a standard 12 degrees of freedom registration to the template. The second stage applies a 12 degrees of freedom registration using MNI 152 subcortical mask to exclude voxels outside the subcortical regions	The image registration and segmentation stages are performed in a unified model. The registration is driven by 12 degrees of freedom and a non-linear registration model. The result is a transformation model composed by an affine matrix and non-linear registration model
Template image used in registration stage	Tailarach space	MNI 152 space at 1mm resolution	MNI ICBM AVG 152 smoothing with 8mm Gaussian kernel
Image segmentation	Removes the skull scrap	No image segmentation stage	The input parameters are the brain tissue probability maps published by ICBM (International Consortium of Brain Mapping). It performs an iterative probability function maximization using a Bayesian framework
Image Labeling	Generated by assigning each voxel to the class for which the probability is the greatest	It performs a shape and appearance active surface model (AAM)	It uses the gray matter mask segmented previously. The brain structures presented in the atlas are mapped to the source image using a simple binary association.
Anatomical atlas used in segmentation and labeling stages	GCA atlas (each structure is represented by a class in a probability function)	Shape appearance models constructed from 336 subjects, consisting of children and adults, normal and subjects with pathologies. The result is a vertex set structured in a mesh	AAL atlas (binary atlas).
Brain structure volumes computation	It counts the labeled structures assigned in the previous stage	Volume is computed filling the meshes obtained from the previous stage	The source image is deformed again, applying the inverse transformation model, and labeled structures are mapped to the original space.

Table 7.4 Amygdala and Hippocampus absolute volumes.

Brain structure	Group	Freesurfer (mm ³)	IBASPM (mm ³)	FSL (mm ³)
Left amygdala (LA)	I	1521 (133) ^a	732 (74) ^a	1765 (339) ^a
	II	1527 (249)	827 (97)	1248 (470)
Right amygdala (RA)	I	1490 (184)	696 (65)	1815 (119)
	II	1502 (212)	814 (85)	1051 (367)
Left hippocampus (LH)	I	4209 (291)	2850 (398)	4912 (566)
	II	3884 (475)	3194 (254)	4226 (558)
Right hippocampus (RH)	I	4075 (339)	2237 (272)	4944 (542)
	II	3766 (533)	2464 (180)	4154 (494)

a. Average volume (standard deviation). Group I is composed by exams scanned in a 1.5 Tesla MR unit and Group II is composed by exams scanned in a 3 Tesla MR unit.

**Fig. 7.4** Amygdala and Hippocampus absolute volumes obtained from Group II (3.0 Tesla MR unit).

The brain structure average volumes are better compared to each other using normalized volumes. The volume normalization is usually obtained dividing the brain structure volume to the assigned subject's intracranial volume. The estimated total intracranial volume (eTIV) obtained automatically from Freesurfer [5] was used. Equation 11.1 shows how the normalized volume for each subject is obtained. Equation 11.2 shows the average normalized volume for each subject group and for each medical images analyzing package.

$$nVol_i = \frac{Vol_i}{eTIV_i} \cdot 10000 \quad (7.1)$$

where Vol_i is the structure volume of a given individual and $eTIV_i$ is the estimated total intracranial volume.

$$average(nVol) = \frac{\sum_{i=1}^n nVol_i}{n} \quad (7.2)$$

where n is the number of subjects per group.

Table 7.5 shows the average normalized volume values got from the three medical image analyzing methods. Comparing to the normalized volumes from Group I and Group II, there should be no significant differences between average pairs, when considering the same brain structure. In order to validate such hypothesis, the confidence interval for each pair difference average was computed, as shown in Table 6. We used a significance level (α) equal to 95%. When the confidence interval does not contain zero, it can be concluded that the averages of Group I and Group II are different. We noticed that left-right hippocampus and right amygdala volumes computed by FSL are different when comparing Group I to Group II, as shown by the highlighted values in Table 11.5.

Table 7.5 Average normalized volumes.

Brain structure	Group	Freesurfer ^a	IBASPM ^a	FSL ^a
LA	I	10.8 (0.9)	5.2 (0.4)	12.6 (2.8)
	II	10.6 (0.7)	5.8 (0.9)	8.8 (3.4)
RA	I	10.5 (0.7)	4.9 (0.5)	12.9 (1.6)
	II	10.5 (0.7)	5.7 (0.6)	7.4 (2.5)
LH	I	29.9 (2.7)	20.1 (1.7)	34.8 (3.5)
	II	27.1 (1.3)	22.6 (3.2)	29.5 (1.6)
RH	I	29.1 (4.0)	15.9 (2.3)	35.4 (5.5)
	II	26.3 (1.8)	17.3 (1.7)	29.1 (2.2)

a. Average normalized volume (standard deviation). Group I is composed by exams scanned in a 1.5 Tesla MR unit and Group II is composed by exams scanned in a 3 Tesla MR unit.

Table 7.6 Confidence interval for each pair group difference average (Group I and Group II).

Brain structure	Freesurfer ^a	IBASPM ^a	FSL ^a
LA	(-3.79; 4.10)	(-4.58; 3.30)	(-0.15; 7.74)
RA	(-3.91; 3.98)	(-4.73; 3.16)	(1.62; 9.50)^b
LH	(-1.17; 6.72)	(-6.38; 1.50)	(1.33; 9.22)^b
RH	(-1; 12; 6.76)	(-5.38; 2.51)	(2.32; 10.21)^b

a. (n1; n2), where n1 and n2 are the limits of confidence interval, respectively, considering a significance level (α) equal to 95%;

b. Group I and Group II are different due to the confidence interval not containing zero.

Many studies have correlated the amygdala and hippocampus volumetric measurements asymmetry with brain diseases and symptom manifestation [22]. In order to measure the degree of asymmetry between the same structures located in different brain hemispheres, two metrics were used: (1) the Pearson's correlation coefficient (p -value) and (2) the mean right-to-left volume ratio. A Pearson's correlation equal to 1 means no difference at all. The smaller the correlation value, the more significant the difference between the two groups is. Equation 11.3 shows how the right-to-left volume ratio was obtained.

$$RLVRi = \frac{RVi - LVi}{RVi + LVi} \quad (7.3)$$

where $RLVRi$ is the right-to-left volume ratio, RVi is the volume obtained for structure located at the right brain hemisphere, LVi is the volume obtained for structure located at the left brain hemisphere, i is the index representing each subject.

As shown in Table 11.6 in highlighted text, the $RLVR$ values are all negative for amygdala, suggesting that the right amygdala is larger than the left amygdala. However, $RLVR$ values are all positive for hippocampus. It suggests that left hippocampus is larger than the right hippocampus considering this image group.

Table 7.7 Right to left volume ratios.

Brain structure	Group	Freesurfer p -value	IBASPM p -value	FSL p -value
RA - LA	I	0.7	0.3	0.7
	II	0.8	0.8	0.3
	RLVR^a	-0.009 (0.44)	-0.016 (0.02)	-0.014 (0.25)
RH - LH	I	0.4	<0.01	0.9
	II	0.6	<0.01	0.8
	RLVR^a	0.46 (0.04)	0.59 (0.027)	0.53 (0.11)

a. Mean right-to-left volume ratio ($RLVR$) computed according to Equation 11.1. The standard deviation is shown between parentheses.

The volume agreement was measured using the relative volume overlay, as shown in Equation 11.4. The volume agreement indexes aim at quantifying the volume agreement of the structures segmented automatically by the three studied medical image packages.

$$ROV(A_1; \dots A_j) = \frac{A_1 \cap \dots A_j}{A_1 \cup \dots A_j} \cdot 100 \quad (7.4)$$

where $A_1, \dots A_j$ are tridimensional vectors representing the spatial coordinates of voxel sets of anatomical structure A considering medical image packages $1 \dots j$, respectively.

Table 11.7 shows the relative overlay volumes (ROV). Freesurfer and FSL presented better relative overlay indexes, as shown by highlighted values in Table 11.7. The highest ROV index was 67.59%. Considering the same brain anatomical struc-

ture, the highest relative overlay volume value was 35.42%, when considering the three medical image segmentation packages.

Table 7.8 Relative Overlay Volumes (ROV).

Brain structure	Group	Freesurfer and IBASPM	Freesurfer and FSL	IBASPM and FSL	Freesurfer, IBASPM and FSL
		ROV (SD) ^a	ROV (SD)% ^a	ROV (SD)% ^a	ROV (SD)% ^a
LA	I	26.71 (4.84)	60.20 (3.40)	62.58 (3.49)	19.28 (3.69)
	II	26.55 (3.03)	53.86 (17.23)	56.25 (14.09)	18.21 (5.20)
RA	I	6.57 (4.82)	63.72 (2.84)	55.59 (4.36)	4.58 (2.92)
	II	6.60 (3.22)	47.91 (11.93)	45.57 (6.17)	2.62 (2.31)
LH	I	43.00 (7.59)	61.75 (4.37)	61.42 (3.61)	30.72 (5.96)
	II	45.35 (4.53)	67.59 (2.72)	60.95 (9.16)	35.42 (3.97)
RH	I	23.23 (5.48)	61.48 (2.44)	56.26 (3.11)	14.85 (4.70)
	II	20.35 (2.36)	63.61 (5.60)	52.92 (5.08)	12.33 (2.50)

a. Relative overlay volume average (ROV) and standard deviation (SD)

Figure 7.5 shows the left-right amygdala areas obtained by the three image segmentation packages in three different image slices. The areas are displayed on 2D coronal images. We observed that Freesurfer has included part of superior region of hippocampus and CSF (cerebrospinal fluid), as indicated in Figure 7.5a1. Also, it has included the white matter (WM), as indicated in Figure 7.5a2. IBASPM has not included many anatomical regions that belongs to the amygdala, as shown in Figure 7.5b. That explains why amygdala volumes from IBASPM were much smaller than the others (see Table 11.4). Regarding FSL, drawn volumes overestimated amygdala real area, including the WM and cerebral spinal fluid (CSF), as shown in Figure 7.5c.

Figure 7.6 shows the left-right hippocampus areas obtained by the three image segmentation packages in four different image slices. Regarding Freesurfer, we observed that when the sulcus between the hippocampus and the parahippocampus gyrus is not clearly defined, the upper cortex of the parahippocampus gyrus was included in the hippocampus volume segmentation. Probably, that problem will not occur in the elderly population as that sulcus will be more pronounced between those two structures, allowing better boundary discrimination. The dentate gyrus, subiculum and upper area of parahippocampal gyrus (entorhinal cortex) seem to have been included in the hippocampus volume, considering the images belonging to both groups studied, as shown by the red arrows in Figure 7.6a3. Freesurfer also included the inferior cornu of the lateral ventricle. In the most posterior portion (tail), Freesurfer included part of the fornix, as shown by the red arrow in Figure 7.6a1. The Freesurfer hippocampus region of interest segmentation criteria were not clearly present in the program technical manual [26]. In FSL, the limits of the hippocampus were overestimated, including adjacent white matter area and other neighbouring structures, providing volumes significantly inflated. FSL also included the fornix and entorhinal cortices. We noticed that the hippocampal volumes obtained from IBASPM have shown discontinuous areas, as shown in Figure 7.6b3

and b4, The errors in hippocampal volume from IBASPM arose from inaccurate image registration and use of the MNI single-subject for the manually predefined hippocampus region of interest [28].

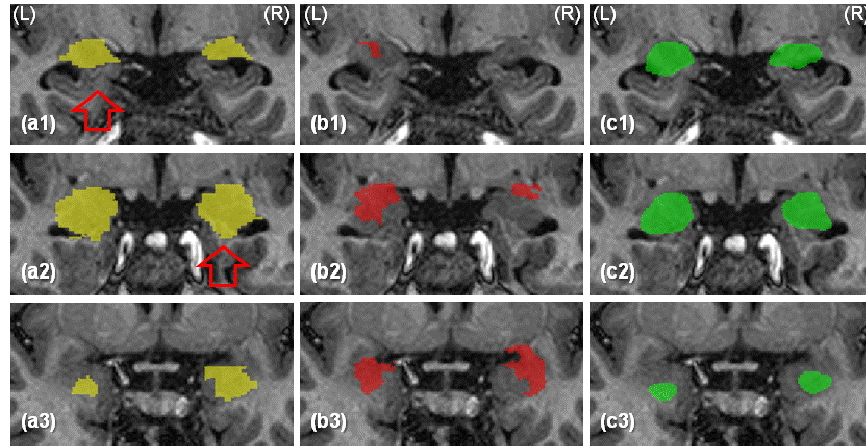


Fig. 7.5 Images slices showing the left (L) and right (R) amygdala slices of a normal subject: 1 and 3 indicate the most-posterior and most-anterior slices, respectively; (a), (b) and (c) show the results of Freesurfer (yellow), IBASPM (red) and FSL (green), respectively.

7.5 Conclusion

This chapter compared volumetric measures obtained from three well-known and public-domain medical image analyzing packages: Freesurfer, IBASPM and FSL. The amygdala and hippocampus were evaluated comparing the volumes obtained automatically for each structure anatomical definition. Volumetric measures have shown some disagreements among the three studied packages. The same regions labeled incorrectly were observed in all images used in the study. Thus, we conclude that the errors were replicated in all images.

Despite being not trustful, manual segmentation is commonly used as a reference in statistic evaluations of sensitivity and specificity. This chapter only considered automated image segmentation methods; no manual segmentation results were available for comparison. As a future work, manual segmentation will be considered in the comparison. The final objective is selecting and/or improving a brain segmentation method in order to obtain amygdala and hippocampus volumes closer to their anatomical definitions. Those results can be used in a clinical decision support system or computer-aided diagnosis tool for AD.

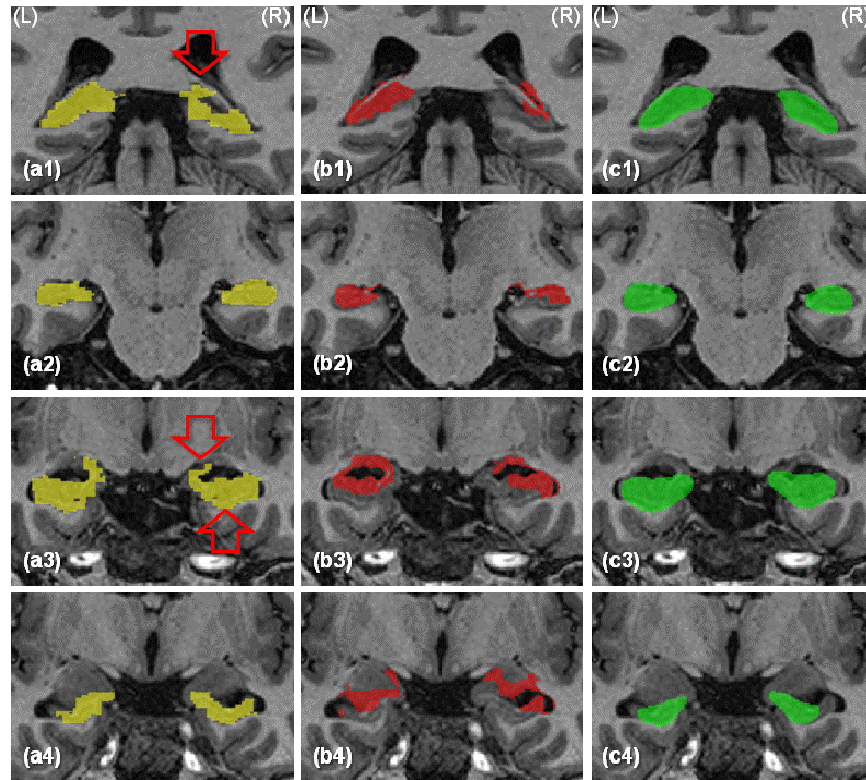


Fig. 7.6 Image slices showing the left (L) and right (R) hippocampus of a normal subject: 1 and 4 indicate the most-posterior and most-anterior slices, respectively; (a), (b) and (c) show the results from Freesurfer (yellow), IBASPM (red) and FSL (green), respectively.

Acknowledgements

We thank the LABS DOr Network, Radiology Department, Rio de Janeiro, Brazil for providing the magnetic resonance images used in this study. This work has been partially supported by CAPES, CNPq, FAPERJ (Brazilian agencies to promote scientific research) and INCT-MACC (National Institute of Science and Technology for Medicine Assisted by Scientific Computing).

References

- [1] Alezn-Gmez, Y., Melie-Garca, L., Valds-Hernandez, P.: Ibaspm: Toolbox for automatic parcellation of brain structures. pp. 11–15 (2006)
- [2] Andersen, P.: The hippocampus book. Oxford University Press, USA (2007)

- [3] Bottino, C.M.C., Azevedo, D., Tatsch, M., Hototian, S.R., Moscoso, M.A., Folquitto, J., Scalco, A.Z., Bazzarella, M.C., Lopes, M.A., Litvoc, J.: Estimate of dementia prevalence in a community sample from so paulo. *Dementia and Geriatric Cognitive Disorders* **26**(4), 291–299 (2008)
- [4] Braak, H., Braak, E.: Neuropathological staging of alzheimer-related changes. *Acta neuropathologica* **82**(4), 239–259 (1991)
- [5] Buckner, R., Head, D., Parker, J., Fotenos, A., Marcus, D., Morris, J., Snyder, A.: A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **23**(2), 724–738 (2004)
- [6] Carmichael, O., Aizenstein, H., Davis, S., Becker, J., Thompson, P., Meltzer, C., Liu, Y.: Atlas-based hippocampus segmentation in alzheimer's disease and mild cognitive impairment. *Neuroimage* **27**(4), 979–990 (2005)
- [7] Chen, Y., Tagare, H., Thiruvankadam, S., Huang, F., Wilson, D., Gopinath, K., Briggs, R., Geiser, E.: Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision* **50**(3), 315–328 (2002)
- [8] Collins, D.: 3d model-based segmentation of individual brain structures from magnetic resonance imaging data. Ph.D. thesis (1994)
- [9] Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P.: Research criteria for the diagnosis of alzheimer's disease: revising the nincds-adrda criteria. *Lancet Neurology* **6**, 734–746 (2007)
- [10] Firbank, M., Barber, R., Burton, E., O'Brien, J.: Validation of a fully automated hippocampal segmentation method on patients with dementia. *Human brain mapping* **29**(12), 1442–1449 (2008)
- [11] Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S.: Whole brain segmentation:: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3), 341–355 (2002)
- [12] Hill, D., Batchelor, P., Holden, M., Hawkes, D.: Medical image registration. *Physics in medicine and biology* **46**, R1 (2001)
- [13] Jack Jr, C., Petersen, R., Xu, Y., Waring, S., O'Brien, P., Tangalos, E., Smith, G., Ivnik, R., Kokmen, E.: Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. *Neurology* **49**(3), 786–794 (1997)
- [14] Jack Jr, C., Slomkowski, M., Gracon, S., Hoover, T., Felmlee, J., Stewart, K., Xu, Y., Shiung, M., O'Brien, P., Cha, R.: Mri as a biomarker of disease progression in a therapeutic trial of milameline for ad. *Neurology* **60**(2), 253–260 (2003)
- [15] Kantarci, K., Jack Jr, C.: Neuroimaging in alzheimer disease: an evidence-based review. *Neuroimaging Clinics of North America* **13**(2), 197 (2003)
- [16] Klppel, S., Stonnington, C., Barnes, J., Chen, F., Chu, C., Good, C., Mader, I., Mitchell, L., Patel, A., Roberts, C.: Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain* **131**(11), 2969–2974 (2008)
- [17] Maguire, E., Spiers, H., Good, C., Hartley, T., Frackowiak, R., Burgess, N.: Navigation expertise and the human hippocampus: a structural brain imaging analysis. *Hippocampus* **13**(2), 250–259 (2003)
- [18] McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis. pp. 171–180. *IEEE* (1996)
- [19] Morey, R., Petty, C., Xu, Y., Pannu Hayes, J., Wagner II, H., Lewis, D.: A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* **45**(3), 855–866 (2009)
- [20] Nitirini, R., Bottino, C.M., Albala, C., Custodio Capunay, N.S., Ketzoian, C., Llibre Rodriguez, J.J., Maestre, G.E., Ramos-Cerqueira, A.T., Caramelli, P.: Prevalence of dementia in latin america: a collaborative study of population-based cohorts. *Int Psychogeriatr* **21**(4), 622–30 (2009). Using Smart Source Parsing Aug; Epub 2009 Jun 9

- [21] Patenaude, B., Smith, S., Kennedy, D., Jenkinson, M.: A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* (2011)
- [22] Pedraza, O., Bowers, D., Gilmore, R.: Asymmetry of the hippocampus and amygdala in mri volumetric measurements of normal adults. *Journal of the International Neuropsychological Society* **10**(5), 664–678 (2004)
- [23] Petersen, R.: *Mild cognitive impairment: aging to Alzheimer's disease*. Oxford University Press, USA (2003)
- [24] Powell, S., Magnotta, V., Johnson, H., Jammalamadaka, V., Pierson, R., Andreasen, N.: Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage* **39**(1), 238–247 (2008)
- [25] Symms, M., Jger, H., Schmierer, K., Yousry, T.: A review of structural magnetic resonance neuroimaging. *Journal of Neurology, Neurosurgery and Psychiatry* **75**(9), 1235–1244 (2004)
- [26] Tae, W., Kim, S., Lee, K., Nam, E., Kim, K.: Validation of hippocampal volumes measured using a manual method and two automated methods (freesurfer and ibaspm) in chronic major depressive disorder. *Neuroradiology* **50**(7), 569–581 (2008)
- [27] Talairach, J., Tournoux, P.: *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*. Thieme (1988)
- [28] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* **15**(1), 273–289 (2002)
- [29] Whalen, P., Phelps, E.: *The human amygdala*. The Guilford Press (2009)

Chapter 8

Adaptive mesh generation for fluid flow simulation in arteries

Alexandre De Lacassa, Antonio Castelo Filho, Mario A. Lizier, Ícaro L. L. da Cunha

8.1 Introduction

The use of finite element and finite volume methods is common in science and engineering problems. Such problems often involve numerical simulation of partial differential equations that govern phenomena such as mechanical deformation, heat transfer, fluid dynamics, propagation of electromagnetic waves, quantum mechanics, among others. A common feature in some of these problems is that most of the objects of interest, which characterize the problem domain, contains complex shapes. To allow the simulation of these objects, the domain is decomposed in a set of simple geometric shapes called *elements*, which are usually triangles or tetrahedrons, but may also be quadrilaterals, prisms, pyramids or hexahedrons. These elements and their vertices represent the object, allowing the calculations of the finite element / finite volume. The set of all nodes and elements representing the domain is called *mesh*.

Alexandre De Lacassa

LMACC - Laboratório de Matemática Aplicada e Computação Científica an Associated Laboratory of INCT-MACC, ICMC - Instituto de Ciências Matemáticas e de Computação - USP, Av. Trabalhador são-carlense, 400, Centro, 13566-590, São Carlos, SP, Brazil, e-mail: lacassa@gmail.com

Antonio Castelo Filho

LMACC - Laboratório de Matemática Aplicada e Computação Científica an Associated Laboratory of INCT-MACC, ICMC - Instituto de Ciências Matemáticas e de Computação - USP, Av. Trabalhador são-carlense, 400, Centro, 13566-590, So Carlos, SP, Brazil, e-mail: castelo@gmail.com

Mario A. Lizier

DC - Departamento de Computação - UFSCAR, Rodovia Washington Luís, Km 235, 13565-905, São Carlos, SP, Brazil, e-mail: lizier@gmail.com

Ícaro L. L. da Cunha

Laboratório NatalNet and Departamento de Engenharia de Computação e Automação, Centro de Tecnologia - UFRN, Caixa Postal 1524 - Campus Universitário Lagoa Nova, 59078-900, Natal, RN, Brazil, e-mail: illdacunha@gmail.com

Mesheres are generally categorized into structured and unstructured. The structured meshes have uniform topology, which does not happens with unstructured meshes. In unstructured meshes, the neighbors are represented by topological data structures.

This paper will focus on unstructured volumetric meshes formed by tetrahedrons, for applications of fluids simulation in human arteries.

The quality of results involving finite element methods is closely related to the shape of the mesh elements. For tetrahedral meshes, the dihedral angle characterizes the effect that the element has on the solution of the problem. Tetrahedrons with large dihedral angles (180°) [28] cause large errors of interpolation and discretization, which compromises the accuracy of the numerical simulation, and small dihedral angles (0°) cause ill conditioning in the matrices associated with the finite element method. We can then represent the quality of a tetrahedron as a number which represents its effect on interpolation errors, discretization errors, and on the conditioning of matrices of finite element method. Because of restrictions that elements with very large angles or very small angles cause in a simulation, we say that the quality of a mesh is closely related to the smallest and greatest angles. Therefore, we say an element has good quality if its shape is close to equilateral.

Methods for mesh generation are always looking for elements with good quality. One technique that offers quality assurance on elements of the mesh is the Delaunay triangulation [26, 31, 18, 9]. Another well known technique for mesh generation is the Advance Frontier [21, 29, 20, 22]. In this approach, a front formed only by triangles is first built on the domain edge, and in every step this front grows with the creation of new elements in order to represents the whole domain. The elements created in each step follow a minimum quality criterion.

The meshing methods based on octrees [25, 24, 33, 32] use this spatial structure to form the mesh elements. These elements are constructed using the vertices of the octree and the intersections between the octree edges and domain, which is usually represented by an implicit function. As these methods are not always able to generate meshes with good elements, strategies to improve the quality of these elements are required. The Delaunay refinement algorithm [26] seeks to improve the quality of the elements through insertion of points at strategic locations, in order to remove the low quality elements, considering predetermined value.

However, there are techniques to improve mesh quality, Delaunay meshes or not, that do not use this approach of adding new points. These techniques move the vertices to new locations, so that improve the mesh elements quality. These methods are called *mesh smoothing* [19, 11]. Methods based on dynamic meshes (with vertices that move) try to insert a elastic feature at the edges, using the concept of spring [2, 10, 35, 34, 30]. These techniques move the vertices of the mesh, by finding the balance between the springs. Smoothing methods based on optimization also displace the vertices to new positions [1, 23]. However, the displacements are made through a process of maximization or minimization of functions that provide information about the quality of the elements affected by the displacements of the vertices.

We present an approach for isosurfaces reconstruction in conjunction with a smoothing step, in order to develop a mesh volumetric generator. The development of this mesh generator is an extension of the work done by Gois-Junior and Polizelli in [13, 14] that discusses only the reconstruction of isosurfaces represented by clouds of points.

The goal is to extend the work [14], inserting elements within a closed mesh surface to obtain a volumetric mesh. The work presented in [7] seeks to fulfill this goal, but with the use of the ideas presented by Labelle and Shewchuk in [17]. In [17] elements of the decomposition of the space are used as internal elements of the mesh surface. The same idea is used in the work presented here. The mesh elements (in this case, the tetrahedra of J_1^a triangulation) will be used as internal elements of the model represented by the isosurface mesh. Thus, we obtain an initial mesh as shown in [7].

The main feature of these meshes is adaptivity, i.e. a greater number of elements near the surface, and fewer elements towards the middle region of the volume. However, these meshes still contains a large number of elements with a quality that can be improved. In this sense, we have developed a process of post-processing that seeks to smoothing the mesh, improving the quality of its elements. This post-processing consists in a strategy for smoothing mesh through virtual springs [30, 2, 10, 34]. This smoothing process is applied firstly on the surface elements and then the internal elements of a volumetric mesh. The edges of the mesh have springs between vertices, generating a spring system. In this spring system, there are forces that act on vertices that need to move them to maintain the static equilibrium of the entire system. Thus, both the inner elements as the surface are smoothed. The tool *Stellar*, created by Klingner and Shewchuk in [16], is used to further improve the quality of the final mesh.

8.2 The J_1^a triangulation

Proposed by [6, 5], the J_1^a triangulation is an algebraically defined structure that can be efficiently constructed in any dimension. To accommodate local features it is naturally capable of handling refinements. Some of its main features are the existence of a mechanism for representing simplices and the existence of algebraic rules for traversing the triangulation. The use of these rules prevents the structure from needing to store connectivity, thus, enabling a more efficient storage.

The three-dimensional J_1^a triangulation for volumetric mesh generation can be classified as an Octree-based method for the construction of the domain partitioning through the J_1^a triangulation is similar with the domain partitioning made through Octree-based methods. However, the algebraic characteristic of the J_1^a triangulation prevents the data structure that such methods require. Another difference with respect to such methods is that the tetrahedral elements that will be part of the final mesh are already present in the description of the J_1^a triangulation.

The structure provided by J_1^a triangulation allows it to be used in conjunction with an implicit function that represents an object of study.

In [14] the isosurface reconstruction by cloud points makes use of a tetrahedral spatial decomposition method based on J_1^a triangulation with a method of partition of unity implicits (PUI method) proposed originally in [13]. This method gives implicit function for approximation of a surface sampled by a cloud of points. The reconstruction is completed by the polygonization process that extracts a surface mesh through the intersection of the implicit function of level 0 with the elements of J_1^a triangulation.

The volumetric mesh generator presented in this work uses the PUI method [13] and the J_1^a triangulation to generate the surface and the internal elements of a mesh.

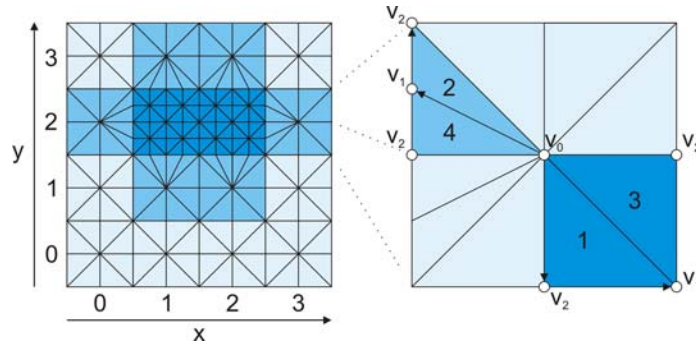


Fig. 8.1 A 2D example of a J_1^a triangulation grid (left) and the details of block $g = (3,2)$, $r = 0$ and two paths for tracing simplices (right).

The J_1^a triangulation consists of a grid formed by n -dimensional hypercubes in a \mathbb{R}^n domain (we will refer to these hypercubes as blocks). Each block is divided into $2^n n!$ n -simplices that can be described algebraically using six values:

$$S = (g, r, \pi, s, t, h) \quad (8.1)$$

The first two elements of S define in which block the simplex is contained, where g is an n -dimensional vector that indicates a block's coordinate at a particular refinement level of the grid given by r . Figure 8.1 illustrates, on the left, a two-dimensional J_1^a grid and, on the right, a highlighted block of refinement level $r = 0$ (0-block) and $g = (3,2)$. Also in Figure 8.1, one can notice that the darkest blue blocks present refinement level $r = 1$ – thus are called 1-blocks – and, for that reason, are part of a higher resolution grid.

To handle the different levels of refinement, the J_1^a triangulation splits a block (that is supposed to be refined) into 2^n blocks and applies some changes to its neighboring blocks in order to accommodate the newly created smaller blocks. Then, a different kind of block is created which will be referred to as a transition block. By having this transition block, the J_1^a triangulation imposes that there is not any situa-

tion where two neighboring blocks have a difference in refinement level larger than one. The refinement level of a transition block is defined as the lowest level of its neighboring blocks.

The transition block has some of its k -dimensional faces (where $0 < k < n$) refined so it can accommodate their neighboring refined blocks. We illustrate this in Figure 8.1, where the highlighted transition block only has its left edge refined. The non-transition blocks (in the figure are represented in the lighter and darker blue) are referred to as basic blocks.

To represent each n -simplex of a block, the first to be noticed is that all of them share at least one vertex (the center vertex labeled as v_0). From this center vertex of the n -dimensional hypercube, the next step is taken in the positive or negative direction of a chosen coordinate axis. This will end up at v_1 (which is at the center of a $(n-1)$ -dimensional face), the process continues for vertices v_2, \dots, v_n (which are respectively the centers of $(n-2), \dots, 0$ dimensional faces). So by having the simplex's vertices it can be presented by the path traversed from v_0 to v_n . The vectors π and s are responsible for coding this path, π stores a permutation of n integers from 1 to n that represents the coordinate axis, and s represents the direction (by positive or negative values) to be taken at each axis. In Figure 8.1, the simplex 1 in the highlighted block is represented by $\pi = (2, 1)$ and $s = (1, -1)$, which means that the first path to be taken should be along the y axis ($\pi_1 = 2$) in the negative direction ($s_{\pi_1} = -1$), and then should be along the x axis ($\pi_2 = 1$) in the positive direction ($s_{\pi_2} = 1$).

The π and s vectors are enough information to represent the simplices from basic blocks and simplices that are from transition blocks but do not reach a refined face. For the remaining cases it must be noticed that when a k -dimensional face is reached, there are more than one center. Because of this other information must be stored to represent these simplices, the scalar h is used to inform how many steps are taken before a refined face is reached, and the vector t defines extra signs for axis $\pi_{h+1} \dots \pi_n$ that are used for selecting one center from all possibilities on the refined face. So again in Figure 8.1, simplex 2 is represented by $\pi = (1, 2)$, $s = (-1, 1)$, $h = 1$ and $t = (0, 1)$; h is equal to 1 because only one step is taken before reaching a refined edge and the chosen center for placing v_1 is in the positive direction of π_{h+1} .

A formal way of describing the representation of a simplex inside a hypercube (centered at the origin $(0, \dots, 0)$ and with edge length equal to 2) is given by the following expression:

$$\begin{cases} v_0 = (0, \dots, 0) \\ v_i = v_{i-1} + e_{\pi_i} s_{\pi_i}, \text{ for } 1 \leq i < h \\ v_h = v_{h-1} + e_{\pi_h} s_{\pi_h} + \frac{1}{2} \sum_{k=h+1}^n e_{\pi_k} t_{\pi_k} \\ v_i = v_{i-1} + \frac{1}{2} e_{\pi_i} s_{\pi_i}, \text{ for } h < i \leq n \end{cases}$$

Where e_i is a vector with value 1 in position i and 0 in the remaining positions.

Another important mechanism of the J_1^q triangulation is its pivoting rules for traversing the triangulation without using an auxiliary topological structure. In Figure 8.1 two pivoting operations are illustrated, where simplex 1 is pivoted in relation to the vertex v_1 resulting in simplex 3, and simplex 2 is pivoted in relation to the ver-

tex v_{21} resulting in simplex 4. All of J_1^a 's pivoting rules can be found in the work by [6].

8.3 Mesh Improvement

Most methods of tetrahedral mesh improvement uses three operations:

- a. *Smoothing*: relocation of vertices in order to improve the quality of mesh elements, without changing its topology;
- b. *Topological*: changing the connections between the vertices, without changing vertices position;
- c. *Insertion of vertices*: adding extra vertices in the mesh (splitting edges, faces or tetrahedrons in the affected region).

One of the smoothing procedures is based on the use of springs in the virtual mesh elements. The application of virtual spring mesh is found in studies of dynamic meshes, where meshes are commonly used in dynamic simulation problems on domains whose geometry varies with time. In this work, the application of virtual springs works only as a smoothing operation of the mesh, i.e. the grid points are relocated in accordance with the virtual springs applied to each edge of the mesh.

Below we discuss each of the operations listed above.

8.3.1 Improvement by smoothing

Mesh smoothing methods aim to improve the quality of a mesh by moving its vertices without changing mesh topology (Figure 8.2). In general, the smoothing algorithms use techniques which adjust the local position of each vertex individually, in order to improve the quality of the incidents elements.

We can classify the smoothing methods mainly in two classes: methods based on Laplacian smoothing and methods based on optimization.

Laplacian Laplacian based method [19, 11] move each vertex to the geometric center of its polygon, determined by its incident vertices (Figure 8.2). This method is computationally inexpensive, but operates on a heuristic and does not guarantee the improvement of elements quality. Indeed, it is possible to produce an invalid mesh, that is, inverted elements. The use of this type of smoothing is not good in surface meshes, it can eliminate sharp features of the object.

Optimization Optimization-based methods [1, 23, 25] prevent the creation of invalid elements and seek an optimal location for vertices using an optimization technique. This optimization technique generally determines the local minimum of a function applied to the elements of a region. This function is created from a quality of mesh elements, and are constructed so that its value is minimal if the mesh elements are close to the desired, i.e., are elements with good quality. These tech-

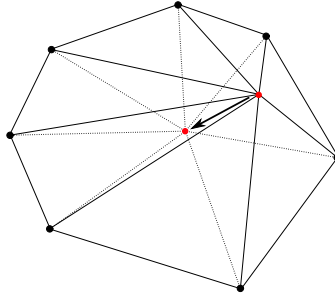


Fig. 8.2 Example of 2D smoothing: the vertex is moved in order to improve the neighbor triangles.

niques have the advantage of offering a guarantee of improvement and validation of the meshes. However, this guarantee requires a higher computational cost compared to the method of Laplace smoothing.

8.3.2 Improvement by topological changes

Improving meshes by topological transformations is based on replacing elements by a different set of elements, with better quality. The filling of the space, left by the erased elements, may be performed in an arbitrary manner, or via a set of *topological operations* as described below:

- 2-3 flip, and its inverse, 3-2 flip (See figure 8.3);
- 4-2 flip, and its boundary version 2-2 flip (See figure 8.3): this operation requires the definition of the pair of vertices which will be flipped;
- Edge removal (See figure 8.4): this is a generalization of 3-2, 4-4 and 2-2 flips. It can be applied on boundary edges. There are many ways to connect the removed edge adjacent vertices, Freitag e Ollivier-Gooch [12] uses templates of triangles, Klingner e Shewchuk [16, 27] uses Klincsek [15] method in order to maximize the minimum quality of the set of created tetrahedrons;
- Multi-faces removal: created by Cougny e Shephard [8] this operation is the inverse of edge removal, shown in figure 8.4. It is a generalization of 2-3 and 4-4 flips, and requires dynamic programming to select and remove the set of faces that are opposed to the two vertices, in order to define a better configuration that improves quality of elements.

Freitag and Ollivier-Gooch [12] uses the first three operations. It may be noted that removal multi-faces can be decomposed into a sequence of a single 2-3 flips 2-3, followed by several 3-2 flips. Klingner and Shewchuk [16] uses in our studies, makes use of all the operations listed above.

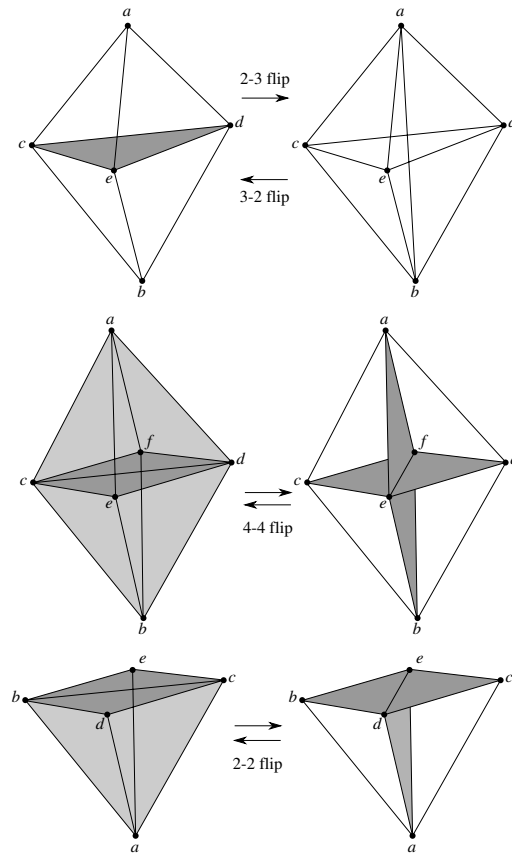


Fig. 8.3 Topological changes: 2-3 flip (and its inverse: 3-2 flip), 4-4 flip and 2-2 flip.

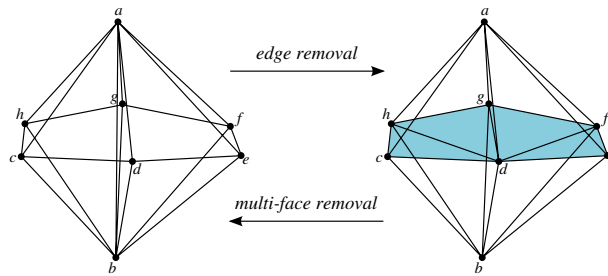


Fig. 8.4 Basic topological changes.

8.3.3 Improvement by vertex insertion

One of the most known improvement strategy based on vertices insertion is the refinement algorithm developed by Delaunay Ruppert [26]. In this method, however, the mesh has to be a Delaunay triangulation and the insertion of a vertex creates new elements that need to fulfill the criterion of the Delaunay mesh.

The main contribution of Klingner and Shewchuk's works [16] is the strategy designed to improve the mesh through the changes made as a result of new vertex insertion. Klingner and Shewchuk in their work showed that the meshes improvement done by transformations caused by the insertion of vertices is much more effective. The proper location of a vertex is a difficult problem to be solved. Klingner and Shewchuk describe a sophisticated and complex algorithm to insert vertices which is similar to inserting vertices of Delaunay algorithm and, along with smoothing operations based on optimization and topological operations, allows the improvement of these meshes, reaching significant results, for example, with all angles between 31° and 149° , or by using an objective function different from 23° and 136° .

8.4 Mesh smoothing through virtual springs

To understand the mesh smoothing process through virtual springs we have to know the types of springs commonly used in the study of dynamic meshes.

There are several types of virtual springs however, the main types are the following:

- Lineal springs: it works on edges and control their lengths.
- Torsional springs: it works on angles between edges and control areas.
- Semi-torsional springs: a simple approach of torsional springs; as well as torsional springs, control areas or volumes.

Soares [30] explored these types of springs and created another to deal with mesh movement and superficial remeshing.

8.4.1 Lineal springs

The lineal springs were proposed by Batina [2]. The author proposed that the spring stiffnesses are inversely proportional to the edge lengths. This way, the smaller the edge, the spring is more rigid, allowing control over its lengths. Thus, to an edge connecting the vertices v_i and v_j , the spring stiffness, denoted by κ_{ij} , is given by (see Figure 8.5):

$$\kappa_{ij} = \frac{a}{(l_{ij})^b}, \quad (8.2)$$

where l_{ij} is the length of an edge that connects v_i to v_j , a and b are constants (generally equal to 1).

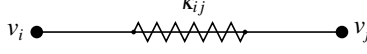


Fig. 8.5 Illustration of a linear spring.

Since the concept of lineal springs is related to the edge lengths, there is no significant difference between using it in triangle meshes or in tetrahedral meshes.

8.4.2 Torsional springs

In a mesh, the torsional springs are a way to control the angles between the edges, and consequently, control the areas of the mesh elements. The use of torsional springs in the problem of dynamic meshes was proposed by Farhat et al. [10].

Consider the triangle t_{ijk} with vertices given by v_i , v_j and v_k . Let θ_i^{ijk} be the angle between the edges e_{ij} and e_{ik} . Let $A(t_{ijk})$ be the area of the triangle t_{ijk} . The spring stiffness in the vertex v_i of the triangle t_{ijk} is given by:

$$C_i^{ijk} = \frac{1}{1 + \cos(\theta_i^{ijk})} \frac{1}{1 - \cos(\theta_i^{ijk})} = \frac{1}{\sin(\theta_i^{ijk})^2} = \frac{(e_{ij})^2 (e_{ik})^2}{4A(t_{ijk})^2} \quad (8.3)$$

Unlike the lineal springs, which work on edges, the torsional springs work on every vertex in the mesh, more precisely in each angle incident to the vertex (2D case).

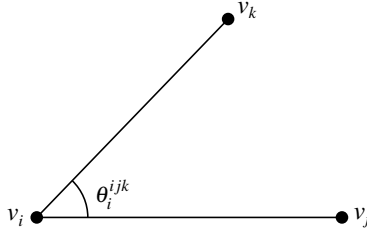


Fig. 8.6 Illustration of a torsional spring [30].

The extension of the idea of the torsional springs in 2D case to 3D case with tetrahedra is not immediate. In the 2D case the triangle area is controlled through the angles in each vertex, however, in the 3D case it is needed to control the volume of the tetrahedron. As tetrahedra have solid angles which are difficult to control, the task of using torsional springs in 3D is not trivial.

8.4.3 Semi-torsional springs

Since torsional springs aggregate angle information on each mesh vertex, incorporating this concept involves a greater complexity than required when using the lineal springs. The idea behind the semi-torsional spring is to make angles influencing the linear spring stiffness, therefore, the system contains, in fact, only linear springs. The semi-torsional springs proposed by Zeng and Ethier [34] imply to add a torsional spring stiffness to the opposite linear spring stiffness. That way, the total lineal stiffness κ_{ij}^{total} of the spring connecting the vertices v_i and v_j is given by:

$$\kappa_{ij}^{total} = \kappa_{ij} + C_k^{ijk} + C_l^{ijl} \quad (8.4)$$

where C is the torsional spring stiffness given by Equation (8.3).

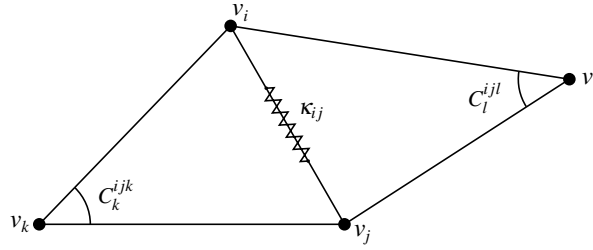


Fig. 8.7 A semi-torsional spring [30].

The use of semi-torsional springs in tetrahedra is simpler than in the case of torsional spring. The calculus of the opposite semi-torsional spring stiffness is done by replacing the angle between the edges by the dihedral angle in the Equation (8.3). This way, in the 3D case, the semi-torsional springs control the angles between faces instead of the angles between edges. Considering the Figure 8.8, where a semi-torsional spring is used to control the dihedral angle between the faces t_{lij} and t_{kji} . The semi-torsional spring ϕ_{ij}^{kl} is present in the dihedral angle on the edge e_{ij} and works over the linear spring presented in the edge e_{kl} , as is the case of Figure 8.8.

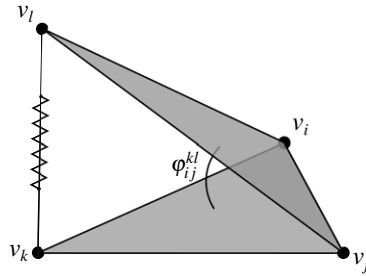


Fig. 8.8 Semi-torsional spring on tetrahedra.

8.4.4 The mesh smoothing scheme using virtual springs

The springs analogy in dynamic meshes involves the forces theory of Physics, which makes use of a system of equations to be solved for maintaining the equilibrium of the spring system. In [30] the author presents the Physic's theory and shows a manner the spring system must be mounted to application in a mesh.

The relocation of the vertices of the mesh by applying the virtual springs acts as a Laplacian smoothing method. In fact, the classic method of Laplacian smoothing is a particular case of the smoothing method using virtual springs, in which the springs have initial lengths of zero and the stiffness is taken constant [35].

In other words, the Laplacian smoothing method can consider the mesh as a spring system, where each edge connecting the vertices of the mesh can be regarded as a linear spring. Given a vertex in the mesh, the elements incident to this vertex form a polygon. If the stiffnesses are constant, the equilibrium point of the spring system will be the geometric center of the polygon.

In a spring system, the forces acting on a vertex can be represented by the mass-spring-damper model. In this model we have the following system of second-order ordinary differential equations, derived from Newton's second law:

$$M\ddot{\Delta} + D\dot{\Delta} + K\Delta = 0 \quad (8.5)$$

where Δ is the displacement vector of the vertices, M is the mass matrix associated with the vertices, D is the coefficient matrix of damping and K is the stiffness matrix of the springs.

Assuming that all corners have the same mass and the same coefficient damping, the problem can be simplified into a system known as quasi-static [10, 4], given by:

$$F_K = K\Delta = 0 \quad (8.6)$$

In this simplified form the system became linear with displacements of the vertices in Δ as unknown factors, and the matrix K is composed of the stiffnesses of the springs in the system. As in [30], we use this formulation to the spring system implemented in this work.

8.4.4.1 The iterative algorithm to mesh smoothing

The methods for solving the system of equations (Equation (8.6)) differ depending on the analogy used for the spring system, which depends on the type of application. Blom [4] presents two different analogies, one for use in dynamic mesh and another for mesh smoothing applications. In [30] the approach is for dynamic meshes, while this paper aims to perform the mesh smoothing. For mesh smoothing, Blom presented an analogy called "vertex spring". In this analogy there is no initial movement of the boundary elements of the mesh. The system can be solved iteratively by the Gauss-Seidel method.

At each iteration, the new position of vertex v_i is calculated as a weighted average of neighboring vertices. The weights are given by κ_{ij} (see Figure 8.9). Thus, the equation to be solved at each iteration is given as follows:

$$v_i^{t+1} = \frac{\sum_{j=0}^k \kappa_{ij} v_j^t}{\sum_{i=0}^k \kappa_{ij}} \quad (8.7)$$

where v_j with $0 < j < k$ are the vertices that form an edge with v_i and κ_{ij} is the stiffness of the spring representing the edge formed by the connection of v_i and v_j .

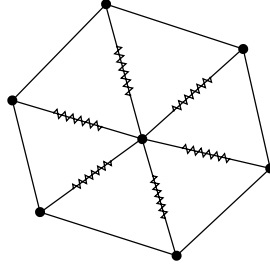


Fig. 8.9 Physical interpretation of the iterative process in a spring system [4].

The algorithm that solves the system through the Gauss-Seidel method is presented below. This algorithm calculates the new positions for all internal vertices of the mesh.

Algorithm 1 Mesh smoothing

Input:

Mesh to be smoothed

Tolerance ε

repeat

for each internal vertex v_i of the mesh **do**

 Compute v_i^{t+1} with Equation (8.7)

end for

until $\max_v \left(\frac{\|v_i^{t+1} - v_i^t\|}{e_{min}} \right) \leq \varepsilon$

The Algorithm 1 can be used to solve a system composed of linear springs or semi-torsional springs. The Equation (8.7) can be changed depending on the type of spring used in the system, i.e., what changes in Equation (8.7) is the value of spring stiffness that composes the system. For example, a system composed of lineal and semi-torsional springs the stiffness calculation is performed according to the Equation (8.4).

The convergence of the algorithm is satisfied if the following condition is true:

$$\max_v \left(\frac{\|v - v_n\|}{e_{min}} \right) \leq \varepsilon \quad (8.8)$$

where e_{min} is the length of the smaller edge incident to the iterated vertex v .

As described by Soares in [30] the choice of this criterion for determining convergence of the algorithm is based on the local characteristic of the mesh. In this manner, the evaluation of the displacement is made considering the size of the smallest edge among the edges incident to the displaced vertex. The displacement will converge if only its length does not exceed a portion of the size of the smallest local edge. This makes the stopping criterion of the algorithm to be more relaxed in regions where the mesh is coarser and more restrictive where the mesh is refined.

In this work we use the algorithm 1 to perform smoothing in the surface mesh and in the volumetric mesh. However, configuration of springs used to perform smoothing on the surface mesh is different from the setting chosen in the volumetric mesh.

To perform the smoothing in the surface mesh we chose only the linear spring, whereas in the smoothing over the volumetric mesh we chose the configuration with linear and semi-torsional springs. The choice of a different configuration for volumetric mesh is because that, in this case, the semi-torsional springs control the volume of tetrahedra. In the following sections we present the strategies to smooth the surface mesh and the volumetric mesh.

8.4.5 Surface mesh smoothing by virtual springs

When performing smoothing on the surface mesh by the Algorithm 1 presented above, the vertices can be displaced so that they no longer represent the surface of the original mesh. This can happen in regions where the surface curvature is high. To eliminate this problem we adopted a simple strategy that projects the vertex in the mesh surface again from the new position calculated by Equation (8.7). This way a step is added in Algorithm 1 right after the step that calculates the new position of the iterated vertex. The Algorithm 2 shows the strategy to smooth surface meshes.

Algorithm 2 Surface mesh smoothing

Input:

Surface mesh

Tolerance ε

repeat

for each internal vertex v_i of the mesh **do**

 Compute v_i^{t+1} with Equation (8.7)

 Updates the position of v_i^{t+1} by projecting it on the surface

end for

until $\max_v \left(\frac{\|v_i^{t+1} - v_i^t\|}{e_{min}} \right) \leq \varepsilon$

Figure 8.10 shows a surface mesh of an ellipse smoothed by the process described in the Algorithm 2. The histogram in Figure 8.10 shows the quality of surface triangles measured by the radius ratio defined by the ratio of the circumscribed circle to the inscribed circle of a triangle. The radius ratio of a equilateral triangle is equal to 1.

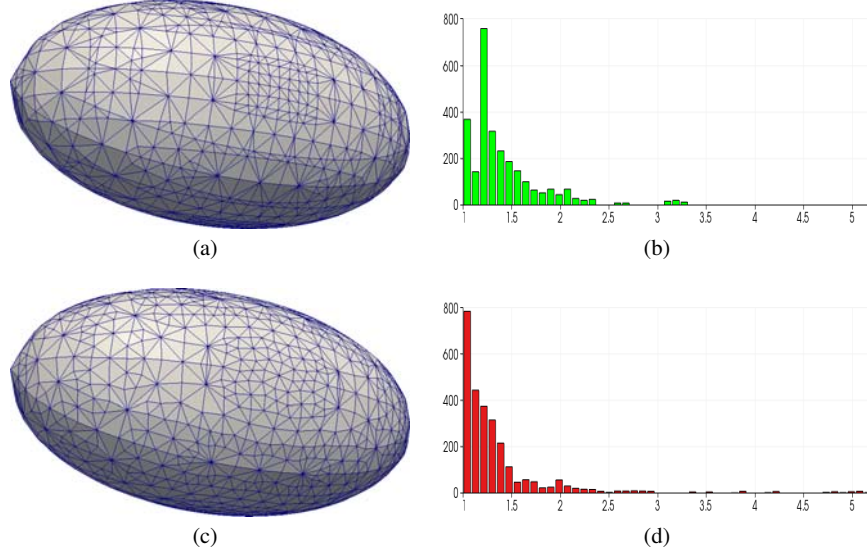


Fig. 8.10 (a) and (b): surface mesh of an ellipse and its quality measure histogram; (c) and (d): smoothed surface mesh and its quality measure histogram.

It is noted by the histograms in Figure 8.10 that the number of elements with a radius ratio close to 1 increased considerably in the mesh smoothed by the Algorithm 2.

8.4.6 Volumetric mesh smoothing by virtual springs

The algorithm described above works for two-dimensional meshes consisting of triangles and three-dimensional meshes consisting of tetrahedrons. In both cases the boundary vertices are not visited by the algorithm. In this work we use the semi-torsional and lineal springs to run the smoothing algorithm on volumetric meshes. Thus, the spring stiffness κ_{ij} used in the Algorithm 1 is the same as described in equation (8.4).

The Figure 8.11 presents a surface mesh of a ellipse smoothed by the process described in the Algorithm 1. In each figure we show the quality histograms of

mesh elements measured by the metric “minimum angle” that calculates the lowest internal dihedral angle of each element.

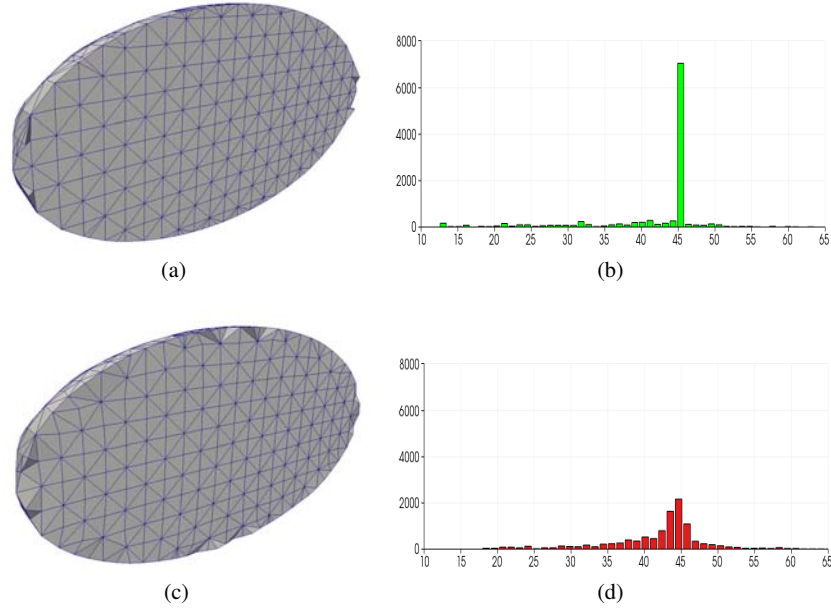


Fig. 8.11 Ellipse model. (a) initial volumetric mesh; (c) the mesh smoothed by the Algorithm 1.

By observing the histograms of Figure 8.11 it is noted that the initial mesh contains most of its elements with dihedral angle close to 45° . In the mesh smoothed by the Algorithm 1, the elements distribution became more uniform, with most of the elements between 35° and 50° . However, the smallest dihedral angle found in the mesh of Figure 11(a) is 12.5° whereas the smallest dihedral angle found in the smoothed mesh of the Figure 11(c) is 17.9° .

8.5 Results

The purpose of our study is the three-dimensional mesh generation for fluid flow simulations in arteries obtained from patient-specific medical images. For this type of application, our approach does not take into account sharp features of the models, considering there are few models that require treatment in this regard. Generally, for fluid flow simulations in arteries the surface of models is smooth.

In our study of mesh generation for medical applications, we used particularly geometries of arterial segments. In this case, the intent is to use the generated meshes for simulating blood flow in these arteries.

The meshes used as starting point for the improvement strategies in this work were obtained by three-dimensional mesh generator presented in [7]. There, the volumetric mesh is obtained with the ideas presented in [17], however, using the J_1^a triangulation presented in the section 8.2, as spatial decomposition, and uses a Partition of Unity Implicit method created by Gois [13, 14] to generate the implicit function of the model.

In [7] the mesh is obtained from a point set in space that represents the model. This set of points is called point cloud. A point cloud can be obtained through 3D scanner or, in our case, from medical imaging (Magnetic Resonance Imaging).

In our study the point cloud used to generate the initial volumetric meshes were obtained from surface meshes. From each surface mesh we take only the points, discarding the other elements of the mesh, and getting a point cloud. A implicit function is created from the point cloud through the Partition of Unity Implicit method in [14]. The implicit function of level 0 represents the surface of the model. From the implicit function we obtain a discretization of the surface by analyzing the function at the tetrahedra vertices of the J_1^a triangulation which decompose the domain. The elements of J_1^a triangulation in the interior of the model are used as the internal elements of the resulting volumetric mesh. However, the elements of J_1^a triangulation affected by the intersection with the implicit function of level 0 must be analyzed and regenerated so that all the final mesh only contains tetrahedra. This process is done in [7] in accordance with the ideas presented in [17].

The mesh improvement created by Klingner and Shewchuk in [16] is used in our studies because it is the state-of-the-art in volumetric mesh improvement. The tool created by them is called *Stellar* and integrates the mesh improvement techniques presented in section 8.3 to generate an improved mesh. In our work we combine the *Stellar* tool with our smoothing technique through virtual springs, getting some interesting results.

The diagram of Figure 8.12 shows the pipeline used for obtaining each mesh presented here.

The examples presented in this section were obtained from the models which were called “renal artery”, “abdominal aortic”, “basilar artery”, “iliac aneurysm”, “abdominal aortic aneurysm”, provided by the project HeMoLab (Hemodynamics Modeling Laboratory) at LNCC - National Laboratory for Scientific Computing.

In order to better compare the results, the meshes were obtained with the same parameters for the J_1^a triangulation. Only in the *abdominal aortic aneurysm* model

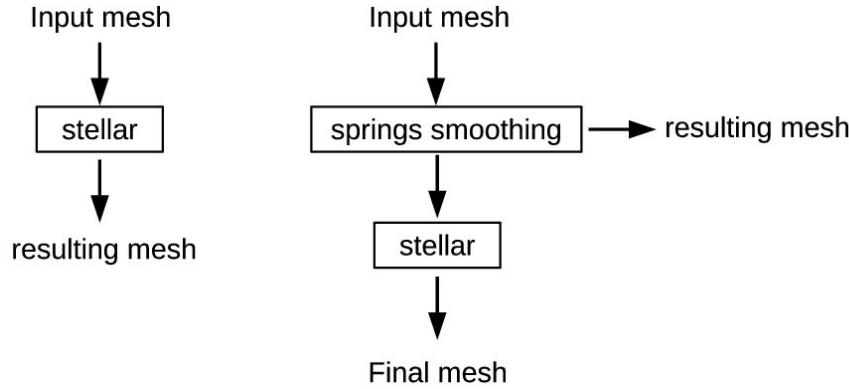


Fig. 8.12 Diagram of execution of tests to obtain the mesh presented in this section.

the maximum refinement level of the J_1^a triangulation is 3, in the other models presented here this parameter is equal to 4.

The Figures 8.13, 8.14, 8.15, 8.16 and 8.17 show the meshes in the following order: (a) complete model; (b) zoom of the initial mesh; (c) zoom of the initial mesh improved by the *Stellar* tool; (d) zoom of the initial mesh improved by our smoothing method through virtual springs and; (e) zoom of the mesh in the Figure (d) improved by the *Stellar* tool.

The table 8.1 shows the minimum and maximum dihedral angles for each model in each mesh improvement strategy, namely, the input mesh used, the mesh improved by the *Stellar* tool, the mesh smoothed by applying virtual springs and a mesh obtained with the *Stellar* tool after smoothing through virtual springs.

	input mesh		<i>Stellar</i>		<i>smoothing</i>		<i>Stellar + smoothing</i>	
	<i>min</i>	<i>max</i>	<i>min</i>	<i>max</i>	<i>min</i>	<i>max</i>	<i>min</i>	<i>max</i>
<i>renal artery</i>	7.4117	162.19	14.96	158.84	7.2185	164.8	17.988	154.57
<i>abdominal aortic</i>	7.0378	161.08	12.597	154.87	1.7102	177.37	16.927	155.11
<i>basilar artery</i>	7.219	162.49	14.492	156.17	7.1605	163.08	19.124	153.32
<i>iliac aneurysm</i>	6.544	160.52	15.545	155.72	3.7584	169.55	16.881	153.24
<i>abdominal aortic aneurysm</i>	6.7657	160.21	19.818	151.06	3.8039	167.03	22.836	147.89

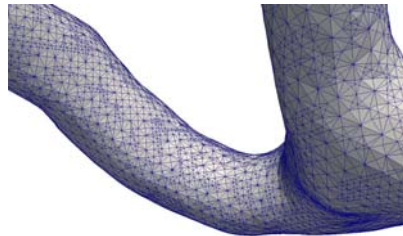
Table 8.1 Meshes quality. The values in the tables correspond to minimum and maximum dihedral angles found in each of the listed meshes.

The Figure 8.18 shows a cut view of the *abdominal aortic aneurysm* model to visualize the internal elements of the mesh for each step illustrated in the diagram of Figure 8.12.

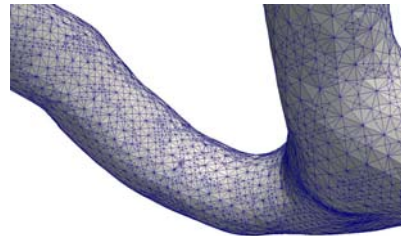
We can see from this figure that near to the model surface the mesh is thinner (with smaller elements) and coarser in the interior of the model (with larger elements). This is due to adaptivity provided by J_1^a triangulation.



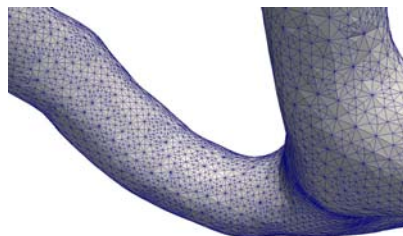
(a)



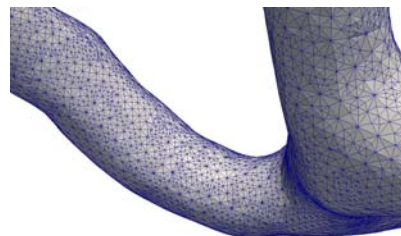
(b)



(c)



(d)



(e)

Fig. 8.13 *Renal artery model.*

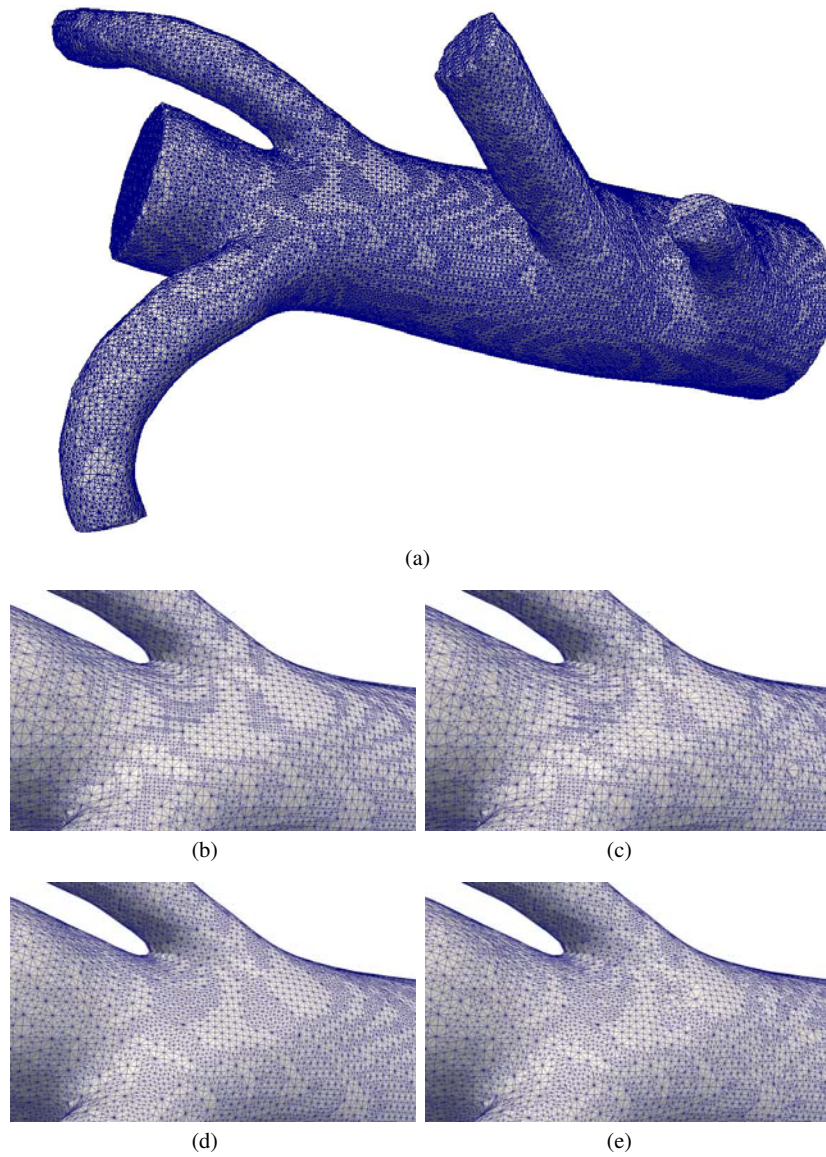


Fig. 8.14 *Abdominal aortic model.*

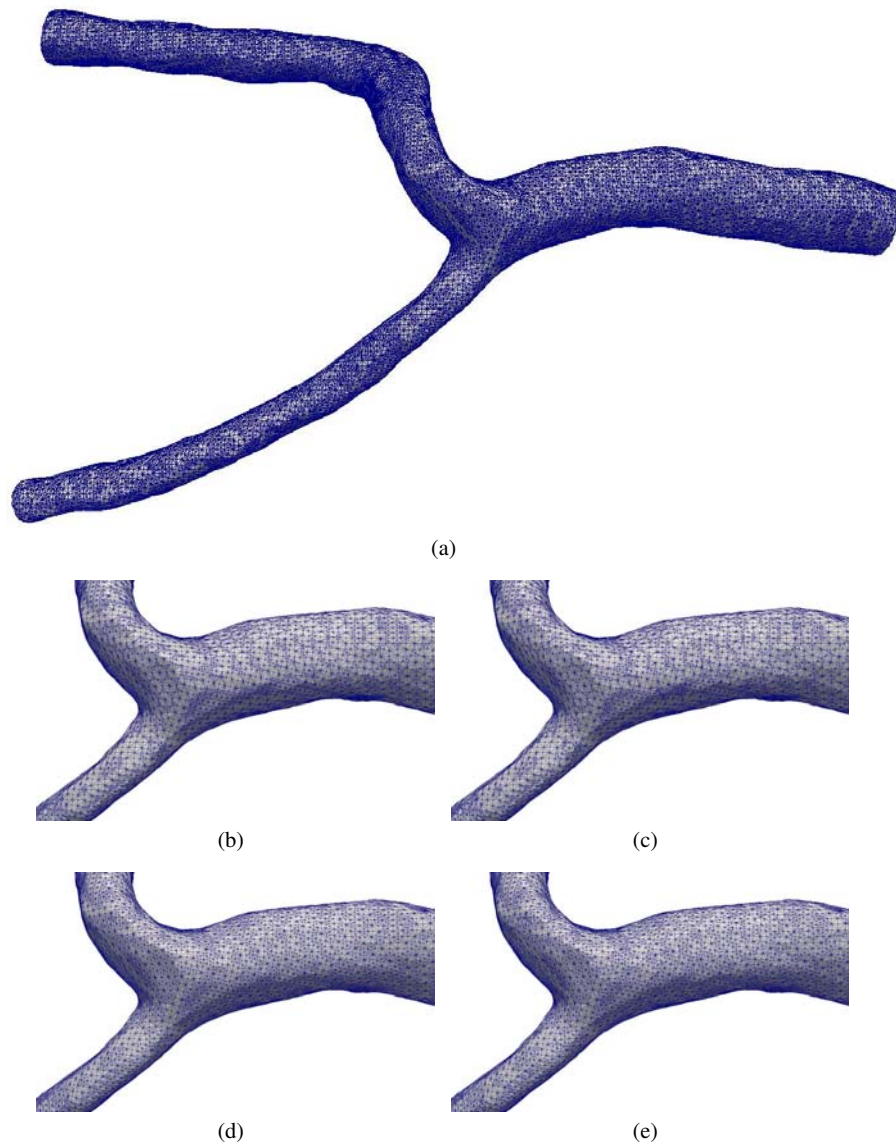


Fig. 8.15 *Basilar artery model.*

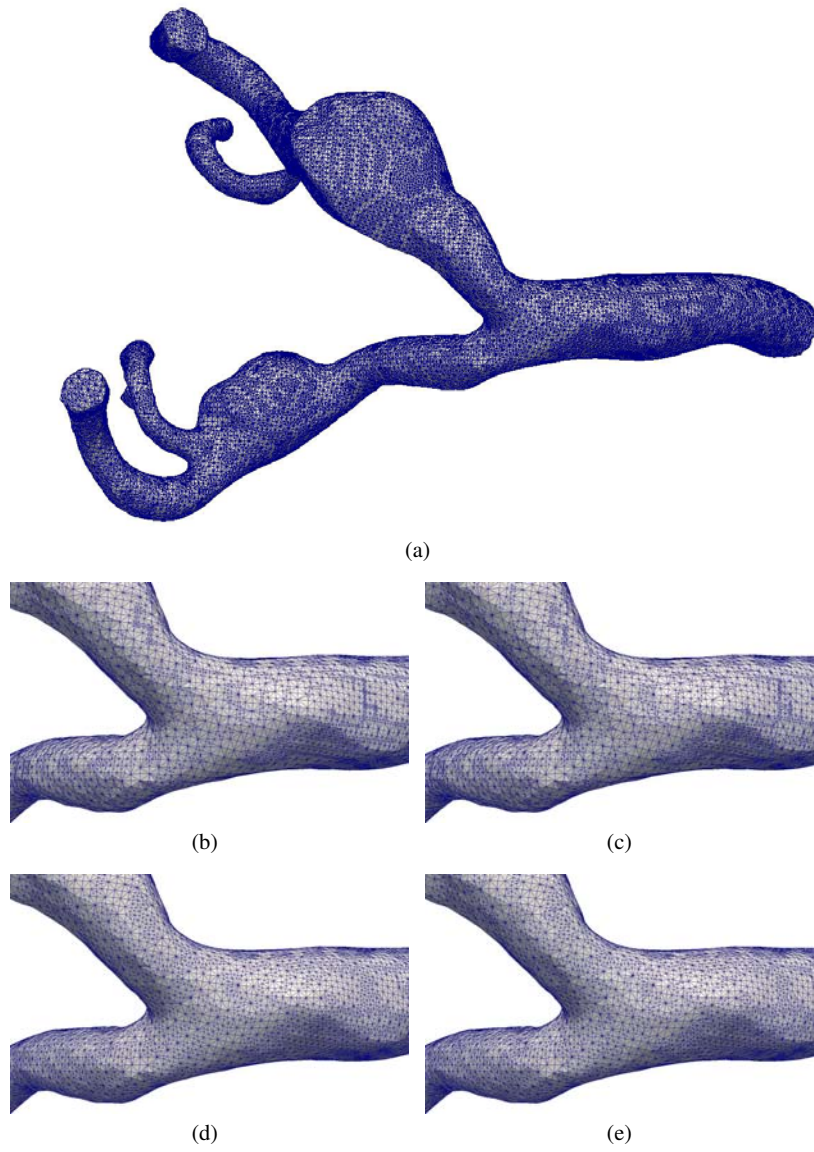


Fig. 8.16 *Iliac aneurysm* model.

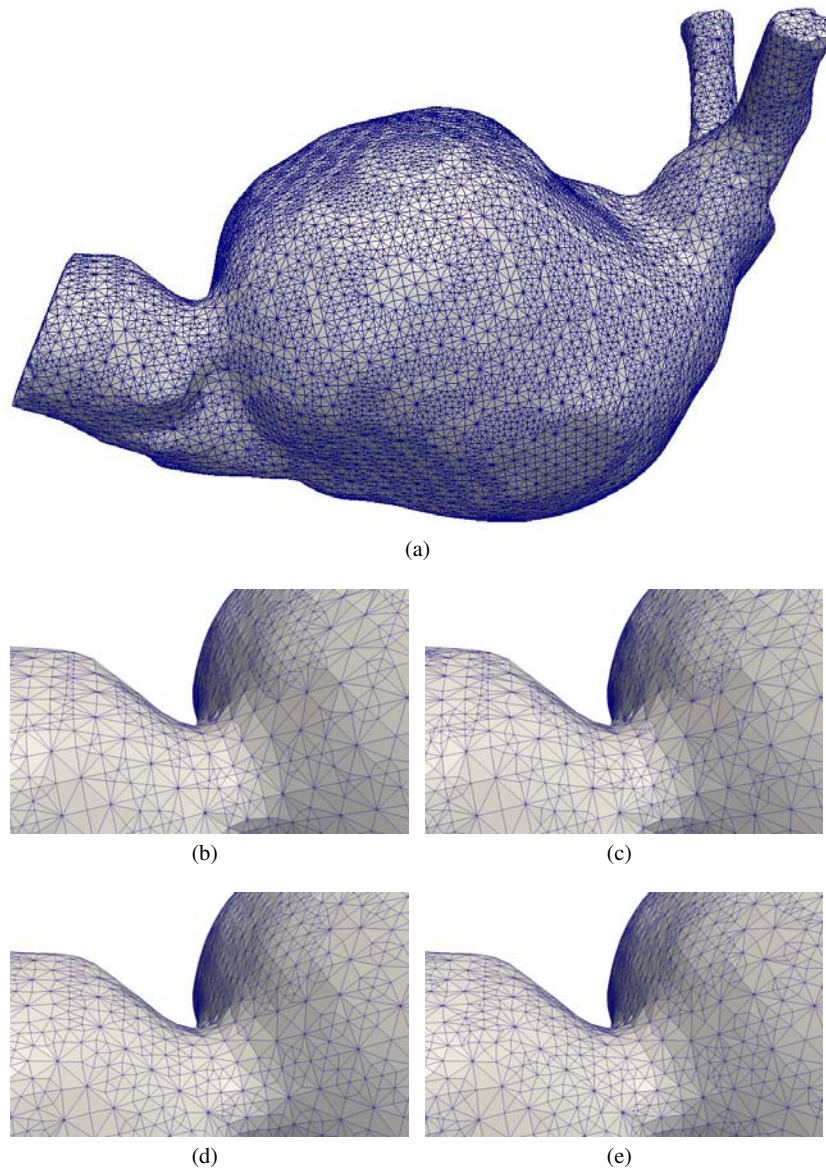


Fig. 8.17 *Abdominal aortic aneurysm model.*

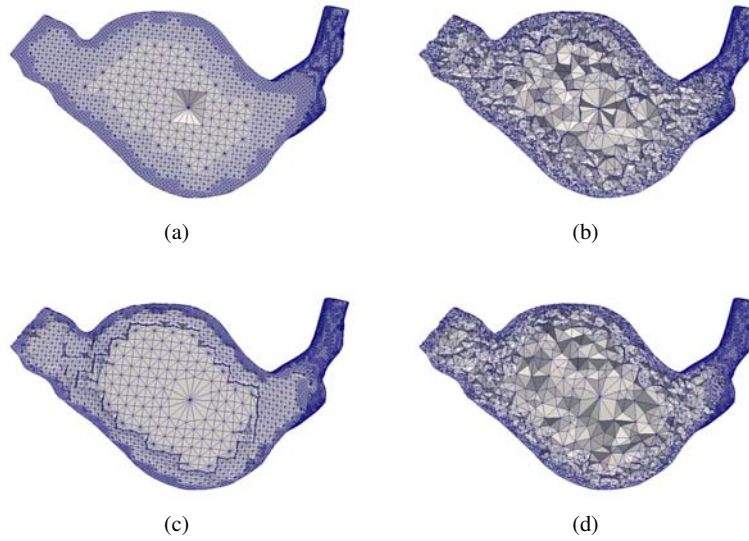


Fig. 8.18 A cut view of *abdominal aortic aneurysm* mesh to visualize the internal elements of the mesh. The Figure (a) is the input mesh; the Figure (b) is the input mesh improved by the *Stellar* tool; the Figure (c) is the input mesh improved by smoothing through virtual springs and the Figure (d) is the mesh of the Figure (c) after the improvement made by the *Stellar* tool.

By observing the figures of all models, we note that the surface elements in the mesh smoothed with our strategy presents a more uniform aspect, although there are neighbors elements with a visible size difference. Besides, we can note that *Stellar* introduces few visual changes on the mesh smoothed with our strategy.

However, looking at table 8.1 we note that *Stellar* improved significantly the smallest and largest dihedral angle present in the mesh. In this same table (table 8.1) we note that the smoothing process through virtual springs applied in the input mesh does not bring many advantages, however, applying *Stellar* to the smoothed mesh we get better results than applying *Stellar* directly to the input mesh.

Although *Stellar* slightly modified surface elements, it performed a large modification in internal elements of the mesh. Observing Figure 8.18 we note that *Stellar* changed dramatically the interior of the mesh. This difference is apparent when we observe the number of elements in each mesh.

In the example shown in Figure 8.18 the number of elements in the smoothed mesh after the application of the *Stellar* tool decreased 44.6 % from the input mesh. Table 8.2 shows the number of elements for each mesh obtained in the tests herein. Unlike improvement strategy applied by *Stellar*, the strategy of smoothing through virtual springs does not change the number of elements of the input mesh.

	renal artery		abdominal aortic		basilar artery		iliac aneurysm		abdominal aortic aneurysm	
	#	%	#	%	#	%	#	%	#	%
input mesh	300823	-	750338	-	277536	-	382820	-	235208	-
<i>smoothing</i>	300823	0	750338	0	277536	0	382820	0	235208	0
<i>Stellar</i>	189257	37,1	443015	41,0	174001	37,3	234422	38,8	135467	42,4
<i>smoothing + Stellar</i>	178428	40,7	419013	44,2	156400	43,6	232110	39,4	130274	44,6

Table 8.2 Number of mesh elements. For each column the value in the left side (#) represents the number of elements (tetrahedra) in the mesh; the value in the right side (%) represents the percentage of decrease in the number of elements related to the number of elements in the input mesh.

8.6 Conclusion

In this paper we presented a technique of adaptive mesh generation, where the input data is a point cloud, and a group of techniques to improve surface and volumetric mesh. The geometrical data used here were provided by the project HeMoLab (Laboratory of Modeling in Hemodynamics). These techniques should be integrated into a framework for computer simulation of the Human Cardiovascular System [3], developed in the project HeMoLab.

The quality verification is done by measuring the smallest and largest dihedral angles found in tetrahedral meshes. However, the only method that guarantees quality delivers angles between 10° and 165° [17].

These results show that only one technique for mesh improvement is not sufficient to obtain meshes with an acceptable quality. It is better to combine techniques to improve volumetric or surface meshes, in order to generate a final mesh with good quality.

On synthetic models, we had good results in terms of adaptivity and quality of elements. However, on real models of arterial vessels we observed that, in most cases, only mesh adaptivity was achieved without any improvement in the quality of elements. The use of the method *Stellar*, developed by Klingner and Shewchuk [16], improved the mesh quality, however, when combined with the virtual springs, this improvement can reach better results.

The adaptivity of J_1^a triangulation reduces the number of elements needed to generate the volumetric mesh. The method *Stellar* reduces considerably the number of volume elements of the mesh. The use of *Stellar* in the volumetric meshes obtained by J_1^a triangulation reduced the number of mesh elements, and the adaptivity achieved by J_1^a triangulation was not lost.

Thus, for models of arterial segments, we can conclude that the use of a volumetric mesh generator based on J_1^a triangulation, with the improvement of meshes by virtual springs and *Stellar* provides an adaptive volume meshing with good quality.

Acknowledgements

The authors would like to thank Pablo Javier Blanco for his collaboration and assistance. This work was partially supported by the Brazilian agencies CNPq, CAPES and FAPESP. The support of these agencies is gratefully acknowledged.

References

- [1] N. Amenta, M. Bern, and D. Eppstein. Optimal point placement for mesh smoothing. *Journal of Algorithms*, 30(2):302–322, 1999.
- [2] J. T. Batina. Unsteady euler airfoil solutions using unstructured dynamic meshes. *AIAA Journal*, 28(8):1381–1388, 1990.
- [3] P. J. Blanco, M. R. Pivello, S. A. Urquiza, N. A. D. S. e Silva, and R. A. Feijoo. Coupled models technology in multi-scale computational haemodynamics. *International Journal of Biomedical Engineering and Technology*, 5(2):132–156, 2011.
- [4] F. J. Blom. Considerations on the spring analogy. *International Journal for Numerical Methods in Fluids*, 32(6):647–668, 2000.
- [5] A. Castelo. *Aproximao adaptativa de variedades implcitas com aplicaes na modelagem implcita e em equaes algbrico-diferenciais*. PhD thesis, Departamento de Matemática, Pontifícia Universidade Católica, Rio de Janeiro, Brasil, 1992.
- [6] A. Castelo, L. G. Nonato, M. Siqueira, R. Minghim, and G. Tavares. The j1a triangulation: An adaptive triangulation in any dimension. *Computer & Graphics*, 30(5):737–753, 2006.
- [7] Í. L. L. da Cunha, A. D. Lacassa, V. Polizelli-Junior, J. P. Gois, L. G. Nonato, and A. C. Filho. Adaptive algebraic mesh generation from implicit functions. In *Iberian Latin American Congress On Computational Methods in Engineering*, 2008.
- [8] H. L. de Cougny and M. S. Shephard. Refinement, derefinement, and optimization of tetrahedral geometric triangulations in three dimensions. 1995.
- [9] B. N. Delaunay. Sur la Sphre Vide. *Izvestia Akademia Nauk SSSR, VII Seria, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7:793–800, 1934.
- [10] C. Farhat, C. Degand, B. Koobus, and M. Lesoinne. Torsional springs for two-dimensional dynamic unstructured fluid meshes. *Computer Methods in Applied Mechanics and Engineering*, 163(14):231 – 245, 1998.
- [11] D. A. Field. Laplacian smoothing and delaunay triangulations. *Communications in Applied Numerical Methods*, 4(6):709–712, 1988.
- [12] L. A. Freitag and C. Ollivier-Gooch. Tetrahedral mesh improvement using swapping and smoothing. *International Journal for Numerical Methods in Engineering*, 40:3979–4002, 1997.
- [13] J. P. Gois, V. Polizelli-Junior, T. Etienne, E. Tejada, A. Castelo, T. Ertl, and L. G. Nonato. Robust and adaptive surface reconstruction using partition of unity implicits. In *SIBGRAP '07: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing*, pages 95–104. IEEE Computer Society, 2007.
- [14] J. P. Gois, V. Polizelli-Junior, T. Etienne, E. Tejada, A. Castelo, L. G. Nonato, and T. Ertl. Twofold adaptive partition of unity implicits. *The Visual Computer*, 24(12):1013–1023, Dezembro 2008.
- [15] G. Klintsek. Minimal triangulations of polygonal domains. In P. L. Hammer, editor, *Combinatorics 79*, volume 9 of *Annals of Discrete Mathematics*, pages 121 – 123. Elsevier, 1980.
- [16] B. M. Klingner and J. R. Shewchuk. Agressive tetrahedral mesh improvement. In *Proceedings of the 16th International Meshing Roundtable*, pages 3–23, october 2007.
- [17] F. Labelle and J. R. Shewchuk. Isosurface stuffing: fast tetrahedral meshes with good dihedral angles. *ACM Trans. Graph.*, 26(3):57, 2007.

- [18] C. L. Lawson. Software for C^1 surface interpolation. *Mathematical Software III*, pages 161–194, 1977.
- [19] S. H. Lo. A new mesh generation scheme for arbitrary planar domains. *International Journal for Numerical Methods in Engineering*, 21:1403–1426, 1985.
- [20] R. Löhner. Extensions and improvements of the advancing front grid generation technique. *Communications in Numerical Methods in Engineering*, 12(10):683–702, 1996.
- [21] R. Löhner and E. Oñate. Advancing front techniques for filling space with arbitrary separated objects. *Finite Elem. Anal. Des.*, 46(1-2):140–151, Jan. 2010.
- [22] R. Löhner and P. Parikh. Generation of three-dimensional unstructured grids by the advancing-front method. *International Journal for Numerical Methods in Fluids*, 8(10):1135–1149, 1988.
- [23] V. N. Parthasarathy and S. Kodiyalam. A constrained optimization approach to finite element mesh smoothing. *Finite Elem. Anal. Des.*, 9(4):309–320, 1991.
- [24] W. J. Schroeder and M. S. Shephard. A combined octree/delaunay method for fully automatic 3-d mesh generation. *International Journal for Numerical Methods in Engineering*, 29(1):37–55, 1990.
- [25] M. S. Shephard and M. K. Georges. Automatic three-dimensional mesh generation by the finite octree technique. *International Journal for Numerical Methods in Engineering*, 32(4):709–749, 1991.
- [26] J. R. Shewchuk. *Delaunay Refinement Mesh Generation*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, May 1997. Available as Technical Report CMU-CS-97-137.
- [27] J. R. Shewchuk. Two discrete optimization algorithms for the topological improvement of tetrahedral meshes. In *Unpublished manuscript*, 2002.
- [28] J. R. Shewchuk. What is a good linear finite element? interpolation, conditioning, anisotropy, and quality measures. In *Eleventh International Meshing Roundtable*, pages 115–126, 2002.
- [29] A. A. Shostko, R. Löhner, and W. C. Sandberg. Surface triangulation over intersecting geometries. *International Journal for Numerical Methods in Engineering*, 44(9):1359–1376, 1999.
- [30] I. P. Soares. *Movimento de malhas e remalhamento de malhas superficiais*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2007.
- [31] D. F. Watson. Computing the Delaunay tessellation with application to Voronoi polytopes. *The Computer Journal*, 24:167–172, 1981.
- [32] M. Yerry and M. Shephard. A modified quadtree approach to finite element mesh generation. *IEEE Computer Graphics and Applications*, 3(1):39–46, 1983.
- [33] M. A. Yerry and M. S. Shephard. Automatic three-dimensional mesh generation by the modified-octree technique. *International Journal for Numerical Methods in Engineering*, 20(11):1965–1990, 1984.
- [34] D. Zeng and C. R. Ethier. A semi-torsional spring analogy model for updating unstructured meshes in 3d moving domains. *Finite Elem. Anal. Des.*, 41(11-12):1118–1139, June 2005.
- [35] T. Zhou and K. Shimada. An angle-based approach to two-dimensional mesh smoothing. In *Proceedings, 9th International Meshing Roundtable*, pages 373–384, 2000.

Chapter 9

Virtual Reality Applied to Medicine

Jauvane C. de Oliveira, Selan R. dos Santos, Bruno M. de Carvalho, Liliane dos S. Machado, Ronei M. Moraes, Rosa M. E. M. da Costa, Vera M. B. Werneck, Márcio S. Pinho, Fátima L. S. Nunes, Hélio Pedrini, Ildeberto Rodello, Marcio E. Delamaro and Romero Tori

Jauvane C. de Oliveira

LNCC - National Laboratory of Scientific Computing and INCT-MACC, Av. Getúlio Vargas 333, Quitandinha, 25651-075, Petrópolis, RJ, Brazil, e-mail: jauvane@lncc.br

Selan R. dos Santos and Bruno M. de Carvalho

DIMAp - Department of Informatics and Applied Mathematics, UFRN - Federal University of Rio Grande do Norte, Campus Lagoa Nova, 59072-970, Natal, RN, and INCT-MACC, Brazil, e-mail: {selan,motta}@dimap.ufrn.br

Márcio Sarroglia Pinho

Faculdade de Informática da PUCRS Av Ipiranga, 6681, Prédio 32. Sala 607, 90.619-900, Porto Alegre, RS, and INCT-MACC, Brazil, e-mail: pinho@pucrs.br

Liliane dos S. Machado and Ronei M. Moraes

Laboratory of Technologies for Virtual Teaching and Statistics (LabTEVE) Federal University of Paraíba (UFPB) Cidade Universitária s/n 58051-900 João Pessoa, and INCT-MACC, Brazil e-mail: liliane@di.ufpb.br, ronei@de.ufpb.br

Rosa M. E. M. da Costa and Vera M. B. Werneck

DICC - Departamento de Informática e Ciência da Computação IME - Instituto de Matemática e Estatística UERJ - Universidade do Estado do Rio de Janeiro Rio de Janeiro, RJ, Brazil e-mail: {rcosta,vera}@ime.uerj.br

Fátima L. S. Nunes

Health Informatics Applications Laboratory, School of Arts, Sciences and Humanities, University of São Paulo, and INCT-MACC, Brazil, e-mail: fatima.nunes@usp.br

Hélio Pedrini

Informatics Visual Laboratory, University of Campinas, and INCT-MACC, Brazil, e-mail: helio@ic.unicamp.br

Ildeberto Rodello

Ribeirão Preto School of Economics, Administration and Accounting, University of São Paulo, Brazil e-mail: rodello@fearp.usp.br

Marcio E. Delamaro

Software Engineering Laboratory, Institute of Mathematical and Computer Sciences, and INCT-MACC, Brazil, e-mail: delamaro@icmc.usp.br

Romero Tori

Interactive Technologies Laboratory, Polytechnic School and LPAI - Research Lab in Interactive Environments, Senac University Center, and INCT-MACC, Brazil, e-mail: tori@usp.br

Abstract Virtual Reality (VR) has become an invaluable asset in supporting medical applications. The simultaneous stimulation of multiple sensory channels—such as vision, touch, and sound—typically found in VR applications, is responsible for engaging users within a lively and reactive three-dimensional virtual environment. In this type of environment it is possible to control the users' exposition to simultaneous stimuli. This situation is ideal for supporting, for instance, an immersive medical training system, a visualization room, or even a rich virtual rehabilitation environment fully controllable by a therapist. The INCT-MACC's Virtual Reality Group is currently investigating innovative ways of applying VR to assist medical applications. Our research ranges from the design of new VR hardware, fundamental algorithms and techniques to support medical research and applications, new methodologies to assess VR systems, the design of collaborative VR systems, to the development of fully functional VR systems that supports a variety of medical applications in training, rehabilitation and visualization, just to name a few. In this chapter we present our contribution to these fields, focusing on both the fundamental finding in VR field as well as describing the most important system prototypes developed so far.

9.1 Introduction

In this chapter we introduce a number of Virtual and Augmented Reality prototypes recently developed within the National Institute of Science and Technology of Medicine Assisted by Scientific Computing. Such prototypes have been developed in various research laboratories, namely the ACiMA/LNCC Lab., LAPIS/USP Lab., LabTEVE/UFPB Lab., LVRV/UFRN Lab., GRV/PUCRS Lab. and LATVI/UERJ Lab. We further introduce each research lab in the following paragraphs.

The Applied Multimedia and Virtual Environments Laboratory (ABoVE, aka ACiMA) started activities in May 2003. It concentrates the research efforts on Collaborative Virtual Environments, Virtual Reality and Multimedia Systems at the National Laboratory for Scientific Computing (LNCC), a research institute that belongs to the Brazilian Ministry of Science, Technology and Innovation.

LAPIS (acronym in Portuguese for Health Informatics Applications Laboratory Laboratório de Aplicações de Informática em Saúde) is a research laboratory located at the School of Arts, Sciences and Humanities of the University of São Paulo. Its focus is virtual reality, content-based image retrieval (CBIR), and image processing techniques in the health area.

LabTEVE is a research laboratory that belongs to the Statistics Department at the Federal University of Paraíba (UFPB). It started activities in 2000, concentrating efforts in the Virtual and Distance Learning areas.

LVRV (Visualization and Virtual Reality Laboratory) is a research laboratory at the Federal University of the Rio Grande do Norte. Its research ranges from fundamental algorithms to support medical research and application, navigation issues in virtual reality environments to serious game for rehabilitation. It has strong interdis-

ciplinary research collaboration with, for example, the department of neuroscience and education.

The Virtual Reality Group (GRV) carries out research in Virtual Reality at the Pontifical Catholic University of Rio Grande do Sul (PUCRS).

The Laboratory of Virtual and Interactive Technologies LATVI was created to support research activities focusing on the Mathematics, Statistics, and Computer Science and experimentations of Software Engineering, Complex Systems, Parallel and Distributed Computing, Virtual Reality, Neuroscience, Ubiquitous Applications, among others. The inter-institutional projects from this laboratory promote interdisciplinary cooperation with physicians, nutritionists, biologists, physicists and professionals from various fields of Engineering. This research lab is in the State University of Rio de Janeiro.

9.1.1 Virtual Reality in Medicine

Virtual Reality (VR) has been broadly used in training applications in general. It is an idea medium for that, as it allows a user to be immersed into a customized Virtual World that allows such user to gain useful experience related to similar situations in real life. A flight simulator is a very good example of that, as a pilot gain expertise and hours of flight in simulators. That both allow one to improve his/her skills without risking his/her life or equipment, as well as allows one to go through a standard set of simulations, gaining experience even in very rare cases, which is at times impossible in real life. In Flight simulators that would include awkward weather conditions, whilst in the medical context it could be rare medical conditions or cases. The same way a pilot earns experience in a simulator, a physician can gain experience in a simulation, so that when he/she is to deal with a real patient, such experience allows a better experience for the patient, the same way in a flight simulator for a pilot who then goes to fly the real plane.

Other than training, VR has also been quite useful for physical rehabilitation. Boring and repetitive exercises can become exciting and involving. Physiotherapy, which requires some specific movements to be performed by a patient that is recovering from some trauma, can be greatly improved. If the virtual environment is appealing enough for the patient, he/she will be a lot more willing to perform the required tasks, especially if we consider youngsters. A game-like setup can mask what would be otherwise tedious movement repetitions.

Since 1992 a conference called “Medicine Meets VR” has been bringing together the two areas, with a lot of results published. Major VR journals, such as MIT Press’ “Presence – Teleoperators and Virtual Environments” often have special issues in VR Medical applications and technologies, as that is a quite appealing theme for the VR community (and vice-versa).

VR systems consist of a simulation that exploits the human senses to provide a rich and realistic experience to the user. Such experience is, ideally, equivalent to that the user would have in a similar setup in real life. That includes a 3D imagery

displayed, positional sound and often some tactile setup as well. Some works already exploit taste and smell, but traditionally sight, hearing and tactile interfaces are more ubiquitous. In order to provide such rich experience, VR systems rely on specialized hardware. Tactile feedback is especially important when the system simulates some medical procedure, such as surgeries or exams. The benefits of haptic systems can be observed in situations in which the comprehension of a problem depends on or are complemented by tactile sensations of objects or scenes [25, 20]. Touch and manual dexterity to perform procedures are also factors that have promoted the development of techniques and processes involving haptics.

9.2 Fundamentals

In this section, we discuss fundamental algorithms, techniques, VR setups and concepts that are later used, or have the potential to be used, in various prototypes, reason why they are considered to be the foundation for some prototypes.

9.2.1 Collaborative Systems

Performance evaluation of Distributed Haptic Virtual Environments (DHVEs) became important to understand the new Internet requirements for supporting multi-sensorial and real-time collaborative applications. In this context, experiments were implemented to evaluate the real conditions of CyberMed [19] when executed over a non-dedicated hybrid network, like the Internet, comparing its results with other similar works found in the literature.

It was performed a representation of haptic networked applications according to well-defined methodology which involved four sequential phases. Such phases are: 1) the requirements analysis; 2) the simulator selection; 3) learning a simulation tool and the CybCollaboration's (CyberMed collaboration module) protocol implementation, and; 4) simulation and analysis.

The CybCollaboration module was added to the CyberMed architecture. It allows programmers to build a complete medical CVEs with just a few lines of code. Among several functions, the main roles of this module are: a) initializing the CVEs visualization; b) manage the exchanging of messages between all participants through a collaborative protocol; and c) it enables users to utilize VR interactive devices during the collaborative task. The module support CVEs with multiple users collaborating with each other in different communication approaches (one-to-one, one-to-many, many-to-many). The most trivial way of collaboration permits users to freely manipulate the virtual objects with the aid of the mouse device. Another option is the tutoring collaboration (assisted collaboration) [26], implemented by the `CybHapticAssistedCollaboration` class, where one host is designated as a guide and its haptic device must guide the remote users movements by

their own devices [23]. Such particular way of collaboration is frequently used by systems with educational purposes where the tutor assumes the leading role within the one-to-many communication approach.

The OPNET Modeler (OM) was chosen to implement the network behavior of the CybCollaboration setup. OM is a simulator of discrete events which allows the creation of models that properly simulate the network behavior of new architectures, network protocols and alike. The chosen simulation tool also enables the use of Probability Density Functions (PDFs) for producing and setting up different kinds of data traffic patterns such as those common in multimedia systems, as well as typical VR applications. In the CybCollaboration's simulation experiment, it was created a node responsible for generating haptic data traffic and for processing all packets according to CybCollaboration's protocol.

The types of DHVE explored by the created scenarios utilizes both unidirectional and bidirectional tele-haptics architectures. Therefore, in the unidirectional tele-haptics, the data flows are always transmitted from just one source node and received by the remote clients, and the bidirectional tele-haptics enables that all nodes transmits and receives the data flows in a concurrent manner (Fig. 9.1). The unidirectional tele-haptics are used, for instance, in medical simulators in which there is a professional of a medicine area acting like a guide to train remote students in a medical procedure, such as blood collection applications that needs the use of haptic devices. The bidirectional tele-haptics are used in collaborative environments where all hosts are able to perform modifications on it.



Fig. 9.1 Concurrent collaborative application using the CybCollaboration module.

There is still not a consensus in studies related to quantifying or qualifying network requirements such as scalability and reliability for collaborative medical simulators [34, 23]. These types of requirements are still not quite well defined and can change according to specific experiments and studies performed. Thus, it is important the use of mechanisms to provide better QoE for CVEs, especially those for medical simulations.

In order to obtain better haptic transmissions in the CyberMed framework, was implemented peer-to-peer unicast and multicast architectures. It was observed that the requirement to allow a programmer to develop CVEs with few lines of code was still achieved, and CybCollaboration is extensible, allowing one to add further features.

A comparative performance evaluation study between the use of multicast and unicast protocol verified that the traffic of peer-to-peer unicast increases exponentially as N becomes higher, overloading bandwidth as well as imposing higher processing requirements at final hosts. Thus, based on the results obtained and comparing them with the network requirements for haptic transmission found in the literature, was observed that the CybCollaboration's protocol can satisfactorily support the development of CVE with unicast and multicast protocols.

9.2.2 Online Assessment

The training assessment based on VR consists on systems which must continuously monitor all user's interactions and compare their performance with predefined expert's classes of performance to recognize users level of training. Nowadays those systems are evolving of seminal works proposed by Dinsmore [8] in the 90's. Two kinds of training assessment can be found in the literature: single user's assessment system (SUAS) and multiple users' assessment system (MUAS). Basically, SUAS can be divided in off-line and on-line. Off-line SUAS can be defined as methods not coupled to VR systems, whose assessment results are provided some time (which can be minutes, hours or days) after the end of the VR-based training. On the other hand, on-line SUAS are coupled to the training system and collect user data to provide a result of their performance at the end of the simulation [13]. An on-line SUAS works coupled to a VR simulator, as showed in the Fig. 9.2.

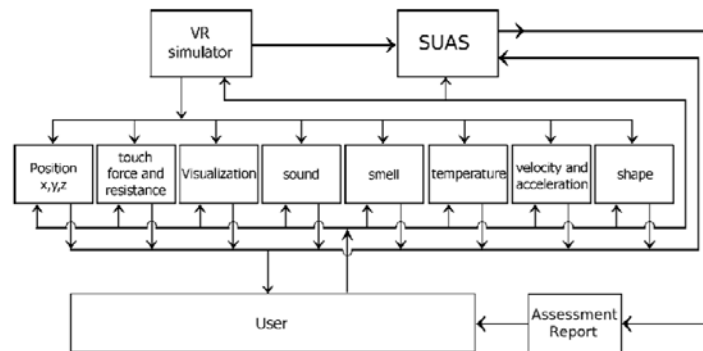


Fig. 9.2 Diagram of a VR simulator with a SUAS, Adapted from [13].

A SUAS should be capable to monitor user's interactions with the VR simulator by variables such as position (of user and of instruments used in the VR simulation), touch force and resistance and user's angle of visualization, among others. All the information are sent to the SUAS which analyzes the data and emits, at the end of the training, an assessment report about the user's performance according predefined classes of performance. An on-line SUAS must have low complexity so that it does not compromise VR simulations performance, but it must have high accuracy to does not compromise the assessment.

The main differences of VR training systems for multiple users from VR training systems for single users are: increase of complexity, use of clusters of computers or a computer capable to generate realistic multiple views, support to changes in virtual environments for all users, support in assessment system, speed of network for communication among computer without compromising the simulation. Eventually, more than one interaction device, as haptic devices, can be installed in a computer and/or tracking systems for each user in training. A MUAS must be interconnected with all users and must receive from them synchronized information about all variables of interest. A MUAS works as a SUAS, and it should be capable to monitor all interactions among users. In order to reach that, it is necessary to collect information about user's position in the space using tracking systems. User's interactions in the simulation should respect the physics laws and train their proper positions in the procedure. To collect some information as force, force feedback, angles and torques, it is necessary to use specific devices to provide them. This information will be collected for each user in training system, as well as for all group, to be used to feed the MUAS. Additionally, synchronization in time and space is necessary for all users to measure interactions among them, to determine the ordering of tasks and to provide details of user's performance [18]. Then, each application will have their own set of relevant variables which will be monitored for each user and also another set of relevant variables which measure interactions among users in the group. All those information is used by MUAS, which analyzes the data and emits, at the end of the training, an assessment report about the individual user's performance according pre-defined classes of performance, as well as, another assessment report about the group performance. There is no classification for MUAS as on-line or off-line, because MUAS are not be able to generate reports immediately after of the end of training session.

In addition to those two assessment types, there are two others: continuous assessment for a single user in sequences of trainings conducted and training assessment executed by Web (used for distance training or distance learning). In the first one, the main focus is use techniques of continuous evaluation to improve trainee performance. Continuous evaluation is an educational methodology used in present and distance learning to help the construction of the knowledge and the cognitive learning. The goal is to construct a profile to help users to understand their difficulties and to improve their performance. The system should be able to provide an assessment report and a continuous assessment report, showing the performance of trainee in the last training and in all trainings performed by him/her, respectively. In the second one, several measures from user interaction in training executed by

Web are used as input for an assessment system. Networks limitations should take into account to design that kind of assessment system. LabTEVE researchers have worked in the design and implementation of assessment systems (SUAS, MUAS, continuous assessment and assessment on Web systems), compatible with the CyberMed framework.

9.2.3 Visual Programming for VR

The CyberMed framework was developed to allow fast integration of virtual reality based simulators. It was designed to be used by programmers. However, was observed that professionals of the health area could develop some simple applications if they know how to use CyberMed. A solution based on previous works pointed out to the use of visual programming. Visual Programming (VP) can solve the usability problem providing improvement in interaction as much as in time reducing of the development process.

Thus, CyberMedVPS [19] was designed to be used by health professionals to allow them to develop their own VR applications. This VP system was projected to be composed by the CyberMed, as much as provide a visual interface for other frameworks for VR development, as ViMET [22]. Its conception as a VPS extinguishes any necessity of textual code input to generate results and interaction happens only in graphical mode based on flowcharts. Thus, textual programming is not necessary for generation or execution of VR applications. Through several graphical components, users can connect boxes to make a flowchart representing steps of instructions required to execute a VR application.

CyberMedVPS was designed to be composed by two layers: the graphical interface layer has all visual elements that can be manipulated by users to produce the medical applications and the integration of this layer to CyberMed or other framework depends on the communication layer that will relate the graphical elements of the flowchart to the framework commands. Thus, this communication layer allows bidirectional flow between the graphical interface and the framework.

Nowadays, the graphical interface of the system has been tested with professionals of the health area to evaluate their efficiency. In Fig. 9.3 is possible to observe the graphical elements that can be selected and added in the flowchart that can be executed directly from the visual system.

9.2.4 Low-Cost CAVE

A low-cost CAVE has been developed at the LNCC. Two setups are used with it: a) a cluster of 4 computers (each one controlling one of the three walls and floor), and b) a single computer with a nVidia QuadroPlex 2100 D4 which provides 8 video outputs. Passive circular polarizing stereoscopic projection is used, with two off-the-

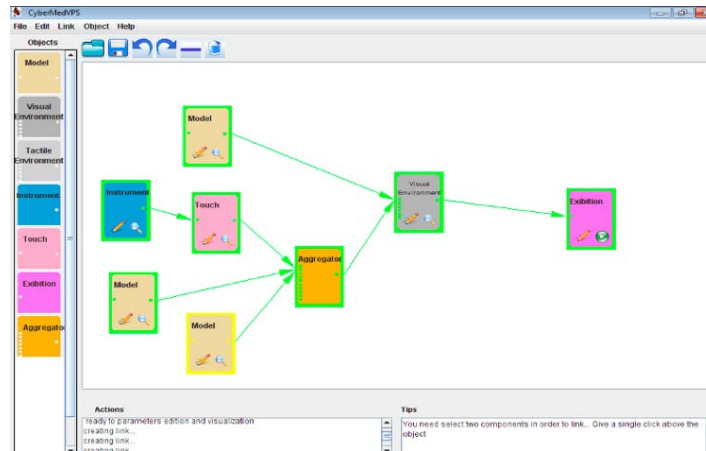


Fig. 9.3 CyberMedVPS interface.

shelf NEC LT245 DLP projectors for each wall. The CAVE structure is built out of PVC pipes. The hardware is controlled by InstantReality plus some custom plug-ins. The performance has been quite acceptable. Fig. 9.4a shows the setup, while Fig. 9.4b shows the CAVE in use.

9.2.5 Object Reconstruction

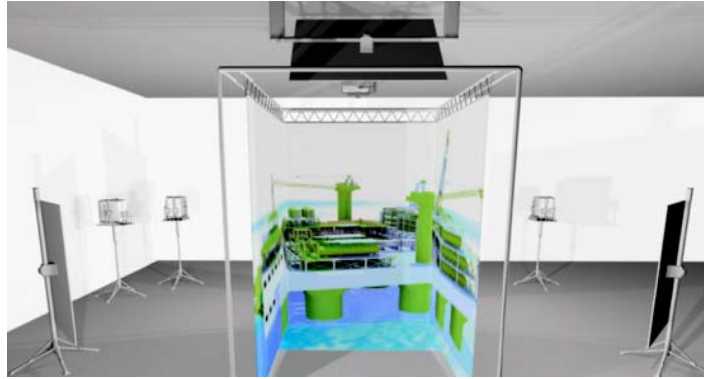
9.2.5.1 Moiré Image Analysis

The goal of this project is to infer the topography of a surface using Moiré shadows effect. The capture of these structures generates information that can be used in various fields such as medical image analysis, reconstruction of 3D objects, analysis of products in a production line, among others. In the following image we can see the projection of Moiré shadows on the face of a mannequin (Fig 9.5).

In this project the Moiré shadows are being projected on the backs of human beings in order to evaluate spinal deformities and deviations, such as scoliosis or lordosis, as can be seen on Fig. 9.6.

9.2.5.2 Texture Fuzzy Segmentation using Adaptive Affinity Functions

Digital image segmentation is the process of assigning distinct labels to different objects in a digital image. The level of detail indicated by the labeling is related to the application at hand; thus, characterizing the image segmentation process as an ill-posed problem. The task of segmenting out an object from its background in



(a) 3D model of the low-cost CAVE.



(b) CAVE displaying blood vessels with flow speed in colors.

Fig. 9.4 The CAVE immersive device.

an image becomes particularly hard for a computer when, instead of the intensity values, what distinguishes the object from the background is some textural property, or when the image is affected by noise and/or inhomogeneous illumination.

In this work we are concerned with the development of robust methods for segmenting objects with textural properties. The interactive region-growing method used to segment the images is based on the fuzzy connectedness, and it computes for every spel (short for spacial element such as pixels and voxels and every object, a number between 0 and 1, that gives the grade of membership of a spel to an object. The fuzzy affinity functions commonly used for image segmentation using this technique are replaced by adaptive functions, whose areas of characterization or scales

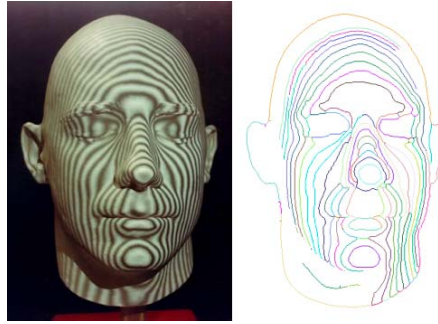


Fig. 9.5 Moiré Shadows example.

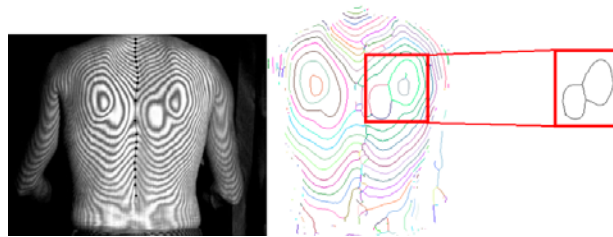


Fig. 9.6 Moiré Shadows on human back.

are defined according to the statistics collected in the areas surrounding the seed spels selected by the user. An algorithm devised for selecting this scale was presented [3], as well as the usage of new affinity functions, that, instead of using the intensities of the spels involved in the computation, uses information (intensities) collected over a neighborhood area defined by the scale chosen by our method.

An example of the applications of this method can be seen on Fig. 9.7, where an image of a skin lesion was segmented using adaptive fuzzy affinity functions.

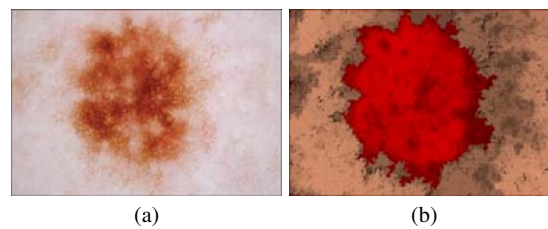


Fig. 9.7 Dermoscopic image of a lesion (a) and a segmentation of it (b) produced by using adaptive fuzzy affinity functions.

9.2.5.3 Fuzzy Segmentation-based Transfer Functions for Volume Rendering

The task of creating transfer functions for volume rendering can be seen as similar to producing the segmentation of an image in the sense that one wants to assign optical properties to voxels that share certain characteristics, with the aim of identifying or visually emphasizing some objects.

The idea of this work is to produce transfer functions from the connectedness maps produced by the fuzzy segmentation algorithm [4]. The connectedness maps, besides coding the pertinence of the voxels, give information about the grade of membership of the voxels, i.e., how likely is that one particular voxel belongs to one or more objects.

By using these connectedness maps, we give users more power to create different transfer functions, using the pertinence information as well as the grade of membership to compute the opacity and color characteristics of the voxels. Fig. 9.8 shows the volume rendering of a tooth produced using this method.

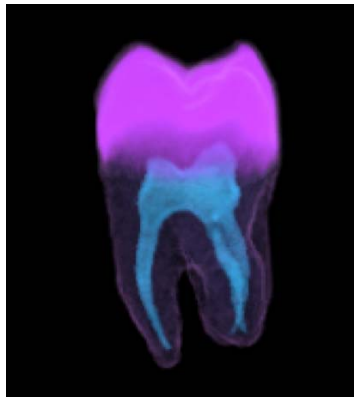


Fig. 9.8 Volume rendering of a tooth generated with fuzzy segmentation-based transfer functions.

9.2.5.4 ViMeT Framework

The *ViMeT* (Virtual Medical Training) Framework is an open source Virtual Reality object-oriented framework to generate tools for simulating biopsy exams [22]. ViMeT provides functionalities and a wizard tool to create applications for simulating generic biopsy exams. Among the available functionalities are automatic creation of a virtual environment, stereoscopy, collision detection with accuracy, deformation and support for non-conventional devices. Examples of applications generated by using ViMeT are presented in Figure 9.9. ViMeT Framework is available at <http://www.each.usp.br/lapis/>.

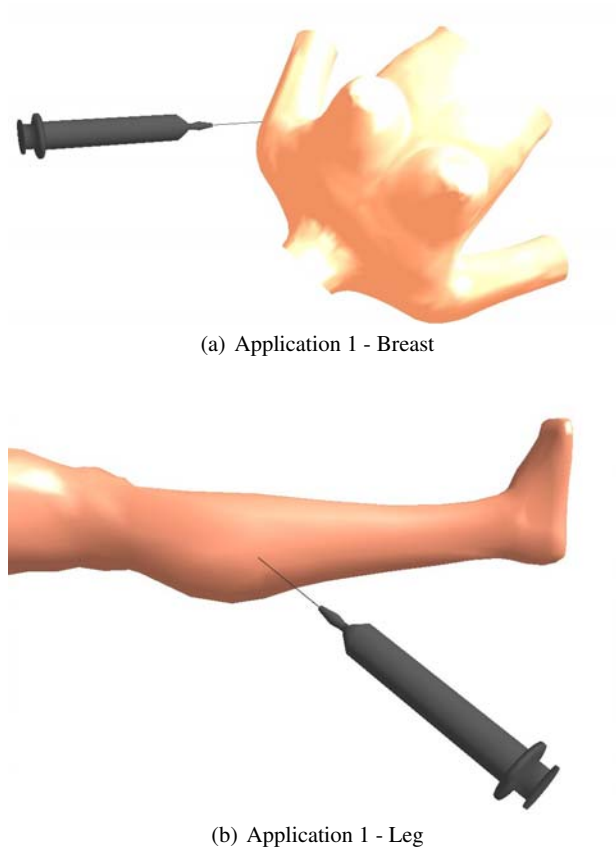


Fig. 9.9 Examples of applications generated by *ViMeT* Framework.

Some current works related to this project are: realism in relation to visualization, methods for providing precise collision detection in real time, realistic deformation of complex objects, use of serious games to evaluate students and definition of methodologies for evaluating knowledge and abilities acquisition by users when trained in three-dimensional virtual environments.

9.2.5.5 Content-based Image Retrieval

Content-based Image Retrieval (CBIR) is a relatively new topic in computer research and is defined as any technology that helps to organize digital image files based on their visual content. In a general way, the studies on CBIR address mainly the problem of querying, in a database, images which are similar to a given reference image provided as a model. The criteria for similarity are obtained by extracting fea-

tures of the image, such as color, texture and shape. The automated CBIR systems involve several areas of computer science, mainly image processing and databases. In a simple way, this type of system is essentially composed of four parts: extractors, similarity functions, indexing structures and sorting algorithms.

In this context, LApIS is developing techniques to retrieve medical images considering two-dimensional images and three-dimensional models. Also researches to evaluate the behavior of similarity functions are been conducted, as well as the application of CBIR concepts to evaluate computer-aided diagnosis systems. The main product in this research line is related to the use of CBIR in the software testing of programs with graphical outputs [7], which produced a framework named O-FIm (Oracle For Images), following the open source paradigm¹.

The mentioned framework helps the tester generate an application (Java program) that is able to compare two images (usually stored in files) responding whether they are similar or not, according to the chosen characteristics. Also to facilitate the creation of an oracle, the O-FIm framework has a wizard tool that generates the textual description of the oracle. This is a graphical interface that allows the tester's intervention to create a description as the one presented in Fig. 9.10. In the figure, the tester creates an oracle which can decide whether a segmentation algorithm, applied to mammographic images behaves as expected.

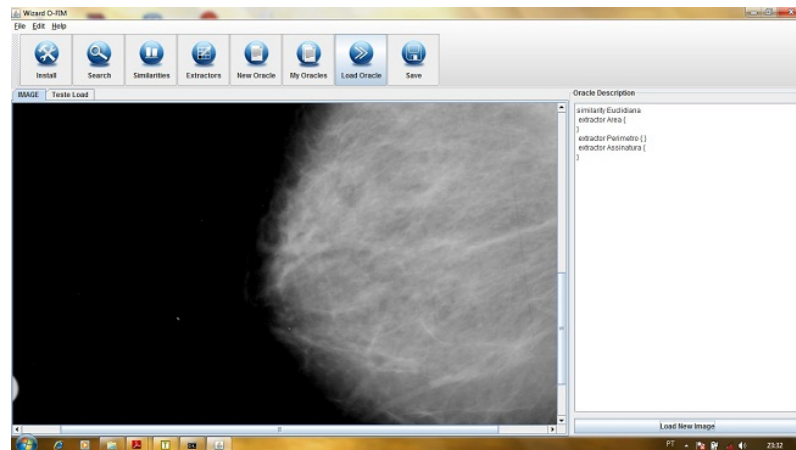


Fig. 9.10 Graphical Interface for *O-FIm Framework*.

9.2.5.6 Virtual Environments Evaluation

The research line of virtual environments evaluation considers both technical aspects of the tools and the user interaction. Along this line, assessment models

¹ <http://ccsl.icmc.usp.br/pt-br/content/O-FIm-oracle-images>

are been developing in order to measure knowledge and skill acquisition in three-dimensional virtual environments.

A theoretical model for the learning evaluation that can be applied in three-dimensional learning virtual environments (LVEs), as well as a proposal for its computer architecture are being developed. The intention is providing a way to investigate the contribution that LVEs effectively provide to learners' knowledge acquisition and how this learning process can be evaluated [21].

Studies on the contributions from the sensorimotor skills (SMS) are essential to the medical training area. From this consideration, a initial Systematic Review was conducted in order to retrieve the state of the art in relation to methodologies to evaluate the skill acquisition and also identify metrics that are used to evaluate strategies of SMS acquisition, both in real environments as well as in virtual environments. After identifying these metrics, an analysis was developed to categorize them according to their similarities and application context. This research is still in an initial phase, and aims to develop a model to acquire data from virtual environments and use them to evaluate the user performance [1].

A last theme related to user evaluation in virtual environments is the use of serious games. An initial application was implemented (Fig. 9.11) in order to investigate how much playful aspects can help users to learn concepts or improve certain skills. The game uses the ViMeT Framework and includes playful aspects like score, soundtrack, difficulty levels and ranking of players. The idea is to try to relate the player's score with the level of learning provided by the virtual environment.

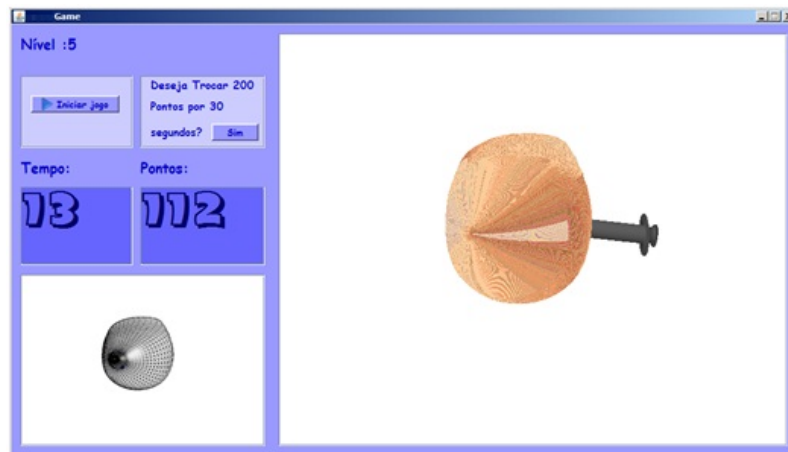


Fig. 9.11 Example of a serious game interface used to evaluate student's learning.

9.3 Prototypes

In this section we describe a number of prototypes developed throughout the last few years by the research laboratories listed in Section 11.1. Such prototypes range from training VR systems through VR-based rehabilitation setups.

9.3.1 MiniVR — Portable VR System

This prototype aims at allowing users which would not otherwise have access to 3D content, to take advantage of interactive 3D applications [12]. Considering the way it was designed, most applications developed to work in the CAVE system also work in the MiniVR system.

The MiniVR consists of an aluminum box with approximately 30 cm on each side. It contains everything that is needed to project 3D content on a screen. The idea is that just a power cable connection is required to have all working. Internally the MiniVR has a processing unit, two compact projectors, polarizing filters as well as all internal connections and cooling solution. Fig. 9.12 shows the MiniVR system, including a quick glance on its internal components.

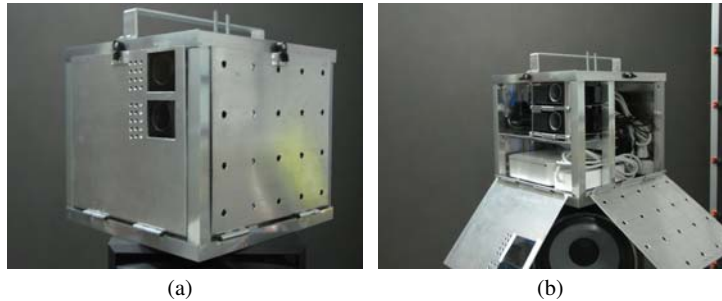


Fig. 9.12 MiniVR system (left) and internal components (right).

MiniVR allows VR applications to be taken where the users are, rather than requiring the users to move to a installation which would provide 3D interaction. The unit weights 8Kg (17.5 pounds). The image resolution is of 1024×768 pixels. The MiniVR works best in an environment which is not too bright, in which case the image can be enlarged considerably. As the environment gets brighter the scene needs to be smaller (projectors closer to the screen) so that the content is bright enough to be properly seen.

We use passive stereoscopy, more precisely circular polarizing filters and matching glasses. The two projectors project images addressed to either the right or left eye. The polarization allows each eye to see only its own image, which provides

the brain with two separated views of the 3D scene, which is automatically used to position the objects in the 3D space. Fig. 9.13 shows a 3D scene being generated by MiniVR. The scene needs to be projected in a surface which keeps the polarization of the light. We use a piece of Stewart Silver Screen (Fig. 9.13 left), but any surface that maintains the polarization is useable.

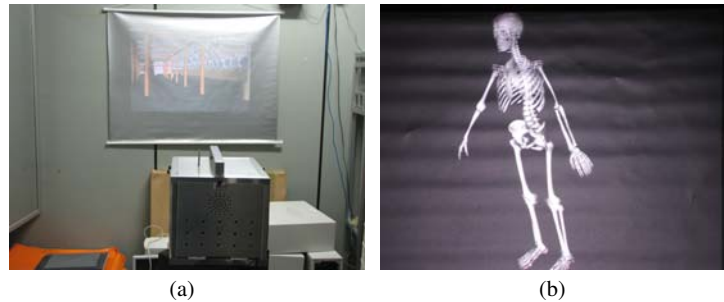


Fig. 9.13 MiniVR working (left) and detail in dark environment (right).

The MiniVR system uses X3D based on InstantReality, which is the same setup used in the CAVE at LNCC. For interaction, we often use a wireless keyboard and mouse, a wireless joystick Logitech Rumblepad. When a more precise control is called for we use a (wired) 3D mouse (3D Connexxion Space Navigator). The setup has been demonstrated to the public in two opportunities: a) FAPERJ's Science, Technology and Innovation Fair (Rio de Janeiro, RJ) – FAPERJ is the State of Rio de Janeiro Research Support Agency, and b) SBPC ExpoT&C Conference and Exposition (Goinia, GO) – SBPC is the Brazilian Society for the Advancement of Science.

We are designing a yet smaller unit, with built-in power, which is to increase the portability even further.

9.3.2 Automatic Evaluation of the Quality of Cardiopulmonary Resuscitation Procedure

The training of Cardiopulmonary Resuscitation (CPR) is often performed in physical manikins which mimic the behavior of the human body. One example of such models is the Laerdal Little Anne (Fig. 9.14). The manikin allows one to practice the CPR procedure, getting used to the pressure to be applied. One may manually calculate the frequency used in the procedure, but the system itself does not do it.

We have added to this training setup some hardware and software that automatically measures three parameters: a) The pressure the user makes, with a historic in a timeline graph; b) The frequency in the procedure, with a historic in a timeline graph, and c) The angle at which the pressure is being made. One shall remember

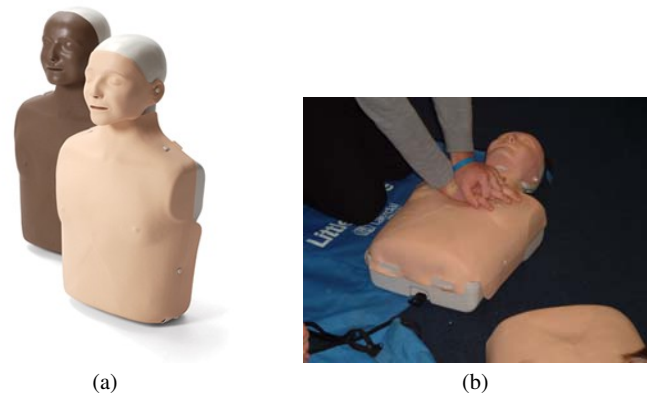


Fig. 9.14 CPR manikin Laerdal Little Anne (left) and CPR Procedure (right).

that the proper procedure uses pressure which is normal to the surface, i.e. it is performed with the pressure down, without inclination. There is also a historic timeline graph which documents this data.

The system has been implemented using a Nintendo Wii Fit Plus scale, which is placed under the manikin. The Wii Fit scale is connected to a computer through Bluetooth. As the pressure is performed in the manikin, the weight is distributed in the corners of the scale. If the pressure is properly applied the weight is uniformly distributed. If it is performed at an angle the corners of the scale will indicate it. The system shows the information in an easy to check interface (Fig. 9.15).

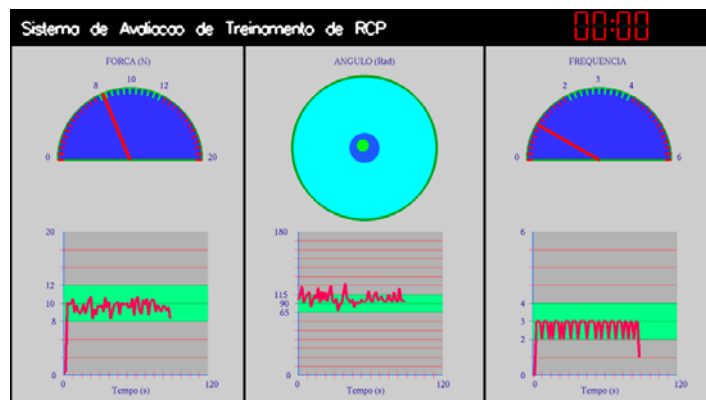


Fig. 9.15 CPR training interface: pressure, angle and frequency charts.

The interface is composed of three sections. The section to the left has a dial which shows the pressure applied at that instant. The green level displays the acceptable values. The section to the right has a dial which shows the frequency at

which the massage is being applied, with the values in the green range showing the ideal frequency. Finally, the section in the center has two concentric circles. If the massage is made properly, the mark stays within the smaller circle, whereas if the massage is being made at a wrong angle, the mark will move to the external circle and eventually outside of it (proportional to the angle applied).

At the lower part of all sections we have a graph that shows how the user has done throughout the procedure. By looking at the graphs shown, one can quickly see how efficient the procedure was performed, if it kept performance consistency or if it faded as time went by (due to tiredness). Fig. 9.16 shows the system being used in the FAPERJ Science, Technology and Innovation fair.

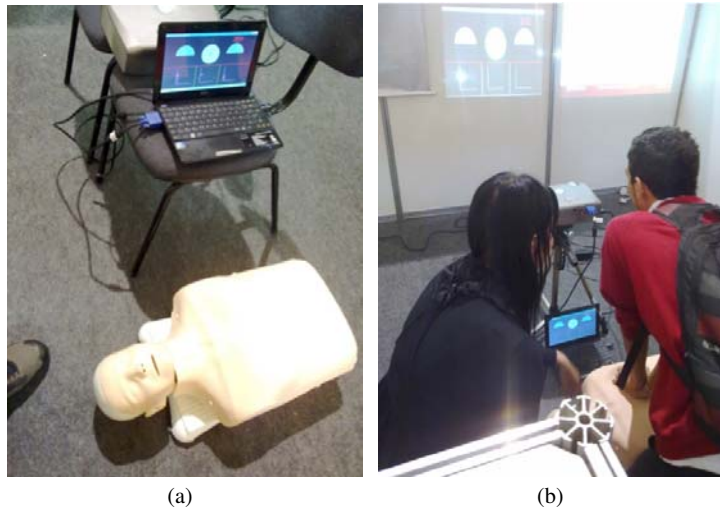


Fig. 9.16 CPR system setup (left) and system being used (right).

This prototype has been shown to many individuals who actually perform CPR in their daily tasks. The results are very positive so far. We expect to add to the system a way to record the data for later evaluation of ones performance.

9.3.3 *Virtual Grasping in the CAVE*

This prototype aimed at allowing a user to grasp a virtual object with his/her own hand to move it around in the CAVE. The virtual world consists of a 3D surgery room (see Fig. 9.17 left) in which we have a set of items on a table. Using a Cyber-Glove Wireless II (Fig. 9.17 right) the user can reach out and grasp an object. The system detects the collision of the users hand with the virtual object and allows the user to push the virtual object, as well as to grasp it and move it around.

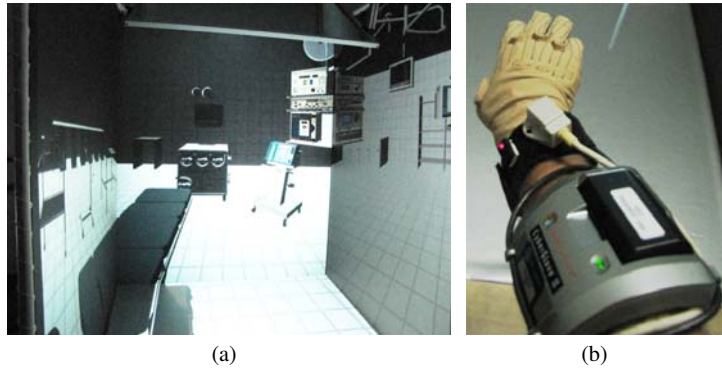


Fig. 9.17 Virtual surgery room (left) and CyberGrasp Wireless II (right).

The system also makes use of a magnetic position tracker (Ascension Nest of Birds). One sensor is attached to the user hand so that the system knows where the hand is located as well as the movements performed (through the data glove). The CyberGlove allows the system to know the configuration of the user's hand, which includes the grasping and release movements. All motions performed by the user are mimicked by a virtual hand model which is within the virtual world.

This system enables fine interaction between a real user and the virtual world. Fig. 9.18 shows a hand that represents the user's hand approaching a virtual model of some objects (left) as well as the user grasping a bottle (right) and moving it. The bottle moves along with the user hand as it would in real life.



Fig. 9.18 Hand approaching virtual object (left) and grasping it (right).

9.3.4 AVIDHa – 3D Haptic Anatomy Atlas

The application, called AVIDHa is a 3D human body atlas for the purpose of anatomy study [16]. AVIDHa allows students to interactively explore several human body systems through the senses of touch and stereoscopic vision. The human body systems are available as high definition 3D models with photo-realistic textures, as shown in Fig. 9.19.



Fig. 9.19 AVIDHa main interface.

The application allows the anatomy student to fly through and move inside of the human body. The flight and exploration modes are done with either 3D mouse or a joystick. The student may also choose to investigate each system separately, change organ's opacity to visualize internal parts, capture screenshots for later examination, manipulate clipping planes to explore the inner parts of a given system, or even use a haptic device, such as the Sensable's Phantom Omni, to feel the organ's density and contours.

AVIDHa may also run as a distributed collaborative application, allowing users geographically apart to one another to interact through the network. In this case a mediator may drive the simulation and share his/her knowledge with other participants. The system also works in the CAVE. AVIDHa was developed using a rendering engine developed at the ACiMA Lab. Such engine, called EnCIMA (An Engine for Collaborative and Immersive Multimedia Applications), will not be discussed here, but can be found in [14, 9, 15]. A WebGL version of the atlas is under development. Such setup will bring comprehensive anatomical information to a web enabled interface.

9.3.5 Head Mounted Display Based Medical Application

This prototype aims at allowing a user to view a 3D scene through a Head Mounted Display (HMD) and interact with such world. Fig. 9.20 shows a user wearing the Sensics xSight 3120 panoramic HMD. We can notice that the users vision is blocked by the HMD. The user only sees what is displayed in the HMD screens.



Fig. 9.20 User wearing the Sensics xSight HMD.

In this application the user sees a surgery room, with a patient lying in the surgery table. Fig. 9.21 shows the interface the user sees. There is a number of displays with information controlled by the application. A menu allows a user to set transparency in the patient model, so that internal structures can be seen (Fig. 9.22). The user can then move a needle around and apply it to the patient. A message in one of the virtual world screen shows the target reached by the needle.

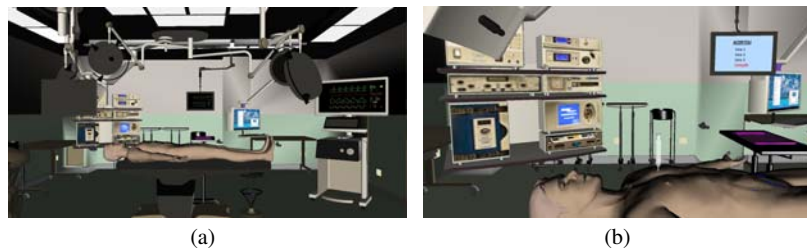


Fig. 9.21 User interface seen within the HMD.

This is an early stage of a prototype which shows some functionality that we expect to use in a virtual surgery application.

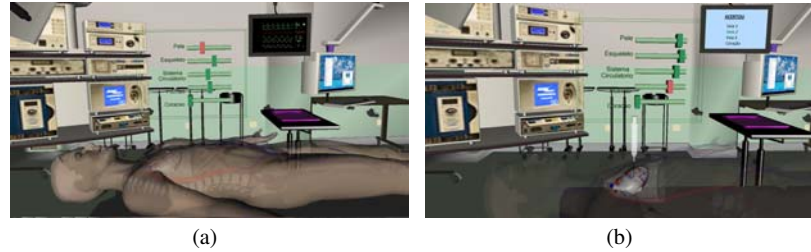


Fig. 9.22 Setting transparency in the model. Internal structures in display.

9.3.6 Volumetric-Based Simulators

In order to allow volumetric approaches, an expansion of the Cybermed has been proposed. Since all data related to the models used in simulation are managed by a data structure (DS) in the Core module, the replacement or expansion of this DS must be considered to support cut and suture in a realistic way.

Online seminars occurred between UFPB and UP (University of Porto) to discuss previous works related to volumetric manipulation for simulation of medical procedures. This discussions resulted in the visit of the researcher PinJing Xia, from the University of Porto, to the LabTEVE where he stayed in the month of March, 2011. During this period, a researcher from Federal University of Rio Grande do Norte (PhD candidate) integrated the group and started the migration of the CyberMed data structure.

The preliminary results obtained allow the use of volumetric models in VTK format with CyberMed. After that, new additions to the CyberMed framework were integrated to compose a new version. In Fig. 9.23 it is possible to observe a volumetric model visualized in an application developed with CyberMed. Those results allowed to start a discussion related to methodologies for cutting and suture in medical simulators. Video-conferences among UFPB and FEUP have occurred in order to study and discuss previous works related to this topic.

9.3.7 Evaluating Dental Drilling Procedure

In order to help the evaluation of the learning process of dental drilling procedures we have developed, in cooperation with the School of Dental Medicine of PUCRS,

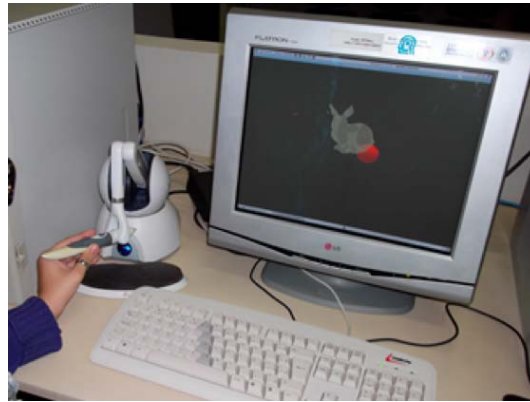
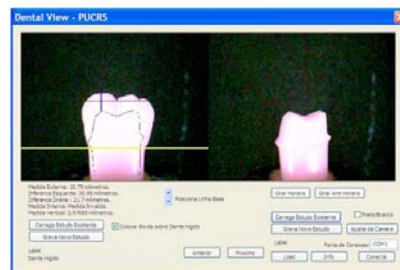


Fig. 9.23 Application developed with CyberMed using a volumetric model (bunny) and haptics for visualization and tactile feedback.

an image processing tool (hardware and software) for facilitating the students to visualize their class work.

The system captures teeth images from many angles and presents these images to the student, allowing him to see in details how was the performed drilling. Besides that, the system can measure the drilled teeth to provide more accurate feedback. Fig. 9.24 shows the evaluation software.



(a)



(b)

Fig. 9.24 Dental Evaluation Tool.

9.3.8 Pre-Surgical training system

Medical education has undergone major changes along the years. Until the mid 19th-century, academic medicine was based on the observation of facts; possible treatments were derived from this observation. Currently, the teaching of Medicine in-

volves the continuous search for new methods to accommodate novel requirements. Parallel to this search, one of the biggest challenges for the teaching of Medicine is the evolution of technology. In surgery, for example, the advent of laparoscopic techniques has translated into the need for abilities that are quite different than those applied in conventional procedures, and the training of surgeons has relied increasingly more on virtual reality (VR) tools.

Considering this context, the objectives of the present work were to develop a VR environment for the teaching of surgery at the undergraduate level, to reflect on the impact of this type of tool for the education of medical students and to consider the feasibility of establishing a graduate program focused on the development VR environments. For that, a multidisciplinary team was formed, and a pre, trans, and postsurgical VR environment was developed. This environment allows students who have never been in contact with a surgical unit to complete various tasks relating to the performance of a thoracotomy and to learn about the rules and routines associated with a surgical unit.

A preliminary assessment with 15 medical students and 12 professionals (five physicians, five computer scientist, and two education specialists) showed that both these groups considered the overall virtual experience as satisfactory or very satisfactory (scores 7–9 in a 10-point scale). The VR environment developed in this study will serve as a basis for other applications, such as additional surgical modules to replace the thoracotomy module. The environment may also be adapted for the training of other target publics, such as nurses or nursing students, nursing assistants or others. Despite the complexities associated with the development of VR tools, the undeniable need to provide students with more opportunities for training, the inexorable ingrowth of technology into medicine, and the importance of integrating all this into an opportunity for education, lead to the conclusion that this project was successful and that the research line established with the present work is extremely promising.

The pictures in Fig. 9.25 show the virtual environment created to the project.



Fig. 9.25 Virtual Environment for Medical Training.

9.3.9 Image Segmentation Systems

In this context we have worked in two different projects. The first one deals with semi automatic tool for teeth segmentation in CT images. The main goal in this case is to allow a three-dimensional reconstruction of tooth structures. Using the developed system it is possible to do automatic and manual segmentation from DICOM images and generates as output a 3D file. This file can be loaded in CAD (Computer Aided Design) systems that are able to create the three-dimensional models from plane sections. Fig. 9.26 shows some images from the system.

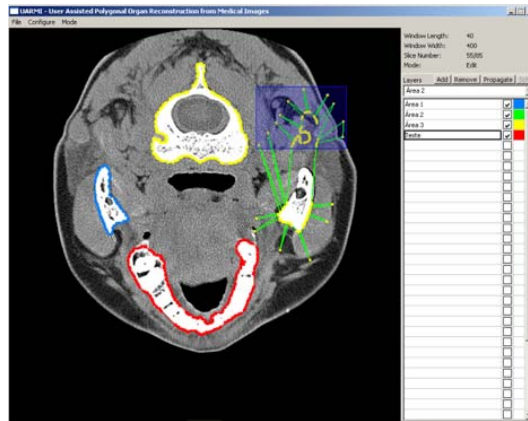


Fig. 9.26 Tooth segmentation system.

The second project in this subject deals with lung segmentation and automatic detection of lung tumors. Lung diseases victimize a huge amount of people by year. Manual diagnosis requires a vast number of tomography images analysis, by radiologists. An alternative to this is the computer's aid to this large scale of analysis, possibly related to subjectivity due to emotional factors. The computer provides a "second-opinion" to the radiologist, notifying him in case of something suspicious in the analysis. The present work proposes a computational system that uses co-occurrence matrixes descriptors, for lung diseases aided detection. Using these information, the system is instructed with the values of the intervals of these descriptors, for each pattern category provided by the doctor. These values consist in the basic rule to define if a lung region is highlighted or not, in a thorax image. In the final tests execution, the capacity of the system, as a proposal of aided detection of lung diseases, was evaluated, and the results were satisfactory. Fig. 9.27 shows the some images from the system.

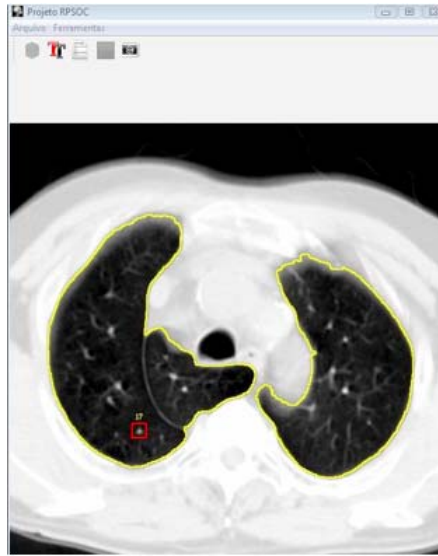


Fig. 9.27 A screenshot of the lung segmentation system.

9.3.10 Medical Emergency Training

The area of Medical Qualification in Life Support training is being constantly improved. However, many problems still have to be faced in the training sessions. During these sessions the students or physicians can repetitively practice patient care procedures in simulated scenarios using anatomical manikins, especially de-signed for this type of training. Current manikins have several resources incorporated to allow and facilitate qualified training, such as pulse, arrhythmia and auscultation simulator. However, some deficiencies have been detected in the existing LS training structure. For example: automatic feedback to the students in consequence of their actions on the manikin, images like facial expressions and body injuries, and their combination with sounds that represent the clinical state of the patient.

The main goal of the ARLIST project is to qualify the traditional training environment currently used for LS training, introducing image and sound resources into the training manikins. Through these features we can simulate some aspects such as facial expressions, skin color changes and scratches and skin injuries through image projection over the manikin body, and also play sounds like cries of pain or groans of an injured man [24]. Fig. 9.28 shows some images from the ARLIST manekin.

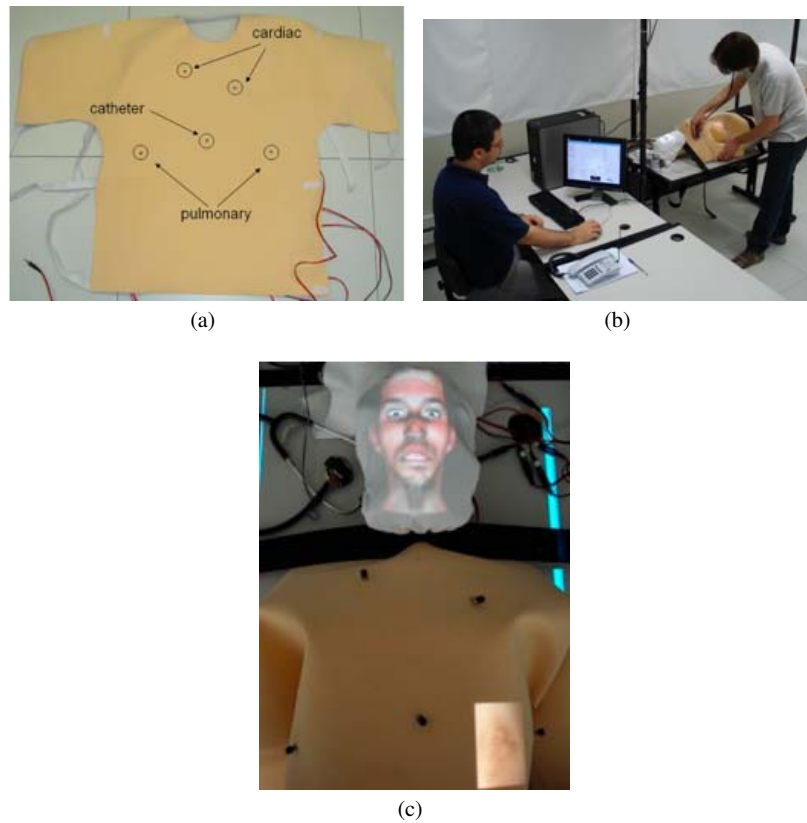


Fig. 9.28 The ARLIST Manikin.

9.3.11 Multi-touch Immersive Volume Visualization Device

This project developed a study of interaction tasks contained in medical imaging visualization systems. Based on this study, a device was developed for interaction in medical imaging visualization virtual environments. This device combines the advantages of direct mapping of the actions of the user to interact in a virtual environment with a method for input values through a multi-touch sensor. As a result, despite the users spend more time to complete the tasks in the virtual environment, in terms of accuracy the device was as effective as the desktop interface [32] (Fig. 9.29).

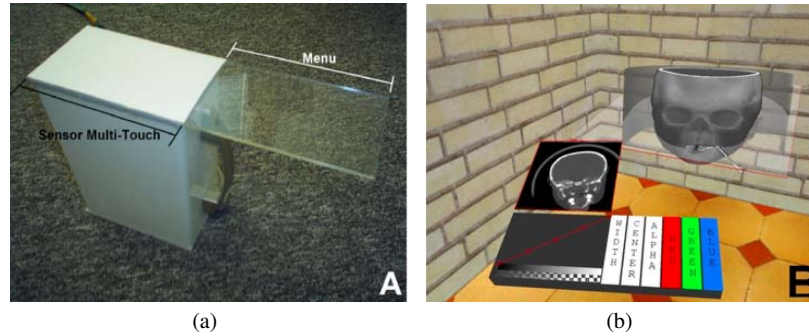


Fig. 9.29 Multi-touch device and its graphical representation inside the immersive environment.

9.3.12 Virtual Reality Game in Cognitive Rehabilitation

Recently, Virtual Reality (VR) technology has been widely used in the Cognitive Rehabilitation (CR) area, obtaining positive results [10, 11, 17]. The introduction of VR in the CR area is considered as a new opportunity to diminish the difficulties involved in the CR process, making the associated activities more motivational and fun.

In many applications there is a strong dependence on the therapist's intervention to control the patient's navigation, to change the level of difficulty of a task, to go back to an earlier stage, or to control the sequence of activities in a synthetic environment. The LATVI research group has developed some experiments to integrate intelligent strategies in the VR environments in order to control the users' navigation, changing the level of difficulty of a task and proposing tasks in an automated way. The inclusion of intelligent agents in the virtual environments could reduce the need of therapists involvement in the activity that can improve treatment. The therapists can assist the activity without interaction.

An agent can be considered an autonomous system seeking different ways to reach pre-established goals in a real or virtual environment [33]. Every agent must have autonomy, which means that an agent must have the ability to manage its internal state and its actions to achieve their goals without human intervention. A multi-agent system is composed of two or more agents who have a set of skills and plans.

The first prototype developed in this context is a serious game which aims at improving the cognitive functions of patients with neuropsychiatric disorders [5]. The game integrates agents and Virtual Reality technologies. It controls the patients right and wrong answers, changing the difficulty level of the tasks according to the user performance. Fig. 9.30 shows an example of the game scene.

The first version of this game was implemented to validate the requirements and verify the game usability with an expert in the CR area. Agents control the levels of difficulty of the proposed activities.



Fig. 9.30 Example of the Smeck-3D scene. Game stimulates cognitive functions such as attention and memory.

The games are distributed into two rooms. The first game has bookshelves (Fig. 9.30), with different objects on them. Objects are shown on a table randomly and the patient must walk to a shelf and click on an object, which is similar. An agent will monitor the time and the user interactions with the environment, controlling the rights and wrongs answers, changing the difficulty level of the tasks according to their performance.

9.3.13 Glycemic Monitor

The Multi-Agent Systems allow experiments which model and implement complex problems. They have requirements that demand solutions for problems of distribution, privacy, autonomy, pro-activity and sociability. The agent-oriented paradigm for software development emerged as an approach that facilitates the transition from prototypes to the development of large-scale commercial applications. In this context we developed some multi-agent systems in Medical areas. The Glycemic Monitor System aims at aiding the diabetes treatment [28].

The Glycemic Monitor system has been developed in the agent-oriented approach focused on Diabetes Control, based on the Guardian Angel Project [27]. This project was designed as an information system centered on the patient, rather than the service provider. The software agents group explains the name “guardian angels” (GA). This “guardian angels” support functions for the patient’s health, including the patient’s medical considerations, legal and financial information. The diabetes treatment requires daily monitoring of the patient, systematic follow-up by the physician and in many cases, adjustments according to day-to-day variations. Two prototypes of the Glycemic Monitor system were developed: the system in web and in mobile platforms. Fig. 9.31 presents an interaction, which shows the calculated dosage of insulin and the result of the patient evaluation.

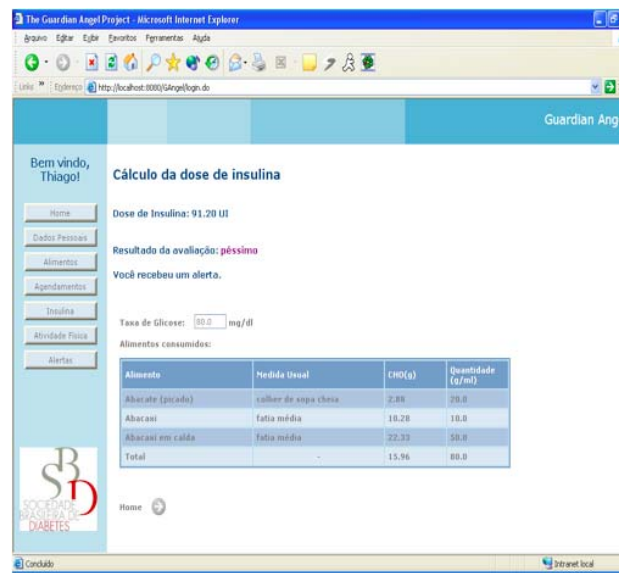


Fig. 9.31 The Glucose Monitor – Result of insulin dosage Evaluation.

The first version of the environment was implemented in JAVA, JSP, HTML, JavaScript and SQL. The agents were implemented in JADE. The PDA version was developed using the API of Java Micro Edition (J2ME). In this case, we had to change the interface and optimize some algorithms (Fig. 9.32).

The Glucose Monitor opens new possibilities to patients to monitor their own progress and have their treatment continuously assessed. The person with diabetes can submit basic queries, with no need to consult a doctor [28].

9.3.14 A.L.I.C.E. in the land of the hands

In this project we have developed a serious game designed to support the recovery of the motor function in patients who have suffered an Encephalic Vascular Accident (EVA) recently and have lost partial or total use of their upper limbs, especially the use of the hands. The game design is based on the concept of Motor Imagery, a process by which an individual mentally simulates or rehearses a given action. In our prototype the motor imagery is stimulated by presenting the user with a series of line drawing representations of the left and right human hand in different pose and rotations. Then, the patient has to guess whether the hand being displayed corresponds to a left or right hand. To do that, they instinctively “imagine” the real hand movement to match the hand pose on the display. This stimulus is known to promote cognitive and motor recovery in patients. Every time the patient correctly guess the handiness (i.e. left or right) he or she progresses the game, scoring points

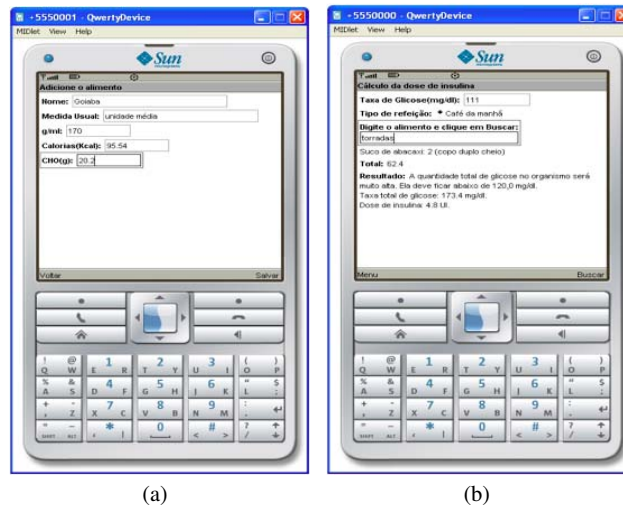


Fig. 9.32 The Glucose Monitor for PDA. Data Base of Food (left) and result of Insulin Dosage Evaluation (right).

and interacting with mini-games that stimulates motor imagery in different forms, as shown in Fig. 9.33.

Currently, the A.L.I.C.E. is begin tested at the UFRNs general hospital, with good results and encouraging feedback from the patients. The next step is to collect all data and proceed to the data analysis and discussion stages.

9.3.15 Handcopter: Development of Multi-modal Serious Games for Rehabilitation

People suffering from stroke can survive with sequels that usually involve loss of vocal ability, restriction of movement or even total paralysis on one side of the body, according to the brain area affected by the stroke. To recover from these sequels, patients must undergo intense physical therapy sessions involving several repetitive tasks. But this long-term recovery usually is not straightforward. The patients often have to go through many physiotherapy sessions in order to recover their lost or impaired movements. These sessions are usually long, tiring and tedious. These factors associated with a fragile psychological state commonly found in stroke survivors are responsible for a high number of patients abandoning treatment before its completion.

One way of tackling this problem is to introduce serious games in the treatment. The premise is that games may bring in an extra element of motivation to a physiotherapy session and, hopefully, make it more interesting for the patient.



Fig. 9.33 Several screenshots of the A.L.I.C.E. system.

In particular, we were looking for a natural body movement tracking input device that would be able to track small range movements of the individual fingers of a hand. Our goal was to design a game targeted at patients in their first stages of treatment, in which the physiotherapy needs to recover the patients basic hand movements. This means the ability to perform flexion and extension of the fingers and thumbs.

With this motivation in mind, we designed the Handcopter Game, a serious game that receives players input from a low cost vision based finger-tracking device. Our goal was to provide a tool that would supplement the regular hand physiotherapy for patients recovering from a stroke. By developing such a solution, we hope to disseminate the importance of close cooperation between serious game and traditional physiotherapy treatments.

Our low-cost tracking solution is based on a regular webcam, a support for the camera, a pair of latex gloves, and computer vision software capable of recognizing small changes in the area span by the patient's hands, even with varying light conditions. Fig. 9.34 shows the computer vision hand detection, the camera support, and a trail patient.

The Helicopter Game is a game designed so that the patient can play it using a small amount of finger movement, captured by our own vision-based data glove.

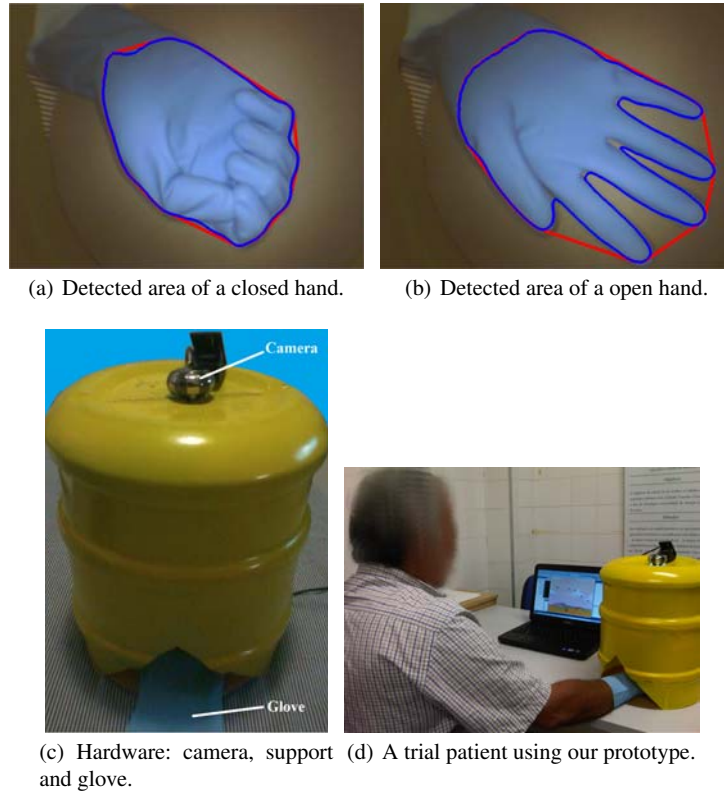


Fig. 9.34 The handcopter games software (a-b) and hardware (c) solution. In (d) we show a picture from one of the trial patients using our prototype during a physiotherapy session.

The patients must close and open his/her hands to, respectively, lower and elevate a helicopter and escape some obstacles (see Fig. 9.35). We believe that the engagement provided by the game might stimulate patients to a faster recovery, motivating them to improve their hand movements [6].

9.3.16 GPU-Based Vascular Segmentation based on Height Ridges

The vascular segmentation is important in diagnosing vascular diseases like stroke, blood clots, etc. The segmentation is often hampered by noise in the image and very thin vessels that can pass unnoticed. One way to accomplish the segmentation is extracting the centerline of the vessel with height ridges, which uses the intensity as features for segmentation. This process can take from seconds to minutes, depending on the current technology employed. In order to accelerate the segmentation



Fig. 9.35 Screenshot of the Handcopter game. The patient controls the helicopters flight level (top left) by flexing extending the hand.

method proposed by Aylward [2] we have adapted it to run in parallel using CUDA architecture.

The performance of the segmentation method running on GPU is compared to both the same method running on CPU and the original Aylward's method running also in CPU. The improvement of the new method over the original one is twofold: the starting point for the segmentation process is not a single point in the blood vessel but a volume, thereby making it easier for the user to segment a region of interest, and; the overall performance of the method running on GPU is about 4 orders of magnitude higher than the same method running on CPU, which almost granted a real time status for the segmentation method proposed (see Fig. 9.36).

9.3.17 VIDA - Virtual Interactive Distance-learning on Anatomy

Associated with ViMeT project, VIDA consists of a virtual anatomic atlas, based on three-dimensional direct manipulation and visualization of anatomical structures, in real-time [30, 31, 29]. It is intended to provide for training students a basic knowledge of the anatomical structures to be manipulated in subsequent training sessions using ViMeT.

To provide users with an “holographic” effect and the illusion of having virtual anatomical objects floating between their hands, VIDA uses stereoscopic projection and head tracking (Fig. 9.37). With this innovative interaction design it is expected that students can experience a more realistic interface and have a better sensation of presence, thus performing their learning activities with more involvement and motivation.

Acknowledgements

This research is funded by the National Institute of Science and Technology in Medicine Assisted by Scientific Computing (Grant CNPq 181813/2010-6 and FAPERJ

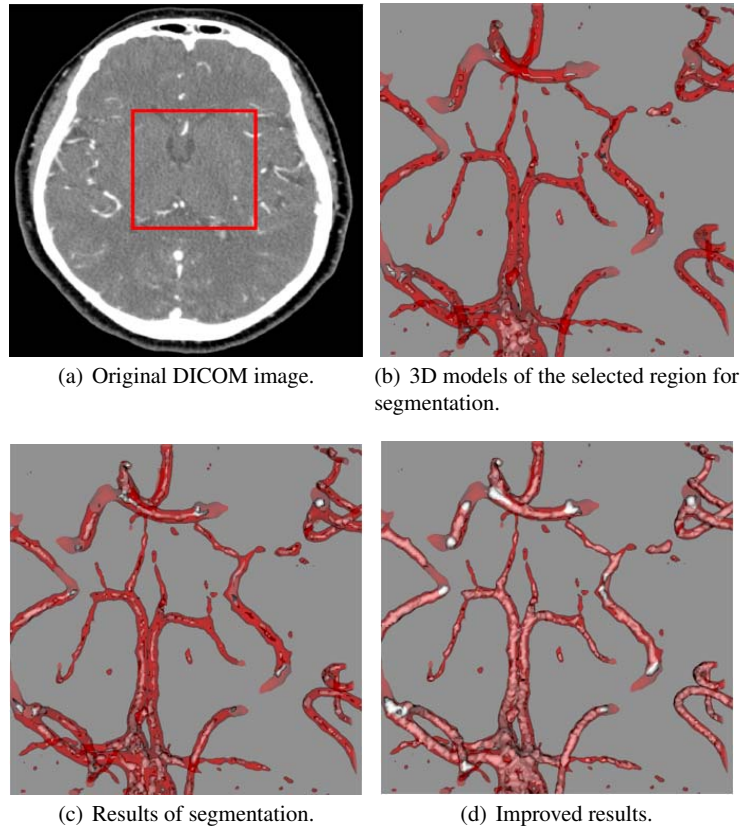


Fig. 9.36 Results of the vascular segmentation.

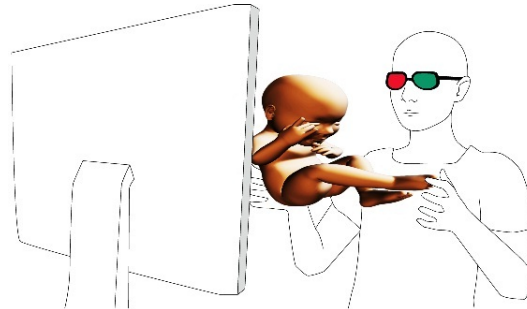


Fig. 9.37 Conceptual illustration of *VIDA* Project.

E-26/170.030/2008), as well as grants CNPq 308857/2008-9, 309318/2011-4, 476045/2011-8, 482119/2009-8, 309311/2009-8, 482119/2009-8 and FAPERJ E-26/103.106/2008 and E-26/103.241/2011.

References

- [1] A. M. Anjos, F. L. S. Nunes, and R. Tori. Evaluation of sensorimotor skills in virtual reality environments for medical training: a systematic review. *Journal of Health Informatics*, 4(1):28–34, 2012.
- [2] S. Aylward and E. Bullitt. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *Medical Imaging, IEEE Transactions on*, 21(2):61–75, feb. 2002.
- [3] B. M. Carvalho, T. S. Santos, and E. Garduno. Texture fuzzy segmentation using adaptive affinity functions. In *ACM Symposium on Applied Computing*, page to appear, 2012.
- [4] C. Ceja, E. Garduno, and B. M. Carvalho. Fuzzy segmentation-based transfer functions, in preparation, 2012.
- [5] P. F. De Abreu, L. A. V. De Carvalho, V. M. B. Werneck, and R. M. E. Moreira Da Costa. Integrating multi-agents in a 3d serious game aimed at cognitive stimulation. In *Proceedings of the 2011 international conference on Virtual and mixed reality: systems and applications - Volume Part II*, pages 217–226, Berlin, Heidelberg, 2011. Springer-Verlag.
- [6] A. M. de C. Souza and S. R. dos Santos. Handcopter game: A video-tracking based serious game for the treatment of patients suffering from body paralysis caused by a stroke. In *Proceedings of the 2012 XIV Symposium on Virtual Reality, SVR '12*, Washington, DC, USA, 2012. IEEE Computer Society.
- [7] M. E. Delamaro, F. L. S. Nunes, and R. A. P. Oliveira. Using concepts of content-based image retrieval to implement graphical testing oracles. *Software Testing, Verification and Reliability*, pages n/a–n/a, 2011.
- [8] M. Dinsmore. Virtual reality simulation: training for palpation of subsurface tumors. Master's thesis, Department of Mechanical and Aerospace Engineering, Rutgers University, October 1996.
- [9] S. dos Santos, J. de Oliveira, L. Fraga, P. Trenhago, and S. Malfatti. Using a rendering engine to support the development of immersive virtual reality applications. In *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2008. VECIMS 2008. IEEE Conference on*, pages 74–79, july 2008.
- [10] E.A. Attree, M.J. Turner, and N. Cowell. A virtual reality test identifies the visuospatial strengths of adolescents with dyslexia. *CyberPsychology & Behavior*, 12(2):163–168, 2009.
- [11] N. Josman, E. Klinger, and R. Kizony. Performance within the virtual action planning supermarket (VAP-S): an executive function profile of three different populations suffering from deficits in the central nervous system. In *Proceedings of the VII International Conference Disability, Virtual Reality & Associated Technologies (ICDVRAT)*, pages 33–38, Maia, Portugal, Maio 2008. University of Reading, UK.
- [12] M. Küllberg, J. de Oliveira, and P. Rosa. MiniVR: Low cost VR projection system. In *Virtual Reality (SVR), 2011 XIII Symposium on*, pages 134–143, may 2011.
- [13] L. S. Machado and R. M. Moraes. *Computational Intelligence in Complex Decision Systems*, chapter 4: Intelligent Decision Making in Training Based on Virtual Reality. Da Ruan (org.), Atlantis Press, Paris, France, 2010.
- [14] S. Malfatti, S. dos Santos, L. Fraga, C. Justel, P.F.F. Rosa, and J. de Oliveira. The design of a graphics engine for the development of virtual reality applications. *Revista de Informática Teórica e Aplicada*, 15(3):25–45, 2008.
- [15] S. M. Malfatti, S. R. dos Santos, L. M. Fraga, C. M. Justel, and J. C. Oliveira. EnCIMA: A graphics engine for the development of multimedia and virtual reality applications. In *Proceedings of the X Symposium on Virtual and Augmented Reality - SVR2008*, volume 1, pages 311–321, João Pessoa, PB. Brasil, Maio 2008. Sociedade Brasileira de Computação.
- [16] S. M. Malfatti, L. M. Fraga, J. C. de Oliveira, S. R. dos Santos, and P. F. F. Rosa. Um atlas 3d háptico para o estudo de anatomia. In *Workshop de Informática Médica (WIM), Belém, PA, Brazil*, July 2008.

- [17] F. Meijer, B. L. Geudeke, and E. L. van den Broek. A virtual reality test identifies the visuospatial strengths of adolescents with dyslexia. *CyberPsychology & Behavior*, 12(5):517–521, October 2009.
- [18] R. M. Moraes and L. S. Machado. Multiple assessment for multiple users in virtual reality training environments. In *Proceedings of the Congress on pattern recognition 12th Iberoamerican conference on Progress in pattern recognition, image analysis and applications*, CIARP'07, pages 950–956, Berlin, Heidelberg, 2007. Springer-Verlag.
- [19] R. M. Moraes and L. S. Machado. *Virtual Reality*, chapter 15: Development of Medical training Systems with Integration of Users Assessment. Jae-Jin Kim (ed.), Intech, Croatia, 2011.
- [20] D. Morris, C. Sewell, F. Barbagli, K. Salisbury, N. H. Blevins, and S. Girod. Visuohaptic simulation of bone surgery for training and evaluation. *IEEE Computer Graphics and Applications*, 26:48–57, 2006.
- [21] E. P. S. Nunes, F. Nunes, and R. Tori. Um modelo de avaliação da aquisição de conhecimento em ambientes virtuais de aprendizagem tridimensionais. In *Virtual Reality (SVR), 2012 XIII Symposium on*, pages 56–65, Niterói, RJ, Brasil, may 2012.
- [22] A. C. M. T. G. Oliveira and F. L. S. Nunes. Building an Open Source Framework for Virtual Medical Training. *Journal of Digital Imaging*, 23(6):706–720, 2010.
- [23] P. V. F. Paiva, L. S. Machado, and J. C. d. Oliveira. Simulation and performance evaluation of a distributed haptic virtual environment supported by the cybermed framework. In *Proceedings of the 2011 XIII Symposium on Virtual Reality*, SVR2011, pages 188–196, Washington, DC, USA, 2011. IEEE Computer Society.
- [24] F. Pretto, I. H. Manssour, M. H. I. Lopes, E. R. da Silva, and M. S. Pinho. Augmented reality systems related to medical education and training. In *Proceedings of the 2009 ACM symposium on Applied Computing*, SAC '09, pages 836–841, New York, NY, USA, 2009. ACM.
- [25] G. Robles-De-La-Torre. The importance of the sense of touch in virtual and real environments. *IEEE MultiMedia*, 13(3):24–30, July 2006.
- [26] B. R. A. Sales, L. S. Machado, and R. M. Moraes. Interactive collaboration for virtual reality systems related to medical education and training. In *6th International Conference on Technology and Medical Sciences (TMSI 2010)*, pages 157–162, Porto, Portugal, July 2010.
- [27] P. Szolovits, J. Doyle, W. Long, I. Kohane, and S. G. Pauker. Guardian angel: Patient-centered health information systems. Technical report, Technical Report MIT/LCS/TR-604. <http://groups.csail.mit.edu/medg/projects/ga/manifesto/GAtr.html>, 1994.
- [28] D. G. Tavares, J. Eicheler, A. L. Fasano, R. F. A. Martins, V. M. B. Werneck, and L. M. Cysneiros. Sistema multi-agentes monitor glicêmico. In *Anais do XII Congresso Brasileiro de Informática em Saúde*, volume 1, pages 1–6, Recife, Brazil, 2010. Wave Media.
- [29] R. Tori, D. M. Tokunaga, R. Nakamura, F. L. S. Nunes, C. G. Corra, and F. L. Picchi. Visualização Não-fotorrealística no Design de Interação do Ambiente VIDA. In *Anais do IX Congresso Brasileiro de Pesquisa e Desenvolvimento em Design*, volume 1, pages 1–8, 2010.
- [30] R. Tori, F. L. S. Nunes, V. H. P. Gomes, and D. M. Tokunaga. VIDA: Atlas Anatômico 3D Interativo para Treinamento a Distância. In *Anais do WIE 2009 - X Workshop de Informática na Escola - XXIX Congresso da Sociedade Brasileira de Computação*. Sociedade Brasileira de Computação, Porto Alegre (RS), 2009.
- [31] R. Tori, F. L. S. Nunes, R. Nakamura, J. L. B. Junior, C. G. Corra, and D. M. Tokunaga. Design de Interação para um Atlas Virtual de Anatomia Usando Realidade Aumentada e Gestos. In *Proceedings of Interaction 2009 - South America*, volume 1, pages 1–8, 2009.
- [32] A. B. Trombetta and M. S. Pinho. Projeto e desenvolvimento de um dispositivo de interação para visualização de imagens médicas em ambientes imersivos. In *Proceedings of the XI Symposium on Virtual and Augmented Reality - SVR2009*, volume 1, pages 278–288, Porto Alegre, RS, Brasil, Maio 2009. Sociedade Brasileira de Computação.

- [33] M. J. Wooldridge. *An Introduction to Multi-Agent Systems*. John Wiley and Sons Limited, Chichester, 1st edition, 2009.
- [34] K. M. Yap, A. Marshall, and W. Yu. Providing qos for distributed haptic virtual environments in ip networks. In *Proceedings of the First International Conference on Immersive Telecommunications*, ImmersCom '07, pages 13:1–13:6, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Chapter 10

Model-driven development of healthcare applications

Regina Maria Maciel Braga, Luciana Tricai Cavalini, Carlos Eduardo Cirilo, Timothy Wayne Cook, Bruno de Souza Pinto Marques Correa, Sérgio Miranda Freire, Antônio Tadeu Azevedo Gomes, Vinícius Moreira Macedo, Anderson Menezes, João Luís Cardoso de Moraes, Antonio Francisco do Prado, Wanderley Lopes de Souza, Iuri Malinoski Teixeira, and Artur Ziviani.

Abstract The application of information and communication technologies in health-care has long been seen as a promising endeavor. The last couple of years have seen nevertheless a significant growth in the skepticism about the health information system (HIS) effectiveness in supporting more efficient healthcare services. As regards the technological inhibitor factors to such effectiveness, electronic health records (EHRs) have been often mentioned as a solution. The expectations in this field are yet to be met though, particularly because of interoperability issues in EHR data models. The ‘data dimension’ is, however, part of a larger problem. There is

Regina M. M. Braga and Iuri M. Teixeira
UFJF - Federal University of Juiz de Fora, Campus Universitário s/n, Cidade Universitária, 36036-330, Juiz de Fora, MG, Brazil, e-mail: regina.braga@ufjf.edu.br, malinoski.iuri@gmail.com

Luciana T. Cavalini and Sérgio M. Freire
DTIES-FCM-UERJ - Department of Health Information Technology, Medical Sciences College / Rio de Janeiro State University and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Rua Prof. Manuel de Abreu, 444/2o andar - LAMPADA, Vila Isabel, 20551-030, Rio de Janeiro, RJ, Brazil, e-mail: lutricav, sergio}@lampada.uerj.br

Carlos E. Cirilo, João L. C. Moraes, Antonio F. Prado, and Wanderley L. Souza
UFSCar - Federal University of São Carlos and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Via Washington Luís Km 235, Monjolinho, 13565-905, São Carlos, SP, Brazil, e-mail: ducirilo@gmail.com, joao_moraes, prado, desouza}@dc.ufscar.br

Timothy W. Cook
International Colaborator – INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, e-mail: timothywayne.cook@gmail.com

Bruno S. P. M. Correa, Antônio T. A. Gomes, Vinícius M. Macedo, Anderson Menezes, and Artur Ziviani
LNCC - National Laboratory for Scientific Computing and INCT-MACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Av. Getúlio Vargas 333, Quitandinha, 25651-075, Petrópolis, RJ, Brazil, e-mail: brunos, atagomes, vmacedo, menezes, ziviani}@lncc.br

also an ‘architecture dimension’ that gains importance as the HISes get bigger in managerial and technical complexity. Until this date, however, the application of software architecture techniques in the realm of HISes has been largely overlooked. This chapter aims first at discussing about the above modeling domains—viz. informational and architectural—in the realm of HIS development. Building upon the provided discussion, we also propose an innovative model-driven method that seeks a higher degree of HIS maintainability and reuse through the transversalization and harmonious weaving of data and architecture models.

10.1 Introduction

The application of information and communication technologies in healthcare has long been seen as a promising endeavor [52]. In this context, Cantrill [15] reflects upon this promise and related challenges, highlighting that the potential advantages include: (i) improved communication between a single patient’s multiple healthcare providers; (ii) elimination of needless medical testing; (iii) decrease in medical errors, improved quality of care, and improved patient safety; and (iv) decreased paperwork, and even improved legibility.

The last couple of years have seen nevertheless a significant growth in the skepticism about the health information system (HIS) effectiveness in supporting more efficient healthcare services [89, 106]. It is generally understood that the inhibitor factors for physicians, clinics, or hospitals to adopt HISes fall into three main clusters, viz., organizational, environmental, and technological [106]. Herein we focus on technological inhibitor factors, which can be summarized—in line with Raghupathi and Umar [89]—into: (i) high development cost; (ii) lack of open-standardization; (iii) technology-orientation; and (iv) troublesome maintainability.

The implementation of electronic health records (EHRs) has been proposed to increase the effectiveness of healthcare, but the expectations in this field are yet to be met. Since 1961, when the first computerized health record system was installed at the Akron General Hospital [1], and over the more than 50 years since that time, software companies of all types have sought the ability to integrate various systems in order to provide a coherent healthcare information platform [30, 63]. Nevertheless, the majority of the systems developed worldwide in this period has been focused on the business and management side of healthcare. Only in the past two decades there has been a concerted effort to computerize the clinical aspects of healthcare [110].

The challenges related to recording clinical information on computer applications are primarily associated to the fact that healthcare is a complex and dynamic system. Regarding complexity, it is known, for instance, that SNOMED-CT, the most comprehensive terminology for healthcare, has more than 311,000 terms, connected by more than 1,360,000 links [105]. The dynamism observed in healthcare information is essentially related to the speed of scientific evolution and technology incorporation, which is a main feature of the field. Furthermore, the healthcare sys-

tem is by definition hierarchical and decentralized. It is therefore expected that the patients will access the system through primary care institutions and then ascend to higher complexity levels of care such as hospital and specialized care units. Crucially, each healthcare institution adopts specific workflows that are adapted to its form and function [109]. This process reflects on the specificity of the information collected, stored, and processed inside a given facility. However, no healthcare institution is isolated from the others; because of the configuration of the HIS, patients commonly circulate across more than one healthcare unit. Thus, ideally, every patient's record should be kept longitudinal, since any piece of information might be important at any moment of the patient's life across the healthcare system.

The achievement of such high levels of interoperability between electronic health records still remains a challenge [59]. Currently, there is a multiplicity of companies and governmental institutions whose mission is to develop HISes; each one of them implemented its own relational data model, which is specific for that HIS [90]. Such data models are not only different from system to system, but they are also ever changing as the scope of the systems change, which includes the continuous changes in medical science, insurance company regulations, and government policies [82]. This constant change is a costly component of managing healthcare information and creates a situation in which much of the semantic context of the healthcare data is embedded into the structure of the database, as well as on the programming language source code. Thus, when sharing data between HISes is attempted, even in the simplest case (when the data types are the same), the complete context in which the data was recorded remains unknown to the receiving system.

Many solutions have been proposed to the problem of interoperability in HISes, which include a vast and variable set of knowledge representation models, especially terminologies and ontologies [13]. Nevertheless, the high implementation and maintenance costs (either financial or not) of the available EHRs have slowed down their widespread implementation; even some throwbacks have been observed over the last years [47, 85]. Until this date, the only development method that has proven in software the achievement of semantic interoperability is a model-driven methodology: the multi(or dual)-level modeling approach originally proposed by the openEHR Foundation [68] and similar projects [17]. With this approach, clinical data models are never built from scratch, thus also lessening the efforts on HIS creation, evolution, and maintenance.

Another important modeling aspect regards the software architecture of HISes. For historical and economic reasons, primary care settings are typically located closer to the user's household, while more complex healthcare institutions (such as hospitals) are usually built in central areas. The functions of primary care and hospitals are clearly different, which determines a high level of variability regarding their architectural structure. Nevertheless, typical communication and control patterns appear within system families (e.g., prehospital emergency systems, epidemiological surveillance systems), which can be captured using well-established software architecture techniques.

Until this date, however, the application of software architecture techniques in the realm of HISes has been largely overlooked. Even though attention has been

drawn to the potential of model-driven methodologies for overcoming inhibitor factors to the deployment of maintainable HISes through the reuse of good architectures [83, 104], the existing initiatives are rather limited with regard to the specification of novel architectures. One example is the ubiquitous network support typically taken for granted by current HISes demanding data communication, which may be scarce or even absent in disaster scenarios in urban areas (e.g. the ones caused by the Indian Ocean tsunami in December 2004, the hurricane Katrina in August 2005, or the Japanese earthquake and tsunami in March 2011). Such scenarios ask for the seamless integration of on-site mobile devices—following the concept of hastily formed networks [31]—so that they can collect and automatically process information about groups of injured people (e.g. for triage) and thus better allocate medical resources accordingly. Adapting existing HISes demanding data communication to operate in intermittent network environments such as the ones depicted above is likely to be challenging. Usually, such systems are unable to gracefully handle disconnection conditions, forcing users to retry connections manually. Using intermittent connectivity requires (i) new algorithms to detect when a network connection has come and gone as well as to route data accordingly; (ii) new data structures to temporarily store in-transit data; (iii) and above everything, innovative software architectures that glue together all such algorithms and data structures into a coherent software system. The community involved in the development of HISes seems yet rather unaware of the possibilities of such technologies, which is arguably due to the lack of development platforms that allow for the specification of yet-to-be-implemented software architectures.

This chapter aims first at discussing about the above modeling domains—viz. informational and architectural—in the realm of HIS development. Building upon the provided discussion, we also propose an innovative model-driven method that seeks a higher degree of HIS maintainability and reuse through the transversalization and harmonious weaving of data and architecture models. The transversalization looks for a separation of concerns with regard to the data to be generated, processed, and stored in a HIS: health domain specialists focus on the *what*, whereas software architects focus on the *how*. The weaving is a special kind of model transformation that combines the two domain models into a single model that can be used for generating HIS skeletons.

10.2 Review of some health information systems

10.2.1 *Emergency support for acute myocardial infarction patients*

Acute myocardial infarction (AMI) is among the leading causes of death and physical incapacity worldwide: in 2005, for instance, cardio-vascular diseases accounted for 30% of deaths globally [96]. Furthermore, it is also known that decreasing the interval from the symptom onset and its recognition to treatment lessens morbid-

ity and mortality of patients experiencing AMI [28]. Nevertheless, in both industrialized and developing countries, the so-called ST-segment elevation myocardial infarction (STEMI) programs—which depend on delivering earlier (typically pre-hospital) emergency medical services (EMS)—are still being studied and revised in order to be efficiently implemented [70, 80]. As a result, AMI patients still experience prolonged delays before receiving primary medical care (PMC) [102].

Among the causes of AMI (which is in general STEMI), the most common is ischemia, which is a sudden reduction or interruption of the blood flow to a tissue because of an arterial constriction or obstruction. To date, there are two main treatments in cardiology to resume the blood flow in a previously constricted or obstructed artery: angioplasty and thrombolysis. Crucially, the latter has the benefits over the former of being cheaper and feasible to be implemented by the paramedic him/herself at the emergency scenario. Nevertheless, the call for the application of thrombolysis requires a careful technical judgment on whether or not the AMI patient is thrombolytic eligible. Often only a cardiologist can make such a judgment. This specialized decision is based on an analysis of an (usually ten-second) electrocardiogram (ECG) in addition to some information on the recent medical history of the patient. Nevertheless, such a specialist is most likely not available at the place where the first EMS is provided to the patient, thus rendering rates of prehospital thrombolytic administration lower than expected by clinical guidelines.

The transmission of ECG data to a central station or a hospital dates back to 1987, when Grim et al. [48] experimented a cellular telephone transmission of 12-lead electrocardiograms from ambulances to a hospital. The main objectives of such a study were testing whether the ECG signal would suffer any distortion or corruption during data transmission as well as quantifying the transmission delay. From then on, different technologies have been put to test for transmitting ECG records from ambulances and emergency units in general. With the rapid dissemination of mobile and wireless technologies in recent years, a multitude of ECG transmission systems have been developed in prehospital settings [58, 73, 29]. Such systems, however, typically focus on the remote monitoring of the patient heart and/or other vital signs in order to allow the cardiologist at a Coronary Care Unit (CCU) or a hospital to track his/her condition and then prepare for his/her arrival. Examples are the monitoring of arrhythmias in moderate-risk patients [50] and the early classification of AMI patients during their transportation to a hospital with CCU [23].

For a general prehospital teleconsultation purpose, Kwak et al. [71] have developed a portable device that allows two-way, real-time communication between hospital doctors and emergency medical technicians. In this initiative, the data communication is carried out by means of audio and video transmission through a wireless Internet connection with an average data transmission rate of 1 Mbps. Such audio and video data transmission in fact requires a connection bandwidth possibly higher than the one that might be available in certain regions in countries like US, Brazil, and others. We are referring here to countries characterized by large territorial areas, some of which possibly lacking advanced telecommunications capability. For an overview of the current situation of US, for instance, the reader may refer to the FCC's 7th 706 report [37]. Besides, transmitting audio and video data complicates

data storage and retrieval for the sake of auditability, a fundamental concern for the controlled delivery of prehospital thrombolysis.

To address the problem of prehospital delivery of thrombolysis to AMI patients in Brazil, the MARTIN Lab at LNCC has developed a teleconsultation system that enables the remote interaction of paramedics who are providing the AMI patient with prehospital EMS and a cardiologist available at a CCU. The system is named AToMS, standing for AMI Teleconsultation & Monitoring System. AToMS [24] takes benefit of ubiquitous (mobile and wireless) technologies to be an effective telemedicine resource anytime, anywhere. Nevertheless, wireless connectivity and bandwidth limitations in Brazil, and also the concern for auditability, have rendered a system in which ECG records are transmitted as files and the paramedics-cardiologist communication are transmitted as structured data.

10.2.2 Maternal pre-hospital care

In the 20th century, worldwide, there were more cases of maternal deaths (related to pregnancy, childbirth, and the postpartum period) than the combined deaths by tuberculosis, suicides, traffic accidents, and AIDS. Even today, two decades after the signing of “Safe Motherhood Initiative”, which pledged to halve the number of women dying of maternal causes by 2000, almost the same number of women still dies from maternal causes as in 1987 [103]. During this period, 10.5 million women died during pregnancy. Other 420 million women had unsafe abortions and 315 million had complications, such as fistulas. Every day, 1,600 women die during pregnancy, 90% of them living in developing countries [92].

However, some progress has been recorded, which may be partly due to the revival of political will, with the global commitment to the Millennium Development Goals (MDGs), which include reducing by 75%, from 1990 to 2015, the rate of maternal mortality. Meanwhile, many of the 529,000 deaths per year are avoidable, since 80% of them result from 1 to 5 well-known and relatively common obstetric complications, which can be easily dealt with extremely cost-effective interventions [54].

Additionally, the complexity of collecting data on maternal mortality is widely recognized. The conventional approach to control the level of maternal mortality using rates as indicators is inherently complex: The deaths have undercount and misclassification, while traditional methods of data collection are difficult to handle and require significant resources [72].

It is known that women prefer modern obstetric medical care, but the access to this modality of service is blocked by many factors, especially where is most needed: Poor urban areas or remote rural zones of developing countries [36, 88]. There is evidence that the indicators related to maternal health are improved by the adoption of obstetrics emergency care policies [86], and that includes good handling of information by proper healthcare applications [46].

The implementation of EMS for maternal health is a globally recognized need. There is evidence from developed countries about improvements in the outcome of cases attended by this type of service, in terms of survival, time and cost of care, and such effects are most apparent in populations with poor access to health services. However, there is a uniform concern about the need to emphasize the integration of care activities to emergency care to the health system as a whole to maximize its effectiveness [100].

10.2.3 Epidemiological and syndromic surveillance systems

The efficiency of epidemiological surveillance actions can lead to the reduction of multiple diseases, globally recognized as health priorities [20]. Epidemics can only be avoided if the local health professionals are aware of the existing cases, which enable them to timely implement the appropriate prevention and control measures. The earlier the recognition of the cases happens, the more effective the prevention. In order to allow health professionals to learn about the existence of suspected cases of an epidemic disease, there is a need for the collection of health information at the point and moment of care. Moreover, there is a need for greater accuracy in the diagnosis of severe cases and atypical epidemic and endemic diseases [78].

Lack of complete and accurate information at the PMC is the number one enabler of the wide spread of infectious diseases. The slowness in the identification of epidemics has pushed up the public health expenditure and contributed to the worsening of the quality of care. The fundamental factors of this process are: (i) the excessive time between the arrival of the patient to the point of care and the arousal of the diagnostic suspicion; and (ii) the incorrect indication of laboratory testing. In addition, this situation makes difficult the identification of severe and atypical cases of epidemic cases. This often misguides the public opinion and makes proper health education efforts more difficult.

The traditional approach to epidemiological surveillance adopts retrospective case definitions, which has reduced the applicability in the current emerging-reemerging pandemic scenarios [108]. The approach regarded as more adequate to the current epidemiological complexity is called syndromic surveillance. This modality of epidemiological surveillance allows the early suspicion of an epidemic disease case based on a small set of signs and symptoms and a simple diagnostic algorithm. These epidemiological surveillance algorithms can be implemented in small application specific healthcare systems in mobile devices [56]. An interface with geographic information systems and statistical time-series analysis tools allows the analysis of the collected information among diverse, remote systems. Thus, at the local level, the control measures can be implemented immediately, blocking the disease transmission chain, and at the governance level, wider areas can be monitored and priority regions can be identified for disease control.

The epidemiological guidelines should be embedded in the same decision support system that contains the syndromic surveillance algorithms. This enables the

healthcare professional to be supported, synchronously, for the syndromic diagnosis and the adoption of the disease prevention and control measures, avoiding the consultation of the guidelines in large paper books or PDF files.

10.3 Discussion about system domains

10.3.1 Health information modeling

The effectiveness of healthcare is related to the adequate response to the dynamic changes of health conditions of the communities. The resulting epidemiological profile of human populations have increased significantly in speed and complexity by globalization, as it can be seen during the occurrence of epidemics and pandemics as well as in the daily overcrowding of emergency services [55]. The time required to respond to such changes may extend for weeks or even months, as analysts collect data from incompatible information systems or paper records, which has limited effectiveness. The incorporation of information technology in healthcare has been proposed as a strategy to overcome the current situation, but there are obstacles derived from the significant complexities of health information in dimensions of space, time, and ontology [57].

Regarding the spatial dimension, there is specificity related both to the population and the healthcare services at the local level. Given a larger geographical area (e.g., a metropolitan area) composed by a set of smaller areas (e.g., municipalities), it can be observed that the smaller areas are usually heterogeneous in terms of population size, demographic and epidemiological profile, and contextual determinants of health conditions such as income distribution, access to general social goods (sanitation, water supply, employment), as well as specific access to healthcare. Thus, the proximity or even the adjacency of small areas does not prevent the existence of a significant variability of population-specific healthcare needs. Depending on the design of the healthcare system from the national to the local level, the political response to those healthcare needs may vary from one small area to another, thus generating additional complexity.

The temporal component of complexity is associated to the technological evolution of the field. The intensity of this process was followed by the development of medical goods such as diagnostic equipment and techniques, pharmaceutical products and other therapeutic technologies and, in consequence, it has been a challenge for healthcare professionals to keep their practice up to date with the innovation in their field of expertise. One of the consequences of the technological development in medical sciences is the continuous creation of new (and change of the previously existent) concepts, with consequences for the practice [110].

The complexity of the ontological component is partially derived from the temporal dynamics described above. However, there is a specific feature of the way that healthcare knowledge is created and disseminated, which adds complexity to the

definition of medical concepts. Since medical science is located in the interface of biological, social, and exact sciences, there may be different definitions for a single medical concept; in other words, the likelihood for reaching consensus in healthcare is rather low. The variability of concepts between medical schools, combined with the independent variability of healthcare policies and medical cultures, among other determinants, may decrease the level of consensus about a given set of medical concepts that it is possible to achieve [95].

In addition, in the typical healthcare provider spectrum, each provider has different information needs. Therefore, the applications or at least the views into applications need to be very specific in order to improve usability. Large standardized systems are usually slow to change and adapt to the rapid rate of change dictated by the adoption of new emerging medical technologies. Thus, healthcare information systems will not long remain interoperable unless the standards and terminologies evolve quickly, which is hardly doable, given its ontological complexities.

The end result of the presence of such complexity in healthcare information systems is that they are usually not interoperable and have high maintenance costs. These issues have a significant impact on the low level of adoption of information and communication technology in the dynamic and emergent healthcare systems worldwide, in particular when compared to other sectors of the global economy [79].

10.3.1.1 Healthcare information standards

The challenges related to the semantic interoperability and high cost maintenance of healthcare information systems have led to the development of several healthcare information standards. The main categories of healthcare standards information are specifications and terminologies. Unique identifiers such as health cards are part of the healthcare information ecosystem, usually associated to terminologies, but both lay beyond the scope of this Chapter.

Healthcare information specifications define data models, data structures, and knowledge modeling procedures for the representation of clinical concepts. The most adopted (or at least studied) specifications are HL7 and the three versions of multilevel modeling originated from the *openEHR* specifications. There is an ISO Technical Committee in Health Informatics, the ISO TC215, and some components of the HL7 and *openEHR* specifications have been converted into ISO standards, such as the ISO 13606 series and the ISO/HL7 family of standards.

HL7

HL7 has two versions: the HL7 Version 2.x, a set of versioned message standard specifications for clinical and administrative transactions [74], and the HL7 Version 3, a messaging standard specification for clinical and administrative transactions, based on the HL7 Reference Information Model [9]. The Clinical Document Architecture (CDA) is the HL7v3 XML-based model for clinical documents [35].

HL7 version 2 has a set of messaging specifications to support claims reimbursement using electronic exchange of health invoices. One of the HL7v2 challenges relates to the fact that the process of message definition is not standardized across all implementations [81]. This limitation in using version 2 has led to the development of version 3, based on a Reference Information Model (RIM), defined as a graphical representation of the HL7v3 standard, being clinical data a RIM-based object model [66]. Thus, the HL7 RIM approach to modeling is to exclude all classes not needed to create a specific message model. This is a cross-sectional modeling approach that creates sub-domain models from it. Therefore, this approach is not different, from an interoperability standpoint, an entity-relationship modeling, which has proven not to be a solution in healthcare and other complex domains [87, 19]. This approach means that each message model must be implemented in software in each system that uses the message instances. Even though HL7v3 is conceptually an object model, the normative artifacts are distributed as a group of Microsoft Access database files [97]. The constraint mechanisms in HL7v3 have been in near constant redefinition and renaming throughout the life-time of this model, which makes difficult to define univocally how any clinical concept should be modeled. For instance, in HL7v3, the seemingly simple concepts such as marital status and gender might have ambivalent representations, due to the choice for making the display of information easier by excessively limiting representations at the data level, which has led to the arousal of questions on to what extent semantic interoperability is attainable by the adoption of HL7 [21].

openEHR

An important approach for specifying and implementing EHR systems based on an advanced architecture is the *openEHR* project.¹ *openEHR* [8] is an open, detailed, and tested specification for a comprehensive interoperable health information computing platform for supporting EHR systems.

The *openEHR* model and specifications [6] provide both functional interoperability (humans can read transmitted health information) and the far more demanding semantic interoperability (computers can understand and automatically process transmitted health information). Semantic interoperability is essential to enable local processing of the shared data; this is a pre-requisite for intelligent decision support and care planning, which are arguably the most important added-value applications of adopting EHRs. *openEHR* provides a specification for the communication of EHR Extracts at various levels of complexity and also a full specification for the creation, storage, maintenance, and querying of EHRs.

The models are expressed using the Unified Modeling Language (UML) notation and formal textual class specifications. The abstract specification consists of the Reference Model (RM), the Service Model (SM), and Archetype Model (AM). The first two correspond to the ISO RM/ODP information and computational view-

¹ <http://www.openehr.org>

points, respectively. The latter formalizes the bridge between information models and knowledge resources. These models constitute the primary references for all *openEHR* semantics. The presentation style of these abstract specifications is deliberately intended to be clear and semantically close to the ideas being communicated. Accordingly, the specifications do not follow idioms or limitations of particular programming languages, schema languages, or other formalisms.

The specifications concerning implementation technology, on the other hand, correspond to the expression of abstract specifications in various programming and schema languages, each of which represents an imperfect and usually partial transformation from the specification models. There are numerous implementation technologies, ranging from programming languages, serial formalisms such as XML, to database and distributed object interfaces. Each of these has its short-comings and strengths. The approach to implementing any of the *openEHR* abstract models in a given implementation technology is to firstly define an implementation technology specification (ITS) for the particular technology, then use it to formally map the abstract models into expressions in that particular technology.

One of the key paradigms on which *openEHR* is based is known as the “two-level” modeling approach, as shown in Figure 10.1: On the first level, a common RM is established, using a predefined set of classes that model the structure of the electronic record; and, on the second level, specific concepts are established by restricting the RM classes, in the form of archetypes, expressed in the Archetype Definition Language (ADL) [5], which can be translated to any language. An archetype constitutes a formal model of a domain concept easily understandable by a domain expert. For example, in order to represent the demographic data of patients, the demographic archetypes available in the *openEHR* repository—called Clinical Knowledge Manager (CKM)²—may be used since they provide all data structures that are needed. In the two-level modeling, data input from users are stored according to the RM, but should also conform to the concepts expressed by the archetypes. The archetypes are designed by domain experts, and not by information technology professionals, which avoids the need for interpretation of the knowledge extracted from an ad hoc interaction [7].

The archetypes, represented in ADL language, consist in three parts: identification, definition (structure, rules, and cardinality), and ontology. To illustrate this, Figure 10.2 shows the definition section of the archetype for data of the patients concept based on the ITEM_TREE class.³ In this archetype, the attribute items on line 48 are of CLUSTER type. The attribute items on line 50 are of ELEMENT type. The attribute value on ELEMENT on line 51 is DV_DATE type, on line 54 is DV_BOOLEAN type, and the others are DV_CODED_TEXT.

As a consequence of the two-level modeling approach of *openEHR*, systems have the possibility of being far smaller and more maintainable than single-level systems. They are also inherently self-adapting, since they are built to consume archetypes and templates as they are developed in the future. Archetypes also act as a well-

² <http://www.openehr.org/knowledge>

³ *openEHR-DEMOGRAPHIC-ITEM_TREE.person.details.v1* available in CKM.

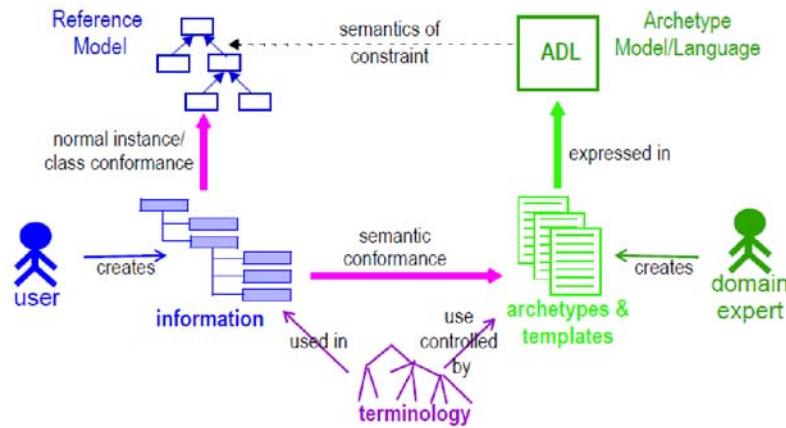


Fig. 10.1 Archetypes meta-architecture [4].

```

46 definition
47   ITEM_TREE[at0000] occurrences matches {0..1} matches { -- demographic data about a person and his/her biometric identifier
48     items cardinality matches {1..*: ordered} matches {
49       CLUSTER[at0001] matches { -- Birth data
50         items matches {
51           ELEMENT[at0010] occurrences matches {1..1} matches { -- Birth date
52             value matches {
53               DV_DATE matches {*}}
54           ELEMENT[at0011] occurrences matches {0..*} matches { -- Date of birth follow up indicator
55             value matches {
56               DV_BOOLEAN matches {*}}
57           ELEMENT[at0012] occurrences matches {0..1} matches { -- Country of birth
58             value matches {
59               DV_CODED_TEXT matches {
60                 defining_code matches {[ac0001]}}}
61           ELEMENT[at0013] occurrences matches {0..1} matches { -- State
62             value matches {
63               DV_CODED_TEXT matches {
64                 defining_code matches {[ac0002]}}}
65           ELEMENT[at0014] occurrences matches {0..1} matches { -- City
66             value matches {
67               DV_CODED_TEXT matches {
68                 defining_code matches {[ac0003]}}}
69           ELEMENT[at0015] occurrences matches {0..1} matches { -- Plurality
70             value matches {
71               DV_TEXT matches {*}
72               DV_CODED_TEXT matches {
73                 defining_code matches {[ac0004]}}}

```

Fig. 10.2 ADL language – person details concept.

defined semantic gateway to terminologies, classifications and computerized clinical guidelines. The use of archotyping in *openEHR* engenders new relationships between information and models, as shown in Figure 10.1. The archetype language works as a “constraint model”, or as an AM, formally related to the RM.

Templates are defined locally, on the contrary to archetypes that define widely re-usable components of information, and describe local usage of archetypes and relevant references. In other words, templates are used to limit archetypes to what

they are intended to do for a particular application. For an examination of blood pressure of a pregnant woman for example, the complete archetype of blood pressure is unnecessary. Instead the archetype is limited to the fields and allowed values actually needed for this particular kind of examination.

For example, a restriction to the first level concept of “Observation” can be made by the blood pressure archetype, which represents a description of all the information a clinician might want to report about it. Archetypes enable binding to different terminologies [101, 38] and are designed to represent maximal data sets, that is, they are designed to represent concepts as generically as possible. Archetypes are then combined in templates in order to generate forms, messages, etc. The templates may further restrict the archetype elements and select the terminologies they will use in order to meet the requirements of the specific context they apply.

Figure 10.3 shows a simplified view of the *openEHR* reference model. In summary, an EHR is a set of compositions (COMPOSITION) whose contents (ENTRY) may be of a clinical or an administrative type. The entries can be organized in sections (SECTION). The ENTRY class is divided into two categories: healthcare data (CARE_ENTRY) and administrative data (ADMIN_ENTRY). The CARE_ENTRY class covers the entire process of patient care and is divided into OBSERVATION, EVALUATION, INSTRUCTION and ACTION that actually relates to past, present and future events. The ADMIN_ENTRY class refers to the administrative data filled in by doctors or nurses. For instance, it can represent admission, scheduling, and requests. Differently from the CARE_ENTRY class, it is a generic class without any other conceptual subclasses. Every data value is recorded in an ELEMENT object. ELEMENTs may be organized in structures such as tables (ITEM_TABLE), trees (ITEM_TREE), lists (ITEM_LIST) and a single ELEMENT (ITEM_SINGLE).

The *openEHR* platform represents healthcare information with full meaning and interoperability [40]. In order to help users design good quality archetypes in a friendly and graphical way, there are some free tools available.⁴⁵

10.3.1.2 Discussion

Besides *openEHR*, there are two other multilevel modeling (MM) specifications available: the ISO 13606 Standard [67], and the Multilevel Healthcare Information Modeling (MLHIM) specifications [17]. Despite the open availability of such specifications, there are few records of MM implementations in real healthcare settings, mostly from academia [2]. We discuss about some possible reasons for that below.

⁴ <http://pangea.upv.es/linkehr>

⁵ <http://www.imt.liu.se/mi/ehr/tools>

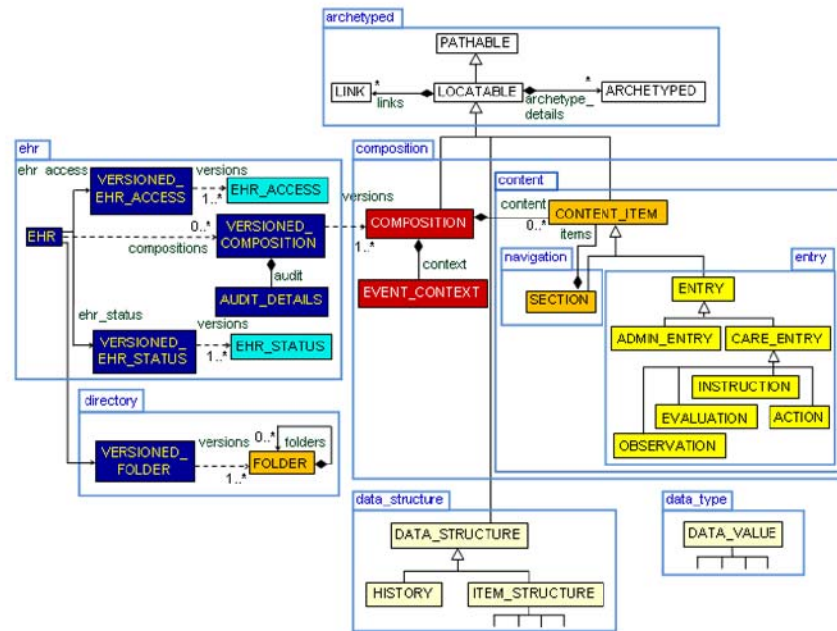


Fig. 10.3 *openEHR* reference model [6].

Domain specific languages in MM

The wider adoption of MM-based applications in real healthcare settings depends on the definition of a standard architecture for the representation of DM concepts. In *openEHR* and ISO 13606, the solution was given by the development of a Domain-Specific Language (DSL) to express the restrictions to the RM in computational structures called archetypes. In both projects, this DSL is the Archetype Definition Language (ADL) [5, 60].

The ADL is the ideal theoretical choice for knowledge representation in healthcare, since a DSL is the most adequate way to represent domain concepts. Like any DSL [32], ADL allows knowledge representation at the exact level of abstraction of the domain, which, in principle, allows healthcare domain experts to develop and validate ADL files. However, the evident benefits of ADL have been surpassed by the costs of educating healthcare domain experts to use it for knowledge modeling in the *openEHR* and the ISO 13606 projects [69]. While this educational barrier is not broken, programmers will very likely be responsible for writing, editing, and parsing ADL files for practical *openEHR* and ISO 13606 implementations. Thus, the adoption of a pure DSL in MM may represent a barrier to entry of programmers and analysts in real large-scale implementation scenarios, because it presupposes

the overcoming of an additional learning curve, since ADL is usually not covered in regular computer science curricula.

Another important aspect is not related to the technical features of ADL, which are, as stated above, unquestionable: the fact that, in both *openEHR* and ISO 13606, the archetype governance model adopted is a top-down, centralized type approach. That centralization is even expressed in one aspect of the formalism: the definition of the namespace for a given archetype. Since archetype namespaces are human-readable, in principle there should be a univocal relationship between a clinical concept and its correspondent archetype, only specializations of the same archetype being allowed in certain cases [5]. This choice for a centralized control of knowledge modeling might be one of the causes for the slow adoption of MM [69] and certainly crashes with the difficult achievement of consensus, which is an intrinsic aspect of healthcare knowledge [49].

On the other hand, at the same time that ADL was been established as the DSL for knowledge modeling in the *openEHR* specifications, the XML technologies became ubiquitous throughout the informatics industry, being essentially the backbone of the emerging Semantic Web [27]. As a consequence, a growing number of informatics professionals are familiar with the use of XML technologies [18]. Concomitantly, XML technologies have been proposed as partial solutions for healthcare data standardization of single-level applications [98]. Nevertheless, for MM-based systems, they showed similar performance when compared to ADL [91].

Since there is a perceived need to increase the adoption of MM in healthcare information systems, it may be recommended to increase the number of potential developers by adopting information technology industry standards, which can ensure sustainability of the implementations. Taking into account this historical process, in the MLHIM specifications, the XML Schema format was adopted as the constraint language. The XML Schema that specializes the MLHIM Reference Model is denominated Concept Constraint Definition (CCD). The CCD defines the necessary and sufficient combination and restriction of MLHIM RM classes for the representation of a given healthcare concept. Therefore, the MLHIM CCDs are analogous to the *openEHR* and ISO 13606 archetypes [17].

Finally, the MLHIM governance model of CCDs has been defined as bottom-up and decentralized, therefore aligned with multilingual, multicultural, and global sociopolitical requirements. Thus, the CCD governance approach is to allow for many CCD implementations of the same concept, the interoperability being maintained in technical aspects by the validation of the different CCDs to the MLHIM RM due to the ability of making all CCDs available through a web-based distributed network, the Healthcare Knowledge Component Repository (HKCR). Although it is anticipated that there will be many CCD implementations for a given concept in the beginning of the modeling process, as the repository grows, one may expect the best CCDs to be the most used, thus naturally rising to the top, as has been seen in many cases of the establishment of *de facto* open source standards [94].

Lesser multilevel models

Another perceived need is related to the expansion of the MM scope beyond the traditional EHR-centric features of MM. This is an important aspect of healthcare information, since the number of devices that collect biological and biomedical data from living subjects is rapidly increasing [14]. Many, if not most, of these devices do very little processing of this data and certainly not in the context of any other device. Thus, for healthcare mobile applications, the same need for a standardized way to represent this data within its semantic context needed for EHRs is required, so that aggregated data on an individual can be processed via reasoning engines.

Since they are usually purpose-specific, medical mobile devices do not need to be designed to support the processing requirements of complete EHR platforms, although the data they collect should be regarded as part of the complete health record of the individual. Therefore, such devices must be able to transmit their collected data in a consistent, industry standard way that allows a variety of other healthcare applications to process that data.

As stated previously, the typical healthcare provider spectrum has different information needs. In addition, the change rate of medicine concepts dictates rapid change in HISes. As a consequence, large monolithic systems are usually slow to change and adapt, and systems cannot be made interoperable unless the standards and terminologies also evolve quickly, which is seldom the case.

In order to overcome the current challenges in the health informatics field, the MLHIM specifications adopted the positive aspects of all available healthcare information standards. Therefore, in MLHIM, the MM principles were adopted, as well as the framework structure of the specifications. Additionally, taking into account that the ISO 13606 standard defines a much smaller model, allowing less semantics to be embedded into the RM classes, that approach was applied to the MLHIM specifications. However, the ISO 13606 standard RM is not comprehensive enough for the development of healthcare applications, being more adequate for defining message exchange between systems [61]. The solution for the MLHIM specifications was thus to get inspiration from the ISO 13606 standard for the reduction of RM classes and the elimination of the excess of semantic context still remaining on the *openEHR* RM classes.

In short, the adoption of purpose-specific, MM-based healthcare applications has the potential to support e-health initiatives to reaching the expected goals of improving effectiveness in healthcare. In order to achieve that goal, the multiplicity of solutions being implemented in the ecosystem of healthcare applications should be taken into account. Thus, at the point-of-care, the proper measures can be more opportunely implemented, improving the cost-effectiveness of healthcare interventions and, at the governance level, wider areas can be monitored and priority regions can be identified for community interventions.

10.3.2 Architectural modeling

Software architecture is an important discipline of software engineering, particularly in the realm of (either technically or managerially) complex system development. The software architecture of a system is the overall structure or structures of the system, which comprise design elements, the externally visible properties of such elements (such as their provided and required services, performance characteristics, fault handling behavior, and so on), and the relationships among them [3]. An architecture specification uses high-level representations such as Architecture Description Languages (ADLs)⁶ or DSLs (such as UML profiles) to embody information about a system's structure while abstracting away details about the internal implementation of its elements. Such a specification allows the software architect to reason about structural properties early in the development process.

Software architecture may also denote principles and guidelines that govern the system design and evolution over time [43]. Architectural styles [41] are the most common way to describe such principles in software architecture modeling. These are defined in terms of formally described software element types, as well as rules that govern the composition of such elements. For instance, a pipeline style might comprise a basic set of pipe and filter element types and rules that permit only non-circular compositions of such elements in a system's structure. A system following a style can thus be seen as an instance of such style at the implementation level.

Architecture description languages such as Acme [42] and Rapide [76] allow one to specify architectural models according to a chosen architectural style. Since these languages have been built for the specific purpose of architectural design, their constructs are meant to map in a one-to-one fashion the elements of the universe of discourse (in this case, elements of software architecture). Moreover, syntactic rules are embedded into architecture design tools (e.g., for Acme, the AcmeStudio) in order to warrant semantic appropriateness—e.g., for a system meant to follow a client-server style one cannot model in such tools a client element that connects to another client element. In Acme, such constraints over architectures are defined using the Armani extensions. Armani [84] is a sub-language based on first-order predicate logic that can be used to express constraints on system structure and behavior.

Within the HIS domain, one can envision typical communication and control patterns that vary across system families. For instance, the typically strict real-time requirements in prehospital emergency healthcare do not necessarily bother in epidemiological surveillance. In the latter, however, the robustness and reliability of the available information is of foremost importance. The particularities for each of these system families can be clearly expressed as architectural styles.

To illustrate how architecture descriptions and styles can be used in the HIS domain, Listing 10.1 presents a partial description of the AToMS system presented in Section 10.2.1 using the Acme ADL. Figure 10.4 gives a pictorial representation of such architecture. As can be seen from these figures, the system follows a pub-sub style at its outermost level. The ETSubsys and VCSubsys components (represent-

⁶ Not to be confused with the Archetype Definition Languages presented in Section 10.3.1.

ing the emergency teams and a teleconsultation center, respectively) are internally composed as client-server subsystems. Within ETSubsys, ETCLI components represent 'rich' client applications running on the portable devices held by the emergency teams, whereas the ETSrv component represents a web service. These components communicate with each other through an asynchronous SOAP connector deployed over a wireless medium. Within TCSubsys, TCCLI components represent web applications accessed by cardiologists through standard web browsers (synchronous HTTP), whereas the TCSrv component represents another web service. Teleconsultation requests/ replies and monitoring notifications are exchanged asynchronously between the two aforementioned web services through two message bus connectors (TelBus e MonBus), which are typically deployed over a wired network.

10.3.2.1 Domain engineering

In spite of the discussion in the previous section, the application of software architecture techniques in the realm of HISes has been to date largely overlooked. Nevertheless, attention has been drawn to the potential of domain engineering in general, and model-driven engineering (MDE) in particular, for overcoming inhibitor factors to the deployment of maintainable HISes. The application of MDE in healthcare has been recently discussed by a few initiatives in the medical informatics literature, most of them focusing on OMG's standardized model-driven architecture (MDA) approach [83]. The MDA approach lies in the translation of a platform-independent model (PIM) into different platform-specific models (PSMs) and then generating code from the latter. This allows one to manage, say, functional changes, directly in the PIM to then regenerate PSM and code, rather than (at worst) handling changes as fixes in the code. This means a better management of the so-called roundtrip problem [51], i.e., handling multiple artifacts that represent different levels of abstraction in the development lifecycle in such a way to avoid redundancy and inconsistency.

Two key techniques are employed in MDE: metamodeling and model transformation:

- A metamodel is a model that represents the types of modeling elements, as well as of their associations, which can be used for building models that instantiate the metamodel. This building process may be also used for defining the metamodel itself, i.e. a metamodel may have one or more underlying metametamodels. In theory this is an endless process. The MDA approach defines 4 levels of model building: metameta-model (M3), metamodel (M2), model (M1), and executable instance—the software system itself (M0). The M3 level is self-built based on OMG's MOF (Meta Object Facility) standard. This technique for building models allows a better design reuse and increased interoperability (guaranteed by construction) of the resulting software. Some MDE frameworks loosely follow the MDA standard, such as the Eclipse Modeling Framework (EMF), which defines a subset of MOF called Ecore, above which differ-

```

System AToMSSystem: PubSubFam = new PubSubFam extended with {
Component EHRSubsystem1: ParticipantT = new ParticipantT extended with {
Ports { annTeleconsultation, annMonitoring: p_announce = new p_announce;
      rcvResult: p_receive = new p_receive; };
Representation rep = { System rep: ClientAndServerFam = new ClientAndServerFam extended with {
Component EHRClient1: ClientT = new ClientT extended with {
Port annMonitoring;
Representation rep = { System rep: TieredFam = new TieredFam extended with {
Component EHRGUI: TierNodeT = new TierNodeT extended with {
Ports { useSensorData: useT = new useT; sendRequest; }; };
Component SensorController: TierNodeT = new TierNodeT extended with {
Ports { provideSensorData: provideT = new provideT; monitoringPort; }; };
Connector conn: LocalConnT = new LocalConnT;
Attachments { EHRGUI.useSensorData to conn.callee;
              SensorController.provideSensorData to conn.caller; }; };
Bindings { SensorController.monitoringPort to EHRClient1.annMonitoring;
          EHRGUI.sendRequest to EHRClient1.sendRequest; }; };
Component EHRServer: ServerT = new ServerT extended with {
Ports { annTeleconsultation; rcvResult; }; };
Connector conn1: CConnT = new CConnT extended with { Property protocol: string = "SOAP"; };
Attachments { EHRClient1.sendRequest to conn1.clientSide;
              EHRServer.receiveRequest to conn1.serverSide; ... }; };
Bindings { EHRClient1.annMonitoring to EHRSubsystem.annMonitoring;
          EHRServer.annTeleconsultation to EHRSubsystem.annTeleconsultation;
          EHRServer.rcvResult to EHRSubsystem.rcvResult; ... }; };
Component VTCSubsystem1: ParticipantT = new ParticipantT extended with {
Ports { rcvTeleconsultation: p_receive = new p_receive;
      annResult: p_announce = new p_announce; };
Representation rep = { System rep: ClientAndServerFam = new ClientAndServerFam extended with {
Component VTCCClient: ClientT = new ClientT extended with {
Representation rep = { System rep: ClientAndServerFam = new ClientAndServerFam extended with {
Component VTCGUI1: ClientT = new ClientT;
Component VTCWebServer: ServerT = new ServerT extended with {
Ports { sendRequest; }; };
Connector conn1: CConnT = new CConnT extended with { Property protocol: string = "HTTP"; };
Attachments { VTCGUI1.sendRequest to conn1.clientSide;
              VTCWebServer.receiveRequest to conn1.serverSide; ... }; };
Bindings { VTCWebServer.sendRequest to VTCCClient.sendRequest; }; };
Component VTCServer: ServerT = new ServerT extended with {
Ports { rcvTeleconsultation; rcvMonitoring; annResult; }; };
Connector conn: CConnT = new CConnT extended with { Property protocol : string = "SOAP"; };
Attachments { VTCCClient.sendRequest to conn.clientSide;
              VTCServer.receiveRequest to conn.serverSide; }; };
Bindings { VTCServer.rcvTeleconsultation to VTCSubsystem.rcvTeleconsultation;
          VTCServer.annResult to VTCSubsystem.annResult; }; };
Component CSSubsystem: ParticipantT = new ParticipantT extended with {
Ports { annMonitoring, annResult: p_announce = new p_announce;
      RcvResult: p_receive = new p_receive; }; };
Component CCUSubsystem: ParticipantT = new ParticipantT extended with {
Ports { rcvMonitoring: p_receive = new p_receive; }; };
Connector TeleconsultationBus: EventBusT = new EventBusT extended with {
Roles { pubTeleconsultation: r_publisher = new r_publisher;
      subTeleconsultation: r_subscriber = new r_subscriber; }; };
Connector MonitoringBus: EventBusT = new EventBusT extended with {
Roles { pubMonitoring: r_publisher = new r_publisher;
      subMonitoring: r_subscriber = new r_subscriber; }; };
Connector ResultBus: EventBusT = new EventBusT extended with {
Roles { pubResult: r_publisher = new r_publisher;
      subResult: r_subscriber = new r_subscriber; }; };
Attachments { EHRSubsystem1.annTeleconsultation to TeleconsultationBus.pubTeleconsultation;
              EHRSubsystem1.annMonitoring to MonitoringBus.pubMonitoring;
              EHRSubsystem1.rcvResult to ResultBus.subResult;
              VTCSubsystem1.rcvTeleconsultation to TeleconsultationBus.subTeleconsultation;
              VTCSubsystem1.annResult to ResultBus.pubResult;
              CSSubsystem.annResult to ResultBus.pubResult2;
              CSSubsystem.rcvResult to ResultBus.subResult2;
              CSSubsystem.annMonitoring to MonitoringBus.pubMonitoring;
              CCUSubsystem1.rcvMonitoring to MonitoringBus.subMonitoring; ... }; };

```

Listing 10.1 Architecture description in Acme (from [44]).

ent metamodels can be defined and processed by tools offered as plug-ins in the Eclipse open platform.⁷

- Model transformation aims at generating (usually more concrete) models from another (usually more abstract) models based on a set of transformation rules.

⁷ <http://www.eclipse.org>

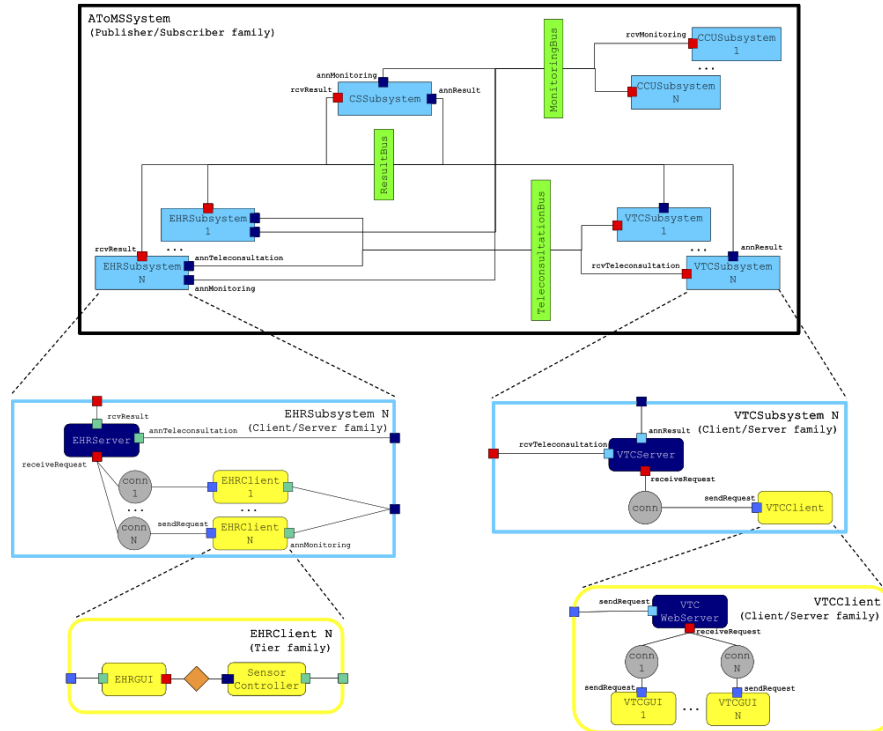


Fig. 10.4 Pictorial representation of the AToMS architecture in Acme (from [44]).

There are different approaches to defining such rules. The most common of such approaches is defining the transformation rules over the metamodels used to build the source and target models. This is the approach adopted by the OMG standards. Figure 10.5 illustrates the relationships among source and target models, and transformation rules. Note that the transformation rules themselves are built on a metamodel that follows the MOF metamodel. In the MDA approach, model transformation is usually employed to generate PSMs from a PIM. Jouault et al. [65] define a transformation language, and associated tools, based on the same operational context as the one depicted in Figure 10.5, which is called ATL (Atlas Transformation Language). ATL is offered as a plug-in to the Eclipse platform, being able to process metamodels based on Ecore.

10.3.2.2 Model-driven engineering in HISes

Exploratory papers like [89, 104] have set an initial basis in terms of the potential that MDA could offer for the systematic engineering of HISes, an important step to

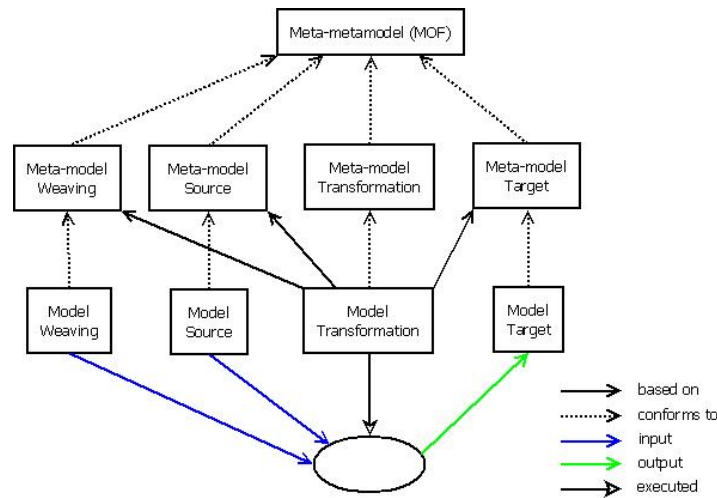


Fig. 10.5 Model transformations according to OMG [65].

reduce the high development cost of HIS. In particular, Raghupathi and Umar [89] specifically argue that MDA allows for (i) low development cost and (hence) reducing the backlog of potential applications; (ii) enabling standards and interoperability; (iii) platform-independence and lessened vendor lock-in; and (iv) enough abstraction to deal with complexity and maintenance. Nevertheless, Tuomainen et al. [104], despite reporting an increase in the use of model-centric approaches in healthcare, are critical about the difficulties in reaching rationale for general design decisions, since non-trivial tuning is often required.

López and Blobel [75] in turn discuss several software engineering methodologies (e.g., RUP – Rational Unified Process), modeling languages (e.g., UML), architectural frameworks (e.g., MDA, EMF) and healthcare standards (e.g., HL7, *openEHR*) in the pursuit of a feasible configuration for enabling semantic interoperability among healthcare systems. The main assumption behind this initiative seems to be a need for integrating methodologies and standards from software engineering in general to those born in health informatics in particular. It is worth highlighting in this sense the difficulty in integrating and evolving proprietary artifacts such as RUP and HL7. On a different path, Janamanchi et al. [62] discuss the potential of dissemination of open artifacts, such as the EMF and the *openEHR* specifications. The authors conclude there exists an active and thriving open-source software development community that is focusing on health and medical informatics.

Actual model-driven HIS implementations have also been reported, though more scarcely. Two European initiatives stand out in this sense. The first one is a Dutch project part of the larger European HealthService 24 initiative,⁸ which is concerned with the application of MDA in the development of Body Area Network (BAN)

⁸ <http://www.healthservice24.com>

healthcare applications [64]. Three generic BANs (viz., LongQT, Diabetes, and Epilepsy) have been modeled in order to derive patient-customized BANs from a combination of them. Special attention is paid to a formal validation of the conformance between a given platform-independent model and its derived implementation. The second initiative is a Norwegian project named MPOWER⁹ that aims at combining Agile Software Development and MDA to design and implement an open platform for simplifying and speeding up the task of developing and deploying services for elderly and people with cognitive disabilities [107]. Their main target end-user applications are individual plan management and smart home and sensor connectivity. The MDA approach is beneficial to MPOWER as it facilitates integration by separating business and application logic from underlying technology [107].

From this brief review on the literature, we highlight that the application of MDE in healthcare is in progress, but the existing initiatives are in general still restricted to the standard MDA approach. The MDA approach itself can be a valuable resource for developing HISes with a separation between healthcare processes and technology. Nevertheless, even though MDA is able to ease the reuse of a good architecture, it still cannot systematically help one to come up with new good architectures. MDA should therefore be augmented in order to improve reuse in terms of application-independent modeling.

10.3.2.3 Discussion

Currently, architecture description languages and corresponding tools, like Acme and AcmeStudio, are provided with exporting features such that a given architecture model can be easily transformed, say, into an Ecore metamodel (in fact, AcmeStudio is implemented as a plug-in to Eclipse platform). Therefore, domain engineering based on software architecture concepts is, at least from a tooling point of view, straightforward to achieve, e.g. by using the Acme language and its Armani extensions to formally describe system families.

10.4 The SPLiCE approach

In our vision, there are two main challenges that should yet be addressed for model-driven HISes to be able to stand for longer life spans:

- a. **Inherent complexity of clinical data modeling.** As we discuss in Section 10.3.1, this is an issue—specially if the clinical data model is to be built from scratch—either when developing a single HIS, or a family of HISes as is often the case for MDE. It requires too much attention from the development team, being it only one piece of the HIS development effort.

⁹ <http://www.sintef.no/mpower>

- b. **Ad-hoc architectural modeling.** As we argue in Section 10.3.2, an effective use of architectural design techniques is specially required in the domain of HISes. Typically, families of HISes (e.g., prehospital emergency, epidemiological surveillance, intrahospital patient monitoring) vary within styles of system communication and control. We argue that these styles deserve research attention in order for us to develop HISes for maintainability and reuse.

Altogether, these two challenges have motivated us in the design of an MDE method that seeks a higher degree of maintainability and reuse. To the best of our knowledge, our proposed method is the first to address HIS data and architecture models in a comprehensive and integrated way, by adhering to three main principles:

- a. **Transversalization of data and architecture models**, so that health domain specialists do not care about how data is to be conveyed, whereas software architects do not bother with what data is to be conveyed.
- b. **Dualization of information and clinical data models**, to lessen the development team's effort on HIS maintenance and evolution.
- c. **Stylization of architecture models**, to capture commonalities within families of HISes as reusable design artifacts.

To support the envisioned MDE method on the generation of HIS skeletons, we are progressing work on the development of a software product line—called SPLiCE (Software Product Line in healthCarE) [45] —that is fully based on open artifacts. The overall architecture of SPLiCE is shown in Figure 10.6. All tools presented in the figure have been prototyped as plug-ins for the Eclipse platform and are discussed throughout the following subsections.

10.4.1 Weaving of data and architecture models

In the SPLiCE approach, clinical data modeling and architectural data modeling are carried out independently and then weaved together in a single, “fully-fledged” model that is then used for generating HIS skeletons, as described in Section 10.4.2.

10.4.1.1 Model transversalization

For the clinical data modeling, we have adopted in SPLiCE the dual-modeling approach proposed by *openEHR*. With this approach, clinical data models are never built from scratch, thus lessening the efforts on HIS maintenance and evolution. Moreover, the dual-modeling approach allows achieving semantic interoperability, which is crucial in recording information in purpose-specific HISes that need to synchronize to other HISes or larger public health databases.

In SPLiCE, both the reference model and the clinical data archetypes of the *openEHR* specifications serve as input to a Data Model Build tool. This tool also receives from the Architecture Model Build tool an indication of the kinds of clinical

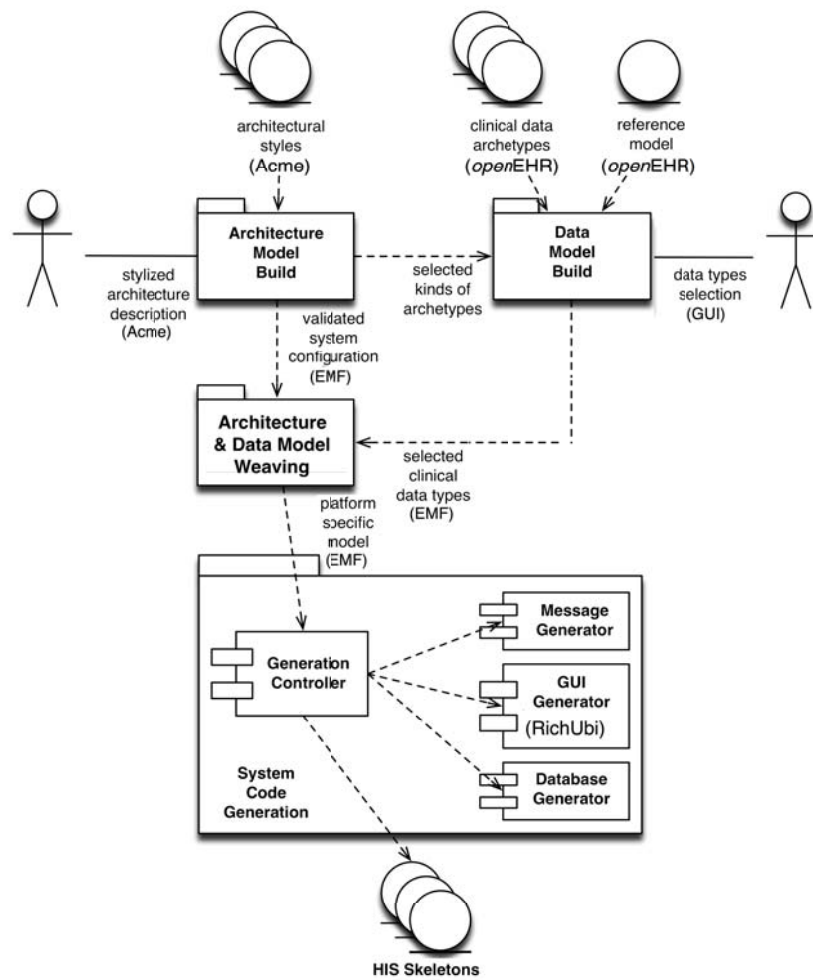


Fig. 10.6 The SPLiCE architecture [45].

data archetypes of interest—observation, instruction, evaluation, or action—for the specific family of HIS the system to be generated makes part of (see below). From these inputs, the Data Model Build tool generates a complete clinical data model, which is presented to the health domain specialist. This professional may then select the data types in the complete clinical data model that are of interest for the specific HIS to be generated—e.g. the “Muscle Tone” data type is crucial in obstetrics (for the Apgar score method) but does not make sense for emergency services to infarction patients. Finally, the Data Model Build tool translates the selected data types (resembling the template idea of *openEHR*) into an EMF-compliant format. In or-

der to achieve compatibility with EMF, the selected data types exported by the Data Model Build tool is persisted by following the XML Metadata Interchange (XMI)¹⁰ schema compliance recommendation from OMG. This standard is used to serialize system models as XML documents so as to allow their interpretation, processing and exchange between different modeling facilities. The XMI schema establishes a valid XML structure, which reflects the relationship between the models' data and their corresponding meta-data—expressed by means of a metamodel. This structure facilitates the models' interpretation and mapping onto code or other refined models when defining ATL-based transformation engines in the Eclipse platform.

For the architectural modeling, we have adopted in SPLiCE the Acme architecture description language. Architectural styles described in Acme serve as input to an Architecture Model Build tool. Such styles represent the rules governing the configuration of systems pertaining to different families of HISes—currently we have been working on Acme styles for prehospital emergency healthcare and epidemiological surveillance systems. The Architecture Model Build tool: (i) validates a system configuration specified by a software architect according to the style rules of the intended family of HISes; (ii) indicates to the Data Model Build tool the kinds of clinical data archetypes of interest for this family—e.g. emergency systems handle mostly observation and instruction data, whereas surveillance systems are usually interested in observation and evaluation data; and (iii) translates the validated system configuration into an EMF-compliant format using XMI schemas, similar to what is done by the Data Model Build tool.

10.4.1.2 Model weaving

The Architecture & Data Model Weaving tool implements a composition operation that combines the validated system configuration exported by the Architecture Model Build tool and the selected data types exported by the Data Model Transformation tool into a single PSM, which serves as input to the System Code Generation tool described in Section 10.4.2.

The composition operation is a special type of transformation that generates a composition model from two different input models. In SPLiCE, we employ model weaving to achieve model composition [34]. Model weaving imports the idea of aspect-oriented software development (AOSD) into the realm of model transformations. The primary objective of model weaving is to associate fine-grained relationships (links) among the elements of distinct models, and establish the links between models as part of the transformation rules. Such links can be defined in the Eclipse platform by using the AMW (ATLAS Model Weaver) plug-in.¹¹ AMW is a model composition framework that uses model weaving and model transformations to produce and execute composition operations [10]. The model resulting from a composition may contain parts or all of the elements of the input models, and it may also

¹⁰ <http://www.omg.org/spec/XMI/2.1.1>

¹¹ <http://www.eclipse.org/gmt/amw>

have new elements. In AMW, the production of a composition operation is divided in two phases. First, a weaving model captures the links between the input model elements. The weaving model conforms to a weaving metamodel, which is dedicated to composition scenarios. It contains elements such as “rename”, “override”, “merge”, and elements specifying how to solve conflicts between the input models. Second, the weaving model is used to generate a transformation rule in ATL, which is then employed in the transformation as the final composition operation.

In the AMW approach, the ATL transformation code complexity is not an issue because the transformation rule is automatically produced. Nevertheless, such complexity may be shifted to the definition of the weaving model depending on the input models being considered in the composition. This is typically the case with the input models generated by the Data Model Build and Architecture Model Build tools.

In the current prototype of the Architecture & Data Model Weaving tool, we adopted the strategy of not dealing with the validated system configuration exported by the Architecture Model Build tool as an input model. Instead, we regard such configuration as “extra data” that annotates the selected data types exported by the Data Model Build tool, similar in concept with what is proposed by de Castro et al. [16]. In this strategy, we define a new weaving metamodel that extends the core weaving metamodel of AMW with architectural constructs. A weaving model instantiating this extended weaving metamodel comprises architectural information exported by the Architecture Model Build tool. This weaving model is used to parameterize a set of general transformation rules defined in ATL that feed the transformation engines of the System Code Generation tool, as described in Section 10.4.2. In this sense, the PSM generated by the current prototype of the Architecture & Data Model Weaving tool is in fact a set comprising the selected data types originally exported by the Data Model Build tool and the weaving model generated from the validated system configuration exported by the Architecture Model Build tool.

10.4.2 GUI and messaging generators

The System Code Generation tool implements an EMF-based transformation engine that generates HIS skeletons from the PSM exported by the Architecture & Data Model Weaving tool.

Each part of the PSM may be of particular interest for different parts of the HIS being generated: (i) for the definition of a suitable GUI layout for the interaction of end users with the HIS according to the clinical data archetypes to be manipulated in the system; (ii) for the structure of messages and protocol entities used for the transportation of clinical data; and (iii) the system entities that represent the persistent clinical data previously selected by the health domain.

The internal structure of the System Code Generation tool reflects this separation of concerns, which comprises four components. The Generation Controller component parses the PSM and invokes each of the three specific generator components: (i) the Message Generator component for generating protocol entities and mes-

sages (e.g. WSDL definitions, XMPP stanzas); (ii) the GUI Generator component, which is based on the Model Driven RichUbi process' support tooling, a third-party, EMF-compliant tool set that generates code for GUIs of context-sensitive applications [22]; and (iii) the Database Generator component for generating database-specific code (e.g. schemas, O/R mapping files). Currently, our prototype implementation of SPLiCE only covers the three first components. In the following, we describe the GUI and Message Generator components.

10.4.2.1 GUI Generator component

The component GUI Generator is implemented as a two-step pipeline. The first step is as a transformation engine that employs the ATL and AMW plug-ins of the Eclipse platform. Figure 10.7 illustrates the main inputs and outputs of this step. Basically, the GUI Generator component consumes the clinical and annotation models that comprise the PSM exported by the Architecture & Data Model Weaving tool, and employs the transformation rules defined in ATL that are stored in its rules repository to generate a GUI clinical model which is based on the Model Driven RichUbi [22], a process that provides support for the development of rich interfaces for context-sensitive applications.

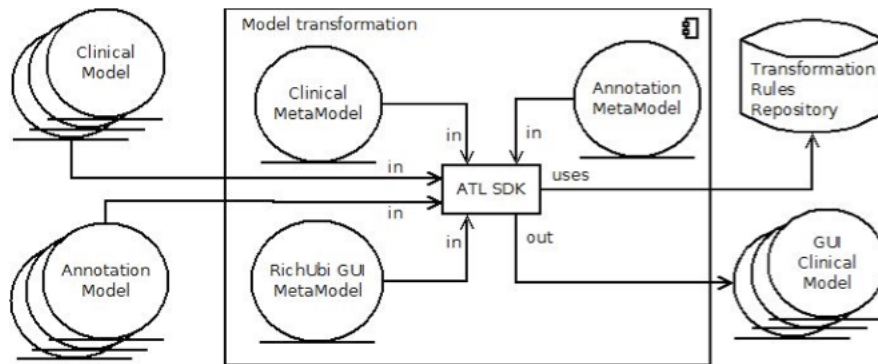


Fig. 10.7 The GUI Generator component.

The second step of the pipeline employs Model-to-Code (M2C) transformations built in the Model Driven RichUbi process. RichUbi includes: (i) a Rich Interfaces Domain metamodel, which expresses the abstract syntax of a DSL [93] to aid rich interfaces modeling; (ii) M2C transformations for code generation from the rich interfaces models; and (iii) dynamic content adapters to provide adaptive behavior to the developed interfaces. The M2C transformations were incorporated into the GUI Generator component in order to enable it to generate the system's user interfaces by applying the preset code generation rules from the transformations. The interface's structure is defined earlier in the phase in which the clinical and annotation

models are weaved together. The clinical model indicates the suitable components for interaction of end users, whether for input of data or for visualization of clinical records. For instance, in an interface for entering clinical information associated with a particular openEHR archetype—such as “Blood Pressure”—all pieces of information associated with data types that restrict the values of the more general “Quantity” data type from the openEHR reference model—such as “Systolic” and “Diastolic”—are exported as Text Field primitives from the Rich Interfaces Domain metamodel. The minimum and maximum value constraints for these are also configured so as to conform to the restrictions imposed by the archetype modelers. On the other hand, in an interface for visualization of blood pressure records, those same pieces of information are exported as simple Label primitives. What determines whether the interface to be generated involves Text Field or Label primitives (cf. the paramedics and specialist endpoints of the AToMS system presented in Section 10.2.1) is the annotation model, which conveys architectural information from the validated system configuration exported by the Architecture Model Build tool. Table 10.1 shows the correlation between the main archetype data types and their corresponding interaction primitives—which ones to be used, input data or visualization, being determined by the annotation model—in RichUbi.

Table 10.1 Correlation between clinical and annotation models and RichUbi primitives.

Archetype data type in the <i>openEHR</i> Reference Model	Component primitive in RichUbi	
	Input data	Visualization
CComplexObject.rmTypeName = 'ELEMENT'	FieldSet	Div
CComplexObject.rmTypeName = 'DV_TEXT'	TextArea + Label	Span
CComplexObject.rmTypeName = 'ELEMENT'	FieldSet	Div
CComplexObject.rmTypeName = 'DV_CODED_TEXT'	Select	Span
CCodedPhrase.codeList = List<String>	SelectItem	Span
CDVQUANTITY	TextField + Label	Span
CQUANTITYTEN	PhraseElement	Span
IntervalOfReal	PhraseElement	Span

The M2C transformations of RichUbi are implemented as transformation templates, which are text files instrumented with constructs for selecting and expanding code. These constructs perform queries on a given input (e.g. a XMI file representing a model) and use the outcome (e.g. the data read from the input model) as parameter to produce custom code in any textual language [26]. In this sense, templates are usually composed of fixed parts, which always are included in the output code, and variable parts, which depend on the data contained in the input model to be generated. To implement the transformations, the Java Emitter Templates (JET)¹² framework has been used. This framework provides a library of metaprogramming markups that implement conditional, looping and formatting statements, and other useful functions to query the models given as input and generating code.

¹² <http://www.eclipse.org/modeling/m2t/?project=jet\#jet>

As an example, Figure 10.8 shows a fragment of the JET template that generates the eXtensible Hypertext Markup Language (XHTML)¹³ code of the Tabbed-Panel interaction component, one of the primitives defined in the Rich Interfaces Domain metamodel. This figure also illustrates the execution process of the template, in which each template's statement is interpreted in order to query the input model (in XMI format) and then producing its corresponding code. Thus, whenever the mechanism that executes transformations (or template processor) finds the node representing the tabbed panel component in the XMI input model, the execution of the template starts. The template's lines 1-12 generate the custom JavaScript code to render the tabbed panel on the user's Web browser. Line 14 produces a `<div>` markup that makes up the tabbed panel's structure. This markup references a pre-set style class named *demo*. In lines 16-19 and 21-28 iterations are made on the panel's tabs defined in the model to generate their content in the output code.

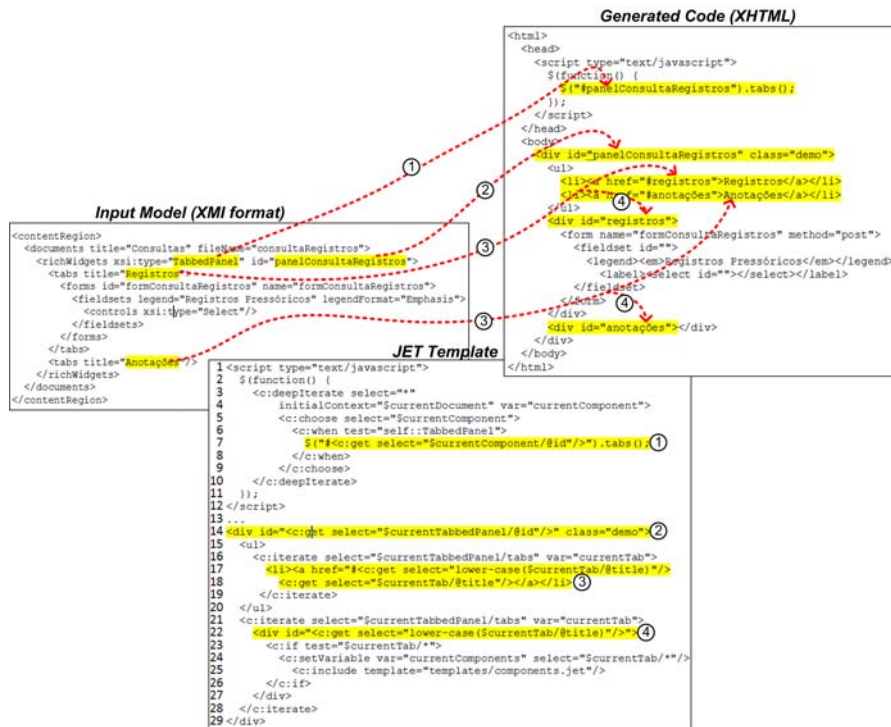


Fig. 10.8 Fragment of the JET template for generating XHTML code in RichUbi.

¹³ <http://www.w3.org/TR/xhtml1>

10.4.2.2 Messaging Generator component

To ensure the interoperability between the SPLiCE compliant applications, and especially between them and third party applications, the component Messaging Generator has its development strictly based on specific communication protocols of the healthcare domain. This component is responsible for the semiautomatic generation of protocol entities and messages, providing the partial code of these artifacts.

In our SPLiCE approach, we consider two strategies for the development of the Messaging Generator component: (i) employing the *openEHR* extracts specifications; or (ii) using the HL7 standard to communicate *openEHR*-compliant data.

openEHR extracts

In this strategy, the Messaging Generator component uses the *openEHR* RM Java implementation to allow the implementation of the new classes in the extended RM, and the extract package in order to allow the communication of EHR extracts. This strategy consists of two parts: RM, delivering the EHR information container needed on the one hand, and the EHR meta-models, called AM for expressing the clinical content on the other hand. Each part has the corresponding models that are also called Reference Model Schema (RM-XMLBeans) based on XMLBeans and Archetype Model Schema (AM-ADL). XMLBeans¹⁴ is a technology for accessing XML by binding it to Java types, and uses XML Schema to compile Java interfaces and classes that are used to access and modify XML instance data.

The Figure 10.9 shows the generated XMLBeans code based on the Java *openEHR* RM implementation. Arrow (1) in this figure points to the attribute generated based on RM implementation, and Arrow (2) to the type of such attribute.

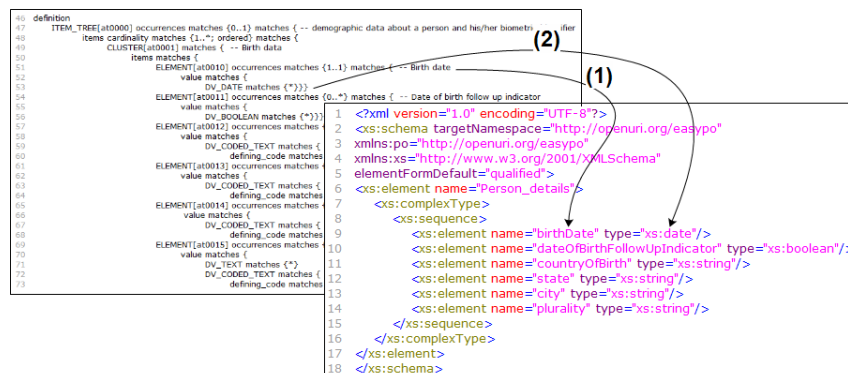


Fig. 10.9 Generated XMLBeans code.

¹⁴ <http://xmlbeans.apache.org/index.html>

The generated code represents an *openEHR* extract containing data about the patient, which can be sent among distributed components of a HIS as representational states using RESTful web services.

HL7 Messaging

In this strategy, the Messaging Generator component is also implemented as a two-step pipeline, in a similar fashion as the GUI Generator component presented in Section 10.4.2.1. The differences lie, for the first step, in the way the weaving model is built and, for the second step, in the target metamodel used (and, consequently, the transformation rules used in the first step).

In the first step of the pipeline, it was initially intended that Model-to-Model (M2M) transformations were made with the least possible human intervention, where the clinical and weaving models would be given as input and the target model (herein called communication model) would be generated as output. However, while the clinical metamodel, which is based on *openEHR*, shows totally generic structures, the communication metamodel, which is based on HL7 v3 (as described below), brings structures that already contain a semantic content. This divergence between the metamodels brought barriers to implementation of a fully automatic approach, since we have not found yet a way to find out beforehand the semantics of a given clinical model structure, which would allow the correct mapping between it and its related structure in the communication metamodel. The found solution to the standoff was to insert the user in the process, which should provide the semantic links between the source and target models as a weaving model. Then, this weaving model will serve as the basis for generating the correct transformations.

For the second step of the pipeline, a communication metamodel has been developed specifically to meet the communication needs of the applications, and reflects the specifications of the RIM defined in the HL7 v3 standard. As such, the communication metamodel provides an explicit representation of relationships between the objects that make up the healthcare domain, specifying the basic components, the allowed relationships and the data types used in the messages development.

Figure 10.10 shows part of the communication metamodel. The basis of such model is made up of three main metaclasses—Act, Role, and Entity—whose relationships are represented by ActRelationship, Participation, and RoleLink metaclasses and several specializations. These main metaclasses are grouped into Subject Areas, which are distinguished by color and represent the functions of the metaclasses in the metamodel. The Acts Subject Area, in red, includes the Act metaclass and its specializations. This group is related to actions and events that make up the services of the healthcare. Entities Subject Area, in green, includes the Entity metaclass, their specializations and other related meta-classes. This group represents the stakeholders and other objects of interest to the Healthcare. Finally, Roles Subject Area, in yellow, includes the Role metaclass and its specializations. This group focuses on the roles that participants can play in the interactions of the healthcare [53].

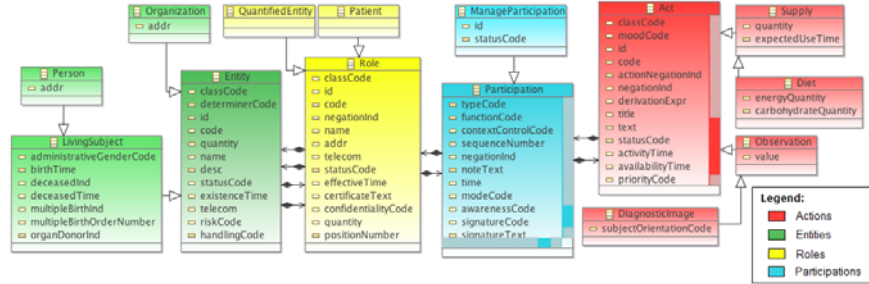


Fig. 10.10 Communication metamodel [53].

Like the GUI Generator component, the M2C transformations used in the second step of the pipeline are based on templates. The execution of the templates used in the partial code generation produces Model Interchange Format (MIF) codes, used by the HL7 committee in representation of all its artifacts schemes [53]. A MIF file is a set of XML formats used to support the storage and exchange of artifacts from the HL7 v3 development process. Validation rules are also present in MIF files, allowing besides the validation of the artifacts, the creation of a documentation that describes what the semantics of the different parts of these artifacts represent [99]. Figure 10.11 shows an excerpt of the JET template responsible for generating the MIF code related to the Observation meta-class. As we can see in lines 4 and 6, iterations are performed in the input clinical model in order to find variable information and assign them to the final code. Calls to other templates can also be performed during processing, as exemplified in the line 19. In this case, the template “participation.jet” was invoked due to the inclusion of a Participation in the model, which will be processed and included in the part where it was called.

Figure 10.12 shows an excerpt of the MIF code generated from the JET template shown in Figure 10.11. This excerpt illustrates how the partial generation of the messages schemas will be done based on models developed. The excerpts detached on the images show the changes made, where the fixed parts of the template were properly copied to the output code (e.g., line 1 of Figure 10.11), and the flexible parts were written according to the information of the input model (e.g., codes from the “participation.jet” template processing of the line 19 of Figure 10.11).

Upon completion of this prototypical stage of development, the aim is also to make this version of the Message Generator component comprise other communication protocols, either specific of this domain or more general, such as HTTP or Jabber. The software architect must make the selection of the best protocol for each case in the early stages of use of SPLiCE, reaching the Message Generator component as architectural information. This will provide greater flexibility to SPLiCE, allowing it to meet the varied communication needs of the applications.

```

[...]
```

- 1 <class name="Observation" isAbstract="false">
- 2 <derivationSupplier staticModelDerivationId="1" className="Observation"/>
- 3 <annotations/>
- 4 <c:deepIterate select="\$currentObs/*" var="currentElement">
- 5 <c:choose select="\$currentElement">
- 6 <c:when test="local-name(\$currentElement) = 'classCode'">
- 7 <attribute name="classCode" sortKey="" isStructural="true" minimumMultiplicity=
- 8 "<c:get select='\$currentElement/@minMult'/">" maximumMultiplicity="<c:get select=
- 9 "\$currentElement/@maxMult"/>" isMandatory="true" conformance="M" defaultValue=
- 10 "<c:get select='\$currentElement/@code'/">">
- 11 <annotations/>
- 12 <derivationSupplier staticModelDerivationId="1" className="Act" attributeName="classCode"/>
- 13 <type name="CS"/>
- 14 <supplierDomainSpecification codingStrength="CNE" domainName="ActType" mnemonic=
- 15 "<c:get select='\$currentElement/@code'/">">
- 16 </attributes>
- 17 </c:when>
- 18 </targetConnection>
- 19 <c:include template="templates/participation.jet"/>
- 20 </targetConnection>
- 21 [...]

Fig. 10.11 JET template.

```

[...]
```

- 1 <class name="Observation" isAbstract="false">
- 2 <derivationSupplier staticModelDerivationId="1" className="Observation"/>
- 3 <annotations/>
- 4 <attribute name="classCode" sortKey="1" isStructural="true" minimumMultiplicity="1" maximumMultiplicity="1"
- 5 isMandatory="true" conformance="M" defaultValue="OBS">
- 6 <annotations/>
- 7 <derivationSupplier staticModelDerivationId="1" className="Act" attributeName="classCode"/>
- 8 <type name="CS"/>
- 9 <supplierDomainSpecification codingStrength="CNE" domainName="ActType" mnemonic="OBS"/>
- 10 </attribute>
- 11 [...]
- 12 <association sortKey="1">
- 13 <targetConnection name="obsSbj" minimumMultiplicity="1" maximumMultiplicity="1" isMandatory="true" sortKey="1">
- 14 <annotations/>
- 15 <derivationSupplier staticModelDerivationId="1" className="Act" associationEndName="participation"/>
- 16 [...]

Fig. 10.12 MIF code.

10.5 Summary and Outlook

The information related to human health is inherently complex at the intersections of space, time, and knowledge. This complexity precludes the feasibility of one, single, all encompassing electronic health record for individuals in a population, since various pieces of information from a number of different applications may be needed at any point in time. Semantic interoperability is crucial in recording information in purpose specific applications that need to synchronize to larger databases. This is the way to provide useful information from the point of care to the healthcare services in a timely manner.

Regarding public health, the typical turnaround time now for acquiring the healthcare status of a population can be weeks or even months while experts pour over the data and try to merge pieces from various applications and paper forms into something meaningful. By basing applications on a common information model

and using a constraint-based approach to define the knowledge components, we can achieve semantic interoperability and near real-time information regarding the healthcare status of a population and its individuals so that faster early action can be taken to meet the needs in that area.

The aforementioned approach may also help empowering application developers at the local level when combined with model-driven software engineering methods, so they can develop healthcare applications fitted to the very specific local needs without losing semantic coherence and interoperability with other local services and with the highest levels of the healthcare systems. From a software engineering viewpoint, however, appropriate information modeling is necessary but insufficient to render maintainable and reusable HISes. In particular, this approach is limited with regard to situations that press for unconventional connections among the software components of a HIS, as for instance in disaster scenarios. In this chapter, we argued that for allowing the specification of innovative software architectures we must leverage the capabilities of current model-driven approaches.

Building upon this rationale, we proposed in this chapter a model-driven method that combines health information and software architecture modeling in a single workflow of M2M and M2C transformations, which leads to a comprehensive software product line for healthcare applications. We have progressed work on the implementation of a workbench called SPLiCE that supports our proposed method. So far, we have prototyped key components of our workbench, including: a translator of Acme specifications to EMF-based system configurations; a translator of openEHR archetypes in ADL to EMF-based clinical data models; the specification of a weaving metamodel based on AMW to annotate EMF-based clinical data models with architectural information extracted from EMF-based system configurations; and the specification of ATL transformation rules that generate RichUbi and HL7-based communication models from annotated EMF-based clinical data models.

To sum up, adopting 21st Century healthcare based on 20th Century information systems will not easily solve inequalities. There are no remaining significant obstacles related to hardware, including mobile computing and pervasive healthcare, but software based on traditional data models and without principled architectural support during its design is not fitted to deal with the significant spatial and temporal complexities of the healthcare domain. We expect that our proposed model-driven method and related workbench will help considerably in tackling such complexities and achieving higher levels of maintainability and reuse of HIS design and code.

Acknowledgements

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. Iuri Teixeira is supported by the Brazilian agency FAPEMIG. Carlos Cirilo is supported by the Brazilian agency CAPES. Wanderley de Souza is supported by the Brazilian agency FAPESP. The support of these agencies is gratefully acknowledged.

References

- [1] ACMI: 1961 electronic medical records (2008), <http://www.youtube.com/watch?v=t-aiKlIc6uk>
- [2] Atalag, K., Yang, H., Tempero, E., Warren, J.: Model driven development of clinical information systems using openEHR. *Stud Health Technol Inform* 169 (2011)
- [3] Bass, L., Clements, P., Kazman, R.: *Software architecture in practice*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1998)
- [4] Beale, T., Heard, S.: openEHR Architecture Overview (2007)
- [5] Beale, T., Heard, S.: Archetype Definition Language (2008), <http://www.openehr.org/releases/1.0.2/architecture/am/adl.pdf>
- [6] Beale, T., Heard, S., Kalra, D., Lloyd, D.: EHR Information Model (2008), http://www.openehr.org/releases/1.0.2/architecture/rm/ehr_im.pdf
- [7] Beale, T.: Archetypes and the EHR. *Stud Health Technol Inform* 96, 238–44 (2003)
- [8] Beale, T.: openEHR Architecture (2008), <http://www.openehr.org/releases/1.0.2/architecture/overview.pdf>
- [9] Beeler, G.: HL7 Version 3—An object-oriented methodology for collaborative standards development. *International Journal of Medical Informatics* 48(1-3), 151–161 (Feb 1998)
- [10] Bézivin, J., Bouzitouna, S., Del Fabro, M., Gervais, M.P., Jouault, F., Kolovos, D., Kurtev, I., Paige, R.F.: A Canonical Scheme for Model Composition. pp. 346–360 (2006)
- [11] Blobel, B.: Ontologies, Knowledge Representation, Artificial Intelligence - Hype or Prerequisites for International pHealth Interoperability? *Stud Health Technol Inform* 165, 11–20 (2011)
- [12] Blobel, B., Oemig, F., Gonzales, C., Lopez, D.: What is missing in health informatics standardization for pHealth? *Stud Health Technol Inform* pp. 3–12 (2010)
- [13] Blobel, B., Pharo, P.: Analysis and evaluation of EHR approaches. *Methods Inf Med* 48(2), 162–9 (2009)
- [14] Boulos, M., Wheeler, S., Tavares, C., Jones, R.: How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. *BioMedical Engineering OnLine* 10(24) (Apr 2011)
- [15] Cantrill, S.V.: Computers in patient care: The promise and the challenge. *Commun. ACM* 53(9), 42–47 (2010)
- [16] de Castro, V., Vara, J.M., Marcos, E.: Model transformation for service-oriented web applications development. In: Koch, N., Vallecillo, A., Houben, G.J. (eds.) MDWE. *CEUR Workshop Proceedings*, vol. 261. CEUR-WS.org (2007)
- [17] Cavalini, L.T., Cook, T.W.: Health informatics: The relevance of open source and multi-level modeling. In: Hissam, S.A., Russo, B., de Mendonça Neto, M.G., Kon, F. (eds.) OSS. IFIP Publications, vol. 365, pp. 338–347. Springer (2011)
- [18] Chen, M.: Factors affecting the adoption and diffusion of XML and web services standards for E-business systems. *Int. J. Hum.-Comput. Stud.* 58(3), 259–279 (mar 2003)
- [19] Chen, R.S., Nadkarni, P., Marengo, L., Levin, F., Erdos, J., Miller, P.L.: Exploring performance issues for a clinical database organized using an representation. *Journal of the American Medical Informatics Association* 7(5), 475–487 (2000)
- [20] Choi, B.C.K.: Perspectives on epidemiologic surveillance in the 21st century. *Chron Dis Can* 19(4) (2000)
- [21] Ciottl, V.V.: Generation HL7. of the 15 founding fathers of HL7, just one remains today, leaving many to question whether interoperability is attainable. *Health Inform* 27(1), 48–48 (2010)
- [22] Cirilo, C.E., do Prado, A.F., de Souza, W.L., Zaina, L.A.M.: Model driven RichUbi: a model driven process for building rich interfaces of context-sensitive ubiquitous applications. In: Anacleto, J.C., de Mattos Fortes, R.P., Costa, C.J. (eds.) SIGDOC. pp. 207–214. ACM (2010)

- [23] Clemmensen, P., Sillesen, M., Sejersten Ripa, M., Strange, S., Loumann-Nielsen, S.: Pre-hospital diversion to hospital with acute PCI set-up using wireless 12-lead ECG transmission. *Journal of electrocardiology* 38(4), 194–198 (2005)
- [24] Correa, B.S.P.M., Gonçalves, B., Teixeira, I.M., Gomes, A.T.A., Ziviani, A.: AToMS: a ubiquitous teleconsultation system for supporting AMI patients with prehospital thrombolysis. *Int. J. Telemedicine Appl.* 2011, 2:1–2:12 (jun 2011)
- [25] Correndo, G., Terenziani, P.: Towards a flexible integration of clinical guideline systems with medical ontologies and medical information systems. In: *Proc. of the Symposium on Computerized Guidelines and Protocols* (April 2004)
- [26] Czarnecki, K., Eisenecker, U.W.: *Generative programming - methods, tools and applications*. Addison-Wesley (2000)
- [27] Daconta, M.C., Obrst, L., Smith, K.T.: *The semantic web - a guide to the future of XML, web services, and knowledge management*. Wiley (2003)
- [28] Danchin, N., Durand, E., Blanchard, D.: Pre-hospital thrombolysis in perspective. *Eur Heart J* 29(23), 2835–42 (2008)
- [29] Dağtas, S., Pekhteryev, G., Sahinoğlu, Z., Çam, H., Challa, N.: Real-time and secure wireless health monitoring. *Int. J. Telemedicine Appl.* 2008, 1:1–1:10 (Jan 2008)
- [30] De Leon, S., Connelly-Flores, A., Mostashari, F., Shih, S.C.: The business end of health information technology. can a fully integrated electronic health record increase provider productivity in a large community practice? *J Med Pract Manage* 25(6), 342–9 (2010)
- [31] Denning, P.J.: Hastily formed networks. *Commun. ACM* 49(4), 15–20 (Apr 2006), <http://doi.acm.org/10.1145/1121949.1121966>
- [32] van Deursen, A., Klint, P., Visser, J.: *Domain-specific languages: an annotated bibliography* (2004), <http://www.cis.uab.edu/courses/managed/cs593/spring2004/DSLAnnotatedBib.pdf>
- [33] Dias, R., Cook, T., Freire, S.: Modeling healthcare authorization and claim submissions using the openEHR dual-model approach. *BMC Medical Informatics and Decision Making* 11(1), 60+ (Oct 2011)
- [34] Didonet, M., Fabro, D., Bèzivin, J., Valduriez, P.: Weaving models with the eclipse amw plugin. In: *In Eclipse Modeling Symposium, Eclipse Summit Europe* (2006)
- [35] Dolin, R.H., Alschuler, L., Beebe, C., Biron, P.V., Boyer, S.L., Essin, D., Kimber, E., Lincoln, T., Mattison, J.E.: The HL7 Clinical Document Architecture. *Journal of the American Medical Informatics Association* 8(6), 552–569 (Nov 2001)
- [36] Essendi, H., Mills, S., Fotso, J.C.: Barriers to formal emergency obstetric care services' utilization. *Journal of urban health bulletin of the New York Academy of Medicine* 88 Suppl 2, S356–S369 (2010)
- [37] Federal Communications Commission: Seventh 706 report. Tech. rep., Federal Communications Commission (2010), <http://www.fcc.gov/broadband/706.html>, retrieved on February 23rd, 2011
- [38] Filgueira, R., Odriozola, A., Simini, F.: Using OpenEHR in SICTI an electronic health record system for critical medicine. *Journal of Physics: Conference Series* 90 (dec 2007)
- [39] Garde, S., Chen, R., Leslie, H., Beale, T., McNicoll, I., Heard, S.: Archetype-based knowledge management for semantic interoperability of electronic health records. In: *Adlassnig, K.P., Blobel, B., Mantas, J., Masic, I. (eds.) MIE. Studies in Health Technology and Informatics*, vol. 150, pp. 1007–1011. IOS Press (2009)
- [40] Garde, S., Knaup, P., Schuler, T., Hovenga, E.: Can openEHR archetypes empower multi-centre clinical research? *Stud Health Technol Inform* 116, 971–976 (2005)
- [41] Garlan, D., Allen, R., Ockerbloom, J.: Exploiting style in architectural design environments. In: *Proceedings of SIGSOFT'94*. ACM Press (December 1994)
- [42] Garlan, D., Monroe, R.T., Wile, D.: Acme: An architecture description interchange language. In: *Proceedings of CASCON'97*. pp. 169–183. Toronto, Ontario (November 1997)
- [43] Garlan, D., Perry, D.E.: Introduction to the special issue on software architecture. *IEEE Trans. Softw. Eng.* 21(4), 269–274 (Apr 1995)

- [44] Gomes, A.T.A., Batista, T.V., Joolia, A., Coulson, G.: Architecting dependable systems IV. chap. Architecting dynamic reconfiguration in dependable systems, pp. 237–261. Springer-Verlag, Berlin, Heidelberg (2007)
- [45] Gomes, A.T.A., Ziviani, A., Correa, B.S.P.M., Teixeira, I.M., Moreira, V.M.: SPLiCE: a software product line for healthcare. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. pp. 721–726. IHI '12, ACM, New York, NY, USA (2012)
- [46] Goodburn, E.A.: Using process indicators to monitor and evaluate obstetric services in developing countries. *J Am Med Womens Assoc* 57(3), 145–8 (2002), <http://www.biomedsearch.com/nih/Using-process-indicators-to-monitor/12146605.html>
- [47] Google Inc.: Google health has been discontinued (2012)
- [48] Grim, P., Feldman, T., Martin, M., Donovan, R., Nevins, V., Childers, R.: Cellular telephone transmission of 12-lead electrocardiograms from ambulance to hospital. *The American journal of cardiology* 60(8) (09 1987)
- [49] Grimshaw, J., Russell, I.: Achieving health gain through clinical guidelines. i: Developing scientifically valid guidelines. *Quality in health care : QHC* 2(4) (12 1993)
- [50] Hadzievski, L., Bojovic, B., Vukcevic, V., Belicev, P., Pavlovic, S., Vasiljevic-Pokrajcic, Z., Ostojic, M.: A novel mobile transtelephonic system with synthesized 12-lead ECG. *IEEE Transactions on Information Technology in Biomedicine* 8(4) (12 2004)
- [51] Hailpern, B., Tarr, P.L.: Model-driven development: The good, the bad, and the ugly. *IBM Systems Journal* 45(3), 451–462 (2006)
- [52] Haux, R.: Medical informatics: Past, present, future. *I. J. Medical Informatics* 79(9), 599–610 (2010)
- [53] HL7: HL7 Version 3 Ballot (2011)
- [54] Hogan, M.C., Foreman, K.J., Naghavi, M., Ahn, S.Y., Wang, M., Makela, S.M., Lopez, A.D., Lozano, R., Murray, C.J.L.: Maternal mortality for 181 countries, 1980–2008: a systematic analysis of progress towards millennium development goal 5. *Lancet* 375(9726), 1609–23 (2010)
- [55] Hoot, N., Aronsky, D.: Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine* 52(2), 126–136.e1 (Aug 2008)
- [56] Hripcsak, G., Soulakis, N.D., Li, L., Morrison, F.P., Lai, A.M., Friedman, C., Calman, N.S., Mostashari, F.: Syndromic surveillance using ambulatory electronic health records. *Journal of the American Medical Informatics Association : JAMIA* 16(3), 354–361 (2009)
- [57] Hudson, D., Cohen, M.: Uncertainty and complexity in personal health records. In: Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. pp. 6773–6776 (31 2010-sept 4 2010)
- [58] Hung, K., Zhang, Y.T.: Implementation of a WAP-based telemedicine system for patient monitoring. *IEEE Transactions on Information Technology in Biomedicine* 7(2), 101–107 (2003)
- [59] Hyman, W.A.: When medical devices talk to each other: the promise and challenges of interoperability. *Biomedical Instrumentation & Technology* pp. 28–31 (2010)
- [60] ISO: IS 13606:2008. Electronic health record communication - Part 1: Reference model (2008)
- [61] ISO: IS 13606:2008. Electronic health record communication - Part 2: Archetype interchange specification (2008)
- [62] Janamanchi, B., Katsamakas, E., Raghupathi, W., Gao, W.: The state and profile of open source software projects in health and medical informatics. *I. J. Medical Informatics* 78(7), 457–472 (2009)
- [63] Javitt, J.C.: How to succeed in health information technology. *Health Aff (Millwood) Suppl Web Exclusives*, W4–321–4 (2004)
- [64] Jones, V., Rensink, A., Brinksmma, E.: Modelling mobile health systems: an application of augmented mda for the extended healthcare enterprise.. In: EDOC. pp. 58–69. IEEE Computer Society (2005)

- [65] Jouault, F., Allilaire, F., Bézivin, J., Kurtev, I.: Atl: a model transformation tool. *Science of Computer Programming* 72(1-2), 31–39 (June 2008)
- [66] Kalra, D.: Clinical Foundations and Information Architecture for the Implementation of a Federated Health Record Service. Master's thesis, University of London (2002)
- [67] Kalra, D.: Electronic health record standards. *Yearbook of medical informatics* pp. 136–144 (2006)
- [68] Kalra, D., Beale, T., Heard, S.: Regional Health Economies and ICT Services: The PICNIC Experience, chap. The openEHR Foundation, pp. 153–173. *Studies in Health Technology and Informatics*, IOS Press (2005)
- [69] Kashfi, H.: The intersection of clinical decision support and electronic health record: A literature review. In: FedCSIS. pp. 347–353 (2011)
- [70] Keeley, E., Boura, J., Grines, C.: Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *The Lancet* 361(9351), 13–20 (Jan 2003)
- [71] Kwak, M.J., Kim, J.M., Shin, I.H., Shin, S.D., Song, K.J., Suh, G.J., Kim, H.C.: Real-time medical control using a wireless audio-video transmission device in a pre-hospital emergency service in Korea. *J Telemed Telecare* 15(8), 404–8 (2009)
- [72] Laurenti, R., Jorge, M.H.P.d.M., Gotlieb, S.L.A.D.: Maternal mortality in Brazilian State Capitals: some characteristics and estimates for an adjustment factor. *Revista Brasileira de Epidemiologia* 7, 449 – 460 (12 2004)
- [73] Lin, Y.H., Jan, I.C., Ko, P.C.I., Chen, Y.Y., Wong, J.M., Jan, G.J.: A wireless PDA-based physiological monitoring system for patient transport. *IEEE Transactions on Information Technology in Biomedicine* 8(4), 439–447 (2004)
- [74] López, D.M., Blobel, B.: Architectural approaches for hl7-based health information systems implementation. *Methods Inf Med* 49(2), 196–204 (2010)
- [75] López, D.M., Blobel, B.: A development framework for semantically interoperable health information systems. I. *J. Medical Informatics* 78(2), 83–103 (2009)
- [76] Luckham, D.C., Kenney, J.J., Augustin, L.M., Vera, J., Bryan, D., Mann, W.: Specification and analysis of system architecture using rapide. *IEEE Transactions on Software Engineering* 21, 336–355 (1995)
- [77] Martínez-Costa, C., Tortosa, M.M., Fernández-Breis, J.T.: An approach for the semantic interoperability of ISO EN 13606 and openEHR archetypes. *Journal of Biomedical Informatics* 43(5), 736–746 (2010)
- [78] Massad, E., Burattini, M., Lopez, L., Coutinho, F.: Forecasting versus projection models in epidemiology: the case of the SARS epidemics. *Medical hypotheses* 65(1) (2005)
- [79] McGinn, C., Grenier, S., Duplantie, J., Shaw, N., Sicotte, C., Mathieu, L., Leduc, Y., Legare, F., Gagnon, M.P.: Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. *BMC Medicine* 9(1), 46 (2011)
- [80] McLean, S., Wild, S., Connor, P., Flapan, A.D.: Treating ST elevation myocardial infarction by primary percutaneous coronary intervention, in-hospital thrombolysis and prehospital thrombolysis. an observational study of timelines and outcomes in 625 patients. *Emerg Med J* 28(3), 230–6 (2011)
- [81] Mead, C.: Data interchange standards in healthcare it—computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? *Journal of Healthcare Information Management* 20(1), 71–78 (2006)
- [82] Metaxiotis, K.S., Ptochos, D., Psarras, J.E.: E-health in the new millennium: a research and practice agenda. *IJEH* 1(2), 165–175 (2004)
- [83] Miller, J., Mukerji, J.: MDA Guide Version 1.0.1. Tech. rep., Object Management Group (OMG) (2003)
- [84] Monroe, R.T., Overview, A.: Capturing software architecture design expertise with Armani. Technical report cmu-cs- 98-163, Carnegie Mellon University (1998)
- [85] National Health Service Media Centre: Dismantling the NHS national programme for IT (2011)

- [86] Nyamtema, A., Urassa, D., van Roosmalen, J.: Maternal health interventions in resource limited countries: a systematic review of packages, impacts and factors for change. *BMC Pregnancy and Childbirth* 11(1), 30 (2011)
- [87] Paul, R., Hoque, A.S.L.: Search efficient representation of healthcare data based on the HL7 RIM. *JCP* 5(12), 1810–1818 (2010)
- [88] Prytherch, H., Massawe, S., Kuelker, R., Hunger, C., Mtatifikolo, F., Jahn, A.: The unmet need for emergency obstetric care in Tanga region, Tanzania. *BMC Pregnancy and Childbirth* 7(1), 16 (2007)
- [89] Raghupathi, W., Umar, A.: Exploring a model-driven architecture (MDA) approach to health care information systems development. *I. J. Medical Informatics* 77(5), 305–314 (2008)
- [90] Rath, D.: Shifting away from silos. the interoperability challenges that hospitals face pale in comparison to the headaches plaguing state departments. *Healthc Inform* 27(1), 32–3 (2010)
- [91] Rinner, C., Janzek-Hawlat, S., Sibinovic, S., Duftschmid, G.: Semantic validation of standard-based electronic health record documents with w3c xml schema. *Methods Inf Med* 49(3), 271–80 (2010)
- [92] Sachs, J., McArthur, J.: The millennium project: a plan for meeting the millennium development goals. *The Lancet* 365(9456), 347–353 (1 2005)
- [93] Sadilek, D.A.: Prototyping domain-specific language semantics. In: *OOPSLA Companion '08: Companion to the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications*. pp. 895–896. ACM, New York, NY, USA (2008)
- [94] Shah, R., Kesan, J.P.: An empirical examination of open standards development. In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*. pp. 212–. *HICSS '08*, IEEE Computer Society, Washington, DC, USA (2008)
- [95] Shalom, E., Shahar, Y., Taieb-Maimon, M., Martins, S.B., Vaszar, L.T., Goldstein, M.K., Gutnik, L., Lunefeld, E.: Ability of expert physicians to structure clinical guidelines: reality versus perception. *Journal of evaluation in clinical practice* 15(6), 1043–1053 (Dec 2009)
- [96] Smith, A.M., Hardy, P.J., Sandler, D.A., Cooke, J.: Paramedic decision making: prehospital thrombolysis and beyond. *Emergency Medicine Journal* 28(8), 700–702 (2011)
- [97] Smith, B., Ceusters, W.: HL7 RIM: an incoherent standard. *Studies in health technology and informatics* 124, 133–138 (2006)
- [98] Sokolowski, R., Dudeck, J.: Xml and its impact on content and structure in electronic health care documents. *Proceedings of AMIA Annual Symposium* (1999)
- [99] Spronk, R.: The HL7 MIF - Model Interchange Format (2010), http://www.ringholm.de/docs/03060_en_HL7_MIF.htm, Ringholm Whitepaper
- [100] Stiell, I.G., Wells, G.A., Field, B., Spaite, D.W., Nesbitt, L.P., De Maio, V.J., Nichol, G., Cousineau, D., Blackburn, J., Munkley, D., Luinstra-Toohey, L., Campeau, T., Dagnone, E., Lyver, M.: Advanced cardiac life support in out-of-hospital cardiac arrest. *N Engl J Med* 351(7), 647–656 (Aug 2004)
- [101] Sundvall, E., Qamar, R., Nystrom, M., Forss, M., Petersson, H., Karlsson, D., Ahlfeldt, H., Rector, A.: Integration of tools for binding archetypes to SNOMED CT. *BMC Medical Informatics and Decision Making* 8(Suppl 1), S7+ (2008)
- [102] Terkelsen, C.J., Sorensen, J.T., Maeng, M., Jensen, L.O., Tilsted, H.H., Trautner, S., Vach, W., Johnsen, S.P., Thuesen, L., Lassen, J.F.: System delay and mortality among patients with STEMI treated with primary percutaneous coronary intervention. *JAMA* 304(7), 763–71 (2010)
- [103] Travis, P., Bennett, S., Haines, A., Pang, T., Bhutta, Z., Hyder, A.A., Pielemeier, N.R., Mills, A., Evans, T.: Overcoming health-systems constraints to achieve the Millennium Development Goals. *The Lancet* 364(9437), 900–906 (2004)
- [104] Tuomainen, M.P., Mykkanen, J.A., Luostarinen, H., Pöyhölä, A., Paakkanen, E.: Model-centric approaches for the development of health information systems. In: Kuhn, K.A.,

- Warren, J.R., Leong, T.Y. (eds.) *MedInfo. Studies in Health Technology and Informatics*, vol. 129, pp. 28–32. IOS Press (2007)
- [105] U.S. National Library of Medicine: 2011aa SNOMED CT source information (2011), <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/SNOMEDCT>
 - [106] Vest, J.R.: More than just a question of technology: Factors related to hospitals' adoption and implementation of health information exchange. *I. J. Medical Informatics* 79(12), 797–806 (2010)
 - [107] Walderhaug, S., Stav, E., Mikalsen, M.: Experiences from Model-Driven Development of Homecare Services: UML Profiles and Domain Models. In: Chaudron, M.R.V. (ed.) *MoD-ELS Workshops. Lecture Notes in Computer Science*, vol. 5421, pp. 199–212. Springer (2008)
 - [108] Wu, X., Hu, R., Zhang, Y., Dong, G., Rupprecht, C.E.: Reemerging rabies and lack of systemic surveillance in people's republic of China. *Emerg Infect Dis* 15(8), 1159–64 (2009)
 - [109] Zusman, E.E.: Form facilitates function: Innovations in architecture and design drive quality and efficiency in healthcare. *Neurosurgery* 66(6), N24 (2010)
 - [110] Zywiets, C.: Communication and interoperability for serial comparison in continuous health care—the new challenges. *Studies In Health Technology And Informatics* 108, 172–180 (2004)

Chapter 11

A decision support system for Alzheimer's disease diagnosis

Flávio L. Seixas, Bianca Zadrozny, Jerson Laks, Aura Conci, and Débora C. Muchaluat-Saade

Abstract Clinical decision support systems (CDSSs) have been designed to assist physicians with decision making tasks. This chapter discusses a CDSS for Alzheimer's disease diagnosis using as inference engine a Bayesian network and an influence diagram as the decision-making model. We design a Bayesian network with fixed structure and probabilities automatically obtained using a machine learning algorithm called Expectation-Maximization. As a training base, we use a patient dataset provided by the Consortium to Establish a Registry for Alzheimer's disease (CERAD). Each Bayesian network node is related to attributes from the patient database. The patient database is composed by several types of information using rating scales for assessing the various aspects of the Alzheimer's disease. We evaluate the Bayesian network using discrimination and probability performance measurements. The results are compared to other classifiers.

Flávio L. Seixas

UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: fseixas@ic.uff.br

Bianca Zadrozny

IBM Research Brazil, Av. Pasteur, 138, 22296-903, Rio de Janeiro, RJ, Brazil

Jerson Laks

Center for Alzheimer's Disease, Psychiatry Institute, Federal University of Rio de Janeiro, Av. Venceslau Brás, 71, 22290-140, Rio de Janeiro, RJ, Brazil, e-mail: jlaks@centroin.com.br

Aura Conci

UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: aconci@ic.uff.br

Débora C. Muchaluat-Saade

MidiaCom Lab/UFF-Telemedicine Group an Associated Laboratory of INCT-MACC, Computer Science Institute, Fluminense Federal University, Rua Passo da Pátria, 156, 24210-240, Niterói, RJ, Brazil, e-mail: debora@ic.uff.br

11.1 Introduction

There is a worldwide concern about quality of service in healthcare. In 2001, two reports published by Institute of Medicine (IOM) exposed the problems with patient safety occurred in healthcare systems in the United States and abroad. These stimulated a broad array of stakeholders to engage in patient safety and motivated organizations and hospitals to adopt safe practices, particularly in integrating the medical knowledge and advances into the clinical setting [30]. Clinical decision support systems (CDSSs) assist physicians in applying new information to patient care through the analysis of patient-specific clinical variables [43]. CDSS are computer systems designed to support high-level cognitive functions involving medical tasks [31]. According to Agency for Healthcare Research and Quality (AHRQ), CDSSs hold great potential to improve health care quality, increase efficiency and reduce health care costs [5]. AHRQ reports also evidences suggesting that CDSS are effective in preventing of medical errors. These systems may be most practical when coupled with computerized physician order entry (CPOE) and electronic health records (EHR). There are three classical medical tasks: (1) diagnosis, (2) prognosis and (3) therapy [49]. Additional contemporary tasks include (4) monitoring, (5) prevention, (6) epidemiology and (7) biomedical research.

Figure 11.1 depicts an overview of the CDSS architecture. The electronic health records (EHRs) are stored in a comprehensive, lifelong and patient-centric database called health data repository. The health data repository provides an integrated view of the complete health and healthcare history of each patient, and contains all sorts of clinical data items from the patient, e.g. the clinical data, physical and neuropsychological tests results, medical images, demographic data, and the therapy outcome measures. There are a number of International Institutions that have promoted a reference model for EHR as HL7 (www.hl7.org), EN13606 (www.en13606.org), ISO 20514 (www.iso.org) and openEHR (www.openehr.org). They mainly concern about semantic interoperation of patient health data among different healthcare institutions [3]. The knowledge base (KB) contains the clinical diagnosis criteria of the disease. All medical concepts and clinical conditions involving the diagnosis criteria are stored in the KB. The inference engine performs the reasoning process and computes the most probable diagnosis, given the observed dataset. The system tries to match the patient health records to medical concepts coded in the knowledge base. This matching process may be more practical when the EHR and KB refer to a common terminology code, e.g. SNOMED (Systematized Nomenclature of Medicine-Clinical Terms). The CDSS communication with the Physician is commonly facilitated by a graphical user interface (GUI) designed for the system. The CDSS communication interface can be used to display the most probable diagnosis, the most relevant unobserved data for confirming or not the diagnosis hypothesis, and the patient EHR.

In this chapter, we design a framework for CDSS used for supporting the Alzheimer's disease (AD) diagnosis. AD is a degenerative disease causing lesions in the brain, and is the leading cause of dementia [21]. According to the annual report published by the Alzheimer's Association (www.alz.org), about 5.4 million peo-

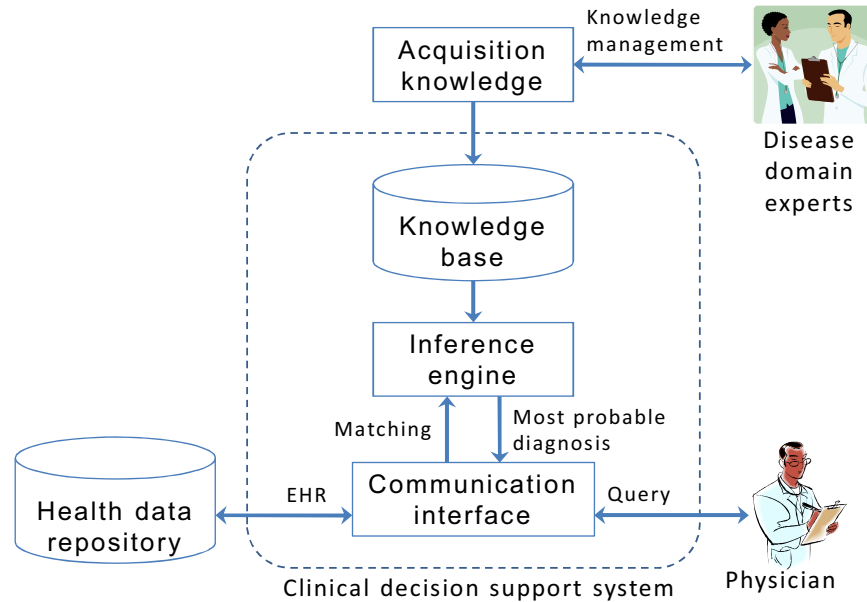


Fig. 11.1 An overview of the clinical decision support system architecture.

ple are affected by AD in the United States. There are nearly 15 million Alzheimer's and dementia caregivers providing 17 billion hours of unpaid care valued at 202 billion US dollars.

This chapter aims at proposing a CDSS for AD diagnosis. The inference engine will be based on Bayesian networks (BN). The decision-making will be modeled as an influence diagram (ID). The CDSS will assist the physicians displaying a message if the patient has or does not have AD with a correspondent degree of uncertainty and the most relevant information to be gathered from patient, in order to confirm or refute the initial diagnosis hypothesis. The Bayesian network structure was constructed manually based on broadly accepted diagnostic criteria and expert elicitation. The Bayesian network parameters were obtained automatically using a supervised machine learning algorithm. As training base we use the patient cases set provided by the Consortium to Establish a Registry for Alzheimer's disease (CERAD). The training base is composed of several types of information using rating scales for assessing the various aspects of the AD disorder, including cognitive impairment, activities of daily living, and functional and behavioral symptoms.

The chapter is organized in five more sections. Section 11.2 details the general concepts involving Bayesian networks and influence diagrams. Section 3 presents the knowledge representation for Bayesian networks. Section 4 proposes the Bayesian

network structure built and the parameter learning algorithm. Section 5 evaluates the learning performance. The conclusion is presented in Section 6.

11.2 Clinical decision support systems

There are three domain-specific systems for CDSS: (1) rule-based systems, (2) guideline-based systems, and (3) probabilistic models. Rule-based systems use rules (e.g., IF-THEN clauses) for making deductions and clinical decisions. Guideline-based systems determine the most likely clinical decision or pathway from a set of predetermined options. Probabilistic models are also referred as Bayesian networks (BNs), and they are represented in a graphical way with numerical information to represent a probability distribution over a set of random variables in a domain. Table 11.1 lists some tools by domain-specific system for CDSS.

Table 11.1 Clinical decision support systems applied to different disease domains.

Inference engine	Clinical decision support systems	Authors
Rule-based systems	CASNET (Causal-Associational Network)	Weiss <i>et al.</i> (1976) [53]
	MYCIN	Buchanan and Shortliffe (1984) [7]
	CPOL (Care Plan Online)	Beliakov and Warren (2001) [4]
Guideline-based systems	EON	Musen <i>et al.</i> (1996) [40]
	Asbru	Shahar <i>et al.</i> (1998) [51]
	GLIF (Guideline Interchange Format)	Peleg <i>et al.</i> (2000) [45]
	Prodigy	Johnson <i>et al.</i> (2000) [24]
	PROforma	Fox <i>et al.</i> (2006) [13]
Probabilistic-based systems	QMR/INTERNIST-1	Miller <i>et al.</i> (1986) [39]
	Mentor	Mani <i>et al.</i> (1997) [36]
	Diaval	Diez <i>et al.</i> (1997) [10]
	BayPAD	Luciani <i>et al.</i> (2007) [33]
	Promedas	Wemmenhove (2007) [54]
	PDS	Pyper <i>et al.</i> (2008) [47]
	TiMeDDx	Denekamp and Peleg (2010) [9]

The earliest CDSSs were rule-based, such as CASNET [53] applied to glaucoma and MYCIN [7] designed to identify a bacteria causing severe infections. The QMR/INTERNIST-1 [39] is the first CDSS using a BN as inference engine. That system includes over 570 diseases and 4400 symptoms. In MENTOR [36], the Bayesian model was built from a medical dataset regarding mental retardation. DI-AVAL [10] uses a BN as knowledge base for the diagnosis of heart diseases, and computes the a posteriori probabilities of each diagnosis. BayPAD [33] uses a BN for the diagnosis of pulmonary embolisms. Besides, the system applied case-based reasoning for the management of hospital resources. PROMEDAS [54] is based on Bayesian inference rules. It takes as input the patient's group of symptoms and test

results, and gives as an output the main diagnoses to take into account, with their corresponding certainty levels. PDS [47] proposes a BN to generate a list of possible diagnoses based on symptoms (diagnostic signs and test results) reported by the patient. Besides, the system has a graphical user interface (GUI) to patient cases entry. TiMeDDx [9] uses a BN to assist the physicians to identify the abstract diagnosis group and the appropriate data that should be collected during the clinical investigation process. That system is based on hypothetic-deductive clinical reasoning process of differential diagnosis. These cited systems show that multiple diagnoses are possible, and others may be included when new evidences are reported. In contrast to that, we focus on AD diagnosis. Furthermore, in this work, the (conditional) probabilities are obtained automatically using a Bayesian learning algorithm, which makes the BN modeling much simpler.

Systems applied to classify individuals with or without AD are usually referred to Computer-Aided Diagnosis (CAD) systems [16]. CAD usually uses as input data medical images, e.g. Magnetic Resonance Images (MRI), Computer Tomography (CT), Single-Photon Emission Computer Tomography (SPECT), and involves an image segmentation algorithm for extracting image features, automatically or semi-automatically. Table 2 shows some classifiers for AD.

Table 11.2 List of classifiers for Alzheimer's disease patients.

Classification method and selected features	Authors
Principal component analysis (PCA) and Artificial neural network (ANN) using MRI (volumetric analysis of hippocampus)	Yang <i>et al.</i> (2010) [57]
Support vector machine (SVM) and classification trees using SPECT (Single Photon Emission Computer Tomography)	Salas-Gonzalez <i>et al.</i> (2010) [50]
SVM algorithm and Bayesian computing to discriminate patients with Alzheimer and normal controls using MRI	Plant <i>et al.</i> (2010) [46]
SVM classification using whole-brain MRI	Magnin <i>et al.</i> (2009) [35]
Principal component analysis and decision trees classifiers using fMRI (functional Magnetic Resonance Images)	Tripoliti <i>et al.</i> (2011) [52]
Fuzzy neural network (FNN) using MRI	Anand <i>et al.</i> (2009) [2]

Regarding other diseases related to AD, Yan *et al.* [56] showed a BN for mild cognitive impairment (MCI) classification. MCI diagnosis is assigned to pre-clinical AD. Lindgren *et al.* [32] proposed a dementia management support system using clinical practice guidelines (CPG) and decision tables as the knowledge and inference model. Anand *et al.* [2] proposed a fuzzy neural network (FNN) using MRI for early AD diagnosis. Besides the classification model, in this work, we propose a CDSS that involves various subfields from artificial intelligence (AI), such as reasoning, knowledge representation, machine learning and classification.

11.3 Bayesian networks

BNs allow representing the uncertainty in a natural way. The uncertainty exists in almost every stage of a clinical decision making process. Sources of uncertainty may include (1) the inaccuracy by patients in reporting their symptoms, (2) inaccurate results reported by laboratories, (3) incomplete understanding of the disease etiologies [27]. Therefore, our KR schema is based on a BN.

The Bayesian network (BN) is a subtype of semantic knowledge representation that allows representing facts with uncertainty using concepts from the Probability Theory and the Bayes' theorem. BNs represent a domain in terms of random variables and explicitly model the interdependence of the random variables in terms of a directed acyclic graph (DAG) [22]. The DAG nodes represent the random variables and arcs represent direct dependencies.

Mathematically, suppose a set of variables X and their conditional dependencies. A BN is defined as a pair $B = (G, P)$, where $G = (V(G), A(G))$ is a DAG with a set of vertices (or nodes) $V(G) = \{X_1, \dots, X_n\}$ and arcs $A(G) \subseteq V(G) \times V(G)$, and a joint probability distribution P defined on the variables corresponding to the vertices, as shown in Equation 11.1.

$$P(X_{V(G)}) = \prod_{v \in V(G)} P(X_v | X_{\pi(v)}) \quad (11.1)$$

where $X_{\pi(v)}$ denotes the variables that correspond to the parents of node X_v . Matrix G denotes the BN structure and P denotes its parameters.

The inference algorithms for BNs are also called Bayesian inference. The Bayesian inference is responsible for computing the posterior probability distribution of the query node, given some evidence or observed nodes. The Bayesian inference can perform both reasoning from symptoms to cause (diagnostic support) and from new information about causes to new beliefs about effects (predictive support) [28]. There are exact algorithms and stochastic sampling algorithms for Bayesian inference depending on the number of nodes, complexity and other computational resource limitations. Examples of exact algorithms are the clustering or junction tree algorithm [34], and the poly-tree algorithm [44]. Examples of stochastic sampling algorithms are probabilistic logic sampling algorithm [18], the likelihood sampling algorithm [15] and the estimated posterior importance sampling algorithm [58]. Let suppose independent and identically distributed events $E \in e_n$. For each event $M \in m_m$, the prior $P(M)$ is updated to the posterior $P(M|E)$. From Bayes' theorem, we have Equation 11.2.

$$P(M|E) = \frac{P(E|M)}{\sum_m P(E|m_m) \cdot P(M_m)} \cdot P(M) \quad (11.2)$$

where $P(E|M) = \prod_k P(e_k|M)$.

The generalization of BNs to represent and solve decision problems under uncertainty are called influence diagrams [19]. Influence diagrams (IDs) offer several strengths for representing medical decisions [41]. IDs provide an explicit represen-

tation for both the actions and the value or utility of the resultant outcomes gives us decision networks. A utility function quantifies preferences, reflecting the usefulness of the outcomes by mapping them to a real number. The best decision will be the one that maximizes the expected utility [28]. The expected utility (\bar{U}) is given by Equation 11.3.

$$\bar{U}(A|E) = \sum_i P(O_i|E, A) \cdot U(O_i|A) \quad (11.3)$$

where E is the available evidence, A is a non-deterministic action with possible outcome states O_i , $U(O_i|A)$ is the utility of each outcome state, given that action A is taken, $P(O_i|E, A)$ is the conditional probability distribution over the possible outcome states, given that evidence E is observed and action A is taken.

An Influence diagram consists of three types of nodes: (1) chance nodes have an oval shape and represent random variables, exactly as in Bayesian networks. In our model, the chance node that represents the probability of the patient having a certain disease is also called query node. (2) The decision nodes have a rectangular shape and represent the decision being made at a certain point in time. In our model, the decision nodes refer to diagnostic decisions. (3) The utility nodes, also called value nodes, have a diamond shape and represent utility functions. The parents of a utility node are the variables describing the outcome state that directly affects the utility function. In our model, the parents of the utility node are the query node and decision node. Figure 11.2 depicts an example of influence diagram for clinical decision. Arrows represent conditioning influences leading into a chance node. Tables show the variables to which the probability assignment to the chance node variable will be conditioned. Figure 11.2 gives a generic example of an influence diagram. The decision is represented by the decision node *Takeaspirin*. The *Flu* node is a parent of the Fever (an effect). That symptom can be measured by a thermometer (*Therm*). If the aspirin is taken, it is likely to get rid of the fever. Fever later node represents the change over time. The adverse reaction to taking aspirin is represented by *Reaction*. The utility node U shows that the utilities depend on whether or not the fever is reduced, and whether the person has an adverse reaction.

11.4 Bayesian modeling

The BN modeling phase can be subdivided into two sequential tasks [28]: (1) define the variables and their causal relationship; (2) define the joint probability distribution over the variables. The variables and their causal relationships are represented by the BN structure. The BN parameters are represented by joint probability distribution. We model our BN according to the process depicted in Figure 11.3.

We performed a data preprocessing in order to combine and integrate the data (patient cases) from different data sources, as well as removing irrelevant data. The BN structure was designed in three levels as depicted in Figure 11.4. Level B (Background information) contains the BN variables related to patient demographic data,

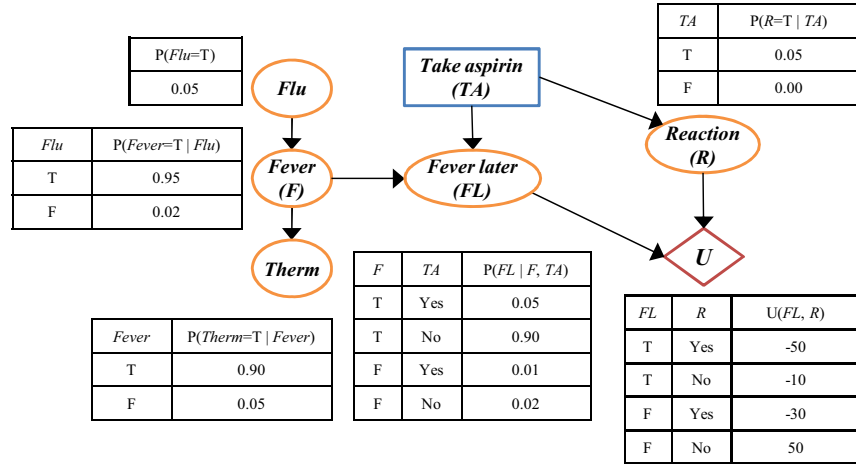


Fig. 11.2 Example of influence diagram used in decision-making if the patient should or not take aspirine.

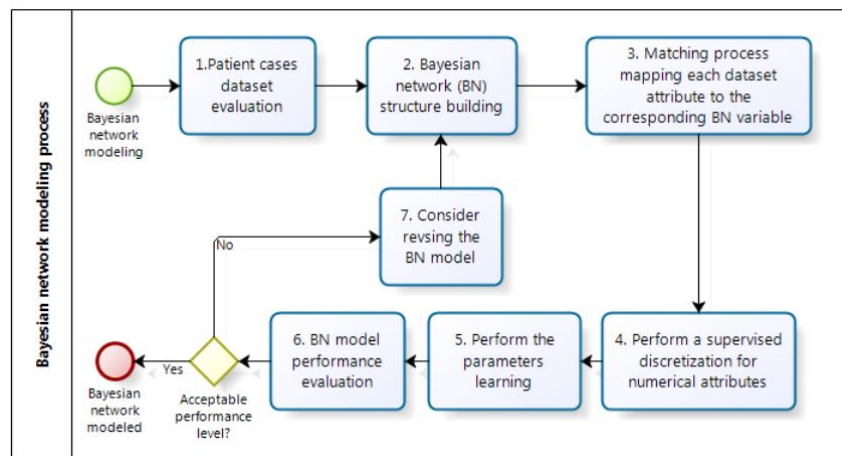


Fig. 11.3 Bayesian network modeling process.

as well as all the factors considered as predisposing and exclusion to AD. Level F (Findings) contains the symptoms, signs, test results, and other disease manifestations. Each variable from level B has a causal relation to the AD query node, and the AD node has a causal relation to each variable from level F . The BN structure was designed with the aid of a physician.

The criteria for the clinical diagnosis of probable AD were settled by the National Institute of Neurological and Communicative Disorders and Stroke and the AD and

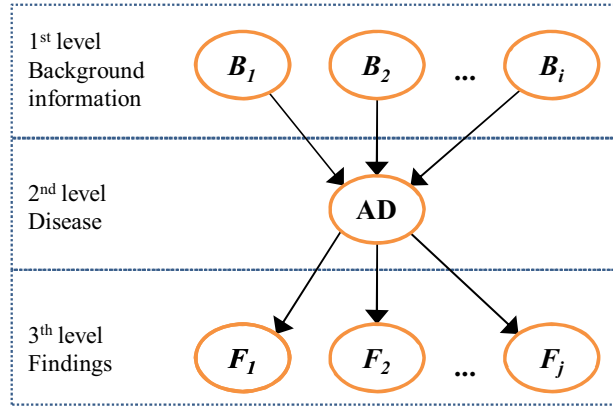


Fig. 11.4 BN structure for clinical decision support system.

Related Disorders Association (NINCDS-ADRDA) workgroup in 1984 [37]. These criteria were universally accepted by medical community. However, important advances in our understanding of AD have been occurring for more than a quarter of a century. In 2011, the National Institute on Aging (NIA) and the Alzheimer's Association presented a criteria revision, and organized them into three documents: the core clinical criteria of the recommendations regarding all-cause dementia [20], dementia due to AD [38], and MCI (mild cognitive impairment) due to AD [1]. According to criteria, the clinician usually applies dementia diagnosis criteria at a first moment, and then, the probable or possible AD diagnosis criteria.

We used the patient case dataset provided by CERAD. CERAD was funded by the National Institute on Aging (NIA) in 1986 with the objective of developing standardized procedures and validated measures for the assessment of AD. Those measures include clinical, neuropsychological, neuropathological and behavioral assessments of AD, and also assessment of family history [12]. We used 72 instances (patient cases), and 16 attributes. In the patient cases set, 15 patients had negative AD diagnosis, and 57 patients had positive diagnosis; all these diagnoses were confirmed by histopathological post-mortem examination. There are also missing values (fields with blank or invalid values) and they correspond to 40% of total value set from training base.

Table 3 lists the BN variables. The BN variables were identified based on features from AD diagnosis criteria. Each BN variable was classified whether belonging to the B level or F level. Then, each BN variable was semantically matched to attributes from patient cases set. There were BN variables that could not be matched to patient attributes, as shown in column *Matched*.

We have two data types: numeric types (e.g., tests score), and nominal or binary types. The data discretization aims at converting numeric data types to nominal. We performed a supervised discretization using the algorithm by Fayyad and Irani [11],

Table 11.3 Bayesian network variables related to Alzheimer's disease.

Level ⁽¹⁾	Description	Matched? ⁽²⁾
B	Diagnosis of dementia	Y
B	Cerebrovascular disease	Y
B	Presence of multiple or extensive infarcts	Y
B	Dementia of the Lewy body type	Y
B	Dementia of frontal lobe type	Y
B	Primary progressive aphasia	Y
B	Evidence for another concurrent, active neurological disease	Y
B	Dementia due to Huntington's disease	Y
B	Age	Y
B	Years of education	Y
B	Sex	Y
B	Race	Y
B	Socioeconomic status	N
F	Impaired ability to acquire and remember new information	Y
F	Difficulty using the elements of language	Y
F	Difficulty using visuospatial reasoning	Y
F	Executive dysfunction	Y
F	Sudden onset of cognitive impairment	N
F	Low CSF	N
F	Decreased fluorodeoxyglucose (FDG)	N
F	CSF tau	N
F	Positive FDG-PET	N
F	Hippocampal volume or medial temporal atrophy	Y

(1) B=Background information (predisposal or exclusion factors);
F=Findings (symptoms, signs, test results)

(2) If yes, this variable was matched with attributes of training cases set and used in the BN parameters learning process

and the implementation available within the WEKA data-mining tool by Waikato University (www.cs.waikato.ac.nz/ml/weka/).

The BN parameters were estimated using a Bayesian learning method called EM (Expectation-Maximization) algorithm [8]. We have missing values in the training base, and the EM algorithm allows learning using partial observations. EM algorithm is also appropriate to deal with BN composed by random variables having discrete states. EM algorithm is an iterative method for finding maximum likelihood estimates (MLE) of parameters θ . EM alternates between performing an expectation E step, which computes the expectation of the log-likelihood evaluated using the current estimate for the parameters, and maximization M step, which computes parameters maximizing the expected log-likelihood found on the E step. We used the EM implementation available at BNT (Bayesian Network Toolbox) for Mathworks Matlab (code.google.com/p/bnt/). The EM learning algorithm con-

verged after 10 iterations with 2 minutes of processing time approximately. Figure 11.5 depicts the BN structure.

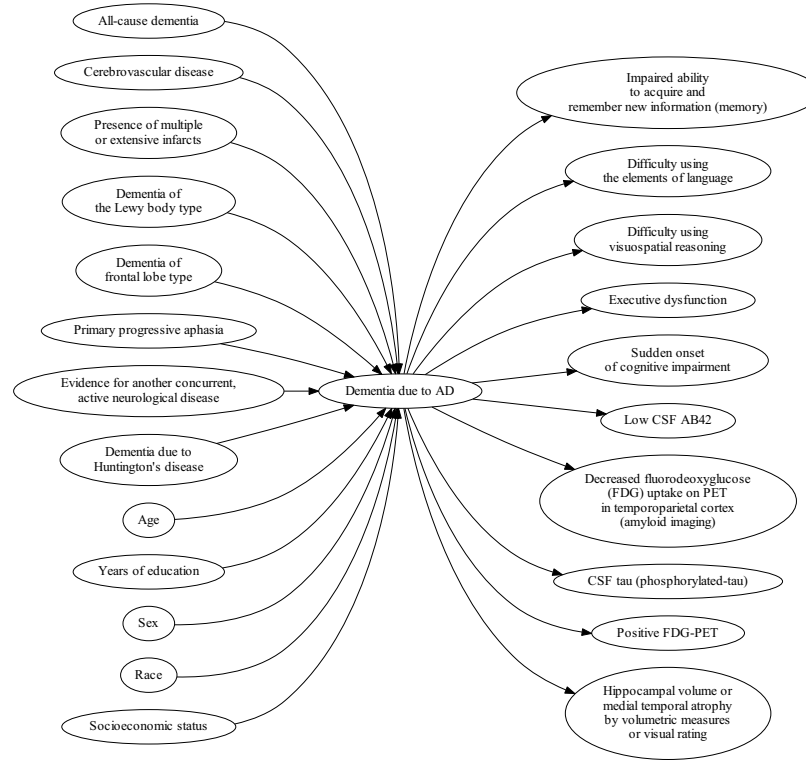


Fig. 11.5 Bayesian network structure for Alzheimer's disease diagnosis.

In Section 11.5, we evaluate the Bayesian network.

11.5 Learning performance evaluation

The performance of the BN model for predicting the probability of a patient having AD, given his (her) clinical observations, was evaluated using two measurements of discrimination and two probability measurements. Assume that the target variable is a binary type that represents the AD diagnosis. The discrimination measurement evaluates how well the BN model differentiates between two values of the target variable, i.e. negative (*neg*) and positive (*pos*) states. We used the area under the

Receiver Operating Characteristic curve (AUC) and harmonic mean of precision and recall (F_1score) as discrimination measures. The classification uses Bayesian classification rule as shown in Equation 11.4.

$$decideZ = positive \text{ if } P(z_{pos}|E) = \max_{i=neg,pos} P(z_i|E) \quad (11.4)$$

where Z is the target variable, E is the evidence set assigned to BN variables and $P(z_i|E)$ is the posterior marginal probability computed to each state i .

The probability measurements are minimized when the predicted value for the target of each case coincides with the true probability of that case taking that target value. We used the mean squared error (MSE) and mean cross-entropy (MXE) for evaluation. A brief description of those measurements is given in Table 11.4.

Table 11.4 Brief description of the performance measurements used for evaluation of performance of the prediction models.

Performance measure	Abbreviation	Range	Best score
Area under the ROC curve	AUC	[0, 1]	1
Harmonic mean of precision and recall	F_1	[0, 1]	1
Mean square error	MSE	[0, 1]	0
Mean cross-entropy	MXE	[0, ∞)	0

Assume TP stands for the number of true positive cases, TN the number of true negative cases, FP the number of false positive cases, FN the number of false negative cases, as represented in the form of the table called confusion matrix shown in Table 11.5. The Receiver Operating Characteristics (ROC) curve is a plot of sensitivity (Equation 11.5) versus $(1 - specificity)$ (Equation 11.6) for all possible thresholds on the probability of belonging to the positive class [17]. The F_1score (Equation 11.8) measures a test accuracy, and considers both that precision and recall to compute the score. It can be used as a single test measure [42].

Table 11.5 Confusion matrix model.

		Actual class	
		Positive	Negative
Predicted class	Positive	TP (true positive)	FP (false positive)
	Negative	FN (false negative)	TN (true negative)

$$sensitivity = recall = \frac{TP}{TP + FN} \quad (11.5)$$

$$specificity = \frac{TN}{TN + FP} \quad (11.6)$$

$$precision = \frac{TP}{TP + FP} \quad (11.7)$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (11.8)$$

Let a binary target variable Y take values in $[0, 1]$, and let y_i denote an indicator variable such that $y_i = 1$ if $Y_i = 1$ and $y_i = 0$ if $Y_i = 0$. Let p_i denote the predicted probability that Y_i for case i takes the value 1. The mean square error for a dataset of n cases is defined as shown in Equation 11.9 [6].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2 \quad (11.9)$$

Mean logarithmic loss or mean cross entropy is another probability measurement. The mean cross-entropy for a dataset of n cases is defined as shown in Equation 11.10 [55].

$$MXE = \frac{1}{n} \sum_{i=1}^n -y_i \cdot \log(p_i) - (1 - y_i) \cdot \log(1 - p_i) \quad (11.10)$$

Our BN model was compared to the following classifiers: *i.* Naïve Bayes [23], *ii.* Logistic regression model with a ridge estimator [29], *iii.* backpropagation multilayer perceptron network [55], *iv.* Adaboost M1 method [14], *v.* simple decision tables majority [25], and *vi.* C4.5 decision tree (J48) [48]. The performance of each classifier was estimated using the k -fold cross-validation [26]. In the k -fold cross-validation, the original sample is partitioned into k subsamples. Of the k subsamples, a single subsample is retained as the training data, and the remaining ones are used as validation data for testing the model. The elements of each subsample are selected such that the sample retains their original characteristics. We used 4-fold cross validation in order to obtain training and test sets that are large enough.

Table 11.6 Comparison of prediction models performance for Alzheimer's disease.

Model	$AUC^{(1\uparrow)}$	$F_1^{(2\uparrow)}$	$MSE^{(3\downarrow)}$	$MXE^{(4\downarrow)}$
Bayesian network ⁽⁵⁾	0.63	0.70	0.23	0.27
NaïveBayes	0.46	0.87	0.19	0.26
Logistic	0.39	0.82	0.23	0.22
Multilayer perceptron	0.42	0.81	0.28	0.38
AdaBoost M1	0.45	0.84	0.22	0.33
Decision table	0.35	0.81	0.20	0.26
J48	0.48	0.86	0.19	0.26

(1 \uparrow) AUC = Receiver Operating Characteristics (ROC) area. Best score is 1.

(2 \uparrow) F_1 = F-score. Best score is 1.

(3 \downarrow) MSE = Mean square error. Best score is 0.

(4 \downarrow) MXE = Mean cross entropy. Best score is 0.

(5) Prediction model proposed by this work.

We computed the average of performance values of each subsample (fold) in order to obtain a single value. We obtained a relative standard deviation lower than 6%, confirming that subsamples have similar characteristics. The prediction model based on BN showed relatively higher values for AUC and lower values for F_1 -score. The AUC and F_1 -score are both discrimination measures. The BN model has a graphic-oriented representation, whose causal relationships are explicitly represented. In contrast to other prediction models, the diagnosis criteria in BN can be represented in a human-oriented causal diagram, which facilitates the communication between domain experts and knowledge engineers. Furthermore, the addition of new medical findings or modifying existing ones in a BN model can be handled in an easy way.

Another prediction model evaluation is to analyze how sensitive the BN is to changes in parameters or inputs; this is called sensitivity analysis [28]. So, we performed the sensitivity to evidence analysis. The sensitivity to evidence analysis can be applied to identify, among other evidences, which one most decreased the uncertainty of the diagnosis, i.e., which most relevant evidences may be used to confirm or refute the initial diagnosis hypotheses. Entropy is the common measure for uncertainty. The entropy of a distribution over variable X is given by Equation 11.11.

$$H(X) = - \sum_{x \in X} P(x) \cdot \log[p(x)] \quad (11.11)$$

The entropy was computed on the target variable, i.e. the random variable that represents the probability of patient having AD diagnosis. Table 7 shows the first five most relevant evidences for AD, according to the BN model. Variables are ordered by delta entropy. Delta entropy was calculated by subtracting the entropy computed without evidences and the entropy computed with evidence: considering only the variables under analysis.

Table 11.7 Sensitivity analysis by evidence that shows the first five most relevant evidences for AD.

N	Variable ⁽¹⁾	Entropy $H(X)$	$\Delta Entropy$ ⁽²⁾
1	Difficulty using visuospatial reasoning	0.92	0.07
2	Difficulty using the elements of language	0.95	0.04
3	Hippocampal volume or medial temporal atrophy	0.97	0.03
4	Impaired ability to acquire and remember new information	0.97	0.03
5	Years of education	0.98	0.02

(1) Random variable from disease-specific prediction model

(2) $\Delta Entropy$ is obtained by subtracting the entropy computed with no evidences and entropy computed with evidence for variables under analysis.

According to Table 7, the difficulty using visuospatial reasoning performing is relevant to confirm (or refute) the AD diagnosis hypothesis. The hippocampal volumetric assessment from Magnetic Resonance Images (MRI) appeared in the third position, followed by impaired ability to acquire and remember new information, a symptom commonly associated to the memory.

11.6 Conclusion

The BN models work well with partial observations, handle uncertainties and have a human-oriented knowledge representation that facilitates the communication between knowledge engineers and domain experts. In this chapter, we proposed a BN for assisting physicians in AD diagnosis. The BN model was evaluated based on performance measurements and the results were compared to other well-known classifiers. The BN model obtained competitive performance results.

AD is a subtype of dementia. In the clinical reasoning process, a physician firstly applies the dementia screening criteria; and then, the probable or possible AD diagnosis criteria. The AD diagnosis decision must occur after the dementia suspicion. The MCI diagnosis has special interest in the scientific community since its treatment reduces the chance of the patient developing Alzheimer's. On going works are the development of BN-based prediction models for dementia and MCI.

As future works, we will extend the Bayesian model including a decision-making modeled as influence diagram. We will also extend the training base adding patient cases from other Institutions. Finally, we will design a knowledge representation model for diagnostic criteria representation, and a Graphical User Interface (GUI) for the CDSS, aiming at deploying it in a daily clinical routine.

Acknowledgements

We thank the Consortium to Establish a Registry for Alzheimer's disease (CERAD) for kindly providing the cases dataset used in this study. This work has been partially supported by CAPES, CNPq and FAPERJ (Brazilian agencies to promote scientific research) and is part of INCT-MACC (National Institute of Science and Technology for Medicine Assisted by Scientific Computing).

References

- [1] Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., Phelps, C.H.: The diagnosis of mild cognitive impairment due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic

- guidelines for alzheimer's disease. *Alzheimer's dementia : the journal of the Alzheimer's Association* **7**(3), 270–279 (2011)
- [2] Anand, S.M., Rao, M.M., Prabhu, N.S., D. Simeon, S., D, K., Rashmi, S.: Automated diagnosis of early alzheimers disease using fuzzy neural network (2009)
 - [3] Beale, T., Heard, S.: An ontology-based model of clinical information. *Studies in health technology and informatics* **129**(1), 760–764 (2007)
 - [4] Beliakov, G., Warren, J.: Fuzzy logic for decision support in chronic care. *Artificial Intelligence in Medicine* **21**(1), 209–213 (2001)
 - [5] Berner, E.S.: *Clinical decision support systems: theory and practice*. Springer, Birmingham (2007)
 - [6] Brier, G.: Verification of forecasts expressed in terms of probability. *Monthly weather review* **78**(1), 1–3 (1950)
 - [7] Buchanan, B., Shortliffe, E.: *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Reading, MA (1984)
 - [8] Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(1), 1–38 (1977)
 - [9] Denekamp, Y., Peleg, M.: Timeddx - a multi-phase anchor-based diagnostic decision-support model. *Journal of Biomedical Informatics* **43**, 111–124 (2010)
 - [10] Diez, F.J., Mira, J., Iturralde, E., Zubillaga, S.: Diaval: a bayesian expert system for echocardiography. *Artificial Intelligence in Medicine* **10**(1), 59–73 (1997)
 - [11] Fayyad, U., Irani, K.: Multi-interval discretisation of continuous-valued attributes for classification. In: *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pp. 1022–1027. Morgan Kaufmann, San Francisco CA (1993)
 - [12] Fillenbaum, G., van Belle, G., Morris, J., Mohs, R., Mirra, S., Davis, P., Tariot, P., Silverman, J., Clark, C., Welsh-Bohmer, K.: Consortium to establish a registry for alzheimers disease (cerad): The first twenty years. *Alzheimer's dementia: the journal of the Alzheimer's Association* **4**(2), 96–109 (2008)
 - [13] Fox, J., Patkar, V., Thomson, R.: Decision support for health care: the proforma evidence base. *Informatics in Primary Care* **14** (2006)
 - [14] Freund, Y., Schapire, R.: Experiments with a new boosting algorithm. pp. 148–156. *MORGAN KAUFMANN PUBLISHERS, INC.* (1996)
 - [15] Fung, R., Chang, K.: Weighing and integrating evidence for stochastic simulation in bayesian networks. p. 220. North-Holland Publishing Co. (1990)
 - [16] Giger, M.L., Chan, H.P., Boone, J.: History and status of cad and quantitative image analysis: the role of medical physics and aapm. *Med Phys* **35**(12), 5799–820 (2008). Using Smart Source Parsing Dec
 - [17] Hanley, J.: Characteristic (roc) curvel. *Radiology* **743**, 29–36 (1982)
 - [18] Henrion, M.: Propagating uncertainty by logic sampling in bayes' networks. pp. 317–324 (1988)
 - [19] Howard, R., Matheson, J.: Influence diagrams. *Decision Analysis* **2**(3), 127–143 (2005)
 - [20] Jack, C.R., Albert, M.S., Knopman, D.S., McKhann, G.M., Sperling, R.A., Carrillo, M.C., Thies, B., Phelps, C.H.: Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's dementia : the journal of the Alzheimer's Association* **7**(3), 257–262 (2011)
 - [21] Jalbert, J.J., Dalello, L.A., Lapane, K.L.: Dementia of the alzheimer type. *Epidemiologic reviews* **30**, 15–34 (2008)
 - [22] Jensen, F.V., Nielsen, T.D.: *Bayesian networks and decision graphs*. Springer Verlag (2007)
 - [23] John, G., Langley, P.: Estimating continuous distributions in bayesian classifiers. pp. 338–345. *Morgan Kaufmann Publishers Inc.* (1995)
 - [24] Johnson, P.D., Tu, S., Booth, N., Sugden, B., Purves, I.N.: Using scenarios in chronic disease management guidelines for primary care. *Proc AMIA Symp* pp. 389–393 (2000)
 - [25] Kohavi, R.: The power of decision tables. *Machine Learning: ECML-95* pp. 174–189 (1995)

- [26] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. pp. 1137–1145. Lawrence Erlbaum Associates Ltd (1995)
- [27] Kong, G., Xu, D., Yang, J.: Clinical decision support systems: a review on knowledge representation and inference under uncertainties. *International Journal of Computational Intelligence Systems* **1**(2), 159–167 (2008)
- [28] Korb, K., Nicholson, A.: Bayesian artificial intelligence. Chapman Hall/CRC, Clayton, Victoria, Australia (2004)
- [29] Le Cessie, S., Van Houwelingen, J.: Ridge estimators in logistic regression. *Applied statistics* pp. 191–201 (1992)
- [30] Leape, L.L., Berwick, D.M.: Five years after to err is human: what have we learned? *Jama* **293**(19), 2384–90 (2005). Using Smart Source Parsing May 18
- [31] Lindgren, H.: Decision support in dementia care: developing systems for interactive reasoning. Ph.D. thesis (2007)
- [32] Lindgren, H.: Dmss: a dementia management and support system for providing tailored advice in the dementia workup. Tech. rep., Umea University (2010)
- [33] Luciani, D., Cavuto, S., Antiga, L., Miniati, M., Monti, S., Pistolesi, M., Bertolini, G.: Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. *Emergency Medicine Journal* **24**(3), 157–164 (2007)
- [34] Madsen, A., Jensen, F.: Lazy propagation: a junction tree inference algorithm based on lazy evaluation. *Artificial intelligence* **113**(1-2), 203–245 (2004)
- [35] Magnin, B., Mesrob, L., Kinkinghuh, S., Plgrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehty, S., Benali, H.: Support vector machine-based classification of alzheimers disease from whole-brain anatomical mri. *Neuroradiology* **51**(2), 73–83 (2009)
- [36] Mani, S., McDermott, S., Valtorta, M.: Mentor: a bayesian model for prediction of mental retardation in newborns. *Res. Dev. Disabil.* **18**(5), 303–318 (1997)
- [37] McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group under the auspices of department of health and human services task force on alzheimer's disease. *Neurology* **34**(7), 939–944 (1984)
- [38] McKhann, G., Knopman, D., Chertkow, H., Hyman, B., Jack Jr, C., Kawas, C., Klunk, W., Koroshetz, W., Manly, J., et al.: The diagnosis of dementia due to alzheimer's disease: recommendations from the national institute on aging and the alzheimer's association workgroup. *Alzheimer's and Dementia* (2011)
- [39] Miller, R.A., McNeil, M.A., Challinor, S.M., Masarie F. E., J., Myers, J.D.: The internist-1/quick medical reference project—status report. *West J Med* **145**(6), 816–22 (1986). Using Smart Source Parsing Dec
- [40] Musen, M., Tu, S.W., Das, A.K., Shahar, Y.: Eon: a component-based approach to automation of protocol directed therapy. *J Am Med Inform Assoc* **3**(6), 367–388 (1996)
- [41] Nease, R., Owens, D.: Use of influence diagrams to structure medical decisions. *Medical Decision Making* **17**(3), 263–275 (1997)
- [42] Olson, D., Delen, D.: Advanced data mining techniques. Springer Verlag (2008)
- [43] Payne, T.H., Hoey, P.J., Nichol, P., Lovis, C.: Preparation and use of preconstructed orders, order sets, and order menus in a computerized provider order entry system. *J Am Med Inform Assoc* **10**(4), 322–9 (2003). Using Smart Source Parsing Jul-Aug; Epub 2003 Mar 28
- [44] Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
- [45] Peleg, M., Boxwala, A.A., Ogunyemi, O., Zeng, Q., Tu, S., Lacson, R., Bernstam, E., Ash, N., Mork, P., Ohno-Machado, L., Shortliffe, E.H., Greenes, R.A.: Glif3: the evolution of a guideline representation format. *Proc AMIA Symp* pp. 645–9 (2000). Using Smart Source Parsing 645-9
- [46] Plant, C., Teipel, S.J., Oswald, A., Bohm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M.: Automated detection of brain atrophy patterns based on mri for

- the prediction of alzheimer's disease. *Neuroimage* **50**(1), 162–74 (2010). Plant, Claudia Teipel, Stefan J Oswald, Annahita Bohm, Christian Meindl, Thomas Mourao-Miranda, Janaina Bokde, Arun W Hampel, Harald Ewers, Michael Research Support, Non-U.S. Gov't United States NeuroImage Neuroimage. 2010 Mar;50(1):162-74. Epub 2009 Dec 2.
- [47] Pyper, C., Frize, M., Lindgaard, G.: Bayesian-based diagnostic decision-support for pediatrics. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp. 4318–4321 (2008)
 - [48] Quinlan, J.: C4. 5: programs for machine learning. Morgan kaufmann (1993)
 - [49] Sackett, D.: Evidence-based medicine+. In: Seminars in perinatology, vol. 21, pp. 3–5. Elsevier (1997)
 - [50] Salas-Gonzalez, D., Griz, J.M., Ramirez, J., Lopez, M., Ivaréz, I., Segovia, F., Chaves, R., Puntónet, C.G.: Computer-aided diagnosis of alzheimer's disease using support vector machines and classification trees. *Physics in Medicine and Biology* **55**, 2807–2817 (2010)
 - [51] Shahar, Y., Miksch, S., Johnson, P.: The asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine* **14**(1-2), 29–51 (1998)
 - [52] Tripoliti, E.E., Fotiadis, D.I., Argyropoulou, M.: A supervised method to assist the diagnosis and monitor progression of alzheimer's disease using data from an fmri experiment. *Artif Intell Med* **53**(1), 35–45 (2011). Using Smart Source Parsing Sep; doi: 10.1016/j.artmed.2011.05.005. Epub 2011 Jun 23
 - [53] Weiss, S.M., Kulikowski, C.A., Safir, A.: A model-based consultation system for the long-term management of glaucoma. *Specialized Systems* **3**, 826–832 (1976)
 - [54] Wemmenhove, B., Mooji, J., Wiegerinck, W., Leisink, M., Happen, H., Neijt, J.: Inference in the promedas medical expert system. *Artificial Intelligence in Medicine* pp. 456–460 (2007)
 - [55] Witten, I., Frank, E., Hall, M.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2011)
 - [56] Yan, S., Shipin, L., Yiyuan, T.: Construction and application of bayesian network in early diagnosis of alzheimer disease's system. In: Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on, pp. 924–929 (2007)
 - [57] Yang, S.T., Lee, J.D., Huang, C.H., Wang, J.J., Hsu, W.C., Wai, Y.Y.: Computer-aided diagnosis of alzheimers disease using multiple features with artificial neural network, pricai 2010: Trends in artificial intelligence. pp. 699–705. Springer Berlin / Heidelberg (2010)
 - [58] Yuan, C., Druzdzel, M.: Importance sampling algorithms for bayesian networks: Principles and performance. *Mathematical and Computer Modelling* **43**(9), 1189–1207 (2006)

Chapter 12

Global Teleradiology services and education: Brazilian projects to improve service and education

Alexandra Maria Vieira Monteiro, Cláudio Márcio Amaral de Oliveira Lima and
Alair Augusto Sarmet dos Santos

Abstract Teleradiology was first described in 1972. Since then, the new technologies and the increase in velocity and in the infrastructure of the Internet has allowed the globalization of teleradiology. Teleradiology is already a reality; policymakers need to address its implications as it spreads. Teleradiology is the most economically viable of the telemedical fields, making radiologists in Brazil increasingly fearful of foreign competition. This chapter goals are to summarize the Brazilian actions focused on ethics harmonized with World Health Organization (WHO) and to introduce Brazilian projects to improve service and e-education.

12.1 Telemedicine and Teleradiology

The concept of Telemedicine, in its fullest essence, dates of the last century, when health professionals already used to share their experiences at distance using a simple telephone. Presently, with the advent of new technologies of communication and information (ICTs), this concept is being enlarged and now embraces online interaction among health professionals combined with the high quality transmission of

Alexandra Maria Vieira Monteiro

Associate Professor of Radiology at Medical School of Rio de Janeiro State University (UERJ). Coordinator of Teleradiology Committee of the Brazilian College of Radiology (CBR). Coordinator of UERJ TeleHealth Laboratory an Associated Laboratory of INCT-MACC.
e-mail: monteiroamv@gmail.com

Cláudio Márcio Amaral de Oliveira Lima

Radiologist Physician at Labs D'or Hospital.
e-mail: cmaolima@gmail.com

Alair Augusto Sarmet dos Santos

Associate Professor of Radiology at Medical School of Federal Fluminense University (UFF). Member of Teleradiology Committee of the Brazilian College of Radiology (CBR). Member of UFF-Telemedicine Group an Associated Laboratory of INCT-MACC.
e-mail: alairsarmet@globo.com

data, images, audio and video. The technology chosen depends largely on the frequency being used, the speed and the broadband resources available for use [15, 18].

Telemedicine is defined as the provision of services related to health care, in cases in which distance is a critical factor; such services are promoted by health professionals using information and communication technologies for the exchange of information for the diagnosis, prevention and treatment of diseases and also for the continuing education of health service providers, as well as for research and evaluation purposes [2, 11]. The basic difference between conventional medicine and medicine allied to the use of ICTs is that in Telemedicine what "travels" is information or patient data. This technology allows so that there is greater scope and better service to patients in remote areas and in cases of difficult diagnosis, as it is necessary specialized care in excellence centers, through second opinion sessions using teleconference systems [11], for example.

The use of technology in health was first described in 1959 by a group of psychiatrists from The College of Medicine of Nebraska University, who monitored a patient at distance [14, 15]. In 1960, NASA (U.S.A. National Aeronautics and Space Administration) adopted more complex systems of telecommunication for sending the first human being to the outer space using monitoring by telemetry [15]. In 1968, still in U.S., there was a video call between the Massachusetts General Hospital and the Boston's Logan Hospital for providing a medical consulting at distance [15]. At that same year, in December, the Dartmouth Medical College together with the Vermont University were the pioneers of INTERACT Program, which started a distance program of medical consulting and education through the network among 10 rural points in Vermont and New Hampshire. The impact was positive; however it had to be interrupted by deficiencies of resources. It was restarted by VTMEDNET, in 1995 [15]. Between 1972 and 1975, NASA also launched another teleintegration program, in Papago Indian Reservation, Arizona, this time using radio waves. Two paramedics in a van equipped with an electrocardiography and an X-ray machine were connected bidirectionally with the medical staff of the Public Health Service Hospital, with very promising outcomes.

Since 1989, NASA implemented the first international Telemedicine program, after a terrible earthquake in Armenia. The teleconsulting involved the transmission of audio, video and similar, and put together staffs from the U.S. and the former Soviet Union. In 1993, the Washington University was added to the project with the use of satellites for sending radiology images (Teleradiology) [15]. These good outcomes on the impact and also concerning coverage promoted, in 1994, the union of NASA, the University of Pittsburgh, the WHO (World Health Organization), the PAHO (Pan American Health Organization), USAID (U.S. for International Development) and finally the World Bank to form GHN - the Global Health Network - with the main goal of implementing a preventive Telemedicine. In 1995, the U.S. Army Research Center, the Trident Inc., and the Cedars Sinai Medical Center, in Los Angeles, joined the GHN [15].

The fast evolution of technology has caused a quick and progressive increase of Telemedicine centers which, in 1993, resulted in the creation of ATA (American Telemedicine Association) [1, 18]. Thus, a new concept of eHealth was then

brought out and defined by Gunther Eysenbach, editor in chief and publisher at Journal of Medical Internet Research: eHealth as a field in the intersection of medical informatics, public health and health professionals, providing health services and information delivered or enhanced through the Internet and other related technologies. In a practical way, it features not only the use of technology, but also a new concept about the thinking and acting of physicians and other health professionals. Into this new broader context, Telemedicine is concerned to the use of technologies for health attendance and exchange of information with remote areas [3, 41].

The impact of using health technologies in Brazil - a continental like country - leads to the improvement in health education and development, and also reduces the costs for population care [38], once there is no need of displacement of users, nor need of repeating images, added to it the fact the referral is made correctly, including to high complexity hospitals, when necessary. Otherwise, the costs would be immeasurable [5, 47, 48].

In 2002, the Brazilian Council of Telemedicine and Telehealth was set up in Brazil, adding to the concept of Telemedicine the use of technologies as a support tool for health careers as a whole [8].

Teleradiology was first described in 1972 [16]. Its commercial appeal was made available just in 1980s, when some suppliers used a photographic camera as a tool, with many technical limitations. The approach consisted of taking photos or recording them for later digitalization and transference. Other systems, more sophisticated, were used with the same purpose, e.g laser digitalizers; all those methods, however, presented serious limitations, because the pictures had to be handled one by one, making the process very slow. With the development of technology, those techniques were abandoned; the rapid evolution of last decade techniques changed Teleradiology drastically, in special in what concerns to the introduction of low-cost communication tools through Internet, what improved the cost-benefit analysis (cost versus performance) of computers in the large adoption of this system of communication and image archiving (picture archiving and communications system - PACS) and of RIS system (Radiology Information System).

In addition to archiving and image transmission, supported by PACS, the maintenance and availability of medical records data associated to the pictures is made in RIS. The use of these digital systems offers several advantages for radiologists, requestors and patients. In medium term, it also provides costs reduction and lesser use of films and chemicals, stimulating the decrease of test repetitions, for technical reasons or for allowing an easier access to previous examinations of a particular patient. These systems, when inter-connected through the Internet, may offer intrinsically the possibility of lighter exchange of examinations and results putting immediately available the reports among clinical centers and also providing the creation of telediagnosis centrals [29]. Moreover, all data and radiologic image archiving may be integrated to the electronic report of the patient in systems that permit interface, such as through HIS (Hospital Information System) [29].

Shy actions in Teleradiology are being realized in Brazil since the 90's, and they are being intensified last years. A so large country earns a lot with the rise and consolidation of collaborative integrated networks of education and distance medical

assistance. Benefits such a reduced displacement of professionals and patients, as well as the possibility to taking specialized Medicine to remote and impoverished localities of the country really makes an expressive difference. Yet the focus on medical ethics must be kept considering the doctor-patient relationship. For this reason, in January 19, 2009, the Brazilian Federal Council of Medicine (CFM) through a resolution (CFM n.1890) ruled and standardized the use of Teleradiology in Brazil [17]. Paradoxically, until now there are no protocols defined on a fundamental question related to this model: the safety needed for the transmission and electronic availability of digital images and reports on radiology. That is why civil societies are work-ing on the issue, cooperating to accelerate and legitimize this process.

Many studies [19, 22, 23, 24, 27, 28, 29, 35, 36, 37] were conducted to evaluate the pros and cons of this new way of looking at radiology. There is vast scientific evidence demonstrating clinical effectiveness and the nice cost-benefit analysis that contribute even more for its acceptability as a daily practice in the diagnostic imaging. Recent studies have proved that hospitals and clinics applying Teleradiology have got a capacity of performing tests 30% higher than those places which do not use Telemedicine. That is why at least 55% of hospitals and clinics in U.S. now adopt Teleradiology [23, 42]. Telemedicine is, at present, a worldwide reality; it is being employed, by the way, in the daily practice of some medical specialties, and also in collaborative virtual environments for medical education at distance [19, 22, 23, 24, 27, 28, 35, 36, 37]. The 21st century doctor must be, thus, updated, inserted in and able to handle with the ICTs' universe; he must also be aware of the new ethical questions which may arise as a result of this fast process of technology development [19, 22, 23, 24, 27, 28, 35, 36, 37]. This chapter aim is to present some Brazilian initiatives in Teleradiology services and education.

12.2 Brazilian College of Radiology and Diagnostic Imaging and the Brazilian Teleradiology Committee

Brazilian College of Radiology (CBR) was founded in September 15, 1948, in So Paulo, during the first Brazilian Conference of Radiology. The CBR is a permanent non-profit civil society of national scope, whose nature is essentially educational, scientific, cultural and social. CBR is affiliated with the Brazilian Medical Association (AMB) and represents the Department of Diagnostic Imaging and Radiotherapy into the Scientific Council. CBR's mission is spreading scientific knowledge, defending its members, stimulating the improvement of the professionals, uniting and orienting its affiliates and giving support to the principles of excellence in methods and procedures of imaging and therapeutic diagnosis [4, 6, 9].

CBR joins specialized physicians who lead activities of diagnosis and treatment through image methods and/or ionizing radiation, and it defends the interests of the doctors of the following areas: Radiology and Diagnostic Imaging, Nuclear Medicine and Radiotherapy. It also maintains periodic scientific publications and works together with the residency programs, applying the tests for the acquirement

of specialist degree and for the certification of the areas of activity. Moreover, it also offers refreshment courses in Radiology and promotes congresses, conferences, forums, and meetings about the several methods. Hence, in 64 years of activity, CBR has become a fundamental tool of refreshment and evaluation of the abilities of the doctor involved with Radiology and Diagnostic Imaging. As a whole, there are 25 local Radiology societies that work complementing CBR's performance; their work is centered in the doctors of every hospital. CBR is responsible for all Brazilian doctors performing the specialties mentioned before. The distribution of those professionals entitled by CBR is asymmetric, as shown in Figure 12.1, [4, 6, 9].



Fig. 12.1 Map of distribution of the physicians with a specialist degree in Radiology and Diagnostic Imaging granted by CBR (Source CBR, 2010).

Brazilian College of Radiology embraces around 15 thousands of doctors who deal with imaging (radiologists and similar) and it performs largely at Professional Defense Department, the entity that cares for medical fees, doctors' interests *versus* clients' interests, professional skills recognition, respect and dignity in the medical practice. CBR also implanted Brazilian Classifying Rank of Medical Procedures. In search of quality in service provider, CBR develops programs of qualification in mammography, ultrasound, X-ray computed tomography, and magnetic resonance imaging (MRI). Each program has got a quality control, made by a National Quality Committee, that makes evaluations and inspections before granting the quality seal for the specific area as well as the CBR quality certification - a valuable plus of quality accredited for patients and clients, like health insurance operators [4, 6, 9], for instance.

This random distribution of radiologists possibly follows the availability of new technologies, including the approach in Teleradiology. The Brazilian Roll of Health Institutions (CNES) is the basis for implementing the health information systems, and it has specially been acquiring high-tech equipment, with an asymmetric geographic distribution (Figure 12.2) [4, 6, 9].

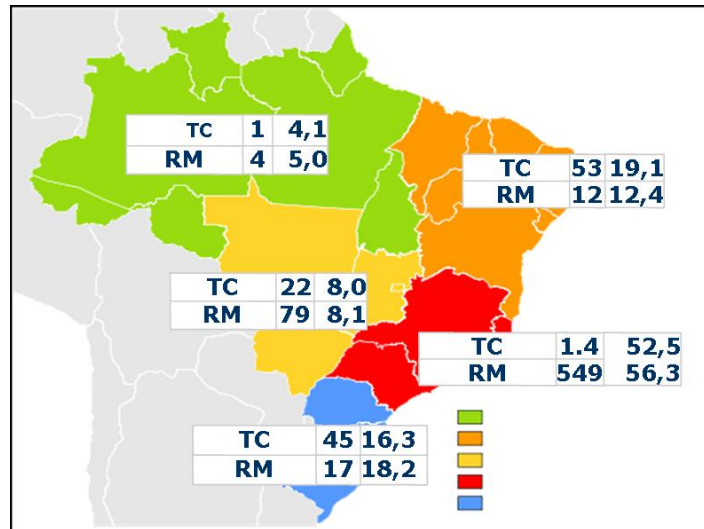


Fig. 12.2 High-resolution equipment distribution (Source CNES, 2011).

For the good practice of Teleradiology CBR created, in 2005, the Teleradiology Committee, whose goals are stimulating teleintegration under the perspective of good practice and ethics in Medicine; monitoring of public and private actions in Teleradiology; promoting the exchange with other representative entities with emphasis to the use of ICTs and development of teleeducation in Radiology. This Committee is acting in different national and international fronts and it took part actively in the regulation of the Brazilian Federal Council of Medicine (CFM) through a resolution (n. 1890/2009) that defined Teleradiology as a medical practice whose critical factor is the distance; the resolution also states that Teleradiology makes use of electronic transmission of radiologic images for appointment and report purposes [4, 6, 9].

12.3 Teleradiology Resolution

In January 19, 2009, the CFM resolution 1890 was published in Federal Official Gazette of Brazil (DOU), defining Teleradiology as a medical practice with a critical factor which is the distance and also pointing that Teleradiology makes use of

electronic transmission of radiologic images for appointment and report purposes. There is another previous resolution of CFM (n. 1643/2002) defining and ruling the rendering of services through Telemedicine. Summarizing, it detaches the incontestability of the doctor-patient relationship, as well as the obligatoriness of the free and clear consent signed by the patient or responsible and the ethical responsibility of the on-site doctor, recognizing the co-responsibility of the distance consulting doctor. Brazilian laws do not allow the medical appointment to patients through ICTs. The second-opinion among doctors (through teleconsulting) and the emission of reports, concerning specialties as Teleradiology, are legal practices.

Another relevant points in this resolution:

- Article 2 - Services rendered by Teleradiology must present the proper technological infrastructure and obey to CFM technical and ethical rules concerning guard, handling, data transmission, confidentiality, privacy and professional secrecy obligation guarantee.
- Article 3 - Examinations transmission must contain the relevant patient clinical data, collected by the requestor doctor, for reports drafting.
- Sole Paragraph. The patient must permit the electronic transmission of the images and data with free, clear and informed consent.
- Article 5 - In case of general non-contrast Radiology, excepting mammography and emergencies, whether there is no specialist in the city, the doctor responsible for the patient may ask to the specialist the support needed for distance diagnosis.
- Article 6 - In remaining cases - the presence of the specialist doctor at the local of its execution is obligatory.
- Article 7 - Professional responsibility support rests with the patient's physician assistant specialist, who performed the examination.

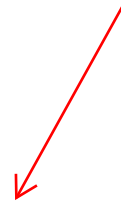
S1 - The physician specialist who issued the distance report is supportive in this responsibility.

S2 - The verification of eventual ethical infringement of those services is going to be made by the Regional Council of the jurisdiction where the procedure was performed.

- Article 8 - In the issue of the report shall bear the medical record number, in the respective Regional Councils of Medicine, of the doctors involved in attendance.

S2 - In case of physical persons, it must be a doctor holding a specialist title or certificate in his area, in accordance with article 4, duly registered in the Regional Council of Medicine of its jurisdiction.

Also, in April 13, 2010, came into force the *Brazilian New Code of Medical Ethics* [30] which in its sole paragraph states that *the distance care, along the lines of Telemedicine or any other method, is going to be placed under regulations of the Brazilian Federal Council of Medicine.*



Such resolutions define the current situation for the practice of Teleradiology into Brazilian territory. However, when one examines the international scenario under the perspective of globalization of Teleradiology, it should be considered that the geographical barriers may and must be broken, yet irrefutably ethics and resolutions must be above any financial interests in the use of technologies. With this vision, the Interamerican College of Radiology [7] has drawn up a document, with the Member States, with recommendations on good practice of Teleradiology. Similarly, the European Society of Radiology [10, 40] and the American Society of Radiology [39] are mobilizing for financial markets of large companies do not overlap the good practice of Telemedicine. In 2004, the *American College of Radiology* (ACR) published the technical standards for Teleradiology, which among other issues, determined that the radiologist who officially performs this service must regularly be licensed and still accredited by the hospital team, when this is the source of an exam. Some other standards were ignored in all of the United States, such as the one established by the *Federation of State Medical Boards*, which created a "model" for addressing the practice of Teleradiology, where a special permit was issued when such practice was frequent and regular, i.e., the activity of Teleradiology was greater than 1% of the daily practice of a physician or whether occurring at least once or more per week or still when more than 10 patients per year were involved. Currently, in the United States, the radiologists have got a permit for each State, in which they receive pictures and provide interpretations, as well as a permit for the State in which the interpretations physically occur. Similarly, they are accredited in each hospital for which they provide services, all in accordance with the standards of ACR [45]. Such procedures are also mandatory in Brazil after the settlement of the resolution for Teleradiology practice previously mentioned. Also, increasingly, systems are developed and improved for the synchronous and interactive communication among groups which are already facilitating the discussion of exams with colleagues or institutions of reference, fairly quickly, as well as the formation of collaborative groups and a new approach to medical education in Radiology. These major innovations point out a change in which radiology shall work very differently, in a flexible environment, at any time and in any location [10].

12.4 Low-cost Teleradiology: A Brazilian Project

The original concept of digital radiology was introduced in the late 1970's and early 1980's, but this technology has not been developed up to the first International Conference and Workshop Picture Archiving and Communication Systems (PACS), held in Newport Beach, California (USA) in January 1982. With the introduction of the digital image, a new era in the field of Radiology has begun: the use of digital files instead of conventional films which has brought new facilities for the handling, transmission, storage and visualization of X-rays [46].

In a service of Radiology and Diagnostic Imaging, the benefits of digitization range from improved efficiency, write-off of losses (with reduction or elimination

of film and chemicals) and increased savings products, ensuring the decentralization of information, allowing the sending of images to all locations in or outside the hospital, favoring the exchange of information and research with others services, ensuring thus a higher quality associated with the excellence of default images [46].

General radiography continues representing 70% of examinations in a radiology department, especially in intrahospital services. Because of the importance of this type of examination, its updating (formerly documented only in conventional X-ray films) to a digital system, in which images may be manipulated, edited, displayed, archived and and/or documented [46] became necessary.

Digital images in Radiology can be acquired in three ways. The first one is the use of conventional screen/film systems and the subsequent image scanning via a special scanner, digital cameras and electronic cards for capturing images. In this process there is no product economy and it is not applicable to a Radiology service with great movement. The second way is by acquisition of digital X-ray equipment (direct detection systems-DR) in which the image of the examination appears almost instantly in the workstation equipment; this system possess a wide exposure latitude, making it possible to obtain images with suitable density on a wide range of exposure levels, eliminating the problems concerning over and under exposition to the radiation dose. Disadvantages: the high cost of purchase and maintenance of DR equipment as well as the obsolescence of conventional X-ray equipment, already in service. The third way of acquisition of digital radiographic images is through systems using phosphor screens (CR = Computed Radiology). The benefits described are: these systems take advantage of the equipment already installed (the X-ray rooms and their equipment do not suffer any changes, keeping the hospital/service routine); they offer a wide exposure latitude, making it possible to obtain images with suitable density on a wide range of exposure levels, eliminating the problems of over and under exposition of the images, and finally, the acquisition cost is much lower than that of DR equipment (at least currently and in our environment) [46].

There are several types of scanners that use CR, from table linen equipment of slow performance up to multicassete equipment able to read more than 8-16 phosphor screens simultaneously, used in services with high number of radiographic examinations. These equipment may scan not only conventional and contrast radiology examinations, but mammography as well. In this case, equipment with higher definition and specific software for mammography [46] is needed.

A disadvantage of using phosphor screens is that, despite their acquisition cost is not very expensive, they have a short shelf life and after a certain number of exhibitions, the acquisition of new screens becomes necessary for using in CR equipment. Exchange time is variable according to the manufacturer, the number of examinations, and the careful use of the screens by the technicians [46].

In addition, it is necessary to recall that the human being is daily exposed to ionizing radiation of natural ($\sim 70\%$) and artificial ($\sim 30\%$) origin. Among artificial sources, it is the Medicine that contributes to human exposure, with about 85% of irradiance coming from diagnostic radiology procedures. Digital Radiology has huger value compared to the conventional technique, because it allows the effective reduction of radiation dose. This is also one of the reasons for the increasing

diffusion of Digital Radiology in all of the radiological services seeking excellence in quality. Finally, digital radiology is said to be “environmentally correct”, once it provides to eliminate the use of developers and fixers of conventional radiographic film processing machines, that are highly polluting and create waste. With digital radiology, we step to improve the quality of life of our surroundings and the whole planet [46].

The digital representation of X-rays offers numerous advantages for the Brazilian system of health, such as the proven efficiency of systems of second opinion via web and even the use of digital material in the process of education and training. To take full advantage of the digital representation of an X-ray, as well as Digital Radiography (DR) and computed radiography (CR), whose cost of implementation is considerably higher when comparing to the cost of the conventional method, it is possible to use the conventional radiographic film scanning with low-cost solutions. Currently, this can be made through two low-cost scanning radiographic film methods: one, using digital photography; other, a scanner for transparencies [19, 22, 23, 24, 27, 28, 35, 36, 37, 46].

The reality of Brazil as a country of continental dimensions and the limitation of resources make it rather difficult and costly the replacement of conventional radiography equipment by DR or CR equipment on a large scale. On the other hand, the small number of radiologists in geographically distant areas and the prevalence of lung diseases turn low-cost scanning solutions for general radiology an essential means to Teleradiology between remote points and centers of excellence, reducing geographical barriers, displacement of patients and costs in Brazilian Single Health System (SUS) [19, 22, 23, 24, 27, 28, 35, 36, 37, 46].

Routine low-cost images scanning using digital camera has been employed, but with limitations on technical features for image acquisition, such as: need for adequate lighting for capturing and adjustments in camera; use of tripod to avoid moving the machine during the image acquisition; and need for post-editing (cropping, rotate, grayscale etc.) of the captured images. New mobile devices technologies have also been adopted sporadically in the practice of the radiologist, although they must not be incorporated into routine services. In addition, there is the need for image compression for sending by a system that meets the existing resolutions. For this purpose and with this critical analysis, a multidisciplinary team of the Computing Core of High Performance of the Rio de Janeiro Federal University (NACAD-COPPE/UFRJ), together with radiologists from Rio de Janeiro State University (UERJ), Federal Fluminense University (UFF) and the team of Telehealth UERJ Laboratory has developed another low-cost solution that uses a commercial scanner adapted for partial scanning of radiographic film, also using a sewing software and image compression, and a teleconsulting system in Teleradiology [19, 20, 21, 22, 23, 24, 25, 27, 28, 35, 36, 37, 26, 46]. This project had got the support of the Financing of Studies and Projects (FINEP) of the Science and Technology Ministry.

The functioning of a common scanner or flatbed scanner is based on the capture, by a sensor, of the light reflected in the media placed in the tray into the scanner. This process is done continuously and the end result is a digital representation of the

scanned image. This process is virtually the same in all scanner models, however in transparency scanners the light is placed behind the image and not in front of it, as it happens with traditional scanners. For this reason, a transparency scanner is applied for radiographic film. The most common transparency scanners on the market are A4-size (21 x 29.7 cm, small in area, however, to the majority of scanning radiographic films used in medical radiology). In a scanner A3-size (29.7 x 42 cm) it is possible to scan almost completely all films evaluated, but the high price of the equipment (up to six times more than the A4) limits its use within the criteria for low-cost. In the solution developed, after they have been scanned with a scanner A4-size, a free image stitching software is employed. This software, called *scanRX*, sews the partially scanned images automatically to form a complete picture with invisible seam (Figure 12.3 to Figure 12.6) [19, 20, 21, 22, 23, 24, 25, 27, 28, 35, 36, 37, 26].

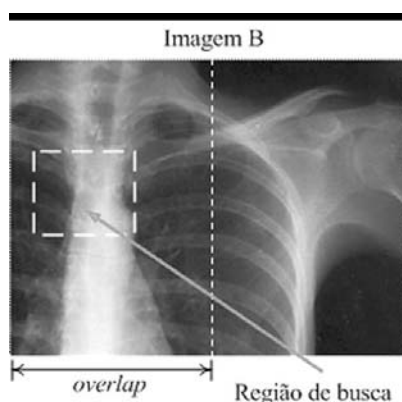


Fig. 12.3 Showing the overlapping A4 size areas to be superposed.

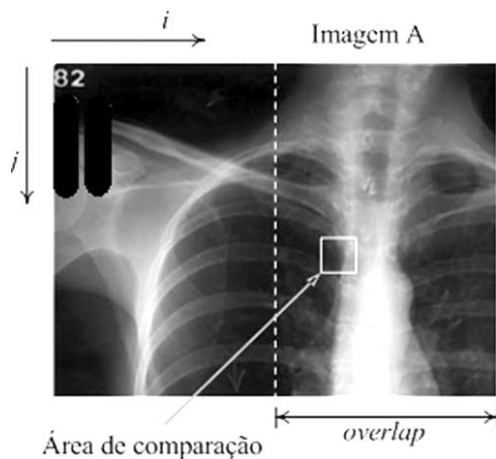


Fig. 12.4 The comparison area chosen in the left partial image.

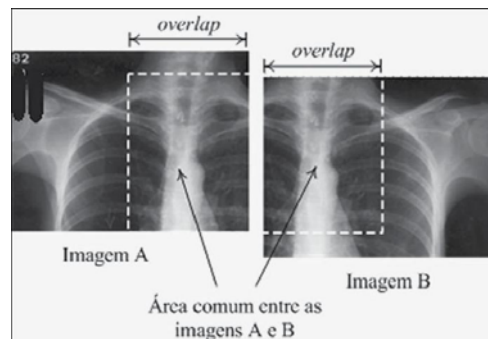


Fig. 12.5 The area which is searched in the right partial image to achieve a match with the left image comparison area.



Fig. 12.6 Final stitched image.

The use of *scanRX* software does not require any specific knowledge (of Radiology or photography), ensures reproducibility, excellent quality of the digital image, without problems of heterogeneity of focus, brightness, or picture framing, that usually occur with images made by a digital camera. Surely, the scanning process is slower than that of digital photography, however the cited advantages justify widely the longer time for acquiring each image. Tests show that the software can also be used to generate high-quality digital images (high resolution) appropriate for mammograms, as well as printed films of CT scans and magnetic resonance imaging (MRI). This solution specifically, when compared with scanning using a camera and having as a gold standard a radiology equipment, has shown a larger sensibility CR (83.33%-95% CI 70.87 -95.79) in relation to the digital cam (80.95% IC 95% 67.89 -94.02) and the specificity was compared between scan RX-93.7% (95% CI -

96.99 90.4) and digital cam-93.86% (95% CI -96.67 89.89). In comparison with the gold standard, the specialists analyzed the agreement by kappa coefficient, which was better for CR method (kappa = 0.76), followed by scanRx method (kappa = 0.66) and at last by digital cam (kappa = 0.57) [20, 21, 25, 26].

This approach is already being used in basic health units and emergency units (UPAs) in Rio de Janeiro city [21, 23, 35]. In UPAs, for example, by the lack of radiologists physicians on duty in some units, around 734 teleconsultings have already been held in a six-month period, resulting in detection and forwarding of 13% of these patients to specific reference centers [12, 13, 21, 23, 35] with diagnostic criteria for tuberculosis. The system used for sending pictures and also for teleconsulting is available at <http://www.telessaude.uerj.br/riotelrx/>.

12.5 Globalizing medical education in Radiology

New practices in Telemedicine and Teleradiology are progressively changing paradigms in continuing education, in upgrading and professional recertification [23, 27, 35, 37].

Several initiatives are being taken by Brazilian Medical Association (ABM) and by the Brazilian College of Radiology (CBR), specifically with respect to medical radiology [6, 17, 23] mainly through distance learning courses. But Brazilian limitations on connectivity (Figure 12.7), despite Broadband National Program [32] be in deployment, still restrict large advances in globalization of knowledge and distance services.

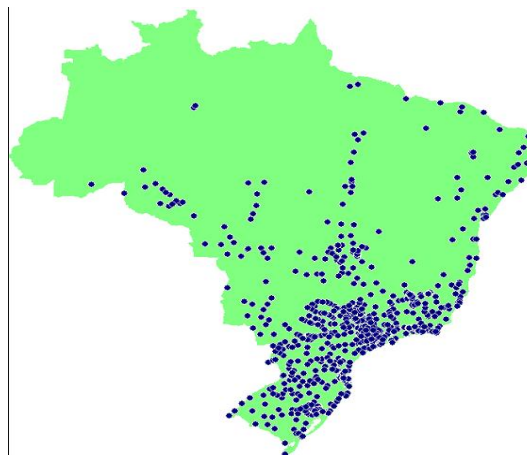


Fig. 12.7 Broadband distribution in Brazil (Source: Brazilian Communication Ministry, 2006).

Given the continental dimension of Brazil, the use of information and communication technologies in Medicine has had a growing importance. Recently, the Brazilian Government has created two nationwide projects entitled, respectively, Telemedicine University Network (RUTE) [33], by the Science and Technology Ministry and also the Telehealth Brazil Networks Program [43], by the Health Ministry. Their goal is to create a health care network joining the University hospitals with other national and international centers of excellence and with basic health units in remote locations, for teleeducation, teleconsulting and telediagnosics. In these health care networks, different working groups were created and among them some of Radiology: Radiology and Diagnostic Imaging of Chest Radiology, Diagnostic Imaging of the Abdomen, Neuroradiology, Radiology and Diagnosis in Pediatrics (RADPED), this one with greatest breadth (Figure 12.8) and residency in Radiology [34].



Fig. 12.8 RADPED network reach map.

In this network of globalization of radiology education, different methodologies are applied, as update-themed teaching teleconferences, interactive virtual seminars, discussion of clinical cases and discussion of difficult diagnosis cases. The meetings are held monthly, with remote access by the target audience, which is composed of undergraduate and postgraduate Medicine students, general practitioners and radiologists physicians. Some working groups, as the RADPED group, make use of two technologies for teleconferencing, which expand the scope of the activities: video conferencing and web conferencing. In addition to the infrastructure of government projects, formerly cited, Radiology networks also have the help of Telehealth UERJ lab's infrastructure to support, recording and provision of educational materials produced.

12.6 Telehealth UERJ's Lab

Telehealth UERJ Lab, of Rio de Janeiro State University [31] acts on different projects of implementation, development and sustainability in Telemedicine and Telehealth. It is a laboratory attached to the Brazilian National Institute of Science

and Technology - Computer-Assisted Medicine by Science Computing Technology (INCT-MACC), and it has, up to this point, national and international partnerships. Recently, by the way, Telehealth UERJ Lab was responsible for launching the Global eHealth Ambassadors Program [44].

In distance education via Moodle platform (Martins Dougiamas, Perth, Australia), that allocates all the recorded educational materials, there are about 19,170 health professionals registered, of whom 0.6% are foreigners. The map of spanning is illustrated below (Figure 12.9).



Fig. 12.9 Telehealth UERJ Lab spanning map.

Sixty-seven per cent of these professionals are active on the basic health network, and the remainder in other levels, in assistance. The distribution of registered professionals, by professions, is shown below (Figure 12.10).

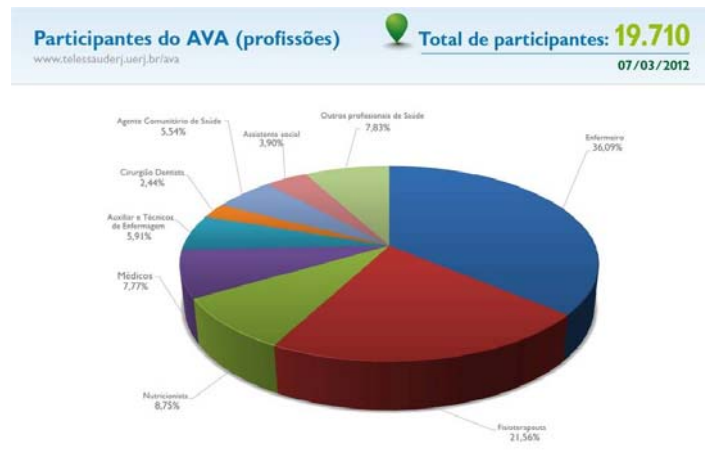


Fig. 12.10 Distribution map, by profession, of professionals registered in distance education Platform of Telehealth UERJ's Lab.

Since its inclusion in INCT-MACC, Telehealth UERJ's team intends contribute for the evaluating tools that can support the research on Computer-Assisted Medicine, particularly in what regards to Radiology.

12.7 Conclusion

Radiology appropriates increasingly of the internet and Telemedicine to its improvement concerning tele-education and services. The expansion of infrastructure and access to the Internet is finally, and in fact, going to make possible the globalization of education and services in Teleradiology.

Acknowledgements

This work was partially supported by the Ministry of Health and Brazilian agencies CNPq, CAPES, FAPERJ, FINEP and RNP. The support of these agencies is gratefully acknowledged.

References

- [1] American Telemedicine Association, February 2012. <http://www.atmeda.org/>.
- [2] Ângelo M. F., Schiabel H.: Uma ferramenta para treinamento na avaliação de imagens mamográficas via internet. *Radiol Bras*, **35**(5):259 – 265, 2002.
- [3] Belda T. E.: Computers in patient education and monitoring. *Respiratory Care*, **49**(5):480 – 488, 2004.
- [4] Cadastro Nacional de Estabelecimentos de Saúde, January 2012. http://cnes.datasus.gov.br/Mod_Ind_Equipamento.asp?VEstado=00.
- [5] Chao L. W., Cestari T. F., Bakos L.: Evaluation of an internet-based teledermatology system. *Journal of Telemedicine and Telecare*, **9**(Suppl. 1):S1:9 – 12, 2003.
- [6] Colégio Brasileiro de Radiologia, January 2012. <http://www.cbr.org.br/>.
- [7] Colégio Interamericano de Radiologia, February 2012. <http://www.webcir.org/etica.php#>.
- [8] Conselho Brasileiro de Telemedicina e Telessaúde, January 2012. <http://www.cbtmps.com.br/>.
- [9] Conselho Federal de Medicina. Resolução de telerradiologia, February 2012. http://www.portalmédico.org.br/resolucoes/CFM/2009/1890_2009.htm.
- [10] Diagnostic Imagin. Teleradiology promises real benefits for radiologists, March 2012. <http://www.diagnosticimaging.com/display/article/113619/1500153>.
- [11] Faculdade de Ciências Médicas-Telemedicina. Universidade do Estado do Rio de Janeiro, March 2012. <http://www.lampada.uerj.br/Telemedicina/Telemedicina.htm>.

- [12] Ferreira L. V., Grande J., Bhaya A., Monteiro A. M. V.: SCANRX: software para digitalização de radiografias a baixo custo para Telemedicina. In *IV Congresso Brasileiro de Telemedicina e Telessaúde*, 2009.
- [13] Ferreira L. V., Grande J., Bhaya A., Monteiro A. M. V.: TIPIRX: Teleintegração por imagens de raios-x. In *IV Congresso Brasileiro de Telemedicina e Telessaúde*, 2009.
- [14] Katsuyuki M., Suzuki Y., Sakai H.: Interactive communication in high-technology home care: Videophones for pediatric ventilatory care. *Pediatrics*, **99**(1):1 – 6, 1997.
- [15] Kim Y. S.: Telemedicine in U.S.A. with focus in clinical applications and issues. *Yonsei Medical Journal*, **45**(5):761 – 775, 2004.
- [16] Lee S. K., Peng C. H., Wen C. H.: Consulting with radiologists outside the hospital by using java. *Radiographics*, **19**:1069 – 1075, 1999.
- [17] Lima C. M. A. O., Monteiro A. M. V., Ribeiro E. B.: Videoconferências: Sistematização e experiências em telemedicina. *Radiol. Bras.*, **40**(5):341 – 344, 2007.
- [18] Mizushima H., Uchiyama E., Nagata H., et al.: Telemedicine comes of age. *Jpn. J. Clin. Oncol.*, **30**(1):3 – 6, 2000.
- [19] Monteiro A. M. V.: A Teleradiologia sem o PACS. O projeto inter-institucional entre Centro de Ensino Nacionais em Radiologia Pediátrica. In *10th Congress of the International Society for Telemedicine and e-health and 2nd Congress of Brazilian Council of Telemedicine and tele-health*, page 15, 2005.
- [20] Monteiro A. M. V., Bahia P., Santos A. A. S. M. D., et al.: Relato da experiência do uso de tecnologias de baixo-custo para a digitalização de imagens radiológicas. In *V Congresso Brasileiro e Internacional de Telemedicina e Telessaúde 2011*, 2011.
- [21] Monteiro A. M. V., Bhaya A., Pacheco L., Corrêa D. G.: Teleintegração por imagens de RX. *Revista Hospital Universitário Pedro Ernesto*, (Supl.):192 – 192, 2008. Also in 46th Congresso Científico do HUPE.
- [22] Monteiro A. M. V., Corrêa D. G., Santos A. A. S. M. D.: Programa de educação a distância em radiologia pediátrica. In *Congre's France Amérique Latine de Radiologie. 39^a Jornada Paulista de Radiologia*, volume Panéis e Temas Livres, page 36, 2009.
- [23] Monteiro A. M. V., Corrêa D. G., Santos A. A. S. M. D.: Telemedicine and pediatric radiology: A new environment for training, learning, and interactive discussions. *Telemedicine Journal and e-Health*, **17**:753 – 756, 2011.
- [24] Monteiro A. M. V., João Junior M., Santos A. A. S. M. D.: Prospective evaluation of a teleducation program by videoconferencing to enhance radiology resident education. a medical school of state university of rio de janeiro, teleteaching initiative. In *22nd ICDE World Conference on Distance Education. Promoting Quality in On-line Flexible and Distance Education ABED 2006*, 2006.
- [25] Monteiro A. M. V., Santos A. A. S. M. D., Bahia P., et al.: Project TIPRIX (integration of teleradiology using low-cost digitalization of x-ray films): a brazilian solution. In *European Congress of Radiology ECR 2011*, 2011.
- [26] Monteiro A. M. V., Santos A. A. S. M. D., Bahia P., et al.: Low-cost teleradiology: A brazilian solution for emergency care units. In *RSNA 2011*, pages 38 – 38, 2011.
- [27] Monteiro A. M. V., Santos A. A. S. M. D., Cavalcanti S.: The brazilian telemedicine pediatric radiology network. In *European Congress of Radiology ESR 2010*, 2010.
- [28] Monteiro A. M. V., Taborda M., Dias C.: Rede RUTE Núcleo UERJ: utilização de ambientes virtuais de aprendizagem como apoio a reutilização das teleconferências e a pesquisa colaborativa. In *IV Congresso Brasileiro de Telemedicina e Telessaúde*, 2009.
- [29] Nobre L. F., von Wangenheim A., Maia R. S., Ferreira L., Marchiori E.: Certificação digital de exames em telerradiologia: um alerta necessário. *Radiol. Bras.*, **40**(6):415 – 421, 2007.
- [30] Novo Código de Ética Médica. Conselho Federal de Medicina, February 2012. <http://www.portalmedico.org.br/novocodigo/index.asp>.
- [31] Portal de Telessaúde da UERJ, February 2012. <http://www.telessaude.uerj.br/site/>.
- [32] Programa Nacional de Banda Larga (PNBL) Brasil conectado, February 2012. <http://www4.planalto.gov.br/brasilconectado/pnbl>.

- [33] Rede Universitária de Telemedicina (RUTE), February 2012. <http://rute.rnp.br/>.
- [34] Rede Universitária de Telemedicina (RUTE) Grupos de Interesse Especial, February 2012. http://rute.rnp.br/sigs?p_p_id=buscasig&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column1&p_p_col_pos=2&p_p_col_count=3&_buscasig_text=&_buscasig_instituicaoId=1&_buscasig_especialidadeId=&_buscasig_struts_action=%2Fnp%2Fsig%2Fbusca%2Fview&page=3).
- [35] Santos A. A. S. M. D., Monteiro A. M. V., Messina L. A., Nobre L. F., Dippolito G.: The use of teleradiology to promote education and training of physicians and residents in radiology in brazil. In *European Congress of Radiology ESR 2010*, 2010.
- [36] Santos A. A. S. M. D., Monteiro A. M. V., Ziviani A.: Participação da radiologia no projeto MACC-RIO. In *Congr's France Amérique Latine de Radiologie. 39ª Jornada Paulista de Radiologia*, volume Panéis e Temas Livres, page 74, 2009.
- [37] Santos A. A. S. M. D., Monteiro A. M. V., Ziviani A., Feijóo R. A.: MACC-RIO Project: A network program of research and technical scientific cooperation between institutions in Brazil. In *European Congress of Radiology ESR 2010*, 2010.
- [38] Silva A. M. M., Furlan A. M., Haygert C. J.: Projeto de validação de um sistema de teleradiologia de baixo custo. 2012. <http://www.sbis.org.br/cbis/arquivos/1035.pdf>.
- [39] Sociedade Americana de Radiologia, March 2012. http://www.myesr.org/cms/website.php?id=/en/ESR_ECR_news.htm<http://www.rsna.org/>.
- [40] Sociedade Européia de Radiologia, March 2012. http://www.myesr.org/cms/website.php?id=/en/ESR_ECR_news.htm.
- [41] Spooner S. A., Gotlieb E. M., and The Steering Committee on Clinical Telemedicine: Telemedicine: Pediatric applications. *Pediatrics*, **113**:639 – 643, 2004.
- [42] Telerradiologia, January 2012. <http://pt.wikipedia.org/wiki/Telerradiologia>.
- [43] Telessaúde Brasil, February 2012. <http://www.telessaudebrasil.org.br/>.
- [44] The International Society for Telemedicine and eHealth (ISfTeH). Launch of the Global eHealth Ambassador's Program, February 2012. http://www.isfteh.org/ambassadors/launch_of_the_global_ehealth_ambassadors_program.
- [45] Thrall J. H.: Teleradiology part I. History and clinical applications. *Radiology*, **243**:613 – 617, 2007.
- [46] Vela J. G., Bhaya A., Monteiro A. M. V.: Digitalização de filmes radiográficos com costura de imagens. *Radiol. Bras.*, **44**(4):233 – 237, 2011.
- [47] Wechsler R., Anç ao M. S., Campos C. J. R., Sigulem D.: Computing in medical practices. *J. Pediatr.*, **79**(Suppl. 1):S03 – S12, 2003.
- [48] Young T. L., Ireson C.: Effectiveness of school-based telehealth care in urban and rural elementary schools. *Pediatrics*, **112**:1088 – 1094, 2003.

Chapter 13

Distributed Cyberenvironments Supporting Computing Assisted Medical Applications

Bruno Schulze, Giacomo V. McEvoy, Antonio R. Mury, Ramon Costa, Felipe J. Fernandes, Mariza Ferro, Fabio Licht, Matheus Bandini, Henrique Kloh, Daniel Yohoyama, Luis de Bona, Marcos Castilho, Marcos Sunye, Eduardo Todt, Daniel Weingaertner, Jose Neuman de Souza, Rossana Andrade, Helder Borges, Maria Cristina S. Boeres, Vinod Rebello, Fabio Porto

Abstract In this chapter we introduce the distributed cyberenvironments enabling the execution of scientific applications in general, and in the context of Medical Applications Assisted by Scientific Computing (MAACC) in particular. These cyberenvironments provide infrastructure and middleware support to execute high performance, resource demanding tasks related to medical applications, such as model solving and scientific visualization. Here we introduce the concept of virtual

Bruno Schulze, Giacomo Mc Evoy, Antonio R. Mury, Ramon Costa, Felipe J. Fernandes, Mariza Ferro, Fabio Licht, Matheus Bandini, Henrique Kloh, Daniel Yohoyama, Jonathan Goncalves, Vinicius Alvarenga

ComCiDis - Distributed Scientific Computing Laboratory, LNCC/MCTI and INCT-MAACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Petrópolis, RJ (<http://comcidis.lncc.br>), e-mail: [schulze,giacomo,aroberto,ramon,felipejf,mariza,licht,mbandini,henrique,yokoyama,jonathan, valvarenga]@lncc.br

Luis de Bona, Marcos Castilho, Marcos Sunye, Eduardo Todt, D.Weingaertner

C3SL - Center for Scientific Computing and Open Software and INCT-MAACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Federal University of Paraná (UFPR), Curitiba, PR (<http://www.c3sl.ufpr.br>), e-mail: [bona,marcos,sunye,todtdanielw]@inf.ufpr.br

Jose Neuman de Souza, Rossana Andrade, Helder Borges

GREAt - Group of Computer Networks, Software Engineering and System and INCT-MAACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Federal University of Ceará (UFC), Fortaleza, CE (<http://www.great.ufc.br/index.php?lang=en>), e-mail: [neuamn,rossana,helder]@ufc.br

Maria Cristina S. Boeres, Vinod Rebello

SGL - Smart Grid Lab, Computing Institute and INCT-MAACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Fluminense Federal University (UFF), Niterói, RJ (<http://sgclab.ic.uff.br/sgclab>), e-mail: [boeres,vinod]@uff.br

Fabio Porto

DEXL - Extreme Data Lab, LNCC/MCTI and INCT-MAACC - National Institute of Science and Technology in Medicine Assisted by Scientific Computing, Petrópolis, RJ (<http://dexl.lncc.br/>), e-mail: fporto@lncc.br.

workspaces, which can be defined as computing environments that can be tailored to execute a set of applications, based on the Cloud Computing paradigm. In regard to MACC requirements, the virtual workspaces enable the user to specify the computational resources and software for parallel applications in a straightforward fashion. This being possible without the hassle of cluster middleware configuration, while also leveraging the benefits of Clouds, such as elasticity and ease of access.

13.1 Introduction

Managing the huge amount of data produced by numerical simulation is a challenge for high performance computing platforms, which are the standard for high demanding scientific applications. In this context, some initiatives are successfully using cloud environments for scientific applications, as for example the *Magellan: a Cloud for Science* project [19] from the Argonne National Laboratory, and the *Nebula Cloud Computing Platform* project [16] from NASA. These projects arise in the context of private clouds, i.e. based on their own computing resources.

As a motivational example from a MACC application, the HemoLab simulation of the human cardiovascular system, when attempting to solve 3D or coupled models, can generate a linear equation system with over 3.500.000 variables [3]. In such cases, the simulation of just one heart beat cycle can take many days to solve on a single multi-core machine, so the development of high performance environments for MACC applications are crucial.

The cyberenvironments are actively developed and maintained by the following organizations: the Distributed Scientific Computing Laboratory (ComCiDis), the Center for Scientific Computing and Open Software (C3SL), the Group of Computer Networks, Software Engineering and System (GREat), the Smart Grid Lab Computing Institute (SGL) and the Extreme Data Lab (DEXL). The implementation of these services depends on the deployment of network services and autonomic systems that provides control and medical management, as well of the integration of these systems supported by a distributed, high performance computing (HPC) infrastructure.

These cyberinfrastructures thrive mainly because of the need of sharing the current medical knowledge and ongoing research, which implies the manipulation of high volumes of medical data. In that context, a computing system should offer basic services —such as managing medical databases and transmission of information between them— as well as advanced services, such as: a) data acquisition via remote monitoring and sensing; b) processing and transmission of biological signals and medical images; c) surgery planning via computational simulation of surgical procedures and via environments of virtual and augmented reality and d) teletraining of medical staff and video-conference medical sessions.

The rest of this chapter is as follows. Section 13.2 introduces concepts related to Distributed High Performance Computing, with a focus on Cloud Computing. Section 13.3 describes virtual workspaces, presents some additional services, working

projects and infrastructure. Section 13.4 describes the use cases in medical assisted applications. Finally, Section 13.5 concludes the chapter.

13.2 High Performance Distributed Computing

Science and technology are becoming increasingly dependent on resources with high computing power and storage capacity. This results from the development of research activities in various areas of science, generating even larger and complex data sets, making the use of high-performance computing essential to data analysis, storage and the quickly and efficiently sharing of this data and results.

In this section, we will review the state-of-the-art of HPC and Distributed computing, and explore their relevance to MACC applications.

13.2.1 High Performance Computing

While the peak performance of modern large scale computing platforms has grown dramatically, perceived application performance has not matched this [6]. An expert programmer has to apply an intensive effort to fine-tune an application to be efficiently executed on such high performance systems. This task still levels of specialized knowledge on software techniques, computer architectures, and programming languages by the scientist. Added to that, the growing desire to tackle highly interdisciplinary problems on complex computing platforms increases the difficulty faced by programmers and users.

In order to satisfy the computational requirements of e-Science applications, several strategies have been adopted, and are subject to continuous optimizations. Several MACC applications have requirements that can be expressed by parameters like: a) turn-around time, the interval between a task submission and its termination; b) throughput, number of tasks completed per time unit; c) waiting time, total time that a task has waited for execution since its submission; d) response time, time interval between a task submission and the first response in interactive systems; e) makespan, defined as the maximum completion time of any job in the set of jobs under consideration [17].

Parallel Computing is a strategy for achieving high performance. It represents a vision of how computation can seamlessly scale from a single processor to virtually limitless computing power, by scaling the execution of an application to more processors. If N identical tasks that comprise a job have no inter-dependency, then they all can be executed in parallel, theoretically dividing the makespan of the job by N . Parallel jobs usually run on multi-core, high-performance computers with shared memory that have highly optimized processor communication and high bandwidth I/O. The main downside of these supercomputers is the price.

Still, efficient scalability and load balance are very hard to implement in parallel HPC, which depends on the tasks' properties and their inter-dependency. Because of this, many HPC applications that run on supercomputers are tuned to execute on that particular hardware with exclusive access, thus losing features like flexibility and resource sharing. One of the challenges for the software development is the successful management of nowadays high performance computing platforms and at the same time, making available strategies and proper hardware resources.

One can note that the current trends of HPC are oriented towards extreme-scale complex infrastructures with a great deal of heterogeneity in the computing and communication resources. In HPC environments, an efficient scheduling of an application on these heterogeneous large systems is critical for achieving high performance. Since it is difficult to statically predict an application behavior, dynamic scheduling and load balance techniques need to be employed. With this view, a scheduling model must identify both application and target systems characteristics relevant to guide the scheduling decisions.

In modern HPC platforms, both shared and distributed memories are an opportunity to improve the execution of scientific and data-intensive applications, which often require analysis on large datasets. However, in order to fulfill these expectations, data locality must be explored. The dynamic data analysis and the memory availability are mechanisms used to schedule computing tasks close to its data. On the other hand, if the amount of data leads to poor performance due to more frequent cache misses, distributing data over the network will provide higher benefits.

Because dedicated supercomputers are very expensive, alternatives to obtain high performance were soon pursued, hence clusters were realized. While there is no precise definition of a computing cluster [12], it can be understood as a collection of commodity components that are interconnected to form a system able to perform parallel computations in a fairly efficient manner. These features present the cluster as an adequate environment for the execution of parallel and high performance applications, where the main scalability enablers are data and task parallelism. A cluster is usually deployed with one master node, and several computing nodes. The master node is responsible of: a) providing a job submission interface to the user; b) managing and distributing the application tasks among the computing nodes and c) collecting execution results and making them available to the user. The cluster approach provides several benefits, such as high performance, scalability, fault tolerance and low economical cost.

13.2.2 Distributed Computing

Distributed computing as alternative to parallel computing, the main strategy being the agglomeration of several distributed sites to match computational requirements of resource-demanding e-science applications. This general strategy is different as of the more restricted parallel computing, because it relies in the distributed memory paradigm, as opposed to shared memory. A distributed system is a collection of

independent computers that appear to the user of the system as a single computer. It offers great computational power to its users by means of resource aggregation. Because there is no global memory among processors, they also do not share a universal clock. This hardens the synchronization of processes, making the design of distributed applications from scratch inherently difficult.

Distributed systems provide several theoretical benefits: speed, reliability, resource sharing, flexibility. Their main advantage is that the cost-benefit is higher when compared to tightly coupled multiprocessors (parallel supercomputers), because several commodity devices can be gathered to collectively produce high-performance. On the downside, distributed systems are very complex and their realization is far from trivial, also are subject to new types of faults and problems, e.g. the uncertainties behind of network delay.

13.2.2.1 Grid Computing

A Grid can be defined as a system that coordinates resource sharing and problem solving in a dynamic, multi-institutional Virtual Organization (VO). Computing Grids are a type of distributed systems, the main characteristic being a high degree of heterogeneity in the existing hardware, networks and administrative domains. Grids are focused on large-scale sharing of heterogeneous computational resources, which are geographically distributed and offer high performance [11].

The Grid was envisioned as a distributed system that would allow e-Science applications to seamlessly scale, in such a way that it would appear as extending the local hardware into a large computing infrastructure, accessing massive amounts of data over a high bandwidth network. It refers to specific capabilities:

- Find and share data: Information residing on different institutional domains, but in the same VO, should be transparently accessible to users that have the proper credentials and permissions.
- Find and share applications: The underlying infrastructure should support sharing applications easily, simplifying their composition and customization.
- Share computing resources: A Grid should let users both offer their free resources to others and use available resources from the community.

13.2.2.2 Cloud Computing

Cloud Computing is a recent and popular trend, where the data is informally stored and processed remotely or “in the Cloud”. Cloud systems emerged as a new type of distributed solution. Its main feature is called elasticity, which can be defined as the capacity to offer high flexibility and scalability of the available resources, expressed for instance, in the ability to acquire a large amount of resources in a relatively small amount of time, with the possibility of resizing these resources as the demand changes. In this context, Cloud Computing can be regarded as a new resource provisioning model [20].

The Cloud model is becoming more prevalent, and is currently being utilized and studied in data processing centers worldwide. Besides elasticity, it incorporates advantages such as application scalability, resource sharing, and the possibility of accessing large-scale processing power and storage capacity. This however, does not imply that the Cloud approach is the best solution for all computational requirements. Issues like security, connectivity and some performance aspects are still referenced as challenges for this paradigm.

Clouds arose mainly from the business sector, but currently it has gained a lot of attention from the scientific community, that find Cloud Computing attractive because it simplifies the process of requesting and accessing computing power. However, the Cloud paradigm has important implications for the types of applications and their performance, and they are a matter of current research. Since Clouds form the basis of the proposed cyberenvironments for MACC, Section 13.2.3 will discuss Cloud Computing in more detail.

13.2.3 Foundations of Cloud Computing

Cloud Computing can be defined as ‘a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction’ [15]. Cloud Computing can be regarded as a business and provisioning model, while a Cloud system can be regarded as a new type of distributed system which attempts to implement that model [20]. This section presents the features of the Cloud Computing paradigm and the properties of Cloud systems.

13.2.3.1 Virtualization

In this context, virtualization refers to machine virtualization, which stands for the creation of a virtual version of the physical hardware of a computer. One such virtual version is called Virtual Machine (VM), and is achieved by a software layer on top of the real machine, which creates a VM with a (potentially) different vision of the underlying resources. Machine virtualization allows the execution of several independent VM instances, each with its own operating system and with apparent exclusive access to the hardware. In this way, functional isolation is achieved naturally, making the execution of applications more reliable, as failure in a VM is contained in its own execution sandbox. The main drawback of virtualization is the performance impact of the additional resource management layer. The impact of virtualization on standalone and distributed and/or high-performance applications is still being researched [2], [7], [21]. While virtualization has been available for many years, recent technological advancements, such as hardware support for virtualization functions, have made VMs an attractive solution for efficient resource management.

The filesystem of a VM instance contains the operating system and the user applications for that instance, and is called a VM image (VMI). When a VMI contains a single application that defines the purpose of the VM, then this VM is called a Virtual Appliance (VAP). A VAP is like a physical appliance but without the hardware; as such, a VAP is like software and can be shipped and stored electronically [18].

The Virtual Machine Monitor (VMM) is a software that makes it possible to instantiate VMs, by providing a layer between the operating system and the hardware. The VMM provides an interface that supports the management of the VM life-cycle operations: create, modify, migrate and destroy a VM. Hardware virtualization allows dynamically reshaping the hardware profile observed by the VM. This means that variable resource availability requested by an application can be satisfied by a VM resizing. Also, VM migration is a transparent process—meaning that the application execution is not interrupted—and can be leveraged for fault tolerance and performance improvement.

From an implementation point of view, computing resources in a Cloud system generally comprise Virtual Machines (VMs) running on heterogeneous hardware under the control of an homogeneous virtualization solution in the form of a deployed VMM. Hardware virtualization has a fundamental role in Cloud systems, offering the advantages of VAP networks, such as: a) server consolidation, i.e., the agglomeration of workloads from underloaded servers to loaded servers; b) legacy-code support without affecting other applications; c) improved security through the sandboxing for applications that may alter the execution of other processes and d) the ability to create virtual clusters. A virtual cluster the result of applying machine virtualization to the concept of a cluster (Section 13.2.1). It can be implemented as a collection of similar VMs, often instantiated from the same VMI, that have been contextualized at runtime to provide support for a parallel environment, thus emulating a physical cluster. Using virtual clusters in a Cloud environment is a powerful way to easily deploy parallel applications with the benefits of clusters, that can scale automatically by requesting additional VM instances.

In summary, virtualization is a key enabling technology for the realization of the Cloud Computing paradigm, as it provides a useful abstraction that can be used for both managing and provisioning computing resources according to the Utility Computing model.

13.2.3.2 Cloud Computing Models

Current Cloud taxonomy considers three layers or service models, and four deployment models [10] [15]. The service models define how the resources are delivered, while the deployment models distinguish the purpose of the infrastructure. Here we present a brief description of these models, starting by the service models:

- **Infrastructure as a Service (IaaS):** in this layer, the processing and storage resources are dynamically delivered as a service, with the purpose of providing ad-hoc systems to users, mainly through a solution based on machine virtualization. The IaaS model offers the most flexibility of the three, because it enables

the capacity of provisioning computational resources that can be used to deploy and execute an arbitrary software.

- Platform as a Service (PaaS): in this layer, a software development platform is delivered as a service to deploy, test, maintain and scale applications in an integrated development environment. A PaaS solution is usually designed on top of an IaaS infrastructure. It goes beyond traditional hosting services, because it includes some of the IaaS features, such as automatic scaling, provisioning with elasticity and a specific execution environment for each virtual workspace.
- Software as a Service (SaaS): in this layer, applications are delivered as a service in order to be accessed on-demand by the users. Interaction with the applications is usually restricted by a data presentation layer, as through a Web browser.

According to the deployment aspect, Cloud systems are classified in the following models:

- Private Cloud: the infrastructure is operated for a single organization, either by the same organization or a third party, with the intent of maximizing control over data. Several current software developments allow the deployment of private Cloud infrastructure.
- Community Cloud: the infrastructure is shared by several organization and supports a community with common interests.
- Public Cloud: the infrastructure is made available to the general public.
- Hybrid Cloud: a composition of two or more Clouds.

13.2.3.3 Utility Computing

Utility Computing can be best described as a pay-as-you-go business model, and is closely related to the Cloud Computing paradigm. Under this model, computational resources are provisioned on demand, and are metered according to the utilization rate. In this way, resources can be accessed in a fine-grained manner (e.g. in units of computation), and priced over effective consumption. This is ideal for scenarios where applications need to be executed several times on demand with different scales, such as when alternating between testing simulation parameters and performing a full-sized computation. The Cloud leverages the VMM capabilities of fine-grained management of distributed resources, and also implements resource monitoring to detect the real consumption of computing power, storage capacity and network transmission.

The users and the Cloud provider are bound to their obligations through a Service Level Agreement (SLA), which is a contract between a user and a provider. In a private Cloud environment, there is only a single provider responsible for all resources, thus the provider has the important responsibility of maintaining the resources available and, for PaaS, having the platform operational. This is implied in the SLA contract, in the form of QoS parameters that the infrastructure/and or platform should satisfy. Under the Utility Computing model, the provider may undergo penalties for violating the SLA.

13.2.3.4 Evolving from Grids to Clouds

The appearance of Cloud Computing systems in the last years has come not from the scientific community but from the business sector. Both of these distributed systems allow access to computational resources and present common features like scalability, fault tolerance, access to infrastructure and application interfaces. In the earliest stages of Cloud Computing, there was a lot of buzz around this new paradigm and often the terms “Grid” and “Cloud” were used interchangeably. Now that Clouds are being shaped and becoming more identifiable as systems, it is possible to compare them with Grids.

Cloud Computing has the same fundamental goal as Grid Computing: “To reduce the cost of computing, increase reliability, and increase flexibility by transforming computers from something that we buy and operate ourselves to something that is operated by a third party” [10]. While the main problems remain very similar as before, the Cloud approach leverages the recent technology breakthroughs that allow for low-cost, efficient use of virtualization to provide access to computing infrastructure with commodity hardware.

The general approach on the Cloud Computing paradigm has been the implementation of the Utility Computing business model. This implies that the user expects the delivery of processing and storage units, which appear to the user as dedicated resources. This is an illusion, because Cloud providers use internally a multi-tenant architecture to serve different users in the same piece of hardware at the same time, and the effect is provided by virtualization down to the hardware level. This approach is different from the general approach in Grids, which mostly rely on either: a) time sharing and task queuing, and b) complex and intrusive middleware that allows shared access to the resources.

The architecture of Clouds and Grids is also similar but different. The following list represents a brief comparison of the architectural layers of Grids and Cloud, where the layers have been adapted for proper comparison.

- **Fabric:** Both systems share this lowest layer, which is composed of heterogeneous, inter-connected resources: computational, storage and network resources, among others. In Clouds, the computational resource is provided through at least one virtualization solution running on computing nodes, which provides homogeneous hardware abstraction.
- **Connectivity:** Consists of communication and authentication protocols. The communication protocols are standard, like TCP/IP and UDP. In the case of authentication protocols, the Grid middleware provides support for single sign-on and delegation to access shared resources. The notion of single sign-on is present in the Cloud, because once the user logs on, he has access to all the resources that were delivered. On the other hand, the delegation is unnecessary because the resources are not shared.
- **Resource:** In Grid systems, this layer is implemented by the middleware and provides protocols for secure initiation, negotiation, monitoring, control, accounting and payment of sharing operations on individual resources. In Cloud Computing, these aspects are managed in a similar way but concern solely to

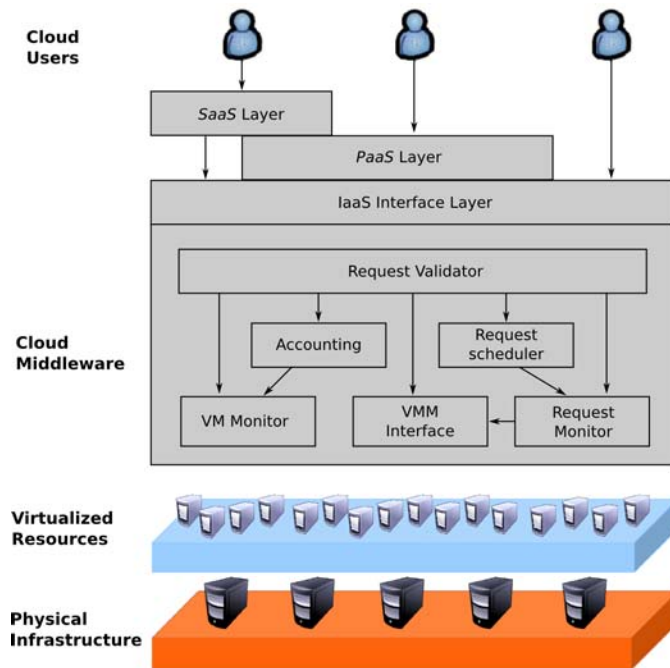


Fig. 13.1 Layered view of a general purpose Cloud.

the Cloud Provider and remain hidden to Cloud users like application developers. In Cloud systems, most of the functionalities for this layer relate to the aggregated management of VMs on the computing nodes.

- **Collective:** This layer deals with protocols and services that are not associated to a particular resource, but rather are global in nature. Current Grid middleware focuses on this layer and provides extensive support for resource discovery and allocation, job scheduling and monitoring, and so forth. The Cloud provider also deals with these aspects, but exposes a more concise interface to the user, in accordance to the Utility Computing paradigm and the available service models.
- **Interface layers:** As presented in Section 13.2.3.2, Clouds may support different service models, which are represented in the Cloud architecture as interface layers built on top of each other. The lower interface is the IaaS layer, then comes the PaaS layer, and finally the SaaS layer. While there is no direct, layer by layer comparison of the Grid interfaces, Grids may expose implementation details as a trade-off for flexibility (similar to PaaS); on the other hand Grids may provide more ad-hoc solutions to support applications, such as Grid Portals [10]. Cloud interfaces tend to be more narrow and expose less details, trading loss in flexibility for user-friendliness.

Figure 13.1 shows a layered view of a general-purpose Cloud system. The physical infrastructure stands for the Fabric and Connectivity mentioned above. The Vir-

tualized Resources represent the Resource layer, here the VMM provides an homogenized abstraction of the underlying resources, along with management of the VM life-cycle. The modules in the Cloud middleware implement the Collective layer, and provide the necessary services that enable the management of the distributed resources. These service includes resource discovery and monitoring, resource aggregation, request scheduling and execution monitoring. Finally, the interface layers are represented explicitly in the figure.

Here follows a comparison of additional characteristics between Grids and Clouds:

- Both are massive, scalable distributed systems built by the aggregation of heterogeneous resources and the principles of service orientation.
- The implementation and maintenance of such infrastructures is a complex task. In particular, the flexibility of the IaaS and PaaS layers are hard to implement, but many solutions currently exist that provide IaaS out-of-the-box.
- The management of resources should be easier than the case of Grids, because of the aggregated virtualization middleware.
- Scalability is achieved seamlessly in Grids and Clouds: in Grids, by increasing the number of compute nodes; in Clouds by resizing virtualized resources as well.
- Grids offer new functionalities that emerge from the aggregation of low level resources, they are general-purpose systems that show bottom-up design. In contrast, A Cloud appears to be designed top down, the interface it exposes is the goal of the design and the purpose of aggregating resources.

Finally, the difference of design priorities between Grids and Clouds have implications on the types of applications that are best supported. Large Grid systems tend to be constructed by aggregating high-performance resources, while striving to satisfy QoS requirements related to both performance and time of execution. On the other hand, Clouds current design is aimed to satisfy varying workload for many users. This preference is a trade-off for performance. While virtualization shows a negligible impact on computation, it may produce considerable overheads, as in I/O access and intensive inter-VM communication (ref). This consequence is an important obstacle for wide deployment of HPC applications in Clouds, and is a matter of current research.

In summary, Cloud Computing can be seen as the next step in the evolution of Distributed Computing, following Grid Computing. Clouds can be used in collaboration with Grids, for instance to satisfy hard-to-predict spikes in resource demand.

13.3 Virtual Workspaces for Scientific Applications

We define a Virtual Workspace as an computing environment where the user is capable of either specifying the characteristics of computational resources and applications in order to execute a particular job, or choosing a pre-defined execution

environment for that purpose. This section elaborates on the motivations behind the proposal, the design requirements and also the implementation of the infrastructure.

The main objective is the development of a scientific computing environment and its nation-wide deployment, integrated by a scalable infrastructure composed of a collection of Clouds, Grids and Clusters, spread over different institutes in the country and able to allow access to computing resources in a trusted, consistent, pervasive and affordable manner.

Besides medical applications, the cyberinfrastructure must support the development of other research areas, such as databases, collaborative environments and data mining, among others. Considering the continuous improvement of available information and communication technologies, it is expected that the design of computational systems can be based mainly on the abundance of available infrastructure resources and geographically distributed computing potential. These advancements will lead to the great challenge of defining and leveraging dynamically configured systems, which interact in new ways with its environment in order to execute or manage their computational tasks.

The infrastructure aims to satisfy the growing demand, by the MACC network, of high performance computational resources and data visualization. It is a responsibility of the supporting organizations to continuously upgrade its resources in order to support the computational demands, as well as to participate in the development of computer and model simulations related to MACC.

13.3.1 Purpose and Requirements

The main purpose of providing virtual workspaces is to provide MACC users with access to a computing platform for their scientific applications, without depending on specific knowledge on informatics in order to use and manage those platforms. Researchers, specialists and students would have remote access to scientific applications, which are running on dedicated and/or high-performance infrastructure and are configured according to their needs. Users should not require any previous software in their workstations, using instead the computational power available from a dedicated environment. This also implies that the infrastructure should be able to store application data.

Currently, due to the nature of their scientific applications, MACC users require environments that provide explicit support for parallel processing. Thus, the virtual workspace should support the form of a cluster (Section 13.2.2). The cluster environment should explicitly support the following quality requirements:

- **Usability:** MACC users are only expected to use their scientific applications and not manage complex environments. The environment interface must be user-friendly, possibly in the form of a portal with step-by-step tutorial support, in order to allow intuitive and practical utilization of HPC resources.
- **Flexibility:** The users should be able to leverage the flexibility from virtual workspaces, i.e. choose parameters that produce a customized execution en-

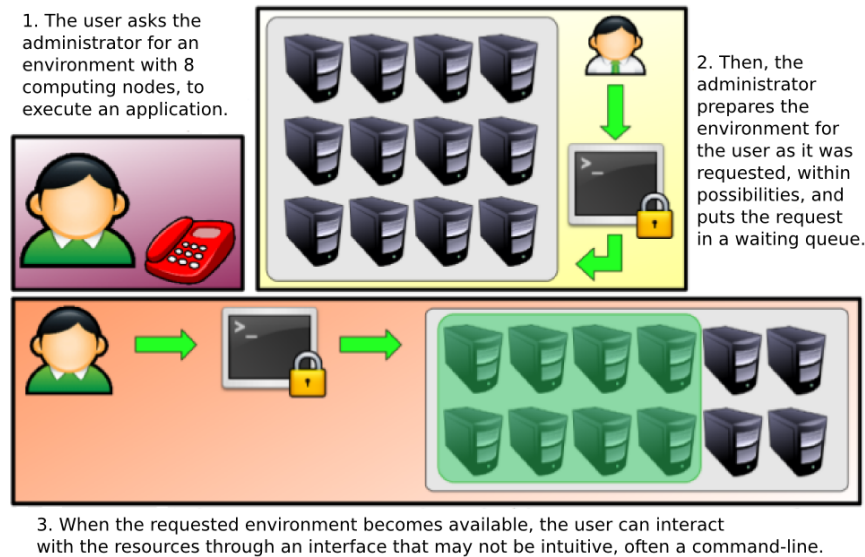


Fig. 13.2 Use case presented as motivation for Virtual Workspaces.

vironment for an application. From the service provider's perspective, these heterogeneous environments should be provisioned independently on the same hardware in a straightforward and reliable manner.

- **Security:** Due to the sensitivity of medical data, the communications should be secure, meaning that users should be authenticated against the infrastructure and authorized to interact with a particular virtual workspace.

In order to design an environment that provides such capabilities and requirements, we observed the routine and procedures of users from MACC-related fields, who interact with scientific cyberenvironments. The procedures include: i) the request for computational resources, ii) accessing the environment, iii) job submission and iv) result retrieval. After observing these routines, we perceived that some of the steps could be optimized, speeding the user's work, maybe even making the tasks more intuitive by a more user-friendly interface. These improvements could also be applied to the infrastructure managers, by the standardization of procedures and automatization of resource deployment for users.

Figure 13.2 presents a scenario that can be used as an example of the previous procedures for requesting resources at MACC. Because the environment must be prepared for the execution of a particular application with specific hardware and software requirements, the explicit interaction with the administrator is necessary. This step is also involved with the authorization and authentication process of the user and the request. Since the environment is configured for the application and functional isolation is required, there are very few opportunities for resource sharing, and the infrastructure relies instead on request queuing to attend several users.

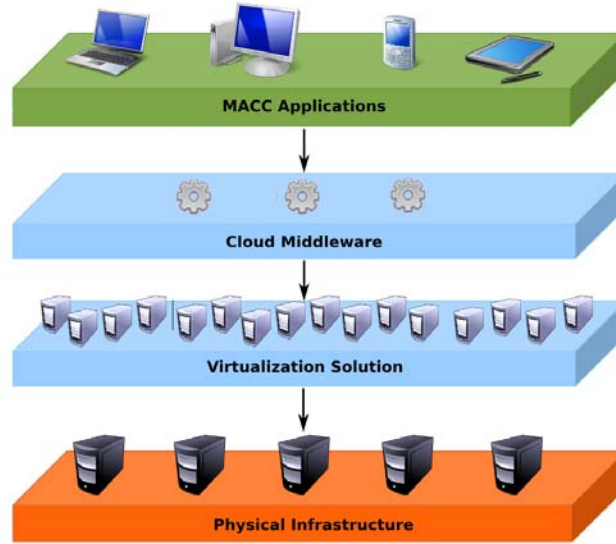


Fig. 13.3 Conceptual layers of the Virtual Workspace model.

13.3.2 Design Considerations

Based on the aforementioned requirements, we chose to implement the virtual workspaces under the Cloud paradigm, which offers the capabilities and advantages presented in Section 13.2.3. Broadly speaking, the idea is to provide such an infrastructure that allows the deployment of different types of execution environments, ranging from a simple, pre-configured virtual workstation to a complex virtual cluster with task management capabilities; all these with a simplified request interface and rapid response. Figure 13.3 shows a high-level conceptual diagram of the architecture: the user applications will interact mainly with the virtualized infrastructure of the cyberenvironment, masking the implementation details of the physical infrastructure.

The proposed solution involves the usage of virtualized resources under a Cloud environment, the concept of virtual workspaces and the configuration of execution platforms as “turnkey appliances”, which are minimalistic environments for the execution of a specific application. With the turnkey appliance approach, the user is not concerned about the installation, configuration, maintenance or minimum execution requirements, thus simplifies the interaction by activating an instance, using it and shutting it down when done. From the application point of view, the execution environment is tailored to have only the minimum requirements, thus has less performance overhead. From the infrastructure point of view, the server utilization is optimized by the reduced computational footprint of the VMs. The current proposal aims to manage and deploy virtualized resources on a large scale and as a service. This approach enables environment customization, optimization of the

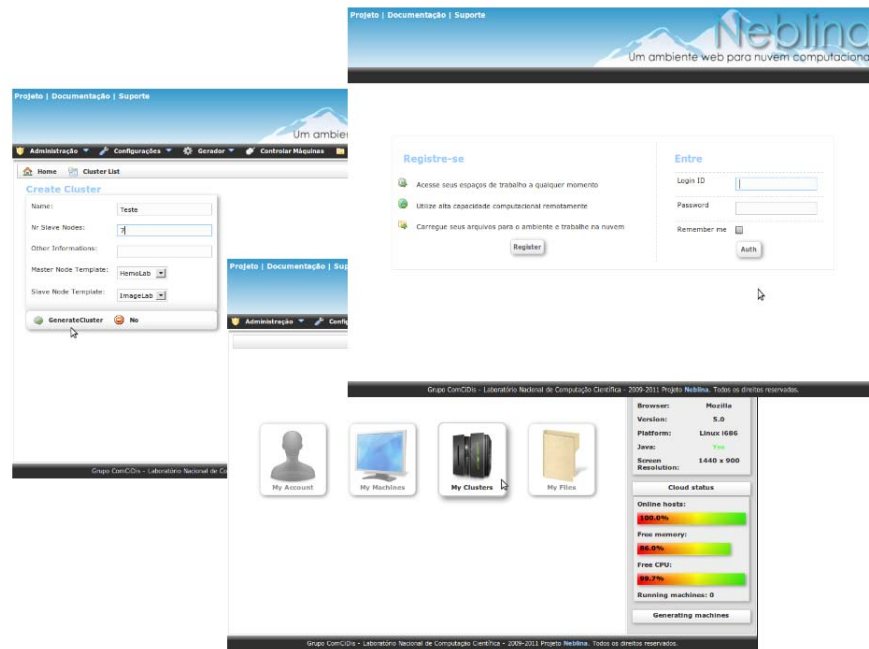


Fig. 13.4 Overview of the Neblina web environment.

usage of computational capacity, and information sharing to promote interactions between researchers.

The current effort on developing a cyberinfrastructure on virtual workspaces was partly based on previous ideas from the Virtual Community Grid (VCG) project [1]. That project proposed the development of a Grid Computing infrastructure to attend the computational requirements of the community involved in the project. The VCG featured a web portal as the main user interface for job submissions, and included a set of tools that aided in the deployment of clusters inside the infrastructure. The current initiative also relies on a graphical portal to strive for simplicity. Figure 13.4 shows some views of the web interface that allows system managers to monitor and control the infrastructure. Through the management interface, information about the state of the resources and submissions can be obtained, as well as deploying more resources in a simple and intuitive fashion. The portal also includes tutorials to speed up the user learning curve.

The Cloud infrastructure makes the creation of the virtual workspace straightforward and with reduced administrative interaction, while also presenting an end-to-end solution to the configuration and deployment of an execution environment for specific applications. These features greatly enhance the usability of the system by users and administrators. Finally, since the VMM can provide functional isolation through sandboxing, the system is able to attend more simultaneous users, improving flexibility and resource sharing.

The development of a cyberinfrastructure that enables virtual workspaces was also motivated by ongoing research focused on evaluating the utilization of Cloud systems to support the execution of scientific applications. For example, [19] evaluated several benchmarks and scientific workloads on resources related to the U.S. Department of Energy, in order to investigate the role of computational Clouds in HPC. The obtained results indicate that the Cloud paradigm provides to users the flexibility to configure their own execution platforms and applications, but often these configurations are complex and demand some knowledge of system management. Also, the work of [4] describes an implementation that enables the creation of customized virtual platforms for a particular application. This application is composed of a web interface for user interaction, a task manager and background daemons. The task manager handles the incoming user requests, while the daemons monitor the state of the infrastructure. The application then generates an execution environment for a request, providing a link to access the corresponding VMI.

13.3.3 Providing Virtual Infrastructure

This section introduces the Neblina project [8] [9], an initial effort that seeks to enable the installation, configuration and management of a private Cloud environment for virtual appliances, currently under development. Neblina was developed from the need to provide a flexible computing environment to applications based on scientific visualization, and is currently deployed in the ComCiDis physical infrastructure. Based on the paradigm of Cloud Computing, it aims to provide solutions through the use of virtual platforms for testing, development and deployment of scientific applications; providing to its users a dedicated and customized environment. Neblina can be regarded as a Web Service that only needs to be installed in a master node, while the communication with the computing nodes is performed with the remote VMM instances through the exchange Linux security keys. In this sense, it differs from other IaaS solutions, and can be better described as a Cloud middleware, because it provides an infrastructure so that applications can interact with distributed software.

Neblina optimizes the use of existing resources through its balancing and sharing, and also incorporating the concept of “turnkey solutions” to provide ready-to-use environments for dedicated applications. In this context, ready-to-use stands for an execution environment composed of one or more virtual machines, where each VM includes an adequate operating system, previously tested libraries and a correctly configured appliance, aimed specifically to deploy the user’s application or execute the user’s jobs.

The implementation of Neblina is enabled by a number of software modules that provide the functionalities required to support a Cloud infrastructure. Figure 13.5 represents a layered module view of Neblina’s architecture, that shows how Neblina can build support for virtual workspaces. The following is a short description of each module:

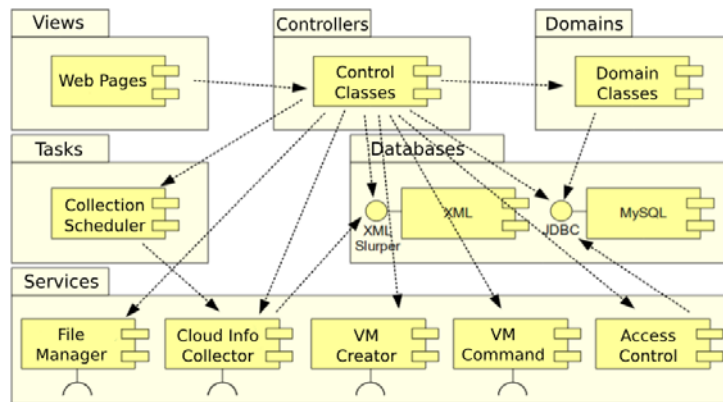


Fig. 13.5 Layered view of the Neblina Virtual Workspace modules.

- *Controller, Views and Domains*: this first layer embodies the Model View Controller (MVC) design pattern. These modules are responsible for the management and presentation of the Cloud's status data, including the deployed virtual workspaces and uploaded files. The views provide an interface for user applications to the system in the form of HTML pages.
- *Databases*: this module provides a persistence layer for the data managed by the MVC modules. It is based on the Grails Object-Relational Mapping (GORM) library and Hibernate, and is responsible of populating a transactional database. A XML module stores dynamic data, such as the state of the VMs and the computing nodes, as to optimize data access.
- *Tasks*: this module is a collection of small services and daemons that are executed automatically, and are aimed to monitor the state of the infrastructure and relay this information to the interface.
- *Services*: this module is responsible for the execution of procedures that support the VM life-cycle, storing files, collecting information on the infrastructure and enforcing security constraints. The *VM Creator* and *VM Command* submodules interact with the remote VMMs, which execute on the computing nodes, through standard protocols.

In order to provide infrastructure to support the Virtual Workspaces environments, Neblina offers the following functionalities:

- *User management*: two types of users are supported, the clients and the administrators. The role of the administrator is to manage the environment: monitor the status of the Cloud, deploying virtualized computing nodes, and if necessary, managing the VM instances. The client is able to interact with his deployed virtual workspaces and upload/access shared files.
- *Host management*: after successfully configuring a new machine for the environment, the administrator chooses the role of the machine, which can be either

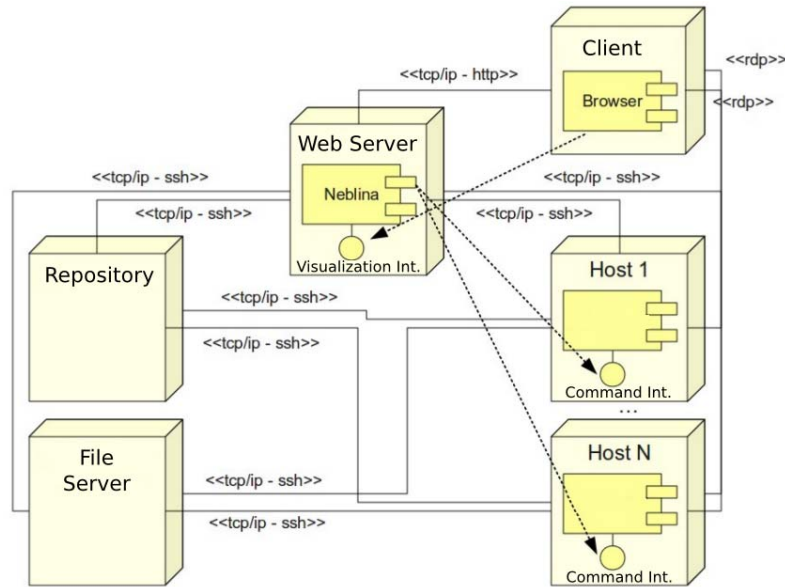


Fig. 13.6 Deployment View for Neblina.

a computing node, a VMI repository or a file server. After that, the host appears in the monitoring interface.

- **Virtual workspace management**: each virtual appliance contains the operating system and software necessary to run a single application on demand, without the need of configuration. The corresponding VMIs are stored in a repository, and are cloned to a computing node when a virtual workspace is requested. After the appliance is deployed on the node, it is made available to the client.
- **File management**: here the requirement of storage of application data is met. In order for the data to be available, clients should migrate the data to the Cloud. Thus, the infrastructure includes basic abstractions for file management, that provides one private folder for each client. The included functionalities are uploading, downloading, navigation, compression and extraction of files. As a result, the Cloud supports the input and output of applications.
- **Virtual Machine Creation**: the VMs themselves are created by the administrator and made available to the clients. From a VM template, a client is able to customize a virtual workspace, by specifying the resource parameters (amount of memory, CPU cores, storage capacity and others). This strategy leverages the VM model to reuse environment specifications among similar workspaces.
- **Workspace management**: after a virtual workspace is deployed, a graphical interface can be accessed by using either: a) the web interface via an applet that implements the Remote Desktop Protocol (RDP); or b) any RDP session client. When using the applet, the graphical desktop is customized to show only the application in a running state.

The current deployment strategy of the Virtual Workspaces infrastructure is shown in the deployment view on Figure 13.6. All servers were configured with *Ubuntu Linux 10.04 Server* operating system. A server was selected to work as a repository for the image files for the models of the virtualized clusters. The repository is accessed internally by the service responsible for VM creation, in order to fetch and transfer the VMI file that is required by a requested virtual workspace. Another server was chosen to execute the following components: a) the core web application (deployed in *Apache Tomcat 6.0.29*); b) the storage for the Cloud's state (*MySQL Database 5.0* and XML files) and also c) the file server for the client files.

13.3.4 Execution and Resources Homogenization Architecture

The ERHA proposal attempts to address some important challenges in running scientific applications in Clouds, such as: a) initial application deployment; b) application execution and c) the provisioning of computing resources in a heterogeneous physical infrastructure, where the VM performance depends on physical characteristics. The proposal is intended for sequential and parallel batch applications, which are common in e-science. The solution aims to provide mechanisms to submit and execute these applications in private IaaS Clouds, using homogeneous virtual environments created over heterogeneous physical infrastructure.

The ERHA architecture meets these requirements by proposing three elements: i) a language to describe resources and execution parameters; ii) a method to deploy and execute applications based on the described settings, and iii) a mechanism that allocates VMs using processing units (PUs) as a metric that represents effective processing power of each machine. The architecture can be represented with three layers, which are placed between the IaaS Cloud middleware and applications. Figure 13.7 shows a layered view of the architecture.

The top layer is called the Resource Description layer. Deployment and execution metadata is inserted into the application's source code, as a block of attribute-value pairs (Resource Description Block, or RDB). The RDB parser in this layer reads all fields and converts them to a specific Cloud middleware format.

The middle layer is called the Execution Management layer. It handles the entire execution process and comprises three components: the Client, the Server and the VM-Daemon. The Client component is installed on each user machine and provides an interface to submit applications to the Cloud through ERHA. The Server component is deployed on the Cloud front-end. It receives the request from the RDB parser, relays the request to the Allocation Layer and interacts with the VMs to provide the proper execution context for the application. The VM-Daemon component is inserted in the VMI and deployed when the VM boots. It sends and receives data relevant to the VM's deployment.

The bottom layer is called the Allocation layer, and it satisfies the requirement of provisioning homogeneous resources using PUs. A PU is the abstraction used to represent processing power, and supposes a fixed VM processing power, even

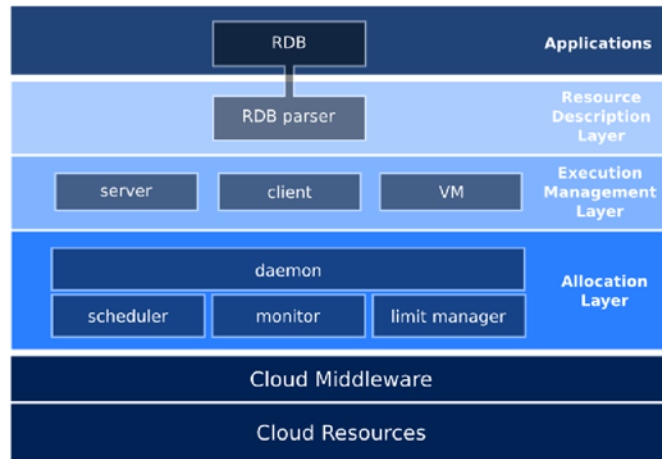


Fig. 13.7 Layered view of ERHA architecture.

with underlying resource heterogeneity. The Limit Manager applies limits on CPU usage, so that the VM processes with the requested amount of PUs. Monitor is the module that captures information related to the CPU, memory, storage and network from the infrastructure. The Scheduler module decides on the allocation of VMs in the physical resources. The decision is based on the amount of PUs required by a VM and the amount of PUs available in the PMs. The Daemon module manages the other three modules and controls the infrastructure. The Daemon receives requests from the Execution Layer and invokes the Scheduler when there are pending VMs. When the Daemon receives the VM-PM mapping from Scheduler, it allocates these VMs. The Daemon is also responsible for storing information about the PUs and invoking the Limit Manager to apply the limit of CPU usage.

The ERHA proposal presents a solution to automate the deployment and execution of batch applications in Clouds, providing mechanisms to create homogeneous virtual environments over private Cloud middleware. It focuses on reducing the impact of the data center heterogeneity in the VMs performance, also provides mechanisms to enable running applications in Clouds in an easy and uniform way, abstracting the infrastructure and middleware complexities.

13.3.5 Analytical Data Management for Numerical Simulations

Numerical simulation of natural phenomena is being fostered by recent advances in powerful high processing computing platforms. Scientists at LNCC, INCT-MACC [13] model a phenomenon being studied through a set of mathematical equations. The latter are transformed into a computing model, using one of the available numerical methods. As scientists strive to obtain a more realistic simulation, a huge

amount of data is produced. Unfortunately, there has been little work on supporting numerical simulation data management, which leaves simulation scientists with huge standard text files and complex analytical programs that eventually extract some meaningful information to validate scientific hypotheses.

In this context, we are trying to bridge this gap by raising some issues involved in numerical simulation data analysis [5]. A representation for numerical simulation data is presented that considers a multidimensional model, for dimensional variables, and their corresponding physical quantities. The SciDB system, a science-oriented DBMS, is used to implement a first prototype supporting the human cardiovascular system simulation developed at the LNCC. Additionally, a cloud service to interface with the numerical simulation data manager is proposed and its integration with Neblina is explored. We focus on the support to the analysis of simulation results, taking a sample of a simulation output and loading it onto a cloud data service modeled to manage space-time multi-scale dependent data.

13.3.5.1 Using SciDB for storing simulation data

Our first effort to represent numerical simulation data uses SciDB, which is a data manager that uses multidimensional arrays structures as the basis of data representation. A user specifies multidimensional structures by providing the range values for each dimension (Δ dimensions) and a list of attribute values (Π) to compose a cell. Figure 13.8 shows how each cell element is a coordinate in the multidimensional space and has values for each attribute. For each cell, the attributes shown in the figure are velocity (three components), pressure, and displacement (three components).

In addition, a versioning mechanism keeps historical values for each attribute. The simulation instance can also be used as a dimension in the array. This strategy can be leveraged to execute many related simulations, each with different parameters, and then compare the simulation results. In this context, the following mapping strategy has been defined. For each set of physical quantities corresponding to a phenomenon being simulated in a given scale: *i*) define the set of Δ dimensions involved; *ii*) specify the list of physical quantities Π to be computed; *iii*) create an array having the dimensions as Δ and attributes as Π .

13.3.5.2 Analytical data management service

From an architectural point of view, we expect to develop a service to interface with the solvers - producing simulation data - and with scientists - submitting analytical queries to the system. Fig. 13.9 shows the architecture of this service where the Simulation Data Management Service (SDMS) is responsible for providing such an interface as a cloud service. The SDMS should manage the storage and retrieval of the simulation data making it transparent to scientific applications. The SDMS may

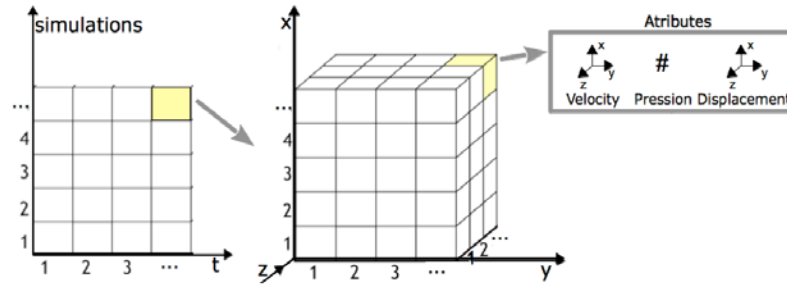


Fig. 13.8 Representation of a multidimensional array and its attributes.

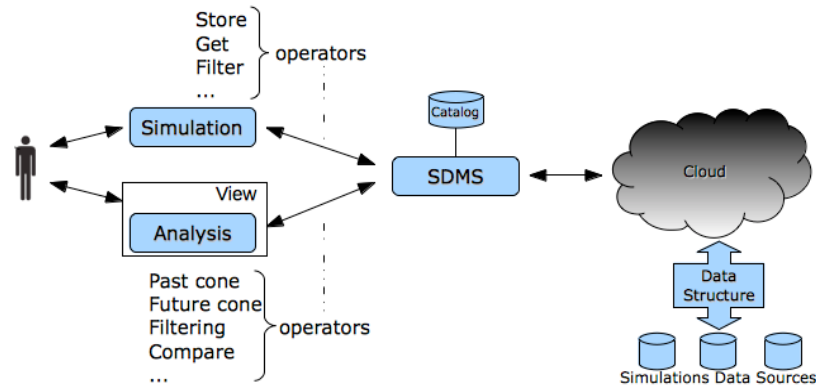


Fig. 13.9 Simulation Data Management Service (SDMS) Architecture

then be queried to study the effects of the input parameters on attributes of interest of a given simulation, and to compare simulations with different parameters.

13.3.6 Opportunities for Computer Assisted Medical Applications

The rapid provisioning of independent and isolated resources, hardware and software customization, quick access to resources as well as on-demand scalability, have made cloud computing an attractive model for the scientific community. In fact, many scientists have adopted this new paradigm, moving their data and performing *in silico* experiments in the cloud.

However, the deployment and execution of scientific applications are generally not straightforward, comprising a set of complex hardware and software configurations. Other issues must be considered in a cloud scenario, e.g., particular control aspects of different clouds, interfaces to be used, number of virtual machines (VMs) to be deployed and what resources will be needed by them.

Currently, Hemolab is built using the Message Passing Interface (MPI) library to enable the execution of parallel computing simulations on clusters. Among the main challenges of migrating a local cluster solution to a distributed environment, we highlight: a) adapting the execution from a homogeneous to a heterogeneous environment; b) building software layers that can interface between the application and the underlying infrastructure and c) secure data transfer between domains. Besides those points, more research is needed in order to establish parameters —such as the best number of computing processors and cores— for the execution of hemodynamic simulations and parallelization of mathematical methods.

Regarding the SDMS, an important issue is the research and development of mechanisms that would enable its deployment in a private cloud as a Service (SaaS) [22]. Additionally, the use of computational clouds becomes relevant due to many other features, such as: scalability; service orientated; flexibility; location transparency; and availability to a large number of users as an appliance. In this context, we should highlight the software Neblina, presented in [9]. Neblina is a middleware developed at LNCC that offers users an interface to cloud resources. Through Neblina a cloud infrastructure, including an application, may be accessed and managed. Typical functionalities include: resources capacity provision, user management, virtualized and physical resources management interface, remote access to the resources and their monitoring.

The SDMS has been integrated into Neblina. This integration makes the cloud environment transparent to SDMS, enabling for instance the activation of its services. In Fig. 13.10 the architecture of the integrated environment is shown: 1) First of all, the simulation software interfaces with Neblina in order to create a virtualized environment; 2) Neblina sets up the environment, creating a virtual cluster to support the simulations data; 3) Neblina sends resource informations to the SDMS to allow the storage of data generated by numerical solvers; 4) the simulation software interfaces with numerical solvers to compute the simulation; 5) numerical solvers requests the SDMS to store the computed results; 6) the SDMS interfaces with SciDB to store the data; 7) once the data has been stored, the simulation software can use the SDMS to answer to analytical queries or retrieve simulation data. 8) finally, the SDMS communicates with SciDB to obtain the requested informations.

An important goal of e-Science was to provide such a computing environment, that scientists have access to additional resources inside their institution or beyond,

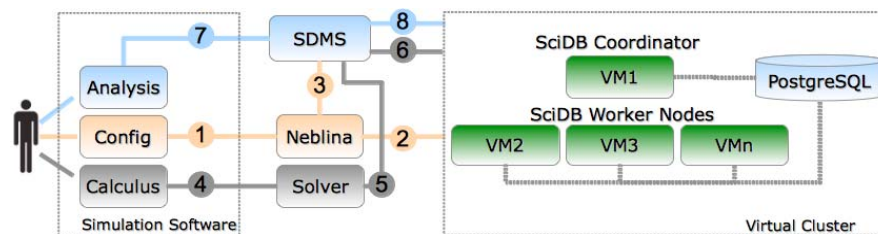


Fig. 13.10 Numerical simulation environment

while at the same time ignoring the complex details of scaling their application in a distributed platform. Cloud can somehow change the usage pattern of HPC environments and its structure. The search for high-performance environments is likely to coexist with increasingly on demand and cost effective scientific oriented environments, more adapted to the needs and skills of its users and more compatible with their applications, even with some loss of performance [14].

a) User specifies desired environment via a form request. A service leverages the internal infrastructure to create the environment in a short period of time.



b) When the requested environment is ready, the user may use a web interface to submit jobs that will execute on the environment.

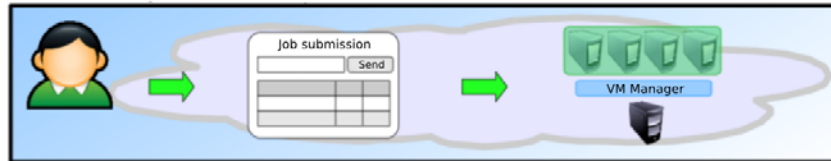


Fig. 13.11 Proposed utilization model via a portal.

13.3.7 Ongoing and Future Projects

Figure 13.11 shows a use case scenario for accessing computational resources via the proposed portal. Firstly, the user must perform a one-time registration with the cyberinfrastructure, so that the user can later login into the system with authorization and authentication. Once the user is logged in, he is able to use the portal to request a virtual workspace. A web form allows the user to specify the characteristics of the desired environment as a set of configuration parameters. Once submitted, the Cloud infrastructure will take a brief amount of time to process the request and inform the user that the environment is ready for use. Depending on the requested environment, the user may have many options to interact with the environment: access the VMs via the traditional command-line; access a VM graphical interface with the desired application already configured and running; or submit jobs to a virtual cluster via a web interface.

13.3.8 Current Physical Resources

The ComCiDis laboratory currently has the computational resources presented at Table 13.1 under its administration, and available for MACC applications. Table 13.2 shows the computational resources that are currently in the installation and configuration phase, and will be available shortly in the future.

Table 13.1 Current physical infrastructure available for MACC applications.

Name	# Nodes	Core/Node	Σ Cores	Mem/Node	Additional Description
ALTIX XE340	30	2 x Xeon 5520 (8)	240	24GB	1TB disk, 2 GbE
Sun Blade x6250	72	2 x Xeon 5540 (8)	576	16GB	
ALTIX 3700 BX2	1	32 Itanium 2	32	64GB	
Supermicro Servers	22	2 x Xeon 5520 (8) + HyperThreading	176	12GB	1TB disk, 2 GbE
ALTIX XE1300	47+1	2 boards x 2 x Xeon 5650 (24) + HT	1140 total	24GB	500GB disk, 2 GbE, 1 controller w/4 GbE
ALTIX XE1300 GPU	12	2 x Xeon 5650 (12) + HyperThreading	144CPU 10752GPU	24GB	500GB disk, 2 GbE, 2 x Tesla Fermi GPU

Table 13.2 Additional planned physical infrastructure available for MACC application.

Name	# Nodes	Core/Node	Σ Cores	Mem/Node	Additional Description
BULL HPC Cluster System	68+1	2 x Xeon 5650 (12) + HyperThreading	828 total	36GB	250GB disk, 2 GbE/Inf., 1 controller, 12TB storage
BULL HPC Cluster Expansion	3 + 1	2 x Xeon 5650 (12) + HyperThreading	48CPU 2688GPU	36GB	250GB disk, 2 GbE/Inf., 2 x Tesla Fermi GPU, 1 controller w/1 GbE/Inf.

13.4 Uses Cases

This section presents use cases of the proposed virtual workspaces system. The figures refer to the client's visualization of the web interface, and show how the main functionalities can be accessed. First of all, the administrator should authorize access to the new client with a user role. After that, the user can login into the web application remotely, using the credentials he received.

The client's home page is shown on Figure 13.12. This page displays the main actions he can perform with the user role. The main panel contains four buttons, hereby described:

- *My Account*: here the client can check and update his profile, including access password.
- *My Machines*: this button shows the *My Machines* page, described below.
- *My Clusters*: here the user can interact with any virtualized cluster he has created.
- *My Files*: this button shows the *My Files* page, described below.

The right panel shows relevant characteristics of the local platform, and also presents an overview of the current state of the Cloud infrastructure, with data collected by the underlying services.

Figure 13.13 shows the view of the *My Files* page, where the user can perform actions related to the file management functionality (Section 13.3.3). The interactions are restricted to a working directory, which is private to the client and is stored inside the Cloud. The main panel now contains a submission form and an interactive table. The form opens a dialog box for uploading a local file to the working directory. After either submitting a file or generating a file output from an application, it is displayed in the table below. The client has the following actions on a table element: a) download the file to the local machine; b) delete the file; c) compress the file or extract a compressed file.

The main view is the *My Machines* page, where the user can interact with the virtual workspaces he has already created. For a machine to become available, he should have configured an appliance based on a previously available VMI, and specify some basic execution parameters that define the VM resource model. On Figure 13.14, the list of available appliances are shown in a table, with each row presenting information about an appliance. The table columns are mostly passive views that describe each appliance, except the last column, named Actions. If an appliance is

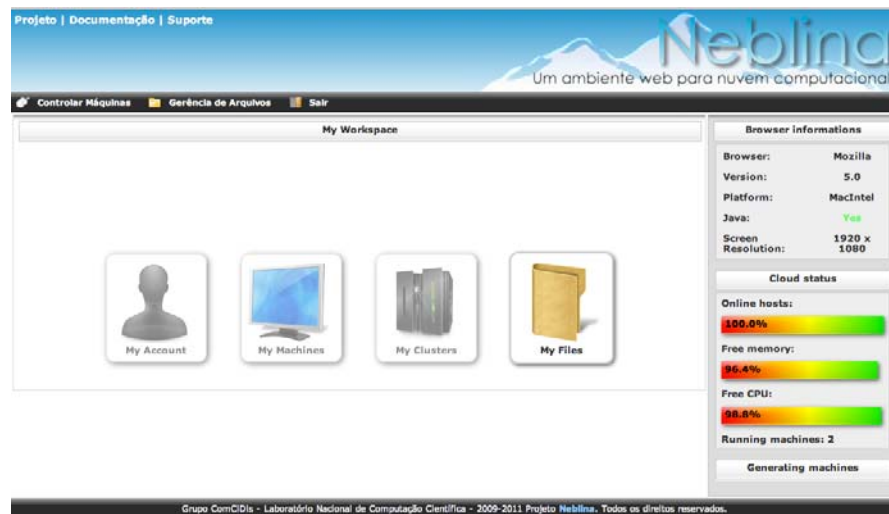
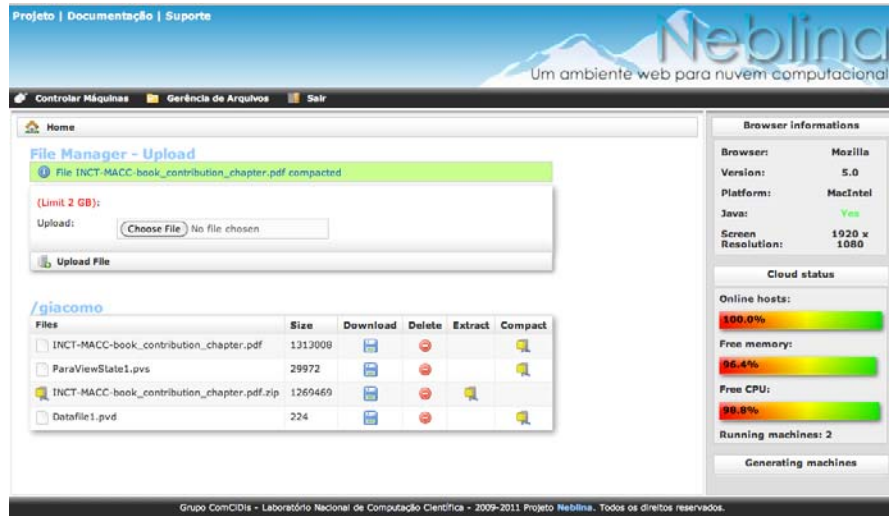
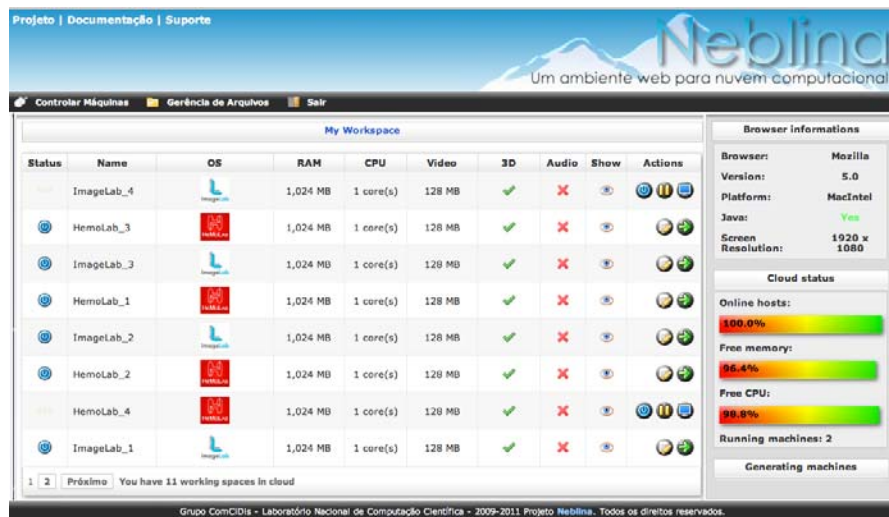


Fig. 13.12 Home page of Neblina after user authentication.

Fig. 13.13 Visualization of the *My Files* page.Fig. 13.14 Visualization of the *My Machines* page.

offline, the action column presents two controllers, to edit and deploy an appliance. If an appliance has already been deployed in the Cloud, the action column presents three controllers, to stop, pause and provide remote access to the appliance. On Figure 13.14, the client has already deployed two appliances, one for the ImageLab application and the other for HemoLab application.

At this moment, the client is able to access the created virtual workspace with a graphical interface via the RDP protocol. Figure 13.15 presents the result of ac-

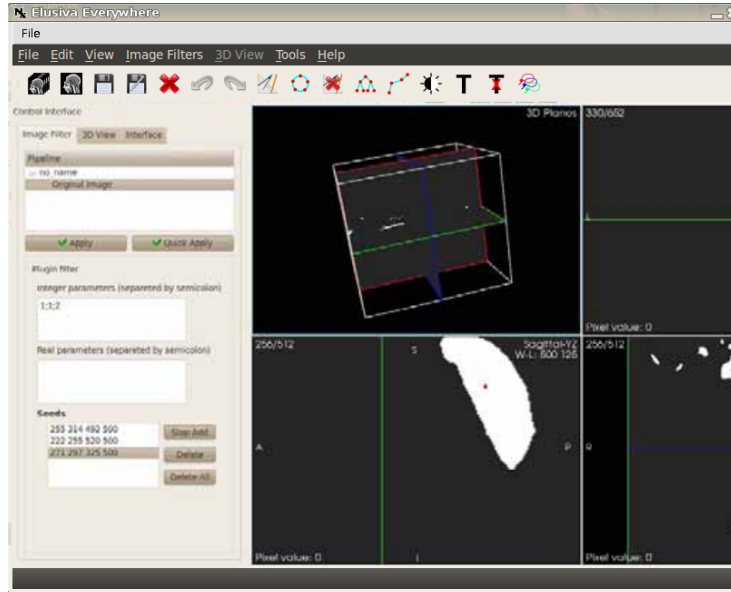


Fig. 13.15 ImageLab running on a virtual workspace.

cessing the ImageLab appliance using the built-in applet. When the applet appears, the application is already running on the VM, and the interface restricts the user's interaction to the application only, effectively embodying a PaaS solution. The application is fully functional, providing all the ImageLab functionalities. However, the application will be in the state that was configured when the appliance was configured. The filesystem is mostly volatile, meaning that the VM state will be lost after the appliance is shutdown. Nonetheless, the client can use the file server feature to save the state of an application.

Figure 13.16 shows the graphical interface of an HemoLab appliance. The application is also fully functional, and can store output on either the VM's volatile filesystem or the shared Cloud storage. Figure 13.16 shows how the storage can be used to load a previous HemoLab session. The content of the client's *My Files* directory is transparently available as `$USER/neblina_files` for any deployed appliance.

13.5 Conclusion

The evolution of computing technologies and specifically, the cloud computing paradigm, created the necessary conditions for the development of middleware applications and tools by the group, concerning the use of distributed cyberenvironments in supporting scientific and medical applications. This was possible by the

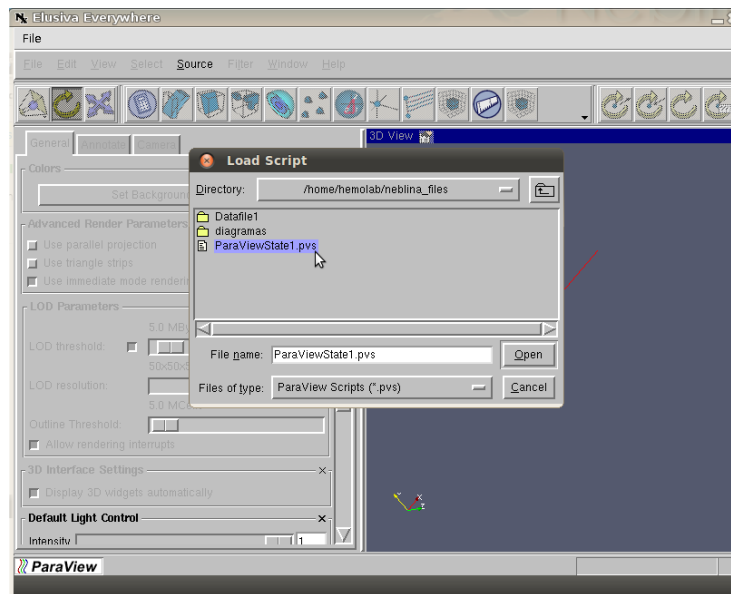


Fig. 13.16 HemoLab running on a virtual workspace.

research, use and work done in conjunction with the INCT-MACC group. The tools developed so far presented in this section had the main purpose of creating environments capable of providing virtualized resources, where the users are able to gain access to applications running over a high performance computing environment.

This type of approach creates the conditions for the research and development of algorithms dedicated to the most diverse study within INCT-MACC mission. We introduced distributed cyberenvironments enabling the execution of scientific applications and in the context of Medical Applications Assisted by Scientific Computing (MACC). These cyberenvironments providing infrastructure and middleware support to execute high performance, resource demanding tasks related to medical applications, such as model solving and scientific visualization. We introduced the concept of virtual workspaces, defining computing environments that can be tailored to execute a set of applications, based on the Cloud Computing paradigm. The virtual workspaces offer solutions in the form of remotely access to applications and platforms, more friendly to researchers and students; contributes to the improvement of human resources capacity in scientific research and in the development of mathematical, simulation models and applications in supporting of many areas especially computer assisted medicine.

This multidisciplinary work and the capacity to prospect new technologies are essential to the continuous enrichment of knowledge, by creating new solutions and the ability to develop new solutions to society.

Acknowledgments

This work was partially supported by the Brazilian agencies CNPq and FAPERJ. The support of these agencies is gratefully acknowledged.

References

- [1] Projeto Virtual Community Grid. Tech. rep. (2008). http://www.rnp.br/_arquivo/gt/2007/Proposta_GT-VCG_Fase2.pdf
- [2] Bientinesi, P., Iakymchuk, R., Napper, J.: HPC on Competitive Cloud Resources. Handbook for Cloud Computing pp. 493–516 (2010)
- [3] Blanco, P.J., Pivello, M.R., Urquiza, S., Feijóo, R.: On the potentialities of 3D-1D coupled models in hemodynamics simulations. *Journal of Biomechanics* **42**(7), 919–930 (2009)
- [4] Chen, W., Cao, J., Li, Z.: Customized Virtual Machines for Software Provisioning in Scientific Clouds. *Networking and Distributed Computing, Intl. Conference on* **0**, 240–243 (2011)
- [5] Costa, R.G., Porto, F., Schulze, B.: Towards analytical data management for numerical simulations. Tech. rep., National Laboratory of Scientific Computing (2012)
- [6] Dongarra, J., White, A., Kennedy, K.: Introduction. In: J. Dongarra, I. Foster, G. Fox, W. Gropp, K. Kennedy, L. Torczon, A. White (eds.) *The Sourcebook of Parallel Computing*. Morgan Kaufmann (2002)
- [7] Ekanayake, J., Fox, G.: High Performance Parallel Computing with Clouds and Cloud technologies. Tech. rep., Indiana University (2009)
- [8] Fernandes, F.J., Mury, A.R., Schulze, B.: Neblina - uma aplicação web para nuvens computacionais. In: XXVI Simp. Bras. de Sistemas Multimídia e Web. Florianópolis, Brazil (2011)
- [9] Fernandes, F.J., Schulze, B., Mury, A.R.: Neblina - espaços virtuais de trabalho para uso em aplicações científicas. In: XXIX Simp. Brasileiro de Redes de Computadores e Sistemas Distribuídos, pp. 965–972. Campo Grande, Brazil (2011)
- [10] Foster, I., Zhao, Y., Raicu, I., Lu, S.: Cloud Computing and Grid Computing 360-Degree Compared. In: *Grid Computing Environments Workshop, 2008. GCE '08*, pp. 1–10 (2008)
- [11] Foster, I.T., Kesselman, C., Tuecke, S.: The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *Int. J. High Perform. Comput. Appl.* **15**(3), 200–222 (2001)
- [12] Gropp, W., Stevens, R., Catlett, C.: Parallel Computer Architectures. In: J. Dongarra, I. Foster, G. Fox, W. Gropp, K. Kennedy, L. Torczon, A. White (eds.) *The Sourcebook of Parallel Computing*. Morgan Kaufmann (2002)
- [13] Laboratório Nacional de Computação Científica: Medicina Assistida por Computação Científica (2012). <http://macc.lncc.br/>
- [14] Lee, C.A.: A perspective on scientific Cloud Computing. In: *Proc. of the 19th ACM Intl. Symposium on High Performance Distributed Computing, HPDC '10*, pp. 451–459. ACM, New York, NY, USA (2010)
- [15] Mell, P., Grance, T.: The NIST definition of Cloud Computing (v15). Tech. rep., National Institute of Standards and Technology (2009)
- [16] NASA: Nasa nebula cloud computing platform (2012). <http://www.nasa.gov/open/plan/nebula.html>
- [17] Parsons, E.W., Sevcik, K.C.: Implementing multiprocessor scheduling disciplines. In: *Proc. of the Job Scheduling Strategies for Parallel Processing, IPPS '97*, pp. 166–192. Springer-Verlag, London, UK (1997)

- [18] Sapuntzakis, C., et al.: Virtual Appliances for Deploying and Maintaining Software. In: LISA '03: Proc. of the 17th USENIX Conference on System administration, pp. 181–194. USENIX Association, Berkeley, CA, USA (2003)
- [19] U.S. Department of Energy: The Magellan Report on Cloud Computing for Science. Tech. rep. (2011). URL {<http://magellan.alcf.anl.gov>}
- [20] Vaquero, L.M., Roderio-Merino, L., Caceres, J., Lindner, M.: A break in the clouds: Towards a Cloud definition. SIGCOMM Comput. Commun. Rev. **39**(1), 50–55 (2009)
- [21] Younge, A.J., Henschel, R., Brown, J.T., von Laszewski, G., Qiu, J., Fox, G.C.: Analysis of Virtualization Technologies for High Performance Computing Environments. In: Proc. of the 4th Intl. Conference on Cloud Computing (CLOUD 2011). IEEE, Washington, DC (2011)
- [22] Zhang, Qi, et al.: Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications **1**(1), 7–18 (2010)



printed by
EDITORA GRÁFICA JORNAL DA CIDADE LTDA.
Rua Antonio Francisco Magalhães, 106 - Bingen
Petrópolis - Rio de Janeiro - CEP 25680-020

Tels: 24 2249.9000 / 2006.0900 / 2006.0901 / 2006.0902
Fax: 24 2006.0909
www.graficajornaldacidade.com.br
E-mail: comercial.graficajc@gmail.com
E-mail: graficajc@gmail.com

End of Printing: July 2012

Scientific Computing Applied to Medicine and Healthcare

Current State and Future Trends at the INCT-MACC the Brazilian National Institute of Science and Technology in Medicine Assisted by Scientific Computing

Raúl A. Feijóo, Artur Ziviani and Pablo J. Blanco (Eds.)

This book offers the current state and the future trends at the Brazilian National Institution of Science and Technology in Medicine Assisted by Scientific Computing (INCT-MACC). It is organized into thirteen chapters providing the current state and future trends in the five R&D and human resource formation fields covered by the INCT-MACC:

- Computational modeling and simulation of the human cardiovascular system and bone trauma and its applications to diagnosis, treatment and planning of several medical procedures.
- Advanced medical image processing including visualization and three dimensional reconstruction of structures of medical relevance and its applications in computational modeling and simulation and novel automated computer methods for image-based diagnosis.
- Collaborative environments of virtual and augmented reality for human resource training and surgical planning.
- Information systems in healthcare, with applications in emergency medical treatment and public health surveillance.
- High performance distributed computing cyber-environments for medical applications.

Each of these topics introduces challenging modeling and scientific computing problems demanding efficient implementation techniques in order to meet the requirements of physicians and healthcare professionals in the use of this emerging and groundbreaking technology in daily practices. This book is addressed to graduate students and researchers in the field of engineering, bioengineering, applied mathematics, computer science and medicine as well as other healthcare domains, wishing to get engaged in the fascinating task of developing novel scientific computing applications for medicine and healthcare activities.



 macc.lncc.br

ISBN 978-85-99961-17-9



9 788599 961179