

Applied Mathematics in Biomedical Sciences and Engineering 2013

Lead Guest Editor: Chang-Hwan Im

Guest Editors: Kiwoon Kwon, Hang Joon Jo, and Pedro Serranho





**Applied Mathematics in Biomedical Sciences
and Engineering 2013**

Journal of Applied Mathematics

**Applied Mathematics in Biomedical
Sciences and Engineering 2013**

Lead Guest Editor: Chang-Hwan Im


Guest Editors: Kiwoon Kwon, Hang Joon Jo, and
Pedro Serranho



Copyright © 2013 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Applied Mathematics." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Theodore E. Simos , Russia






Academic Editors

Saeid Abbasbandy , Iran
Waleed Adel , Egypt
Ali R. Ashrafi , Iran
Tudor Barbu , Romania
Ali Bassam , Mexico
Aliaa Burqan , Jordan
Bruno Carpentieri , The Netherlands
Han H. Choi , Republic of Korea
Carlos Conca , Chile
Meng Fan , China
Elisa Francomano , Italy
Bernard J. Geurts, The Netherlands
Keshlan S. Govinder , South Africa
Ferenc Hartung , Hungary
Ying Hu , France
Dan Huang , China
Lucas Jodar , Spain
Waqar A. Khan , Pakistan
Jong Hae Kim, Republic of Korea
Kannan Krithivasan, India
Kamal Kumar , India
Mirosław Lachowicz, Poland
Peter G. L. Leach, Cyprus
Chong Lin , China
Chongxin Liu , China
Yansheng Liu , China
Oluwole D. Makinde , South Africa
Panayotis Takis Mathiopoulos, Greece
Michael McAleer, Taiwan
Alain Miranville , France
Arvind Kumar Misra , India
Ram N. Mohapatra , USA
Donal O'Regan, Ireland
Turgut Öziş , Turkey
KIRAN PALLE, India
Juan Manuel Peña , Spain
Hector Pomares , Spain
Hui-Shen Shen , China
Fernando Simoes , Portugal
Abdel-Maksoud A. Soliman , Egypt
Qiankun Song , China
Nasser-Eddine Tatar , Saudi Arabia
Mariano Torrisi , Italy

Mehmet Ünver , Turkey
Qing-Wen Wang , China
Frank Werner , Germany
Man Leung Wong, Hong Kong
Zihua Zhang , China
Yun-Bo Zhao , China
Jian G. Zhou , United Kingdom

Contents

Regularised Model Identification Improves Accuracy of Multisensor Systems for Noninvasive Continuous Glucose Monitoring in Diabetes Management

Mattia Zanon , Giovanni Sparacino , Andrea Facchinetti , Mark S. Talary, Andreas Caduff , and Claudio Cobelli 



Research Article (10 pages), Article ID 793869, Volume 2013 (2013)

Analysis of a Model for the Morphological Structure of Renal Arterial Tree: Fractal Structure

Aurora Espinoza-Valdez, Francisco C. Ordaz-Salazar, Edgardo Ugalde, and Ricardo Femat

Research Article (6 pages), Article ID 396486, Volume 2013 (2013)

Hemodynamic Features in Stenosed Coronary Arteries: CFD Analysis Based on Histological Images

Mahsa Dabagh , Wakako Takabe, Payman Jalali , Stephen White, and Hanjoong Jo

Research Article (11 pages), Article ID 715407, Volume 2013 (2013)

Noise Suppression in ECG Signals through Efficient One-Step Wavelet Processing Techniques

E. Castillo, D. P. Morales, A. García, F. Martínez-Martí, L. Parrilla, and A. J. Palma

Research Article (13 pages), Article ID 763903, Volume 2013 (2013)

Effective Preprocessing Procedures Virtually Eliminate Distance-Dependent Motion Artifacts in Resting State fMRI

Hang Joon Jo, Stephen J. Gotts, Richard C. Reynolds, Peter A. Bandettini, Alex Martin, Robert W. Cox, and Ziad S. Saad

Research Article (9 pages), Article ID 935154, Volume 2013 (2013)

Improving CT Image Analysis of Augmented Bone with Raman Spectroscopy

J. Charwat-Pessler, M. Musso, K. Entacher, B. Plank, P. Schuller-Götzburg, S. Tangl, and A. Petutschnigg

Research Article (10 pages), Article ID 271459, Volume 2013 (2013)

A Novel Adaptive Probabilistic Nonlinear Denoising Approach for Enhancing PET Data Sinogram

Musa Alrefaya  and Hichem Sahli


Research Article (14 pages), Article ID 732178, Volume 2013 (2013)

Uniqueness, Born Approximation, and Numerical Methods for Diffuse Optical Tomography

Kiwoon Kwon

Review Article (5 pages), Article ID 824501, Volume 2013 (2013)

Corticomuscular Activity Modeling by Combining Partial Least Squares and Canonical Correlation Analysis

Xun Chen , Aiping Liu, Z. Jane Wang, and Hu Peng

Research Article (11 pages), Article ID 401976, Volume 2013 (2013)

Stress Analysis of Osteoporotic Lumbar Vertebra Using Finite Element Model with Microscaled Beam-Shell Trabecular-Cortical Structure

Yoon Hyuk Kim, Mengying Wu, and Kyungsoo Kim

Research Article (6 pages), Article ID 285165, Volume 2013 (2013)

Quantification of Stretching in the Ventricular Wall and Corpus Callosum and Corticospinal Tracts in Hydrocephalus before and after Ventriculoperitoneal Shunt Operation

Hans von Holst and Xiaogai Li

Research Article (10 pages), Article ID 350359, Volume 2013 (2013)

Mathematical Modelling of Biomechanical Interactions between Backpack and Bearer during Load Carriage

Lei Ren , David Howard , and Richard K. Jones 

Research Article (12 pages), Article ID 349638, Volume 2013 (2013)

Forming Mechanism and Correction of CT Image Artifacts Caused by the Errors of Three System Parameters

Ming Chen and Gang Li

Research Article (7 pages), Article ID 545147, Volume 2013 (2013)

Research Article

Regularised Model Identification Improves Accuracy of Multisensor Systems for Noninvasive Continuous Glucose Monitoring in Diabetes Management

Mattia Zanon,¹ Giovanni Sparacino,¹ Andrea Facchinetti,¹ Mark S. Talary,² Andreas Caduff,² and Claudio Cobelli¹

¹ Department of Information Engineering, University of Padova, Via Gradenigo 6B, 35131 Padova, Italy

² Biovotion AG, Technoparkstrasse 1, 8005 Zurich, Switzerland

Correspondence should be addressed to Claudio Cobelli; cobelli@dei.unipd.it

Received 14 March 2013; Accepted 21 June 2013

Academic Editor: Kiwoon Kwon

Copyright © 2013 Mattia Zanon et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Continuous glucose monitoring (CGM) by suitable portable sensors plays a central role in the treatment of diabetes, a disease currently affecting more than 350 million people worldwide. Noninvasive CGM (NI-CGM), in particular, is appealing for reasons related to patient comfort (no needles are used) but challenging. NI-CGM prototypes exploiting multisensor approaches have been recently proposed to deal with physiological and environmental disturbances. In these prototypes, signals measured noninvasively (e.g., skin impedance, temperature, optical skin properties, etc.) are combined through a static multivariate linear model for estimating glucose levels. In this work, by exploiting a dataset of 45 experimental sessions acquired in diabetic subjects, we show that regularisation-based techniques for the identification of the model, such as the least absolute shrinkage and selection operator (better known as LASSO), Ridge regression, and Elastic-Net regression, improve the accuracy of glucose estimates with respect to techniques, such as partial least squares regression, previously used in the literature. More specifically, the Elastic-Net model (i.e., the model identified using a combination of l_1 and l_2 norms) has the best results, according to the metrics widely accepted in the diabetes community. This model represents an important incremental step toward the development of NI-CGM devices effectively usable by patients.

1. Introduction

Diabetes consists of a malfunction of the glucose-insulin regulatory system leading to the onset of long and short term complications, like retinopathy, neuropathy, and cardiovascular and hearth diseases, due to sustained blood glycaemic levels exceeding the normal range of 70–180 mg/dL [1]. According to the International Diabetes Federation, diabetes is estimated to currently affect more than 350 million people worldwide, and this number is rapidly growing [2]. Not surprisingly, in the last few decades, diabetes has received an increasing attention both for its social and economic implications [3].

Standard therapy of type 1 diabetes is based on a combination of diet, physical activity, and exogenous insulin injections, modulated on the basis of individual patient levels

of glucose concentration in the blood. Accurate and frequent monitoring of glucose concentrations by portable sensor devices plays a crucial role in the diabetes therapy [4]. Self-monitoring blood glucose (SMBG) sensors require the collection of a blood sample by pricking the skin with a lancet device. An external pocket device is then used to analyze the blood and determine glucose concentration for instance by the glucose oxidase principle [5]. These sensors are uncomfortable for the patient and are typically used no more than 3–4 times per day. Such sparseness of sampling does not allow the observation of glucose dynamics and glucose excursions outside the safety range, and dangerous nocturnal hypoglycaemic swings are often not detected. To overcome these problems, portable continuous glucose monitoring (CGM) sensors, measuring blood glucose values every 1 to 5 minutes for up to 7 consecutive days, have been proposed in the

early 2000s and are now of great interest for the tuning and optimization of diabetes therapy [6, 7]. In particular, online applications of CGM sensors include prevention of hyper-/hypoglycaemic events; see for example [8, 9], and closed-loop glucose control aimed at determining optimal automatic insulin infusion in the so-called artificial pancreas systems; see [10–14]. Notably, dealing with these online applications requires facing nontrivial signal processing issues connected to denoising [15], calibration [16], and prediction [17–21] (see [22, 23] for reviews) calling for the development of the so-called “smart” CGM sensor architecture [24].

Most of the CGM sensors requiring the placement of a small needle in the subcutaneous tissue use an enzyme-based glucose-oxidase electrode and thus are invasive, albeit minimally. To limit patients’ discomfort, in the last decade, several noninvasive continuous glucose monitoring (NI-CGM) sensors have been prototyped. Many different physical principles have been considered to pursue NI-CGM, but none of them has clearly outperformed the others so far; see for example [25–27]. One major difficulty with NI-CGM is the fact that environmental (e.g., temperature) and physiological (e.g., sweat, blood oxygenation, etc.) processes act as perturbing factors and often allow blood glucose changes to be tracked only in highly controlled conditions (e.g., during in-hospital studies) [28–30]. To tackle this issue, an approach gaining increasing attention in the last years is the multisensor approach to NI-CGM; that is, instead of focusing on a single physical principle, these devices resort to a combination of technologies. For instance, the GlucoTrack [31] exploits a mix of thermal, acoustic, and electromagnetic technologies and compares the three measurements, assuming they all reflect a glucose-related measurement. A different, yet recent, multi-sensor prototype [32] employs a combination of dielectric spectroscopy (DS) and mid-infrared-based sensors. A further example of multi-sensor device, proposed by Solianis Monitoring AG (Zürich, Switzerland, technology taken over by Biovotion AG, Zurich, Switzerland), embeds sensors of different nature for the biophysical characterisation of skin and underlying tissue in order to track glucose-related and perturbing factors separately [30, 33].

Multisensor approaches to NI-CGM require a model to connect the physical quantities measured by the sensors with blood glucose concentrations. For instance, in the Solianis Monitoring AG prototype (from now on named Solianis for the sake of readability) considered in this work, a multivariate linear regression model is used to combine ~ 150 signals recorded noninvasively for estimating a glucose concentration profile (see Section 3.1 for more details). As described in [33, 34], the linear regression model is identified on a dataset collected in a population of subjects and comprising multi-sensor channels measurements and reference blood glucose (RBG) values collected in parallel by a gold standard technique. Previous work [35] has shown that a regularised identification method, based on l_1 norm (least absolute shrinkage and selection operator—LASSO), provided a glucose profile more accurate than that of other models identified with approaches controlling complexity such as subset selection method or partial least squares (PLS). In the present work, by using the same dataset of 45

experimental sessions used in [35], we demonstrate that the accuracy of glucose estimates can be further improved by considering l_2 norm regularization (Ridge regression) and a combination of l_1 and l_2 norms (Elastic-Net—EN regression), providing a further incremental step toward the development of an NI-CGM effectively usable by diabetic patients.

2. Database

The database consists of 45 experimental sessions recorded from 6 type 1 diabetic subjects, during which plasma glucose was induced to vary according to different predetermined profiles to cause different hypo- and hyperglycaemic excursions. During each session, multi-sensor data and RBG were acquired in parallel, with a time sampling of 20 seconds and 10–15 minutes, respectively. The RBG samples were acquired by means of a SMBG device. The study was performed at the University Hospital Zurich according to the requirements of good clinical practice and was approved by the local ethical commission. More clinically related information can be found in [33].

For the analysis, the database was split into two data subsets of 22 and 23 experimental sessions, respectively. If the first data subset is used for identifying the models with the different techniques, the second is used to test the models over “unseen” multi-sensor data (1→2) and vice versa (2→1). For the sake of space, we will not discuss results of the application of the model on the same dataset used for their identification, and only model test results will be considered.

Multi-sensor data in the identification data subset undergoes a preprocessing step described in detail in [35]. Briefly, each multi-sensor channel is normalized to have zero mean and unitary standard deviation. These values are then used to standardize the same channels in the test data subset to permit simulation of real-time glucose monitoring. Moreover, the first RBG value available at the beginning of each recording session is used to calibrate the estimated glucose profiles by the model setting a baseline adjustment, to allow for a real-time consideration/implementation.

3. Methods

3.1. Glucose Determination by a Multisensor System. In stating the problem, we deliberately make only a brief description of the framework we are working on and we refer the reader to the quoted bibliography for details.

Rather than focusing on a single physical principle, the Solianis multi-sensor device resorts to a combination of technologies, embedded into the substrate in contact with the skin for the biophysical characterisation of the skin and underlying tissue in order to account for confounding factors, which can significantly deteriorate the accuracy of glucose measurements [30, 36–38]. In particular, glucose-related signals are obtained from DS fringing field capacitive electrodes, with different geometrical properties, providing a spectrum of the frequency-dependent complex dielectric properties of skin and underlying tissue including blood, which can be easily parameterised by its magnitude and phase [39]. Environmental and physiological processes that can interfere

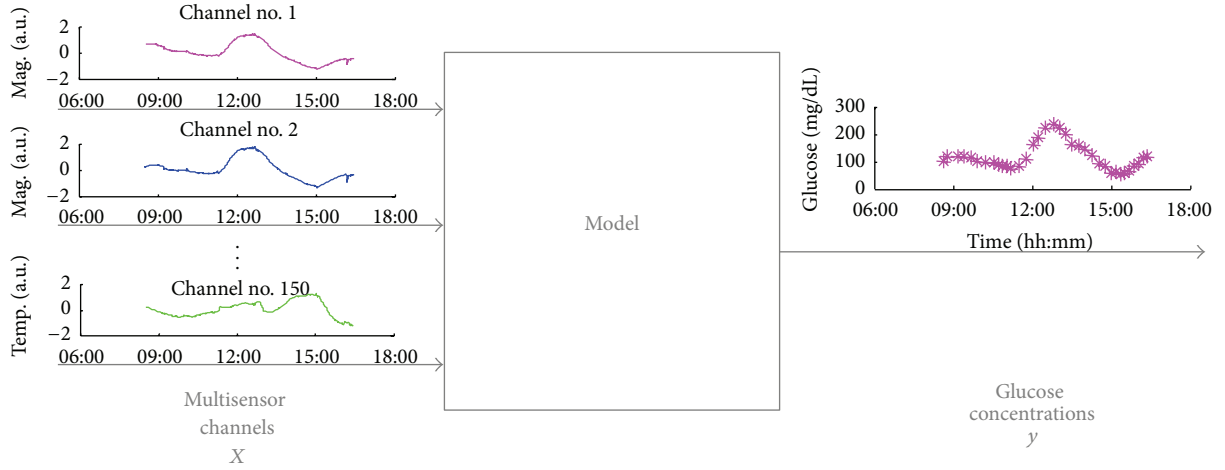


FIGURE 1: A model (center) is needed to properly combine the 150 multisensor channels (left, depicted with lines of different colours) for estimating blood glucose concentration profile (right, magenta stars).

with the measurements of the main glucose-related signals are measured with temperature, optical, humidity, accelerometer, and additional DS sensors incorporated into the same device substrate [30]. About 150 channels are thus provided by the multi-sensor device (Figure 1, left), which have to be properly combined through a mathematical model (Figure 1, middle) for estimating glucose concentrations in the blood (Figure 1, right). Since a mechanistic description comprising the physical principles, linking measured channels with physical/physiological processes, in particular, related to glucose changes, has not yet been fully developed, a “black-box” multivariate linear regression model is used.

Formally, if N is the number of data points available and p is the number of measured channels, the model to identify from the identification dataset is described as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}, \quad (1)$$

where \mathbf{y} represents the $(N \times 1)$ target vector containing RGB values, \mathbf{X} is the $(N \times p)$ matrix collecting the multi-sensor channels, $\boldsymbol{\beta}$ is the $(p \times 1)$ vector containing the coefficients of the linear model, and $\boldsymbol{\nu}$ is the $(N \times 1)$ term representing the error.

The vector $\boldsymbol{\beta}$ in (1) can be identified by minimizing the cost function $F(\boldsymbol{\beta})$ given by the residual sum of squares (RSS):

$$F(\boldsymbol{\beta}) = \text{RSS}(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = \sum_{i=1}^N (y_i - \mathbf{X}_i\boldsymbol{\beta})^2, \quad (2)$$

measuring the distance between data and model predictions. Since this cost function has a quadratic form, a closed-form solution, the so-called ordinary least squares (OLS) estimate, can be obtained. In the case under consideration, OLS suffers from overfitting due to the high dimensionality of the measurement space and to the correlation between subsets of input channels (well visible in the channels showed in Figure 1, left). Recent work [34] showed that further improved results can be obtained by exploiting subset selection techniques. Then, further work [35] investigated the use

of other methods controlling “model complexity,” including PLS (widely used in chemometrics and related fields dealing with spectroscopy data) and the LASSO regularization technique. It has been shown that regularization techniques, in particular, the LASSO, outperform PLS since it sets many coefficients to zero being less sensitive to occasional noise in the multi-sensor channels.

Formally speaking, regularized model identification techniques add a term $L(\boldsymbol{\beta})$ to the cost function of (2), leading to

$$F(\boldsymbol{\beta}) = \text{RSS}(\boldsymbol{\beta}) + L(\boldsymbol{\beta}), \quad (3)$$

where $L(\boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$ reflecting complexity. Depending on the form and on the parameters of $L(\boldsymbol{\beta})$ in (2), the resulting model will present different well-known features, as will be briefly reviewed in the following section.

Thus, the $\boldsymbol{\beta}$ minimizing (3) is identified establishing a trade off between adherence to the data and model complexity, usually by a cross-validation procedure as better discussed next.

3.2. Model Identification by Regularisation Techniques

3.2.1. l_1 Norm: LASSO. In the $L(\boldsymbol{\beta})$ term of (3), the l_1 norm can be used. In the literature, this l_1 norm has been proposed in signal processing (under the name of basis pursuit) [40] and in statistics [41] for its main feature of inducing sparse solutions. Formally, in (3), the l_1 norm leads to define $L(\boldsymbol{\beta})$ as the sum of the absolute values of the coefficients of the model

$$L(\boldsymbol{\beta}) = \lambda \|\boldsymbol{\beta}\|_1 = \lambda \sum_{j=1}^p |\boldsymbol{\beta}_j| \quad (4)$$

multiplied by the scalar nonnegative parameter, hereafter referred to as regularization parameter for the sake of reasoning. Thus, the solution is found by minimizing

$$\hat{\boldsymbol{\beta}}^{\text{LASSO}} = \arg \min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^N (y_i - \mathbf{X}_i\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^p |\boldsymbol{\beta}_j| \right\} \quad (5)$$

and is known as the least absolute shrinkage and selection operator (LASSO), for its properties of shrinking many coefficients to zero selecting only few input variables. In particular, in our application documented in [35], the LASSO was shown to outperform PLS since it sets many coefficients to zero being less sensitive to occasional noise in multi-sensor channels. Equation (5) does not have a closed-form solution because the cost function is not differentiable when some coefficients β_j are zero, and a plethora of methods have been developed to calculate approximate solutions numerically; see [42, 43] for reviews. In particular, an efficient technique for computing the LASSO solution is obtained by modifying the least angle regression algorithm [44]. With this technique, the parameter that has to be fixed represents the number of input variables allowed to enter the model, indicated with j in Section 3.

Remark 1. The regularization parameter weighting the $L(\beta)$ term in (5) is obtained by means of a standard procedure known as K -fold cross-validation [45]. Briefly, the identification dataset is split into K approximately equal parts. Then, the model is identified on $K - 1$ parts and tested over the portion of data not considered for deriving the model, calculating the mean squared error (MSE):

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N_{K-1}} (y_i - \hat{y}_i)^2, \quad (6)$$

where \mathbf{y} and $\hat{\mathbf{y}}$ represent the true and estimated output, respectively, and N_{K-1} represents the number of data points in the $K - 1$ portion of data. This procedure is repeated K times and for a range of values of the parameter that has to be determined. Then, the cross-validation curve is plotted, presenting the MSE as a function of the regularization parameter. Empirical evidence suggests choosing its value in correspondence with a clear drop of the cross-validation curve.

3.2.2. l_2 Norm: Ridge Regression. The l_2 norm involves the penalization of the sum of squares of the coefficients of the model multiplied by a scalar nonnegative parameter:

$$L(\beta) = \lambda \|\beta\|_2^2 = \lambda \sum_{j=1}^p \beta_j^2. \quad (7)$$

The so-called Ridge regression solution, from now on ridge, is thus given by

$$\hat{\beta}^{\text{ridge}} = \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \mathbf{X}_i \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}. \quad (8)$$

The quadratic nature of the cost function in (7) entails a closed-form solution for β dependent on the parameter λ :

$$\hat{\beta}^{\text{ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}_{\text{pxp}})^{-1} \mathbf{X}^T \mathbf{y}. \quad (9)$$

Also, in this case, the regularization parameter λ can be fixed by means of cross-validation. According to [45], as an

indication of the model complexity, the degrees of freedom (df) can be used:

$$\text{df}(\lambda) = \text{tr} \left[\mathbf{X}(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \right] = \sum_{j=1}^p \frac{d_j^2}{d_j^2 + \lambda}, \quad (10)$$

where the d_j 's are the singular values of \mathbf{X} . Thus, to determine the regularization parameter by cross-validation, the MSE is plotted against df in (10).

3.2.3. $l_1 + l_2$ Norms: Elastic Net-Regression. The so-called Elastic-Net regression, from now on EN, resorts to a weighted sum of the two previously described norms, defining $L(\beta)$ as

$$\begin{aligned} L(\beta) &= \lambda (\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2) \\ &= \lambda \left(\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right), \end{aligned} \quad (11)$$

where α weighs the contribution of the two norms while λ sets the amount of regularization [46]. Hence, the EN model parameters are obtained solving the following:

$$\begin{aligned} \hat{\beta}^{\text{EN}} &= \arg \min_{\beta} \left\{ \sum_{i=1}^N (y_i - \mathbf{X}_i \beta)^2 + \lambda \left(\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right) \right\}. \end{aligned} \quad (12)$$

Problem (12) does not have a closed-form solution, and several numerical algorithms have been proposed to compute an approximate solution (some of them are simple adaptations of those developed for the LASSO [46]). The algorithm that has been used in this work is the one based on cyclical coordinate descent originally developed for the LASSO [47] and adapted to problem (12) following [48–50]. The parameters λ and α are determined by cross-validation, following a procedure similar to that of Remark 1.

3.3. Criteria for Model Assessment. The accuracy of estimated glucose profiles in the model test is measured through a set of indexes widely used in the diabetes research community. In particular, we consider the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N ((y_i - \hat{y}_i)^2)}, \quad (13)$$

the mean absolute difference (MAD), indicating how much estimated glucose values are lower or higher than the reference,

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (14)$$

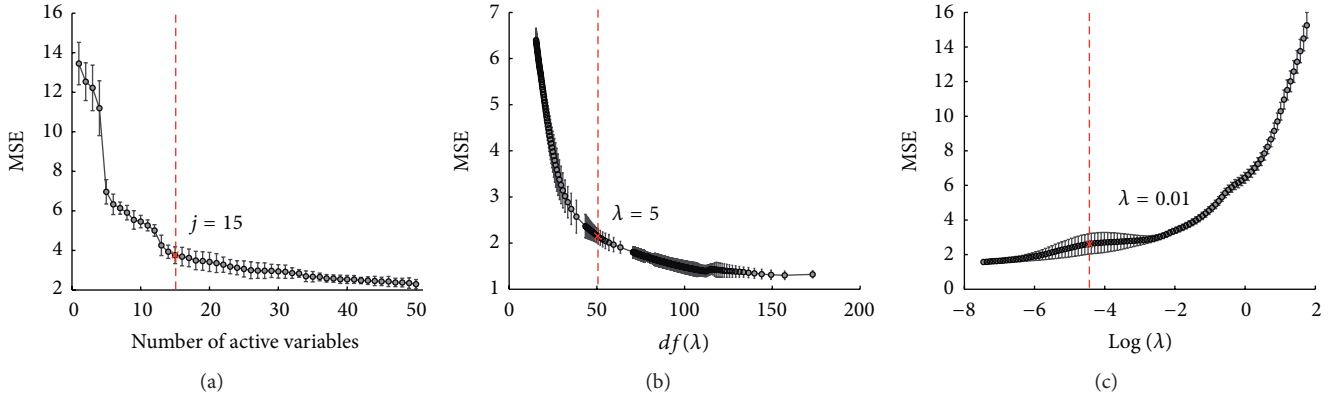


FIGURE 2: 10-fold cross-validation curves for the choice of the “optimal” complexity parameters for LASSO (a), Ridge (b), and EN for $\alpha = 0.4$ (c). The MSE (mean value and one standard deviation) is represented as a function of the model complexity parameter for each method. The red crosses represent the values of the complexity parameter chosen according to the drop in the error curve. The vertical red dashed line is for a better visual identification of the complexity parameters.

and the mean absolute relative difference (MARD), which characterizes the relative errors (in %) of the estimated glucose:

$$\text{MARD} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i}, \quad (15)$$

where y_i and \hat{y}_i , for $i = 1, \dots, N$, are, respectively, the N reference RBG samples and the glucose estimates provided by the models (all the experimental sessions are simultaneously considered). Finally, a popular method used in the diabetes community to judge on the point accuracy of glucose sensors is the error grid analysis (EGA) proposed by Clarke and coworkers [51]. The area where estimated glucose by the model and RBG values are displayed as a scatterplot is broken down into five regions (labeled from A to E). Zone A represents those glucose values within 20% of the RBG values and so on. The most dangerous situations are those where estimated glucose values fall into zones C/D/E because, from a clinical point of view, they will lead to unnecessary or even wrong and potentially dangerous treatments. An evolution of EGA developed for CGM sensors is the continuous EGA (CEGA) that also measures the accuracy of estimated glucose trends by creating a grid which is broken down into regions labeled from A_R to E_R ; see [52] for details.

To give an idea of the values of these indexes for the current state-of-the-art, minimally invasive (needle-based) commercial CGM sensors, a recent study [11] reported mean MARD levels ranging from 11.8 to 20.2% and a percentage of data points in the A+B region ranging from 98.9 to 96.9 under controlled conditions, comparing CGM measurements to gold standard blood glucose sampling.

4. Results

4.1. Regularization Parameter Determination by Cross-Validation. The first step in the analysis is setting the values for λ in (5), (8), and (12) and for α in (12). These were determined by finding where the cross-validation curve presents a clear

drop in slope, as explained in Remark 1. Figure 2 shows the values obtained when identification data subset “part 1” is considered, and the red cross, together with a vertical red dashed line for visualization purpose, highlights the selected “optimal” value (cross-validation plots for identification data subset “part 2” are not showed for the sake of space). Specifically, a K -fold cross-validation strategy has been applied, with $K = 10$, for having a good compromise between bias and variance of the estimated error [45]. The left subplot shows the error curve as a function of the number of latent variables for the LASSO model, indicating a drop of the cross-validation curve around 15. The choice of the regularization parameter for ridge followed a similar approach, with the cross-validation curve (middle panel) presenting a drop when the degree of freedom, defined in (10), approximately equals 50, corresponding to $\lambda = 5$. Similarly for EN, the ending part of the drop in the error curve can be noticed for $\log(\lambda) \cong -4.5$ (Figure 2(c)), corresponding to $\lambda = 0.01$. For EN, different cross-validation curves for different values of α were examined. The most reasonable choice seemed to be that obtained with $\alpha = 0.4$. Indeed, this combination of parameters is the one providing a good trade off between the l_1 and l_2 norms allowing a reasonable compromise for the EN model to be achieved. A value of $\alpha = 0.4$ can suggest that, although it is important to shrink channel weights to zero in order to lower the probability of occasional jumps or spikes entering the model, allowing a grouping effect over correlated predictors is also important for a more robust estimation of glucose profiles.

4.2. Model Test. Figure 3 shows examples of estimated blood glucose concentration profiles (continuous line) versus reference RBG samples (open bullets) for two representative experimental sessions. Visual inspection suggests that the profiles provided by the EN model (bottom panels) outperform, in terms of accuracy, those provided by LASSO (top) and Ridge (middle). The same observation applies for the examples depicted in Figure 4, which represents a more challenging situation because of the presence of wider

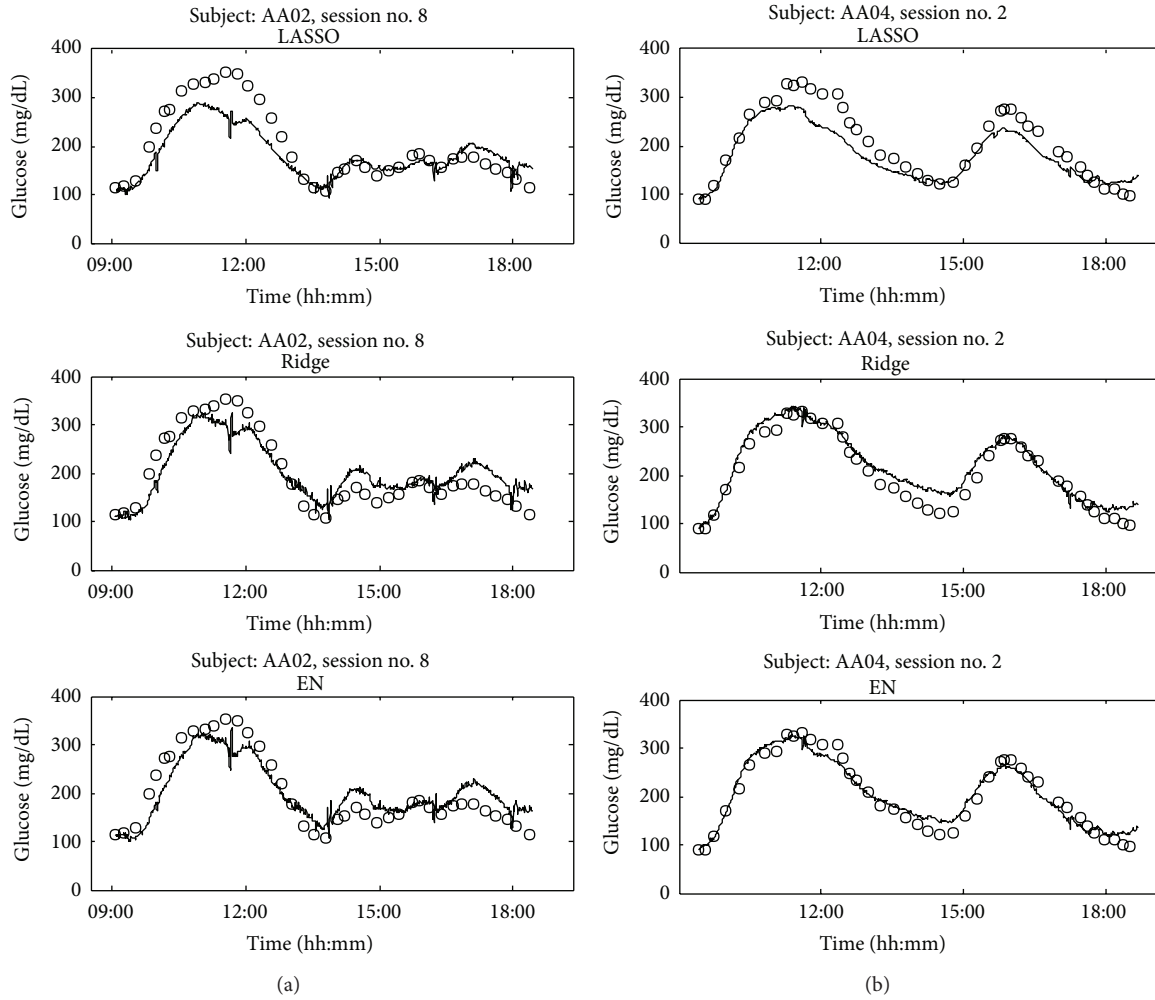


FIGURE 3: Representative recording sessions of Subjects AA02 (a) and AA04 (b). LASSO, Ridge, and EN models test over independent test data subset (continuous lines) versus RBG levels (open bullets). MARD values for the experimental session on the right are of 16.7% (LASSO), 16.9% (Ridge), and 16.5% (EN), while for the experimental session on the left of 12.7% (LASSO), 11.2% (Ridge), and 9.1% (EN).

disturbances, as witnessed by the spikes and jumps affecting the representative multi-sensor raw channels displayed in an additional fourth row of panels. This qualitative observation is supported from the analysis of the MARD values obtained for the representative sessions in Figure 3, that is, 16.7%, 16.9%, and 16.5% (experimental session depicted in the left panels) and 12.7%, 11.2%, and 9.1% for the LASSO, Ridge, and EN models, respectively. When occasional noise is affecting some of the multi-sensor channels, the MARD values slightly worsen, as can be seen for the experiment depicted in Figure 4(a) presenting MARD of 53% for the LASSO, 24.5% for Ridge, and 20.6 for EN. However, the EN model still provides blood glucose estimation profiles more accurately than Ridge and LASSO. This is confirmed, in general, by Table 1, which shows the aggregate results over the test data subset.

By analysing the results in more detail, the LASSO model seems to have a slight advantage in estimating glucose trends (last column of Table 1). The reason is twofold: first,

the regularization performed by the l_1 norm prevents the model coefficients from assuming large values thus predicting glucose profiles that are more flat than the other models, as it happens for example in Figures 3(b) and 4(b); second, channels more sensitive to noise that contain also glucose-related information are considered by Ridge and EN exploiting the effect of the l_2 norm but are less probable to be selected by LASSO, thus yielding smoother estimates. This fact can clearly be seen from Figure 4, where artifacts are present (e.g., in channel no.2) for the session of left data and in channel no.3 for the session of the right data.

Interestingly, the LASSO model seems more robust than the other models to these jumps in the data, preserving the glucose profile with elevated smoothness and reasonably accurate trend. Indeed the l_1 norm shrinks many coefficients to zero allowing an easier interpretation of the results with a reduced number of original variables, representing the strongest effects, considered important for estimating glucose profiles. This is a typical feature of the LASSO to act as

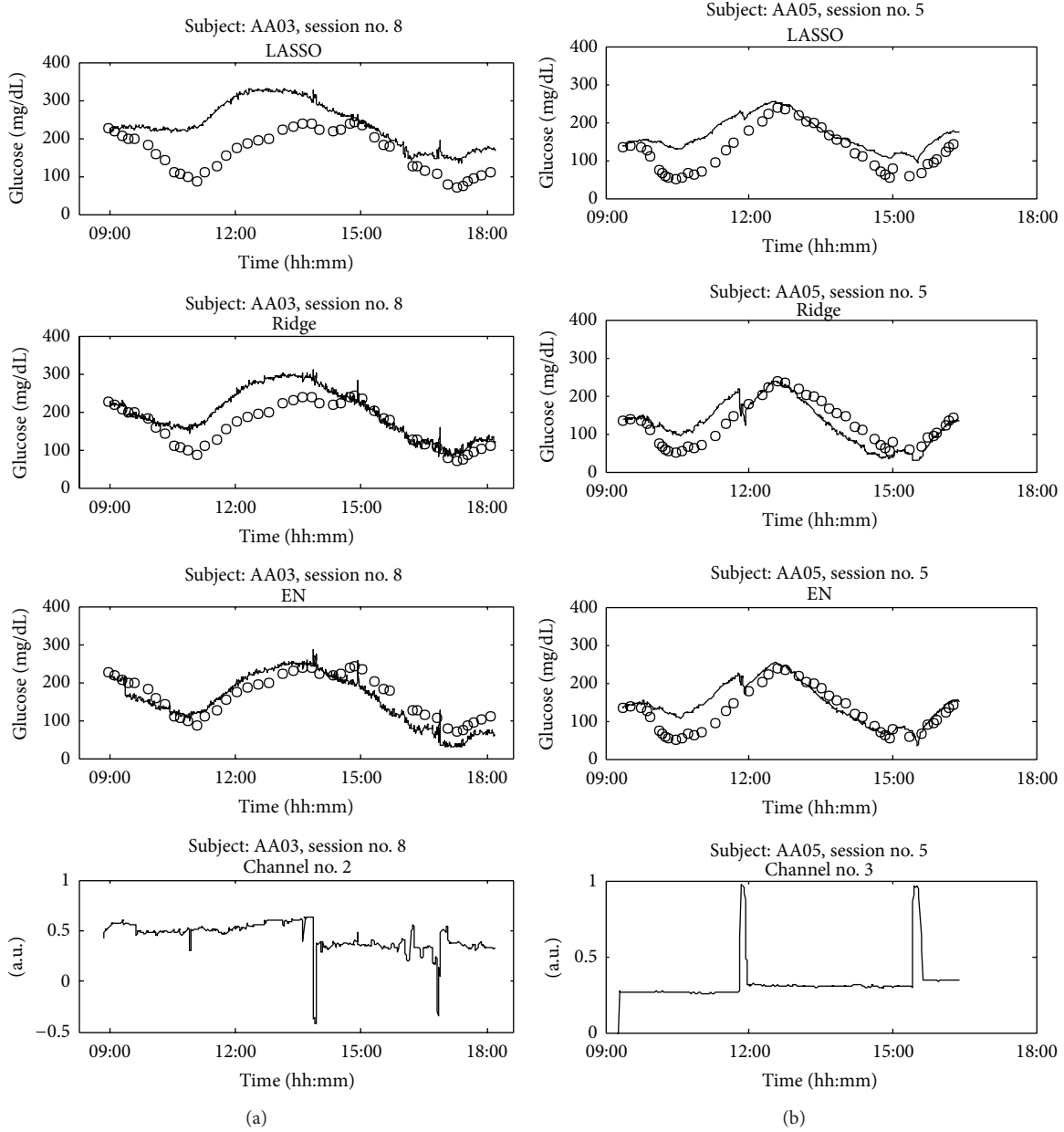


FIGURE 4: Representative recording sessions of Subjects AA03 (a) and AA05 (b). LASSO, Ridge, and EN models test over independent test data subset (continuous lines) versus RBG levels (open bullets). Bottom panels display two representative channels (no.2 and no.3 for subject on the left and on the right, resp.) entering the models, where occasional spikes and jumps are evident. MARD values for the experimental session on the right are of 53% (LASSO), 24.5% (Ridge), and 20.6% (EN), while for the experimental session on the left of 55.7% (LASSO), 34.8% (Ridge), and 34.7% (EN).

a variable selection method. If, from one side, smoother estimates of glucose profiles are obtained with the shrinking properties of the LASSO, sometimes this can lead to biased estimates (see Figure 4(a)).

The Ridge model is identified minimizing the RSS cost function subject to a bound on the l_2 norm of the coefficients. This norm does not have the ability of inducing sparseness on the coefficients of the multivariate linear regression model; thus a parsimonious model is not identified and all the predictors are kept in the model. This might cause the

estimated glucose profiles by the Ridge model to be sensitive to occasional spikes or jumps in the multi-sensor channels, as can be seen in Figure 4(b), where the Ridge model is the one more sensitive among the three. However, estimated glucose profiles by the Ridge model show accuracy indicators slightly better than those of LASSO. This might indicate that (a) channels discharged by the l_1 norm because sensitive to occasional spikes or jumps actually contain useful glucose-related information and (b) that retaining information from all the input channels may help in compensating noisy

TABLE 1: Model test performance when “part 1” of the data set is used for model identification and “part 2” for model test. In brackets is the complexity model parameter chosen by means of cross-validation. Mean and standard deviation (in brackets) over the experimental sessions for root mean squared error (RMSE), mean absolute difference (MAD), mean absolute relative difference (MARD), error grid analysis (EGA (Clarke)) (A + B (A) C/D/E regions whose sum accounts for 100% of data points), continuous error grid analysis (CEGA) ($A_R + B_R$ (A_R) $C_R/D_R/E_R$ regions whose sum accounts for 100% of data points).

	RMSE (mg/dL)	MAD (mg/dL)	MARD (%)	EGA A + B (A) C/D/E	CEGA $A_R + B_R$ (A_R) $C_R/D_R/E_R$
LASSO ($j = 15$)	57.9 (27.1)	48.6 (23.7)	37.8 (20)	89.4 (42.2) 0.9/9.6/0.1	89.2 (62.1) 6.3/2.5/2
Ridge ($\lambda = 5$)	52.3 (22.8)	44.1 (19.2)	35 (17.7)	91 (58.7) 0.1/8.9/0	88 (63) 4.9/4.8/2.3
EN ($\lambda = 0.01$; $\alpha = 0.4$)	51.8 (24.3)	43.9 (20.5)	34.1 (17.2)	92.3 (59.9) 0.1/7.6/0	88.6 (65) 4.9/4.4/2.1

channels thanks to the grouping effect induced by the l_2 norm. Thus, it is reasonable to expect that a combination of the l_1 and l_2 norms could identify a model sharing both properties of sparseness and grouping effect.

Indeed, as mentioned before, the EN model results outperform those of Ridge and LASSO allowing a reasonably accurate estimation of the glucose profile concentrations also when occasional noise is affecting some multi-sensor channels (see Figure 4), presenting lower MARD values than LASSO and Ridge. Thus, EN is the model presenting the best indicators and is only slightly worse than LASSO in accuracy for glucose trends (see CEGA results). Moreover, its clinical accuracy in terms of Clarke error grid, with a percentage of points within the A + B of 92.3 (see Table 1), is substantially close to that of minimally invasive devices, spanning from 98.9 to 96.9 [53].

The good results obtained by the EN model are likely due to the combination of the l_1 and l_2 norms, giving to this model both the advantages of LASSO and Ridge. Indeed, a limitation of the LASSO is that, if there is a group of correlated variables, then it tends to select only one variable from the group and does not care which one is selected, thus lacking the ability of revealing grouping information. On the opposite, the l_2 norm allows all coefficients to enter the model, resulting in more sensitive to noisy channels. Thus, the l_1 norm shrinks channel weights to zero (eliminating multi-sensor channels not useful for predicting glucose), while the l_2 norm encourages a grouping effect (automatically including whole groups into the model once one channel among them is selected). This combination results in indicators outperforming those of the other models and in estimated glucose profiles with a good trade off between sparseness of the model coefficients and robustness due to the grouping effect (see, e.g., Figure 4(a)).

Model test results when data subset “part 2” is used for model identification and data subset “part 1” for model test

are comparable with those in Table 1 (not shown here for the sake of space).

5. Conclusions

In diabetes management, tight monitoring of glycaemic levels by CGM sensors is important for avoiding both long and short term complications related to hyper- and hypoglycaemic excursions. NI-CGM devices are potentially more appealing than the minimally invasive sensors based on needle electrodes, but their development is challenging for several technological and methodological reasons. In the last years, the idea of embedding sensors of different nature within the same device in order to obtain a better biophysical characterisation of the skin and underlying tissues has gained particular attention to develop NI-CGM. In these multi-sensor approaches, a model linking the measured multi-sensor channels to glucose is needed, together with a set of techniques that can be used to identify its parameters. In this work, we investigated the use of regularisation-based methods to identify the linear regression model employed in the multi-sensor device for NI-CGM proposed in [30]. Results on 45 experimental sessions indicate that the EN model generally outperforms the other models: thanks to the combination of the l_1 and l_2 norms, it allows to take the advantage of the LASSO—shrinking many model weights to zero being more robust to possible occasional jumps or spikes occurring on the multi-sensor data—and of the Ridge model—averaging the contribution of correlated channels allowing a more robust estimation of glucose profiles. With respect to the previous sensor literature, where PLS represents the current state of the art (see [34, 54, 55] to mention just a few), we showed that EN can become very useful with multi-sensor data. While retaining information from a group of variables (as PLS does), EN also automatically selects those channels representing the strongest effects, giving more insight into the specific problem at hand.

To conclude, in this work, we showed that further increased point accuracy can be obtained through suitable techniques for the identification of the multivariate model, representing an important incremental step towards the development of NI-CGM devices. While most of the accuracy indices of Table 1 have not yet reached a fully comparable level with those of current enzyme-based needle sensors [53], glucose trends estimated by the considered NI-CGM device exhibit an acceptable accuracy (last column of Table 1). This result could be potentially important in the treatment of diabetes since the glucose trend can be valid adjunctive information to complement standard SMBG devices that measure glucose by fingerstick, for example, helping the diabetic patient in preventing the occurrence of critical events, such as hypoglycaemia, by exploiting the dynamic risk concept recently developed in [56].

References

- [1] M. Brownlee, “The pathobiology of diabetic complications: a unifying mechanism,” *Diabetes*, vol. 54, no. 6, pp. 1615–1625, 2005.

- [2] *IDF Diabetes Atlas*, 5th edition, 2013, <http://www.idf.org/diabetesatlas/>.
- [3] J. E. Shaw, R. A. Scree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030," *Diabetes Research and Clinical Practice*, vol. 87, no. 1, pp. 4–14, 2010.
- [4] B. W. Bode and T. Battelino, "Continuous glucose monitoring in 2010," *International Journal of Clinical Practice, Supplement*, no. 170, pp. 10–15, 2011.
- [5] G. McGarraugh, "The chemistry of commercial continuous glucose monitors," *Diabetes Technology & Therapeutics*, vol. 11, supplement 1, pp. S17–S24, 2009.
- [6] S. Garg, H. Zisser, S. Schwartz et al., "Improvement in glycemic excursions with a transcutaneous, real-time continuous glucose sensor: a randomized controlled trial," *Diabetes Care*, vol. 29, no. 1, pp. 44–50, 2006.
- [7] T. Battelino and J. Bolinder, "Clinical use of real-time continuous glucose monitoring," *Current Diabetes Reviews*, vol. 4, no. 3, pp. 218–222, 2008.
- [8] G. Sparacino, A. Facchinetti, A. Maran, and C. Cobelli, "Continuous glucose monitoring time series and hypo/hyperglycemia prevention: requirements, methods, open problems," *Current Diabetes Reviews*, vol. 4, no. 3, pp. 181–192, 2008.
- [9] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "Reduction of number and duration of hypoglycemic events by glucose prediction methods: a proof-of-concept in silico study," *Diabetes Technology and Therapeutics*, vol. 15, no. 1, pp. 66–77, 2013.
- [10] R. Hovorka, J. M. Allen, D. Elleri et al., "Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial," *The Lancet*, vol. 375, no. 9716, pp. 743–751, 2010.
- [11] S. J. Russell, F. H. El-Khatib, D. M. Nathan, K. L. Magyar, J. Jiang, and E. R. Damiano, "Blood glucose control in type 1 diabetes with a bihormonal bionic endocrine pancreas," *Diabetes Care*, vol. 35, no. 11, pp. 2148–2155, 2012.
- [12] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.
- [13] A. Dauber, L. Corcia, J. Safer, M. S. Agus, S. Einis, and G. M. Steil, "Closed-loop insulin therapy improves glycemic control in children aged <7 years: a randomized controlled trial," *Diabetes Care*, vol. 36, no. 2, pp. 222–227, 2013.
- [14] M. Phillip, T. Battelino, E. Atlas et al., "Nocturnal glucose control with an artificial pancreas at a diabetes camp," *New England Journal of Medicine*, vol. 368, no. 9, pp. 824–833, 2013.
- [15] A. Facchinetti, G. Sparacino, and C. Cobelli, "Online denoising method to handle intraindividual variability of signal-to-noise ratio in continuous glucose monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2664–2671, 2011.
- [16] S. Guerra, A. Facchinetti, G. Sparacino, G. de Nicolao, and C. Cobelli, "Enhancing the accuracy of subcutaneous glucose sensors: a real-time deconvolution-based approach," *IEEE Transaction on Biomedical Engineering*, vol. 59, no. 6, pp. 1658–1669, 2012.
- [17] A. Gani, A. V. Gribok, Y. Lu, W. K. Ward, R. A. Vigersky, and J. Reifman, "Universal glucose models for predicting subcutaneous glucose concentration in humans," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 157–165, 2010.
- [18] M. Eren-Oruklu, A. Cinar, and L. Quinn, "Hypoglycemia prediction with subject-specific recursive time-series models," *Journal of Diabetes Science and Technology*, vol. 4, no. 1, pp. 25–33, 2010.
- [19] A. Facchinetti, G. Sparacino, E. Trifoglio, and C. Cobelli, "A new index to optimally design and compare continuous glucose monitoring glucose prediction algorithms," *Diabetes Technology and Therapeutics*, vol. 13, no. 2, pp. 111–119, 2011.
- [20] V. Naumova, S. V. Pereverzyev, and S. Sivanathan, "A meta-learning approach to the regularized learning-case study: blood glucose prediction," *Neural Networks*, vol. 33, pp. 181–193, 2012.
- [21] C. Zecchin, A. Facchinetti, G. Sparacino, G. de Nicolao, and C. Cobelli, "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 6, pp. 1550–1560, 2012.
- [22] G. Sparacino, A. Facchinetti, and C. Cobelli, "'Smart' continuous glucose monitoring sensors: on-line signal processing issues," *Sensors*, vol. 10, no. 7, pp. 6751–6772, 2010.
- [23] B. W. Bequette, "Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms," *Journal of Diabetes Science and Technology*, vol. 4, no. 2, pp. 404–418, 2010.
- [24] A. Facchinetti, G. Sparacino, S. Guerra et al., "Real-time improvement of continuous glucose monitoring accuracy: the smart sensor concept," *Diabetes Care*, vol. 36, no. 4, pp. 793–800, 2013.
- [25] A. Tura, "Advances in the development of devices for noninvasive glycemia monitoring: who will win the race?" *Nutritional Therapy and Metabolism*, vol. 28, no. 1, pp. 33–39, 2010.
- [26] S. K. Vashist, "Non-invasive glucose monitoring technology in diabetes management: a review," *Analytica Chimica Acta*, vol. 750, pp. 16–27, 2012.
- [27] G. Sparacino, M. Zanon, A. Facchinetti, C. Zecchin, A. Maran, and C. Cobelli, "Italian contributions to the development of continuous glucose monitoring sensors for diabetes management," *Sensors*, vol. 12, no. 10, pp. 13753–13780, 2012.
- [28] K. V. Larin, M. S. Eledrisi, M. Motamedi, and R. O. Esenaliev, "Noninvasive blood glucose monitoring with optical coherence tomography: a pilot study in human subjects," *Diabetes Care*, vol. 25, no. 12, pp. 2263–2267, 2002.
- [29] M. A. Arnold and G. W. Small, "Noninvasive glucose sensing," *Analytical Chemistry*, vol. 77, no. 17, pp. 5429–5439, 2005.
- [30] A. Caduff, M. S. Talary, M. Mueller et al., "Non-invasive glucose monitoring in patients with type 1 diabetes: a multisensor system combining sensors for dielectric and optical characterisation of skin," *Biosensors and Bioelectronics*, vol. 24, no. 9, pp. 2778–2784, 2009.
- [31] I. Harman-Boehm, A. Gal, A. M. Raykhman, E. Naidis, and Y. Mayzel, "Noninvasive glucose monitoring: increasing accuracy by combination of multi-technology and multi-sensors," *Journal of Diabetes Science and Technology*, vol. 4, no. 3, pp. 583–595, 2010.
- [32] C. F. Amaral, M. Brischwein, and B. Wolf, "Multiparameter techniques for non-invasive measurement of blood glucose," *Sensors and Actuators B*, vol. 140, no. 1, pp. 12–16, 2009.
- [33] A. Caduff, M. Mueller, A. Megej et al., "Characteristics of a multisensor system for non invasive glucose monitoring with external validation and prospective evaluation," *Biosensors and Bioelectronics*, vol. 26, no. 9, pp. 3794–3800, 2011.
- [34] M. Mueller, M. S. Talary, L. Falco, O. de Feo, W. A. Stahel, and A. Caduff, "Data processing for noninvasive continuous glucose monitoring with a multisensor device," *Journal of Diabetes Science and Technology*, vol. 5, no. 3, pp. 694–702, 2011.

- [35] M. Zanon, G. Sparacino, A. Facchinetti et al., “Non-invasive continuous glucose monitoring: improved accuracy of point and trend estimates of the multisensor system,” *Medical and Biological Engineering and Computing*, vol. 50, no. 10, pp. 1047–1057, 2012.
- [36] A. Caduff, M. S. Talary, and P. Zakharov, “Cutaneous blood perfusion as a perturbing factor for noninvasive glucose monitoring,” *Diabetes Technology and Therapeutics*, vol. 12, no. 1, pp. 1–9, 2010.
- [37] P. Zakharov, F. Dewarrat, A. Caduff, and M. S. Talary, “The effect of blood content on the optical and dielectric skin properties,” *Physiological Measurement*, vol. 32, no. 1, pp. 131–149, 2011.
- [38] D. Huber, M. Talary, F. Dewarrat, and A. Caduff, “The compensation of perturbing temperature fluctuation in glucose monitoring technologies based on impedance spectroscopy,” *Medical and Biological Engineering and Computing*, vol. 45, no. 9, pp. 863–876, 2007.
- [39] P. Åberg, *Skin cancer as seen by electrical impedance [Ph.D. thesis]*, Division of Medical Engineering, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden, 2004.
- [40] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [41] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.
- [42] M. Schmidt, G. Fung, and R. Rosales, “Optimization methods for ℓ_1 -regularization,” Tech. Rep. TR-2009-19, University of British Columbia, 2009.
- [43] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.
- [44] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [45] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2009.
- [46] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society B*, vol. 67, no. 2, pp. 301–320, 2005.
- [47] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, “Pathwise coordinate optimization,” *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.
- [48] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [49] A. J. van der Kooij, “Prediction accuracy and stability of regression with optimal scaling transformations,” Tech. Rep., Department of Data Theory, Leiden University, 2007.
- [50] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [51] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, “Evaluating clinical accuracy of systems for self-monitoring of blood glucose,” *Diabetes Care*, vol. 10, no. 5, pp. 622–628, 1987.
- [52] W. L. Clarke, S. Anderson, and B. Kovatchev, “Evaluating clinical accuracy of continuous glucose monitoring systems: continuous glucose—error grid analysis (CG-EGA),” *Current Diabetes Reviews*, vol. 4, no. 3, pp. 193–199, 2008.
- [53] E. R. Damiano, F. H. El-Khatib, H. Zheng, D. M. Nathan, and S. J. Russell, “A comparative effectiveness analysis of three continuous glucose monitors,” *Diabetes Care*, vol. 36, no. 2, pp. 251–259, 2013.
- [54] A. M. K. Enejder, T. G. Scecina, J. Oh et al., “Raman spectroscopy for noninvasive glucose measurements,” *Journal of Biomedical Optics*, vol. 10, no. 3, Article ID 031114, 2005.
- [55] M. A. Arnold, L. Liu, and J. T. Olesberg, “Selectivity assessment of noninvasive glucose measurements based on analysis of multivariate calibration vectors,” *Journal of Diabetes Science and Technology*, vol. 1, no. 4, pp. 454–462, 2007.
- [56] S. Guerra, G. Sparacino, A. Facchinetti, M. Schiavon, C. Dalla Man, and C. Cobelli, “A dynamic risk measure from continuous glucose monitoring data,” *Diabetes Technology and Therapeutics*, vol. 13, no. 8, pp. 843–852, 2011.

Research Article

Analysis of a Model for the Morphological Structure of Renal Arterial Tree: Fractal Structure

**Aurora Espinoza-Valdez,¹ Francisco C. Ordaz-Salazar,²
Edgardo Ugalde,³ and Ricardo Femat⁴**

¹ Departamento de Ciencias Computacionales, CUCEI, Universidad de Guadalajara, Avenida Revolución 1500, CP 44430, Guadalajara, JAL, Mexico

² Universidad Politécnica de San Luis Potosí, Urbano Villalón 500, CP 78363, La Ladrillera, SLP, Mexico

³ Instituto de Física UASLP, Universidad Autónoma de San Luis Potosí, Avenida Manuel Nava No. 6., Zona Universitaria, CP 78200, SLP, Mexico

⁴ Laboratorio para Biodinámica y Sistemas Alineales, División de Matemáticas Aplicadas, IPICYT, Apartado Postal 3-90, CP 78231, Tangamanga, SLP, Mexico

Correspondence should be addressed to Aurora Espinoza-Valdez; aurora.espinoza@cucei.udg.mx

Received 7 March 2013; Accepted 12 June 2013

Academic Editor: Kiwoon Kwon

Copyright © 2013 Aurora Espinoza-Valdez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the fields of applied mathematics is related to model analysis. Biomedical systems are suitable candidates for this field because of their importance in life sciences including therapeutics. Here we deal with the analysis of a model recently proposed by Espinoza-Valdez et al. (2010) for the kidney vasculature developed via angiogenesis. The graph theory allows one to model quantitatively a vascular arterial tree of the kidney in sense that (1) the vertex represents a vessels bifurcation, whereas (2) each edge stands for a vessel including physiological parameters. The analytical model is based on the two processes of sprouting and splitting angiogeneses, the concentration of the vascular endothelial growth factor (VEGF), and the experimental data measurements of the rat kidneys. The fractal dimension depends on the probability of sprouting angiogenesis in the development of the arterial vascular tree of the kidney, that is, of the distribution of blood vessels in the morphology generated by the analytical model. The fractal dimension might determine whether a suitable renal vascular structure is capable of performing physiological functions under appropriate conditions. The analysis can describe the complex structures of the development vasculature in kidney.

1. Introduction

The arterial structure of organs has been the subject of many studies [1–5]. However, not all systems have similar functions; for the organs the purpose of arterial structure is to provide the blood required in the metabolic process and other specific functions. The arterial structure is highly nonuniform because it is determined by reasons of anatomy and local flow requirements [4]. The kidney is one of the most complicated organs in terms of structure and physiology because it is highly vascularized and constitutes the main organ for maintaining chemical balance in blood [6]. The kidney consists of three trees: arterial, venous, and ureter. The arterial

vascular tree of the kidney is structured by the renal artery branches into interlobar arteries, arcuate arteries, and interlobular arteries, which are formed by bifurcation [6]. The development of the arterial vascular tree of the kidney can be formed mainly by angiogenesis [1]. The process of angiogenesis is the formation of new blood vessels from preexisting vessels and consists of two different processes: sprouting and splitting. Sprouting refers to the case in which the new branch literally sprouts to some existing branch. In splitting a branch is divided into two new branches [7]. The factors involved at the vascular development are VEGF, renin-expression, ephrins A and B, platelet-derived growth factor-B (PDGF-B), Ets family (as Ets-1 and TEL), and angiotensins

1-2. The VEGF is an essential regulatory factor for both processes of angiogenesis in the development of renal arterial vascularization [1]. Then, we only consider VEGF, because the other factors increase the complexity of the analytical model.

Zamir and Phipps in [8] studied the morphological characteristics of the rat kidney and found that the branching rules of these vessels are determined by considerations of the angiological function. Subsequently, Zamir generated a model based on L-systems (Lindenmayer System) and incorporated some physiological laws, that is, random parameters [4]. Zamir's results suggest that the arterial structure of the tree is determined mainly by the growth rules of arterial tree branching [4]. The structural morphologic reconstruction of renal vasculature from microcomputed tomography (micro-CT) images was presented by Nordsletten et. al. The arterial and venous trees of the rat kidney were generated numerically using micro-TC [3]. These morphological data provide a statistical basis from which renal vascular topology can be generated [5]. Graph theory generates branching tree structures incorporating the physiological laws of the renal artery branching through the process of angiogenesis [5]. Many pathological conditions, such as cancers, arteriovenous malformations, and diabetes, induce changes to vessel's morphology or spatial organization. The fractal analysis has been applied to a large variety of healthy or pathological vascular networks [9–14]. Sabo et al. concluded in [15] that the microvessel fractal dimension as a marker of tumor microvascular complexity might provide important prognostic information as well as shed light on the complex interactions between tumor angiogenesis and growth. Moreover, Cross et al. analyzed the fractal dimension of renal angiograms: a normal kidney, a congenitally dysplastic kidney, a kidney with renal artery stenosis, and a kidney with recurrent thromboembolism lesions (see Figure 3 in [16]). The analysis of fractal dimension may allow for characterization of the vascular morphology.

In this paper we extend the results of [5] by analysing mathematically the structure of the arterial vascular tree of the kidney and by incorporating physiological parameters generated from a computational model. In the implementation of the proposed model the branches are represented by the edges of the graph and the ramification point is represented by the nodes [5]. We expect that this representation implemented in computer methodology will help to understand the mechanism involved in biological systems, specifically the development of the renal arterial bifurcation. The analysis provides the mean of the length in each level, and the average width of the arterial vascular tree of the rat kidney is in the range of what is observed in previous experimental and computational models. The fractal dimension depends on the sprouting angiogenesis that appears in the development of the arterial vascular tree of the kidney. The complex structures of the renal arterial tree are consistent with previous fractal analysis of vascular structures [9, 11–14]. These study opens the possibility of a new taxonomy for normal kidneys and for the pathological injuries related to the vascular morphology of the kidneys.

2. Theoretical Basis

The representation of the renal arterial vascularization in the computational model is as follows: (1) the arterial vascular tree of the kidney is defined by labeled and oriented graphs; we can identify two objects: the branches (edges) and the branching point (nodes). That is, the edge represents a blood vessel and the vertex is where the angiological stimulus appears. (2) The length and diameter of the new branches are smaller than their parent branches [5, 6].

The arterial vascular tree of the kidney down to the interlobular arteries can be structured as follows: renal artery, level 0: interlobar arteries, levels 1-2: arcuate arteries, levels 3-4: interlobular arteries, and levels 5–9. We based in the experiments of Tomanek [3] for the morphological structure of the renal arterial tree. In the model, we develop the arterial vascular tree of the rat kidney. Therefore, the model based on the graph theory for the arterial tree is a binary tree of ten levels and has a total of vessels equal to 1023. The physiological parameters, and rules are based on parent vessel branch level, the parameters, and the morphological structure of their child branches, and this process deploys the structure of the model. The parameters included in the edges of the binary tree model are s , C_{gf} , l , d , and θ as follows (for details see [5]).

- (i) *The Type of Angiogenesis: Sprouting or Splitting (s)*. Angiogenesis is identified by the variable s which can take the values in the set $\{a_b, a_p\}$, where a_b denotes sprouting angiogenesis and a_p denotes splitting angiogenesis. Both sprouting and splitting angiogenesis generate branching tree structures, which can be represented using graph theory. It is conjectured that the sprouting and splitting processes occur with different probability in the renal arterial tree for the morphology of the kidney. In the proposed computational model this assumption is controlled probabilistically. Sprouting angiogenesis is more probably than splitting angiogenesis because it requires only reorganization of existing endothelial cells (not migration), that is, having less energy requirements [1].
- (ii) *Concentration of the VEGF in the Vessel (C_{gf})*. Both processes, sprouting and splitting angiogenesis, depend on the regulation of C_{gf} in the pre-existing vessel [1]. While in sprouting angiogenesis it is known that endothelial cells promotes the differentiation, migration, proliferation, and assembly, the effect on splitting angiogenesis is still in research. In the model C_{gf} is generated randomly in a uniform way within the range $[0, 35]$ ng/mL reported in [5, 17].
- (iii) *Length of the Vessel (l)*. Here, we focus on determining the value of the lengths in the arterial bifurcations depending of the concentration of the VEGF. Although the effect of C_{gf} over length for the splitting angiogenesis is unknown, in the sprouting angiogenesis we consider that the length of the new vessels depends of the C_{gf} . We obtain the value of the dimensionless length $l_e : [C_{gf}, \overline{C_{gf}}] \rightarrow [l_e, \overline{l_e}]$ from

experimental data; the function which can be approximated for VEGF₁₂₁ is [5] $l_e = 0.00878C_{gf}^3 - 0.51326C_{gf}^2 + 8.52128C_{gf} + 81.12064$. The l_e is applied in this model; the length of the vessel l depends on l_e and is compared with the length ranges reported in the literature according to the level [3, 5]. The l formed by splitting is defined by a contraction factor. In the case that l is out of these ranges, the process is repeated finite number of times. If l is still out of range, then it is adjusted to the length of its parent branch.

- (iv) *Diameter of the Vessel (d)*. In this work, we do not discuss the value of the diameters in the blood vessels. This is due to the fact that we do not have sufficient information of C_{gf} in d for sprouting and splitting angiogeneses. Then, in sprouting angiogenesis, the diameter of the sprout is adjusted to a factor of the diameter of its parent branch. In splitting angiogenesis the diameter of every new vessel is adjusted to one-half of the diameter of its parent branch.
- (v) *Angle of Bifurcation (θ)*. In sprouting angiogenesis the angle of bifurcation is adjusted within the range $\pm[60^\circ, 80^\circ]$ with respect to its parent branch [1, 18]. In splitting angiogenesis, one branch is adjusted to $+32.5^\circ$ and the other branch to -32.5° [18].

The physiological parameters are included in the label of the edge and have the form $(s, C_{gf}, l, d, \theta)$. The arterial vascular tree of the kidney has vertices with oriented edges in such a way that from each vertex one edge arrives and two edges leave (the orientation symbolizes the circulation of blood flow in arteries of the kidney). The renal arterial vascular tree of the kidney has 10 levels, because we based in the experiments of Tomanek. Therefore, the arterial vascular tree of the kidney has 2^{10} nodes. The set of integers to name every node is $[0, \dots, 2^{10} - 1]$.

An algorithm generates step by step the physiological characteristics of every branch and saves it as labeled edges with the format $\{s, C_{gf}, l, d, \theta\}$. Just as the tree is generated, the other structure is constructed in which we save the position of every node in \mathbb{R}^2 . The algorithm calculates the position of every node in function of the physiological characteristics of the branches and the position of nodes of its parent branch. The physiological parameters of the root (the initial condition) are determined: $s = a_b$ with probability $P_{ab} = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $s = a_p$, where $P_{ap} = 1 - P_{ab}$, $C_{gf} \in [0, 35]$ ng/mL, $l = 5$, $d = 1$, $\theta = 0$, and they continue in numerical order with the other branches using the information of parent branches. Some characteristics such as s and C_{gf} are generated probabilistically in every step, and the others depend on these parameters, their parent branch, and the rules that were enumerated in the previous section.

3. Main Results

3.1. Length and Width. The length of each level into the kidney depends on a_b and a_p ; that is, the length for each level

j denoted by l_j in the arterial vascular tree of the kidney is analytically determined as follows:

$$l_j = l_{00} (P_{ab}\lambda_b + P_{ap}\lambda_p)^j, \quad (1)$$

where the parameters are defined as in the previous section. Then, the mean length for each segment j is

$$\begin{aligned} l_0 &= 5 \text{ mm}, & l_1 &= 3.095 \text{ mm}, \\ l_2 &= 1.919 \text{ mm}, & l_3 &= 1.191 \text{ mm}, \\ l_4 &= 0.741 \text{ mm}, & l_5 &= 0.462 \text{ mm}, \\ l_6 &= 0.287 \text{ mm}, & l_7 &= 0.179 \text{ mm}, \\ l_8 &= 0.112 \text{ mm}, & l_9 &= 0.070 \text{ mm}. \end{aligned} \quad (2)$$

These data are within the range of rat trees reported in [3]; we compare the results with Nordsletten from depth $j = 0$ to $j = 9$.

Moreover, the analytical average width denoted by A_{G_R} is the root length (renal artery, $j = 0$) to the leaves of the tree (interlobular arteries, $j = 9$), and then

$$A_{G_R} = l_{00} \sum_{j=0}^9 (P_{ab}\lambda_b + P_{ap}\lambda_p)^j, \quad (3)$$

where $0 \leq j \leq 9$; $l_{00} = 5$ mm is the length of the renal artery [5]; P_{ab} is defined in the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $P_{ap} = 1 - P_{ab}$; $\lambda_b = 0.5$ is the average of the contraction factor for a_b ; and $\lambda_p = 0.67$ is the contraction factor for a_p . Then,

$$\begin{aligned} P_{ab} = 0.1, & & P_{ap} = 0.9, & & A_{G_R} = 14.20 \text{ mm}, \\ P_{ab} = 0.2, & & P_{ap} = 0.8, & & A_{G_R} = 13.59 \text{ mm}, \\ P_{ab} = 0.3, & & P_{ap} = 0.7, & & A_{G_R} = 13.01 \text{ mm}, \\ P_{ab} = 0.4, & & P_{ap} = 0.6, & & A_{G_R} = 12.48 \text{ mm}, \\ P_{ab} = 0.5, & & P_{ap} = 0.5, & & A_{G_R} = 11.99 \text{ mm}. \end{aligned} \quad (4)$$

These data are within the range of rat trees reported in [3]; while we consider 10 levels [1, 5], Nordsletten considers 11 levels.

However, if $\lambda_b \in [0.3, 0.7]$, the average with respect to λ_b is as follows:

$$\begin{aligned} E_b \left((P_{ab}\lambda_b + P_{ap}\lambda_p)^j \right) & \\ &= \frac{5}{2P_{ab}} \int_{.3}^{.7} (P_{ab}\lambda_b + P_{ap}\lambda_p)^j P_{ab} d\lambda_b \\ &= \frac{5}{2P_{ab}(j+1)} \left[(.7P_{ab} + P_{ap}\lambda_p)^{j+1} - (.3P_{ab} + P_{ap}\lambda_p)^{j+1} \right]. \end{aligned} \quad (5)$$

Let $a = (.7P_{ab} + P_{ap}\lambda_p)^{j+1}$ and $b = (.3P_{ab} + P_{ap}\lambda_p)^{j+1}$, and then

$$A_{G_R} = \frac{5l_{00}}{2P_{ab}} \sum_{j=0}^9 \left(\frac{a^{j+1}}{j+1} - \frac{b^{j+1}}{j+1} \right) \tag{6}$$

$$\approx \frac{5l_{00}}{2P_{ab}} \sum_{j=0}^{\infty} \left(\frac{a^{j+1}}{j+1} - \frac{b^{j+1}}{j+1} \right),$$

where

$$\sum_{j=0}^{\infty} \frac{a^{j+1}}{j+1} = \sum_{j=0}^{\infty} \left(\int_0^a x^j dx \right)$$

$$= \int_0^a \left(\sum_{j=0}^{\infty} x^j \right) dx \quad (\text{by uniform convergence})$$

$$= \int_0^a \frac{dx}{1-x} = -\ln(1-x) \Big|_0^a = \ln\left(\frac{1}{1-a}\right). \tag{7}$$

Similarly, $\sum_{j=0}^{\infty} (b^{j+1}/j+1) = \ln(1/1-b)$, and then

$$A_{G_R} = \frac{5l_{00}}{2P_{ab}} \ln\left(\frac{1-b}{1-a}\right) \tag{8}$$

$$= \frac{5l_{00}}{2P_{ab}} \ln\left(\frac{1 - (.3P_{ab} + P_{ap}\lambda_p)}{1 - (.7P_{ab} + P_{ap}\lambda_p)}\right) \pm .7^{11},$$

where $\pm .7^{11}$ mm is the error.

Substituting the values l_{00} , P_{ab} , P_{ap} , and $\lambda_p = 0.67$,

$P_{ab} = 0.1,$	$P_{ap} = 0.9,$	$A_{G_R} = 14.42 \pm .7^{11}$ mm,
$P_{ab} = 0.2,$	$P_{ap} = 0.8,$	$A_{G_R} = 13.79 \pm .7^{11}$ mm,
$P_{ab} = 0.3,$	$P_{ap} = 0.7,$	$A_{G_R} = 13.23 \pm .7^{11}$ mm, (9)
$P_{ab} = 0.4,$	$P_{ap} = 0.6,$	$A_{G_R} = 12.73 \pm .7^{11}$ mm,
$P_{ab} = 0.5,$	$P_{ap} = 0.5,$	$A_{G_R} = 12.28 \pm .7^{11}$ mm.

We generate 5000 trees of renal arterial vasculature for different probabilities in a_b and a_p . In Figure 1 the mean \pm SD (standard deviation) width of the kidney has an intersection with these data and previous result where the mean width is 17.454 ± 6.165 mm [3].

Therefore, the A_{G_R} depends on the process of angiogenesis; that is, A_{G_R} is inversely proportional to the probability of the a_b .

3.2. Fractal Dimension of Kidney Vascular Tree. The fractal dimension (D) quantifies through dimension the ability of an object to occupy a space. Different methods exist and are based on the different definitions of fractal dimension [9, 11–14, 19].

The analytical fractal dimension can be derived using the following result [19]. Let $\{\mathbb{R}^m; w_1, w_2, \dots, w_N\}$ be a hyperbolic

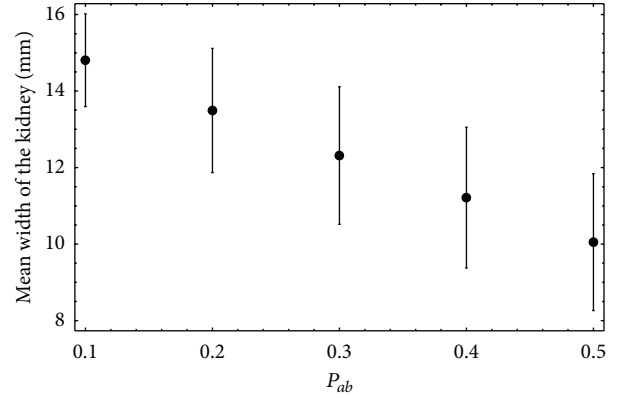


FIGURE 1: The mean \pm SD of width of 5000 trees generated for 5 different probabilities of a_b and a_p .

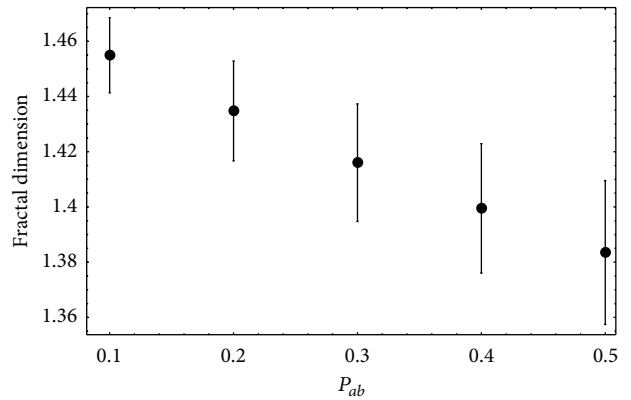


FIGURE 2: The mean \pm SD of fractal dimension for different probabilities of a_b , where a_p has the probability $P_{ap} = 1 - P_{ab}$.

iterative function system (IFS), and let A denote its attractor. Suppose that w_n is a similitude of scaling factor s_n for each $n \in \{1, 2, 3, \dots, N\}$. If the IFS is totally disconnected or just touching, the attractor has fractal dimension $D(A)$, which is given by the unique solution of

$$\sum_{n=1}^N |s_n|^{D(A)} = 1, \quad D(A) \in [0, m]. \tag{10}$$

If the IFS is overlapping, then $\bar{D} \geq D(A)$, where \bar{D} is the solution of

$$\sum_{n=1}^N |s_n|^{\bar{D}} = 1, \quad \bar{D} \in [0, \infty). \tag{11}$$

The morphology of arterial vascular tree of the kidney, produced by an IFS, is composed of n applications of contraction with a factor s_n . Thus, from the subsection of length and width, the contraction factor for the splitting angiogenesis is $\lambda_p = 0.67$ while for sprouting angiogenesis the average

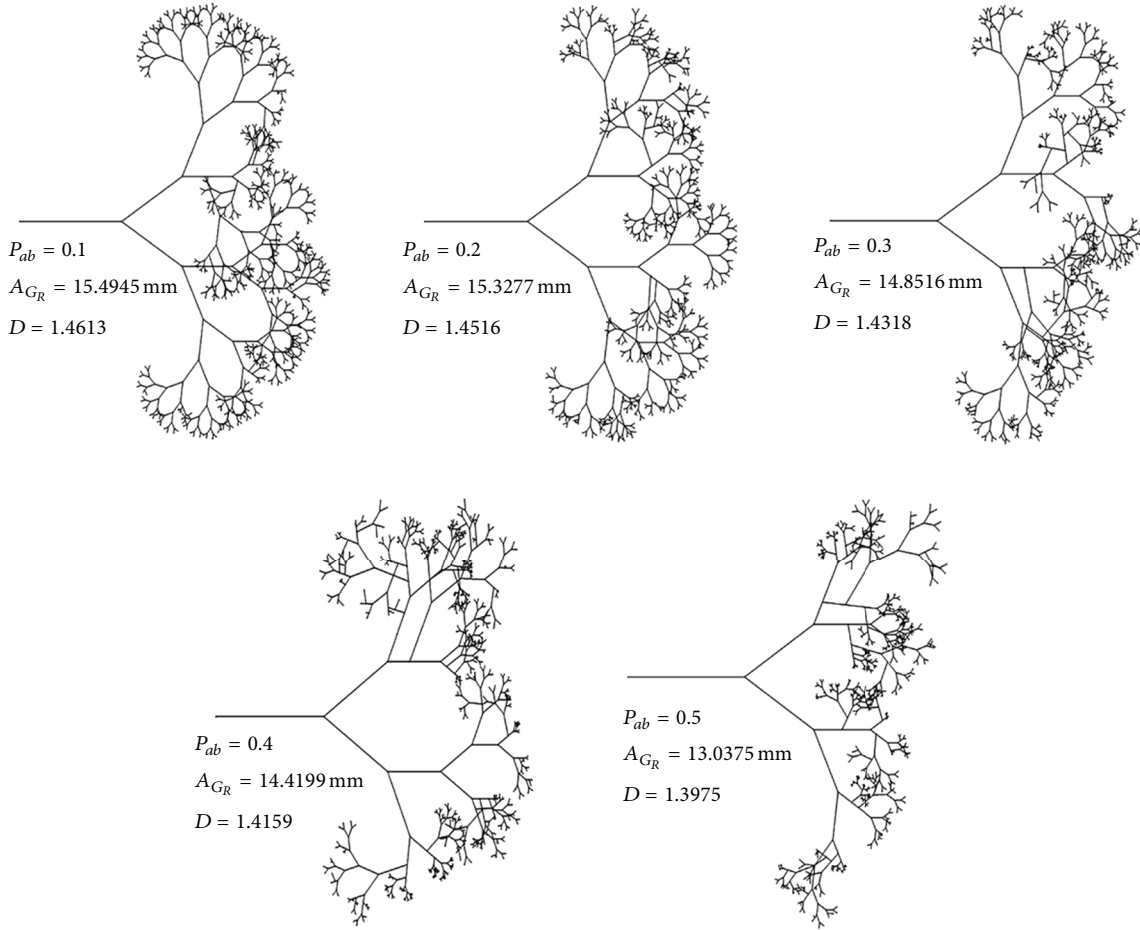


FIGURE 3: Trees generated with different probabilities in a_b and a_p (where $P_{ap} = 1 - P_{ab}$), mean width A_{GR} and fractal dimension D of the kidney.

contraction factor is $\lambda_b = 0.5$. Since we have $P_{ab}\lambda_b^D + P_{ap}\lambda_p^D = 1$, the analytical fractal dimension of the renal arterial tree is

$$D = \frac{\ln(P_{ap}) - \ln((1/2) - P_{ab}\lambda_b^D)}{-\ln(\lambda_p)}, \quad (12)$$

where λ_b is generated randomly within the range defined by $\lambda_b \in [0.3, 0.7]$. In other words, the fractal dimension depends on the number of vessels, the spatial relationships between the vascular components, and the surrounding environment.

3.2.1. Box Counting D of the Arterial Vascular Tree of the Kidney. Additionally, the box counting theorem by Barnsley [19] states that, denoting with $N_n(G_R)$ as the number of boxes of side $1/2^n$,

$$D = \lim_{n \rightarrow \infty} \frac{\ln(N_n(G_R))}{\ln(2^n)}, \quad (13)$$

where the arterial vascular tree of the kidney has fractal dimension D . The box counting results allow one to compute the rate of change in complexity with scale as well as a measure of heterogeneity.

Here, we used the box counting theorem for determining the fractal dimension. Figure 2 shows the average of the fractal dimension in each probability for the arterial vascular tree of the kidney. The fractal dimension decreases with the increase in the probability of sprouting in the development of the arterial vascular tree of the kidney, whereas the mean width of the kidney decreases. Then, the fractal dimension of the arterial vascular tree of the kidney depends on the probability of a_b and a_p , that is, the distribution of blood vessels in the morphology generated by graph theory model.

The results suggest that the fractal dimension is inversely proportional to P_{ab} . This behavior is congruent in the sense that, by symmetry structure, the arterial vascular tree is bigger for small values of P_{ab} ; that is, the renal space is covered more efficiently. The complex structures of the vasculature in kidney are consistent with previous studies of vascular structures about their fractal nature [9–16].

Examples derived from graph model using different probabilities in a_b and a_p , mean width and fractal dimension, are shown in Figure 3. The analysis of these responses may allow for characterization of the vascular morphology on the arterial vascular tree of the kidney.

4. Conclusions

The arterial vascular tree of kidney development was analyzed through angiogenesis. The renal arterial vascular tree of the kidney goes down into the interlobular arteries. We generate 5000 trees of renal arterial vasculature for different probabilities on a_b and a_p . The analytical mean length in each level and the average width of the arterial vascular tree of the kidney have an intersection with previous studies. Then, the graph theory allows the vascular tree to model the vascular growth; that is, it generates the ramification of structures arborescent incorporating physiological laws of arterial branching.

In conclusion, the analytical expression of the fractal dimension depends on the number of vessels, the spatial relationships between the vascular components, and the surrounding environment. The arterial vascular tree of the kidney has a fractal dimension which is inversely proportional to the probability of the occurring sprouting angiogenesis. The fractal dimensions determined for the development of the arterial vascular tree of the kidney by box counting may allow for characterization of the vascular morphology. As a conjecture, the fractal dimension might determine whether a suitable renal vascular structure is capable of performing physiological functions under appropriate conditions (hemodynamics). These studies could be expanded to include those pathologies originating from arterial alterations.

References

- [1] R. J. Tomake, *Assembly of the Vasculature and Its Regulation*, Birkhäuser, Berlin, Germany, 2001.
- [2] E. M. Wahl, L. V. Quintas, L. L. Lurie, and M. L. Gargano, "A graph theory analysis of renal glomerular microvascular networks," *Microvascular Research*, vol. 67, no. 3, pp. 223–230, 2004.
- [3] D. A. Nordsletten, S. Blackett, M. D. Bentley, E. L. Ritman, and N. P. Smith, "Structural morphology of renal vasculature," *American Journal of Physiology*, vol. 291, no. 1, pp. H296–H309, 2006.
- [4] M. Zamir, "Arterial branching within the confines of fractal L-system formalism," *Journal of General Physiology*, vol. 118, no. 3, pp. 267–275, 2001.
- [5] A. Espinoza-Valdez, R. Femat, and F. C. Ordaz-Salazar, "A model for renal arterial branching based on graph theory," *Mathematical Biosciences*, vol. 225, no. 1, pp. 36–43, 2010.
- [6] A. C. Guyton and J. E. Hall, *Tratado de Fisiología Médica*, McGraw-Hill, 10th edition, 2000, in Spanish.
- [7] S. Patan, "Vasculogenesis and angiogenesis as mechanisms of vascular network formation, growth and remodeling," *Journal of Neuro-Oncology*, vol. 50, no. 1-2, pp. 1–15, 2000.
- [8] M. Zamir and S. Phipps, "Morphometric analysis of the distributing vessels of the kidney," *Canadian Journal of Physiology and Pharmacology*, vol. 65, no. 12, pp. 2433–2440, 1987.
- [9] Y. Gazit, D. A. Berk, M. Leunig et al., "Scale-invariant behavior and vascular network formation in normal and tumor tissue," *Physical Review Letters*, vol. 75, no. 12, pp. 2428–2431, 1995.
- [10] Y. Gazit, J. W. Baish, N. Safabakhsh, M. Leunig, L. T. Baxter, and R. K. Jain, "Fractal characteristics of tumor vascular architecture during tumor growth and regression," *Microcirculation*, vol. 4, no. 4, pp. 395–402, 1997.
- [11] J. Panico and P. Sterling, "Retinal neurons and vessels are not fractal but space-filling," *Journal of Comparative Neurology*, vol. 361, no. 3, pp. 479–490, 1995.
- [12] P. G. Vico, S. Kyriacos, O. Heymans, S. Louryan, and L. Cartilier, "Dynamic study of the extraembryonic vascular network of the chick embryo by fractal analysis," *Journal of Theoretical Biology*, vol. 195, no. 4, pp. 525–532, 1998.
- [13] C. Arlt, H. Schmid-Schönbein, and M. Baumann, "Measuring the fractal dimension of the microvascular network of the chorioallantoic membrane," *Fractals*, vol. 11, no. 2, pp. 205–212, 2003.
- [14] S. Lorthois and F. Cassot, "Fractal analysis of vascular networks: insights from morphogenesis," *Journal of Theoretical Biology*, vol. 262, no. 4, pp. 416–433, 2010.
- [15] E. Sabo, A. Boltenko, Y. Sova, A. Stein, S. Kleinhaus, and M. B. Resnick, "Microscopic analysis and significance of vascular architectural complexity in renal cell carcinoma," *Clinical Cancer Research*, vol. 7, no. 3, pp. 533–537, 2001.
- [16] S. S. Cross, R. D. Start, P. B. Silcocks, A. D. Bull, D. W. K. Cotton, and J. C. E. Underwood, "Quantitation of the renal arterial tree by fractal analysis," *Journal of Pathology*, vol. 170, no. 4, pp. 479–484, 1993.
- [17] M. N. Nakatsu, R. C. A. Sainson, S. Pérez-del-Pulgar et al., "VEGF₁₂₁ and VEGF₁₆₅ regulate blood vessel diameter through vascular endothelial growth factor receptor 2 in an *in vitro* angiogenesis model," *Laboratory Investigation*, vol. 83, no. 12, pp. 1873–1885, 2003.
- [18] E. Gabryś, M. Rybaczuk, and A. Kędzia, "Fractal models of circulatory system. Symmetrical and asymmetrical approach comparison," *Chaos, Solitons & Fractals*, vol. 24, no. 3, pp. 707–715, 2005.
- [19] M. Barnsley, *Fractals Everywhere*, Academic Press, 1988.

Research Article

Hemodynamic Features in Stenosed Coronary Arteries: CFD Analysis Based on Histological Images

Mahsa Dabagh,^{1,2} Wakako Takabe,^{2,3} Payman Jalali,¹ Stephan White,⁴ and Hanjoong Jo^{2,3}

¹ LUT School of Technology, Lappeenranta University of Technology, 53851 Lappeenranta, Finland

² Department of Cardiology, Emory University, Atlanta, GA 30322, USA

³ Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

⁴ Faculty of Medicine and Dentistry, University of Bristol, Bristol BS8 1TH, UK

Correspondence should be addressed to Mahsa Dabagh; mahsa@lut.fi

Received 8 March 2013; Revised 30 May 2013; Accepted 30 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Mahsa Dabagh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Histological images from the longitudinal section of four diseased coronary arteries were used, for the first time, to study the pulsatile blood flow distribution within the lumen of the arteries by means of computational fluid dynamics (CFD). Results indicate a strong dependence of the hemodynamics on the morphology of atherosclerotic lesion. Distinctive flow patterns appear in different stenosed regions corresponding to the specific geometry of any artery. Results show that the stenosis affects the wall shear stress (WSS) locally along the diseased arterial wall as well as other adjacent walls. The maximum magnitude of WSS is observed in the throat of stenosis. Moreover, high oscillatory shear index (OSI) is observed along the stenosed wall and the high curvature regions. The present study is capable of providing information on the shear environment in the longitudinal section of the diseased coronary arteries, based on the models created from histological images. The computational method may be used as an effective way to predict plaque forming regions in healthy arterial walls.

1. Introduction

Atherosclerosis in the coronary arteries tends to be localized in regions with curvature and branching associated with complex, unsteady, and turbulent pattern of the flow [1]. In these regions, the fluid shear stress deviates from its normal spatial and temporal distribution patterns in straight vessels. The role of WSS in the localization of the atherosclerosis has been widely accepted [1–6]. Shahcheraghi et al. [3] explained clinically the fact that a low shear stress region is generally found in the downstream of developing stenotic plaques, leading to an increased density of leaky junctions associated with an increased level of low-density-lipoprotein (LDL) flux into the arterial wall. Moreover, the OSI is known as the predictor of formation of atherosclerosis and vulnerability for plaque in coronary arteries [7–9]. High OSI has been widely used as indicators of atherosclerosis because the regions with

high OSI are likely to experience stagnation or backflow [7–11]. This is strongly related to the progression of atherosclerosis because it affects the arrangement of endothelial cells in adjacent tissues. It has been shown that the areas with high values of OSI are usually located in the regions where wall shear stress is low [7]. Taking all together, the creation and development of atherosclerosis depend not only on the biological features but also on the biomechanical factors, including WSS and the cyclic force caused by the pulsatile blood pressure. Therefore, a full description of the flow pattern through the diseased coronaries, with the emphasis on the patient-specific data, is essential for quantifying the wall shear stress that may occur in a stenosed artery during the cardiac cycle. The knowledge will also lead to find the most vulnerable sites for the plaque rupture.

Previous studies have shown the flow pattern through the idealized geometry [6–8] or the reconstructed geometry of

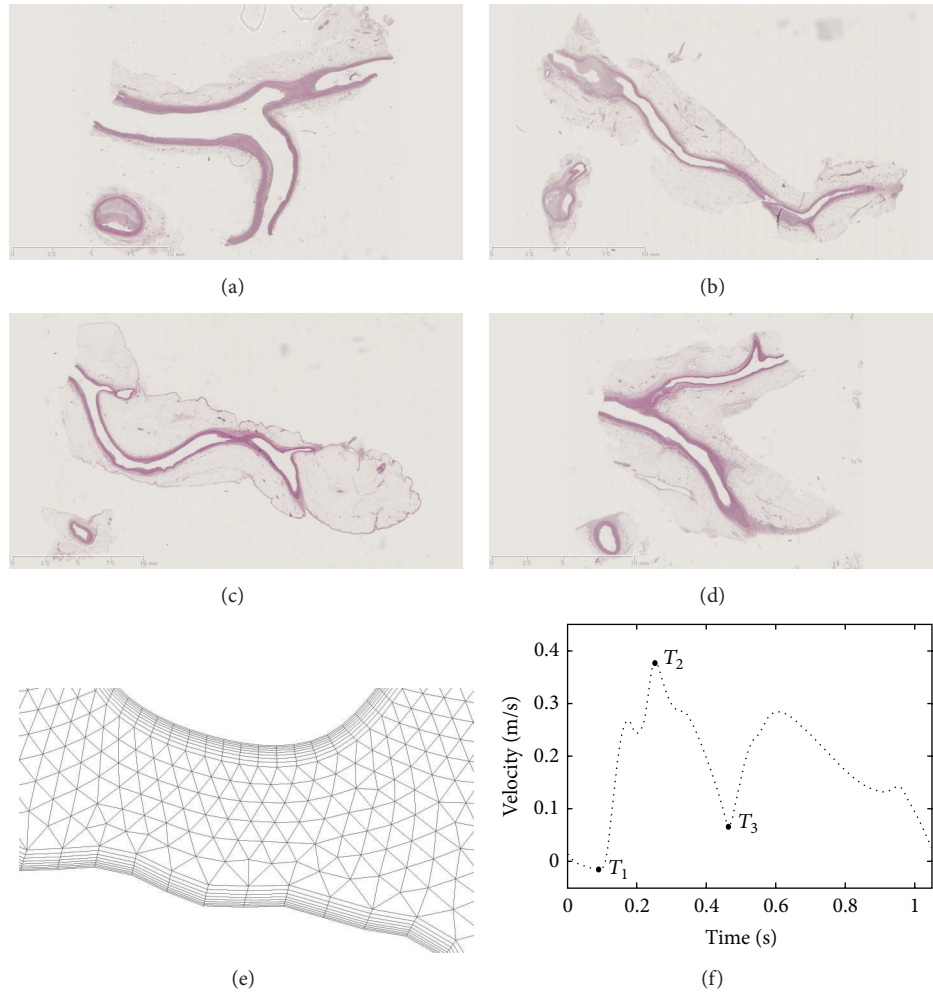


FIGURE 1: (a)–(d) LS images of four diseased coronary arteries, (e) the computational meshes at the stenosis location, and (f) velocity pulse at the inlet.

arteries from a series of slices obtained using computerized tomography (CT) scan images [9, 10], magnetic resonance images (MRI) [11], the spiral CT scan [12], the phase-contrast MRI method [13], or intravascular ultrasound (IVUS) [4]. However, simplifications were applied in these studies; for example, the cross-section of the artery was approximated as circles which fit the centerline to a cubic spline curve [7, 13], or a constant diameter was applied throughout the model [13, 14].

In the present study, the histological images from longitudinal section of four diseased coronary arteries were used, for the first time, to investigate the local hemodynamic more precisely through the stenosed artery models. The geometry is reconstructed directly by graphical reading of points on the luminal surface of the arteries. Any simplifications have been avoided in the construction of the model. The pulsatile blood flow pattern was applied through the reconstructed stenosed coronary arteries, and the blood flow characteristics along the arterial wall was analyzed by means of computational fluid dynamics (CFD).

2. Methods

2.1. Reconstruction of 2D Geometry of Coronary Artery from Histological Image. Two-dimensional (2D) geometries of the human coronaries are reconstructed from the histological images which were acquired from coronary arteries of four donor hearts (Faculty of Medicine and Dentistry, University of Bristol, Bristol, UK). The samples contained regions with mild to severe atherosclerosis. The diseased coronary arteries were cut in longitudinal section, stained with H&E, and imaged on a nanozoomer, namely, NanoZoomer 2.0HT (Hamamatsu photonics, Hamamatsu, Japan) with 40x magnification at the pathology core lab in the Winship Cancer Institute, Emory University.

To construct the coronary artery geometries, all the images were processed individually. The images were converted first to the boundary plots within MATLAB v. 7.11.0 (R2010b). The lumen area was identified via image segmentation. Points representing the surface of the endothelium were obtained as the final output of this step. The vertices

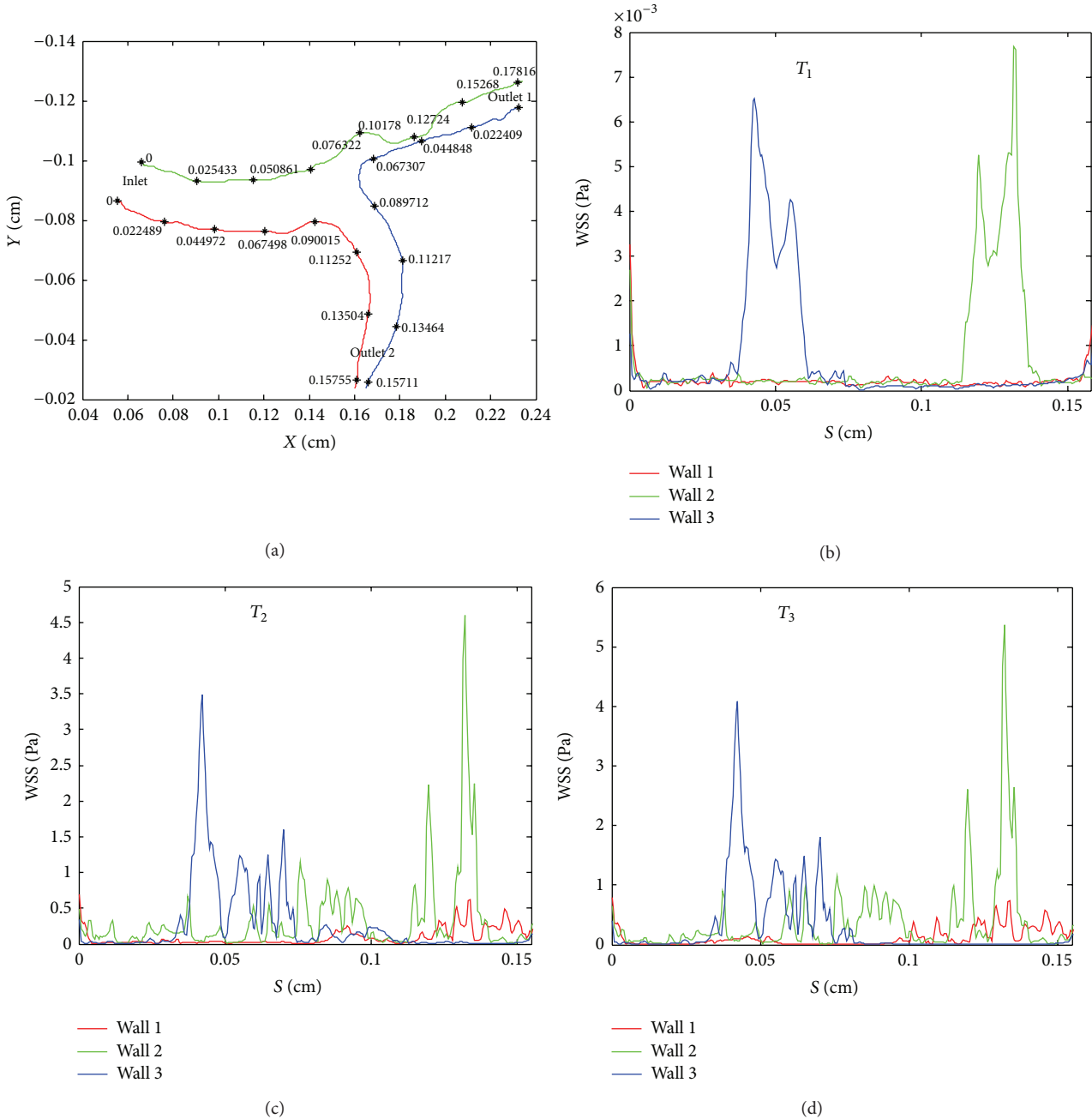


FIGURE 2: (a) Boundary plots of the coronary arterial walls 1, 2, and 3 shown in red, green, and blue, respectively. (b)–(d) WSS distribution along coronary walls at time T_1 , T_2 , and T_3 , respectively.

acquired from the processing of the images were connected to generate a 2D surface representing the lumen of coronary. The 2D model was trimmed with the CFD preprocessing and meshing code Gambit V2.4.6 to generate the flat inlet and outlet boundaries. The computational mesh grid consisted of about 50,000 grid elements. Thin boundary layers were used for the precise calculations near the walls. Several steady-state simulations were performed with different grid sizes to ensure the mesh independence of the results. The boundary layer mesh used in these simulations consists of 6 layers with the first layer attached to the luminal surface with the

thickness of 0.01 mm. The thickness of consecutive layers grows with the factor of 1.2. Figures 1(a)–1(d) demonstrate the original histological image of each coronary artery. Figure 1(e) shows the grids generated for the models.

2.2. Governing Equations and Boundary Conditions. The blood flow is assumed to be incompressible with uniform properties including the viscosity of $3.5 \times 10^{-3} \text{ Pa} \cdot \text{s}$ (Newtonian fluid) and the density of 1067 kgm^{-3} . The assumption of Newtonian fluid for the blood is valid wherever the shear

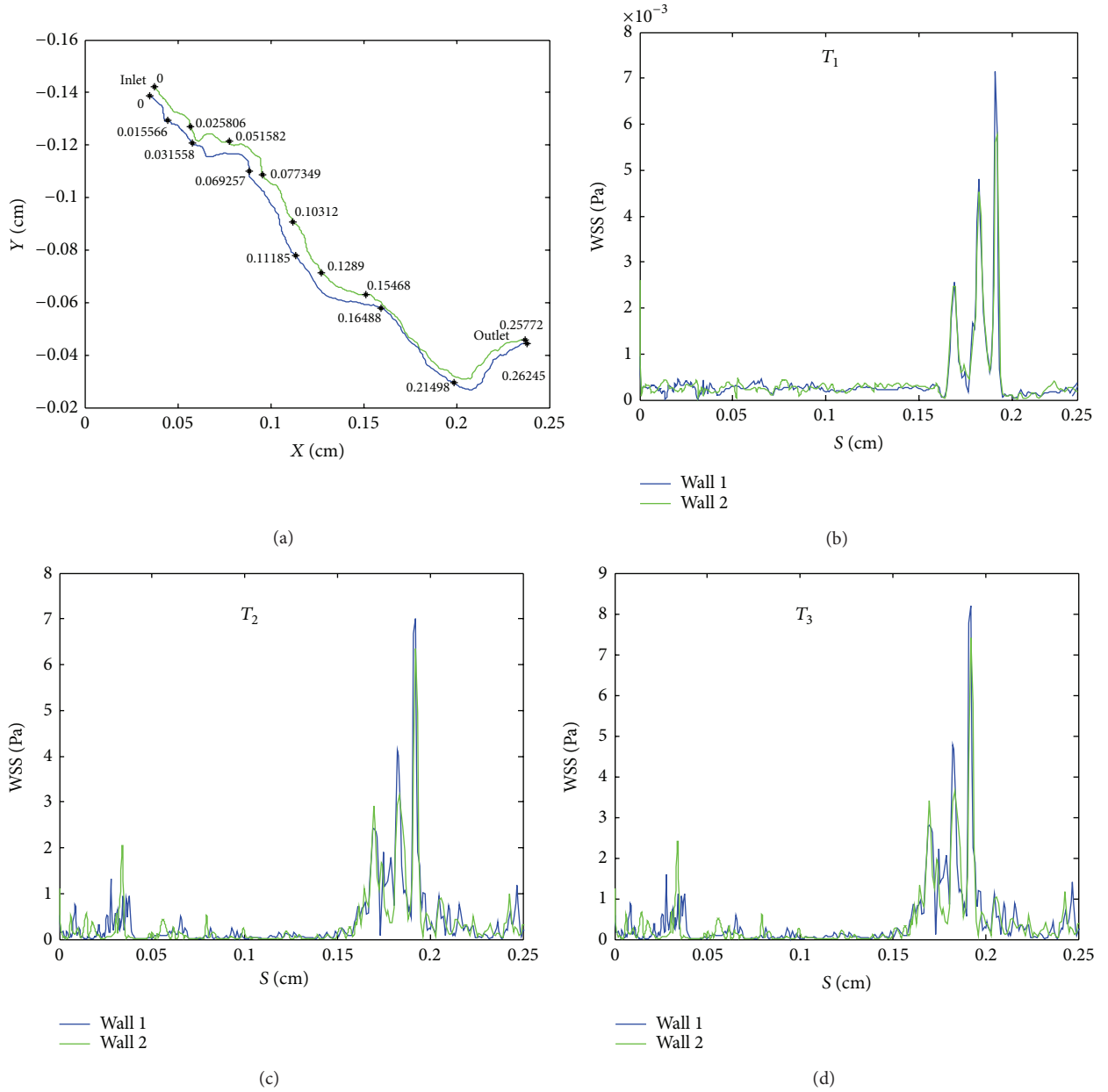


FIGURE 3: (a) Boundary plots of the coronary arterial walls 1 and 2 shown in green and blue, respectively. (b)–(d) WSS distribution along coronary walls at time T_1 , T_2 , and T_3 , respectively.

rate is greater than 100 S^{-1} [9, 10, 12, 13]. This assumption is feasible for coronary artery. The time-dependent Navier-Stokes equations govern the blood flow in the coronary artery, as

$$\begin{aligned} \nabla \cdot \vec{U} &= 0, \\ \rho \frac{\partial \vec{U}}{\partial t} + \rho (\vec{U} \cdot \nabla) \vec{U} &= -\nabla p + \mu \nabla^2 \vec{U}, \end{aligned} \quad (1)$$

where \vec{U} is the velocity vector, while p , ρ , and μ denote pressure, density, and viscosity, respectively. The maximum Reynolds number, $\text{Re} = \rho \bar{U} D / \mu$ (where D is the inlet diameter of coronary and \bar{U} is the maximum pulsating inflow

velocity), is calculated as 232. The inlet diameter of the distal left anterior descending coronary artery is $1.9 \pm 0.4 \text{ mm}$, and the maximum pulsating inflow velocity is taken as 0.4 m/s [14]. Figure 1(e) demonstrates the pulse for inlet velocity. The critical Reynolds number for an unsteady flow ranges from 5000 to 20,000 [15–17]. Therefore, the laminar flow assumption is reasonable in the present simulation.

The pulse of the inlet velocity (Figure 1(f)) applied in the proximal site of the right coronary artery (RCA) is decomposed into a Fourier series of trigonometric basis functions with 3 harmonics. The input velocity shown in Figure 1(f) is based on the velocity waveforms measured with a commercial ultrasound Doppler probe FloWire (Volcano

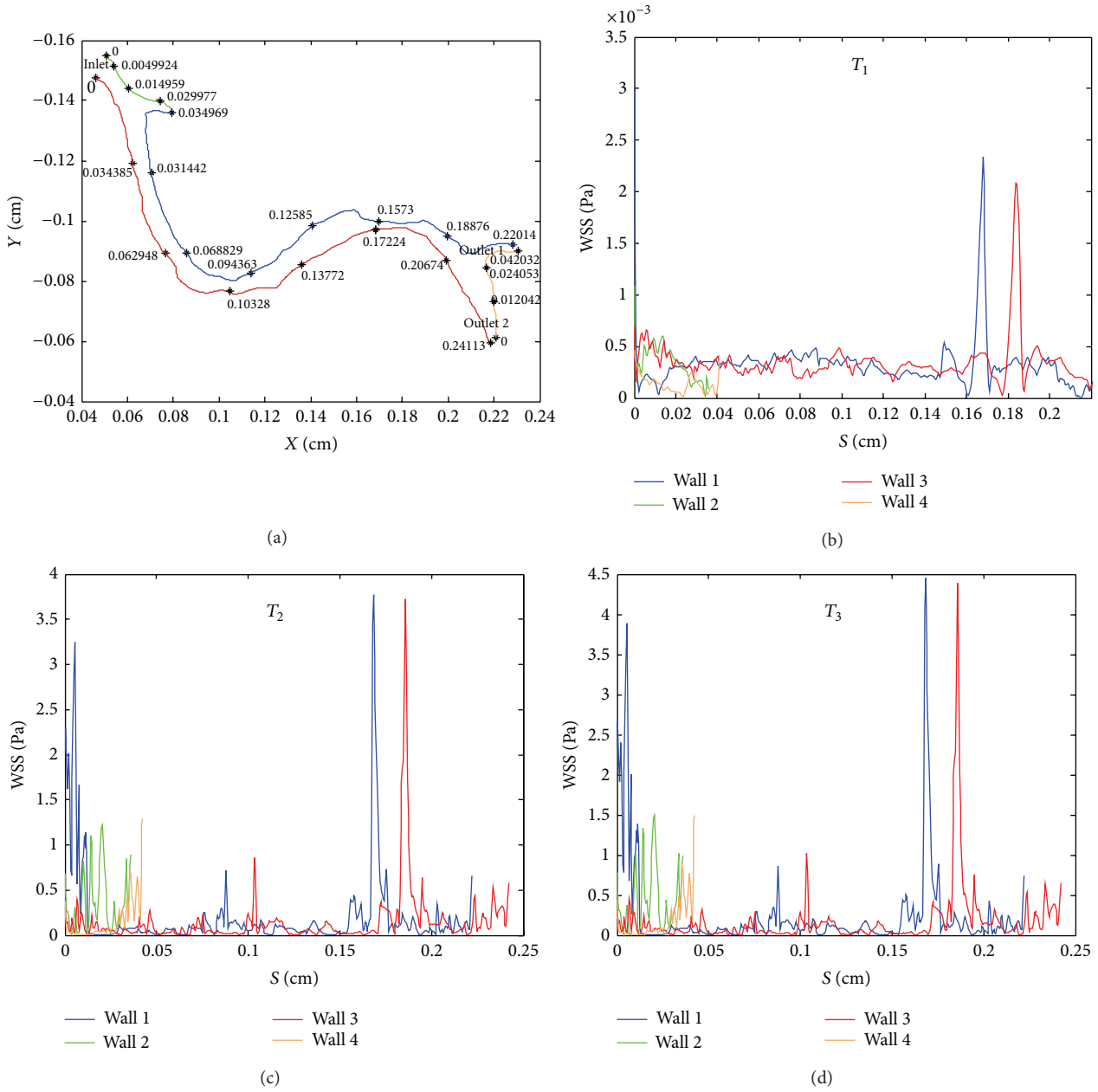


FIGURE 4: (a) Boundary plots of the coronary arterial walls 1, 2, 3, and 4 shown in red, green, blue, and brown, respectively. (b)–(d) WSS distribution along coronary walls at time T_1 , T_2 , and T_3 , respectively.

TM Corporation) with 1kHz sampling rate at the proximal site of RCA, provided by Torii et al. [18]. The inlet velocity was kept constant across the inlet surface at each time instant of the cardiac cycle. The points indicated by T_1 , T_2 , and T_3 in Figure 1(f) correspond to the diastole, the peak systole, and accelerating phase within the cardiac cycle, respectively. The outflow boundary conditions were imposed at the outlets. The coronary arterial walls are assumed to be rigid with no-slip boundary condition on the luminal surface. The governing equations along with the given boundary

conditions are solved with the finite-volume (FV) solver package, Fluent V12.1.4. The SIMPLE algorithm is used for the coupling of the pressure-velocity terms. The simulations are carried out for four inlet cycles to ensure that the spatial and temporal flow dynamics are periodic. The total time of 1s with the time step size of 1ms is taken for each cycle. The simulations are performed for four cardiac cycles, of which only the results of the fourth cycle are demonstrated in the following section at different times to ensure that all the phenomena associated with the initial conditions are damped

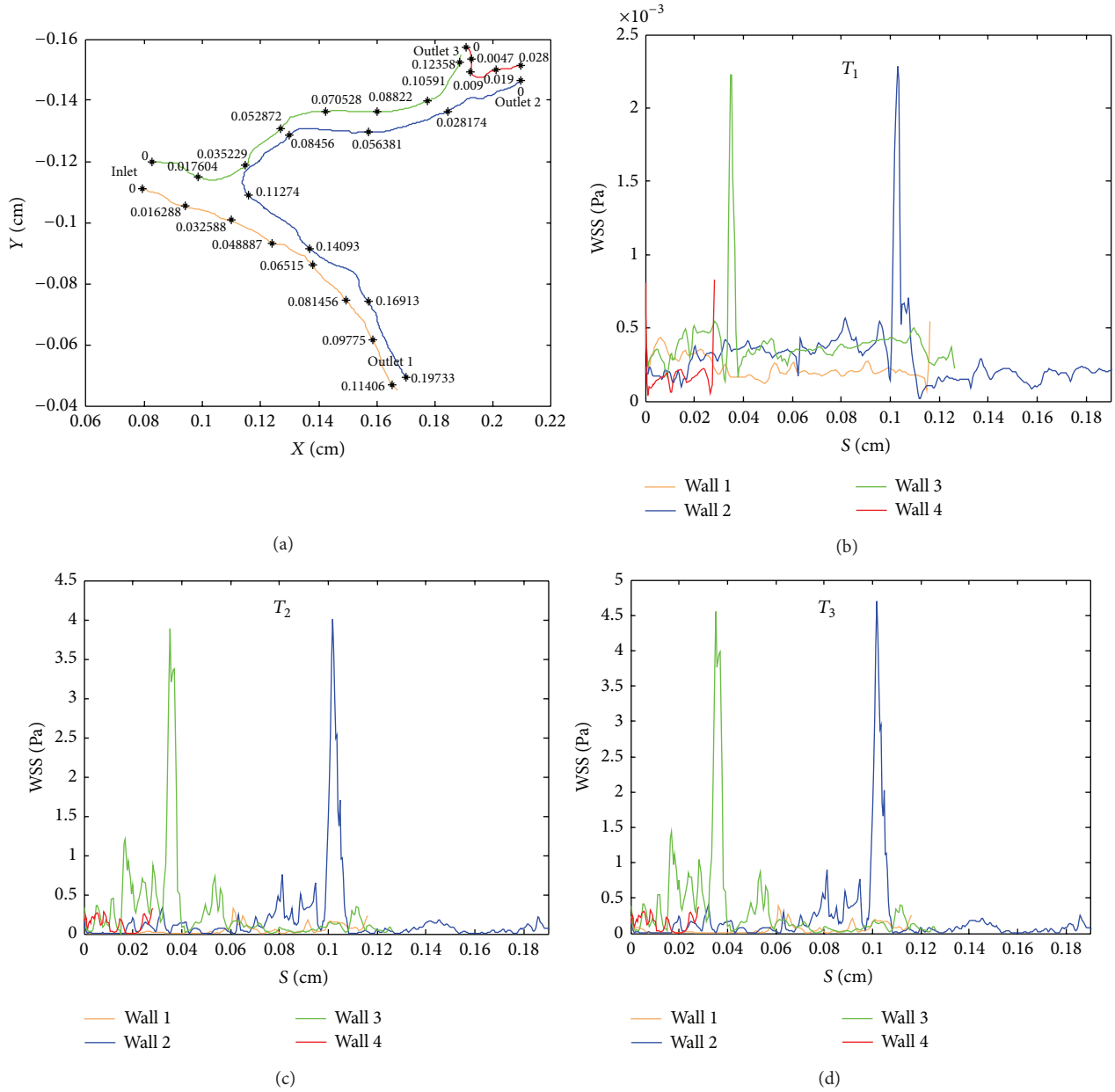


FIGURE 5: (a) Boundary plots of coronary arterial walls 1, 2, 3, and 4 shown in red, green, blue, and brown, respectively. (b)–(d) WSS distribution along coronary walls at time T_1 , T_2 , and T_3 , respectively.

out during the first three cycles. The oscillatory shear index for the pulsatile flow simulation is calculated as [4]

$$\text{OSI} = 0.5 \left[1 - \frac{\left| \int_0^T \bar{\tau}_w dt \right|}{\int_0^T |\bar{\tau}_w| dt} \right]. \quad (2)$$

3. Results and Discussion

3.1. Wall Shear Stress. The wall shear stress is among the first mechanisms proposed to relate the blood flow to the localization of atherosclerosis [18] and the plaque rupture [19, 20]. Figures 2–5 demonstrate the WSS distributions along the

coronary walls versus the curve length (S), at three different time instants, for four different diseased coronary arteries. In the present study, the WSS is defined as the magnitude of all three components at any location on the wall.

Figure 2(a) demonstrates the boundary plots for the images of the first stenosed coronary artery (Figure 1(a)). The WSS distributions along the coronary walls versus curve length at diastole, the peak systole, and accelerating phase are shown in Figures 2(b)–2(d), respectively. The areas with maximum WSS can be observed along the distal side and on the throat of the stenosis at diastole time point. Note that a minimum value of WSS appears at the top of the atherosclerotic plaque at the peak systole and accelerating

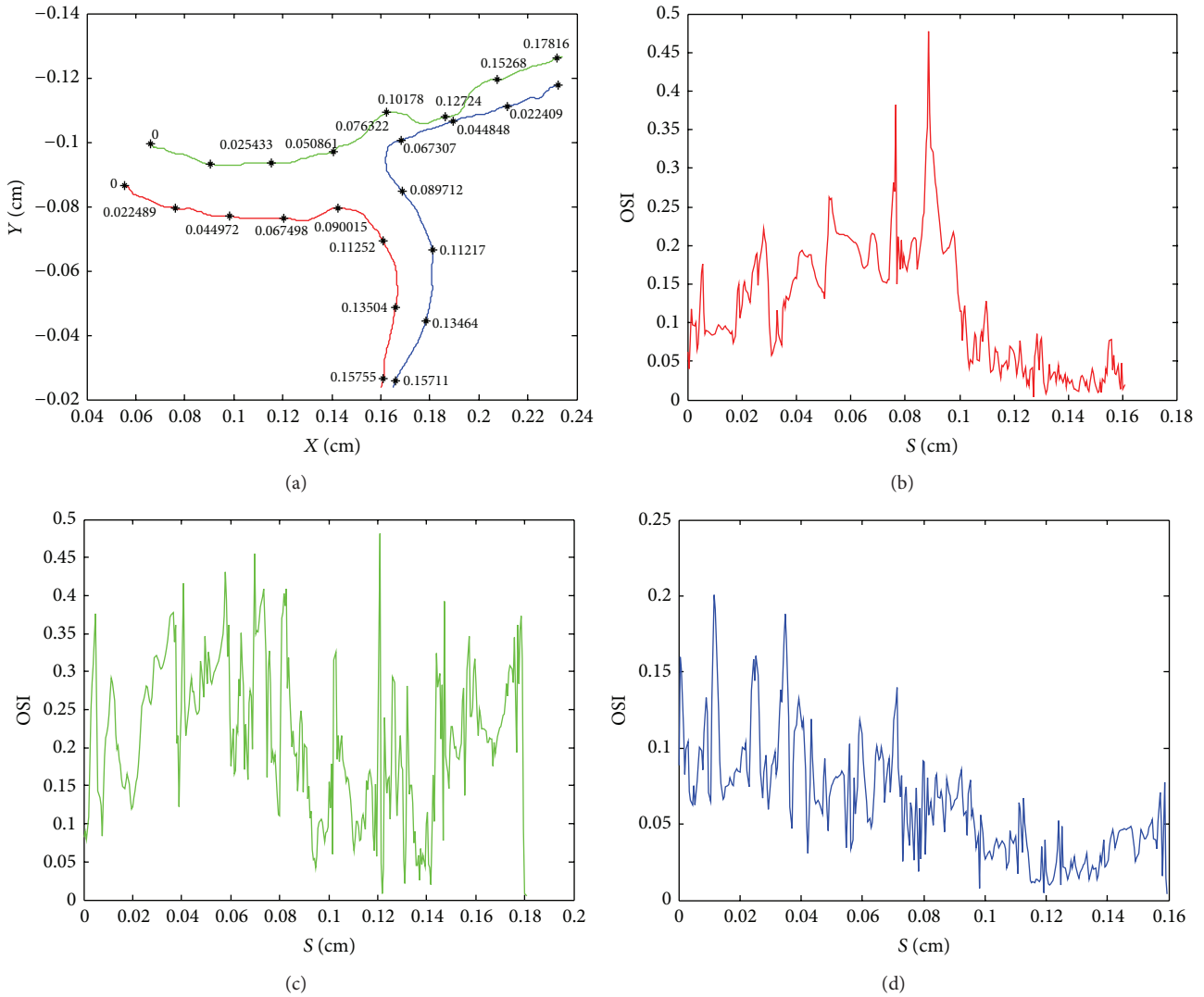


FIGURE 6: (a) Boundary plots of the coronary arterial walls 1, 2 and 3 shown in red, green, and blue, respectively. (b)–(d) OSI distribution along coronary walls 1, 2, and 3, respectively.

time points. It is worth mentioning that the surfaces of the coronary walls, in the present study, are generated based on the original images of patients. Surfaces are not smoothed after being read in MATLAB. The reason is associated with the aim of the current study at the investigation of the connection of atherosclerotic plaque to the local characteristics of the luminal surface. As Figure 2(a) shows, the wall on the top of the plaque is not smooth which is according to the original histological image (Figure 1(a)). This is the region with a minimum magnitude of WSS.

Figures 3(a)–3(d) show the WSS versus curve length along the walls of the second stenosed coronary artery (Figure 1(b)). The maximum magnitude of WSS occurs at the top of the plaque. Another stenosed region shows the second peak of WSS values at systole and accelerating time points.

Figures 4(a)–4(d) demonstrate the WSS along the walls of the third diseased coronary artery (Figure 1(c)), while Figures 5(a)–5(d) demonstrate the WSS along the walls of the fourth

diseased coronary artery (Figure 1(d)). Maximum of the WSS magnitude occurs at the throat of the stenoses in both models, at diastole time point. As seen in Figures 4(c) and 4(d), another maximum value of WSS is observed at systole and accelerating time points in the region with high curvature. In Figures 5(c) and 5(d), the proximal side of the plaque shows relatively high magnitude of WSS at systole and accelerating time points. In this artery the stenosis has been created at the bifurcation.

3.2. Oscillatory Shear Index (OSI). OSI is a known predictor of formation of atherosclerosis and vulnerability for plaque in coronary arteries. Figures 6(b)–6(d) demonstrates OSI for the first stenosed coronary artery. Figure 6(b) shows that OSI has high values at the high curvature region. Therefore, this location may be predicted to experience lesion in the future. As shown in Figure 6(c), OSI is high along the stenosed wall of coronary artery (shown in green in Figure 6(a)). OSI

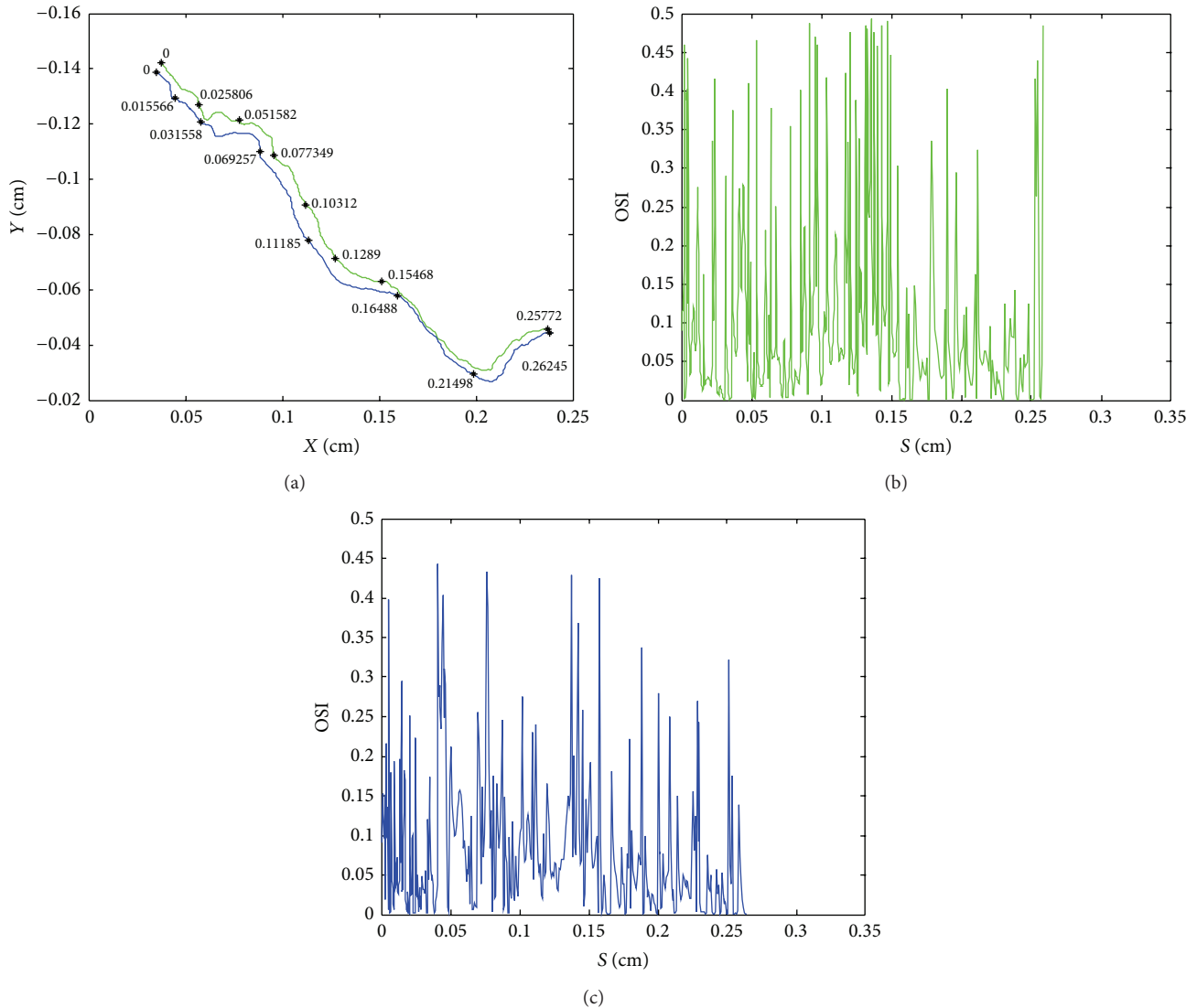


FIGURE 7: (a) Boundary plots of the coronary arterial walls 1 and 2 shown in green and blue, respectively. (b)-(c) OSI distribution along coronary walls 1 and 2.

distribution reveals that the whole length of the green wall is vulnerable to atherosclerosis. Figure 6(d) demonstrates a relatively high OSI values at the curvature region along the wall.

Figures 7, 8, and 9 show OSI for the next stenosed coronary arteries. Maximum values of OSI occur in stenosed regions, in all arteries. The next area with maximum of OSI is the region with high curvature. Note that this behavior is observed even in healthy curves.

4. Conclusions

In this study, pulsatile blood flow is studied in four reconstructed models of the diseased coronary arteries from histological images. Surface smoothing and approximation of the vessels with regular shapes such as straight or curved cylinders are avoided unlike most of earlier studies. The real

geometric model and the conditions of pulsatile flow allow to spot vulnerable areas to the development of atherosclerotic lesions. The results indicate a strong dependence of the hemodynamics on the morphology of atherosclerotic lesion. Distinctive flow patterns appear in different stenosed regions corresponding to the specific geometry of the arteries. The OSI is a well-known predictor of formation of atherosclerosis and vulnerability for plaque in coronary arteries. High OSI has been widely used as indicators of atherosclerosis because the region with high OSI is likely to experience stagnation or backflow. This is strongly related to the progression of atherosclerosis because it affects the arrangement of endothelial cells in adjacent tissues. It has been shown that the areas with high values of OSI are usually located in the regions where wall shear stress is low. In the present study, OSI is high in diseased region as well as healthy regions with low values of WSS which can be predicted to experience

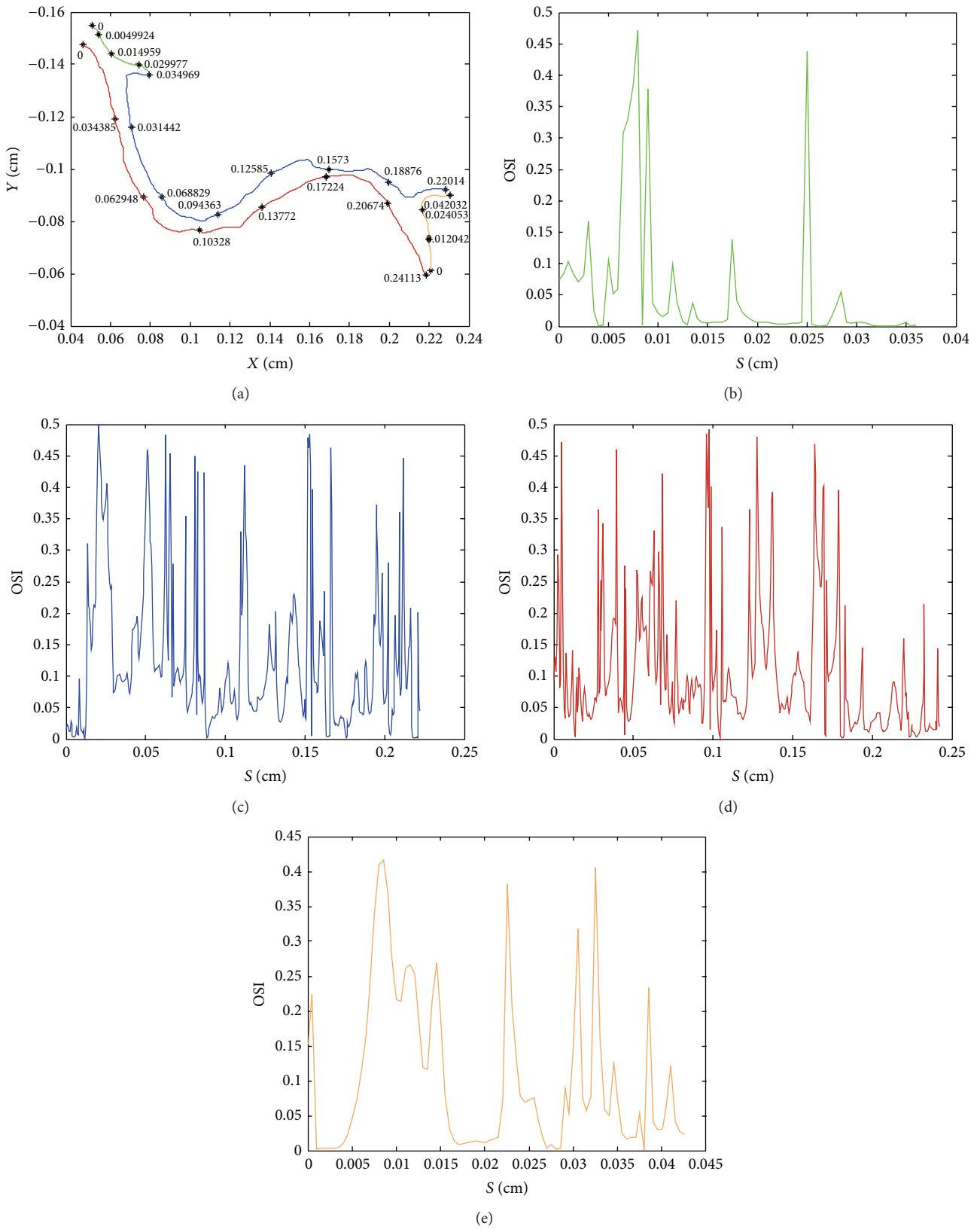


FIGURE 8: (a) Boundary plots of coronary arterial walls 1, 2, 3, and 4 shown in red, green, blue, and brown, respectively. (b)–(e) OSI distribution along coronary walls 1, 2, 3, and 4.

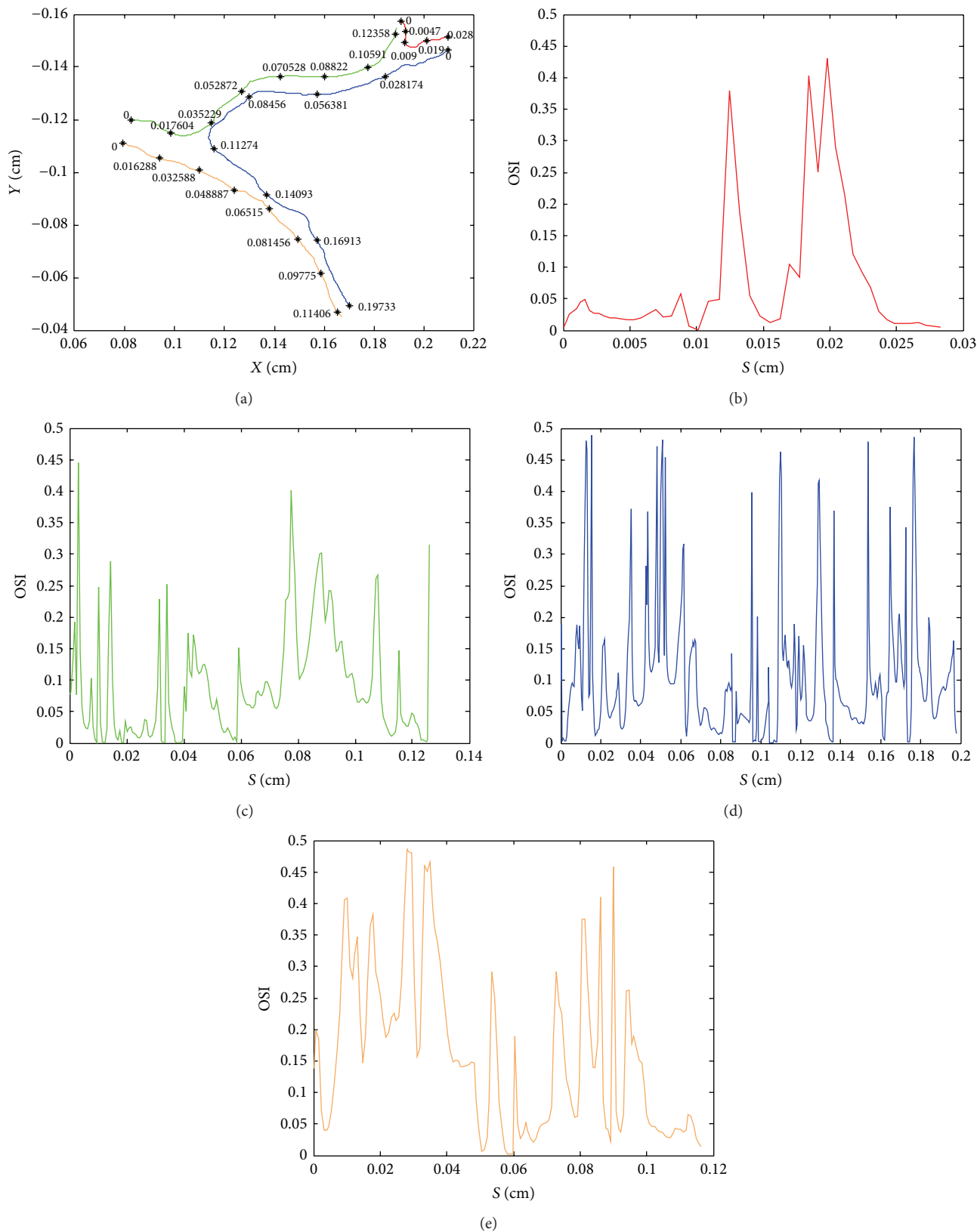


FIGURE 9: (a) Boundary plots of coronary arterial walls 1, 2, 3, and 4 shown in red, green, blue, and brown, respectively. (b)–(e) OSI distribution along coronary walls 1, 2, 3, and 4.

lesion in the future. The findings support the low shear stress and oscillatory shear stress theory which relates abnormal biological effects inside arteries to high OSI. Moreover, the present study demonstrates the shear environment in longitudinal section of the diseased coronary. The stenosis affects the WSS locally along the diseased arterial wall as well as adjacent walls. The maximum value of WSS is observed in the throat of stenosis which highlights the risk of plaque rupture. The arterial plaque can break apart as a result of high stresses acting on or within the arterial wall. The severity and the peripheral distribution of stenosis are the crucial geometric factors influencing the risk of breakage.

It is hoped that computational studies of this type will enable scientists to directly relate the hemodynamics to the formation and progression of atherosclerosis in the coronary arteries. More than 60% of myocardial infarction is caused by the rupture of vulnerable plaques [17]. A significant contribution of the present study is the demonstration of robust and highly efficient computational methods for studying the pulsatile blood flow field in different stenosed coronary arteries. The current study is the first detailed numerical investigation of the flow field within the diseased coronary arteries which are reconstructed from histological images. The spatial and temporal variation of wall shear stress from the present study can be fed into other models to investigate the effect of wall shear stress on stresses and strains that individual endothelial cells may tolerate locally.

Conflict of Interests

None of the authors has a conflict of interests regarding this work.

Acknowledgments

M. Dabagh and P. Jalali would like to thank the financial support from the Faculty of Technology in Lappeenranta University of Technology and Tukisäätiö Foundation. Authors kindly acknowledge the Pathology Core Lab in the Winship Cancer Institute, Emory University, for preparing the images.

References

- [1] M. Dabagh, P. Jalali, and J. M. Tarbell, "The transport of LDL across the deformable arterial wall: the effect of endothelial cell turnover and intimal deformation under hypertension," *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 297, no. 3, pp. H983–H996, 2009.
- [2] J. M. Tarbell, "Mass transport in arteries and the localization of atherosclerosis," *Annual Review of Biomedical Engineering*, vol. 5, pp. 79–118, 2003.
- [3] N. Shahcheraghi, H. A. Dwyer, A. Y. Cheer, A. I. Barakat, and T. Rutaganira, "Unsteady and three-dimensional simulation of blood flow in the human aortic arch," *Journal of Biomechanical Engineering*, vol. 124, no. 4, pp. 378–387, 2002.
- [4] D. Mori, T. Hayasaka, and T. Yamaguchi, "Modeling of the human aortic arch with its major branches for computational fluid dynamics simulation of the blood flow," *JSME International Journal*, vol. 45, no. 4, pp. 997–1002, 2002.
- [5] L. Morris, P. Delassus, A. Callanan et al., "3-D numerical simulation of blood flow through models of the human aorta," *Journal of Biomechanical Engineering*, vol. 127, no. 5, pp. 767–775, 2005.
- [6] C.-Y. Wen, A.-S. Yang, L.-Y. Tseng, and J.-W. Chai, "Investigation of pulsatile flowfield in healthy thoracic aorta models," *Annals of Biomedical Engineering*, vol. 38, no. 2, pp. 391–402, 2010.
- [7] H. S. Ryou, S. Kim, S. W. Kim, and S. W. Cho, "Construction of healthy arteries using computed tomography and virtual histology intravascular ultrasound," *Journal of Biomechanics*, vol. 45, pp. 1612–1618, 2012.
- [8] P. Vasava, P. Jalali, M. Dabagh, and P. J. Kolari, "Finite element modelling of pulsatile blood flow in idealized model of human aortic arch: study of hypotension and hypertension," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 861837, 14 pages, 2012.
- [9] J.-M. Zhang, L. P. Chua, D. N. Ghista, S. C. M. Yu, and Y. S. Tan, "Numerical investigation and identification of susceptible sites of atherosclerotic lesion formation in a complete coronary artery bypass model," *Medical and Biological Engineering and Computing*, vol. 46, no. 7, pp. 689–699, 2008.
- [10] Y. Huo, T. Wischgol, and G. S. Kassab, "Flow patterns in three-dimensional porcine epicardial coronary arterial tree," *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 293, no. 5, pp. H2959–H2970, 2007.
- [11] A. M. Malek, S. L. Alper, and S. Izumo, "Hemodynamic shear stress and its role in atherosclerosis," *Journal of the American Medical Association*, vol. 282, no. 21, pp. 2035–2042, 1999.
- [12] M. Khakpour and K. Vafai, "Critical assessment of arterial transport models," *International Journal of Heat and Mass Transfer*, vol. 51, no. 3-4, pp. 807–822, 2008.
- [13] Ö. Smedby, "Do plaques grow upstream or downstream?: an angiographic study in the femoral artery," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 17, no. 5, pp. 912–918, 1997.
- [14] A. Feintuch, P. Ruengsakulrach, A. Lin et al., "Hemodynamics in the mouse aortic arch as assessed by MRI, ultrasound, and numerical modeling," *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 292, no. 2, pp. H884–H892, 2007.
- [15] K. Jozwik and D. Obidowski, "Numerical simulations of the blood flow through vertebral arteries," *Journal of Biomechanics*, vol. 43, no. 2, pp. 177–185, 2010.
- [16] T. Kim, A. Y. Cheer, and H. A. Dwyer, "A simulated dye method for flow visualization with a computational model for blood flow," *Journal of Biomechanics*, vol. 37, no. 8, pp. 1125–1136, 2004.
- [17] J. T. Dodge Jr., B. G. Brown, E. L. Bolson, and H. T. Dodge, "Lumen diameter of normal human coronary arteries: influence of age, sex, anatomic variation, and left ventricular hypertrophy or dilation," *Circulation*, vol. 86, no. 1, pp. 232–246, 1992.
- [18] R. Torii, N. B. Wood, A. D. Hughes et al., "A computational study on the influence of catheter-delivered intravascular probes on blood flow in a coronary artery model," *Journal of Biomechanics*, vol. 40, no. 11, pp. 2501–2509, 2007.
- [19] J. K. Li, *Dynamic of the Vascular System*, World Scientific Publishing, Hackensack, NJ, USA, 2004.
- [20] R. M. Nerem, W. A. Seed, and N. B. Wood, "An experimental study of the velocity distribution and transition to turbulence in the aorta," *Journal of Fluid Mechanics*, vol. 52, no. 1, pp. 14–1972, 1972.

Research Article

Noise Suppression in ECG Signals through Efficient One-Step Wavelet Processing Techniques

E. Castillo, D. P. Morales, A. García, F. Martínez-Martí, L. Parrilla, and A. J. Palma

Department of Electronics and Computer Technology, ETSIIT, University of Granada, C/ Daniel Saucedo Aranda, 18071 Granada, Spain

Correspondence should be addressed to D. P. Morales; diegopm@ugr.es

Received 16 March 2013; Accepted 13 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 E. Castillo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper illustrates the application of the discrete wavelet transform (DWT) for wandering and noise suppression in electrocardiographic (ECG) signals. A novel one-step implementation is presented, which allows improving the overall denoising process. In addition an exhaustive study is carried out, defining threshold limits and thresholding rules for optimal wavelet denoising using this presented technique. The system has been tested using synthetic ECG signals, which allow accurately measuring the effect of the proposed processing. Moreover, results from real abdominal ECG signals acquired from pregnant women are presented in order to validate the presented approach.

1. Introduction

Electrocardiogram (ECG) is a valuable technique that has been in use for over a century, not only for clinic applications [1]. ECG acquisition from the human skin involves the use of high gain instrumentation amplifiers. This fact makes the ECG signal to be contaminated by different sources of noise [2]. This circumstance is highlighted when the target is the measurement of fetal ECG signals acquired over the mother's abdomen [3]. For processing ECG signals, it is necessary to remove contaminants from these signals that make visual inspection and ECG feature extraction difficult. In general, ECG contaminants can be classified into different categories, including power line interference, electrode pop or contact noise, patient-electrode motion artifacts, electromyographic (EMG) noise, and baseline wandering. Among these noises, the power line interference and the baseline wandering (BW) are the most significant and can strongly affect ECG signal analysis. Except for these two noises, other noises may be wideband and usually involve a complex stochastic process, which also distorts the ECG signal. The power line interference is narrow-band noise centered at 50 Hz or 60 Hz with a bandwidth of less than 1 Hz. Usually the ECG signal acquisition analog hardware can remove the power line interference. However, the baseline wandering and other

wideband noises are not easy to be suppressed by analog circuits. Instead, the software scheme is more powerful and feasible for portable ECG signal processing. Thus, denoising this type of signals is decisive for further parameter extraction in clinic applications.

Wavelet transform (WT) [4] is a useful tool for a variety of signal processing [5, 6] and compression applications [7, 8]; its primary, and most advantageous, application areas are those that have to generate or process wideband signals. This transform produces a time-frequency decomposition of the signal under analysis, which separates individual signal components more effectively than the traditional Fourier analysis. This fact makes WT one of the most used tools for biosignal processing, with ECG being an obvious candidate for this type of analysis. Discrete wavelet transform (DWT) provides a multiresolution, analysis, which allows representing a signal by a finite sum of components at different resolutions so that each component can be processed adaptively based on the objectives of the application.

This paper proposes an arrangement of discrete wavelet transform structures for ECG signal processing on portable, embedded computing real-time implementations [9], focusing on the suppression of different types of noise including DC levels and wandering. This suppression is carried out with a one-step wavelet processing, which reduces computing cost.

Appropriate software models and parameter selection are presented based on mathematical analysis for the mentioned ECG digital processing. Moreover, these software models will also enable very quick tests of the required signal processing. In order to select an appropriate method for the development of the software modeling, noise suppression and feature extraction procedures on ECG signals are studied in the following.

2. Materials and Methods

2.1. DWT Fundamentals. Wavelet transform (WT) is one of the most used tools in multiresolution signal analysis due to its ability to decompose a signal at various resolutions, which allow observing high-frequency events of short duration in nonstationary signals [10]. The continuous WT of a signal $f(t) \in L^2(\mathbb{R})$ is defined as [11]

$$W(a, b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) dt, \quad (1)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right). \quad (2)$$

In (2), $*$ denotes complex conjugate, and a is a scale factor and b a translation factor. The normalization factor $a^{-1/2}$ is included so that $\|\psi_{a,b}\| = \|\psi\|$. Thus, $\psi_{a,b}(t)$ represents a shifted and scaled version of the so-called mother wavelet ψ , which is a window function that defines the basis for the wavelet transformation. A mother wavelet ψ of order m is a function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ which satisfies the following four properties.

- (1) If $m > 1$, then ψ is $(m - 1)$ times differentiable.
- (2) $\psi \in L^\infty(\mathbb{R})$. If $m > 1$, for each $j \in \{1, \dots, m - 1\}$, $\psi^{(j)} \in L^\infty(\mathbb{R})$.
- (3) ψ and all its derivatives up to order $(m - 1)$ decay rapidly: for each $r > 0$ there is a $\gamma > 0$ such that

$$|\psi^{(j)}(t)| < \frac{1}{t^\gamma}, \quad j \in \{1, \dots, m - 1\} \text{ for each } |t| > \gamma. \quad (3)$$

- (4) For each $j \in \{0, 1, \dots, m\}$, $\int t^j \psi(t) dt = 0$.

Practical applications rely on the discrete wavelet transform (DWT), since it provides enough information while requiring reasonable computation time and resources. The DWT of a discrete function $f(n)$ is defined as

$$W(a, b) = C(j, k) = \sum_n f(n) \psi_{j,k}(n), \quad (4)$$

where

$$\psi_{j,k}(n) = 2^{-j/2} \psi(2^{-j}n - k) \quad (5)$$

and $a = 2^j$ and $b = k2^j$, with $j, k \in \mathbb{Z}$. Thus, DWT is computed in practice through a set of two FIR-like filters, lowpass and highpass, at each decomposition level i , followed by a downsampling by 2, which implies a reduction in

the sampling frequency. The result of the low-pass filter is usually known as the approximation and is used as input to the following decomposition level, while the result of the high-pass filter is called the detail. Thus, the usual DWT corresponds to the scheme in Figure 1(a), with

$$a_n^{(i)} = \sum_{k=0}^{N-1} g_k a_{2n-k}^{(i-1)} \quad i = 1, 2, \dots, J, \quad (6)$$

$$d_n^{(i)} = \sum_{k=0}^{N-1} h_k a_{2n-k}^{(i-1)} \quad a_n^{(0)} \equiv x_n$$

for an N -sample input sequence. The coefficients of the low-pass and high-pass filters are defined by the family of wavelet functions used as basis for the transformation: Haar, Daubechies, Quadratic Spline, and so forth [11].

Since every detail is the result of high-pass filtering of the previous approximation and approximations are the result of low-pass filtering, these details and approximations at every decomposition level contain information at different frequencies and time scales; this reflects the multiresolution analysis of the WT. Moreover, it is possible to reconstruct the original signal from the set of details and approximations, through the inverse DWT, so

$$\hat{a}_m^{(i-1)} = \begin{cases} \sum_{k=0}^{(N/2)-1} \bar{g}_{2k} \hat{a}_{(m/2)-k}^{(i)} + \sum_{k=0}^{(N/2)-1} \bar{h}_{2k} \hat{a}_{(m/2)-k}^{(i)} & m \text{ even} \\ \sum_{k=0}^{(N/2)-1} \bar{g}_{2k+1} \hat{a}_{((m-1)/2)-k}^{(i)} + \sum_{k=0}^{(N/2)-1} \bar{h}_{2k+1} \hat{a}_{((m-1)/2)-k}^{(i)} & m \text{ odd,} \end{cases} \quad (7)$$

where, once more, the corresponding details are high-pass filtered and the approximations are low-pass filtered, this time through the appropriate reconstruction filters. Thus, the reconstructed approximation at level $i - 1$ is obtained by addition of the output of the reconstruction filters and an upsampling by 2, as shown in Figure 1(b).

2.2. DWT-Based ECG Processing. The special characteristics of ECG signals and their frequency spectrum make the use of traditional Fourier analysis for the detection of ECG features difficult [12]. Wavelet transform can be applied in many fields, but its primary, and most advantageous, application areas are those that have, generate, or process wideband signals. This is due to the multiresolution analysis that wavelet provides, which allows representing a signal by a finite sum of components at different resolutions, so that each component can be processed adaptively based on the objectives of the application. In this way, this technique represents signals compactly and in several levels of resolution, which is ideal for decomposition and reconstruction purposes. Thus, the

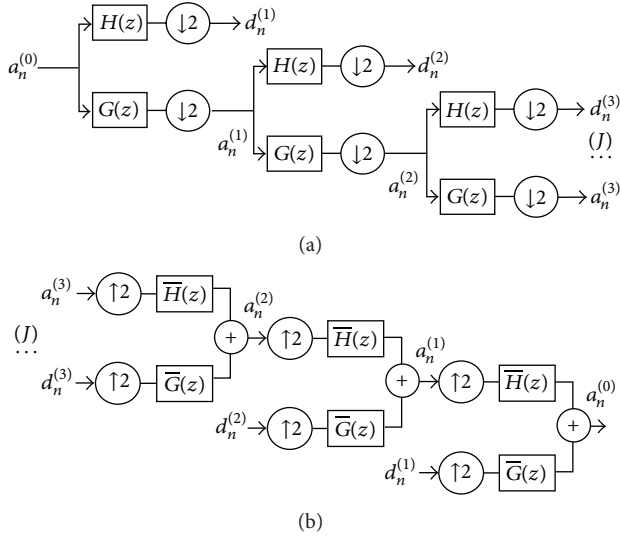


FIGURE 1: Block diagram of the multiresolution decomposition using the wavelet transform, (a) analysis filter bank and (b) reconstruction filter bank.

discrete wavelet transform is an effective way to digitally remove noises within specific subbands for ECG signals.

Several wavelet families have been proposed for ECG processing [12]. Selecting the right wavelet for this application is a task requiring at least a brief discussion. The wavelet type to use for the discrete wavelet analysis is an important decision for this processing. Singh and Tiwari [13] studied different wavelet families and analyzed the associated properties. We can distinguish between two types: the orthogonal wavelets as Haar, Daubechies (db x), Coiflets (coif x), Symlets (sym x), and biorthogonal (bior x), where x indicates the order of the wavelet and the higher the order, the smoother the wavelet. The orthogonal wavelets are not redundant and are suitable for signal or image denoising and compression. A biorthogonal wavelet is a wavelet where the associated wavelet transform is invertible but not necessarily orthogonal. Designing biorthogonal wavelets allows more degrees of freedom than orthogonal wavelets. One of these additional degrees of freedom is the possibility to construct symmetric wavelet functions. The biorthogonal wavelets usually have the linear phase property and are suitable for signal or image feature extraction.

2.2.1. Wavelet-Based Wandering Suppression. Baseline wandering (BW) usually comes from respiration at frequencies varying between 0.15 and 0.3 Hz. Different methods for wandering suppression exist, will the common objective of them to make the resulting ECG signals contain as little baseline wandering information as possible, while retaining the main characteristics of the original ECG signal. One of the proposed methods consists in high-pass digital filtering; for example, a Kaiser Window FIR high-pass filter [14] could be designed, where appropriate specifications of the high-pass filter should be selected to remove the baseline wandering.

In addition to digital filters, the wavelet transform can also be used to remove the low frequency trend of a signal. This wavelet-based approach is better because it introduces no latency and less distortion than the digital filter-based approach. The required steps for the application of this wavelet-based processing for baseline wander correction are the following.

- (1) Decomposition: apply wavelet transform to the signal up to a certain level L in order to produce the wavelet approximation coefficients $a_n^{(L)}$ that captures the baseline wander.
- (2) Zeroing approximation coefficients: replace the approximation vector $a_n^{(L)}$ for an all-zero vector to subtract this part from the raw ECG signal and remove the wandering.
- (3) Reconstruction: compute wavelet reconstruction, based on these zeroing approximation coefficients of level L and the detail coefficients of levels from 1 to L , in order to obtain the BW corrected signal.

The main idea is to remove the low frequency components, which better estimate the baseline wander. This processing is easy to carry out using wavelet decomposition, for which it is necessary to select the proper wavelet function and resolution level.

2.2.2. Wavelet-Based Denoising. The goal in ECG denoising is to try to recover the clean ECG from the undesired artifacts with minimum distortion. The recovered signal is called denoised signal, and it allows further ECG processing, such as in the case of separation of fetal ECG [3], QRS complex detection, and parameter estimation (such as the cardiac rhythm). The underlying model for the noisy signal is basically of the following form:

$$s(n) = f(n) + \sigma e(n), \quad (8)$$

where $s(n)$ represents the noisy signal, $f(n)$ is an unknown, deterministic signal, time n is equally spaced, and σ is a noise level. In the simplest model we suppose that $e(n)$ is a Gaussian white noise $N(\mu, \sigma^2) = N(0, 1)$. The denoising objective is to suppress the noise part of the noisy ECG signal and to recover the clean ECG. From a statistical viewpoint, the model is a regression model over time, and the method can be viewed as a nonparametric estimation of the function f using orthogonal basis.

Wavelet denoising has emerged as an effective method requiring no complex treatment of the noisy signal [15]. It is due to the sparsity, locality and multiresolution nature of the wavelet transform. The wavelet transform localizes the most important spatial and frequential features of a regular signal in a limited number of wavelet coefficients. Moreover, the orthogonal transform of stationary white noise results in stationary white noise. Thus, in the wavelet domain the random noise is spread fairly uniformly among all detail coefficients. On the other hand, the signal is represented by a small number of nonzero coefficients with relatively larger values. This sparsity property assures that wavelet

shrinkage can reduce noise effectively while preserving the sharp features (peaks of QRS complex). The general wavelet denoising procedure basically proceeds in three steps [11].

- (1) Decomposition: apply wavelet transform to the noisy signal $s(n)$ to produce the noisy wavelet coefficients to the level N , by which we can properly distinguish the presence of partial discharges.
- (2) Threshold detail coefficients: for each level from 1 to N , select appropriate threshold limit and apply soft or hard thresholding to the detail coefficients at some particular levels to best remove the noise.
- (3) Reconstruction: compute wavelet reconstruction, based on the original approximation coefficients of level N and the modified detail coefficients of levels from 1 to N , in order to obtain estimated/smoothed signal $\hat{f}(n)$.

Although the application of this denoising method is not conceptually complex, some issues are carefully studied and addressed in the following for getting satisfactory results.

3. One-Step DWT-Based Baseline Wandering (BW) and Noise Suppression Proposal

Due to the similar wavelet structure for the application of BW and noise suppression, we propose here to apply in only one step both wavelet-based techniques. It will save important resources and/or time, which would facilitate any future portable hardware implementation [16, 17]. The required steps for the application of this approach are as follows.

- (1) Decomposition: the wavelet decomposition is applied down to a certain level L in order to produce the approximation coefficients $a_n^{(L)}$ that capture the BW.
- (2) Zeroing approximations: the approximation $a_n^{(L)}$ is replaced by all-zero vector.
- (3) Threshold details: the level M (with $M < L$) allowing to properly distinguish the presence of partial discharges in the noisy details must be selected. Additionally, for each level from $i = 1$ to M , the appropriate threshold limit and rule (soft or hard) are applied to the detail coefficients $d_n^{(i)}$ for better removing the noise.
- (4) Reconstruction: the wavelet reconstruction, based on the zeroing approximations of level L , the modified details of levels 1 to M , and original details of levels from $M + 1$ to L , is computed to obtain the BW corrected and denoised signal.

Thus, simultaneous BW and noise suppressions are easy to get using this wavelet-based technique, for which it is necessary to select the proper parameters. One of the aims of this work is to develop a mathematical processing model oriented to portable hardware implementation. Thus, a deep analysis of the involved parameters is detailed in the following.

3.1. Analysis of Parameters

3.1.1. Wavelet Family. The selection of the *wavelet family* has to be based on the similarities between the ECG basic structure and the wavelet functions and the type of processing to apply. For ECG wandering suppression, the selection of the wavelet family is based on the study of the wavelet that best resembles the most significant and characteristic waveform QRS of the ECG signal. Thus, the detail sequences at different levels of decomposition from 1 to L can capture and keep the detail features that are of interest for properly reconstructing the ECG without baseline wander. This is achieved through the elimination of the approximation sequence at level L . The selection of the wavelet family for ECG denoising is made in a similar way to that for wandering removing; it is also based on the different types of wavelets and their correlation to different signals. The order-6 Daubechies wavelet is a function well suited because of its similarity to an actual ECG. Other features include order-10 Symlets.

3.1.2. Resolution Level for Wandering. Another important parameter for wavelet-based wandering suppression is the decomposition level L , so that $a_n^{(L)}$ can capture the baseline wander without oversmoothing. To select this resolution level, it is important to take into account the maximum number of decomposition levels, N , which is determined by the length of the sampled ECG signal. Decomposition level must be a positive integer not greater than $\log_2(L_s)$, where L_s is the length of the signal, two different methods can be employed for selecting the resolution level.

- (i) Visual inspection: it consists of plotting the approximation sequences for different decomposition levels and to select the resolution level whose approximation better captures the baseline wandering. ECG signal from lead 4 of the DaISy dataset [18] is chosen to illustrate this method. This dataset contains 8 leads of skin potential recordings of a pregnant woman. The lead recordings, three thoracic and five abdominal, were sampled at 250 sps rate and are 10-second long. Figure 2 plots the ECG signal and approximation sequences $a_n^{(i)}$ for $i = 5$ to 9. This figure shows that $a_n^{(6)}$ and $a_n^{(7)}$ are the approximation sequences better capturing the baseline wander. However, this method is not effective for real-time processing.
- (ii) Analytical calculation: this method is based on the calculation of the resolution level for wander suppression as follows. Wavelet decomposition uses half band low-pass and high-pass filters. As it was commented in Section 2, multiresolution decomposition allows representing a signal by a finite sum of components at different resolutions. Each decomposition level contains information at different frequency bands and time scales. The approximation sequence at level i , $a_n^{(i)}$, is decomposed to obtain approximation and detail at level $i + 1$, $a_n^{(i+1)}$, and $d_n^{(i+1)}$. Specifically, approximation $a_n^{(i+1)}$ is the result of low-filtering approximation $a_n^{(i)}$ and a downsampling by 2, so that

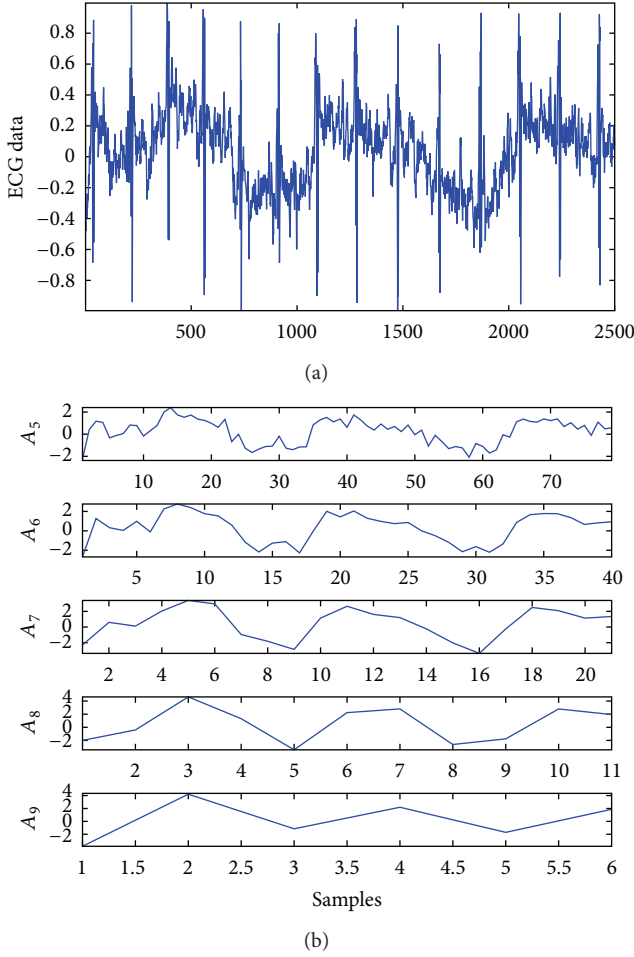


FIGURE 2: ECG data with BW (lead 4 DaSy dataset) and approximation sequences at decomposition levels from 5 to 9.

frequencies in $a_n^{(i)}$ that are above half of the highest frequency component are removed. This means that at every decomposition level the frequency band of the approximation is reduced to the half. Figure 3 shows a study of the frequency subbands of wavelet decomposition of DaSy dataset lead 4. This figure displays approximation FFTs for resolution levels for $i = 1$ to 7. Let us consider that the most important frequency bands in baseline wander are below a certain frequency f_c . For example, $f_c = 1$ Hz for wandering coming from respiration (0.15–0.3 Hz) [19]. Other wandering components such as motion of the patients and instruments have higher frequencies components. To remove wandering, it should be necessary to select the resolution level such as the approximation captures the ECG components for frequencies lower than this f_c . The decomposition level for wandering suppression can be calculated as follows:

$$L_{BW} = \text{int} \left[\log_2 \left(\frac{F_{\max}}{f_c} \right) \right], \quad (9)$$

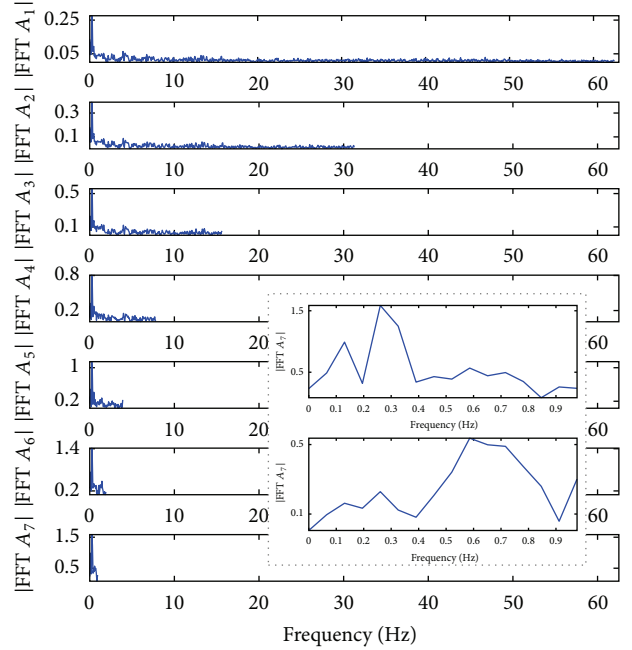


FIGURE 3: FFT of ECG DaSy dataset lead 4 ECG1 and FFTs of approximation sequences, at decomposition levels from 1 to 7 and zoom to FFTs approximation and detail level 7.

where F_{\max} is the highest frequency component. For example, the sampling frequency for DaSy dataset is 250 Hz, so $F_{\max}=125$ Hz (Nyquist Theorem) and for $f_c = 1$ Hz from (9) we can obtain $L_{BW} = 7$. After applying seven low-pass filters and downsampling processes, $a_n^{(7)}$ captures frequencies from 0 Hz to 0.977 Hz and is a good estimation of baseline wander. The visual inspection method also determined that $a_n^{(7)}$ captures the baseline wander. The BW corrected signal can be obtained using wavelet reconstruction based on the detail coefficients of levels from 1 to 7 and zeroing $a_n^{(7)}$. Figure 4 illustrates the proposed method for wavelet-based wandering suppression. This figure shows the original ECG signal (DaSy dataset lead 4), the estimated baseline wander with $a_n^{(7)}$, and the ECG signal obtained after zeroing $a_n^{(7)}$.

3.1.3. Level of Decomposition for Denoising. The maximum level for detail thresholding, M , depends on several factors such as the SNR in the original signal or the sample rate. The level of decomposition specifies the number of levels in the discrete wavelet analysis to take into account for detail thresholding. Unlike conventional techniques, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy will depend on the signal and experience. Most often, the level is chosen based on a desired low-pass cut-off frequency. Measured signal having lower SNR usually needs more levels of wavelet transform to remove most of its noise. On the other hand, a factor that also influences the optimal

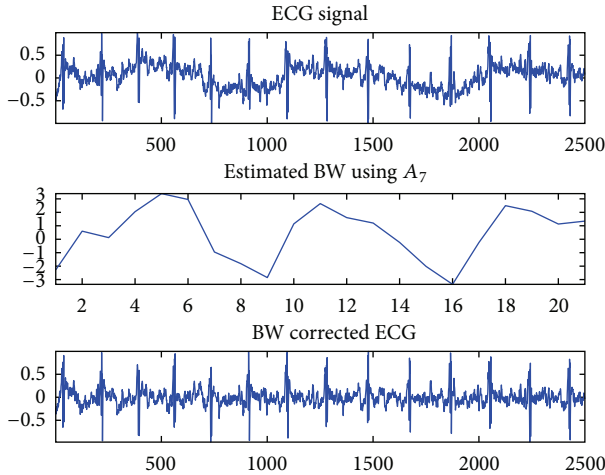


FIGURE 4: Original ECG signal (DaSy dataset lead 4), estimated BW from A_7 , and BW corrected signal.

level decomposition for denoising is the sampling rate of ECG signal. Sharma et al. [15] use a study of the spectrum of the detail coefficient at each level for estimating the optimum level decomposition for denoising. They concluded that noise content is significant in high frequency detail subbands, while most of the spectral energy lies in low frequency subbands. For an ECG signal sampled at $F_s = 500$ Hz, noise content is significant in detail subbands $d_n^{(1)}$, $d_n^{(2)}$, and $d_n^{(3)}$. Thus, in order to avoid losing clinically important components of the signal, such as PQRST morphologies, only detail $d_n^{(1)}$, $d_n^{(2)}$, and $d_n^{(3)}$ should be treated for denoising. Manikandan and Dandapat [20] present a wavelet energy-based diagnostic distortion measure to assess the reconstructed signal quality for ECG compression algorithms. Their work includes a study showing, for a given sampling frequency, the information of the ECG signal and its energy contribution at each frequency subband. This helps to understand how the noise perturbs the detail coefficients and the effect over the energy spectrum. The authors also show several examples that clarify the effect of applying “zeroing” of detail coefficients at different decomposition levels. As an example to clarify the correct selection of this parameter, Figure 5 shows 4th level wavelet decomposition for ECG signal DaSy dataset lead 4. It can be observed that small subbands reflect the high frequency components of the signal, and large subbands reflect the low frequency components of the signal. The effect of high frequency artifacts can be seen in detail subbands $d_n^{(1)}$ and $d_n^{(2)}$. These bands are weighted with small values because the energy contribution to the spectrum is low. For this example, it should be advisable to treat only detail subbands $d_n^{(1)}$ and $d_n^{(2)}$ for wavelet denoising in order to maintain the main features of the ECG signal. It is due, as the figure shows, to the fact that detail subbands higher than level 2 contain most of the significant information for diagnostic. A more detailed study for the determination of the optimum level for wavelet-based denoising is shown in Section 4.

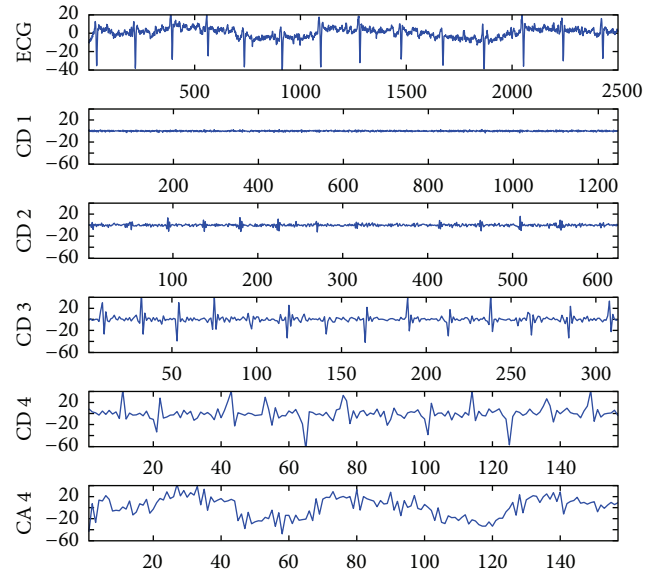


FIGURE 5: fourth level wavelet decomposition for DaSy dataset lead 4.

3.1.4. Threshold Limits. Most algorithms are based on the previous threshold definition established in Donogo’s universal theory [21]. Since this work, modified versions of the universal threshold and new thresholds have been proposed. Matlab functions for denoising (*wden.m*, *thselect.m*) [22] and Labview blocks (Wavelet Denoise Express VI) [23] establish some of these thresholds as predefined options: *rigsure*, *sqtwlog*, *heursure*, and *minimax*. Taking this into account, the aim of this work is to develop processing models for portable computing implementations, so a study of the proposed threshold based on the following criteria was made: computational complexity, delays, latency, and clock frequency. The evaluation of these criteria derives the following threshold classification.

Prefixed Thresholds. For a given signal length, N , and a level of decomposition for wavelet denoising, J , it will be possible to know the coefficient length for each level j , denoted by n_j , and thus, some thresholds proposed in the literature [13] could be calculated using software tools and stored in memory for later usage. Some of them are the following:

- (i) universal threshold:

$$\text{Th}_{\text{uni}} = \sqrt{2 \log N}, \quad (10)$$

- (ii) universal threshold level dependent:

$$\text{Th}_{\text{uni},j} = \sqrt{2 \log n_j}, \quad (11)$$

- (iii) universal modified threshold level dependent:

$$\text{Th}_{\text{uni, mod},j} = \frac{\sqrt{2 \log n_j}}{\sqrt{n_j}}, \quad (12)$$

(iv) exponential threshold:

$$\text{Th}_{\text{exp}} = 2^{((j-1)/2)} \sqrt{2 \log N}, \quad (13)$$

(v) exponential threshold level dependent:

$$\text{Th}_{\text{exp},j} = 2^{((j-1)/2)} \sqrt{2 \log n_j}, \quad (14)$$

(vi) minimax threshold (RefMatlab):

$$\text{Th}_{\text{minimax}} = 0.3936 + 0.1829 * \left(\frac{\log(n_j)}{\log(2)} \right). \quad (15)$$

Minimax threshold uses the minimax principle to estimate the threshold [24].

Nonprefixed Thresholds. There are more threshold proposals that have been positively evaluated [15, 22] and, in some cases, get better results than the prefixed thresholds. However, a previous software calculation would not be possible since samples and/or coefficient data which are initially unknown are needed. On the other hand, the estimation of these thresholds could imply complex operations, and thus an important increment of latency. The following are two examples of this type of thresholds:

(i) maxcoef threshold [22]:

$$\begin{aligned} \text{Th}_{\text{maxcoef}} &= 2^{n-J}, \\ n &= \text{round} \left[\log_2 \left(\max \{ |c_j| \} \right) \right], \end{aligned} \quad (16)$$

(ii) Kurtosis and ECE-based thresholds [15]:

$$\begin{aligned} \text{Th}_{\text{DF},j} &= \frac{1}{\epsilon_j} \times \frac{\max(c_j)}{F_{\text{JSN}}}, \\ \text{Th}_{\widehat{\text{DF}},j} &= \frac{1}{\widehat{\epsilon}_j} \times \frac{\max(c_j)}{F_{\text{JSN}}}, \end{aligned} \quad (17)$$

where ϵ_j is the energy contribution efficiency of j th detail subband, $\widehat{\epsilon}_j$ is detail energy contribution efficiency of j th wavelet subband, and F_{JSN} is the ratio between the Kurtosis value of the signal at subband j to the Kurtosis value of Gaussian noise:

$$\begin{aligned} \epsilon_j &= \frac{E_j}{E_t} \times 100, \\ \widehat{\epsilon}_j &= \frac{E_j}{\widehat{E}_t} \times 100, \end{aligned} \quad (18)$$

$$K' = \frac{m_4}{m_2^2} = \frac{(1/N) \sum_{i=1}^N (x_i - \text{media}(x))^2}{\left((1/N) \sum_{i=1}^N (x_i - \text{media}(x))^2 \right)^2}.$$

Analyzing this type of thresholds, their data and coefficients dependence makes them different from prefixed thresholds. On the hand, the implementation of the involved operations of these nonprefixed thresholds requires a high number of operations, division between them. Thus increasing implementation complexity of these thresholds. On the other hand, it will consume a big number of clock cycles and at least two accesses to each stored sample. It could have devastating effects over the total latency and real-time processing.

3.1.5. Threshold Rescaling. For signal denoising, once the threshold to be applied is selected, this threshold is rescaled using noise variance:

$$\text{th}_{j,\text{rescaled}} = \sigma_j \cdot \text{th}_j. \quad (19)$$

The noise variance is used to rescale the threshold at each level, so other important setting is related to the method for estimating the noise variance at each level. If the noise is white, the standard deviation from the wavelet coefficients at the first level can be used, and the thresholds can be updated using this value. If the noise is not white, best results for denoising are obtained when estimating the noise standard deviation at each level independently and using each one to rescale the associated wavelet coefficients.

The σ is calculated based on median absolute deviation (MAD) [15, 25, 26]:

$$\sigma = C \cdot \text{MAD}(c_j), \quad (20)$$

where C is a constant scale factor, which depends on distribution of noise. For normally distributed data $C = 1.4826$ is the 75th percentile of the normal distribution with unity variance. On the other hand, for a univariate dataset of wavelet details at j th level $c_{j1}, c_{j2}, \dots, c_{jm}$, the MAD is defined as [27]

$$\text{MAD} = \text{median} \left(|c_j - \text{median}(c_j)| \right). \quad (21)$$

Thus, estimated noise variance can be written as

$$\sigma = 1.4826 \cdot \text{median} \left(|c_j - \text{median}(c_j)| \right). \quad (22)$$

Some predefined applications or functions, such as the *wnoisest.m* Matlab function, use a simpler expression for σ :

$$\sigma = 1.4826 \cdot \text{median} \left(|c_j| \right). \quad (23)$$

It must be noted that the rescale factor σ is not a prefixed value since, as (21) and (22) show, it depends on the wavelet coefficient values. However, memory access is reduced if expression (22) is used (expression (21) implies three memory accesses, while (22) requires only one). Depending on the

TABLE 1: BW suppression analysis.

F_s (Hz)	N	F_w (Hz)	SNR for estimated BW								
			A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}
250	5400	0.15	1.866	6.563	12.579	25.986	25.484	7.826	-7.957	-8.948	-9.199
		0.19	1.866	6.563	12.570	25.636	22.445	-0.216	-8.630	-9.072	-9.193
		0.23	1.866	6.563	12.555	24.591	16.422	2.536	-8.981	-9.003	-9.035
		0.27	1.866	6.563	12.548	23.882	10.589	-7.621	-8.874	-8.943	-8.977
		0.31	1.866	6.563	12.561	24.063	6.442	-8.311	-8.886	-9.212	-9.339
500	10800	0.15	0.612	4.491	10.172	16.583	30.755	24.674	5.845	-9.762	-10.373
		0.19	0.612	4.491	10.170	16.541	29.129	21.425	-2.123	-10.440	-10.866
		0.23	0.612	4.491	10.168	16.485	26.036	14.864	0.825	-10.778	-10.796
		0.27	0.612	4.491	10.168	16.469	25.625	8.869	-9.412	-10.669	-10.750
		0.31	0.612	4.491	10.170	16.521	25.237	4.688	-10.16	-10.726	-10.064
1000	21600	0.15	0.104	1.040	5.684	11.679	18.344	33.993	23.436	4.831	-10.805
		0.19	0.104	1.040	5.684	11.678	18.319	31.447	20.792	-3.154	-11.545
		0.23	0.104	1.040	5.684	11.677	18.271	26.531	13.913	-0.238	-11.897
		0.27	0.104	1.040	5.684	11.677	18.255	24.680	7.822	-10.512	-11.771
		0.31	0.104	1.040	5.684	11.679	18.300	25.393	3.629	-11.258	-11.807

TABLE 2: Denoising evaluation using synthetic ECG signals.

M	Threshold	Wavelet	SNR	SNRIMP	M	Threshold	Wavelet	SNR	SNRIMP
Soft thresholding 15 dB					Hard thresholding 15 dB				
3	Th _{exp}	coif3	21.2631	6.4015	4	Th _{minimax}	sym7	20.8648	5.9319
3	Th _{exp}	db7	21.4798	6.3944	4	Th _{minimax}	sym7	21.0110	5.9256
3	Th _{exp}	coif3	21.4773	6.3919	4	Th _{minimax}	sym7	20.7633	5.9017
3	Th _{exp}	db7	21.2360	6.3743	5	Th _{minimax}	sym7	20.8288	5.8958
3	Th _{exp}	sym7	21.4286	6.3432	5	Th _{minimax}	sym7	20.7219	5.8603
3	Th _{exp}	db5	21.4274	6.3420	5	Th _{minimax}	sym7	20.9024	5.8170
3	Th _{exp}	db5	21.2008	6.3391	3	Th _{minimax}	sym7	20.8818	5.7964
Soft thresholding 25 dB					Hard thresholding 25 dB				
3	Th _{exp}	bior3.9	29.5524	4.4960	3	Th _{minimax}	bior6.8	29.5025	4.5809
3	Th _{exp}	bior3.9	29.2865	4.3649	3	Th _{minimax}	bior6.8	29.6610	4.5534
3	Th _{exp}	sym7	29.2657	4.3440	4	Th _{minimax}	bior6.8	29.5801	4.4725
3	Th _{exp}	sym7	29.3678	4.3114	4	Th _{minimax}	bior6.8	29.3823	4.4607
3	Th _{exp}	bior3.9	29.3712	4.2784	3	Th _{minimax}	bior6.8	29.5049	4.4485
3	Th _{exp}	bior6.8	29.2479	4.1915	3	Th _{minimax}	sym6	29.5035	4.4471
3	Th _{minimax}	bior3.9	29.0937	4.1720	5	Th _{minimax}	bior6.8	29.5436	4.4360

system architecture and data processing, a higher number of memory accesses could increase the system latency and thus jeopardize real-time processing.

3.1.6. Thresholding Rules. Thresholding can be done using soft or hard thresholding. Hard thresholding is the simplest method. It can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold and then shrinking the nonzero coefficients towards 0. The hard procedure creates discontinuities, while the soft procedure does not. The soft

and hard thresholding is shown in (24) and (25), respectively:

$$\tilde{d}_n^{(j)} = \begin{cases} \text{sign}(d_n^{(j)}) (|d_n^{(j)}| - th_j) & \text{if } |d_n^{(j)}| \geq th_j \\ 0 & \text{if } |d_n^{(j)}| < th_j, \end{cases} \quad (24)$$

$$\tilde{d}_n^{(j)} = \begin{cases} d_n^{(j)} & \text{if } |d_n^{(j)}| \geq th_j \\ 0 & \text{if } |d_n^{(j)}| < th_j, \end{cases} \quad (25)$$

where $\tilde{d}_n^{(j)}$ and $\tilde{d}_n^{(j)}$ represent the modified values of j th level detail coefficient based on the selected threshold and th_j are an approximation of the detail coefficients of the denoised transform. The new detail coefficients, $\tilde{d}_n^{(j)}$ or $\tilde{d}_n^{(j)}$, have

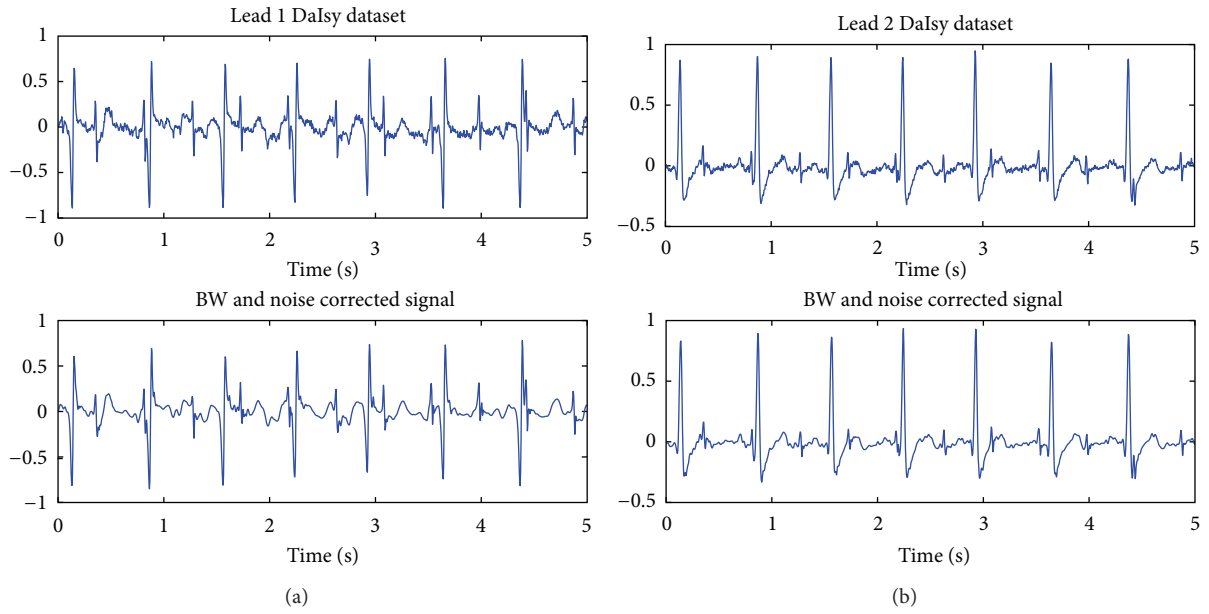


FIGURE 6: DaIisy database signals: (a) lead 1 and noise and BW corrected signal; (b) lead 2 and noise and BW corrected signals.

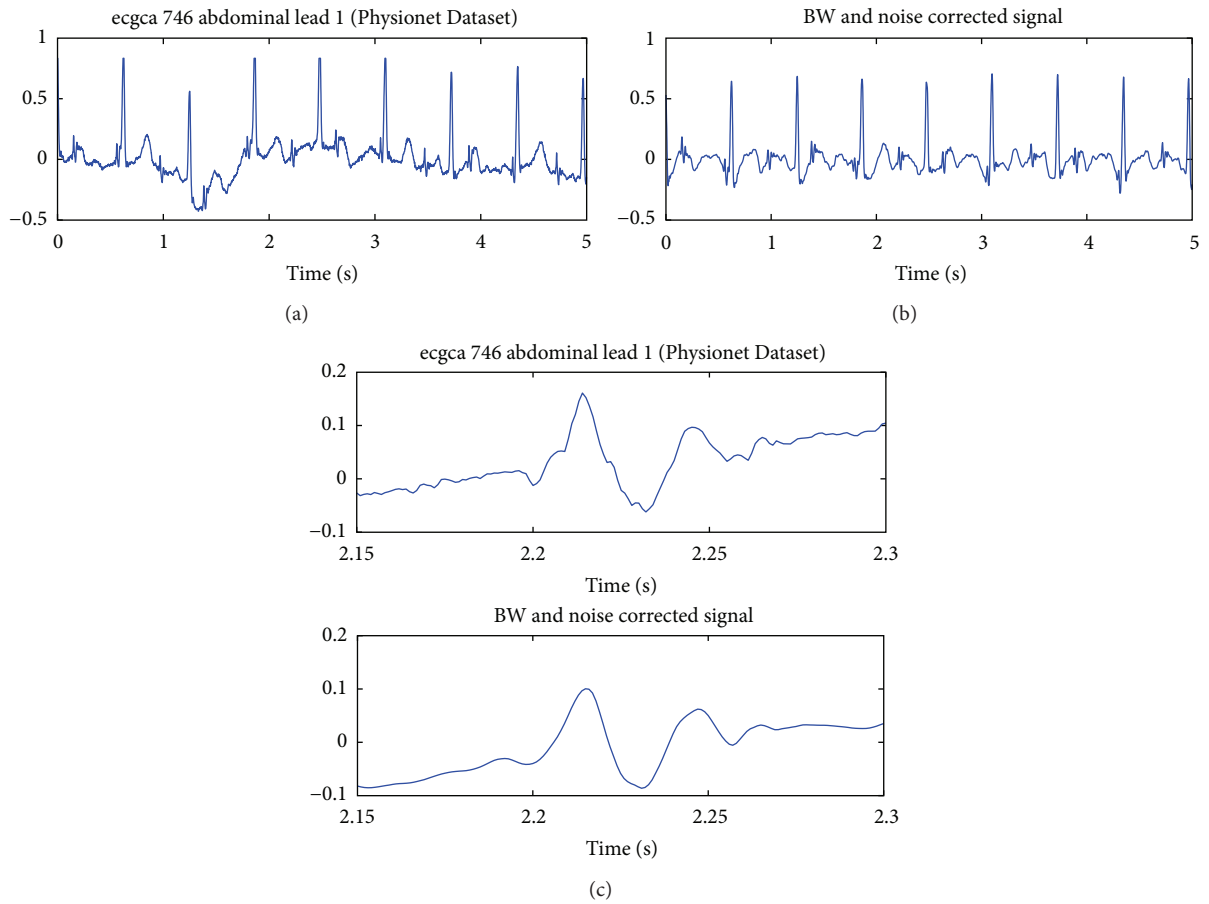


FIGURE 7: ecgca 746 abdomen lead 1 Noninvasive Fetal ECG Database (Physionet Dataset): (a) signal; (b) BW and noise corrected signal; (c) signals detail.

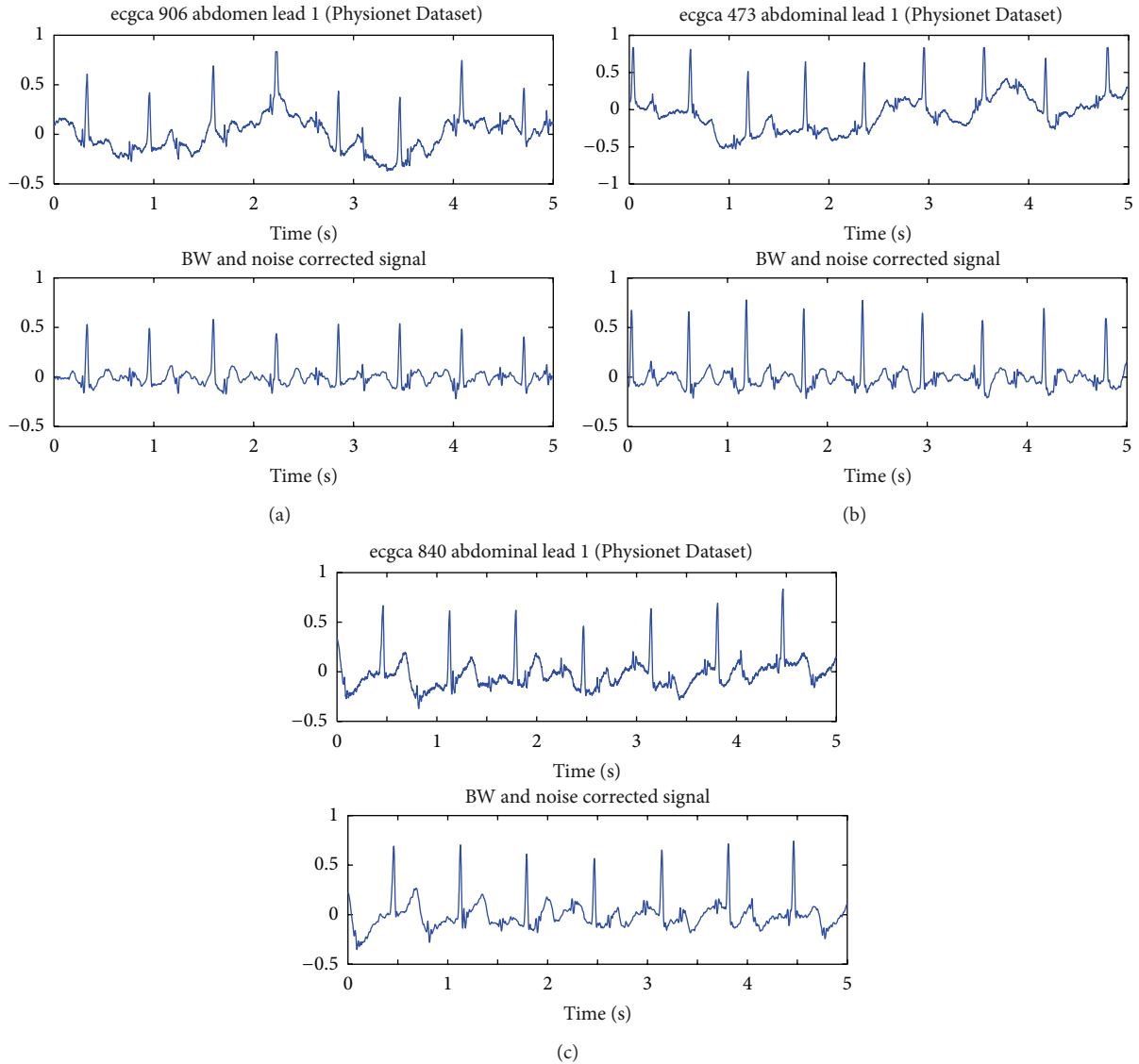


FIGURE 8: Noninvasive Fetal ECG Database (Physionet Dataset): (a) ecgca 906 lead 1 signal, BW and noise corrected signal; (b) ecgca 473 lead 1 signal, BW and noise corrected signal; (c) ecgca 840 lead 1 signal, BW and noise corrected signal.

to be calculated for the wavelet transform levels considered for denoising, as it was pointed previously.

4. Results

This section is devoted to analyze the one-step proposed method through quantitative parameters. When working with real noisy ECG signals, it is not trivial to calculate a parameter that provides a quantitative measure of the benefits of the applied technique. In order to better analyze our proposed one-step model, a separate study of BW and noise suppression has been carried out using synthetic ECG signals. For a quantitative evaluation of the BW suppression, we have employed three types of synthetic ECG signals. These synthetic signals were elaborated from the *ecg.m* Matlab function [28] and are 21.6 second long and contain 5400,

10800, and 21600 samples thus the resulting sample rates are 250, 500, and 1000 sps, respectively. Signals affected by BW are obtained adding a sine wave plus a DC level, using frequencies from 0.15 to 0.31 Hz that fit to the frequency band in real BW. Our study has estimated the BW of the signals as the approximations from level 1 to 12 and has reconstructed the signal removing the estimated BW. Table 1 resumes the main results, showing the SNRs between the synthetic and the BW corrected signal. According to the expression of L , for signals sampled at 250, 500, and 1000 sps, the adequate decomposition level for BW suppression will be 7, 8, and 9, respectively, which is corroborated by Table 1. This study also reflects that better BW suppression (higher SNR) is achieved if the signal is sampled at higher rate.

Synthetic ECG signals were also used to evaluate the performance of the noise suppression. These signals were

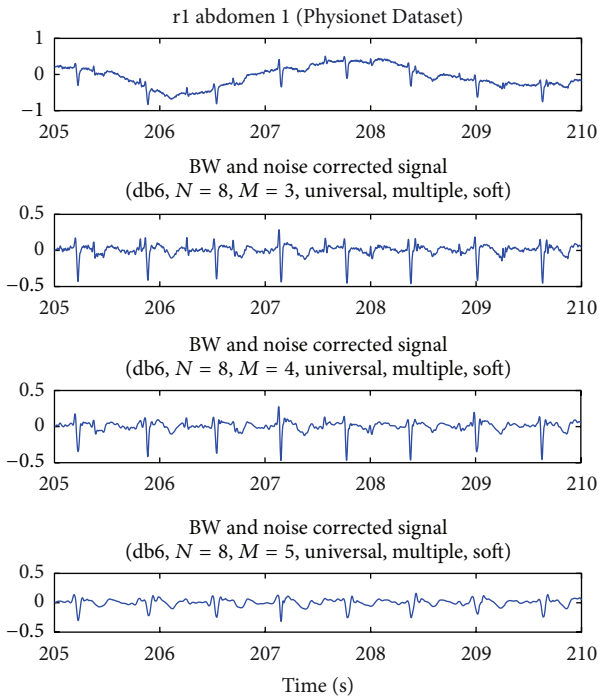


FIGURE 9: r1 abdomen 1 Abdominal and Direct Fetal Electrocardiogram Database (Physionet Dataset) signal and BW and noise corrected signals using different levels for denoising.

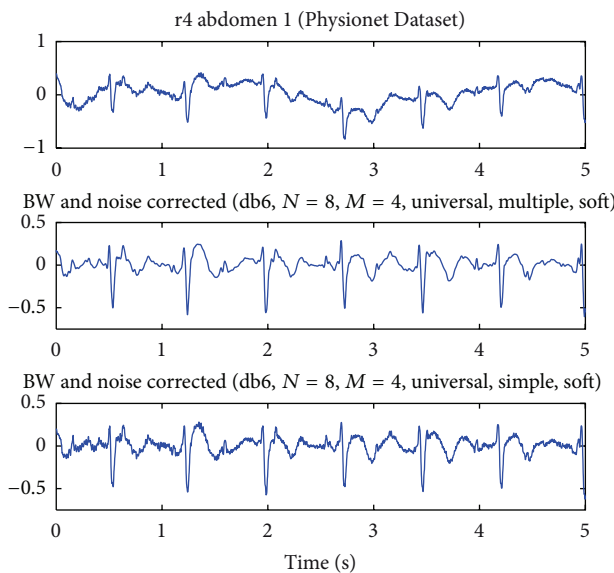


FIGURE 10: r4 abdomen 1 Abdominal and Direct Fetal Electrocardiogram Database (Physionet Dataset) signal and BW and noise corrected signals using *multiple* and *simple rescaling*.

contaminated adding Gaussian white noise. Thus, the noisy signal is processed by the proposed method to obtain the denoised signal. This scenario is used by several authors [29] and allows visual inspection and quantitative evaluation. There are several parameters to measure the quality of the denoised signal [15, 29], as the SNR Improvement Measure

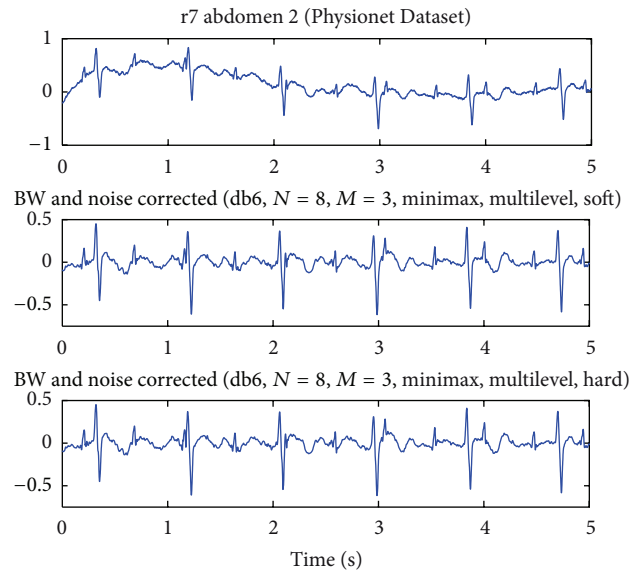


FIGURE 11: r7 abdomen 2 Abdominal and Direct Fetal Electrocardiogram Database (Physionet Dataset) signal and BW and noise corrected signals using the two types of thresholding rules.

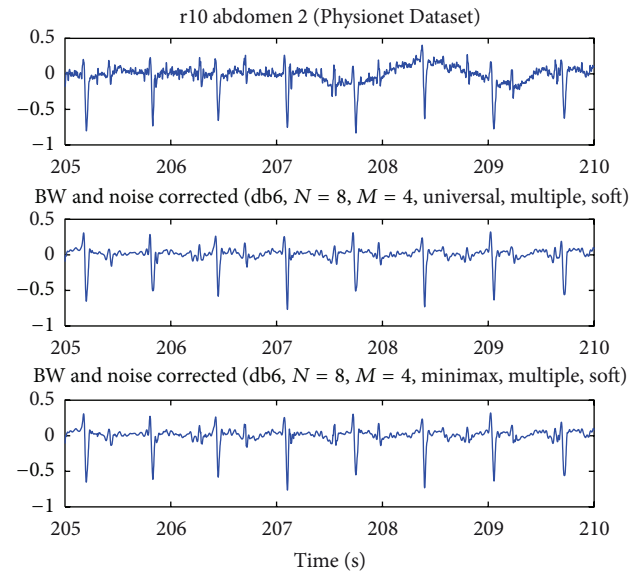


FIGURE 12: r10 abdomen 2 Abdominal and Direct Fetal Electrocardiogram Database (Physionet Dataset) signal and BW and noise corrected signals using two types of thresholds.

SNRIMP [29]. A study using 3, 4, and 5 as maximum levels for wavelet denoising M , *universal*, *exponential*, and *minimax* as threshold limits, and *simple rescaling* and *soft* and *hard thresholding* was carried out. A total of 12 wavelet functions were used for this evaluation. The study also considers two noise levels, approximately 15 dB and 25 dB and a maximum of three attempts for each case (for each attempt, all the parameters are the same, including noise level, but the noisy signal is different due to the random distribution of it over the

original signal). Table 2 shows the best results for denoising. Observing these summarized results and all the generated data, we can conclude that there are no large differences for the SNRIMP values of the different wavelet functions. Comparing *soft* and *hard thresholding*, *soft* gets for both noise levels higher SNRIMP using a less number of levels. Regarding thresholds, Th_{exp} achieves best denoising if it is used along with *soft thresholding*, as it is the case with the combination $Th_{minimax}$ and *hard thresholding*.

To study BW suppression and denoising for our one-step denoising and BW suppression proposal, visual inspection of the obtained signals is also important, and in some cases it is even more conclusive than quantitative measures. DaISy dataset and Physionet Dataset [30] are targeted for evaluating the proposed one-step BW and noise suppression. The recordings from Noninvasive Fetal ECG Database [30] have two thoracic and four abdominal channels sampled at 1 ksp/s, all 60 seconds long. The signal bandwidth is 0.01 Hz–100 Hz. In addition, recordings from the Abdominal and Direct Fetal Electrocardiogram Database [31] have been used. Each recording comprises four differential signals acquired from maternal abdomen and the reference direct fetal electrocardiogram registered from the fetus's head. The fetal R-wave locations were automatically determined in the direct FECG signal by means of online analysis applied in the KOMPOREL system [31]. The recordings, sampled at 1 ksp/s, are 5-minute long, and the signal bandwidth is 1 Hz–150 Hz.

For these signals the selected parameters were wavelet function db6, $M = 3$, *universal threshold*, *soft thresholding*, *simple rescaling* for DaISy dataset, and *multiple rescaling* for Noninvasive Fetal ECG Database. Figure 6 includes the obtained result for lead 1 and lead 2 from DaISy dataset, with BW and noise corrected signals being shown. Figure 7 shows an example of results for ecga 746 signal of Noninvasive Fetal ECG Database, including the detail of one of the fetal QRS complexes before and after processing. These figures show that the abdominal ECG signals are BW corrected and denoised while retaining their main characteristics, as the fetal QRS complexes, which are very important for future parameter extraction [7]. Figure 8 shows more examples of results for signals of Noninvasive Fetal ECG Database.

Figures 9, 10, 11, and 12 show processing examples for signals of Abdominal and Direct Fetal Electrocardiogram Database displayed. Figure 9 uses signal r1 abdomen 1 to make a study of the BW and noise corrected signals using three different levels for signal denoising, $M = 3$, $M = 4$, and $M = 5$; as this level increases, the denoised signal loses its main characteristics (i.e., fetal QRS complexes). Figure 10 shows the denoising of r4 signals using *multiple rescaling* and *simple rescaling*. It can be observed that the best denoising is obtained using *multiple rescaling*. On the other hand, Figure 11 shows the results for r7 abdomen 2 signal using *soft* and *hard* thresholding; *soft* thresholding gets better denoised signal since *hard* thresholding introduces some discontinuities into the denoised signal. Finally, *universal* and *minimax* thresholds are used for denoising r10 abdomen 2 signal, as Figure 12 shows. Similar results are obtained for these two thresholds, but observing signal details, *minimax* threshold provides better results.

5. Conclusion

This paper presents the mathematical bases for electrocardiogram signal denoising by means of discrete wavelet processing. A novel one-step wavelet-based method has been introduced performing both BW and noise suppression, which makes computationally feasible real-time implementations. The presented approach is performed by only one wavelet decomposition and reconstruction step, which is required for eliminating both types of perturbations. This approach has been linked to an exhaustive study of the related parameters, such as number of decomposition levels, threshold edges, rescaling, and rules that allow an optimal signal denoising and meeting specific ECG signal characteristics including signal shape, sample rate, and noise levels. The presented results for synthetic ECG signals validate this method, while applications on real AECG signal from three different databases have led to improved signals that are valid for further analysis and extraction of parameters such as heart rate variability. The defined algorithm will also allow its compact implementation, thus fitting portable ECG applications.

Acknowledgment

This work has been partially supported by the CEI BIOTIC Granada under project 2013/81, "Compromiso con la investigación y el desarrollo."

References

- [1] J. Lee, Y. Chee, and I. Kim, "Personal identification based on vectorcardiogram derived from limb leads electrocardiogram," *Journal of Applied Mathematics*, vol. 2012, Article ID 904905, 12 pages, 2012.
- [2] D. P. Morales, A. García, E. Castillo, M. A. Carvajal, J. Banqueri, and A. J. Palma, "Flexible ECG acquisition system based on analog and digital reconfigurable devices," *Sensors and Actuators A*, vol. 165, no. 2, pp. 261–270, 2011.
- [3] D. P. Morales, A. García, E. Castillo, M. A. Carvajal, L. Parrilla, and A. J. Palma, "An application of reconfigurable technologies for non-invasive fetal heart rate extraction," *Medical Engineering and Physics*, 2013.
- [4] S. G. Mallat, "Theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [5] E.-B. Lin and P. C. Liu, "A discrete wavelet analysis of freak waves in the ocean," *Journal of Applied Mathematics*, vol. 2004, no. 5, pp. 379–394, 2004.
- [6] T. Abualrub, I. Sadek, and M. Abukhaled, "Optimal control systems by time-dependent coefficients using cas wavelets," *Journal of Applied Mathematics*, vol. 2009, Article ID 636271, 10 pages, 2009.
- [7] R. Sameni and G. D. Clifford, "A review of fetal ECG signal processing, issues and promising directions," *The Open Pacing, Electrophysiology & Therapy Journal*, vol. 3, no. 1, pp. 4–20, 2010.
- [8] C.-T. Ku, K.-C. Hung, H.-S. Wang, and Y.-S. Hung, "High efficient ECG compression based on reversible round-off non-recursive 1-D discrete periodized wavelet transform," *Medical Engineering and Physics*, vol. 29, no. 10, pp. 1149–1166, 2007.

- [9] C. Ghule, D. G. Wakde, G. Viridi, and N. R. Khodke, "Design of portable ARM processor based ECG module for 12 lead ECG data acquisition and analysis," in *Proceedings of the 2nd International Conference on Biomedical and Pharmaceutical Engineering (ICBPE '09)*, pp. 1–8, December 2009.
- [10] D. Karadaglić, M. Mirković, D. Milošević, and R. Stojanović, "A FPGA system for QRS complex detection based on Integer Wavelet Transform," *Measurement Science Review*, vol. 11, no. 4, pp. 131–138, 2011.
- [11] J. C. Goswami and A. K. Chan, *Fundamentals of Wavelets Theory, Algorithms, and Applications*, EE.UU: John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2011.
- [12] P. S. Addison, "Wavelet transforms and the ECG: a review," *Physiological Measurement*, vol. 26, no. 5, pp. R155–R199, 2005.
- [13] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital Signal Processing*, vol. 16, no. 3, pp. 275–287, 2006.
- [14] C. B. Mbachu, G. N. Onoh, E. N. Ifeagwu, and S. U. Nnebe, "Processing ECG signal with Kaiser Window-Based FIR digital filters," *International Journal of Engineering Science and Technology*, vol. 3, no. 8, 2011.
- [15] L. N. Sharma, S. Dandapat, and A. Mahanta, "ECG signal denoising using higher order statistics in Wavelet subbands," *Biomedical Signal Processing and Control*, vol. 5, no. 3, pp. 214–222, 2010.
- [16] M. I. Ibrahimy, F. Ahmed, M. A. Mohd Ali, and E. Zahedi, "Real-time signal processing for fetal heart rate monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 2, pp. 258–262, 2003.
- [17] J. J. Oresko, Z. Jin, J. Cheng et al., "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 734–740, 2010.
- [18] B. De Moor, Database for the identification of systems (DaISy), 2010, <http://homes.esat.kuleuven.be/~smc/daisy/>.
- [19] J. G. Webster, *Medical Instrumentation, Application and Design*, John Wiley & Sons, 1995.
- [20] M. S. Manikandan and S. Dandapat, "Wavelet energy based diagnostic distortion measure for ECG," *Biomedical Signal Processing and Control*, vol. 2, no. 2, pp. 80–96, 2007.
- [21] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [22] Inc. The MathWorks, Denoising: Wavelet shrinkage, nonparametric regression, block thresholding, multisignal thresholding, 2013, <http://www.mathworks.es/es/help/wavelet/denoising.html>.
- [23] LabView TM, Advanced Signal Processing Toolkit, Wavelet Analysis Tools User Manual, 2013, <http://www.ni.com/pdf/manuals/371533a.pdf>.
- [24] S. Sardy, "Minimax threshold for denoising complex signals with waveshrink," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1023–1028, 2000.
- [25] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.
- [26] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society B*, vol. 59, no. 2, pp. 319–351, 1997.
- [27] W. H. Swallow and F. Kianifard, "Using robust scale estimates in detecting multiple outliers in linear regression," *Biometrics*, vol. 52, no. 2, pp. 545–556, 1996.
- [28] Matlab, ECG simulation Using atlab, 2013, <http://www.mathworks.com/matlabcentral/fileexchange/10858>.
- [29] H. G. Rodney Tan, A. C. Tan, P. Y. Khong, and V. H. Mok, "Best wavelet function identification system for ECG signal denoise applications," in *Proceedings of the International Conference on Intelligent and Advanced Systems (ICIAS '07)*, pp. 631–634, November 2007.
- [30] A. L. Goldberger, L. A. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.
- [31] J. Jezewski, A. Matonia, T. Kupka, D. Roj, and R. Czabanski, "Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram," *Biomedical Engineering*, vol. 57, no. 5, pp. 383–394, 2012.

Research Article

Effective Preprocessing Procedures Virtually Eliminate Distance-Dependent Motion Artifacts in Resting State FMRI

Hang Joon Jo,¹ Stephen J. Gotts,² Richard C. Reynolds,³ Peter A. Bandettini,¹ Alex Martin,² Robert W. Cox,³ and Ziad S. Saad³

¹ Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-1148, USA

² Section on Cognitive Neuropsychology, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-1366, USA

³ Scientific and Statistical Computing Core, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-1148, USA

Correspondence should be addressed to Hang Joon Jo; joh21@mail.nih.gov

Received 3 April 2013; Accepted 21 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Hang Joon Jo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artifactual sources of resting-state (RS) FMRI can originate from head motion, physiology, and hardware. Of these sources, motion has received considerable attention and was found to induce corrupting effects by differentially biasing correlations between regions depending on their distance. Numerous corrective approaches have relied on the identification and censoring of high-motion time points and the use of the brain-wide average time series as a nuisance regressor to which the data are orthogonalized (Global Signal Regression, GSReg). We replicate the previously reported head-motion bias on correlation coefficients and then show that while motion can be the source of artifact in correlations, the distance-dependent bias is exacerbated by GSReg. Put differently, correlation estimates obtained after GSReg are more susceptible to the presence of motion and by extension to the levels of censoring. More generally, the effect of motion on correlation estimates depends on the preprocessing steps leading to the correlation estimate, with certain approaches performing markedly worse than others. For this purpose, we consider various models for RS FMRI preprocessing and show that the local white matter regressor (WMe_{LOCAL}), a subset of ANATICOR, results in minimal sensitivity to motion and reduces by extension the dependence of correlation results on censoring.

1. Introduction

Resting-State Functional Magnetic Resonance Imaging (RS FMRI) has become a popular methodology for studying brain function with FMRI and holds promise for understanding brain functions without a task or stimulus [1]. A commonly used approach employs the cross correlation between time series to estimate the strength of connection between a pair of voxels or regions of interest after possible artifacts are removed by linear regression (nuisance-removal regression) from the original echo planar imaging (EPI) time series data. Part of the appeal of RS FMRI is the relative ease with which the data can be acquired.

However, drawing valid inferences can be fraught with pitfalls, as illustrated in recent publications that have caused

a considerable stir in the functional neuroimaging field. For example, Power et al. [2] showed that head movement differences between subjects might explain perceived differences in the spatial patterns of brain connectivity and suggested that these motion differences differentially bias short-range versus long-range correlations. This inference was reached by considering the change in interregional correlations after high-motion points were eliminated from the estimation of correlation. Removing high-motion samples differentially affected correlations depending on interregional distance, thus implicating motion as the source of this distance-dependent bias. As a result, the authors conclude that censoring is the recommended approach for reducing the effects of motion. While we agree with the notion that data censoring can be important, we find that the reported

distance-dependent bias is not primarily induced by motion. It is strongly exacerbated by the inclusion of the global signal averaged over whole brain (GS) and related regressors derived by time series averaging over regions containing signals of interest.

In this work, we replicate the bias reported in [2] using the data the authors have very generously made public; we demonstrate how the exclusion of particular tissue-based regressors reduces the distance-dependent bias effect considerably and how the use of a variant on ANATICOR [3] almost entirely eliminates the effect, establishing that our recommended approach is less sensitive to motion induced artifacts. This result is yet another demonstration of why the GS and comparable time series averaged over large brain areas, a practice still widely used, *should not* be projected out of the data in RS-FMRI [3–5]. Finally, we provide an annotated flowchart that presents our recommended data preprocessing pipeline.

2. Materials and Methods

2.1. MRI Data. Image data used in [2] are open to the public at the FCON 1000 project website (http://fcon_1000.projects.nitrc.org/). We used children group data (cohort 1; $N = 22$) that exhibited larger motion effects than the other two groups (adolescent and adult cohorts in the full data set). The details of the cohort are described in [2] and the website (http://fcon_1000.projects.nitrc.org/indi/retro/Power2012.html).

2.2. Preprocessing Pipeline. Overview of the preprocessing pipeline for this work is described in Figure 1. The recommended preprocessing steps for RS FMRI analysis are described towards the end of this work (see Figure 7). We deviate from our recommended pipeline to accommodate the particulars of the data at hand, as detailed in the text below and in the flowchart in Figure 1.

2.2.1. Segmentation of T1-Weighted Images. T1-weighted images of individual subjects were aligned to the first frame of FMRI echo planar imaging (EPI) data of resting scans and segmented into gray, white, and cerebrospinal fluid tissue classes using AFNI’s “3dSeg” program [6].

2.2.2. Despiking, Slice-Timing, and Head-Motion Correction. Despiking was done with AFNI’s “3dDespike” program for Figures 2 and 4(e), as the first step of the preprocessing pipeline. Each voxel’s time series $f(t)$ is L^1 fit to a Fourier series of order L , defaulting to 1/30 of the number of time points:

$$f(t) = a + bt + ct^2 + \sum_{k=1}^L \left\{ d_k \sin\left(2\pi \frac{kt}{T}\right) + e_k \cos\left(2\pi \frac{kt}{T}\right) \right\}, \quad (1)$$

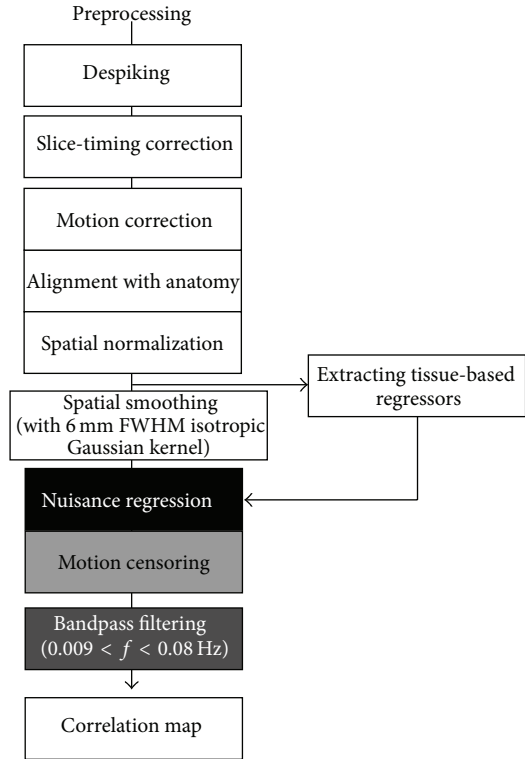


FIGURE 1: Overall preprocessing pipeline for this work (see text for detail of each step).

where T is the duration of time series, the parameters a , b , c , d , and e are chosen to minimize the sum over t of $|v(t) - f(t)|$ (L^1 regression), and $v(t)$ is the EPI time series of each voxel. The value of L used herein is $N/30$, where N is the number of time points. The median absolute deviation (MAD) of the residuals is used to obtain a standard deviation estimate σ that is robust to outliers:

$$\sigma = \sqrt{\frac{\pi}{2}} (\text{MAD}). \quad (2)$$

For each voxel value, define s as follows:

$$s(t) = \frac{v(t) - f(t)}{\sigma}, \quad (3)$$

and values with s greater than the threshold value of s for a spike (c_1) are replaced with a value that yields a modified s' :

$$s' = c_1 + (c_2 - c_1) \tanh\left(\frac{s - c_1}{c_2 - c_1}\right), \quad (4)$$

where c_2 is the upper range of the allowed deviation from $f(t)$. $s = [c_1, \infty)$ is mapped to $s' = [c_1, c_2)$. By default parameters c_1 and c_2 are set to 2.5 and 4, respectively, although program “3dDespike” allows users to modify them. With the default parameters, despiking consists of transforming spike values from the range of $[2.5\sigma, \infty)$ to $[2.5\sigma, 4\sigma)$. The purpose of this transformation is to make the output data be continuous in the input data: small changes in the input (e.g., a value

going from slightly under a threshold to slightly over) will not produce large changes in the despiked output. Slice-timing correction was performed, and motion correction was done by rigid body registration of EPI images to a base image [7]. Alignment of EPI data to the T1 was accomplished via an affine transformation, as was the spatial normalization of the T1 to the MNI avg152 T1 template, in MNI stereotaxic coordinates. All 3 transformations were applied at once to the EPI data to prevent multiple resampling steps.

Despiking was skipped in [2]. In practice, however, we find that despiking appears to improve the results of volume registration over time as illustrated in Figure 2 (also see supplementary video S1 available on line at <http://dx.doi.org/10.1155/2013/935154>). With despiking, motion parameters are less variable and the alignment quality is superior when visually examined.

2.2.3. Nuisance-Removal Regression. Five types of nuisance regression models were compared in this study. All regressors were extracted from motion-corrected EPI data *before* spatial smoothing with an isotropic Gaussian smoothing kernel (full-width-at-half-maximum; FWHM = 6 mm). Extraction of tissue-based regressors prior to any spatial smoothing is essential, to avoid mixing data from different tissue types; this point, while obvious, is often not made in Methods sections of papers. Regressors in the first model, GS + MO, included the 6 motion estimates, the tissue-based averages (global signal, GS; white matter signal, WM; large ventricle signal, LV), and the first time difference of each of the aforementioned regressors. In addition, n th-order Legendre polynomials were used to model slow baseline fluctuations. n is automatically determined by the number of EPI time points in the AFNI program *afni_proc.py* and was set to 4 for the time series analyzed here (http://afni.nimh.nih.gov/pub/dist/doc/program_help/afni_proc.py.html)

$$Y^i = X_{GS}B_{GS}^i + X_{WM}B_{WM}^i + X_{LV}B_{LV}^i + X_{MO}B_{MO}^i + X_{DT}B_{DT}^i + \text{residual} \quad (\text{model GS + MO}), \quad (5)$$

where Y^i is the EPI time series at a voxel i , X_{GS} is a global signal (GS) calculated by averaging the time-series over all brain mask voxels, X_{WM} is the average signal of all white matter voxels, X_{LV} are the averaged time series of lateral ventricles (LV) masks, X_{MO} is the group of six regressors for motion correction parameters (three translation and three rotation), and X_{DT} is the group of n detrending polynomials. The *residual* is the “cleaned” time series after subtracting the L^2 best-fit regression model of the nuisance variables from the original voxel time series. The second model, GS, excluded the 6 motion estimate regressors and their first difference terms as follows:

$$Y^i = X_{GS}B_{GS}^i + X_{WM}B_{WM}^i + X_{LV}B_{LV}^i + X_{DT}B_{DT}^i + \text{residual} \quad (\text{model GS}). \quad (6)$$

The third model, MO, included motion estimates with their first difference terms but omitted any tissue-derived regressors and their first difference terms as follows:

$$Y^i = X_{MO}B_{MO}^i + X_{DT}B_{DT}^i + \text{residual} \quad (\text{model MO}). \quad (7)$$

The fourth model was based on the model MO but included a localized and eroded WM regressor to form a local estimate of nuisance parameters while avoiding gray matter signals in the regions of interest as follows:

$$Y^i = X_{MO}B_{MO}^i + X_{WMe_{LOCAL}}^i B_{WMe_{LOCAL}}^i + X_{DT}B_{DT}^i + \text{residual} \quad (\text{model MO + WMe}_{LOCAL}), \quad (8)$$

where $X_{WMe_{LOCAL}}^i$ is a regressor of local WM signal for each voxel i , which can be calculated by averaging signals in eroded WM with a local sphere mask ($r = 45$ mm) by the AFNI program *3dLocalStat*. The fifth model Depike + MO + WMe_{LOCAL} is based on the model MO + WMe_{LOCAL}, but EPI data Y^i was despiked at the first stage of processing. These final two models are reduced variants of ANATICOR [3], lacking regressors of independently acquired physiological signals (not available in [2]) and the eroded large ventricles (LVE) regressors, as well as the despiking step for the fourth model (MO + WMe_{LOCAL}).

2.2.4. Censoring and Bandpass Filtering. We based the criterion for censoring on the Euclidian norm of the first time differences of motion estimates $\|\mathbf{d}\|_2$. This criterion has been part of the AFNI processing stream (*afni_proc.py*) and while not identical to the frame-wise displacement (FD) in [2], it serves the same function of eliminating data at time points when significant rapid motion is detected. At a $\|\mathbf{d}\|_2$ threshold of 0.25 mm, we censored on average 17.6% of the time series (1.8% and 45.0% at minimum and maximum, resp.).

Though contrary to our recommendation in the Discussion section, we filtered the data with a bandpass filtering kernel ($0.009 < f < 0.08$ Hz) after nuisance regression to avoid the degrees-of-freedom (DOFs) limitation for high movers because the EPI data had less samples than regression model parameters. This filtering was done via linear regression of sine/cosine basis functions, to avoid artifacts that would otherwise arise from the censoring process (e.g., assuming constant time steps or including censored data points).

2.2.5. Spatial Smoothing, and Orders of Preprocessing Steps. Figure 3 illustrates the effects of different processing steps on the spectral content of the time series. The preprocessing pipeline used in [2], shown to the left of Figure 3 as pipeline 1, included spatial smoothing (FWHM = 6 mm) and bandpass filtering (0.009–0.08 Hz) before regression. The first row shows the periodogram of slice-timing and head-motion corrected fMRI data, which were used as the common inputs to both pipelines 1 and 2. The other rows of each column are the periodograms of fMRI data as they are sequentially processed by subcomponents of the pipelines from top to

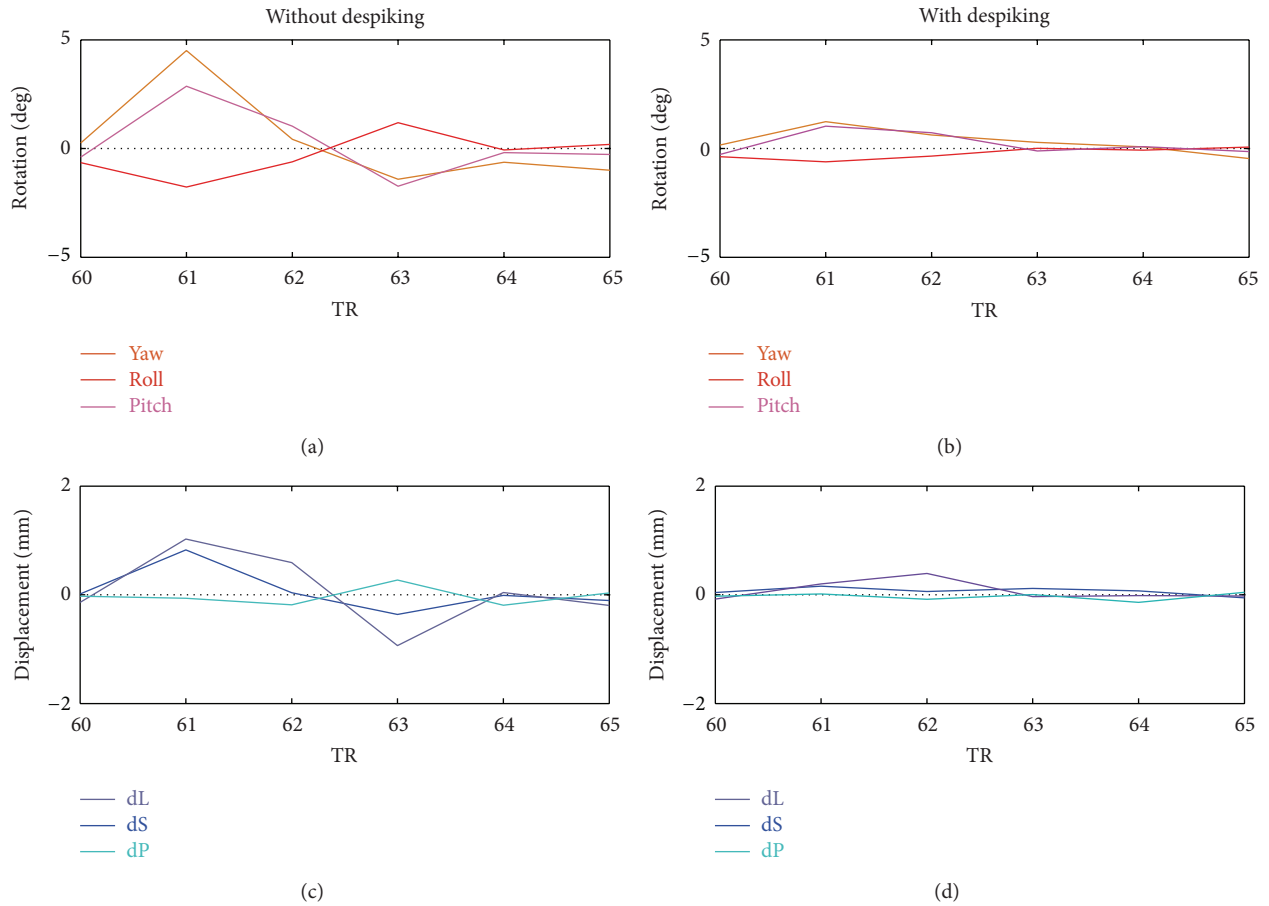


FIGURE 2: Improved motion correction by adding despiking step. The 6 motion estimates are from time frames in the corresponding video file (Supplementary Material S1). The upper and lower rows show the rotation and translation estimates, respectively, and the left and right columns show the motion estimates from the volume registration of fMRI data processed by pipelines starting without and with despiking, respectively. Registration without despiking resulted in visible residual motion between the 60th and 65th time frames. This suggests that the more elevated motion estimates obtained without despiking are less accurate than those with despiking. The subject used is *sub0015004*, who had the largest head movements in the children group.

bottom. Gray, black, blue, and red lines are spectral densities of GS, gray matter (GM), cerebrospinal fluid (CSF), and WM masks, respectively, which were averaged across the subjects. Not surprisingly, spatial smoothing can be done at any of these stages, as long as the tissue-based regressors are derived before spatial smoothing. The regression of nuisance parameters can also be carried out either before or after the bandpass filtering stage as long as the nuisance regressors are subject to the same bandpass filter. Otherwise, the regression step would reintroduce frequency components outside of the bandpass range as shown in the bottom row of column 1 [8].

2.3. Correlation Analysis for Seed Pairs. For each individual subject, the time series of 264 seed locations in standard brain space (MNI 152) were obtained from censored and uncensored data to produce two sets of 34,716 ($264 \times 263/2$) Pearson correlation coefficients [2]. The uncensored correlation coefficients were subtracted from the motion-censored correlation coefficients. The correlation differences are plotted as a function of the Euclidean distance between

the pairs of seed locations in Figure 4, and the nonlinear dependence on distance is referred to as the distance-dependent correlation bias. Note that for all models, we censored the same fraction of time points. We also examined the benefits of replacing Pearson correlation with Spearman's rank correlation, which is more robust to the presence of outliers in the time series that may be induced by motion.

2.4. Fits of Nuisance Regressors. We examined the spatial distribution of variance captured by the nuisance regressors [4] and correlations between them. To this end, we computed (i) the marginal explained variance (R^2 value) maps of the global signal (GS) and six head-motion estimates (MO) at each voxel in the brain (see Figure 5), and (ii) the cross-correlation matrix of the regressors to identify shared variance (see Figure 6). For these types of tests, the regressors were obtained from the time series that were despiked, slice-timing corrected, and volume registered.

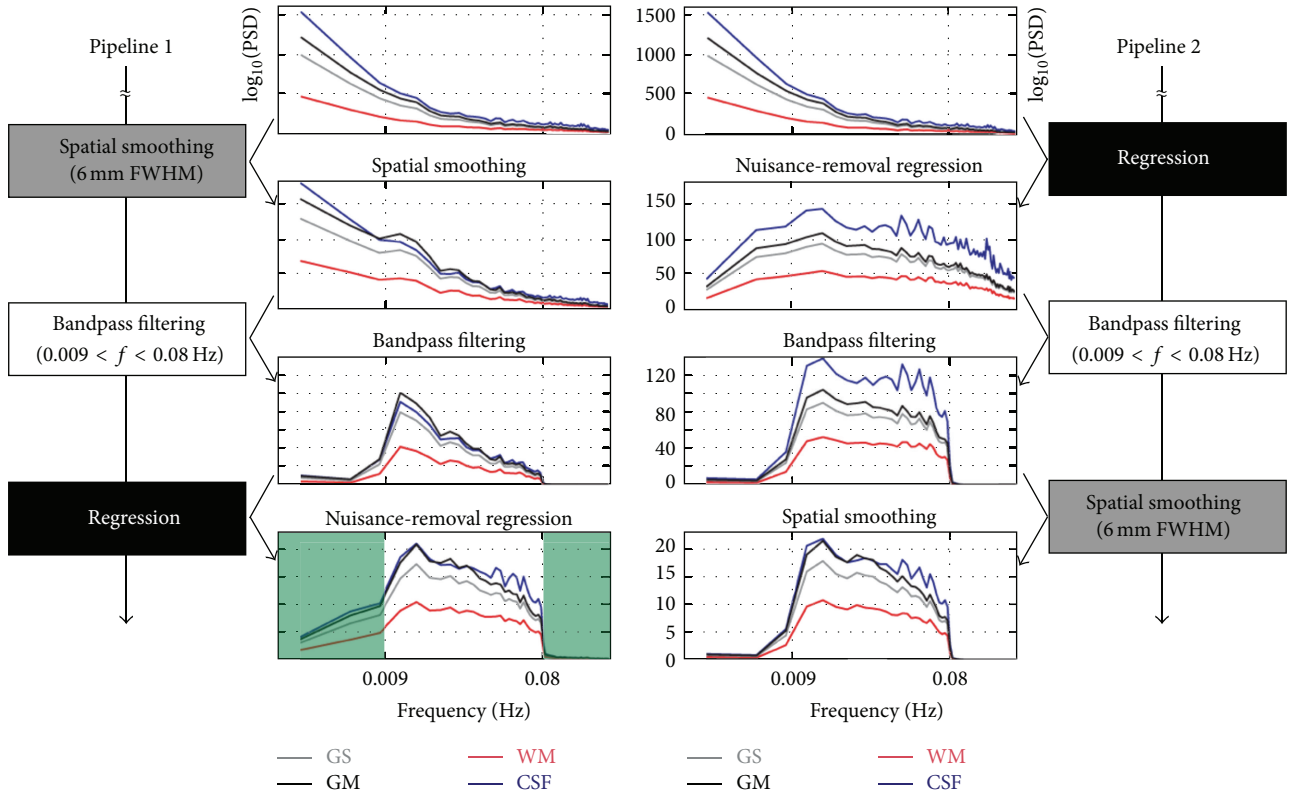


FIGURE 3: Group averaged power spectrum densities (PSD) of resting-state fMRI time series within brain tissues for each step in two different preprocessing orders. The improper processing order (pipeline 1) can reintroduce noise frequency components (signals of no interest) in lower frequency bands ($f < 0.009$ Hz, the green-tinted area) and higher frequency bands (see the text for more details).

3. Results

3.1. Distance-Dependent Correlation Bias after Different Preprocessing Steps. The distance-dependent correlation biases present after different preprocessing steps are shown in Figure 4, with results for the more standard Pearson correlation coefficient shown in the upper row. The distance-dependent bias with the GS + MO model (Figure 4(a)) mimic those obtained in [2]. The distance-dependent bias is captured by the curvilinear blue trace showing the average change in correlation after censoring. What this result indicates is that the correlation estimate can change considerably in the presence of motion and in a manner that depends on the interregional distance. In other words, GS + MO is sensitive to motion and by extension the censoring threshold, since eliminating points of high motion change the correlation values considerably. The desired trend for an estimate in these figures would be a flat line, preferably with zero mean and zero variance as a function of distance. With model GS, where motion estimates with their first difference terms were excluded, the bending of the mean curve was more pronounced than in Figure 4(a) (see Figure 4(b)). With model MO, which included motion estimates and their first differences but excluded tissue-based regressors, the bias was negative throughout and was more constant across interregional distances (Figure 4(c)). Figures 4(b) and 4(c) indicate that while the addition of GS makes the correlation estimate

more sensitive to motion, the use of MO alone is not enough to yield a robust estimate of correlations. Most notably, however, when WMe_{LOCAL} was added as a nuisance regressor to model MO (Figure 4(d)), the change in correlation became considerably less variable with distance and closer to zero. The addition of despiking further reduced the bias fluctuation as shown in Figure 4(e) where the nonlinear dependence of bias on distance was mostly eliminated; the mean bias was near zero and the variance of correlation change with censoring was the smallest of all five models tested. Thus of all models tested, Despike + MO + WMe_{LOCAL} resulted in the correlation estimates with minimal sensitivity to the presence of motion.

The lower row of Figure 4 shows results when Spearman's rank was used to compute the correlations. The trends are largely similar to those in Figure 4, with a small reduction in the scatter of correlation change for (a), (b), (c), and (d) panels where no despiking was performed. Not surprisingly, the use of Spearman's rank had little effect when despiking was included in the processing stages (e).

3.2. GS + MO Regressor Fits

3.2.1. Explained Variances of Nuisance Regressors. The marginal explained variances (R^2 values) of each regressor are presented in Figure 5. For the regression model GS + MO,

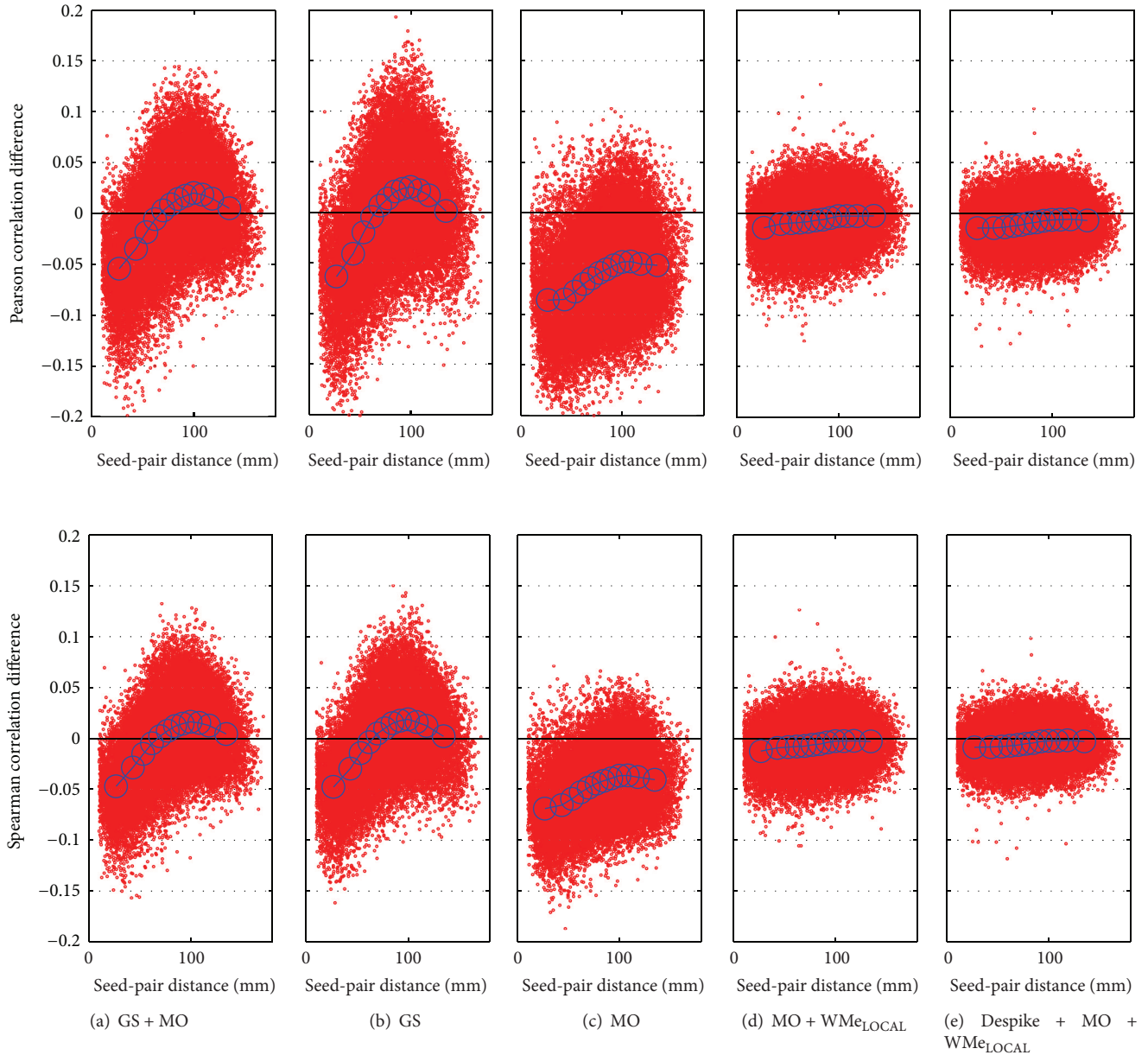


FIGURE 4: The column (a) shows that censoring high-motion frames from RS-fMRI data decreases short-distance correlations and augments long-distance correlations. The Pearson and Spearman correlation differences are plotted as a function of the Euclidean 3D distance between brain locations in the upper and lower rows, respectively. The results for each seed pair averaged over 22 subjects are plotted as red dots. Blue circles are the grand mean of averaged correlation differences for equal numbers of brain location pairs in twelve segments (2,882 pairs per circle), to highlight the trend. In the preprocessing steps, 6 motion estimates with their first difference terms (MO) and tissue-based regressors with their first difference terms (GS; global, eroded white matter, and lateral ventricle signals) were regressed out. Columns (b) and (c) present the distance-dependent correlation biases of nuisance regression models GS and MO, respectively. Column (d) shows results when a localized and eroded WM signal is added in the regression model of (c). Column (e) shows the model of Column (d) with the addition of despiking. The censored time points of fMRI images were determined at $\|d\|_2 > 0.25$ mm in (a), and the same time points were used in the censoring process of all models.

the regressors (GS and MO) fit most brain regions and locations at the outer edge of the brain with high R^2 values ($R^2 > 0.3$) (see the column GS + MO in Figure 5). When we measured R^2 values for each regressor, a different pattern in the spatial locations fit by each regressor could be identified:

(i) GS tended to fit GM, the sinuses, and mid-sagittal locations (yellow to red color overlays in the column GS in Figure 5; $R^2 > 0.7$), and (ii) MO captured variance more uniformly than GS throughout the brain, and the highest R^2 values were observed along the boundary between cortex and

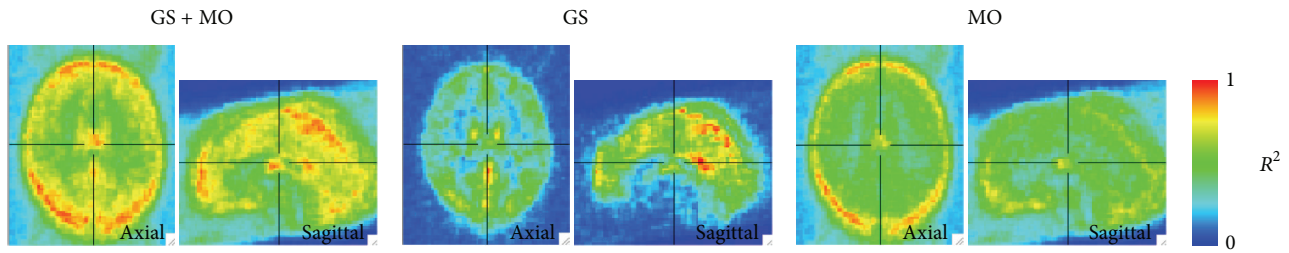


FIGURE 5: Marginal explained variances (R^2 values) of the regressors in nuisance removal regression models GS + MO. The labels GS and MO correspond to global signal and motion regressors, respectively.

nonbrain areas (see the column MO in Figure 5). The areas with the high R^2 values of GS and MO seldom overlapped each other.

3.2.2. Cross-Correlation Matrix between Regressors. The cross-correlation matrix between regressors is shown in Figure 6. GS and its first-difference term (GS') only have high correlations with one partial component (dP; the displacement along the anterior-to-posterior direction) of MO and its first difference term (MO'), respectively.

4. Discussion

4.1. Correlation Bias Observed by Motion Censoring. It was reported in [2] that the presence of motion introduces a distance-dependent distortion of correlations, whereby correlations between neighboring voxels were biased differently than correlation between voxels that are more distant. The authors also proposed a version (dubbed “scrubbing”) of motion censoring as a method to mitigate the bias of motion on correlation estimates. The evidence that the distance-dependent bias was introduced by subject motion was summarized in graphs that show the change in correlation magnitude between a set of brain location pairs (regions-of-interest; ROIs) as time points affected by excessive motion were excluded from the correlation estimation. The censoring process reduced the distance-dependent bias. While we agree that censoring is a valid approach, we highlight the fact that the distance-dependent bias *does not* appear to be driven by the mere presence of motion, and that the particular choice of preprocessing stream considerably exacerbates this distance-dependent bias. To illustrate this effect, we began by reproducing the effects of data censoring on short- versus long-distance correlations. For Figure 4(a), preprocessing included regression of head motion parameters, tissue-based time series including the GS, and their first-order time differences. In a reproduction of the results in [2], we found that censoring differentially affects correlations between ROIs that are close together compared to those that are further apart. This bend in the distribution was considered in [2] as evidence that motion was behind this bias, since lessening the effects of motion through censoring in turn differentially affected correlation values between ROIs depending on their distance. However, this *is not* entirely the case. In Figure 4(b), we recomputed the correlation differences but without including

the 6 motion estimates and their first differences, thereby amplifying the effect of censoring on the correlations. The scatter plot of the correlation difference increased in variance but the distance-dependent bias remained. The nonlinear trends in these scatter plots can be considered as a measure of the sensitivity of particular correlation estimates to motion. The ideal trend for a correlation estimate would be a flattened cloud with small constant variance and a constant bias of 0. In Figure 4(c), we brought back the motion estimates with their first differences but omitted any tissue-derived regressors and their first differences, most notable of which is the GS. With this model, the effect of censoring on the correlations became considerably less dependent on the inter-ROI distance. The correlation changes were also more uniformly negative, implying that sharp head motion tends to increase correlations prior to censoring (see also Gotts et al. [9]). Taking together the results of Figures 4(a)–4(c), we can conclude that the addition of GS to the model exacerbates the distance-dependence of the correlation estimates on motion, with results that are more dependent on the level of motion censoring. For Figure 4(d), we repeated the analysis in Figure 4(c) with the additional inclusion of a local eroded white matter signal, a regressor that intends to measure local manifestations of artifacts (e.g., hardware artifacts resulting from faulty head coil channels, [3]) while avoiding regions with the (gray matter) signals of interest. Not only was the dependence on the inter-ROI distance much reduced, but also the mean and range of the correlation bias were closer to zero. Addition of a despiking step at the very beginning of the preprocessing pipeline (Figure 4(e); see also Satterthwaite et al. [10]) further improved these trends, resulting in correlation estimates that varied little with the censoring of high-movement data points. The despiking procedure is often used to dampen the effects of extreme signal deviations on motion correction and variance estimates, and it is essentially a mild form of censoring. While it is expected that despiked data will always result in smaller changes in correlation after censoring, the two operations are not interchangeable as despiking is performed independently for each voxel. In other terms, not all reduced spikes get flagged as high-motion points. In conclusion, with a preprocessing model including despiking as an initial processing step and WME_{LOCAL} , the correlation estimate was least sensitive to motion artifacts and, by extension, to censoring threshold levels. We emphasize that the fraction of time points censored was the same

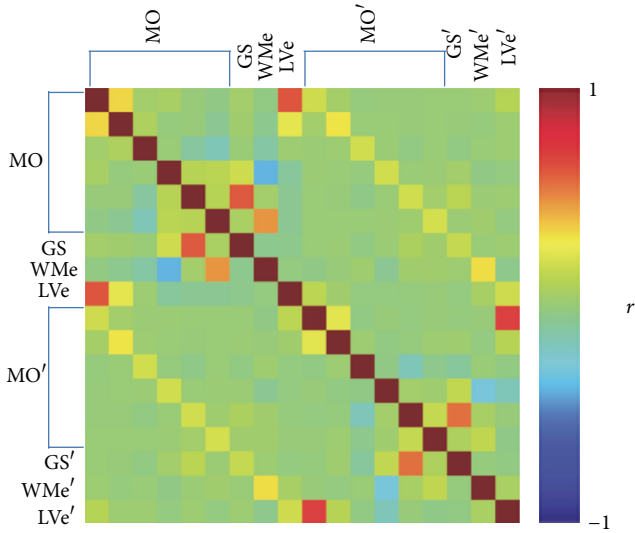


FIGURE 6: Cross-correlation matrix between the regressors. The correlation coefficients were averaged over all subjects ($N = 22$). MO represents the six regressors containing rigid-body motion parameter estimates (three translations and three rotations). GS is the average RS-FMRI time series over all voxels in the brain mask. WMe and LVe refer to the RS-FMRI time series averaged within eroded white matter and eroded lateral ventricles, respectively. Prime marks (') indicate the first differences of the regressors. The regressors were obtained from the time series that were despiked, slice-timing corrected, and then volume registered, and spatial smoothing *was not* applied to avoid mixing signal across different brain-tissue masks.

for all the models tested. Even when despiking was adopted in panel (e), we censored the same fraction of time points as in panel (a), (b), and (c). Therefore the fact that censoring had minimal effect on the correlation estimates suggests that the Despike + WMe_{LOCAL} approach is more robust to motion contamination than all the other models and is consequently least sensitive to censoring threshold levels. These results suggest that ANATICOR, the physiological noise augmented form of Despike + WMe_{LOCAL} , is not only useful for reducing local hardware artifacts, but also local manifestation of motion. While the basis of the benefit of WMe_{LOCAL} in reducing the motion bias is not entirely clear, one possibility is that it provides some adaptation to small local changes in the B_z magnetic field resulting from movement, which will affect the EPI time series. Lastly, we found that using Spearman rank instead of Pearson correlation was of little advantage for despiked time series but was of mild advantage for other conditions.

4.2. Suggested Preprocessing Pipeline. Figure 7 shows the pipeline we recommend for RS-FMRI analysis. Despiking FMRI data at the subject level is recommended to reduce the contribution of sudden spike signals to correlation estimates. Anecdotally, we also found it to improve the accuracy at the volume registration step (see the video in the Supplementary Material and Figure 2). Physiological denoising is carried

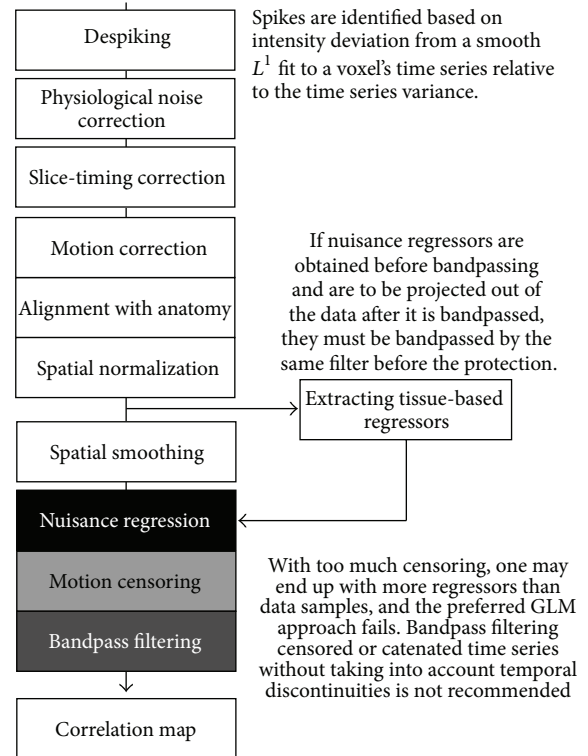


FIGURE 7: Annotated processing flowchart for RS-FMRI analyses. If nuisance regressors are obtained before bandpassing and are to be projected out of the data after it is bandpassed, they must be bandpassed by the same filter before the projection. With too much censoring, one may end up with more regressors than data samples, and the preferred regression approach fails. Bandpass filtering censored or concatenated time series without taking into account temporal discontinuities *is not* recommended. The WMe_{LOCAL} regressor is recommended, particularly for subject cohorts expected to exhibit high levels of head motion; it has the additional benefit of removing hardware artifacts that are hard to detect visually in the time courses of imaged volumes [3]. Slice-based physiological noise models [11] are projected from the data immediately after the despiking step because they are a function of slice acquisition time. A sample pipeline generating command is shown in example 5C at http://nimh.nih.gov/pub/dist/doc/program_help/afni_proc.py.html.

out early in the processing pipeline because RETROICOR [11] nuisance models depend on the acquisition time of each slice relative to the cardiac and respiratory cycles. These nuisance regressors are projected from the time series immediately after the despiking step. Bandpass filtering should be applied to both data and regressors of no interest. Otherwise, frequency components in cut-off bands will be introduced back through the regressors of no interest. It is best to perform censoring, nuisance regression, and bandpass filtering simultaneously in one regression model. By simultaneously doing these three subprocesses in one general linear model, there is no conflict between bandpassing and censoring. Though not carried out in this work for lack of data (physiological measures *were not* taken in [2]), physiological denoising is highly recommended [11, 12], as physiological noise differences amongst the subjects can certainly lead to

false inferences. In our recommended pipeline in Figure 7, we advocate bandpass filtering in the same model for nuisance regression. This manner of censoring can be handled readily, unlike in the pipelines 1 and 2 of Figure 3.

The regression model used here contains 6 motion estimates, their first difference terms, and WMe_{LOCAL} only, since “global” tissue-based regressors (e.g., GS, average gray matter, GM, noneroded LV, and WM) can also cause group differences either by spreading hardware artifacts that are undetectable by visual inspection in fMRI data [3] or by corrupting the correlation matrix, as can be seen when using GSReg [5]. As long as care is taken to prevent the inclusion of gray matter signals of interest, tissue-based regressors such as eroded LV (LVE) and WMe_{LOCAL} can be beneficial at reducing physiological and hardware artifacts and are part of our recommended ANATICOR [3] approach. The results presented here further demonstrate the utility of WMe_{LOCAL} in helping to reduce head motion artifacts. In these data, we were not able to include time series from the LVE mask as a nuisance component because the erosion operation eliminated too many LV voxels in most subjects due to a combination of small brain size and relatively coarse EPI resolution.

4.3. Summary. In this work, we have demonstrated that the distance-dependent bias in correlations between ROIs reported by Power and colleagues [2] is not driven only by motion. It is considerably exacerbated by the regression of nonspecific, tissue-averaged time series such as the GS. Specifically, the use of GS as a nuisance regressor can increase the sensitivity of correlation estimates to motion and motion censoring levels. This constitutes another example of why GS and equivalent regressors should not be projected out of the data in RS-fMRI [5]. We also find that Despike + WMe_{LOCAL} , a reduced version of our denoising approach dubbed ANATICOR [3], resulted in correlation estimates with minimal sensitivity to motion. While many in the field are rightfully concerned about the impact of motion on functional connectivity measures, these concerns can be effectively mitigated by the choice of appropriate preprocessing methods.

Acknowledgments

This study was greatly facilitated by the generous contribution of data by the authors of Power et al. 2012. The authors thank Kelly Barnes for helpful discussions. This research was supported by the NIMH and NINDS Intramural Research Programs of the NIH.

References

- [1] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar MRI,” *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.
- [2] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, “Spurious but systematic correlations in functional

- connectivity MRI networks arise from subject motion,” *NeuroImage*, vol. 59, no. 3, pp. 2142–2154, 2012.
- [3] H. J. Jo, Z. S. Saad, W. K. Simmons, L. A. Milbury, and R. W. Cox, “Mapping sources of correlation in resting state fMRI, with artifact detection and removal,” *NeuroImage*, vol. 52, no. 2, pp. 571–582, 2010.
- [4] K. Murphy, R. M. Birn, D. A. Handwerker, T. B. Jones, and P. A. Bandettini, “The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?” *NeuroImage*, vol. 44, no. 3, pp. 893–905, 2009.
- [5] Z. S. Saad, S. J. Gotts, K. Murphy et al., “Trouble at rest: how correlation patterns and group differences become distorted after global signal regression,” *Brain Connectivity*, vol. 2, pp. 25–32, 2012.
- [6] A. Vovk, R. W. Cox, J. Stare, D. Suput, and Z. S. Saad, “Segmentation priors from local image properties: without using bias field correction, location-based templates, or registration,” *NeuroImage*, vol. 55, no. 1, pp. 142–152, 2011.
- [7] R. W. Cox and A. Jesmanowicz, “Real-time 3D image registration for functional MRI,” *Magnetic Resonance in Medicine*, vol. 42, pp. 1014–1018, 1999.
- [8] J. Carp, “Optimizing the order of operations for movement scrubbing: comment on Power et al.,” *NeuroImage*, vol. 76, pp. 436–438, 2013.
- [9] S. J. Gotts, W. K. Simmons, L. A. Milbury, G. L. Wallace, R. W. Cox, and A. Martin, “Fractionation of social brain circuits in autism spectrum disorders,” *Brain*, vol. 135, pp. 2711–2725, 2012.
- [10] T. D. Satterthwaite, M. A. Elliott, R. T. Gerraty et al., “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data,” *Neuroimage*, vol. 64, pp. 240–256, 2013.
- [11] G. H. Glover, T. Q. Li, and D. Ress, “Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR,” *Magnetic Resonance in Medicine*, vol. 44, pp. 162–167, 2000.
- [12] R. M. Birn, J. B. Diamond, M. A. Smith, and P. A. Bandettini, “Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI,” *NeuroImage*, vol. 31, no. 4, pp. 1536–1548, 2006.

Research Article

Improving CT Image Analysis of Augmented Bone with Raman Spectroscopy

J. Charwat-Pessler,¹ M. Musso,² K. Entacher,¹ B. Plank,³ P. Schuller-Götzburg,⁴
S. Tangl,^{5,6} and A. Petutschnigg¹

¹ University of Applied Sciences Salzburg, Markt 136a, 5431 Kuchl, Austria

² University of Salzburg, Department of Materials Research and Physics, Hellbrunnerstraße 34, 5020 Salzburg, Austria

³ University of Applied Sciences Upper Austria, Stelzhamerstraße 23, 4600 Wels, Austria

⁴ Paracelsus Medical University Salzburg, Strubergasse 21, 5020 Salzburg, Austria

⁵ Medical University of Vienna, Karl Donath Laboratory for Hard Tissue and Biomaterial Research, Department of Oral Surgery, Sensengasse 2a, 1090 Wien, Austria

⁶ Austrian Cluster for Tissue Regeneration, Donaueschingenstraße 13, 1200 Vienna, Austria

Correspondence should be addressed to J. Charwat-Pessler; johann.charwat-pessler@fh-salzburg.ac.at

Received 29 March 2013; Accepted 10 May 2013

Academic Editor: Hang Joon Jo

Copyright © 2013 J. Charwat-Pessler et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, bone graft substitutes have been increasingly used in the medical field, for example, in order to promote new bone formation. Microcomputed tomography (μ -CT) is an image-guided technique used in medicine as well as in materials science, enabling the characterization of biomaterials with high spatial resolution. X-ray-based methods provide density information; however, the question how far conclusions on chemical structures can be inferred from any kind of CT information has not been intensively investigated yet. In the present study, a bone sample consisting of autogenous bone derived cells (ABCs) and bovine bone mineral (BBM) was investigated by μ -CT and Raman spectroscopic imaging, that is, by two nondestructive imaging methods. Thereby, the image data were compared by means of regression analysis and digital image processing methods. It could be found that 51.8% of the variance of gray level intensities, as a result of μ -CT, can be described by different Raman spectra of particular interest for bone composition studies by means of a multiple linear regression. With the better description of μ -CT images by the linear model, a better distinction of different bone components is possible. Therefore, the method shown can be applied to improve CT-image-based bone modeling in the future.

1. Introduction

Bone graft substitutes are widely used in the medical field and their usage covers a broad range of applications as for instance in dentistry or orthopedics. Bone grafting can be required in order to repair skeletal defects either to replace missing bone, filling voids, or to promote new bone formation so as to be able to place for instance an implant. In recent years an increase in demand for such bone graft substitutes could be observed [1–5] and due to the demographic situation especially in the United States and Europe a reversal of this trend is not forecast today. These bone substitutes encompass a variety of tissues and materials sources which can mainly be

distinguished by their source and properties, respectively. In terms of their composition, bone graft substitutes can broadly be divided in autograft, allograft, and xenograft materials. Autograft materials are designated to a graft made of the body's own bone which is usually taken from the hip or ribs, while allograft materials are made of human tissues. Xenograft materials are derived from animals as cows or pigs and are in fact established as well; however, autograft materials are still considered as the gold standard in medicine for reasons of compatibility. Biomaterials can also be differentiated due to their properties as osteogenesis, osteoinduction, and osteoconduction. Osteogenesis describes the formation

of new bone within the graft, osteoinduction refers to a stimulation of cells to convert into cells that are able to form bone, and osteoconduction means that the graft material acts as a scaffold that is ideally resorbed later on [6, 7]. Due to the fact that autogenous bone is not always available in the required quantity, different materials have to be considered. The question how materials with respect to allograft and xenograft materials, respectively, behave after bone grafting in terms of bone regeneration mechanisms (i.e., osteogenesis, osteoinduction, and osteoconduction) and involved risks for the patient, is subject of current research in medicine [8–10]. The mechanical behavior after transplantation is a different question which is of central interest in the medical and engineering field and consequently subject of intense research efforts. Bone tissue and the associated mechanism of bone modeling as well as bone remodeling are not completely understood yet due to the complex hierarchical structure of that material [11, 12]. Numerous different imaging techniques are used in order to address this problem. One of the probably most frequently used technologies in terms of clinical as well as material science is computed tomography (CT) and X-ray microcomputed tomography (μ -CT) whereby μ -CT was developed based on computed tomography (CT) [13]. The benefit of μ -CT is the higher spatial resolution and is considered as a powerful tool in medicine and bone research, respectively, [14, 15]. In matters of micromechanical investigations, X-ray based technologies form the basis for 3D models that are used later for calculations based on finite elements [16–18]. Admittedly the information content of μ -CT is strictly speaking limited to information of X-ray absorption which does not necessarily allow conclusions on tissues, especially if two different materials show either a similar or nearly the same absorption coefficient. These established methods (i.e., CT, μ -CT) are still part of science, but different techniques as Raman spectroscopy make an important contribution to the investigation of such previously mentioned questions [19–21], specifically as soon as Raman spectroscopy is used as imaging tool which is referred to as Raman spectroscopic imaging. Raman spectroscopic imaging can provide valuable clues in terms of the chemical structure, the molecular vibrations, and thus the chemical composition of cells and tissues, clues which are missing when regarding any kind of X-ray based method. Because that Raman spectroscopic imaging is bound by natural limitations on surfaces the implementation of the chemical information in generated 3D models requires as a prior condition the knowledge of how the relationship of these two different image data could be described in a mathematical way. Therefore the present study elaborates the question how far Raman spectra enable a description of a μ -CT image by means of a multiple linear regression.

2. Materials and Methods

In the present study a bone sample consisting of autogenous bone derived cells (ABC) and bovine bone mineral (BBM) was investigated with two different imaging techniques whereas ABC as well as Bio-Oss were compounds of the

appropriate sample [22]. Bio-Oss, produced by Geistlich Pharma AG, is a medical device based on bovine bone that is commonly used in bone regeneration due to its osteoconductive properties with respect to dental implants and periodontal defects, respectively. The sample originated from minipigs (i.e., a crossbred of Minnesota pigs and Vietnamese potbellied pigs) which were previously sinus grafted [22]. It was investigated by Raman spectroscopic imaging on the one hand and μ -CT on the other hand. Because the comparison of the information from the two imaging methods is of central interest, the bone sample in the present study was cut with a saw so as to have a valuable point of reference when matching the images.

The μ -CT data were collected by using a GE phoenix | X-ray nanotom 180 NF with a minimal possible isotropic voxel size (spatial resolution) of $0.5 \mu\text{m}$. For the investigated bone samples, a voxel size of $9.79 \mu\text{m}$ was chosen. The tomograms were generated with 80 kV voltage at the nano focus tube, the measurement current was $300 \mu\text{A}$, and the integration time at the detector was 500 ms. Altogether 1.500 projections were recorded, which led to a total measurement time of about 90 minutes. For image reconstruction a filtered back-projection algorithm was applied by using the Nanotom reconstruction software datos | X. The bone sample dried up to a moderate rate during the recording and was immediately put in an alcoholic solution again afterwards. This fact may have led to a certain degree of imprecision.

The confocal Raman based investigations were carried out with a Thermo Scientific DXR Raman microscope and a laser excitation at 780 nm. The spatial resolution comprised an area of approximately 3 by 4 mm with an increment of $50 \mu\text{m}$. The scanned area comprised an area of approximately 3 by 4 mm with a step size of $50 \mu\text{m}$. In each measurement point the Raman spectrum was collected by accumulating 50 times with a period of 0.5 seconds for each single accumulation. Some of the recorded Raman spectra had a bad signal to noise ratio (SNR) because of fluorescence from impurities on the surface but were not removed from the data set due to data acquisition program limitations. As the acquisition time by confocal Raman spectroscopy lasted almost 50 hours, the bone sample, which was preserved in an alcoholic solution before the measurement procedure, was completely dried up after the Raman measurement, causing shrinkage of the sample.

Raman spectroscopic imaging results in a chemical map comprising a multitude of spectra and thus showing the intensities of the appropriate selected Raman shift. Raman bands of particular interest for compositional studies regarding bone are listed in Table 1 [21]. Each peak can be considered as an indicator for either bone, collagen, or other tissues. The two CH related vibrational modes are present in collagen as well as noncollagenous organic moieties. Bone apatite, in which phosphate is substituted among other with carbonate ions (i.e., type B carbonate substitution), changes various physical properties of hydroxyl(l)apatite (OHAp) and is used in bone graft substitutes due to its chemical similarity to bone mineral [23]. Carbonated hydroxy(l)apatite promotes the osseointegration of hydroxyl(l)apatite [24], but there are important chemical and structural differences between the

TABLE 1: Selected Raman bands of bone and the manually defined maximum intensities.

Raman peak	Peak position (cm ⁻¹)	Intensity (cps)	Presence of
$\nu_1\text{PO}_4$	961	400	Bone
$\nu_2\text{PO}_4$	438	200	Bone
$\nu_4\text{PO}_4$	589	200	Bone
Amide I	1677	150	Collagen
Amide III	1256	150	Collagen
C-H bending	1458	150	Organic moiety
C-H stretching	2937	500	Organic moiety
Type B carbonate substitution	1075	200	Bone mineral

mineral in bone and carbonated apatite [23]. This should be taken into account regarding the indicated presence of type B carbonate substitution in Table 1. It could be found that the intensities of B type carbonate substitution concur extremely well with the Bio-Oss correlation map. Considering the fact that Bio-Oss is reported to contain carbonate apatite [25] the vibrational mode at 1075 cm⁻¹ might explain most notably the presence of Bio-Oss as well.

In addition to these characteristic vibrational modes, reference measurements of known compounds were done in order to be able to compare the spectra of the bone sample with the spectra from a known compound. Correlation maps can help to identify different compounds in the augmented bone sample more easily which is part of the sample.

These materials mentioned previously were Bio-Oss, Bio-Gide (a naturally resorbable bilayer membrane consisting of porcine collagen and produced by Geistlich Pharma AG), tricalcium phosphate powder (Carl Roth GmbH & Co. KG) with a purity grade of 99.9%, and bone as well as fat of pork, both provided by the local butcher.

The Raman spectra obtained with laser excitation at 780 nm showed a fluorescent background, which influences the quality of the data evaluation. All the spectra had therefore to be corrected applying an automatic baseline correction using the software solution of Thermo Scientific Inc. (OMNIC 8) for Raman applications.

There are a variety of ways to achieve a baseline correction using OMNIC 8. The basic idea is to fit a function through selected background points and subtract the resulting background from the trace. This can be achieved either by a linear equation or by polynomials of 1st up to 6th order. In order to find the coefficients of the function a matrix inversion technique is used. In case of logarithmic or exponential functions the equation is linearized prior to this by a natural log transformation. The underlying algorithm for an automatic baseline correction is commonly employed in industry in terms of spectral processing and fits the baseline through an iterative process [26].

On the basis of the chemical information contained in the Raman bands 13 different images could be generated, each showing the intensities related to the appropriate Raman band and reference (i.e., known compounds) spectrum, respectively. The key task of the present study was to determine how the different Raman spectra can describe the μ -CT

image on the basis of a linear relationship; therefore a multiple linear regression according to (1) was chosen for modeling:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_{i-1}X_{j-1} + b_iX_j. \quad (1)$$

With such a model it is not only possible to describe pure materials but also material compositions by means of Raman intensity spectra (term on the right side). These chemical data are used to describe the μ -CT information (term on the left side). The regression parameters b_j of the linear estimation function (1) are assumed by minimizing the residuals following the least square method in which each variable X_j represents a Raman spectrum. For this purpose the μ -CT image was considered as the dependent variable and the Raman spectra consequently as the independent variables. The constant b_0 can be interpreted as a gray level value from where all alterations work. In order to avoid redundant information in the model, variables (Raman spectra) that could be described by other variables within the data set were removed as soon as the coefficient of correlation exceeded a value of ± 0.9 . The coefficients of correlation were calculated following Pearson's ratio which is defined as the covariance of two variables divided by the product of their standard deviation (2):

$$\rho(X, Y) = \frac{\text{COV}(X, Y)}{\sigma(X)\sigma(Y)}. \quad (2)$$

After the data set was screened with the help of a correlation matrix all those spectra which enabled a description of the μ -CT image were determined. The odd data set (i.e., the data set without those spectra according to Table 2) formed the basis for a multiple linear regression. Additionally it can generally be specified that an unreasonable or unnecessary increase of explaining variables possibly leads to an artificially raised coefficient of determination of the model and thereby distorts the result. This fact emphasizes the need for data sets to be screened before performing a multiple linear regression.

The goodness of fit of the model can be verified on the one hand by means of the coefficient of determination and on the other hand by means of an F -test. If there is a true causal relationship between the dependent variables X_j and the independent variable \hat{Y} , respectively, the coefficients b_i are not equal to null. The null hypothesis suggests (3) that there

TABLE 2: Correlations matrix of the Raman bands and the reference spectra following the order: (1) Bio-Oss, (2) tricalcium phosphate, (3) Bio-Gide, (4) pork bone, (5) pork fat, (6) Amide I, (7) Amide III, (8) CH bending, (9) CH stretching, (10) phosphate bending ($\nu_2\text{PO}_4$), (11) phosphate bending ($\nu_4\text{PO}_4$), (12) phosphate stretching ($\nu_1\text{PO}_4$), and (13) type B carbonate substitution.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
(1)	1	0.059	-0.507	-0.488	-0.486	0.152	0.166	0.339	-0.404	-0.054	-0.087	0.283	0.149
(2)	0.059	1	0.562	0.680	0.335	0.251	0.140	0.055	0.396	0.620	0.556	0.804	0.482
(3)	-0.507	0.562	1	0.972	0.828	0.211	0.089	0.046	0.689	0.315	0.280	0.341	0.142
(4)	-0.488	0.680	0.972	1	0.841	0.246	0.120	0.067	0.740	0.440	0.398	0.468	0.261
(5)	-0.486	0.335	0.828	0.841	1	0.196	0.091	0.249	0.783	0.202	0.172	0.207	0.105
(6)	0.152	0.251	0.211	0.246	0.196	1	0.957	0.850	0.555	0.737	0.746	0.579	0.838
(7)	0.166	0.140	0.089	0.120	0.091	0.957	1	0.827	0.441	0.695	0.717	0.480	0.834
(8)	0.339	0.055	0.046	0.067	0.249	0.850	0.827	1	0.497	0.464	0.470	0.411	0.636
(9)	-0.404	0.396	0.689	0.740	0.783	0.555	0.441	0.497	1	0.588	0.566	0.495	0.507
(10)	-0.054	0.620	0.315	0.440	0.202	0.737	0.695	0.464	0.588	1	0.985	0.793	0.923
(11)	-0.087	0.556	0.280	0.398	0.172	0.746	0.717	0.470	0.566	0.985	1	0.729	0.918
(12)	0.283	0.804	0.341	0.468	0.207	0.579	0.480	0.411	0.495	0.793	0.729	1	0.783
(13)	0.149	0.482	0.142	0.261	0.105	0.838	0.834	0.636	0.507	0.923	0.918	0.783	1

is no difference between the regression coefficients b_i of the function and can thus be rejected on the basis of an F -test:

$$H_0 : b_1 = b_2 = \dots = b_i = 0. \quad (3)$$

In this study two multiple linear regressions are presented. The first one, which refers later on to model 1, incorporates all spectra whereas in the second one, which refers to model 2, variables not significant according to the F -test were excluded using the so-called backwards method. This technique operates progressively as it starts with a linear model including at first all variables and excludes variables afterwards step by step. This is done by evaluating the highest empirical F -value and has the benefit of reducing the amount of variables once more.

The coefficients b_i of the hyperplane \hat{Y} (1) are not directly comparable according to their influence on the estimated variables as these explanatory variables (i.e., the independent variables) can be scaled differently. In order to be able to compare the explanatory variables with each other Backhaus et al. [27] suggests to standardize the coefficients following (4) where the so-called Beta values β_i are the standardized coefficients, b_i the unstandardized coefficients, s_{x_i} the standard deviation of the appropriate variable, and finally s_y , which is the standard deviation of the dependent variable \hat{Y} . The higher the influence of a variable, the higher the absolute value of the standardized coefficient:

$$\beta_i = b_i \frac{s_{x_i}}{s_y}. \quad (4)$$

Initially the linear equation was calculated factoring all the remaining Raman spectra regardless whether the coefficients were significant with reference to a t -test or not. In a second stage the regression parameters b_j were tested by means of a two tailed t -test at which the null hypothesis (5) was rejected

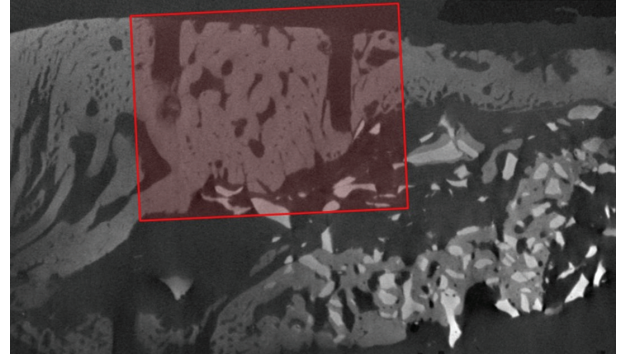


FIGURE 1: The appropriate region of interest related to the μ -CT image.

in favor of the alternative hypothesis (6) at a confidence level of 0.95:

$$H_0 : b_i = 0, \quad (5)$$

$$H_1 : b_i \neq 0. \quad (6)$$

It must be added that all calculations were carried out on the basis of gray level images although the Raman intensities in this paper are displayed in color. This does not have any impact on the results but solely on the perception of the information. Differences can be determined visually much more easily by the reader using color images instead of gray level images. All the charts and the associated statistical calculations were carried out in SPSS 19 whereas all the images were processed using MATLAB 2011a. In terms of a base line correction the spectra resulting from the Raman measurement were processed in OMNIC 8 as the algorithm for automatic base line correction was applied directly.

3. Results

The region of interest (ROI) is denoted by the red rectangle (Figure 1) in which the two kerfs serving as an orientation

mark are located. Due to the fact that Raman spectroscopic imaging is limited to surfaces by natural conditions and μ -CT provides images of the internal structure the very best μ -CT image was selected manually being as close as possible to the surface and preserving an appropriate image quality at the same time.

The microscopic image in Figure 2(a) has been assembled from several individual images, which were stitched together to produce an overview image with high resolution. The squares are a result of this stitching process because the borders of the individual images remain visible. The accuracy of the microscopic evaluation is not affected by this phenomenon. In Figure 2 the microscopic image (Figure 2(a)) is compared with the corresponding Raman image (Figure 2(b)) showing the intensity distribution of the Amide III band, where red indicates a high and blue very low intensity.

The colors in between follow the order of the rainbow colors. In addition the Raman image was overlapped with the μ -CT image (Figure 2(c)) in order to gain a first visual impression of the relationship between these two image data. In entirely practical terms it can be observed that the shrinkage as a consequence of dehydration of the bone sample due to the long term measurement is narrowed to a moderate rate. The generated Raman images (Figure 2(b)) turned out to vary in intensity to such an extent that some of the spectra were enhanced by reducing the maximum value of the range of intensity. This means that values which were at a low level before (f.i. green) are displayed subsequently at a higher level as orange or yellow for instance. Figure 3 shows in an exemplary manner the impact of such a processing. On the left side a chemical map focusing on a Raman peak at 1500 cm^{-1} is given. The intensity maximum of the range is indicated by OMNIC 8 with 420. The maximum value was reduced in a second step to a value of 210. Lower intensities, which were poorly visible before (Figure 3(a)) can now be determined in a better way (Figure 3(b)). At the same time the intensities at the green spot in the upper region are raised accordingly following the order of the rainbow colors. The maximum values, as they were set in the present study, are listed in Table 1.

In an exemplary manner the effect of the performed baseline correction is displayed in Figure 4 by means of the reference spectrum of pork bone. It can be seen how the background is obscured due to fluorescence and how the background is flattened by means of the baseline correction in virtue of the Thermo Scientific software solution for Raman applications.

In a further step all those spectra showing a coefficient of correlation greater than ± 0.9 were removed from the data set and all further analyses were carried out with the reduced data set. In the event that highly correlated data within the data set could not be cleaned up the outcome will inevitably lead to a falsification of the explanatory linear model as explained in the previous chapter. In Table 2 the correlation matrix is given on which basis the data set was filtered.

Table 3 provides an overview of the spectra pairs exceeding the previously defined Pearson's correlation and those

which were retained and removed respectively. It could be found that four data pairs fulfilled the determined criteria, accordingly four data set were removed and the statistical calculations were continued without Bio Gide, the Amide III band and without the two phosphate bending modes.

Table 4 sums up the results with model 1 referring to the forced multiple linear regression and model 2 referring to the regression eliminating all those coefficients having no significant contribution to the linear model according to the t -test. Furthermore the Raman peak at 438 cm^{-1} ($\nu_2\text{PO}_4$) highly correlates with the Raman peak at 589 cm^{-1} ($\nu_4\text{PO}_4$). Both were removed given that each of these Raman peaks highly correlates with the Raman peak at 1075 cm^{-1} and exceeded the previously determined criterion (i.e., ± 0.9).

The global statistical evaluation shows by means of the coefficient of determination and the adjusted coefficient of determination, respectively, the goodness of fit of the total linear model. In conclusion at least 51.8% of the variation can be explained by the linear model. The goodness of fit remains unchanged after exclusion of the variable tricalcium phosphate and thus clearly indicates the very low explanatory amount of the same. All further observations are based on the multiple linear regression in which tricalcium phosphate was removed (model 2).

In Figure 5 the normal probability plot of the standardized residuals is given. In case of a perfect match all the data points would be on the linear equation. The residuals are arranged close to the straight line which emphasizes the normal distribution of the residuals. The assumption of normal distributed standardized residuals is relevant with reference to the coefficients (i.e., b_0 , b_i) because these estimators are assumed to be normal distributed when choosing a multiple linear model and performing an F -test and t -test, respectively. In case of not normal distributed residuals the test statistics used would not be valid.

Table 5 gives an overview of the unstandardized coefficients as well as of the standardized coefficients and the appropriate t -test results for each linear model. The confidence interval is stated for the unstandardized coefficients.

The compound eliminated due to the t -test was the correlation map of tricalcium phosphate showing a significance of 0.593 and thereby exceeding the defined limit of 0.05. Some of the values are low to such an extent that the decimal place cannot be seen in Table 5, but that does not mean that the significance has the value zero. As mentioned previously, the coefficients can be compared to each other directly as soon as they are standardized. The underlying intention thereby is to get rid of the different measurement dimensions which allow a comparison and thus a reasonable interpretation of these values with respect to their importance. In other words, the higher the value, the greater the importance of a variable X_i . A coefficient with a negative algebraic sign means that the constant (i.e., average gray level value) is decreased by the amount indicated by the appropriate coefficient b_i and β_i , respectively. A positive coefficient means vice versa that the constant is increased by the correspondent coefficient b_i within that linear model. However, the coefficients can only be interpreted that way if the 95% confidence interval

TABLE 3: Spectra showing a correlation greater than ± 0.9 .

Retained spectra		Removed spectra	
Raman peak	Peak position (cm^{-1})	Raman peak	Peak position (cm^{-1})
Pork bone	Correlation map	Bio-Gide	Correlation map
Amide I	1677	Amide III	1256
Carbonate subst.	1075	$\nu_2\text{PO}_4$	438
Carbonate subst.	1075	$\nu_4\text{PO}_4$	589

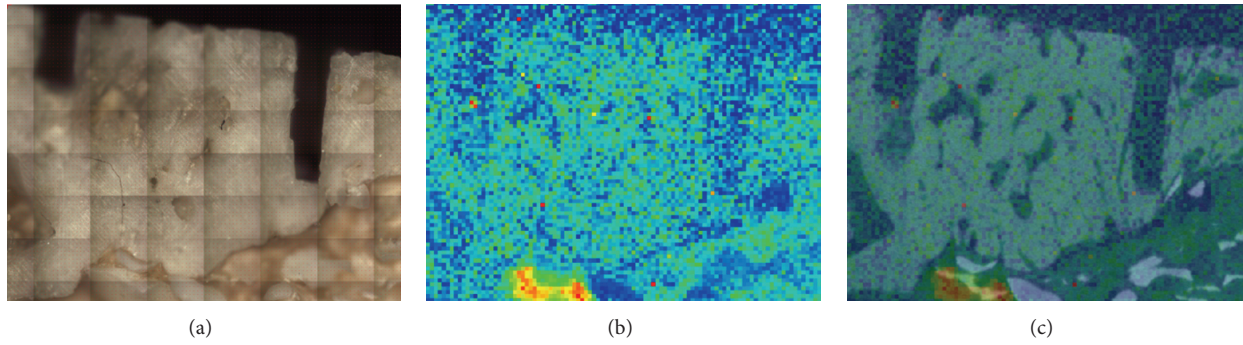
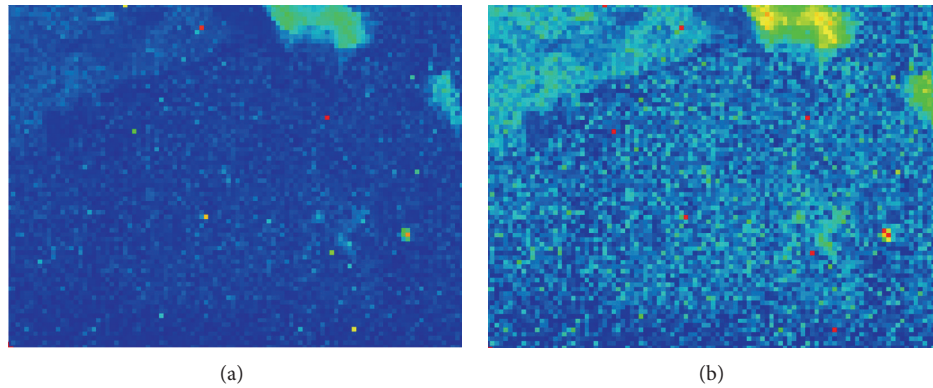
FIGURE 2: (a) Microscopic image, (b) Raman image showing the intensity distribution of the Amide III band, and (c) Raman image overlapped with the μ -CT image.

FIGURE 3: (a) Raman map with its original intensity range [0–420]. (b) Raman map with its reduced intensity range [0–210].

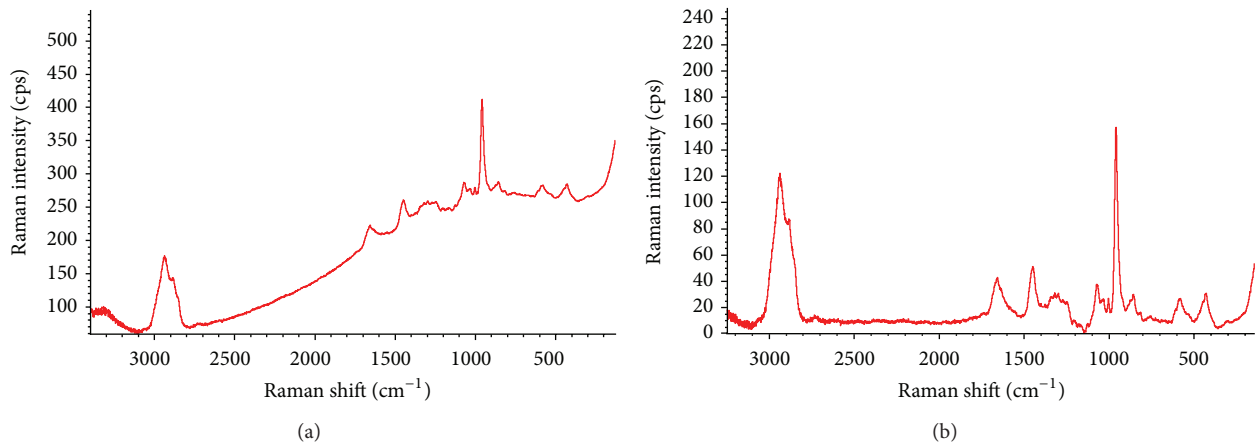


FIGURE 4: (a) Raman spectrum of pork bone showing fluorescence. (b) Corrected Baseline of the same spectrum.

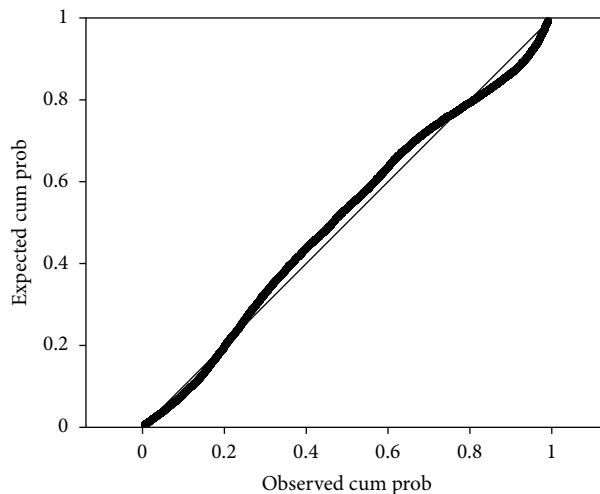


FIGURE 5: Normal probability plot.

TABLE 4: Model summary.

Model	R	R square	Adjusted R square
1	0.720	0.519	0.518
2	0.720	0.519	0.518

does not include the value zero. As soon as zero is part of the interval this statement is not true. Furthermore the confidence intervals provide information with respect to the unstandardized coefficients.

Table 5 offers an overview of the standardized values of both linear models. It can be observed that the CH bending mode together with the pork fat correlation map is the strongest positive standardized estimators in model 1 and the only ones regarding model 2. The confidence interval of these modes is positive as well and does not include the value zero for the unstandardized coefficients. This allows the conclusion that an increasing amount of CH bending mode and pork fat will lead to a higher μ -CT value.

The most important negative estimators are given with the carbonate substitution mode and the correlation map of pork bone. The corresponding intervals are also negative and do not include the value zero. A higher amount of that mode and pork bone, respectively, will lead to a lower μ -CT value. Such a breakdown (Table 5) shows which vibrational mode and component, respectively, influence the estimators (i.e., coefficient) the most.

In order to discuss the error of the presented linear regression model the residual plots are given in the form of scatterplots (Figure 6) in which the residuals of the regression model are plotted against each variable X_j . By means of these figures it can be determined whether the residuals are randomly scattered or not. If it happens that the residuals appear to be not randomly scattered, different additional parameters might be overlooked. Although tricalcium phosphate was excluded in model 2 the appropriate residual plot is nevertheless presented.

It can be observed that the scattering of the residuals appears differently in each case. Although the data seem to be well (i.e., randomly) spread around the red zero line, two main different data structures become visible and conspicuous, respectively, and are marked with two different colors (i.e., green and blue). In Figure 6(i) the intensity values for the type B carbonate substitution are displayed on the horizontal axis and the residuals are layed on the vertical axis. Those values which are framed with the green ellipse are distinctive and were accordingly assigned in the other figures (Figures 6(b), 6(c), 6(d), 6(f), and 6(g)). These data points show an unreasonable SNR which might be caused by impurities on the surface of the bone sample [28].

In Figure 6(h) the intensity values for phosphate (Raman peak at 961 cm^{-1}) and the residuals are displayed in exactly the same manner regarding the axes as previously described. The values which are marked with the blue ellipse in Figure 6(h) show an unreasonable SNR too. In large parts the values marked with the blue ellipse in Figures 6(e) and 6(f) are also affected by a bad SNR. A reason for this might be impurities on the surface of the sample again causing a general degradation of the Raman signal. This hypotheses, however, was not verified for which reason these data could not be excluded. These impurities might have changed their chemical composition due to the energy input of the laser which leads to the conclusion that any testing of that hypotheses is hardly feasible. Therefore a clear statement regarding these data would be too complicated.

Thus the error of the regression model can be traced back to two parameters. Firstly impurities on the surface of the sample are able to cause measurement errors which cannot be explained by the model and thus effect a major scattering. Secondly the model is not up to the task of explaining the overall variance. Different investigation methods could possibly lead to a clearer explanation regarding the material composition.

TABLE 5: Results of the multiple linear regression.

Model	Raman spectra	Unstand. coefficient	Stand. coefficient	Significance (<i>t</i> -test)	Confidence interval (at a level of 0.95)	
1	Constant	182.438	—	—	165.405	199.471
	Bio-Oss	-0.050	-0.049	0.022	-0.092	-0.007
	Tricalcium phosphate	0.009	0.013	0.593	-0.023	0.041
	Pork bone	-0.109	-0.391	0.000	-0.132	-0.085
	Pork fat	0.153	0.172	0.000	0.095	0.210
	Amide I	-0.087	-0.064	0.058	-0.176	0.003
	CH bending	0.547	0.550	0.000	0.483	0.612
	CH stretching	-0.384	-0.179	0.000	-0.482	-0.287
	Phosphate stretching	-0.099	-0.201	0.000	-0.126	-0.071
	Carbonate substitution	-0.709	-0.356	0.000	-0.809	-0.609
2	Constant	183.728	—	—	167.366	200.090
	Bio-Oss	-0.048	-0.047	0.026	-0.090	-0.006
	Pork bone	-0.104	-0.375	0.000	-0.122	-0.087
	Pork fat	0.147	0.165	0.000	0.094	0.199
	Amide I	-0.097	-0.072	0.018	-0.178	-0.016
	CH bending	0.551	0.554	0.000	0.488	0.614
	CH stretching	-0.389	-0.181	0.000	-0.485	-0.293
	Phosphate stretching	-0.096	-0.196	0.000	-0.122	-0.070
Carbonate substitution	-0.701	-0.352	0.000	-0.797	-0.605	

For this study a coefficient of correlation of 0.72 is sufficient in order to evaluate the presented method for which reason further methods do not appear necessary.

4. Discussion and Conclusion

Due to the fact that the sample was drying (moving) over time, for the applied μ -CT scans a compromise of contrast, image quality and measurement time had to be found. Results according to this approach show that hardly any conclusion on the chemical composition of the sample can be made by means of a μ -CT image solely. Alternatively to μ -CT it may be possible to use synchrotron radiation for CT. Indeed, synchrotron radiation offers significant advantages in terms of coherence, brilliance, and divergence of the beam [29] which results in fewer artifacts, improved contrast, and resolution, as well as much lower measurement times. These aspects will be taken into account in further investigations since synchrotron radiation together with Raman spectroscopy might possibly make another important contribution to the characterization of biomaterials.

Raman spectroscopic imaging is a tool that is up to the task of providing valuable additional information and has been successfully used in the present study in order to characterize biomaterials such as bone samples consisting of bone graft substitutes.

The development and generation of the present mathematical model, comprising correlation calculations, multiple regression analysis, and digital image processing methods, showed that 51.8% of the variance of the gray level pixels can

be successfully explained that way. Considering the fact that interpretations on the composition of bone graft substitutes and bone, respectively, are still carried out to a great extent on the basis of CT information, this study emphasizes the lack of deeper knowledge and understanding regarding the characterization of such materials on the one hand and highlights once more the importance of implementation of further imaging techniques on the other hand. Raman spectroscopy and other spectroscopic based examination methods should be embedded in further research efforts intensively since CT data do not provide reliable information on the chemical structure of biomaterials.

In view of the present results it can be stated moreover that this model needs to be improved by further data acquisition. The composition of the material in this study is very complex; thus further research efforts will be carried out on the basis of materials showing a lower degree of complexity with respect to their composition. The development of mathematical descriptions can thus be improved more easily in future.

This method is appropriate in order to improve the accuracy of 3D models which are exclusively based on CT image data and is thereby able to increase the quality of predictions on, for instance, micromechanical properties of bone and bone graft substitutes, respectively. An improved course of action in this regard is probably also of benefit to medical routines since the question how well such substitute materials are able to merge with body's own bone is still of importance in the medical field in terms of evaluation and characterization of such materials.

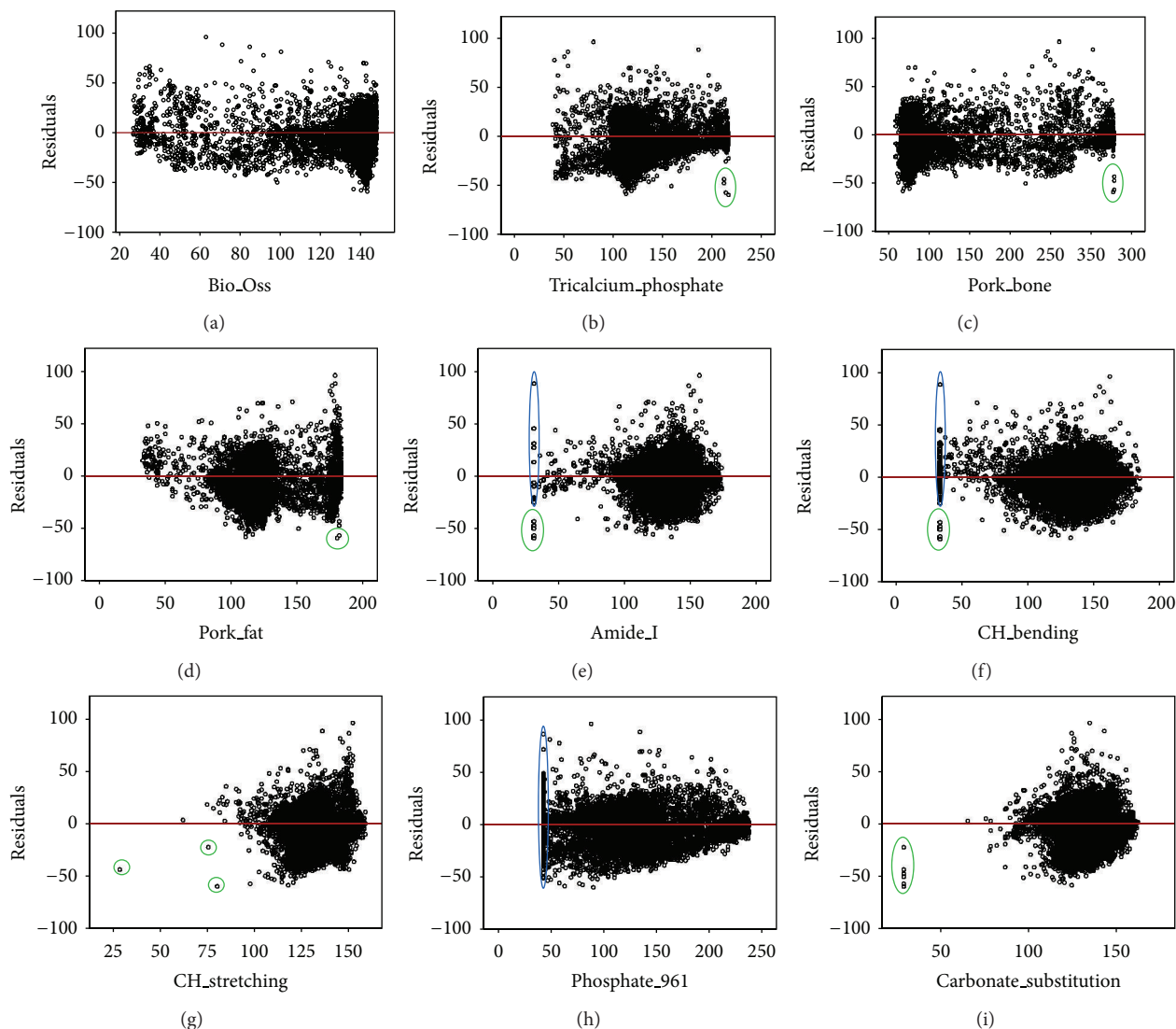


FIGURE 6: Residual plots and marked faulty measurements: (a) Bio Oss, (b) tricalcium phosphate, (c) pork bone, (d) pork fat, (e) Amide I, (f) CH bending, (g) CH stretching, (h) phosphate stretching at 961 cm^{-1} , and (i) carbonate substitution.

The question as to how far chemical maps by means of Raman spectroscopic imaging are allowed or even should be processed in terms of digital image processing techniques remains open and should be generally discussed in a scientific context since every image enhancement possibly affects the result.

Disclosure

The authors have neither a direct nor an indirect financial relation with the commercial identity of the products mentioned in the present paper.

Acknowledgments

The authorship would like to thank the Austrian Science Fund FWF for financing the research project TRP 254-N13. Furthermore the authorship would like to thank the

Medical University of Vienna, Department of Oral Surgery, for providing the bone sample.

References

- [1] A. S. Greenwald, S. D. Boden, V. M. Goldberg, Y. Khan, C. T. Laurencin, and R. N. Rosier, "Bone-graft substitutes: facts, fictions, and applications," *Journal of Bone and Joint Surgery A*, vol. 83, supplement 2, pp. 98–103, 2001.
- [2] C. Ogunsalu, "Bone substitutes and validation," in *Implant Dentistry-The Most Promising Discipline of Dentistry*, I. Turkyilmaz, Ed., InTech, New York, NY, USA, 2011.
- [3] A. L. Dumitrescu, *Chemicals in Surgical Periodontal Therapy*, Springer, Heidelberg, Germany, 2011.
- [4] A. S. Brydone, D. Meek, and S. Maclaine, "Bone grafting, orthopaedic biomaterials, and the clinical need for bone engineering," *Proceedings of the Institution of Mechanical Engineers H*, vol. 224, no. 12, pp. 1329–1343, 2010.

- [5] C. Laurencin, Y. Khan, and S. F. El-Amin, "Bone graft substitutes," *Expert Review of Medical Devices*, vol. 3, no. 1, pp. 49–57, 2006.
- [6] L. S. Pryor, E. Gage, J. C. Langevin et al., "Review of bone substitutes," *Cranial Maxillofac Trauma Reconstruction*, vol. 2, no. 3, pp. 151–160, 2009.
- [7] S. K. Nandi, S. Roy, P. Mukherjee, B. Kundu, D. K. De, and D. Basu, "Orthopaedic applications of bone graft & graft substitutes: a review," *Indian Journal of Medical Research*, vol. 132, no. 7, pp. 15–30, 2010.
- [8] W. R. Moore, S. E. Graves, and G. I. Bain, "Synthetic bone graft substitutes," *ANZ Journal of Surgery*, vol. 71, no. 6, pp. 354–361, 2001.
- [9] X. Wang, Y. Zuo, D. Huang, X. Hou, and Y. Li, "Comparative study on inorganic composition and crystallographic properties of cortical and cancellous bone," *Biomedical and Environmental Sciences*, vol. 23, no. 6, pp. 473–480, 2010.
- [10] F. Schwarz, K. Bieling, T. Latz, E. Nuesry, and J. Becker, "Healing of intrabony peri-implantitis defects following application of a nanocrystalline hydroxyapatite (Ostim) or a bovine-derived xenograft (Bio-Oss) in combination with a collagen membrane (Bio-Gide): a case series," *Journal of Clinical Periodontology*, vol. 33, no. 7, pp. 491–499, 2006.
- [11] P. Fratzl, H. S. Gupta, E. P. Paschalis, and P. Roschger, "Structure and mechanical quality of the collagen-mineral nanocomposite in bone," *Journal of Materials Chemistry*, vol. 14, no. 14, pp. 2115–2123, 2004.
- [12] P. Fratzl and R. Weinkamer, "Nature's hierarchical materials," *Progress in Materials Science*, vol. 52, no. 8, pp. 1263–1334, 2007.
- [13] S. R. Stock, *Microcomputed Tomography: Methodology and Applications*, CRC Press, Boca Raton, Fla, USA, 2008.
- [14] I. Kallai, O. Mizrahi, W. Tawackoli, Z. Gazit, G. Pelled, and D. Gazit, "Microcomputed tomography-based structural analysis of various bone tissue regeneration models," *Nature Protocols*, vol. 6, no. 1, pp. 105–110, 2011.
- [15] H. D. Barth, E. A. Zimmermann, E. Schaible, S. Y. Tang, T. Alliston, and R. O. Ritchie, "Characterization of the effects of x-ray irradiation on the hierarchical structure and mechanical properties of human cortical bone," *Biomaterials*, vol. 32, no. 34, pp. 8892–8904, 2011.
- [16] A. Boccaccio, A. Ballini, C. Pappalettere, D. Tullo, S. Cantore, and A. Desiate, "Finite element method (FEM), mechanobiology and biomimetic scaffolds in bone tissue engineering," *International Journal of Biological Sciences*, vol. 7, no. 1, pp. 112–132, 2011.
- [17] G. Limbert, C. van Lierde, O. L. Muraru et al., "Trabecular bone strains around a dental implant and associated micromotions—A micro-CT-based three-dimensional finite element study," *Journal of Biomechanics*, vol. 43, no. 7, pp. 1251–1261, 2010.
- [18] P. Schuller-Götzburg, K. Entacher, A. Petutschnigg, W. Pomwenger, and F. Watzinger, "Sinus lift with a cortical bone graft block: a patient-specific 3D finite element study," *International Journal of Oral & Maxillofacial Implants*, vol. 27, no. 2, pp. 359–368, 2012.
- [19] Y. Zhang, H. Hong, and W. Cai, "Imaging with Raman spectroscopy," *Current Pharmaceutical Biotechnology*, vol. 11, no. 6, pp. 654–661, 2010.
- [20] M. D. Morris, M. V. Schulmerich, K. A. Dooley, and K. A. Esmonde-White, "Vibrational spectroscopic imaging of hard tissues," in *Infrared and Raman Spectroscopic Imaging*, R. Salzer and H. W. Siesler, Eds., pp. 149–171, Wiley-VCH, Weinheim, Germany, 2009.
- [21] S. Gamsjäger, M. Kazanci, E. P. Paschalis, and P. Fratzl, "Raman application in bone imaging," in *Raman Spectroscopy for Soft Matter Applications*, M. S. Amer, Ed., pp. 227–267, John Wiley & Sons, New Jersey, NJ, USA, 2009.
- [22] G. Fuerst, S. Tangl, R. Gruber, A. Gahleitner, F. Sanroman, and G. Watzek, "Bone formation following sinus grafting with autogenous bone-derived cells and bovine bone mineral in minipigs: preliminary findings," *Clinical Oral Implants Research*, vol. 15, no. 6, pp. 733–740, 2004.
- [23] B. Wopenka and J. D. Pasteris, "A mineralogical perspective on the apatite in bone," *Materials Science and Engineering C*, vol. 25, no. 2, pp. 131–143, 2005.
- [24] A. Porter, N. Patel, R. Brooks, S. Best, N. Rushton, and W. Bonfield, "Effect of carbonate substitution on the ultrastructural characteristics of hydroxyapatite implants," *Journal of Materials Science*, vol. 16, no. 10, pp. 899–907, 2005.
- [25] M. Figueiredo, J. Henriques, G. Martins, F. Guerra, F. Judas, and H. Figueiredo, "Physicochemical characterization of biomaterials commonly used in dentistry as bone substitutes—comparison with human bone," *Journal of Biomedical Materials Research B*, vol. 92, no. 2, pp. 409–419, 2010.
- [26] "Thermo Scientific [homepage on the Internet]. Waltham, Massachusetts: Thermo Fisher Scientific Inc.(NYSE: TMO)," 2013, <http://www.thermoscientific.de/>.
- [27] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, *Multivariate Analysemethoden: Eine Anwendungsorientierte Einführung*, Springer, Berlin, Germany, 13th edition, 2011.
- [28] R. L. McCreery, "Signal-to-noise in raman spectroscopy," in *Raman Spectroscopy for Chemical Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2005.
- [29] D. H. Bilderback, P. Elleaume, and E. Weckert, "Review of third and next generation synchrotron light sources," *Journal of Physics B*, vol. 38, no. 9, pp. S773–S797, 2005.

Research Article

A Novel Adaptive Probabilistic Nonlinear Denoising Approach for Enhancing PET Data Sinogram

Musa Alrefaya^{1,2} and Hichem Sahli^{1,3}

¹ Department of Electronics and Informatics (ETRO-IRIS), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

² Faculty of Information Technology and Computer Engineering, Palestine Polytechnic University, Hebron, Palestine

³ Interuniversity Microelectronics Centre (IMEC), Leuven, Belgium

Correspondence should be addressed to Musa Alrefaya; musa.alrefaya@vub.ac.be

Received 23 March 2013; Revised 4 May 2013; Accepted 4 May 2013

Academic Editor: Hang Joon Jo

Copyright © 2013 M. Alrefaya and H. Sahli. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose filtering the PET sinograms with a constraint curvature motion diffusion. The edge-stopping function is computed in terms of edge probability under the assumption of contamination by Poisson noise. We show that the Chi-square is the appropriate prior for finding the edge probability in the sinogram noise-free gradient. Since the sinogram noise is uncorrelated and follows a Poisson distribution, we then propose an adaptive probabilistic diffusivity function where the edge probability is computed at each pixel. The filter is applied on the 2D sinogram pre-reconstruction. The PET images are reconstructed using the Ordered Subset Expectation Maximization (OSEM). We demonstrate through simulations with images contaminated by Poisson noise that the performance of the proposed method substantially surpasses that of recently published methods, both visually and in terms of statistical measures.

1. Introduction

Positron Emission Tomography (PET) is an in vivo nuclear medicine imaging method that provides functional information of the body tissues. The PET image results from reconstructing very noisy, low resolution raw data, that is, the sinogram, in which important features are shaped as curved structures. Enhancing the PET image spurred a wide range of denoising models and algorithms. Some methodologies focus on enhancing the reconstructed PET image directly, where others prefer enhancing the sinogram prior to reconstruction.

Although developing an appropriate denoising method for filtering the PET images has been widely studied in the last two decades, this problem is still receiving high attention from researchers trying to improve the diagnosis accuracy of this image. Existing methods may suffer drawbacks such as the careful selection of a high number of parameters, smoothing of the important features boundaries, or prohibitive computation. Recently, nonlinear diffusion

techniques have been investigated for PET images. Many researchers did explore the application of the well-known Perona and Malik anisotropic diffusion [1] in combination with diverse diffusivity functions on PET images [2–6], as well as on sinograms [7–9].

In [5], authors introduced a postreconstruction adaptive nonlinear diffusion (Perona and Malik) filter based on varying the diffusion level according to a local estimation of the image noise. Applying the nonlinear anisotropic filtering method on the whole body, PET image and the sinogram were presented in [4, 8, 10], respectively. Results showed that the anisotropic diffusion filtering algorithm helps improve the quantitative aspect of PET images.

In the study of [9], combining the anisotropic diffusion method with the coherence enhancing diffusion method for filtering the sinogram as a preprocessing step was proposed. However, the considered cascading approach is time consuming, and the results are highly dependent on the parameters selection criteria. Zhu et al. [11] built the diffusivity function using fuzzy rules that were expressed in a linguistic form.

The mean curvature and the Gauss curvature diffusion algorithms for filtering the PET images during reconstruction were investigated in [12]. An anatomically driven anisotropic diffusion filter (ADADF) as a penalized maximum likelihood expectation maximization optimization framework was proposed in [2]. This filter enhances a single-photon emission computed tomography images using anatomical information from magnetic resonance imaging as a priori knowledge about the activity distribution. Authors in [3] proposed a reconstruction algorithm for PET with thin plate prior combined with a forward-and-backward diffusion algorithm. The main drawback of the above filter, with respect to sinogram images is that the diffusion produces important oscillations in the gradient. This leads to a poorly smoothed image [11, 12]. Moreover, the adopted diffusivity functions do not consider the special properties of the sinogram, which are high level of Poisson noise and curved-shape features.

Happonen and Koskinen [13] proposed filtering the sinogram in a new domain, that is, stackgram domain, where the signal along the sinusoidal trajectories can be filtered separately. They applied this method using the Gaussian and nonlinear filters. Radon transform is applied for inverse mapping of the filtered data from the stackgram domain to the sinogram domain. The above described method is impractical for noise reduction of the medical images such as PET, since the corresponding 3D stackgram requires an enormous space of computer memory. Furthermore, it is a complex method not suitable for clinical applications where timely reconstruction of the PET image is a very important issue.

Most of the above studies considered a global image noise level. Local or varying noise level has been used in [14], where a nonlinear diffusion method for filtering MR images with varying noise levels was presented. The authors assumed that the MR image can be modeled as a piecewise constant (slowly varying) function and corrupted by additive zero-mean Gaussian noise. Pizurica et al. [15] proposed a wavelet domain denoising method for subband-adaptive and spatially adaptive image denoising approach. The latter approach is based on the estimation of the probability that a given coefficient contains a significant noise-free component called “signal of interest.” The authors of [15] found that the spatially adaptive version of their proposed method yielded better results than the existing spatially adaptive ones.

In this work, based on the following PET sinogram characteristics:

- (i) the important features in the sinogram are curved structures with high contrast values and these represent the regions of interest in the reconstructed PET image, for example, tumor,
- (ii) the weak edges in the sinogram are the edges that contain low contrast values,
- (iii) the noise in the sinogram is a priori identified as a Poisson noise,

we propose filtering the sinogram by means of a curvature constrained filter. The amount of diffusion is modulated according to a probabilistic diffusivity function that suits images contaminated with Poisson noise, being the known

noise distribution of PET sinograms. Further, considering the sinogram characteristics, we propose a probabilistic edge-stopping function based on Chi-square prior for the ideal sinogram gradient with a spatially adaptive algorithm for calculating the prior odd at each pixel. We show that this method is improving and denoising sinogram data which leads to enhanced reconstructed PET image.

The remainder of the paper is organized as follows, Section 2 briefly reviews the notions of curvature motion, edge affected diffusion filtering, and self-snakes. The proposed adaptive filtering scheme is introduced in Section 3. Finally, Section 4 discusses the experimental results, while the conclusions and future work are given in Section 5.

2. Geometry Driven Scale-Space Filtering

This section reviews the formulations for Mean Curvature Motion (MCM), Edge Affected Variable Conductance Diffusion (EA-VCD) and self-snakes. Also we recall the Gauge Coordinates notions. Extensive discussion can be find in [16]. Let f be a scalar image defined on the spatial image domain Ω , then the family of diffused versions of f is given by

$$U(f) : f(\cdot) \longrightarrow u(\cdot, t) \quad \text{with } u(\cdot, 0) = f(\cdot), \quad (1)$$

where U is referred to as the scale-space filter, u is denoted by the scale-space image, and the scale $t \in \mathbb{R}^+$ [6]. The denoised or enhanced version of f is a given $u(\cdot, t)$ that is the closest to the unknown noise-free version of f . In the following we denote $u(\cdot; t)$ by u_t .

2.1. Curvature Motion. Mean Curvature Motion (MCM) is considered as the standard curvature evolution. MCM allows diffusion solely along the level lines. In Gauge coordinates (see Section 2.4) the corresponding PDE formulation is (k is Euclidian curvature) as follows:

$$u_t(\cdot, t) = u_{vv} = k |\nabla u| = \operatorname{div} \left[\frac{\nabla u}{|\nabla u|} \right] |\nabla u|. \quad (2)$$

Hence diffusion solely occurs along the v -axis.

2.2. Edge Affected Variable Conductance Diffusion. Variable Conductance Filtering (VCD) is based on the diffusion with a variable conduction coefficient that controls the rate of diffusion [16]. In the case of Edge Affected-VCD (EA-VCD), the conductance coefficient is inversely proportional to the edgeness. Consequently it is commonly referred to as the edge-stopping function (g), in which the edgeness is typically measured by the gradient magnitude. The EA-VCD is governed by

$$u_t = \operatorname{div} [g(|\nabla u|) \nabla u]. \quad (3)$$

The above PDE system together with the initial condition given in (1) is completed with homogenous von Neumann boundary condition on the boundary of the image domain. Note that the Perona and Malik’s antitropic diffusion [1] is an EA-VCD.

2.3. Self-Snakes. Self-snakes are a variant of the MCM where an edge-stopping function is introduced [17]. The main goal is preventing further shrinking of the level-lines once they have reached the important image edges. For scalar images, self-snakes are governed by

$$u_t = |\nabla u| \operatorname{div} \left[g(|\nabla u|) \frac{\nabla u}{|\nabla u|} \right]. \quad (4)$$

This equation adopts the same boundary condition as (3). Furthermore, it can be decomposed into two parts [16, 17]:

$$\begin{aligned} u_t &= g|\nabla u| \operatorname{div} \left[\frac{\nabla u}{|\nabla u|} \right] + (\nabla g) \cdot \nabla u \\ &= gk|\nabla u| + (\nabla g) \cdot \nabla u. \end{aligned} \quad (5)$$

The first part describes a degenerate forward diffusion along the level lines that is orthogonal to the local gradient; it allows preserving the edges. Additionally, the diffusion is limited in areas with high gradient magnitude and encouraged in smooth areas. Actually the first term is the constraint curvature motion. The second term can be viewed as a shock filter since it pushes the level-lines towards valleys of high gradient, acting as Osher's shock filter.

2.4. Gauge Coordinates. An image can be thought of as a collection of curves with equal value, the isophotes. At extrema an isophote reduces to a point, at saddle points the isophote is self-intersecting. At the noncritical points; a Gauge coordinate system [18] is defined by two orthogonal axes v and w , which correspond to the directions of the tangent and normal at the isophote. The second order Gauge derivatives of the image in the vv and ww directions have the following second-order structures:

$$\begin{aligned} u_{vv} &= \frac{u_{xx}u_y^2 - 2u_xu_yu_{xy} + u_{yy}u_x^2}{(u_x^2 + u_y^2)}, \\ u_{ww} &= \frac{u_{xx}u_y^2 + 2u_xu_yu_{xy} + u_{yy}u_x^2}{(u_x^2 + u_y^2)}. \end{aligned} \quad (6)$$

These gauge derivatives can be expressed as a product of gradients and a 2×2 matrix with second-order derivatives [18]:

$$\begin{aligned} u_{ww}u_w^2 &= (u_x, u_y) \begin{pmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}, \\ u_{vv}u_w^2 &= (u_x, u_y) \begin{pmatrix} u_{yy} & -u_{xy} \\ -u_{xy} & u_{xx} \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}. \end{aligned} \quad (7)$$

For u_{ww} the matrix equals the Hessian H . For u_{vv} it is $\det H \cdot H^{-1}$. Note that the expressions are invariant with respect to the spatial coordinates. The above expression can also be obtained as follows [16]:

$$\begin{aligned} u_{vv} &= u_{xx} \sin^2 \theta + u_{yy} \cos^2 \theta - 2u_{xy} \sin \theta \cos \theta, \\ u_{ww} &= u_{xx} \cos^2 \theta + u_{yy} \sin^2 \theta + 2u_{xy} \sin \theta \cos \theta, \end{aligned} \quad (8)$$

where θ is given as $\theta = \arctan(u_y/u_x)$ and $u_v = -u_x \sin \theta + u_y \cos \theta = 0$.

The two expressions of (6) can be combined linearly in a PDE setting. In this way, an image u_0 is evolved according to a weighted combination of these two invariants:

$$u_t = g_1(|\nabla u|)u_{vv} + g_2(|\nabla u|)u_{ww} \quad (9)$$

with $u(t=0) = u_0$.

Equation (9) comprises a diffusion modulated by g_1 along the image edges vv (a smoothing term) and a diffusion adjustable by g_2 across the image edges ww (a sharpening term). Careful modeling of these terms allows efficiently denoising the PET sinograms, whilst keeping their interesting features.

3. The Probabilistic Curvature Motion Filter

The proposed probabilistic curvature motion filters are based on the idea of the *probabilistic diffusivity function* [19], where the diffusivity function is expressed as the probability that the observed gradient presents no edge of interest under a suitable marginal prior distribution for the noise-free gradient. In [19], the probabilistic diffusivity function has been defined as

$$g_{\text{pr}}(x) = A(1 - P(H_1 | x)), \quad (10)$$

where the normalizing constant A is set to $A = 1/(1 - P(H_1 | 0))$ to ensure that $g_{\text{pr}}(0) = 1$, the hypothesis H_1 describes the notion whether an edge element of interest is present given the considered noise and H_0 an edge element of interest is absent. For a noisy gradient model $x = y + n$, we set

$$H_0 : y \leq \sigma_n, \quad H_1 : y > \sigma_n \quad (11)$$

with y being the ideal, noise-free, gradient magnitude, and σ_n being the standard deviation of the noise n . In [19] it has been demonstrated that

$$g_{\text{pr}}(x) = (1 + \mu\eta(0)) \frac{1}{1 + \mu\eta(x)} \quad (12)$$

with the prior odds defined as

$$\mu = \frac{P(H_1)}{P(H_0)} \quad (13)$$

and the likelihood ratio

$$\eta(x) = \frac{P(x | H_1)}{P(x | H_0)}. \quad (14)$$

Considering a Laplacian prior $p(y) = (\lambda/2)e^{-\lambda|y|}$, we have [19]:

$$\mu = (e^{\lambda\sigma_n} - 1)^{-1}, \quad (15)$$

and the parameter λ can be estimated as

$$\lambda = [0.5(\sigma^2 - \sigma_n^2)]^{-1/2} \quad (16)$$

with σ^2 denoting the variance of the noisy image and σ_n as defined above. Due to limited space, the reader is referred to [19] for the detailed expression of $\eta(x)$ in (14).

It has to be noted that the diffusivity function (12) fits in the cluster of backward-forward diffusivities, and it has no free parameters to be set. Moreover, for the considered PET sinograms, the noise standard deviation, σ_n , in (16) is estimated as $\sigma_n^2 = \text{Var}(f_{Ln})$, where the image noise, f_{Ln} , is reconstructed via wavelet decomposition of f , from the two finest resolution levels coefficients, using Daubechies (4) wavelet.

3.1. Probabilistic Self-Snakes (PSS). It can be demonstrated that the diffusion of scalar images via EA-VCD can be decomposed into (5), which can be rewritten as follows [16]:

$$u_t = g(|\nabla u|) u_{vv} + [g(|\nabla u|) + g'(|\nabla u|)|\nabla u|] u_{ww} \quad (17)$$

consolidating the properties of both the self-snakes and the EA-VCD into a single diffusion schema.

Considering (9) and the probabilistic diffusivity function, if we set $g_1(x) \doteq g_{pr}(x)$, and the sharpening term, $g_2(x) \doteq g_{pr}(x) + xg'_{pr}(x)$, we obtain the probabilistic self-snakes (PSS) [7]. By its nature, the PSS enhances the weak edges and/or features in the sinogram. The second term in (32) allows the sharpening. In this way, weak but important edges are enhanced whilst the noise is removed efficiently.

PSS proved to be very effective and flexible for the sinogram image where the high contrast regions, which represent a tumor in the reconstructed PET, should be smoothed wisely without blurring the poor edges [7]. Like EA-VCD, the main advantage of this filter is that the average gray value of the image is not altered during the diffusion process which is a significant issue in the sinogram.

3.2. Adaptive Probabilistic Diffusivity Function Based on a Chi-Square Prior for a Sinogram. The probabilistic diffusivity function (12) does not take into account the spatially varying noise levels such as the case for the sinograms. It has a global threshold parameter, which is related to the image noise standard deviation $T = \sigma_n$. Hence, in such formulation, if two pixels/voxels have equal gradient magnitude, they will have the same $g_{pr}(x)$ values, no matter the noise level at these pixels. In this work, the probabilistic diffusivity function is improved by considering the local statistical noise at each element. We adopt the ideas of [15] and adapt the estimator to the local spatial context in the image.

3.2.1. Spatially Adaptive Probabilistic Diffusivity Function. The most appropriate way to achieve a spatial adaptation is to estimate the prior probability of signal presence $P(H_1)$ adaptively for each pixel instead of fixing it globally. This can be achieved by conditioning the hypothesis H_1 on a local spatial activity indicator such as the locally averaged magnitude or the local variance of the observed gradient.

To estimate the probability that “signal of interest” is present at the position i , we consider a local spatial activity indicator, z_i , at each position. z_i is defined as the locally averaged gradient magnitude within a relatively small window, $w(i)$, with a fixed size, N , around the position i in the gradient image.

Starting from the prior odd, (13), we replace the ratio of “global” probabilities by a locally adaptive prior, $P(H_1 | z_i)/P(H_0 | z_i)$; that is, $P(H_1)$ and $P(H_0)$ are conditioned on the local spatial indicator

$$\frac{P(H_1 | z_i)}{P(H_0 | z_i)} = \frac{P(z_i | H_1)}{P(z_i | H_0)} \cdot \frac{P(H_1)}{P(H_0)} = \xi_i(z_i) \cdot \mu, \quad (18)$$

where (ξ) is the local likelihood ratio:

$$\xi_i(z_i) = \frac{P(z_i | H_1)}{P(z_i | H_0)} \quad \mu = \frac{P(H_1)}{P(H_0)}. \quad (19)$$

Considering Bayes’ rule, the probability that an “edge of interest” is present at position i , $P(H_1 | x_i)$, is given as

$$P(H_1 | x_i) = \frac{\mu \eta_i(x_i) \xi_i(z_i)}{1 + \mu \eta_i(x_i) \xi_i(z_i)}. \quad (20)$$

The spatially adaptive probabilistic diffusivity function can then be formulated as

$$g_{apr}(x_i, z_i) = A(1 - P(H_1 | x_i, z_i)), \quad (21)$$

$$g_{apr}(x_i, z_i) = (1 + \mu \eta(0)) \left(\frac{1}{1 + \mu \eta_i(x_i) \xi_i(z_i)} \right)$$

with

$$\eta_i(x_i) = \frac{P(x_i | H_1)}{P(x_i | H_0)}. \quad (22)$$

We ensure that $g(0) = 1$, because the minimum of $P(H_1 | x)$ is at $x = 0$ and thus $(1 - P(H_1 | x))$ peaks at $x = 0$.

Intuitively, the proposed method considers an “observed gradient” at a given location as how probable this location presents useful information compared to its neighborhood, based on

- (1) the likelihood ratio via $\eta_i(x_i)$,
- (2) a measurement from the local surrounding via $\xi_i(z_i)$,
- (3) the global prior odd via μ .

The local spatial activity indicator z_i is defined as

$$z_i = \frac{1}{N} \sum_{l \in w(i)} x_l, \quad (23)$$

where x_l is the gradient magnitude at location $l \in w(i)$. Assume that all the elements within the small window are equally distributed and conditionally independent. With these simplifications, the conditional probability of z_i given H_1 in a window $w(i)$ of size N , which is denoted as $P_N(z_i | H_1)$, is given by N convolutions of $P(x_i | H_1)$ with itself as follows:

$$P_N(z_i | H_1) = P(x_i | H_1) \text{Conv}_N P(x_i | H_1), \quad (24)$$

while the conditional probability $P_N(z_i | H_0)$ of z_i given H_0 is given by N convolutions of $P(x_i | H_0)$ with itself:

$$P_N(z_i | H_0) = P(x_i | H_0) \text{Conv}_N P(x_i | H_0). \quad (25)$$

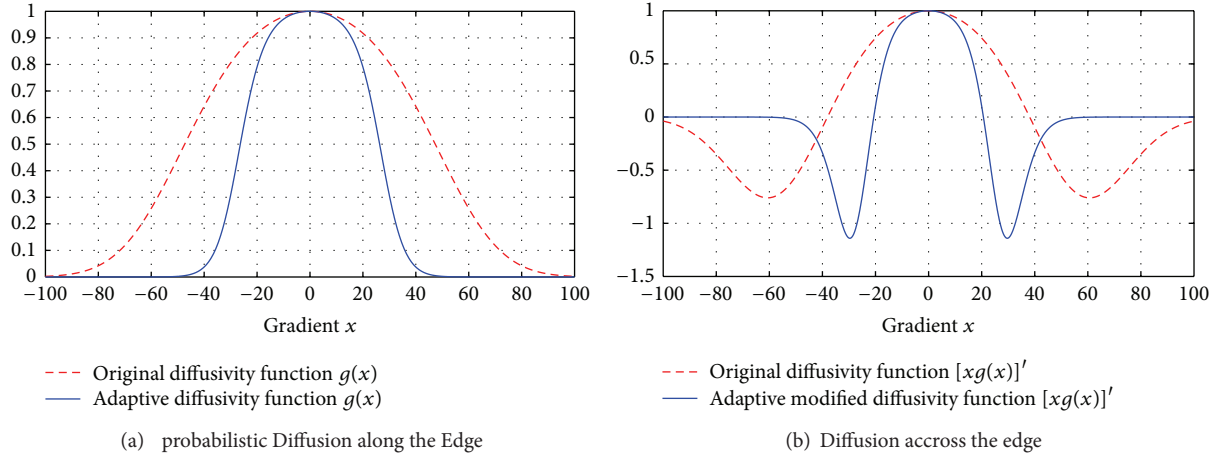


FIGURE 1: The Adaptive probabilistic diffusivity function $g_{\text{apr}}(\cdot)$ versus the original probabilistic diffusivity function $g_{\text{pr}}(\cdot)$. (a) Diffusivity functions along the edge ($g(x)$). (b) Diffusivity functions across the edge ($[xg(x)]'$).

Due to limited space, the reader is referred to [19] for the detailed expression of $P(x_i | H_0)$, $P(x_i | H_1)$, and $\eta(x_i)$.

Figure 1 illustrates the original probabilistic diffusivity function and adaptive probabilistic diffusivity function. The adaptive function has lower values, as shown in Figure 1(a), which allows better preservation of important edges than the original function. The negative peaks of the proposed function occur usually at larger gradient magnitude than the original one, which indicates stronger edge enhancement capability. This property indicates a quick smoothing of the nearly uniform areas while maintaining and enhancing the strong edges.

3.2.2. A Chi-Square Prior for Noise-Free Sinogram Gradient. The noise in the sinogram is a priori identified as a Poisson noise [12, 20]. The magnitude of Poisson noise varies across the image, as it depends on the image intensity. This makes removing such noise very difficult. In the original probabilistic diffusivity function (Section 3), the Laplacian prior was imposed for the ideal image gradient that is contaminated with Gaussian noise. In order to take into account the sinogram's noise distribution in the filtering scheme, in this section we aim at redefining the diffusivity function for handling Poisson noise. This can be accomplished by finding an appropriate prior for the ideal noise-free sinogram gradient.

In the following, we argue that we can represent Poisson distribution by a Gaussian distribution as $\text{Gauss}(0, \sqrt{2m})$. Afterwards, we demonstrate that the gradient magnitude of the noise-free sinogram follows a Chi-square distribution, and finally we reformulate the probabilistic diffusivity function based on a Chi-square prior for the noise-free sinogram gradient.

In the literature, several studies demonstrated that the Poisson distribution (of probability mass $P(S) = m^S e^{-m}/S!$, with S being the number of occurrences of an event and m being the expected number of occurrences during a given

interval) approaches a Gaussian density function in the case of high number of counts [21, 22]. Moreover, Miller et al. [21] showed that the Gaussian approximation is surprisingly accurate, even for a fairly small number of counts. To illustrate this, we use the logarithmic function to simplify the proof:

$$\ln P(S) = \ln \left(\frac{m^S e^{-m}}{S!} \right). \quad (26)$$

Using Stirling's formula (for large S as we are assuming here) $S! \approx S^S \cdot e^{-S} \sqrt{2\pi S}$, we have

$$P(S) \approx \frac{e^{-(S-m)^2/2m}}{\sqrt{2\pi m}}. \quad (27)$$

Assuming that the sinogram gradient is approximated by absolute difference of neighboring pixel values on a 2-connected grid, we demonstrate that the gradient of Poisson random variables follows a Skellam distribution. Then, we show that the Skellam distribution can be approximated as a Gaussian distribution.

Let s_1 and s_2 be two statistically independent adjacent pixels in the observed sinogram be with a Gaussian distribution $s_1 \sim \text{Gauss}(m_1, \sigma_{s_1})$ and $s_2 \sim \text{Gauss}(m_2, \sigma_{s_2})$, respectively. The distribution of the difference, $a = s_1 - s_2$, of two statistically independent random variables s_1 and s_2 , each having Poisson distributions with different expected values m_1 and m_2 , is denoted as the Skellam distribution [23] and can be given as

$$\text{PD}(a; m_1, m_2) = e^{-(m_1+m_2)} \left(\frac{m_1}{m_2} \right)^{a/2} I_{|a|} (2\sqrt{m_1 m_2}), \quad (28)$$

where $I_k(Z)$ is the modified Bessel function of the first kind.

The difference between two Poisson variables has the following properties: (i) $\sigma_{s_1 s_2}^2 = \sigma_{s_1}^2 + \sigma_{s_2}^2 = 2\sigma^2$ and (ii) $m = m_{s_1 s_2} = m_{s_1} - m_{s_2} = 0$. Considering these properties, the cross-correlation, and the delta function, the

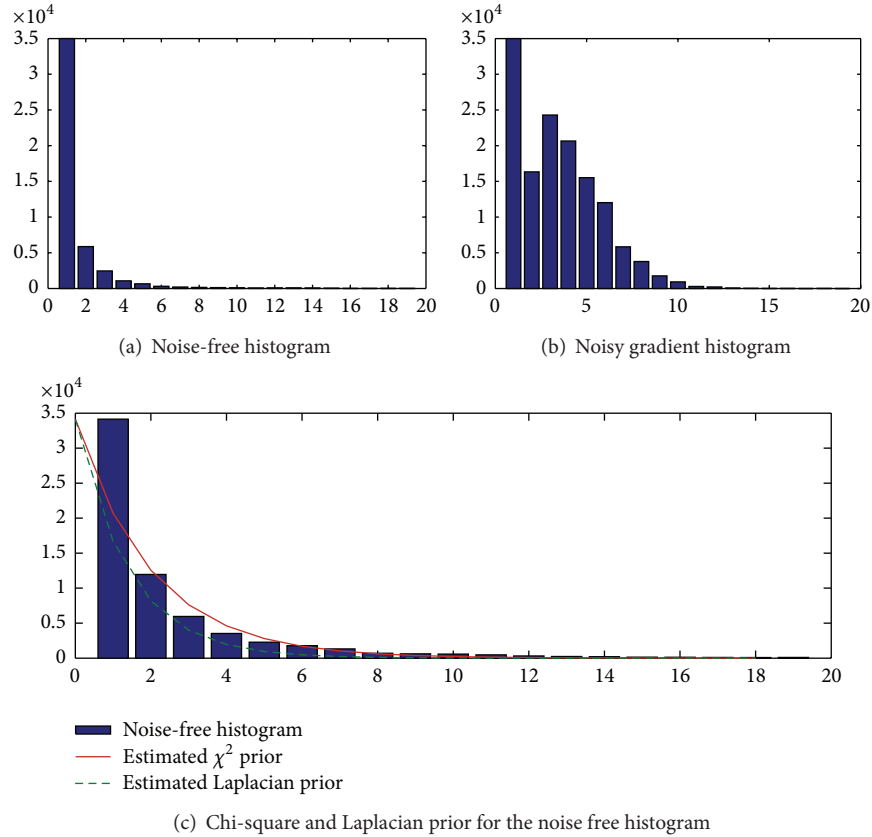


FIGURE 2: Chi-square prior versus Laplacian prior. (a) Histogram of the noise-free gradient. (b) Histogram of the noisy gradient magnitude. (c) Laplacian and Chi-square priors for the noisy-free gradient, estimated from the noisy data.

approximated distribution of the sinogram gradient can be given as $\text{Gauss}(0, \sqrt{2m})$.

Based on the assumption that the sinogram gradient follows the $\text{Gauss}(0, \sqrt{2m})$ distribution, we can show that the distribution of this gradient leads a Chi-square distribution as follows:

$$\begin{aligned} \nabla S_{(i,j)} &\sim \text{Gauss}(0, \sqrt{2m}), \\ \frac{|\nabla S_{(i,j)}|}{\sqrt{2m}} &\sim \text{Gauss}(0, 1), \\ \frac{|\nabla S_{(i,j)}|^2}{2m} &\sim \chi^2. \end{aligned} \quad (29)$$

The Chi-square distribution is defined by the following probability density function:

$$P(y) = \frac{y^{\zeta/2-1} \cdot e^{-y/2}}{2^{\zeta/2} \gamma(\zeta/2)}, \quad (30)$$

where $\gamma(\zeta/2)$ denotes the Gamma function and ζ is a positive integer that specifies the number of degrees of freedom. For the noise gradient model $x = y + n$, the Chi-square with 2

degrees of freedom (2 degrees because we are dealing with 2D images) is given as

$$P(y) = \frac{1}{2} \cdot e^{-(1/2)|y|}. \quad (31)$$

Based on the above, the prior odd, (15), can be reformulated considering Chi-square prior instead of Laplacian prior and the noise $n \sim \text{Gauss}(0, 2\sigma_n^2)$. Note that the Chi-square with 2 degrees of freedom is almost a special case of Laplacian prior with a rate parameter (scale parameter) $\lambda = 1/2$. This parameter determines the ‘‘scale’’ or statistical dispersion of the probability distribution. It indicates that the distribution of the ideal gradient is independent of a rate parameter since the Poisson noise is a pixel dependent (i.e., depends on the number of counts). Therefore, it is more natural to use Chi-square prior for estimating the ideal gradient of Poisson data rather than Laplacian prior with a parameter, λ , which is based on the variances of the image and noise gradients as indicated by (16).

In Figure 2, we illustrate the estimation of the Laplacian and Chi-square priors for the noise-free gradient. Figures 2(a) and 2(b) show the histograms of the gradient magnitudes for the noise-free and noisy images, respectively. The noise-free gradient histogram is typically sharply peaked at zero since the noise-free images typically contain large portions

of relatively uniform regions that produce negligible gradient values. Sharp edges and textured regions produce some relatively large gradients, building in this way long tails of the gradient histogram. From the noisy histogram of Figure 2(b), we estimate the parameter of the prior for the noise-free sinogram gradient. The results are illustrated in Figure 2(c), which shows the estimated Laplacian (dotted curve) and Chi-square (continues curve) priors in comparison to the ideal, noise-free histogram.

3.2.3. Algorithm: The Adaptive Probabilistic Curvature Motion Filter Based on Chi-Square Prior. In summary the general equation of the proposed Adaptive Modified Probabilistic Curvature Motion (AMPCM) filter is given by

$$u_t = g_{\text{apr}}(|\nabla u|) u_{vv} + [g_{\text{apr}}(|\nabla u|) + g'_{\text{apr}}(|\nabla u|) |\nabla u|] u_{ww}. \quad (32)$$

Let $g_1(x_i, z_i) \doteq g_{\text{apr}}(x_i, z_i)$, as given by (21), and the sharpening term $g_2(x) \doteq g_1(x_i, z_i) + xg'_1(x_i, z_i)$; the overall algorithm is given as shown in Algorithm 1.

Algorithm 1. Adaptive Modified Probabilistic Curvature Motion Filter Algorithm

(i) Build the Probabilistic Diffusivity Function:

(a) compute the prior odd based on the Chi-Square prior:

$$P(y) = \frac{1}{2} \cdot e^{-(1/2)|y|}, \quad (33)$$

$$\mu = (e^{(1/2)\sigma_n} - 1)^{-1}.$$

(ii) Build the spatially adaptive diffusivity function $g_{\text{apr}}(x_i, z_i)$, for each pixel $i = 1, \dots, \text{size}_f$:

(a) compute the local spatial activity indicator

$$z_i = \frac{1}{N} \sum_{l \in w(i)} M_l, \quad (34)$$

(b) compute the likelihood ratio for each window:

$$P_N(z_i | H_1) = P(M_i | H_1) \text{Conv}_N P(M_i | H_1), \quad (35)$$

$$P_N(z_i | H_0) = P(M_i | H_0) \text{Conv}_N P(M_i | H_0),$$

(c) compute the normalizing constant A:

$$A = 1 + \mu\eta(0) \quad (36)$$

(d) compute $P(H_1 | x_i)$ as defined by (20)

(iii) compute the diffusivity function $g(x_i, z_i)$:

$$g_1(x_i, z_i) = A(1 - P(H_1 | x_i, z_i)), \quad (37)$$

$$g_2(x_i, z_i) = g_1(x_i, z_i) + xg'_1(x_i, z_i).$$

(iv) Filter the sinogram, $u_0 = f$, based on (8) recursion, for $t = 0, \dots, (\text{number of iterations } 1)$:

$$u(x, y, t + \Delta t) = u(x, y, t) + \frac{1}{\Delta t} \left[g_1 \cdot \sin^2 \theta \cdot u(x-1, y, t) + g_2 \cdot \cos^2 \theta \cdot u(x-1, y, t) + g_1 \cdot \sin^2 \theta \cdot u(x+1, y, t) + g_2 \cdot \cos^2 \theta \cdot u(x+1, y, t) + g_1 \cdot \cos^2 \theta \cdot u(x, y+1, t) + g_2 \cdot \sin^2 \theta \cdot u(x, y+1, t) + g_1 \cdot \cos^2 \theta \cdot u(x, y-1, t) + g_2 \cdot \sin^2 \theta \cdot u(x, y-1, t) - 2 \cdot g_1 \cdot u(x, y, t) - 2 \cdot g_2 \cdot u(x, y, t) + \frac{g_2}{2} \cdot \sin \theta \cos \theta \cdot u(x+1, y+1, t) - \frac{g_1}{2} \cdot \sin \theta \cos \theta \cdot u(x+1, y+1, t) + \frac{g_2}{2} \cdot \sin \theta \cos \theta \cdot u(x-1, y-1, t) - \frac{g_1}{2} \cdot \sin \theta \cos \theta \cdot u(x-1, y-1, t) - \frac{g_2}{2} \cdot \sin \theta \cos \theta \cdot u(x+1, y-1, t) + \frac{g_1}{2} \cdot \sin \theta \cos \theta \cdot u(x+1, y-1, t) - \frac{g_2}{2} \cdot \sin \theta \cos \theta \cdot u(x-1, y+1, t) + \frac{g_1}{2} \cdot \sin \theta \cos \theta \cdot u(x-1, y+1, t) \right]. \quad (38)$$

4. Experiments and Discussion

For the analysis of the proposed denoising approaches, we use a simulated thorax PET phantom, containing three hot regions of interest (tumors) of activities 1.18, 1.8, and 1.57. 100 realizations (noisy sinograms) with added noise of 1×10^6 coincident events have been generated. Each sinogram has a size of 256×256 pixels and its spacing is 2×2 (mm/pixel), with 128 detectors and 128 projection angles. Figure 3(a) illustrates the noise-free sinogram, and Figure 3(b) illustrates a noisy realization.

For reconstructing the PET images, we adopt the Ordered Subset Expectation Maximization (OSEM) reconstruction

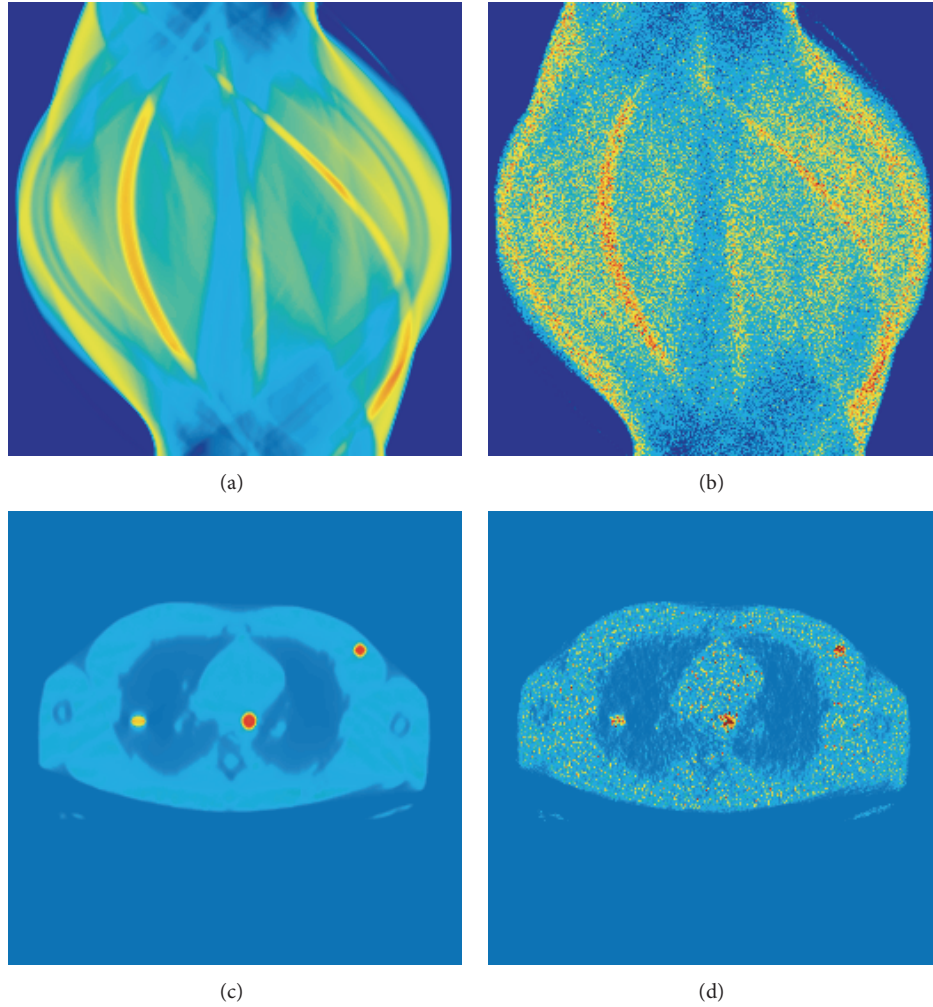


FIGURE 3: Experimental dataset. (a) Noise-free simulated sinogram. (b) Noisy sinogram. ((c)-(d)) Corresponding reconstructions.

algorithm [24]. One way to represent the imaging system is with the following linear relationship:

$$f = Hf_{\text{PET}} + n, \quad (39)$$

where f is the set of observations (sinogram), H is the known system model (projection matrix), f_{PET} is the unknown image, and n is the noise. The goal of reconstruction is to use the data values f (projections through the unknown object) to find the image f_{PET} . Under the Poisson assumption, counts of all the detectors around the object are considered as independent Poisson variables. OSEM groups projection data into an ordered sequence of subsets and progressively processes every subset of projections in each iteration process [24]. The OSEM method results are highly affected by the number of iterations and subsets used. The iterations should be ended before the noise becomes too dominant and not too early to avoid losing important information. In our experiments, the parameters of the OSEM algorithm are set as follows: we use 16 subsets and run them for 4 iterations. Such PET reconstructions are illustrated in Figures 3(c) and 3(d) for the noise-free and noisy sinograms, respectively.

For the experimental assessment of the proposed diffusion filtering, we use two sets of evaluation measures. The first set stems from measuring the quality of the filtering techniques on the sinogram, whilst the second set originates from validating the quality of the PET reconstruction, after filtering the sinogram observations. As ground-truth information, the former uses the noise-free sinogram, while the latter needs prior identification of the important areas by a medical professional.

A fundamental issue with scale-spaces induced by diffusion processes, as the ones proposed in this paper, is the automatic selection of the most salient scale. For PET sinogram denoising application, we use an earlier proposed optimal scale selection approach [25], where the maximum correlation method has been adopted:

$$\begin{aligned} t_{\text{opt}} &= \operatorname{argmax} \left[\widehat{\mathcal{E}}_{\text{mp}}(t) \right] \\ &= \operatorname{argmax} \left[\sigma[u_t] + \frac{\sigma[n_o(t_0)]}{\sigma[u_{t_0}]} \sigma[n_o(t)] \right] \end{aligned} \quad (40)$$

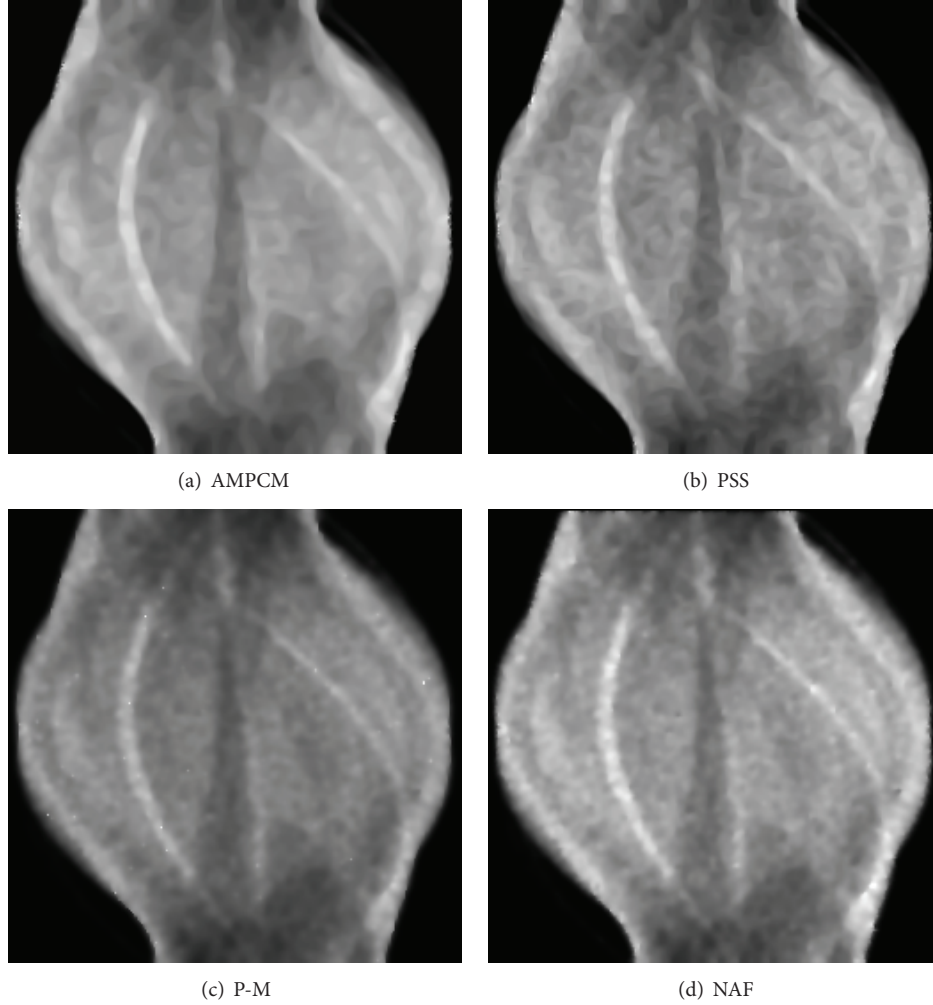


FIGURE 4: Optimal enhanced scale ($u_{t_{opt}}$) of the noisy sinogram of Figure 3(b).

with n_o being the so-called outlier noise estimated using wavelet-based noise estimation. t_0 is the zeroth scale; thus $u_{t_0} = f$ and $n(t_0)$ represents the initial amount of noise.

4.1. Sinogram Denoising Evaluation. In this work, we are interested in comparing the proposed AMPCM filter with filtering approaches from the literature: the Probabilistic self-snakes (PSS) [7], Perona and Malik filter (P-M) [1], and the Noise-Adaptive Nonlinear Diffusion Filtering (NAF) [14].

The Lorentzian diffusivity function $g(X) = 1/(1 + X^2/k^2)$ is used for applying the P-M filter. This function was proposed by Perona and Malik [1] and widely used in the literature. The contrast parameter k is calculated based on the value given by a percentage of the cumulative histogram of the gradient magnitude [16]. The same diffusivity constant is used for all filters ($d_t = 0.2$).

Figure 4 illustrates the optimal enhanced scale of the sinograms for all the considered filtering methods. The filtered sinogram by AMPCM and PSS shows better enhanced results especially the curvy shape features.

For the qualitative assessment of the denoised sinogram, $u_{t_{opt}}$, with respect to the noise-free image I , we adopt the following measures [25]:

DQ1 The Peak Signal to Noise Ratio (PSNR) is a statistical measure of error, used to determine the quality of the filtered images. It represents the ratio of a signal power to the noise power corrupting. Obviously, one sees that the higher the PSNR, the better the quality:

$$\text{PSNR}(t_{opt}) = 10 \log_{10} \frac{\text{Card}(\Omega)}{\sum_{p \in \Omega} |I(p) - u_{t_{opt}}(p)|}. \quad (41)$$

DQ2 The correlation ($\mathcal{E}_{m\rho}$) between the noise-free and the filtered image. The higher this correlation is, the better the quality is:

$$\mathcal{E}_{m\rho}(t_{opt}) = \rho[I, u_{t_{opt}}]. \quad (42)$$

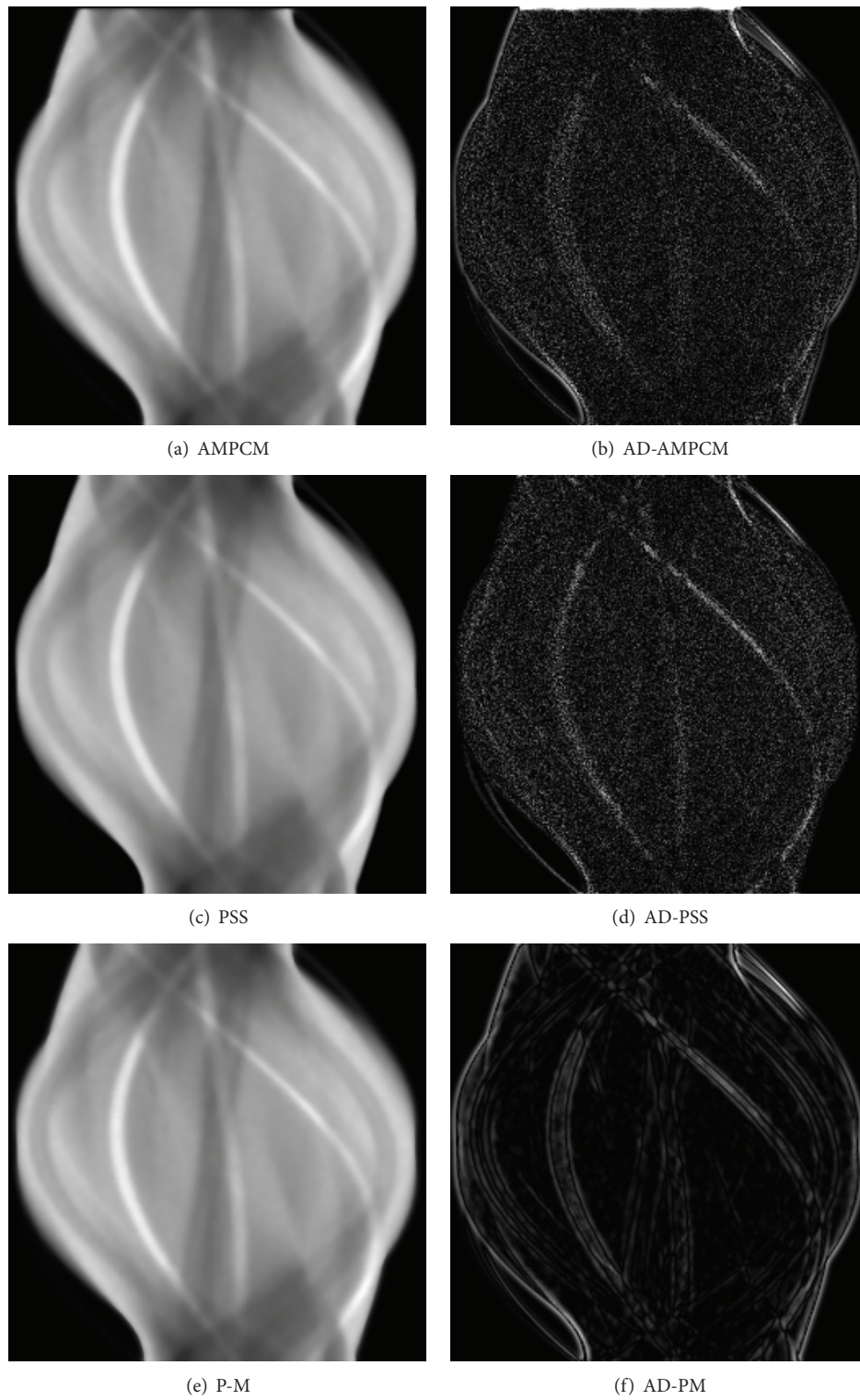


FIGURE 5: Mean of the filtered sinograms (left column). Absolute difference between the mean of the filtered sinograms and the noise-free image (right column).

TABLE 1: Denoising quality measures.

Method	f	AMPCM	PSS	P-M	NAF
PSNR (t_{opt})	15.1	30.8	29.73	22.8	29.1
NV (t_{opt})	0.11	0.02	0.0223	0.225	0.0261
$\mathcal{E}_{m_p}(t_{\text{opt}})$	0.92	0.9955	0.9950	0.9956	0.9911

DQ3 The calculated variance of the noise (NV) describes the remaining noise level. Therefore, it should be as small as possible:

$$\text{NV}(t_{\text{opt}}) = \text{Var}\left(\left|I - u_{t_{\text{opt}}}\right|\right). \quad (43)$$

We are interested in comparing the maximum of each measure for different filtering approaches.

Table 1 lists the above measures for each of the considered filtering approaches. The best performance (maximum of the measure) is displayed in bold. As it can be seen the best performing filtering is achieved when using the AMPCM and PSS. Furthermore, we notice that for all measures, AMPCM and PSS outperform the NAF and (P-M) filters. The performance of the smoothing algorithms proposed in this paper is remarkable in discriminating a true edge from image noise (see Figure 4). We also notice the improved performance of the probabilistic curvature motion algorithms as compared to the standard diffusion algorithms. Table 2 gives the number of iterations to reach the optimal scale t_{opt} as defined in (40).

The absolute difference between the mean of the filtered sinograms and the noise-free one is another indication of the behavior and stability of the filtering methods. The mean results of the filtered realizations by the PSS and AMPCM filters are close to the ideal image, as shown in Figure 5. Comparing the mean results between the AMPCM and PSS with the other methods clearly demonstrates the effect of the sharpening/enhancing term which yields a better enhanced edges. On the other hand, P-M keeps edges unsmoothed, which is appearing as sharper edges in Figure 5. Figure 5 shows more noise remained, represented as larger values in the absolute difference images of P-M and NAF. Furthermore, using AMPCM and PSS, the absolute difference approaches zero (black image), indicating that the noise has been effectively and adaptively reduced from the noisy sinograms.

The heavy noise is clearly eliminated without destroying the texture (curves) by the probabilistic curvature motion filter. From the above and other various examples, we have observed that the proposed algorithm has ability to eliminate the Poisson noise. No stair-casing has been observed, nor severe blurring is introduced. Figure 4(c) shows that P-M filter performs well for relatively low levels of noise, while it results in poor quality of images for high noise levels. However, Perona's operator does not work well when applied to very noisy images because the noise introduces important oscillations of the gradient.

4.2. PET Reconstruction Evaluation. In this section, we use the smoothed sinograms, u_r , for PET reconstruction via the OSEM algorithm. The reconstructed PET image is denoted as

TABLE 2: Optimal number of Iterations.

AMPCM	PSS	P-M	NAF
16	16	22	21

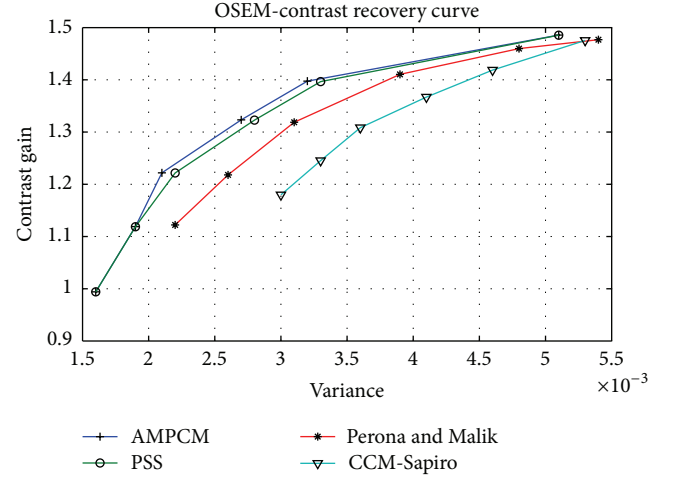


FIGURE 6: The contrast recovery curves for reconstructed PET by OSEM.

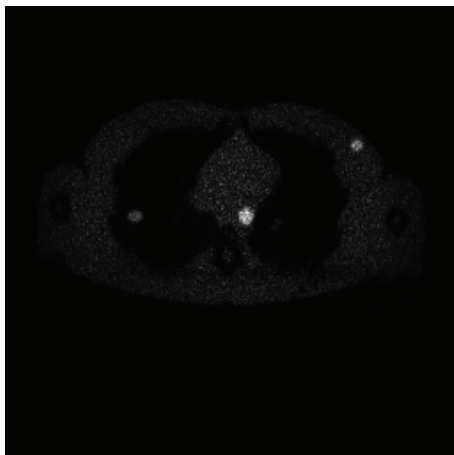
$f_{\text{PET}}(t)$. For evaluating the effect of sinogram predenoising on the PET reconstruction, we use the *contrast recovery method* which aims to measure the quality of $f_{\text{PET}}(t)$ by measuring the contrast recovery of the interesting objects (ROIs) in the image. The contrast recovery is computed based on the contrast gain. The latter measures how much the ROIs (i.e., tumor) in the PET image are discriminated from the background by sharp edges and on the variance of the contrast gain for different realizations. Further, the contrast gain evaluates the stability of the applied algorithm. The contrast recovery curve is estimated using the set of regions of interest that were identified by a medical professional. This curve is widely used in the literature for evaluating the quality of the reconstructed PET images [26].

In Figure 3(c), there are 3 white spots that represent tumors. We calculate the contrast gain CG_i for each realization f_{PET}^i , $i \in [1, N]$ ($N = 100$ denotes the number of realizations), and its overall variance σ_{CG}^2 . Let $R = \{r_1, r_2, \dots, r_{\text{no}}\}$ (no = 3 in our case) be the set of identified ROIs, and let B be a representative background tissue area then

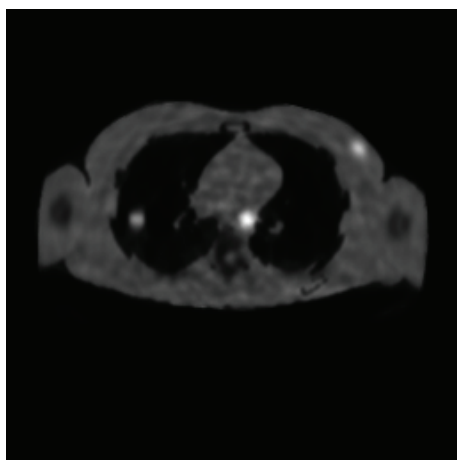
$$\text{CG}_i(t) = \frac{1}{\text{no}} \sum_{r \in R} \left[\frac{1}{\text{Card}(r)} \sum_{p \in r} f_{\text{PET}}^{(i)}(p, t) - C_{\text{bkg}} \right],$$

$$\sigma_{\text{CG}}^2(t) = \frac{1}{N} \sum_{i=1}^N (\text{CG}_i(t) - \overline{\text{CG}})^2, \quad (44)$$

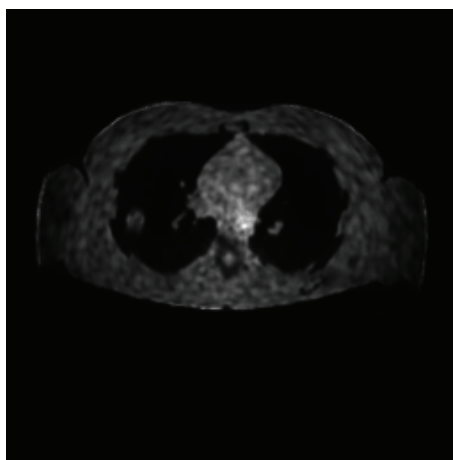
where p denotes a pixel, t the scale number, and $\overline{\text{CG}}$ the contrast mean, $\overline{\text{CG}} = (1/N) \sum_{j=1}^N \text{CG}_j(t)^2$. C_{bkg} represent



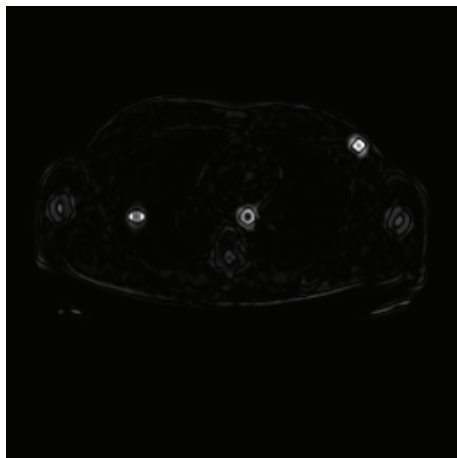
(Var-image (noisy PET))



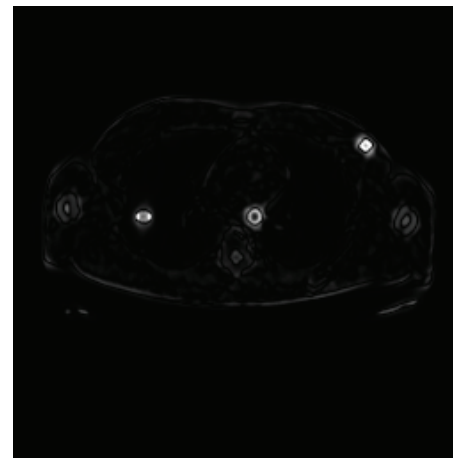
(a) Reconstructed after AMPCM



(b) Variance image of (a).



(c) Mean results of the filtered reconstruction



(d) Difference between (c) and noise-free image

FIGURE 7: Evaluation of the PET reconstruction using enhanced sinograms.

the mean intensities inside background, defined as $C_{\text{bkg}} = (1/\text{Card}(B)) \sum_{p \in B} f_{\text{PET}}^{(i)}(p, t)$.

Plotting the contrast gain in function of the variance given the smoothness parameter, which is in this work, the scale parameter t yields the contrast recovery curve [26].

In order to perform, under the same conditions, the comparison of the contrast curves of the different diffusion scheme, one should identify the scales, t , emanating from different diffusion schemes, having similar information content. For this, we use the scale synchronization proposed by

Niessen et al. [27], being a Shannon-Wiener entropy used to compare nonlinear scale-space filters:

$$\mathfrak{R} [f_{\text{PET}}(t)] \triangleq \frac{\sigma^2(f_{\text{PET}}(0)) - \sigma^2(f_{\text{PET}}(t))}{\sigma^2(f_{\text{PET}}(0)) - \sigma^2(\mathcal{M}(f_{\text{PET}}))}, \quad (45)$$

where \mathcal{M} represents the averaging operator that maps an image f_{PET} onto a constant image, where each pixel equals the average feature vector.

The contrast gain and the variance were computed for 6 scales, for each filtering methods, with similar information content, selected based on Niessen et al. [27] approach. Figure 6 shows the contrast recovery curves for the considered dif-fusion schemes AMPCM, PSS, P-M, and CCM-Sapiro [17]. The best quality PET reconstruction is situated in the upper, that is, high contrast gain, left, that is, high stability, area of the plot. The figure shows that AMPCM has better contrast recovery of the interesting objects (ROIs) in the reconstructed PET images.

Figure 7 illustrates the reconstructed PET starting from the enhanced sinogram, by the AMPCM, along with the variance images of the reconstructed PET, the mean of the reconstructed PET images, and the absolute difference between the mean of the filtered PET and the noise-free image. We immediately notice that the background seems much flatter when adopting the OSEM reconstruction.

Experiments results show that combining the probabilistic diffusivity function with the curvature motion diffusion produces a powerful nonlinear filtering method that is appropriate for the unique characteristics of the PET images. The AMPCM filter preserves the boundaries of the curvy shape features and wisely smooths the regions of interest as well as the other regions. Figure 7 demonstrates that the edge enhancement around image data is significantly improved. The effect of edge strengthening is even more pronounced for weaker edges in PET images or in image areas affected by a high level of noise.

5. Conclusions

Adaptive probabilistic curvature motion filter (AMPCM) for enhancing PET images is developed and discussed in this work. The filter is applied on the 2D sinogram pre-reconstruction. For considering the special characteristics of the sinogram data, a Chi-square is used as a marginal prior for noise-free sinogram gradient in the diffusivity function.

The qualitative evaluation results results show that, among other diffusivity functions, the probabilistic adaptive function seems more effective in smoothing all the homogenous regions that contain high level of noise. Further, AMPCM retain in an accurate way the location of the edges that defines the shape of the represented structures in the sinogram. It preserves the boundaries of the curvy shape features and wisely smooths the regions of interest as well as the other regions. The application of the proposed diffusion scheme results in well-smoothed images and preserves the edges. It gains the advantages of the curvature motion diffusion and the shock filter. Further, it deals better with the

problem of poor and discontinuous edges which are common in PET sinograms.

Applied as a preprocessing step before PET reconstruction, we demonstrated via qualitative measures (the contrast curve, the variance, and the mean images) that the adaptive probabilistic diffusion function has a great capability and stability to detect and enhance the important features in the reconstructed PET image, which make it a reliable and practical approach as it has no free parameters to be optimized. All parameters are image based and are automatically estimated and proved to give the best results.

Acknowledgment

The authors would like to thank Professor M. Defrise, Division of Nuclear Medicine at UZ-VUB, for providing them the PET phantom images, the discussions, and feedback on the Poisson noise and Chi-square prior, as well as on the evaluation of the PET reconstruction results.

References

- [1] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [2] D. Kazantsev, S. R. Arridge, S. Pedemonte et al., "An anatomically driven anisotropic diffusion filtering method for 3D SPECT reconstruction," *Physics in Medicine and Biology*, vol. 57, no. 12, pp. 3793–3810, 2012.
- [3] Z. Quan and L. Yi, "Image reconstruction algorithm for positron emission tomography with Thin Plate prior combined with an anisotropic diffusion filter," *Journal of Clinical Rehabilitative Tissue Engineering Research*, vol. 15, no. 52, 2011.
- [4] O. Demirkaya, "Post-reconstruction filtering of positron emission tomography whole-body emission images and attenuation maps using nonlinear diffusion filtering," *Academic Radiology*, vol. 11, no. 10, pp. 1105–1114, 2004.
- [5] D. R. Padfield and R. Manjeshwar, "Adaptive conductance filtering for spatially varying noise in PET images," *Progress in Biomedical Optics and Imaging*, vol. 7, no. 3, 2006.
- [6] J. Weickert, *Anisotropic Diffusion in Image Processing*, European Consortium for Mathematics in Industry, B. G. Teubner, Stuttgart, Germany, 1998.
- [7] M. Alrefaya, H. Sahli, I. Vanhamel, and D. Hao, "A nonlinear probabilistic curvature motion filter for positron emission tomography images," in *Scale Space and Variational Methods in Computer Vision*, vol. 5567 of *Lecture Notes in Computer Science*, pp. 212–2223, 2009.
- [8] O. Demirkaya, "Anisotropic diffusion filtering of PET attenuation data to improve emission images," *Physics in Medicine and Biology*, vol. 47, no. 20, pp. N271–N278, 2002.
- [9] W. Wang, "Anisotropic diffusion filtering for reconstruction of poisson noisy sinograms," *Journal of Communication and Computer*, vol. 2, no. 11, pp. 16–23, 2005.
- [10] C. Negoita, R. A. Renaut, and A. Santos, "Nonlinear anisotropic diffusion in positron emission tomography kinetic imaging," in *SIAM Conference on Imaging Science*, Salt Lake City, Utah, USA, 2004.
- [11] H. Zhu, H. Shu, J. Zhou, C. Toumoulin, and L. Luo, "Image reconstruction for positron emission tomography using fuzzy nonlinear anisotropic diffusion penalty," *Medical and Biological Engineering and Computing*, vol. 44, no. 11, pp. 983–997, 2006.

- [12] H. Zhu, H. Shu, J. Zhou, X. Bao, and L. Luo, "Bayesian algorithms for PET image reconstruction with mean curvature and Gauss curvature diffusion regularizations," *Computers in Biology and Medicine*, vol. 37, no. 6, pp. 793–804, 2007.
- [13] A. P. Happonen and M. O. Koskinen, "Experimental investigation of angular stackgram filtering for noise reduction of SPECT projection data: study with linear and nonlinear filters," *International Journal of Biomedical Imaging*, vol. 2007, Article ID 38516, 2007.
- [14] A. A. Samsonov and C. R. Johnson, "Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels," *Magnetic Resonance in Medicine*, vol. 52, no. 4, pp. 798–806, 2004.
- [15] A. Pizurica, P. Scheunders, and W. Philips, "Multiresolution multispectral image denoising based on probability of presence of features of interest," in *Proceedings of Advanced Concepts for Intelligent Vision Systems (Acivs '04)*, 2004.
- [16] I. Vanhamel, *Vector valued nonlinear diffusion and its application to image segmentation [Ph.D. thesis]*, Vrije Universiteit Brussel, Faculty of Engineering Sciences, Electronics and Informatics (ETRO), Brussels, Belgium, 2006.
- [17] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, Cambridge, UK, 2001.
- [18] A. Kuijper, "Geometrical PDEs based on second-order derivatives of gauge coordinates in image processing," *Image and Vision Computing*, vol. 27, no. 8, pp. 1023–1034, 2009.
- [19] A. Pižurica, I. Vanhamel, H. Sahli, W. Philips, and A. Katartzis, "A Bayesian formulation of edge-stopping functions in nonlinear diffusion," *IEEE Signal Processing Letters*, vol. 13, no. 8, pp. 501–504, 2006.
- [20] A. Teymurazyan, T. Riauka, H. S. Jans, and D. Robinson, "Properties of noise in positron emission tomography images reconstructed with filtered-back projection and row-action maximum likelihood algorithm," *Journal of Digital Imaging*, vol. 26, no. 3, pp. 447–456, 2013.
- [21] G. Miller, H. F. Martz, T. T. Little, and R. Guilmette, "Using exact poisson likelihood functions in Bayesian interpretation of counting measurements," *Health Physics*, vol. 83, no. 4, pp. 512–518, 2002.
- [22] H. A. Gersch, "Simple formula for the distortions in a Gaussian representation of a Poisson distribution," *American Journal of Physics*, vol. 44, no. 9, pp. 885–886, 1976.
- [23] J. G. Skellam, "The frequency distribution of the difference between two Poisson variates belonging to different populations," *Journal of the Royal Statistical Society A*, vol. 109, no. 3, p. 296, 1946.
- [24] X. Lei, H. Chen, D. Yao, and G. Luo, "An improved ordered subsets expectation maximization reconstruction," in *Advances in Natural Computation*, vol. 4221, pp. 968–971, 2006.
- [25] I. Vanhamel, C. Mihai, H. Sahli, A. Katartzis, and I. Pratikakis, "Scale selection for compact scale-space representation of vector-valued images," *International Journal of Computer Vision*, vol. 84, no. 2, pp. 194–204, 2009.
- [26] C. Comtat, P. E. Kinahan, J. A. Fessler et al., "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels," *Physics in Medicine and Biology*, vol. 47, no. 1, pp. 1–20, 2002.
- [27] W. J. Niessen, K. L. Vincken, J. A. Weickert, and M. A. Viergever, "Nonlinear multiscale representations for image segmentation," *Computer Vision and Image Understanding*, vol. 66, no. 2, pp. 233–245, 1997.

Review Article

Uniqueness, Born Approximation, and Numerical Methods for Diffuse Optical Tomography

Kiwoon Kwon

Department of Mathematics, Dongguk University, Seoul 100-715, Republic of Korea

Correspondence should be addressed to Kiwoon Kwon; kwkwon@dongguk.edu

Received 8 April 2013; Accepted 12 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Kiwoon Kwon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diffuse optical tomography (DOT) is to find optical coefficients of tissue using near infrared light. DOT as an inverse problem is described and the studies about unique determination of optical coefficients are summarized. If a priori information of the optical coefficient is known, DOT is reformulated to find a perturbation of the optical coefficients inverting the Born expansion which is an infinite series expansion with respect to the perturbation and the a priori information. Numerical methods for DOT are explained as methods inverting first- or second-order Born approximation or the Born expansion itself.

1. Introduction

DOT is to find optical coefficients of tissue using near infrared light. DOT is known to be of low cost, portable, nonionized, and nonmagnetized. And DOT has higher temporal resolution and more functional information than conventional structural medical imaging modalities such as magnetic resonance imaging (MRI) and computerized tomography (CT). For the comparison to other functional imaging modalities such as functional MRI (fMRI), photon emission tomography (PET), and electroencephalogram (EEG), see [1]. DOT is used in the area of breast imaging [2–4], functional neuroimaging [5, 6], brain computer interface (BCI) [7, 8], and the study about seizure [9, 10], newborn infants [11, 12], osteoarthritis [13], and rat brain [14, 15].

In this paper, DOT is explained as an inverse problem with respect to a forward problem formulated as an elliptic partial differential equation. Propagation of light in biological tissues is usually described by diffusion approximation equation in the frequency domain, the simplest but nontrivial approximation of the Boltzmann equation, as follows:

$$-\nabla \cdot (\kappa \nabla \Phi) + \left(\mu_a + \frac{i\omega}{c} \right) \Phi = q \quad \text{in } \Omega, \quad (1a)$$

$$\Phi + 2a\nu \cdot (\kappa \nabla \Phi) = 0 \quad \text{on } \partial\Omega, \quad (1b)$$

where ν is an outer unit normal vector, Φ is a photon density distribution, $\kappa = 1/3(\mu_a + \mu'_s)$ is an diffusion coefficient, μ_a is an absorption coefficient, μ'_s is a reduced scattering coefficient, and a is a reflection coefficient.

If $\partial\kappa/\partial\nu = 0$ on $\partial\Omega$, by setting $\Psi = \sqrt{\kappa}\Phi$ and $k = \sqrt{\Delta\sqrt{\kappa}/\sqrt{\kappa} + \mu_a/\kappa + i(\omega/c\kappa)}$ with $\text{Im}(k) \geq 0$, we have

$$-\Delta\Psi + k^2\Psi = \frac{q}{\sqrt{\kappa}} \quad \text{in } \Omega, \quad (2a)$$

$$\Psi + 2a\nu \cdot (\kappa \nabla \Psi) = 0 \quad \text{on } \partial\Omega. \quad (2b)$$

If k is constant and $q(\cdot)/\sqrt{\kappa} = \delta(\cdot, r_s)$ for some source point r_s , we have the following solution of (2a):

$$\Psi(r) = R(r, r_s) = \frac{e^{ik|r-r_s|}}{4\pi|r-r_s|}. \quad (3)$$

DOT is to find the optical coefficients μ_a and/or μ'_s from the measurement information $\Phi_{i,j}$ which is the value of the solution of (1a) and (1b) at $r_i \in \partial\Omega$ when $q(r) = \delta(r, r_j)$, $r_j \in \partial\Omega$. The r_i and r_j are usually called source and detector point, respectively.

In Section 2, the unique determination of the optical coefficients is discussed and many known results are summarized for the uniqueness questions. In Section 3, DOT is

reformulated as to find perturbation of the optical coefficient inverting the Born expansion. The errors of the Born approximation in the Lebesgue and Sobolev norms are given. In Section 4, numerical methods of DOT are mainly described as the methods inverting the first-, second-, and higher-order Born approximation and Born expansion itself.

2. Uniqueness

The research about unique determination of the optical coefficients in DOT is rare except [16], but it is a very important issue for DOT as an inverse problem. The determination of optical coefficients (μ_a, μ_s) in (1a) and (1b) is equivalent to the determination of k in (2a) and (2b) when $\omega \neq 0$.

When $\mu_a, \kappa, \nabla \kappa$ have upper and lower bound and q is contained in $H^{-1}(\Omega)$ or a Dirac delta function, (1a) and (1b) have a unique solution $\Phi \in H^1(\Omega)$ and (2a) and (2b) have a unique solution $\Psi \in H^1(\Omega)$ [17, 18].

Boundary value problem (1a) and (1b) with $q(r) = \delta(r, r_s)$ is equivalent to boundary value problem with (1a) for $q = 0$ and nonzero Robin boundary condition replacing (1b). This argument can be proved using the function H in [19]. Therefore, DOT is redescribed as to find the optical coefficients from the Robin-to-Dirichlet map defined as a map from $H^{-1/2}(\partial\Omega)$ to $H^{1/2}(\partial\Omega)$. Using unique solvability of (1a) with the Dirichlet or Neumann boundary condition replacing (1b), the Robin-to-Dirichlet map is equivalent to the Neumann-to-Dirichlet map and to the Dirichlet-to-Neumann map. Since $\text{Im}(k) > 0$ for $\omega \neq 0$, k is not a Dirichlet eigenvalue of (2a). Therefore, knowing the Dirichlet-to-Neumann map is also equivalent to knowing farfield map in inverse scattering problem [20, 21].

Using above results and many known results for inverse scattering problem, we summarized the uniqueness and nonuniqueness results as follows.

Case 1 ($\kappa = 1$, $\mu_a = 1 + (m - 1)\chi_D$, $m \neq 1$). The cases for positive constant m can be understood as special cases of Case 2. The limiting cases $m = \infty$ and $m = 0$ could be considered as (1a) in $\Omega \setminus \bar{D}$ with the Robin boundary condition (1b) on $\partial\Omega$ and the boundary condition on ∂D as $\Phi = 0$ (sound-soft case) and $\partial\Phi/\partial\nu = 0$ (sound-hard case) on ∂D . The uniqueness for sound-soft and sound-hard obstacle D is considered in [20, 22].

Case 2 ($\kappa = 1$, $\mu_a = 1 + (m(x) - 1)\chi_D$, where $m(x) \neq 1$ on ∂D). This case is called ‘‘inverse transmission problem’’ and the uniqueness is solved in [20].

Case 3 ($\kappa = 1$, $\mu_a = \mu_a(x)$). In [16], nonuniqueness of $\mu_a(x)$ is shown by assuming that refractive index is not determined or we use only continuous wave light source. However, if refractive index is known and $\omega \neq 0$, unique determination of the optical coefficient is possible. This nonuniqueness is proved by using [23].

Case 4 ($\kappa = I_n + (K(x) - I_n)\chi_D$, $\mu_a = 1 + (m(x) - 1)\chi_D$). Here, $K(x)$ is a unknown positive-definite matrix function such that $K(x) \neq I_n$ on ∂D and $m(x)$ is a positive function

such that $m(x) \neq 1$ on ∂D . The uniqueness of D is solved in [21, 24–28] and the nonuniqueness of $K(x)$ is reported in [24, 29, 30]. Therefore, although the domain of nonhomogeneity D can be uniquely determined by infinite measurements, the nonhomogeneous anisotropic diffusion coefficient $K(x)$ cannot be determined uniquely. Similar results are known for the nonuniqueness and illusion for anisotropic nonhomogeneous electric conductivity [24, 31].

3. Born Approximation

Suppose that we know a priori information $x^0 = (\kappa^0, \mu_a^0)$ about the optical coefficients $x = (\kappa, \mu_a)$ to find. Then, DOT is redescribed as to find the perturbation $\delta x = x - x^0 = (\delta\kappa, \delta\mu_a)$.

Let Φ and Φ^0 be the solutions of (1a) and (1b) for x and x^0 , respectively. When $q(r) = \delta(r, r_s)$, Φ and Φ^0 are the Robin function $R(\cdot, r_s)$ and $R^0(\cdot, r_s)$ for the optical coefficients x and x^0 , respectively. $R(\cdot, r_s)$ is expanded as the Born expansion, an infinite series, with respect to $R^0(\cdot, r_s)$ and δx as follows:

$$R(r, r_s) = R^0(r, r_s) + \mathcal{R}R^0(\cdot, r_s) + \mathcal{R}^2R^0(\cdot, r_s) + \cdots, \quad (4)$$

where

$$\begin{aligned} (\mathcal{R}\Psi)(r) &= \mathcal{R}(\delta x)\Psi(r) = (\mathcal{R}_1\Psi)(r) + (\mathcal{R}_2\Psi)(r), \\ (\mathcal{R}_1\Psi)(r) &= \mathcal{R}_1(\delta\mu_a)\Psi(r) \\ &= - \int_{\Omega} \delta\mu_a(r') R(r, r') \Psi(r') dr', \\ (\mathcal{R}_2\Psi)(r) &= \mathcal{R}_2(\delta\kappa)\Psi(r) \\ &= - \int_{\Omega} \delta\kappa(r') \nabla R(r, r') \nabla \Psi(r') dr'. \end{aligned} \quad (5)$$

It is proved that the n th order term in the above Born expansion (4) is the n th order Frechet derivative divided by $n!$ [18]. The bounds for the operators \mathcal{R}_1 and \mathcal{R}_2 are given in the Sobolev, Lebesgue, and weighted Sobolev spaces norms [18]. The estimate for the Lebesgue space norms for \mathcal{R}_1 is also given in [32]. Therefore, there exist positive constants C_1 and C_2 such that

$$\|\mathcal{R}_1\|_{W^{1,p}(\Omega) \rightarrow W^{1,p}(\Omega)} \leq C_1 \|\delta\mu_a\|_{L^\infty(\Omega)}, \quad (6a)$$

$$\|\mathcal{R}_2\|_{W^{1,p}(\Omega) \rightarrow W^{1,p}(\Omega)} \leq C_2 \|\delta\kappa\|_{L^\infty(\Omega)}, \quad (6b)$$

for $p \geq 1$. The detailed estimate for the coefficients C_1, C_2 in special cases is given in [18, 32].

If

$$C_3 = C_1 \|\delta\mu_a\|_{\infty} + C_2 \|\delta\kappa\|_{\infty} < 1 \quad (7)$$

holds, then the error of the m th order Born approximation $B^m R^0(\cdot, r_s)$ is given by

$$\|R(\cdot, r_s) - B^m R^0(\cdot, r_s)\|_{W^{1,p}} \leq C_3^m \|R(\cdot, r_s)\|_{W^{1,p}(\Omega)} \quad (8)$$

using the estimates (6a) and (6b).

4. Numerical Methods

Numerical methods for DOT are explained in terms of solving the Born approximation given in Section 3. Other numerical methods not categorized as inverting Born approximation are commented in the final section.

4.1. Linearized Methods. Linearized DOT to find δx by solving the first-order Born approximation is studied by many researchers in the initial stage of DOT. Although this method lacks exact recovery in most cases, it is used still frequently when faster real-time computing is needed and very good a priori information is given. In this method, the discretized problem is an algebraic equation and it is essential to use efficient matrix solver. In the Jacobian matrix, the number of rows is the number of measurements and the number of columns is the number of elements to be determined. In most cases, the number of elements is larger than the number of measurements in order to obtain higher resolution image. Therefore, the algebraic equation is usually an underdetermined system. Efficient matrix solvers including arithmetic reconstruction technique (ART), simultaneous arithmetic reconstruction technique (SART), simultaneous iterative reconstruction technique (SIRT), and the Krylov space methods are studied [33, 34]. Since the linearized problems also have ill-posed property of DOT, there are many researches about regularization methods including the Tikhonov regularization. The dependence on the discretization error due to ill posedness of the linearized method is reported in [35].

4.2. Nonlinear Methods. Nonlinear method is to find δx by solving the Born expansion [36]. This method is usually formulated as optimization problem and is solved by the Newton-type method including Levenberg-Marquardt method [37, 38]. A few softwares based on nonlinear method with finite element forward solver with corresponding references are as follows:

- (i) TOAST (Time-Resolved Optical Absorption and Scattering Tomography) [39],
- (ii) NIRFAST (Near InfraRed Florescence and Spectral Tomography) [40, 41],
- (iii) PMI (Photon Migration Imaging) Toolbox [42].

Nonlinear method needs heavy computation due to large iteration numbers. To reduce the heavy computation, there are studies about efficient numerical techniques such as multigrid, domain decomposition [43], and adaptive [44] method.

4.3. Inverse Born Approximation of Order Higher Than Two. This method is to find δx by solving second- and higher-order Born approximation. In fact, solving the Born approximation of order higher than two is implicit but can be approximated by explicit inverse Born approximation. Formal inverse of the Born approximation is called inverse Born approximation. The first-order inverse Born approximation corresponds to

linearized DOT and inverse Born expansion itself corresponds to (nonlinear) DOT. Higher-order methods improve the order of convergence for lower-order methods. The error of the inverse Born approximation is given and analyzed in terms of the Lebesgue space norms when $\delta\kappa = 0$ [45] and in terms of the Sobolev space norms for the second order [46].

4.4. Other Methods. The solution of (1a) and (1b) is usually solved by finite element method [36]. Another approach for the forward problem is to compute directly the Robin function using the Fourier-Laplace transform [47–50]. The disadvantage of this method is that it depends on the special geometry of the region of interest and the inverse Fourier-Laplace transform is known to be severely ill posed.

Equations (1a) and (1b) could be replaced by probabilistic approach in the Monte Carlo method [51]. The method takes much more time than finite element method and highly depends on random number generator. The comparison of finite element method and the Monte Carlo method is done by many papers including [52].

The diffusion approximation (1a) and (1b) is the first-order approximation of radiative transfer equation. There are studies about DOT based on radiative transfer equation [53, 54] and its n th order approximation [55].

5. Conclusion

Unique determination of DOT is surveyed. The study about nonuniqueness for anisotropic diffusion coefficients and for unknown refractive index is also surveyed. The perturbation of photon density with respect to the perturbed optical coefficient is expanded using the Born expansion and the error analysis removing higher-order terms is given. The numerical methods for DOT are described by inverting first-, second-, and higher-order Born approximation, and the Born expansion itself is reviewed.

Acknowledgment

This work is supported by Basic Science Research Program through the National Research foundation of Korea (NRF) funded by the Ministry of Education (2010-0004047).

References

- [1] B. White, *Developing high-density diffuse optical tomography for neuroimaging [Ph.D. thesis]*, Washington University, St. Louis, Mo, USA, 2012.
- [2] Q. Zhang, T. J. Brukilacchio, A. Li et al., "Coregistered tomographic X-ray and optical breast imaging: initial results," *Journal of biomedical optics*, vol. 10, no. 2, Article ID 024033, 2005.
- [3] T. Durduran, R. Choe, G. Yu et al., "Diffuse optical measurement of blood flow in breast tumors," *Optics Letters*, vol. 30, no. 21, pp. 2915–2917, 2005.
- [4] V. Ntziachristos, A. G. Yodh, M. Schnall, and B. Chance, "Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 6, pp. 2767–2772, 2000.

- [5] J. P. Culver, A. M. Siegel, M. A. Franceschini, J. B. Mandeville, and D. A. Boas, "Evidence that cerebral blood volume can provide brain activation maps with better spatial resolution than deoxygenated hemoglobin," *NeuroImage*, vol. 27, no. 4, pp. 947–959, 2005.
- [6] M. A. Franceschini and D. A. Boas, "Noninvasive measurement of neuronal activity with near-infrared optical imaging," *NeuroImage*, vol. 21, no. 1, pp. 372–386, 2004.
- [7] R. Sitaram, H. Zhang, C. Guan et al., "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface," *NeuroImage*, vol. 34, no. 4, pp. 1416–1427, 2007.
- [8] S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain-computer interface using a simplified functional near-infrared spectroscopy system," *Journal of Neural Engineering*, vol. 4, no. 3, pp. 219–226, 2007.
- [9] E. Watanabe, A. Maki, F. Kawaguchi, Y. Yamashita, H. Koizumi, and Y. Mayanagi, "Noninvasive cerebral blood volume measurement during seizures using multichannel near infrared spectroscopic topography," *Journal of Biomedical Optics*, vol. 5, no. 3, pp. 287–290, 2000.
- [10] K. Buchheim, H. Obrig, W. V. Pannwitz et al., "Decrease in haemoglobin oxygenation during absence seizures in adult humans," *Neuroscience Letters*, vol. 354, no. 2, pp. 119–122, 2004.
- [11] G. Taga, Y. Konishi, A. Maki, T. Tachibana, M. Fujiwara, and H. Koizumi, "Spontaneous oscillation of oxy- and deoxy-hemoglobin changes with a phase difference throughout the occipital cortex of newborn infants observed using non-invasive optical topography," *Neuroscience Letters*, vol. 282, no. 1-2, pp. 101–104, 2000.
- [12] T. Austin, A. P. Gibson, G. Branco et al., "Three dimensional optical imaging of blood volume and oxygenation in the neonatal brain," *NeuroImage*, vol. 31, no. 4, pp. 1426–1433, 2006.
- [13] Z. Yuan, Q. Zhang, E. Sobel, and H. Jiang, "Three-dimensional diffuse optical tomography of osteoarthritis: Initial results in the finger joints," *Journal of Biomedical Optics*, vol. 12, no. 3, Article ID 034001, 2007.
- [14] A. M. Siegel, J. P. Culver, J. B. Mandeville, and D. A. Boas, "Temporal comparison of functional brain imaging with diffuse optical tomography and fMRI during rat forepaw stimulation," *Physics in Medicine and Biology*, vol. 48, no. 10, pp. 1391–1403, 2003.
- [15] J. P. Culver, T. Durduran, D. Furuya, C. Cheung, J. H. Greenberg, and A. G. Yodh, "Diffuse optical tomography of cerebral blood flow, oxygenation, and metabolism in rat during focal ischemia," *Journal of Cerebral Blood Flow and Metabolism*, vol. 23, no. 8, pp. 911–924, 2003.
- [16] S. R. Arridge and W. R. B. Lionheart, "Nonuniqueness in diffusion-based optical tomography," *Optics Letters*, vol. 23, no. 11, pp. 882–884, 1998.
- [17] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, vol. 224 of *Grundlehren der Mathematischen Wissenschaften*, Springer, Berlin, Germany, 2nd edition, 1983.
- [18] K. Kwon and B. Yazıcı, "Born expansion and Fréchet derivatives in nonlinear diffuse optical tomography," *Computers & Mathematics with Applications*, vol. 59, no. 11, pp. 3377–3397, 2010.
- [19] C. Miranda, *Partial Differential Equations of Elliptic Type*, *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 2*, Springer, New York, NY, USA, 1970.
- [20] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93 of *Applied Mathematical Sciences*, Springer, Berlin, Germany, 2nd edition, 1998.
- [21] V. Isakov, "On uniqueness in the inverse transmission scattering problem," *Communications in Partial Differential Equations*, vol. 15, no. 11, pp. 1565–1587, 1990.
- [22] A. Kirsch and R. Kress, "Uniqueness in inverse obstacle scattering," *Inverse Problems*, vol. 9, no. 2, pp. 285–299, 1993.
- [23] J. Sylvester and G. Uhlmann, "A global uniqueness theorem for an inverse boundary value problem," *Annals of Mathematics. Second Series*, vol. 125, no. 1, pp. 153–169, 1987.
- [24] K. Kwon, "Identification of anisotropic anomalous region in inverse problems," *Inverse Problems*, vol. 20, no. 4, pp. 1117–1136, 2004.
- [25] K. Kwon and D. Sheen, "Anisotropic inverse conductivity and scattering problems," *Inverse Problems*, vol. 18, no. 3, pp. 745–756, 2002.
- [26] M. Piana, "On uniqueness for anisotropic inhomogeneous inverse scattering problems," *Inverse Problems*, vol. 14, no. 6, pp. 1565–1579, 1998.
- [27] R. Potthast, "On a concept of uniqueness in inverse scattering from an orthotropic medium," *Journal of Integral Equations and Applications*, vol. 11, pp. 197–215, 1999.
- [28] D. Sheen and D. Shepelsky, "Inverse scattering problem for a stratified anisotropic slab," *Inverse Problems*, vol. 15, no. 2, pp. 499–514, 1999.
- [29] G. Eskin, "Inverse scattering problem in anisotropic media," *Communications in Mathematical Physics*, vol. 199, no. 2, pp. 471–491, 1998.
- [30] F. Gyls-Colwell, "An inverse problem for the Helmholtz equation," *Inverse Problems*, vol. 12, no. 2, pp. 139–156, 1996.
- [31] K. Kwon, "The uniqueness, cloaking, and illusion for electrical impedance tomography," Submitted.
- [32] S. Moskow and J. C. Schotland, "Convergence and stability of the inverse scattering series for diffuse waves," *Inverse Problems*, vol. 24, no. 6, Article ID 065005, 16 pages, 2008.
- [33] S. R. Arridge and J. C. Schotland, "Optical tomography: forward and inverse problems," *Inverse Problems*, vol. 25, pp. 1–59, 2009.
- [34] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 2005.
- [35] M. Guven, B. Yazici, K. Kwon, E. Giladi, and X. Intes, "Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging. I," *Inverse Problems*, vol. 23, no. 3, pp. 1115–1133, 2007.
- [36] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, no. 2, pp. R41–R93, 1999.
- [37] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, pp. 164–168, 1944.
- [38] D. Marquadt, "An algorithm for least-squares estimation of nonlinear parameters," *IEEE Transactions of Circuits and Systems*, vol. 26, 1979.
- [39] TOAST website, April 2013, <http://web4.cs.ucl.ac.uk/research/vis/toast/>.
- [40] NIRFAST website, April 2013, <http://www.dartmouth.edu/nir/nirfast/index.php>.
- [41] H. Dehghani, M. E. Eames, P. K. Yalavarthy et al., "Near infrared optical tomography using NIRFAST: algorithm for numerical

- model and image reconstruction,” *Communications in Numerical Methods in Engineering with Biomedical Applications*, vol. 25, no. 6, pp. 711–732, 2009.
- [42] PMI Toolbox website, April 2013, <http://www.nmr.mgh.harvard.edu/PMI/toolbox/>.
- [43] K. Kwon, B. Yazıcı, and M. Guven, “Two-level domain decomposition methods for diffuse optical tomography,” *Inverse Problems*, vol. 22, no. 5, pp. 1533–1559, 2006.
- [44] M. Guven, B. Yazıcı, K. Kwon, E. Giladi, and X. Intes, “Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging. II,” *Inverse Problems*, vol. 23, no. 3, pp. 1135–1160, 2007.
- [45] S. Moskow and J. C. Schotland, “Numerical studies of the inverse Born series for diffuse waves,” *Inverse Problems*, vol. 25, no. 9, Article ID 095007, 18 pages, 2009.
- [46] K. Kwon, “The second-order Born approximation in diffuse optical tomography,” *Journal of Applied Mathematics*, vol. 2012, Article ID 637209, 15 pages, 2012.
- [47] V. A. Markel and J. C. Schotland, “Inverse problem in optical diffusion tomography. I. Fourier-Laplace inversion formulas,” *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 18, no. 6, pp. 1336–1347, 2001.
- [48] V. A. Markel and J. C. Schotland, “Inverse problem in optical diffusion tomography. II. Role of boundary conditions,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 19, no. 3, pp. 558–566, 2002.
- [49] V. A. Markel, V. Mital, and J. C. Schotland, “Inverse problem in optical diffusion tomography. III. Inversion formulas and singular-value decomposition,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 20, no. 5, pp. 890–902, 2003.
- [50] V. A. Markel, J. A. O’Sullivan, and J. C. Schotland, “Inverse problem in optical diffusion tomography. IV. Nonlinear inversion formulas,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 20, no. 5, pp. 903–912, 2003.
- [51] L. Wang, S. L. Jacques, and L. Zheng, “MCML—Monte Carlo modeling of light transport in multi-layered tissues,” *Computer Methods and Programs in Biomedicine*, vol. 47, no. 2, pp. 131–146, 1995.
- [52] K. Kwon, T. Son, K. J. Lee, and B. Jung, “Enhancement of light propagation depth in skin: Cross-validation of mathematical modeling methods,” *Lasers in Medical Science*, vol. 24, no. 4, pp. 605–615, 2009.
- [53] G. S. Abdoulaev and A. H. Hielscher, “Three-dimensional optical tomography with the equation of radiative transfer,” *Journal of Electronic Imaging*, vol. 12, no. 4, pp. 594–601, 2003.
- [54] T. Tarvainen, M. Vauhkonen, V. Kolehmainen, and J. P. Kaipio, “Finite element model for the coupled radiative transfer equation and diffusion approximation,” *International Journal for Numerical Methods in Engineering*, vol. 65, no. 3, pp. 383–405, 2006.
- [55] E. D. Aydin, C. R. E. de Oliveira, and A. J. H. Goddard, “A comparison between transport and diffusion calculations using a finite element-spherical harmonics radiation transport method,” *Medical Physics*, vol. 2, pp. 2013–2023, 2002.

Research Article

Corticomuscular Activity Modeling by Combining Partial Least Squares and Canonical Correlation Analysis

Xun Chen,^{1,2} Aiping Liu,¹ Z. Jane Wang,¹ and Hu Peng²

¹ Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

² Department of Biomedical Engineering, School of Medical Engineering, Hefei University of Technology, Hefei, Anhui 230009, China

Correspondence should be addressed to Hu Peng; hpeng@hfut.edu.cn

Received 8 March 2013; Accepted 9 May 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Xun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Corticomuscular activity modeling based on multiple data sets such as electroencephalography (EEG) and electromyography (EMG) signals provides a useful tool for understanding human motor control systems. In this paper, we propose modeling corticomuscular activity by combining partial least squares (PLS) and canonical correlation analysis (CCA). The proposed method takes advantage of both PLS and CCA to ensure that the extracted components are maximally correlated across two data sets and meanwhile can well explain the information within each data set. This complementary combination generalizes the statistical assumptions beyond both PLS and CCA methods. Simulations were performed to illustrate the performance of the proposed method. We also applied the proposed method to concurrent EEG and EMG data collected in a Parkinson's disease (PD) study. The results reveal several highly correlated temporal patterns between EEG and EMG signals and indicate meaningful corresponding spatial activation patterns. In PD subjects, enhanced connections between occipital region and other regions are noted, which is consistent with previous medical knowledge. The proposed framework is a promising technique for performing multisubject and bimodal data analysis.

1. Introduction

Corticomuscular activity modeling is important for assessing functional interactions in the motor control system, that is, studying simultaneous cortical and muscular activities during a sustained isometric muscle contraction. Traditionally, the most common method to assess the interactions between motor-related brain areas and the muscles is magnitude-squared coherence (MSC), which is a normalized measure of correlation between two waveforms or signals in the frequency domain. For instance, in monkeys, coherent oscillations in the 20–30 Hz band could be detected between cortical local field potentials and the rectified electromyography (EMG) from contralateral hand muscles that were modulated during different phases of a precision grip task [1]. In humans, similar findings were discovered in the beta-band corticomuscular coherence between magnetoencephalography (MEG) [2] and electroencephalography (EEG) [3] from the primary motor cortex during isometric contractions.

Although MSC has been popular in studying corticomuscular coupling, it suffers from several limitations. First, addressing the intersubject variability challenge to make a robust group inference is not straightforward with MSC because the exact frequency of maximum coupling may be inconsistent across subjects. Second, MSC emphasizes the role of individual locus in the brain in driving the motor system, while motor activity is known to be more distributed [4]. In fact, recent work has suggested that *interactions between* brain regions correspond more closely to ongoing EMG than activity at discrete sites [5–7]. Moreover, when the brain activity is measured by EEG, applying MSC directly to raw EEG and EMG signals normally yields a very low coherence value, because only a small fraction of ongoing EEG activity is related to the motor control [8]. This implies that extensive statistical testing is required to determine whether the EEG/EMG coherence is statistically significant.

Recently, several data-driven multivariate methods have been developed for analyzing biological data, and they seem

to be appropriate for modeling corticomuscular activity because these methods explore dependency relationships between data sets. These methods include multiple linear regression, principal component regression, partial least squares (PLS), and canonical correlation analysis (CCA) [9]. Among these methods, the latent-variable- (LV-) based approaches, such as PLS and CCA, play a dominating role, probably due to the fact that the extracted LVs could help the biological interpretations of the results.

PLS, first developed for process monitoring in chemical industry, exploits the covariation between predictor variables and response variables and finds a new set of latent components that maximally relate to them [10]. An advantage of PLS is that PLS can handle high-dimensional and collinear data, which is often the case in real-world biological applications. PLS and its variants have been investigated in many medical applications, such as assessing the spatial patterns of brain activity in functional magnetic resonance imaging (fMRI) data associated with behavioural measures [11] and the common temporal components between EEG and fMRI signals [12]. In addition to the ability of handling high-dimensional and collinear data, PLS is sufficiently flexible that it can be extended to perform group level analysis and to accommodate multiway data [5].

CCA is commonly used to seek a pair of linear transformations between two sets of variables, such that the data are maximally correlated in the transformed space. Generally CCA is not as popular as PLS in practical applications [13]. This is probably because real-world data are usually high dimensional and collinear, and thus applying CCA directly to the raw data can be ill-conditioned. However, with some appropriate preprocessing strategies, CCA has been shown to be quite useful in many medical applications. For instance, in [14], Clercq et al. successfully removed muscle artifacts from a real ictal EEG recording without altering the recorded underlying ictal activity. In [15], Gumus et al. found that there were significant correlations at expected places, indicating a palindromic behavior surrounding the viral integration site. CCA can be extended to accommodate multiple data sets simultaneously [16].

Although PLS and CCA have been investigated in many medical applications, to the best of our knowledge no report has profoundly explored their underlying differences, compared their characteristic performances, and combined their advantages to overcome their drawbacks. For corticomuscular activity modeling, as we will elaborate more in Section 2, both PLS and CCA have their advantages and disadvantages, but perhaps more importantly, these two methods can be considered complementary. In this paper, we propose combining PLS and CCA to improve the performance of the joint LV extraction, and the proposed method is denoted as PLS + CCA. More specifically, the proposed PLS + CCA has a two-step modeling strategy: we first adopt PLS to obtain LVs across two data sets and then perform CCA on the extracted LVs. In the first step, PLS is performed for preliminary LV preparation. The aim of this step is to extract LVs which can most explain its own data set and meanwhile are well correlated with the LVs in the other data set. Besides, this step can also prevent the ill-conditioned problem when

applying CCA directly to the raw data. In the second step, CCA is applied to the extracted LVs by PLS to construct the LVs by maximizing the correlation coefficients. With these two steps, it is ensured that the extracted components are maximally correlated across two data sets and meanwhile can well explain the information within individual data sets.

We will evaluate the performance of the proposed method on both synthetic data and real-world data. We first illustrate its performance using simulations, and we then apply the method to concurrent EEG and EMG data collected from patients with Parkinson's disease (PD) and age-matched normal subjects when they perform a dynamic, visually guided tracking task. We note highly correlated temporal patterns between EEG and EMG signals and meaningful spatial activation patterns. While the proposed method is intentionally proposed for corticomuscular coupling analysis, it can also be applied to analyze other types of concurrent signals, including, but not limited to, fMRI, photoplethysmograph (PPG), electrocardiography (ECG), and kinematic data.

2. Materials and Methods

2.1. Methods. In this section, we first analyze the properties of PLS and CCA and demonstrate their complementarity. Based on this observation, we then propose combining the two approaches to have the PLS + CCA method. The two zero-mean data sets are stored in two matrices, the predictor matrix $\mathbf{X}(N \times p)$ and the response matrix $\mathbf{Y}(N \times q)$, where N means the number of observations and p and q indicate the numbers of variables in corresponding matrices.

2.1.1. Partial Least Squares. PLS exploits the covariation between predictor variables and response variables and tries to find a new set of LVs that maximally relate to them [13]. In other words, the covariance between the extracted LVs should be maximized as

$$\begin{aligned} \max_{\mathbf{w}_1, \mathbf{w}_2} & \quad (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2, \\ \text{s.t.} & \quad \mathbf{w}_i^T \mathbf{w}_i = 1, \quad i = 1, 2, \end{aligned} \quad (1)$$

where \mathbf{w}_i 's ($i = 1, 2$) are the weight vectors. A typical PLS can be implemented by the classical NIPALS algorithm [10]. Also, an alternative calculation way is to perform eigenvalue-eigenvector decomposition [17]. Therefore, the maximum of (1) is achieved by having \mathbf{w}_1 and \mathbf{w}_2 as the largest eigenvectors of the matrices $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$ and $\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}$, respectively. To obtain subsequent weights, the algorithm is repeated with deflated \mathbf{X} and \mathbf{Y} matrices. The detailed calculation procedure can be found in the appendix.

The number of components to be extracted is a very important parameter of a PLS model. Although it is possible to extract as many PLS components as the rank of the data matrix \mathbf{X} , not all of them are generally used. The main reasons for this are the following: the measured data are never noise-free and some small components only describe noise, and it is common to ignore small components because of the

problems of collinearity. Therefore, appropriate measures are needed to determine when to stop. Typically, the number of components needed to describe the data matrices is determined based on the amount of variation remained in the residual data [10].

2.1.2. Canonical Correlation Analysis. Different from PLS, CCA is to find linear combinations of both \mathbf{X} and \mathbf{Y} variables which have maximum correlation coefficient with each other. This leads to the same objective function but different constraints compared with (1):

$$\begin{aligned} \max_{\mathbf{v}_1, \mathbf{v}_2} & \quad (\mathbf{v}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{v}_2)^2 \\ \text{s.t.} & \quad \mathbf{v}_1^T \mathbf{X}^T \mathbf{X} \mathbf{v}_1 = 1, \quad \mathbf{v}_2^T \mathbf{Y}^T \mathbf{Y} \mathbf{v}_2 = 1, \end{aligned} \quad (2)$$

where \mathbf{v}_i 's ($i = 1, 2$) are the weight vectors.

The solutions to this problem are the largest eigenvectors of the matrices $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X}$ and $(\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$, respectively. The subsequent weights are the eigenvectors of the same matrix in the order of decreasing eigenvalues. The predictor LVs \mathbf{U}_X can be calculated directly from the original \mathbf{X} matrix as $\mathbf{U}_X = \mathbf{X} \mathbf{V}_1$, the columns of which are uncorrelated with each other. The detailed derivation is shown in the appendix. However, the solution depends heavily on whether or not the covariance matrix $\mathbf{X}^T \mathbf{X}$ is invertible. In practice, it is possible to have $\text{rank}(\mathbf{X}^T \mathbf{X}) < p$ so that the invertibility cannot be satisfied, and directly applying eigenvalue decomposition in the raw data space may lead to the ill-conditioned problem. Therefore, some appropriate preprocessing strategies are needed in practice before applying CCA.

2.1.3. The Combined PLS + CCA Method. Based on the discussion above, we can see that the fundamental difference between PLS and CCA is that PLS maximizes the covariance while CCA maximizes the correlation. The objective of PLS is to construct LVs which could most explain their own data set and meanwhile are well correlated with the corresponding LVs in the other set. In other words, the first priority of PLS is to find the LVs which can explain significant proportion of variance in each data set, and the second priority is to find the LVs with relatively high correlation coefficients between the two data sets. In contrast, the only objective of CCA in the construction of LVs is to maximize their correlation coefficients with the LVs in another data set. From this point of view, the LVs extracted by PLS are able to represent major information for individual data sets while the ones extracted by CCA may be trivial (e.g., noises with similar patterns) even if their correlation coefficient is maximum. This is an advantage of PLS over CCA. Besides, PLS can handle high-dimensional and collinear data, which is often the case in real-world biological applications, while applying CCA directly to the raw data may be ill-conditioned. However, we should note that our goal is to find the relationships between two data sets, not just to explore the information within individual data sets. It is possible that a higher covariance merely results from the larger variance of

LVs, which may not necessarily imply strong correlations. To overcome this, CCA is a powerful tool to ensure that the extracted LVs have similar patterns across the data sets.

For corticomuscular activity modeling, the coupling relationships between EEG and EMG signals are what to be explored. In practice, EEG and EMG signals can be contaminated by other types of signals and are never noise-free. In addition, the signals from adjacent channels generally are similar, which leads to collinear data. By employing PLS, we can deal with the collinear EEG/EMG data sets and extract significant LVs, but it cannot guarantee that the corresponding LVs are highly correlated with each other. With using CCA, we can extract highly correlated LVs from EEG and EMG signals, but it cannot ensure that such LVs are nontrivial and we may face the ill-conditioned problem.

For corticomuscular coupling analysis, both PLS and CCA have their advantages and disadvantages, but perhaps most importantly, these two methods can be considered complementary. It is natural for us to think of combining PLS and CCA to form a two-step modeling strategy. In the first step, PLS is performed for preliminary LV preparation. The aim of this step is to extract LVs which can most explain its own data set and meanwhile are well correlated to the LVs in another data set. In this case, the trivial and irrelevant information across data sets could be removed. Besides, this step can also prevent the ill-conditioned problem when applying CCA directly to the raw data. In the second step, CCA is applied to the prepared LVs by PLS to construct the LVs by maximizing the correlation coefficients. After these two steps, it is ensured that the extracted components are maximally correlated across data sets and meanwhile can well explain the information within each individual data set. The details of the proposed PLS + CCA method are given in the appendix, and the specific implementation procedure is shown in Algorithm 1.

2.2. Synthetic Data. In this simulation, we applied the proposed method to synthetic data and also reported the results of the PLS and CCA approaches for comparison. As an illustrative example, without loss of generality, four sources were generated and analyzed for each data set.

The following four source signals were considered for the data set \mathbf{X} :

$$\begin{aligned} \mathbf{s}_{11} &= 1.5 \sin(0.025(t + 63)) \sin(0.2t), \\ \mathbf{s}_{12} &= 1.5 \sin(0.025t), \\ \mathbf{s}_{13} &= \text{sign}(\sin(0.3t) + 3 \cos(0.1t)), \\ \mathbf{s}_{14} &= \text{uniformly distributed noise in the range } [-1.5, 1.5], \end{aligned} \quad (3)$$

where t denotes the time index vector, valued from 1 to 1000, and \mathbf{s}_i 's ($i = 1, 2, 3, 4$) represent four simulated sources, as shown in Figure 1(a). Note that here \mathbf{s}_i 's are column vectors.

Input: two data sets \mathbf{X} (with size $N \times p$) and \mathbf{Y} (with size $N \times q$)
Output: corresponding LVs matrices \mathbf{U}_X , and \mathbf{U}_Y

The First Step:

- (1) Solve the eigen decomposition problems:
 $(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}) \mathbf{w}_1 = \lambda_1 \mathbf{w}_1$ and $(\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}) \mathbf{w}_2 = \lambda_2 \mathbf{w}_2$.
- (2) Determine R_1 and R_2 , the numbers of LVs extracted, corresponding to the above two problems by the ratio of explained variance.
- (3) Determine the final number of LVs: $R = \min(R_1, R_2)$.
- (4) Set *count* = R .
- (5) Initialize both LVs matrices to be empty, that is, $\mathbf{T}_X = []$ and $\mathbf{T}_Y = []$.
- (6) **while** *count* > 0 **do**
- (7) Set \mathbf{w}_1 and \mathbf{w}_2 to be the largest eigenvectors of the matrices $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$ and $\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y}$, respectively.
- (8) Calculate the LVs as $\mathbf{t}_X = \mathbf{X} \mathbf{w}_1$ and $\mathbf{t}_Y = \mathbf{Y} \mathbf{w}_2$.
- (9) Set $\mathbf{T}_X = [\mathbf{T}_X \ \mathbf{t}_X]$ and $\mathbf{T}_Y = [\mathbf{T}_Y \ \mathbf{t}_Y]$.
- (10) Deflate \mathbf{X} by subtracting the effects of the LV \mathbf{t}_X from the data space:
 $\mathbf{X} = \mathbf{X} - \mathbf{t}_X (\mathbf{t}_X^T \mathbf{t}_X)^{-1} \mathbf{t}_X^T \mathbf{X}$.
- (11) Deflate \mathbf{Y} by subtracting the effects of the LV \mathbf{t}_Y from the data space:
 $\mathbf{Y} = \mathbf{Y} - \mathbf{t}_Y (\mathbf{t}_Y^T \mathbf{t}_Y)^{-1} \mathbf{t}_Y^T \mathbf{Y}$.
- (12) Let *count* = *count* - 1.
- (13) **end while**

The Second Step:

- (14) Solve the following eigen decomposition problems:
 $[(\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y (\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X] \mathbf{v}_1 = \eta_1 \mathbf{v}_1$ and
 $[(\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X (\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y] \mathbf{v}_2 = \eta_2 \mathbf{v}_2$.
- (15) Set \mathbf{V}_1 and \mathbf{V}_2 to be the R associated eigenvectors, respectively.
- (16) The recovered LVs \mathbf{U}_X and \mathbf{U}_Y can be calculated by
 $\mathbf{U}_X = \mathbf{T}_X \mathbf{V}_1$ and $\mathbf{U}_Y = \mathbf{T}_Y \mathbf{V}_2$.

ALGORITHM 1: The combined PLS + CCA method.

TABLE 1: The correlation coefficients between the corresponding source pairs of \mathbf{X} and \mathbf{Y} .

	\mathbf{s}_{11} and \mathbf{s}_{21}	\mathbf{s}_{12} and \mathbf{s}_{22}	\mathbf{s}_{13} and \mathbf{s}_{23}	\mathbf{s}_{14} and \mathbf{s}_{24}
CC*	0.3655	0.8787	0.5520	1.00

*Here CC stands for correlation coefficient between two source signals.

Also, four source signals were considered for the data set \mathbf{Y} :

$$\begin{aligned}
 \mathbf{s}_{21} &= 1.5 \sin(0.025(t + 69)) \sin(0.2(t + 6)), \\
 \mathbf{s}_{22} &= 1.5 \sin(0.025(t + 20)), \\
 \mathbf{s}_{23} &= \text{sign}(\sin(0.3(t + 7)) + 3 \cos(0.1(t + 7))), \\
 \mathbf{s}_{24} &= \text{uniformly distributed noise (the same as } \mathbf{s}_{14}),
 \end{aligned} \tag{4}$$

where the notations are similarly defined. The four simulated sources are shown in Figure 1(b).

Two mixed data sets \mathbf{X} and \mathbf{Y} were generated as follows with each row denoting one observation in their respective data space:

$$\mathbf{X} = \mathbf{S}_1 \cdot \mathbf{A}, \quad \mathbf{Y} = \mathbf{S}_2 \cdot \mathbf{B}, \tag{5}$$

where $\mathbf{S}_1 = [\mathbf{s}_{11} \ \mathbf{s}_{12} \ \mathbf{s}_{13} \ \mathbf{s}_{14}]$ and $\mathbf{S}_2 = [\mathbf{s}_{21} \ \mathbf{s}_{22} \ \mathbf{s}_{23} \ \mathbf{s}_{24}]$ with

$$\begin{aligned}
 \mathbf{A} &= \begin{bmatrix} 0.76 & -0.65 & 0.77 & 0.83 & 0.82 \\ 0.49 & 0.25 & 0.12 & 0.22 & -0.17 \\ 0.28 & -0.21 & 0.11 & 0.19 & -0.11 \\ 0.07 & 0.06 & -0.08 & 0.07 & -0.04 \end{bmatrix}, \\
 \mathbf{B} &= \begin{bmatrix} 0.73 & -0.82 & 0.91 & -0.79 & 0.88 \\ 0.42 & -0.27 & 0.17 & -0.20 & -0.30 \\ 0.27 & 0.26 & -0.18 & 0.17 & -0.24 \\ 0.08 & -0.01 & 0.01 & 0.09 & -0.01 \end{bmatrix}.
 \end{aligned} \tag{6}$$

The patterns of the corresponding sources are similar across the two data sets, representing common information. However, from (3) and (4), we can see that there are some time-shifts between corresponding source pairs, and their correlation coefficients are given in Table 1. The first pair sources have the lowest CC, but in the mixed data sets we intentionally assign the highest weights to this pair of sources, as shown in the mixing matrices \mathbf{A} and \mathbf{B} . This pair can represent the major information within individual data sets but cannot reflect too much the coupling relationships between the two sets. The second and third pairs have relatively high CCs and moderate weights in the mixed data sets. These two pairs generally not only contain the major information within individual data sets, but also represent the coupling relationships across data sets. The fourth pair

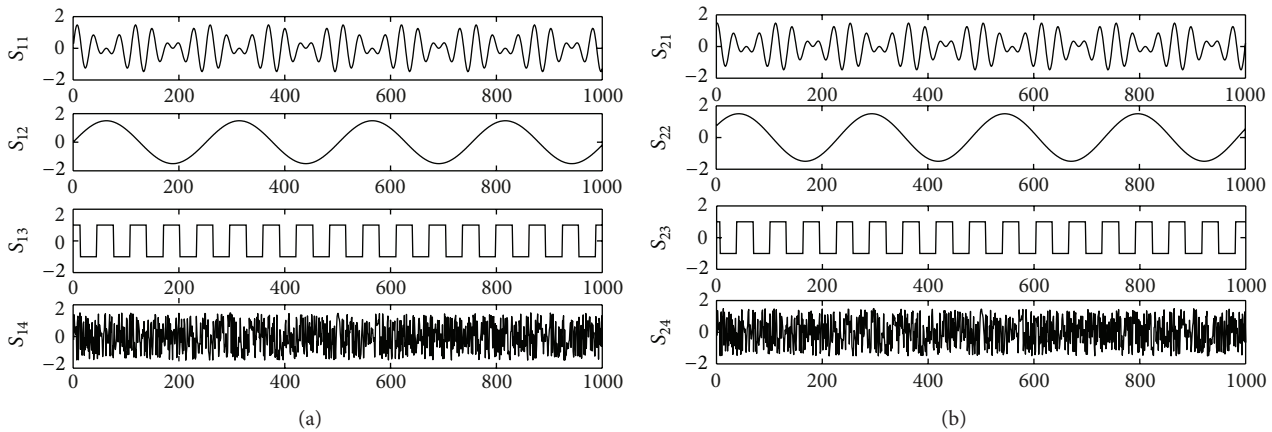


FIGURE 1: The four simulated source signals: (a) for X ; (b) for Y .

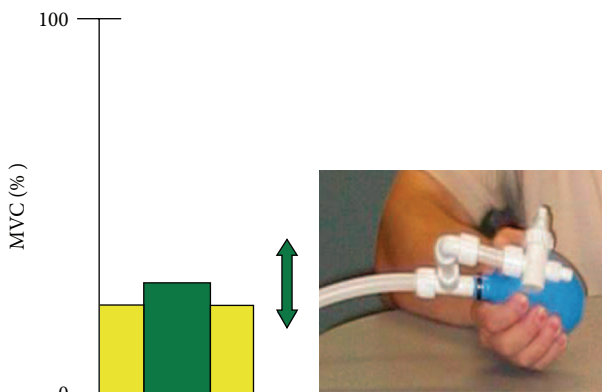


FIGURE 2: The squeezing task: the subject was instructed to follow the target bar (yellow) as closely as possible, and the green bar shows the force exerted by the subject.

sources have the highest CC, but we assign the smallest weights. Although this pair sources have the highest CC, they do not represent significant information due to the small weights. Generally, they could be regarded as trivial information. Moreover, different white Gaussian noise with 10% power was added to each source in each data space.

2.3. Real Data. In many medical applications, the results of analyzing one subject's data cannot be generalized to the population level because of the intersubject variability concern. Therefore, it is necessary to recruit a proper number of subjects to perform a group analysis in many medical applications. For modeling the corticomuscular activity, we apply the proposed method to concurrent EEG and EMG signals collected from normal subjects and patients with PD during a motor task.

2.3.1. Data Collection. The study was approved by the University of British Columbia Ethics Board, and all subjects gave written, informed consent prior to participating. Nine PD patients (mean age: 66 yrs) were recruited from the Pacific

Parkinson's Research Centre at the University of British Columbia (Vancouver, Canada). They all displayed mild to moderate levels of PD severity (stages 1-2 on the Hoehn and Yahr scale) and were being treated with L-dopa medication (mean daily dose of 720 mg). All PD subjects were assessed after a minimum of 12-hour withdrawal of L-dopa medication, and their motor symptoms were assessed using the Unified Parkinson's Disease Rating Scale (UPDRS), resulting in a mean score of 23. In addition, eight age-matched healthy subjects were recruited as controls. During the experiment, subjects seated 2 m away from a large computer screen. The visual target was displayed on the screen as a vertical yellow bar oscillating in height at 0.4 Hz. Subjects were asked to squeeze a pressure-responsive bulb with their right hand. The visual feedback representing the force output of the subject was displayed as a vertical green bar superimposed on the target bar as shown in Figure 2. Applying greater pressure to the bulb increased the height of the green bar, and releasing pressure from the bulb decreased the height of the green bar. Subjects were instructed to make the height of the green bar match the height of target bar as closely as possible. Each squeezing period lasted for 15 seconds and was followed by a 15-second rest period. The squeezing task was performed twice. The force required was up to 10% of each subject's maximum voluntary contraction (MVC), which was measured at the beginning of each recording session.

The EEG data were collected using an EEG cap (Quick-Cap, Compumedics, TX, USA) with 19 electrodes based on the International 10–20 system. The EEG data were sampled at 1000 Hz using SynAmps2 amplifiers (NeuroScan, Compumedics, TX, USA). A surface electrode on the tip of the nose was used as ground. Ocular movement artifacts were measured using surface electrodes placed above and below the eyes (Xltek, ON, Canada). Data were later processed by a band-pass filter (1 to 70 Hz) offline and downsampled to 250 Hz. Artifacts associated with eye blinks and muscular activities were removed using the Automated Artifact Removal in the EEGLAB Matlab Toolbox [18]. The raw time sequences of the electrodes were then normalized to have zero-mean and unit variance. For subsequent analysis, data

collected during the squeezing periods were concatenated in time into a single matrix for each individual subject. Data from the rest periods were excluded from the analysis.

The EMG signals were recorded using self-adhesive, silver, silver-chloride pellet surface electrodes with 7 mm diameter. A bipolar montage was used with a fixed inter-electrode distance of 30 mm. The surface EMG signals were simultaneously collected together with the EEG signals and were amplified and sampled at 1000 Hz. To be consistent with the EEG preprocessing, the EMG signals were downsampled offline to 250 Hz, and only the squeezing periods were used for subsequent analysis.

2.3.2. Feature Extraction. In most existing studies, the analysis for corticomuscular coupling is performed directly on the raw EEG and EMG data. This typically yields quite small correlation values. Nonetheless, with appropriate preprocessing steps, highly correlated EEG and EMG feature(s) can be extracted from the raw signals. In this work, we examine the coupling relationships between time-varying EEG features and amplitudes of the EMG signals, constituting \mathbf{X}_b and \mathbf{Y}_b , respectively, for each subject b (for $b = 1, 2, \dots, B$). We have a total of B subjects ($B = 17$ in this study). To achieve a group analysis, all subjects' data sets are concatenated together as

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_B], \quad \forall b = 1, 2, \dots, B, \\ \mathbf{Y} &= [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_B], \quad \forall b = 1, 2, \dots, B \end{aligned} \quad (7)$$

with the assumption that all subjects share common group patterns in the temporal dimension [19].

EEG Features. Pairwise Pearson's correlations [20] are considered in this study. Pearson's correlation measures the dependency between a pair of EEG signals $\mathbf{e}^* = (e_1^*, e_2^*, \dots, e_n^*)$ and $\mathbf{e}^\circ = (e_1^\circ, e_2^\circ, \dots, e_n^\circ)$ in the time domain as follows:

$$\gamma_{e^*e^\circ} = \frac{\sum_{i=1}^n (e_i^* - \bar{e}^*)(e_i^\circ - \bar{e}^\circ)}{\sqrt{\sum_{i=1}^n (e_i^* - \bar{e}^*)^2 \sum_{i=1}^n (e_i^\circ - \bar{e}^\circ)^2}}, \quad (8)$$

where \bar{e}^* and \bar{e}° are the sample means of \mathbf{e}^* and \mathbf{e}° . In this work, we calculate the time-varying pairwise correlations between EEG channels, using a Hamming window with length 300 and with a 95% overlap. Therefore, the raw EEG information can be represented by a two-dimensional matrix with size $N \times M$, where the rows correspond to the samples at different time points and the columns correspond to the features, that is, pairwise correlations between the EEG channels.

EMG Features. An individual EMG channel signal can be considered as a zero-mean, band-limited and wide-sense stationary stochastic process modulated by the EMG amplitude, which represents the overall muscle activity of individual underlying muscle fibers [21]. While different techniques have been proposed for accurate amplitude estimation, in this study, we employ the root-mean-square (RMS) approach to

calculate the EMG amplitude of short-duration EMG signals $\mathbf{e} = (e_1, e_2, \dots, e_n)$:

$$e_{\text{rms}} = \sqrt{\frac{1}{n} (e_1^2 + e_2^2 + \dots + e_n^2)}. \quad (9)$$

A moving window with length $n = 300$ and a 95% overlap is applied here, the same as in the EEG feature calculation, to ensure that the obtained EEG and EMG features are temporally aligned and matched.

In the above setting, for each subject b (for $b = 1, 2, \dots, B$), \mathbf{X}_b and \mathbf{Y}_b represent the time-varying feature matrices of EEG and EMG, respectively. The length of the time sequences here is 480 associated with the 300-length moving window and a 95% overlap. For the EEG correlation feature, since we have 19 EEG channels based on the International 10–20 system, thus there are a total of $C_2^{19} = 171$ correlation connections. Therefore, \mathbf{X}_b is of size 480×171 . For the EMG amplitude feature, since there are three surface EMG channels, \mathbf{Y}_b is of size 480×3 .

2.3.3. Significance Assessment. To determine the statistical significance levels of the extracted LVs, we employ a nonparametric permutation test [22] in which the temporal order of EEG features \mathbf{X}_b is uniformly permuted for all subjects while keeping the EMG features \mathbf{Y}_b intact. Two hundred random permutations are generated. The proposed PLS + CCA method described in Section 2.1.3 is applied to each of these permutations. The correlation coefficients among the extracted temporal patterns from permuted EEG features and unchanged EMG features are then calculated to form an empirical null distribution. The p value of the original EEG/EMG correlation coefficient is then computed from the null distribution as the proportion of sampled permutations whose correlation coefficients are greater than or equal to the original correlation coefficient. The components with p value being less than 0.05 are considered to be statistically significant, denoted as \mathbf{LV}_{EEG} and \mathbf{LV}_{EMG} , both with size $(N \times K)$, where K means the number of significant components.

2.3.4. Spatial Pattern Extraction. Our goal is to investigate the differences in spatial patterns of EEG channels between the normal and PD patient groups when the subjects perform a motor task. After the identification of significant temporal patterns, we can regress the EEG-related components \mathbf{LV}_{EEG} back onto the EEG features \mathbf{X}_b (for $b = 1, 2, \dots, B$) for each subject as follows:

$$\mathbf{p}_{bk} = \sqrt{\frac{1}{\mathbf{lv}_k^T \mathbf{X}_b \mathbf{X}_b^T \mathbf{lv}_k}} \mathbf{X}_b^T \mathbf{lv}_k, \quad k = 1, 2, \dots, K, \quad (10)$$

where \mathbf{lv}_k is the k th column of \mathbf{LV}_{EEG} and \mathbf{p}_{bk} is the spatial pattern of the k th component for subject b . In addition, we also want to determine which EEG features in the spatial patterns have significant contributions to the corresponding temporal patterns. This is done by identifying EEG features that have weights statistically different from zero. To determine the group-level spatial pattern, for each significant component, we apply a two-tailed t -test to each element of the spatial patterns of all subjects with each group.

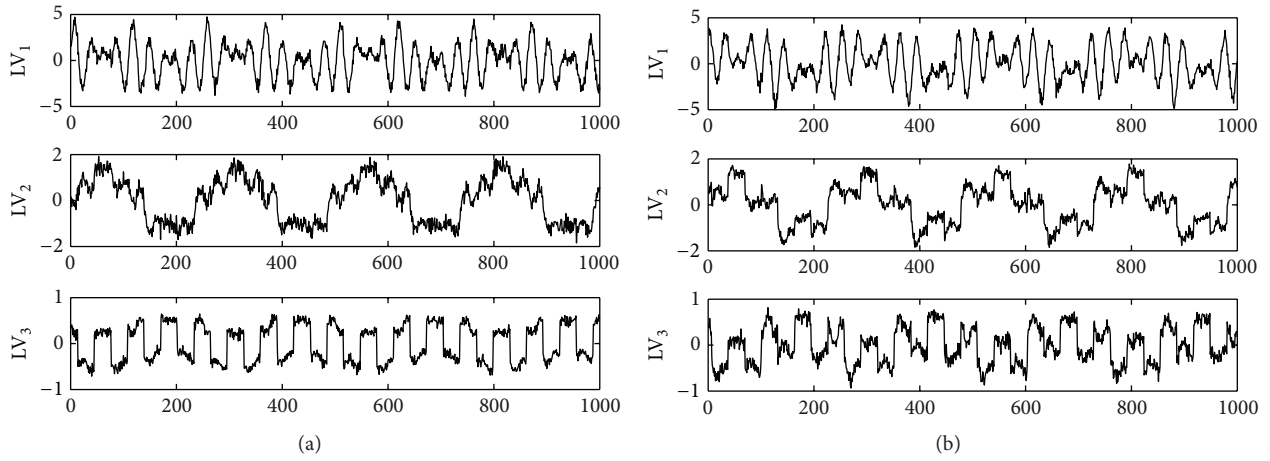


FIGURE 3: (a) The LVs estimated in X using PLS. (b) The LVs estimated in Y using PLS.

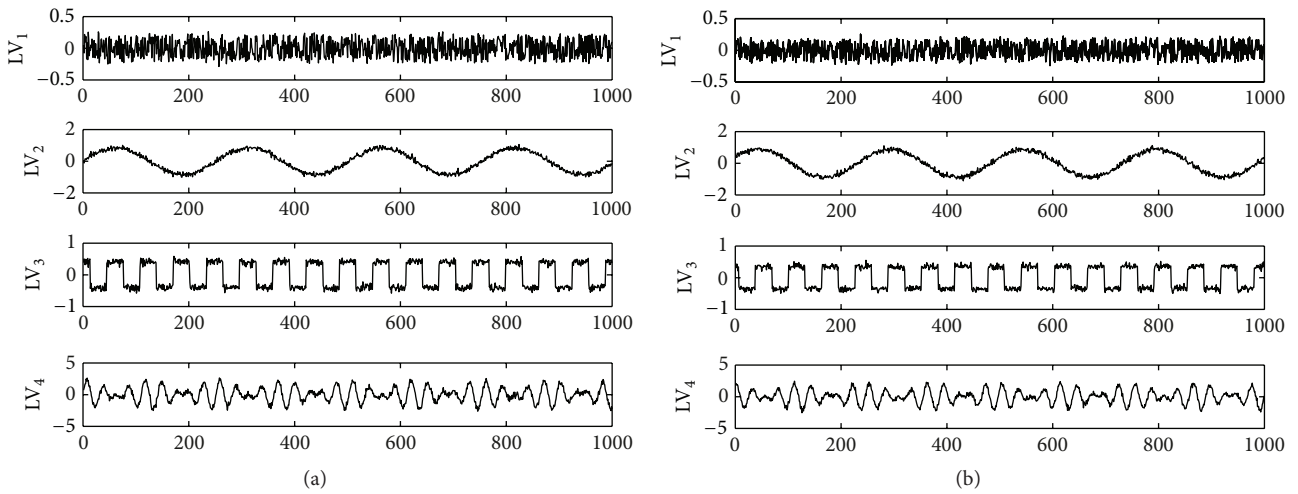


FIGURE 4: (a) The LVs estimated in X using CCA. (b) The LVs estimated in Y using CCA.

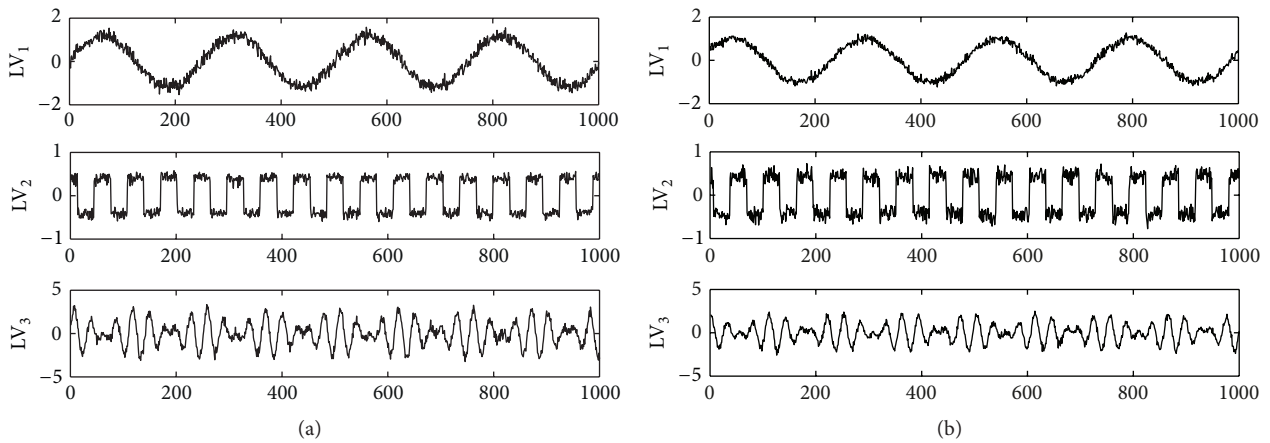


FIGURE 5: (a) The LVs estimated in X using the proposed PLS + CCA. (b) The LVs estimated in Y using the proposed PLS + CCA.

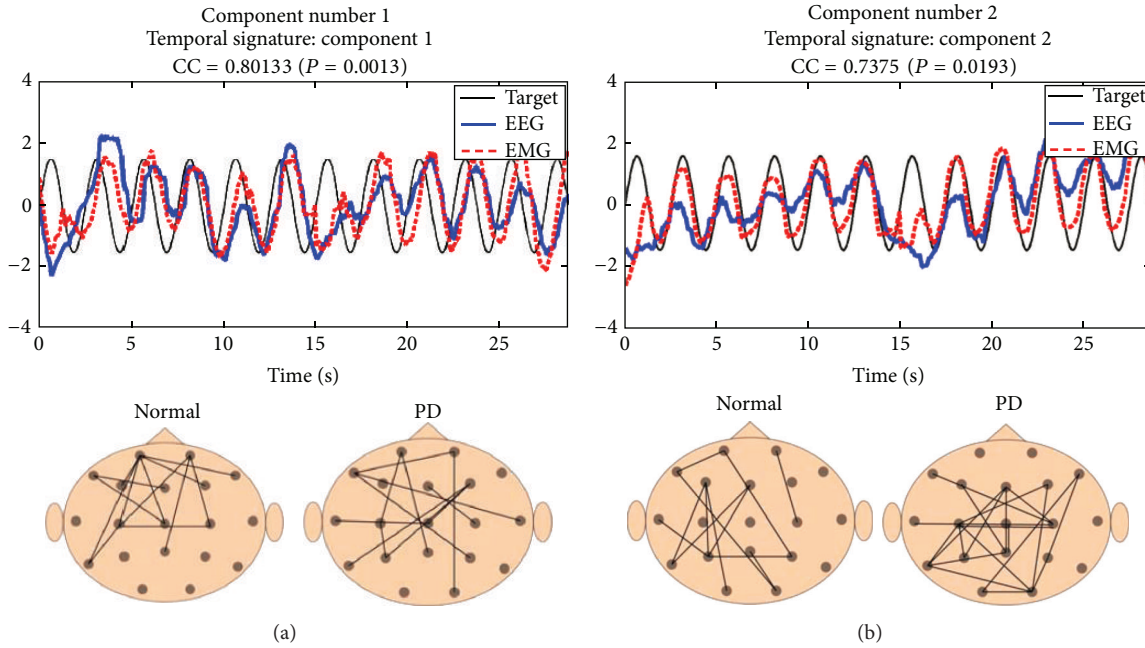


FIGURE 6: The two components from the proposed PLS + CCA method when using the EEG correlation features and the EMG amplitude features as data sets. Top panel: temporal patterns of the EEG (blue, solid) and the EMG (red, dashed). The oscillation of the target bar is also shown (black, solid). Bottom panel: EEG spatial patterns of normal subjects (left) and PD subjects (right). The connections reflect the respective spatial patterns in the two groups. CC means correlation coefficient.

3. Results and Discussion

3.1. The Synthetic Data Case. The extracted components using PLS, CCA, and the proposed PLS + CCA methods are shown in Figures 3, 4, and 5, respectively. The LVs extracted by PLS are automatically ordered in terms of their significance. To some extent, the LVs successfully reflect the corresponding relationships of the underlying sources between X and Y . However, compared with the original sources, the extracted LVs are distorted, suggesting that a higher covariance may merely result from the larger variance of LVs, which may not necessarily imply strong correlations. We can see that CCA can recover the original sources accurately in both data spaces, and the LVs are ordered strictly according to their correlation coefficients, but it completely ignores the influence of the variance and thus the extracted LVs may only reflect trivial information of the data sets (e.g., the 1st LV). For instance, although the first pair of LVs has the highest correlation coefficient, they do not contain major information of the data spaces. In practice, such LVs generally represent the noises with similar patterns simultaneously coupled into the two data modalities. When jointly modeling the data sets, they should be removed. We also note that PLS only extracts three LVs since they are sufficient to describe the data sets. These LVs do not include the first pair recovered by CCA due to their triviality. The above observations motivate us to employ the proposed PLS + CCA method.

When the proposed method is employed, the dominant sources which make significant contributions to both data spaces are first identified and ordered in terms of covariance. At the same time, trivial information is removed. Then,

within the extracted major information, sources that are highly correlated are accurately recovered with the focus on correlation. In this case, it is ensured that the extracted LVs are maximally correlated across two data sets and meanwhile can well explain the information within each individual data set.

3.2. The Real Data Case. In this case study, we applied the proposed method for corticomuscular activity modeling to the EEG and EMG features generated using the procedure described in Section 2.3.2 from 8 normal and 9 PD subjects simultaneously. The joint modeling of normal and PD data allows the identification of common temporal patterns across the groups. Meanwhile, the spatial patterns may be different across subjects, from which we could identify specific correlation connections that are differently recruited by PD subjects during the motor task.

Using the permutation test, two components were deemed significant ($P \leq 0.05$) (Figure 6). Note that in the figure only connections whose weights are statistically different from zero are shown. The results based on real data from PD and normal subjects performing a dynamic motor task are promising. In the past, most EEG/EMG coupling studies have compared EEG activity at a specific locus (e.g., sensorimotor cortex contralateral to the hand performing the task) with the EMG during sustained contractions. However, we found that in normal subjects, correlations between the contralateral sensorimotor cortex and other regions are closely associated with ongoing EMG features during dynamic motor tasks (Figure 6). It is likely that the dynamic nature of the task might require the recruitment of additional regions such as

frontal regions for motor selection [23], contralateral (i.e., ipsilateral to hand movement) sensorimotor cortex for fine modulatory control [24], and occipital regions for posterror adaptations [25].

From Figure 6, we note similar connections between the PD and control groups, especially when comparing connections associated with each component, but we also note significant differences when comparing the PD and control groups. It is noted that PD subjects have increased connectivity between the frontal regions and central and sensorimotor cortices, compared with control subjects. This may reflect the enhanced effort required by PD subjects for motor task switching [26], a problem common in this PD population [27]. In addition, PD subjects have a significant connection between the left sensorimotor and occipital regions, that is, not present in the control group. We note that the connections with occipital regions are prominent in PD subjects. Compared to normal subjects, the PD subjects heavily rely on visual cues for the initiation [28] and ongoing control of movement [29]. Moreover, the increased intra- and interhemispheric connections observed in the PD subjects are consistent with the findings in previous MEG studies [30].

4. Conclusions

In this paper, we combine the advantages of PLS and CCA and propose a PLS + CCA method to improve the performance of the joint LV extraction. We illustrate the performances of the proposed approach using both synthetic data and real-life data. For corticomuscular activity modeling, we note highly correlated temporal patterns between EEG and EMG signals and meaningful spatial activation patterns. The proposed method is a promising analysis technique for multisubject and bimodal data sets, including, but not limited to, fMRI, PPG, ECG, and kinematic data.

Appendix

A. The Derivation of the Algorithm

In this appendix, we show how to mathematically derive the solution of the proposed PLS + CCA method.

A.1. The First Step: PLS. The cost function of PLS is as follows (the same as (1) in Section 2.1.1):

$$\begin{aligned} \max_{\mathbf{w}_1, \mathbf{w}_2} \quad & (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2 \\ \text{s.t.} \quad & \mathbf{w}_i^T \mathbf{w}_i = 1, \quad i = 1, 2, \end{aligned} \quad (\text{A.1})$$

where \mathbf{w}_i 's ($i = 1, 2$) are the weight vectors.

By employing the method of Lagrange multipliers, we rewrite the initial cost function as

$$L(\mathbf{w}_i, \lambda_i) = (\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2)^2 - \sum_{i=1}^2 \lambda_i (\mathbf{w}_i^T \mathbf{w}_i - 1), \quad (\text{A.2})$$

where λ_i 's are Lagrange multipliers.

Now we only present the detailed derivations regarding \mathbf{w}_1 since \mathbf{w}_2 can be similarly derived. Taking the derivatives of $L(\mathbf{w}_i, \lambda_i)$ with respect to \mathbf{w}_1 and λ_1 and setting them to be zero, we have

$$\nabla L_{\mathbf{w}_1} = 2 \left| \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right| \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 - 2\lambda_1 \mathbf{w}_1 = 0, \quad (\text{A.3})$$

$$\nabla L_{\lambda_1} = \mathbf{w}_1^T \mathbf{w}_1 - 1 = 0. \quad (\text{A.4})$$

Left multiplying both sides of (A.3) by \mathbf{w}_1^T , we have

$$2 \left(\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2 - 2\lambda_1 \mathbf{w}_1^T \mathbf{w}_1 = 0. \quad (\text{A.5})$$

According to (A.4), λ_1 can be calculated as

$$\lambda_1 = \left(\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2. \quad (\text{A.6})$$

Through the similar procedure, $\nabla L_{\mathbf{w}_2}$ and λ_2 can be easily derived as

$$\nabla L_{\mathbf{w}_2} = 2 \left| \mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right| \mathbf{Y}^T \mathbf{X} \mathbf{w}_1 - 2\lambda_2 \mathbf{w}_2 = 0, \quad (\text{A.7})$$

$$\lambda_2 = \left(\mathbf{w}_1^T \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 \right)^2. \quad (\text{A.8})$$

Substituting (A.8) into (A.3) and (A.7), respectively, we have the following two expressions:

$$\sqrt{\lambda_2} \mathbf{X}^T \mathbf{Y} \mathbf{w}_2 = \lambda_1 \mathbf{w}_1, \quad (\text{A.9})$$

$$\frac{1}{\sqrt{\lambda_2}} \mathbf{Y}^T \mathbf{X} \mathbf{w}_1 = \mathbf{w}_2. \quad (\text{A.10})$$

By substituting (A.10) into (A.9), we can formulate an eigenvalue-eigenvector decomposition problem:

$$\left(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \right) \mathbf{w}_1 = \lambda_1 \mathbf{w}_1. \quad (\text{A.11})$$

Similarly, we can have the formulation for \mathbf{w}_2 as

$$\left(\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y} \right) \mathbf{w}_2 = \lambda_2 \mathbf{w}_2. \quad (\text{A.12})$$

The above solutions are straightforward. A practical issue is to determine the number of LVs. In our study, we determine the number R by setting a threshold that corresponds to the ratio of explained variance (e.g., 95%). Therefore, the corresponding LVs in \mathbf{X} and \mathbf{Y} can be calculated by

$$\mathbf{T}_X = \mathbf{X} \mathbf{W}_1, \quad \mathbf{T}_Y = \mathbf{Y} \mathbf{W}_2, \quad (\text{A.13})$$

where \mathbf{W}_1 is composed of the first R eigenvectors associated with (A.11) and the columns of \mathbf{T}_X represent the R components extracted from \mathbf{X} . \mathbf{W}_2 and \mathbf{T}_Y are similarly defined.

However, the collinearity problem may exist in the LVs calculated through the above procedure since each data set is used repetitively for each LV's calculation. The extracted LVs are not necessarily uncorrelated to each other. To effectively implement the second step and avoid the ill-conditioned problem, we need to address this uncorrelatedness concern and thus we design a deflation procedure: before extracting

the second common LV in each data space, \mathbf{X} and \mathbf{Y} are deflated by their corresponding first LVs as follows:

$$\begin{aligned}\mathbf{X} &= \mathbf{X} - \mathbf{t}_X(\mathbf{t}_X^T \mathbf{t}_X)^{-1} \mathbf{t}_X^T \mathbf{X}, \\ \mathbf{Y} &= \mathbf{Y} - \mathbf{t}_Y(\mathbf{t}_Y^T \mathbf{t}_Y)^{-1} \mathbf{t}_Y^T \mathbf{Y}.\end{aligned}\quad (\text{A.14})$$

Then the above procedure will be repeated for the further extraction of common LVs. In this way, the following new LVs are uncorrelated to the previous ones.

The purpose of this step is to extract LVs which can most explain the individual data sets and meanwhile are well correlated to the LVs in another data set. With this step, trivial and irrelevant information across data sets could be removed. However, a higher covariance may merely result from the larger variance of LVs, which may not necessarily imply strong correlations. To address this concern, the 2nd step will help further refine the results.

A.2. The Second Step: CCA. Based on the extracted LVs in the first step, the objective function of CCA can be constructed as follows:

$$\begin{aligned}\max_{\mathbf{v}_1, \mathbf{v}_2} & \quad (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_Y \mathbf{v}_2)^2 \\ \text{s.t.} & \quad \mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_X \mathbf{v}_1 = 1, \quad \mathbf{v}_2^T \mathbf{T}_Y^T \mathbf{T}_Y \mathbf{v}_2 = 1,\end{aligned}\quad (\text{A.15})$$

where \mathbf{v}_i 's ($i = 1, 2$) are the weight vectors.

By employing the method of Lagrange multipliers, we rewrite the initial objective function as

$$\begin{aligned}L(\mathbf{v}_i, \eta_i) &= (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_Y \mathbf{v}_2)^2 - \eta_1 (\mathbf{v}_1^T \mathbf{T}_X^T \mathbf{T}_X \mathbf{v}_1 - 1) \\ &\quad - \eta_2 (\mathbf{v}_2^T \mathbf{T}_Y^T \mathbf{T}_Y \mathbf{v}_2 - 1),\end{aligned}\quad (\text{A.16})$$

where η_i 's are Lagrange multipliers. Similar to the derivation in the first step, we can obtain the following eigenvalue-eigenvector decomposition problem:

$$\left[(\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y (\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X \right] \mathbf{v}_1 = \eta_1 \mathbf{v}_1. \quad (\text{A.17})$$

Similarly, for \mathbf{v}_2 , we have

$$\left[(\mathbf{T}_Y^T \mathbf{T}_Y)^{-1} \mathbf{T}_Y^T \mathbf{T}_X (\mathbf{T}_X^T \mathbf{T}_X)^{-1} \mathbf{T}_X^T \mathbf{T}_Y \right] \mathbf{v}_2 = \eta_2 \mathbf{v}_2. \quad (\text{A.18})$$

The solutions to this problem are the R largest eigenvectors of the corresponding matrices. The recovered LVs \mathbf{U}_X and \mathbf{U}_Y can be calculated directly from the matrices \mathbf{T}_X and \mathbf{T}_Y by

$$\mathbf{U}_X = \mathbf{T}_X \mathbf{V}_1, \quad \mathbf{U}_Y = \mathbf{T}_Y \mathbf{V}_2, \quad (\text{A.19})$$

where \mathbf{V}_1 is composed of the R eigenvectors associated with (A.17) and the columns of \mathbf{U}_X represent the R components extracted from \mathbf{T}_X . \mathbf{V}_2 and \mathbf{U}_Y are similarly defined.

After these two steps, it is ensured that the extracted components \mathbf{U}_X and \mathbf{U}_Y are maximally correlated across data sets and meanwhile can well explain the information within each individual data set.

References

- [1] S. N. Baker, E. Olivier, and R. N. Lemon, "Coherent oscillations in monkey motor cortex and hand muscle EMG show task-dependent modulation," *Journal of Physiology*, vol. 501, no. 1, pp. 225–241, 1997.
- [2] B. A. Conway, D. M. Halliday, S. F. Farmer et al., "Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man," *Journal of Physiology*, vol. 489, no. 3, pp. 917–924, 1995.
- [3] D. M. Halliday, B. A. Conway, S. F. Farmer, and J. R. Rosenberg, "Using electroencephalography to study functional coupling between cortical activity and electromyograms during voluntary contractions in humans," *Neuroscience Letters*, vol. 241, no. 1, pp. 5–8, 1998.
- [4] V. N. Murthy and E. E. Fetz, "Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5670–5674, 1992.
- [5] J. Chiang, Z. J. Wang, and M. J. McKeown, "A multiblock PLS model of cortico-cortical interactions in Parkinson's disease during corticomuscular activity," *NeuroImage*, vol. 63, no. 3, pp. 1498–1509, 2012.
- [6] X. Chen, X. Chen, R. K. Ward, and Z. J. Wang, "A joint multimodal group analysis framework for modeling corticomuscular activity," *IEEE Transactions on Multimedia*, 2013.
- [7] X. Chen, C. He, Z. J. Wang, and M. J. McKeown, "An IC-PLS framework for group corticomuscular coupling analysis," *IEEE Transactions on Biomedical Engineering*, 2013.
- [8] R. Bortel and P. Sovka, "EEG-EMG coherence enhancement," *Signal Processing*, vol. 86, no. 7, pp. 1737–1751, 2006.
- [9] B. S. Everitt and G. Dunn, *Applied Multivariate Data Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2010.
- [10] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta C*, vol. 185, pp. 1–17, 1986.
- [11] A. McIntosh, F. Bookstein, J. V. Haxby, and C. L. Grady, "Spatial pattern analysis of functional brain images using partial least squares," *NeuroImage*, vol. 3, no. 3, pp. 143–157, 1996.
- [12] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, 2004.
- [13] H. Yu and J. F. MacGregor, "Post processing methods (PLS-CCA): simple alternatives to preprocessing methods (OSC-PLS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 73, no. 2, pp. 199–205, 2004.
- [14] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2583–2587, 2006.
- [15] E. Gumus, O. Kursun, A. Sertbas, and D. Ustek, "Application of canonical correlation analysis for identifying viral integration preferences," *Bioinformatics*, vol. 28, no. 5, pp. 651–655, 2012.
- [16] N. M. Correa, T. Adali, Y. O. Li, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.
- [17] F. Lindgren, P. Geladi, and S. Wold, "The kernel algorithm for PLS," *Journal of Chemometrics*, vol. 7, no. 1, pp. 45–59, 2005.
- [18] G. Gomez-Herrero, "Automatic Artifact Removal (AAR) toolbox v1.3 (Release 09.12.2007) for MATLAB," 2007.

- [19] V. D. Calhoun, J. Liu, and T. Adali, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, supplement 1, pp. S163–172, 2009.
- [20] S. Weiss and P. Rappelsberger, "Long-range EEG synchronization during word encoding correlates with successful memory performance," *Cognitive Brain Research*, vol. 9, no. 3, pp. 299–312, 2000.
- [21] E. A. Clancy, E. L. Morin, and R. Merletti, "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of Electromyography and Kinesiology*, vol. 12, no. 1, pp. 1–16, 2002.
- [22] P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer Series in Statistics, Springer, Heidelberg, Germany, 2nd edition, 2000.
- [23] P. Praamstra, L. Boutsen, and G. W. Humphreys, "Frontoparietal control of spatial attention and motor intention in human EEG," *Journal of Neurophysiology*, vol. 94, no. 1, pp. 764–774, 2005.
- [24] J. Liepert, C. Dettmers, C. Terborg, and C. Weiller, "Inhibition of ipsilateral motor cortex during phasic generation of low force," *Clinical Neurophysiology*, vol. 112, no. 1, pp. 114–121, 2001.
- [25] C. Danielmeier, T. Eichele, B. U. Forstmann, M. Tittgemeyer, and M. Ullsperger, "Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas," *Journal of Neuroscience*, vol. 31, no. 5, pp. 1780–1789, 2011.
- [26] P. Sauseng, W. Klimesch, R. Freunberger, T. Pecherstorfer, S. Hanslmayr, and M. Doppelmayr, "Relevance of EEG alpha and theta oscillations during task switching," *Experimental Brain Research*, vol. 170, no. 3, pp. 295–301, 2006.
- [27] I. G. M. Cameron, M. Watanabe, G. Pari, and D. P. Munoz, "Executive impairment in Parkinson's disease: response automaticity and task switching," *Neuropsychologia*, vol. 48, no. 7, pp. 1948–1957, 2010.
- [28] P. Praamstra, D. F. Stegeman, A. R. Cools, and M. W. I. M. Horstink, "Reliance on external cues for movement initiation in Parkinson's disease. Evidence from movement-related potentials," *Brain*, vol. 121, no. 1, pp. 167–177, 1998.
- [29] J. K. R. Stevenson, M. M. K. Oishi, S. Farajian, E. Cretu, E. Ty, and M. J. McKeown, "Response to sensory uncertainty in Parkinson's disease: a marker of cerebellar dysfunction?" *European Journal of Neuroscience*, vol. 33, no. 2, pp. 298–305, 2011.
- [30] D. Stoffers, J. L. Bosboom, J. B. Deijen, E. C. Wolters, and C. J. Stam, "Increased cortico-cortical functional connectivity in early-stage Parkinson's disease: an MEG study," *NeuroImage*, vol. 41, no. 2, pp. 212–222, 2008.

Research Article

Stress Analysis of Osteoporotic Lumbar Vertebra Using Finite Element Model with Microscaled Beam-Shell Trabecular-Cortical Structure

Yoon Hyuk Kim,¹ Mengying Wu,¹ and Kyungsoo Kim²

¹ Department of Mechanical Engineering, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Republic of Korea

² Department of Applied Mathematics, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Republic of Korea

Correspondence should be addressed to Yoon Hyuk Kim; yoohkim@khu.ac.kr

Received 8 March 2013; Accepted 18 April 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Yoon Hyuk Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Osteoporosis is a disease in which low bone mass and microarchitectural deterioration of bone tissue lead to enhanced bone fragility and susceptibility to fracture. Due to the complex anatomy of the vertebral body, the difficulties associated with obtaining bones for in vitro experiments, and the limitations on the control of the experimental parameters, finite element models have been developed to analyze the biomechanical properties of the vertebral body. We developed finite element models of the L2 vertebra, which consisted of the endplates, the trabecular lattice, and the cortical shell, for three age-related grades (young, middle, and old) of osteoporosis. The compressive strength and stiffness results revealed that we had developed a valid model that was consistent with the results of previous experimental and computational studies. The von-Mises stress, which was assumed to predict the risk of a burst fracture, was also determined for the three age groups. The results showed that the von-Mises stress was substantially higher under relatively high levels of compressive loading, which suggests that patients with osteoporosis should be cautious of fracture risk even during daily activities.

1. Introduction

Osteoporosis is a disease in which low bone mass and microarchitectural deterioration of bone tissue lead to enhanced bone fragility and susceptibility to fracture [1]. Osteoporosis is one of the most common health problems affecting both men and women [2], and it is becoming increasingly prevalent in our aging society [3]. The degree of osteoporosis is categorized with the T-score, which is the number of standard deviations above or below that of an average young adult: normal, above -1.0 ; osteopenia, above -2.5 and below -1.0 ; osteoporosis, below -2.5 ; severe osteoporosis, the presence of one or more fragility fracture [4, 5]. About 1.5 million fractures due to osteoporosis are reported annually in the United States, including over 700,000 vertebral fractures [6]. Spine fractures in particular result in a high mortality rate: survival is 72% in the first year and only 28% after five years [7].

The human spine is composed of 24 spinal bones, called vertebrae, which are stacked on top of one another to create the spinal column. The spinal column is the body's main upright support and the vertebral bone is the primary compressive load-bearing structure in the spine [8]. The vertebral bone is composed of a porous internal trabecular bone core surrounded by a thin shell of cortex. In osteoporosis, bone mineral density is reduced even in the outer layer, so the cortex is thinner than in normal bones. The structure of osteoporotic trabecular bone is similar to a lattice, while normal bone is plate like.

Due to the complex anatomy of the vertebral body, the difficulties associated with obtaining bones for in vitro experiments, and the limitations on the control of the experimental parameters, finite element models have been developed to analyze the biomechanical properties of the vertebral body [9, 10]. Large-scale voxel-based models have been used to investigate the mechanics of bone, where the trabecular

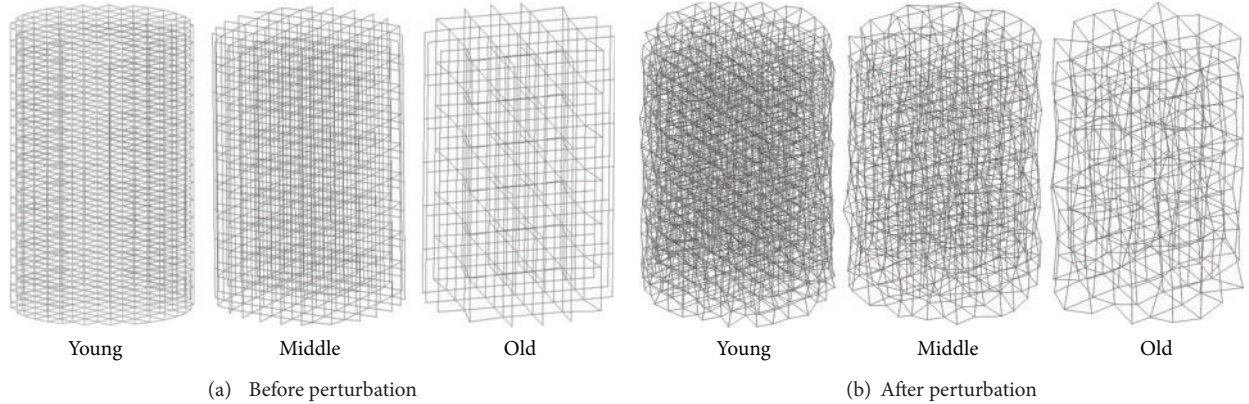


FIGURE 1: Lattice models with a trabecular structure before and after perturbation for three age groups—young (<50 years); middle (50 through 75 years); and old (>75 years).

structure is modeled as a solid body with material properties obtained from previous experiments [11–13]. While the trabecular structure from three-dimensional microcomputed tomography (μ CT) was directly implemented into a finite element using the cubic voxel meshes, an additional surface smoothing process was necessary [14]. Lattice models have been proposed to simulate osteoporotic and normal bone through variation in trabecular thickness, spacing, or random material removal [15–19]. Since these studies only addressed the trabecular structure within a small region, a more recent study combined the lattice beam element model of the trabecular core with a thin layer of shell elements for the cortical part to make a whole vertebra model, and analyzed compressive strength, compressive stiffness, and tissue-level strain [20]. In this paper, finite element models of normal and various grades of osteoporotic lumbar vertebrae that incorporate the microscaled trabecular structure of lattice models and the cortical area of shell elements were developed. The models were validated using the results of previous experimental and computational studies. The von-Mises stress was analyzed to predict the risk of the burst fracture in osteoporotic bones of various grades.

2. Materials and Methods

Trabecular bone was modeled as a lattice composed of many struts, including both vertical (longitudinal) and horizontal (transverse) struts. A single strut model was developed with two quadratic beam elements. By combining single strut models, we were able to develop a cylindrical core lattice model for different age groups: young (<50 years), middle (50 through 75 years), and old (>75 years) (Figure 1(a)). The geometries, which were the horizontal and vertical thicknesses (d_h and d_v) and the horizontal and vertical lengths (l_h and l_v) of each strut, are provided for each age group in Table 1 based on [20, 21]. The elastic-perfectly plastic material properties of the struts were based on those of a previous study [20], in which Young's modulus was 8.0 GPa, the Poisson ratio was 0.3, and the yield stress was 64 MPa. In order to mimic the irregular structure of the trabecular

TABLE 1: Geometries for trabecular lattice models for the young, middle, and old groups.

	Thickness of a single strut (mm)		Length of a single strut (mm)	
	Horizontal (d_h)	Vertical (d_v)	Horizontal (l_h)	Vertical (l_v)
Young	0.150	0.208	0.674	0.633
Middle	0.116	0.187	0.861	1.100
Old	0.107	0.201	1.145	1.668

struts, the lattice models were perturbed by randomly moving vertex nodes with MATLAB (MathWorks Inc., MA, USA) (Figure 1(b)) [20, 22]. The distance that each vertex node was moved ranged between 0% and 30% of trabecular spacing (horizontal length l_h and vertical length l_v) according to a Gaussian distribution. The Gaussian distribution $N(\mu, \sigma)$ was given based on the assumption that the mean $\mu = 0$ and the random values are between -3σ and 3σ with the probability of 99.7%. Since the movement was constrained up to 30% of trabecular spacing, the standard deviation σ was supposed as $\sigma = (1/3) \times 0.3 \times ((l_h + l_v)/2) = (l_h + l_v)/20$. The direction was also randomly generated to prevent the model from having a bias in one direction.

The trabecular bone lattice models were validated by comparing the results for compressive strength F_c , which is the capacity of a material or structure to withstand axial forces, with those in the experimental study [21]. The geometry of the specimen (cylindrical shape, height of 10 mm, and radius of 3.5 mm), the boundary conditions, and the loading conditions were selected based on the experimental study [21]. The bottom nodes of the lattice model were fixed, and the total reaction force on all fixed bottom nodes in the axial direction was calculated until the top nodes were displaced by 2 mm downward in the axial direction, which was regarded as being compressed, using the ABAQUS/Standard (Dassault Systèmes, RI, USA). The maximum value of reaction force during the compression was considered as the compressive strength F_c .

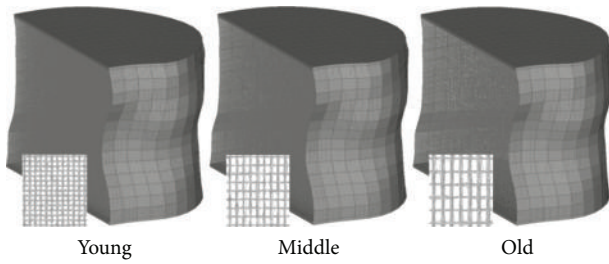


FIGURE 2: Vertebral body models in the three age groups—young (<50 years); middle (50 through 75 years); and old (>75 years).

Computed tomography (CT) images of a 1 mm slice were obtained from L2 of a male subject with a height of 175 cm. Based on these CT images, two endplates and the cortical part of the vertebra were developed (Figure 2). The elements of the S4 type and the C3D8 type composed the cortical shell and the solid endplates. The thickness of the endplates was 0.5 mm for all age groups, while the thickness of the cortical shell was 0.5 mm for the young and middle groups and 0.2 mm for the old group based on the results of a previous study [20]. The material properties of the endplates and the cortical shell were also based on the results of a previous study [20], in which Young's modulus was 8.0 GPa, the Poisson ratio was 0.3, and the yield stress was 64 MPa. The trabecular lattice was tied with the cortical shell and endplates by fixing the ends of trabecular beam elements on the interfaces of the cortical shell and endplate for all translational and rotational degrees of freedom. In this case, the posterior element of the vertebra was not considered because the whole vertebral body has too many elements, and the posterior element plays only a minor role during compressive loading [23].

The nodes on the bottom side of the bottom endplate were fixed, and the nodes of the top side of the upper endplate were controlled by a reference point. The reference point was displaced 6 mm downward in the axial direction. The compressive strength F_c which was the maximum total reaction force on the bottom endplate in the axial direction during the compression up to 6 mm was calculated using the ABAQUS/Standard (Dassault Systèmes, RI, USA). In addition, the compressive stiffness k_c was estimated as $k_c = F_c/\delta_c$, where δ_c was the corresponding axial downward displacement of the reference point at the maximum compressive strength. The compressive strength F_c and stiffness k_c were compared with those from previous experimental and computational studies [20, 24–27] for the validation of the model.

The von-Mises stress in the trabecular lattice was then analyzed for the three age groups (young, middle, and old) under compressive loading. Since the intradiscal pressure in upright standing is around 0.5 MPa and the intradiscal pressure is 1.5 times the average pressure over the endplate, the pressure on the endplate was considered to be 0.3 MPa in upright standing [28–30]. Thus, the compressive loading on the entire upper endplate was assumed to be 0.15 MPa, 0.3 MPa, 0.45 MPa, 0.6 MPa, and 0.75 MPa. The ABAQUS/Explicit (Dassault Systèmes, RI, USA) was used for the analysis.

3. Results and Discussion

In order to mimic a realistic trabecular structure, we developed a model in which perturbation of the vertex nodes occurred within 30% of the trabecular spacing according to a Gaussian distribution $N(\mu, \sigma)$, where the mean μ was 0 mm in all age groups and the standard deviation σ was 0.065 mm in the young group, 0.098 mm in the middle group, and 0.141 mm in the old group. We used a perturbation factor of 30% because Silva and Gibson showed that cross-sectional images of a model with 30% perturbation were comparable to those of specimens [21]. The mechanical anisotropy was also included in the model due to the discrepancy in horizontal and vertical geometries of trabecular lattice models. Therefore, we developed a model with an anisotropic irregular trabecular structure, and this model can be regarded as clinically relevant.

The compressive strength F_c of the trabecular bone lattice model was 1.74 MPa for the middle group, while it was 1.35 ± 0.64 MPa in [22]. For the whole-body model, the compressive strength F_c was 7.35 kN for the young group, 3.80 kN for the middle group, and 1.36 kN for the old group. In a previous computational study with the same age classes, the compressive strength was 5.74 kN for the young group, 4.06 kN for the middle group, and 1.25 kN for the old group [20]. In experimental studies using normal vertebrae, the compressive strengths have ranged from 0.9 to 15.9 kN (0.9 to 5.0 kN in [24], 1.5 to 4.5 kN in [25], 2.0 to 8.0 kN in [26], and 2.0 to 15.9 kN in [27]). In addition, the compressive stiffness k_c was 5.6 kN/mm for the young group, 15.8 kN/mm for the middle group, and 29.4 kN/mm for the old group, respectively. In a previous computational study, the compressive stiffness was 8.0 kN/mm for the young group, 18.7 kN/mm for the middle group, and 29.4 kN/mm for the old group [20]. These results indicate that the presented model could be considered as being validated for the compression.

The highest von-Mises stresses occurred in the middle of the trabecular region (Figure 3). The maximum stress was strongly related to age: maximum values were about 50% higher for the middle group than the young group and about 120% higher for the old group than the young age group (Figures 4 and 5). In addition, the maximum stress was greater than 50% of the yield stress (64 MPa) when compressive loading exceeded 0.45 MPa for middle group and 0.3 MPa for the old group; in contrast, the maximum stress did not reach 50% of the yield stress even under 0.75 MPa for the young group. The 0.45 MPa of compressive loading on the endplate equates to about 0.7 MPa of intradiscal pressure, which is similar to that produced during daily activities, such as standing while bent forward (1.10 MPa), standing up from a chair (1.10 MPa), and lifting or holding a weight of 20 kg (1.10–2.30 MPa) [30]. These results suggest that osteoporosis can affect the stress acting on the vertebra even during routine daily activities.

The vertebra model that incorporates a realistic trabecular structure is advantageous because it permits the simulation of in vivo specimens for the study of osteoporosis. The microscale trabecular structure represented by tiny struts would provide the mechanism that the strut deformation or

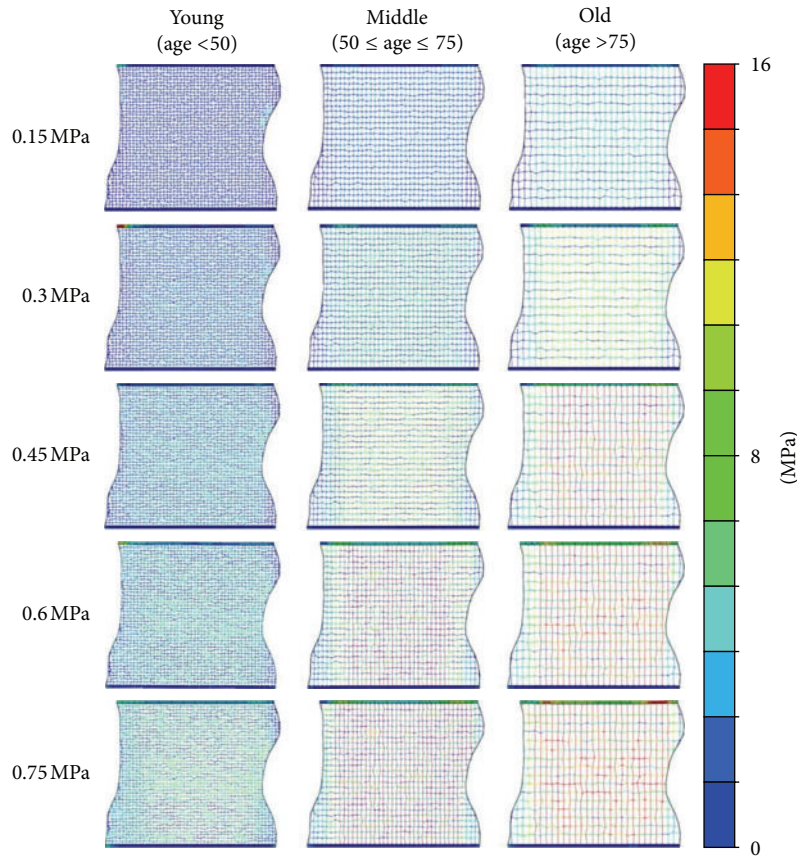


FIGURE 3: Cross-sectional view in the middle sagittal plane of the von-Mises stress distribution for the three age groups under 0.15 MPa, 0.3 MPa, 0.45 MPa, 0.6 MPa, and 0.75 MPa of compressive loading.

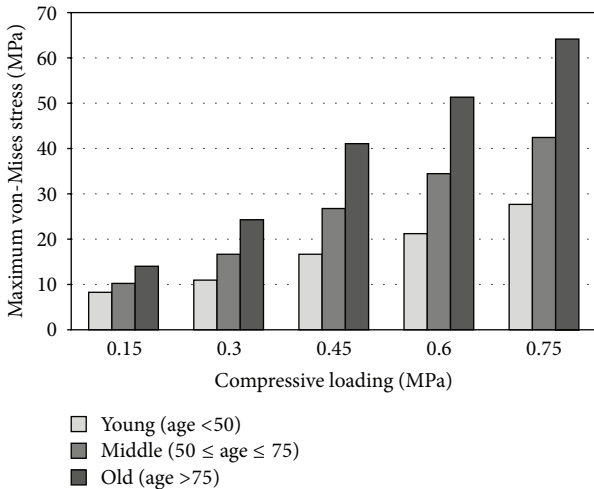


FIGURE 4: Maximum stress for the three age groups under 0.15 MPa, 0.3 MPa, 0.45 MPa, 0.6 MPa, and 0.75 MPa of compressive loading.

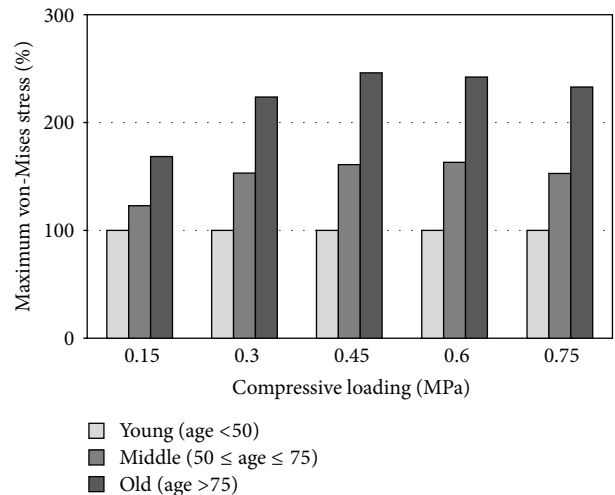


FIGURE 5: Ratios of the maximum stress for the middle and old groups relative to the young group under 0.15 MPa, 0.3 MPa, 0.45 MPa, 0.6 MPa, and 0.75 MPa of compressive loading.

buckling leads to the fracture in a whole vertebra. In addition, various grades of osteoporosis can be incorporated into the model by changing the spacing between the struts.

This study has some limitations. The validity of models was indirectly confirmed by comparing the compressive

strength and compressive stiffness of the developed models to those in the previous experimental and computational studies. The full validation of the model through experiments with the same specimen from which the model was generated can enhanced the confidence of this study. In addition, the

generic lattice model was randomly perturbed, and the thicknesses of the endplates and the cortical shell were assumed to be uniform for all age groups. Patient-specific information on the microstructure and geometry of the bone using data from CT scans would improve the accuracy and relevance of the stress analysis. Finally, the quantitative relationship between the von-Mises stress and the fracture risk was not investigated. Mechanical tests that measure the stress and failure strength of bone, as well as clinical observations that accurately identify the region of the fracture site, would improve the prediction of vertebral fracture in osteoporotic patients.

4. Conclusions

Osteoporosis is a major contributor to the increased risk of fracture with age due to low bone mass and structural change. We developed finite element models of the L2 vertebra, which consisted of the endplates, the trabecular lattice, and the cortical shell, for three age-related grades of osteoporosis. The compressive strength and stiffness results revealed that we had developed a valid model that was consistent with the results of previous experimental and computational studies. The von-Mises stress, which was assumed to predict the risk of a burst fracture, was also determined for the three age groups. The results showed that the von-Mises stress was substantially higher under relatively high levels of compressive loading, which suggests that patients with osteoporosis should be cautious of fracture risk even during daily activities.

Conflict of Interests

All authors, Yoon Hyuk Kim, Mengying Wu, and Kyung-soo Kim, disclose that there are no financial nor personal relationships with other people or organizations that could inappropriately influence (bias) this work.

Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0005167) and 2013 National Agenda Project (NAP) funded by Korea Research Council of Fundamental Science & Technology (NAP-09-2/P-13-JC-LU01-C01).

References

- [1] W. A. Peck, P. Burckhardt, C. Christiansen et al., "Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis," *American Journal of Medicine*, vol. 94, no. 6, pp. 646–650, 1993.
- [2] B. L. Riggs and L. J. Melton III, "The worldwide problem of osteoporosis: insights afforded by epidemiology," *Bone*, vol. 17, no. 5, pp. 505S–511S, 1995.
- [3] J. T. Lin and J. M. Lane, "Osteoporosis: a review," *Clinical Orthopaedics and Related Research*, no. 425, pp. 126–134, 2004.
- [4] J. A. Kanis, "Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report," *Osteoporosis International*, vol. 4, no. 6, pp. 368–381, 1994.
- [5] J. A. Kanis, E. V. McCloskey, H. Johansson, A. Oden, L. J. Melton III, and N. Khaltav, "A reference standard for the description of osteoporosis," *Bone*, vol. 42, no. 3, pp. 467–475, 2008.
- [6] S. R. Cummings, D. Bates, and D. M. Black, "Clinical use of bone densitometry: scientific review," *Journal of the American Medical Association*, vol. 288, no. 15, pp. 1889–1897, 2002.
- [7] J. A. Kanis, L. J. Melton III, C. Christiansen, C. C. Johnston, and N. Khaltav, "The diagnosis of osteoporosis," *Journal of Bone and Mineral Research*, vol. 9, pp. 1137–1141, 1994.
- [8] W. C. Hutton, B. M. Cyron, and J. R. R. Stott, "The compressive strength of lumbar vertebrae," *Journal of Anatomy*, vol. 129, no. 4, pp. 753–758, 1979.
- [9] U. Wolfram, L. O. Schwen, U. Simon, M. Rumpf, and H. J. Wilke, "Statistical osteoporosis models using composite finite elements: a parameter study," *Journal of Biomechanics*, vol. 42, no. 13, pp. 2205–2209, 2009.
- [10] V. Kosmopoulos and T. S. Keller, "Finite element modeling of trabecular bone damage," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 6, no. 3, pp. 209–216, 2003.
- [11] R. Müller, T. Hildebrand, and P. Rüeeggesser, "Non-invasive bone biopsy: a new method to analysis and display the three-dimensional structure of trabecular bone," *Physics in Medicine and Biology*, vol. 39, pp. 145–164, 1994.
- [12] W. M. Park, Y. S. Park, K. Kim, and Y. H. Kim, "Biomechanical comparison of instrumentation techniques in treatment of thoracolumbar burst fractures: a finite element analysis," *Journal of Orthopaedic Science*, vol. 14, no. 4, pp. 443–449, 2009.
- [13] P. K. Zysset, M. S. Ominsky, and S. A. Goldstein, "A novel 3D microstructural model for trabecular bone: the relationship between fabric and elasticity," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 1, pp. 321–331, 1998.
- [14] P. Arbenz and C. Flaig, "On smoothing surfaces in voxel based finite element analysis of trabecular bone," *Lecture Notes in Computer Science*, vol. 4818, pp. 69–77, 2008.
- [15] O. C. Yeh and T. M. Keaveny, "Biomechanical effects of intraspecimen variations in trabecular architecture: a three-dimensional finite element study," *Bone*, vol. 25, no. 2, pp. 223–228, 1999.
- [16] X. E. Guo and C. H. Kim, "Mechanical consequence of trabecular bone loss and its treatment: a three-dimensional model simulation," *Bone*, vol. 30, no. 2, pp. 404–411, 2002.
- [17] D. Dagan, M. Beery, and A. Gefen, "Single-trabecula building block for large-scale finite element models of cancellous bone," *Medical and Biological Engineering and Computing*, vol. 42, no. 4, pp. 549–556, 2004.
- [18] I. Diamant, R. Shahar, and A. Gefen, "How to select the elastic modulus for cancellous bone in patient-specific continuum models of the spine," *Medical and Biological Engineering and Computing*, vol. 43, no. 4, pp. 465–472, 2005.
- [19] I. Diamant, R. Shahar, Y. Masharawi, and A. Gefen, "A method for patient-specific evaluation of vertebral cancellous bone strength: in vitro validation," *Clinical Biomechanics*, vol. 22, no. 3, pp. 282–291, 2007.
- [20] K. McDonald, J. Little, M. Pearcy, and C. Adam, "Development of a multi-scale finite element model of the osteoporotic lumbar vertebral body for the investigation of apparent level vertebra mechanics and micro-level trabecular mechanics," *Medical Engineering and Physics*, vol. 32, no. 6, pp. 653–661, 2010.

- [21] M. J. Silva and L. J. Gibson, "Modeling the mechanical behavior of vertebral trabecular bone: effects of age-related changes in microstructure," *Bone*, vol. 21, no. 2, pp. 191–199, 1997.
- [22] L. Mosekilde, "Sex differences in age-related loss of vertebral trabecular bone mass and structure—biomechanical consequences," *Bone*, vol. 10, no. 6, pp. 425–432, 1989.
- [23] M. Hongo, E. Abe, Y. Shimada, H. Murai, N. Ishikawa, and K. Sato, "Surface strain distribution on thoracic and lumbar vertebrae under axial compression: the role in burst fractures," *Spine*, vol. 24, no. 12, pp. 1197–1202, 1999.
- [24] E. M. Lochmüller, F. Eckstein, D. Kaiser et al., "Prediction of vertebral failure loads from spinal and femoral dual-energy x-ray absorptiometry, and calcaneal ultrasound: an in situ analysis with intact soft tissues," *Bone*, vol. 23, no. 5, pp. 417–424, 1998.
- [25] K. Imai, I. Ohnishi, M. Bessho, and K. Nakamura, "Nonlinear finite element model predicts vertebral bone strength and fracture site," *Spine*, vol. 31, no. 16, pp. 1789–1794, 2006.
- [26] R. P. Crawford and T. M. Keaveny, "Relationship between axial and bending behaviors of the human thoracolumbar vertebra," *Spine*, vol. 29, no. 20, pp. 2248–2255, 2004.
- [27] D. L. Kopperdahl, J. L. Pearlman, and T. M. Keaveny, "Biomechanical consequences of an isolated overload on the human vertebral body," *Journal of Orthopaedic Research*, vol. 18, no. 5, pp. 685–690, 2000.
- [28] F. Magerl, M. Aebi, S. D. Gertzbein, J. Harms, and S. Nazarian, "A comprehensive classification of thoracic and lumbar injuries," *European Spine Journal*, vol. 3, no. 4, pp. 184–201, 1994.
- [29] A. Nachemson and J. M. Morris, "In vivo measurements of intradiscal pressure. discometry, a method for the determination of pressure in the lower lumbar discs," *The Journal of Bone and Joint Surgery*, vol. 46, pp. 1077–1092, 1964.
- [30] H. J. Wilke, P. Neef, M. Caimi, T. Hoogland, and L. E. Claes, "New in vivo measurements of pressures in the intervertebral disc in daily life," *Spine*, vol. 24, no. 8, pp. 755–762, 1999.

Research Article

Quantification of Stretching in the Ventricular Wall and Corpus Callosum and Corticospinal Tracts in Hydrocephalus before and after Ventriculoperitoneal Shunt Operation

Hans von Holst^{1,2} and Xiaogai Li²

¹ Department of Neurosurgery, Karolinska University Hospital, 171 76 Stockholm, Sweden

² Division of Neuronic Engineering, School of Technology and Health, Royal Institute of Technology (KTH), KTH-Flemingsberg Alfred Nobels Allé 10, Huddinge, 141 52 Stockholm, Sweden

Correspondence should be addressed to Hans von Holst; hans.vonholst@karolinska.se

Received 6 March 2013; Accepted 10 April 2013

Academic Editor: Hang Joon Jo

Copyright © 2013 H. von Holst and X. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we establish a quantitative model to define the stretching of brain tissue, especially in ventricular walls, corpus callosum (CC) and corticospinal (CS) fiber tracts, and to investigate the correlation between stretching and regional cerebral blood flow (rCBF) before and after ventriculoperitoneal shunt operations. A nonlinear image registration method was used to calculate the degree of displacement and stretching of axonal fiber tracts based on the medical images of six hydrocephalus patients. Also, the rCBF data from the literature was analyzed and correlated with the strain level quantified in the present study. The results showed substantial increased displacement and strain levels in the ventricular walls as well as in the CC and CS fiber tracts on admission. Following shunt operations the displacement as well as the strain levels reduced substantially. A linear correlation was found to exist between strain level and the rCBF. The reduction in postoperative strain levels correlated with the improvement of rCBF. All patients improved clinically except for one patient due to existing dementia. These new quantitative data provide us with new insight into the mechanical cascade of events due to tissue stretching, thereby provide us with more knowledge into understanding of the role of brain tissue and axonal stretching in some of the hydrocephalus clinical symptoms.

1. Introduction

Hydrocephalus is the consequence of various causes such as congenital malformations, subarachnoid hemorrhage, traumatic brain injury, and benign and malignant brain tumors. Regardless of etiology the common denominator among the patients is obstruction of cerebrospinal fluid interfering with the brain tissue function. The symptoms and signs are somewhat different depending on whether the patient is an infant, is child, or receives the hydrocephalus in mature age.

The expansion of the ventricular system influences the surrounding areas in the brain tissue in various intensities. The clinical condition is usually defined as the Hakim triad of symptoms including gait apraxia, cognitive disturbance, and incontinence. Initially the white matter suffers more than the grey matter with reversible changes following shunt operation which reduces the expanding ventricles.

Depending on the age at which hydrocephalus develops together with the magnitude of ventricular expansion, the neuropathological pattern is different in severity. When long-standing the mechanical distortion results in more advanced neuropathological signs such as damage to the axons and myelin. Also, when the cerebral arteries and capillaries are substantially distorted, there is a potential of reduced regional cerebral blood flow which may further interfere with the clinical condition [1, 2].

Stretching/compression of the ventricular wall [3], the corpus callosum (CC) [4, 5] and corticospinal (CS) [6, 7] fiber tracts in hydrocephalus patients has been widely recognized in many previous research investigations. Also an early hypothesis suggested that the compression and/or deformation of the CS tracts due to enlargement of the ventricles in hydrocephalus patient may contribute to the gait disturbance [8]. However, there is a lack of study providing

quantitative data of the axonal stretching level, and how the mechanical stretching may influence the axonal functions remains unclear until now. A quantitative data should also provide us more insight to understand the involvement of axonal fiber stretching in clinical symptoms.

Thus, the aim of the present study was

- (i) to analyze the displacement and strain levels in patients with hydrocephalus by using a numerical simulation method focusing on the ventricular walls, CC and CS fiber tracts before and after shunt operations and
- (ii) to estimate the regional cerebral blood flow after treatment and the correlation to the strain level after the neurosurgical procedure.

2. Material and Methods

2.1. Medical Images. Medical images from computed tomography (CT) and magnetic resonance (MR) imaging of the six patients with hydrocephalus were investigated retrospectively before (Figure 2) and after (Figure 3) a ventriculoperitoneal shunt operation. In this study, only the geometry of the brain is of concern to the strain level quantification. Therefore, both CT and Resonance MR Images are acceptable. This study was approved by the local research ethical committee of Karolinska University Hospital, Stockholm County.

2.2. Image Registration to Obtain Displacement Field. Image registration aims at finding a displacement field between a *fixed image* and *moving image* to align both images as accurate as possible [9]. A nonlinear registration method, the Diffeomorphic Demons (DD) algorithm [9] implemented in the open-source software Slicer 3D, was used in this study which has been used to quantify the displacement field occurred during decompressive craniotomy in previous studies [10, 11]. The core components of the DD registration algorithm including similarity measures and the regularization model which will be presented.

Given a fixed image $F(\cdot)$ and a moving image $M(\cdot)$, intensity-based image registration is posed as an optimization problem that aims at finding a spatial mapping that will align the moving image to the fixed image. The transformation $s(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}^D, p \mapsto s(p)$ models the spatial mapping of a particular point p from the fixed image space to the moving image space [12]. The similarity criterion $E_{\text{sim}}(F, M \circ s)$ measures the quality or the goodness of the matching of a given transformation. In this study, binary images are used, and the sum of the square difference (SSD) suffices for our application as defined as

$$\begin{aligned} E_{\text{sim}}(F, M \circ s) &= \frac{1}{2} \|F - M \circ s\|^2 \\ &= \frac{1}{2\Omega_p} \sum_{p \in \Omega_p} |F(p) - M(s(p))|^2, \end{aligned} \quad (1)$$

where $M \circ s$ represents the morphed moving image, Ω_p is the region of overlap between F and $M \circ s$. The problem now

becomes an optimization problem to find a transformation s over a given space that minimizes the similarity energy function E_{sim} .

In order to end up with a global minimization of a well-posed criterion, the similarity energy function in (1) was modified by introducing an auxiliary variable (i.e., correspondence c) in the registration process [13]. The introduction of this auxiliary variable c decouples the complex minimization into simple and efficient steps by alternating optimization over c and s . Considering a Gaussian noise on the displacement field, the global energy then becomes [12]

$$\begin{aligned} E(c, s) &= \frac{1}{\sigma_i^2} E_{\text{sim}}(F, M \circ s) + \frac{1}{\sigma_x^2} \text{dist}(s, c)^2 \\ &\quad + \frac{1}{\sigma_T^2} E_{\text{reg}}(s), \end{aligned} \quad (2)$$

where σ_i accounts for the noise on the image intensity, σ_x accounts for a spatial uncertainty of the correspondences, and σ_T controls the amount of regularization that is needed: $\text{dist}(s, c) = \|c - s\|$ and $E_{\text{reg}}(s) = \|\nabla s\|$.

The model assumes that so-called “demons” at every voxel from the fixed image are applying forces that push the voxels of the moving image to match up with the fixed image. The transformation s is driven by a “demons force” derived from the assumption of image intensity conservation [14]. The original demons algorithm has been modified and allows retrieving small and large dense displacement field defined as [15]

$$\mathbf{U}(p) = \frac{F(p) - M(s(p))}{\|\nabla F(p)\|^2 + \alpha^2(F(p) - M(s(p)))^2} \nabla F(p), \quad (3)$$

where $\mathbf{U}(p)$ is the displacement at point p defined in the fixed image, α is a positive homogenization factor. The updated unconstrained dense displacement field is computed based on an optical flow computation at each voxel at every iteration. The resulting updated field is added to the global deformation field, and the displacement field is regularized by applying a Gaussian smoothing filter.

In order to solve the displacement field \mathbf{U} , the global energy defined in (2) needs to be minimized. This energy function allows the whole optimization procedure to be decoupled into two simple steps. The first step solves for the correspondence c by optimizing $(1/\sigma_i^2)E_{\text{sim}} + (1/\sigma_x^2)\|s - c\|^2$ with respect to c and with s given. The second step solves for the regularization by optimizing $(1/\sigma_x^2)\|s - c\|^2 + (1/\sigma_T^2)E_{\text{reg}}(s)$ with respect to s and c given [9].

The transformation s in the original Demons method [14] is not constrained and does not provide diffeomorphic transformations. A diffeomorphic extension to the demons framework was proposed by adapting the optimization procedure to a space of diffeomorphic transformations. It is performed by using an intrinsic update step which computes the vector field exponentials of the Lie group of diffeomorphisms (see [9, 12] for details).

2.3. Strain Level Quantification. From the diffeomorphic Demons registration, the transformation s , corresponding

to a displacement field $\mathbf{U}(\mathbf{X})$, is obtained from (3). The displacement field is defined on every voxel in the fixed image and morphs the fixed image to the moving image. The strain tensor could then be calculated based on the theory of continuum mechanics [16]. A Lagrangian reference frame was assumed at the fixed image space:

$$s : \mathbf{x} = \mathbf{X} + \mathbf{U}(\mathbf{X}). \quad (4)$$

The displacement field $\mathbf{U}(\mathbf{X}) = [U_X, U_Y, U_Z]$ for each voxel, obtained from the image registration, gives a reasonable alignment to bring the fixed image at point p with coordinates $\mathbf{X}(X, Y, Z)$ to its corresponding point in the moving image at coordinates $\mathbf{x}(x, y, z)$ [9]. The displacement gradient is defined as [16]

$$\text{grad}(\mathbf{U}) = \frac{d\mathbf{U}}{d\mathbf{X}} = \begin{pmatrix} \frac{dU_X}{dX} & \frac{dU_X}{dY} & \frac{dU_X}{dZ} \\ \frac{dU_Y}{dX} & \frac{dU_Y}{dY} & \frac{dU_Y}{dZ} \\ \frac{dU_Z}{dX} & \frac{dU_Z}{dY} & \frac{dU_Z}{dZ} \end{pmatrix}. \quad (5)$$

The deformation gradient is defined as

$$\mathbf{F} = \frac{d\mathbf{x}}{d\mathbf{X}} = \frac{d(\mathbf{X} + \mathbf{U}(\mathbf{X}))}{d\mathbf{X}} = \mathbf{I} + \frac{d\mathbf{U}}{d\mathbf{X}} = \mathbf{I} + \text{grad}(\mathbf{U}). \quad (6)$$

The Lagrangian finite strain tensor \mathbf{E} is defined as

$$\mathbf{E} = \frac{1}{2} (\mathbf{F}^T \mathbf{F} - \mathbf{I}). \quad (7)$$

The Lagrangian finite strain tensor, \mathbf{E} , is a 3×3 symmetric tensor representing the deformation of a point. Different scalar indices that can be derived from the tensors, especially the principal strains which represent the maximum and minimum normal strains experienced at a point, are characterized by the tensor eigenvalues.

Strain level represents the deformation of brain tissue due to the occurrence of a disease. In order to quantify the brain tissue deformation, a dense displacement field representing the brain tissue motion from healthy stage to hydrocephalus stage is needed. For this, the main premise is the images acquired at two distinct states: at healthy state and at diseased state. Since the brain images of the patient under healthy state were not available, we proposed a strategy to recover a healthy brain for a specific patient by combing the information from the diseased brain geometry together with the brain atlas by a nonlinear image registration method [10].

The lateral ventricles and brain tissue were firstly manually segmented from the recovered healthy brain and hydrocephalic brain images. A rigid registration step was first used to centre the images about the same point which were the input volumes to nonlinear registration process. In principle, the segmented healthy brain should be chosen as the *fixed image* so as to capture the motion of brain tissue from the healthy to hydrocephalic state. In this study, since the registration performs better from healthy brain to the hydrocephalic brain image, we chose the hydrocephalic brain image

as the *fixed image*, and the obtained displacement field was then inverted by an `itkIterativeInverseDisplacementFieldImageFilter` algorithm implemented in ITK (see the description for the approach on ITK's website <http://www.itk.org>). The final obtained displacement for further processing is the inverted displacement field defined on the recovered healthy brain space which represents the structural brain change that occurred from the healthy to hydrocephalus state.

From the obtained displacement field, a quantitative description of the nerve tissue deformation that occurred during hydrocephalus development can be derived in the form of the finite Lagrange strain tensor [16, 17]. Different scalar indices can be derived from the strain tensor, especially, the 1st principal strain, defined as the maximum eigenvalue of Lagrange strain tensor \mathbf{E} , representing the maximum stretching was used to describe the stretching of brain tissue.

2.4. Axonal Fiber Tract Extraction and Healthy Brain Shape from Atlas. From a series of diffusion weighted (DW) images, the effective diffusion tensor, \mathbf{D} , can be estimated using the relationship between the measured signal intensity at each voxel and the applied magnetic field gradient sequence [18, 19]. The diffusion tensor is symmetric (i.e., $D_{ij} = D_{ji}$), and thus it contains only six unique values. Given this, at least six noncollinear diffusion gradient directions are required to determine the diffusion tensor [20, 21].

Once the diffusion tensor \mathbf{D} is determined, different indices can be derived to provide information for each voxel in the image. The diffusion tensor is a real, symmetric second-order tensor. Mathematically, this entails a linear rotation of the diffusion tensor to diagonalize it thereby setting off-diagonal elements to zero [22]:

$$\begin{aligned} \mathbf{D} &= \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \\ &= [\mathbf{e}_1 \mathbf{e}_1 \mathbf{e}_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{e}_1 \mathbf{e}_1 \mathbf{e}_3]^T, \end{aligned} \quad (8)$$

where λ_1 , λ_2 , and λ_3 are eigenvalues ($\lambda_1 > \lambda_2 > \lambda_3$), and the corresponding eigenvectors are \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 . The eigenvalues of the diffusion tensor provide diffusion coefficients along the orientations defined by its respective eigenvectors [22]. λ_1 represents the greatest diffusion value along a fiber axis, denoted by the direction vector \mathbf{e}_1 . λ_2 , λ_3 represent the diffusion value along two axes perpendicular to \mathbf{e}_1 . Because \mathbf{D} is symmetric and positive definite, its three eigenvectors (principal coordinate directions) \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 are orthogonal [23].

The Fractional Anisotropy (FA) is the most commonly used anisotropy measure and is a normalized expression of the tensor eigenvalues [24] according to

$$\text{FA} = \sqrt{\frac{3 \left((\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2 \right)}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}, \quad (9)$$

where $\bar{\lambda}$ is the mean of the eigenvalues of the diffusion tensor.

TABLE 1: Summary of six hydrocephalic patients including age, clinical symptoms before treatment with a ventriculoperitoneal shunt postoperative image performance and outcome evaluated in days after the operation.

Patient/age	Clinical symptoms			
	Gait disturbance	Cognitive deficit	Incontinence	Improved outcome/days after treatment
KP/45	Yes	Yes	No	Yes/63
KA/47	Yes	Yes	No	Yes/96
KB/72	Yes	Yes	Yes	Yes/30
KF/70	No	Yes	Yes	Yes/27
PN/57	No	Yes	No	Yes/19
SB/71	Yes	Yes	No	No/21

Once the fractional anisotropy (FA) and the preferred diffusion direction have been created, it is possible to perform fiber tracking or diffusion tensor tractography. Streamline tractography is one of the methods for fiber tract estimation [25, 26]. Streamline tractography uses the maximum orientation described by the eigenvector \mathbf{e}_1 associated with the largest eigenvalue λ_1 , as an estimate of local tract orientation. It assumes that λ_1 associated with the maximum eigenvector is the main fiber direction.

In this study, white matter fiber tracts were extracted from ICBM DTI-81 atlas provided by the international consortium for brain mapping (ICBM) [27]. The T2-weighted image from the same dataset which represents an average adult brain was used for recovery of a healthy brain as described in a previous section. The atlas is based on probabilistic tensor maps obtained from 81 normal adults ranging from 18 to 59 years of age. From the diffusion tensor imaging, the Fractional Anisotropy (FA) could be calculated which is a scalar value representing the anisotropic property. The color-coded FA value calculated from the DTI tensor is displayed in (Figure 1, left) where the color reflects the white matter orientation (red (right-left), green (anterior-posterior), and blue (superior-inferior)). The fiber tracts which pass through the predefined region of interests (ROI) are CC and CS fiber tracts. The corresponding sagittal and coronal sections of the T2-weighted images from the same dataset are shown in Figure 1 upper row on the right.

Streamline method [26, 28] was used to extract white matter fiber tracts. The final extracted white matter tracts in the whole brain contain the polylines with a corresponding diffusion tensor at each fiber point (Figure 1, lower row: left and middle), from which the CC and CS fiber tracts were isolated (Figure 1, lower row: right). The obtained fiber tracts from the atlas were then adapted to the patient brain by applying the obtained displacement field which represent the cranial structural shape difference between the patient and atlas. This gave the fiber tracts for the recovered healthy brain for a specific patient.

3. Results

3.1. Patient Information. The average age of the six patients with hydrocephalus was calculated to 60 years. Clinical symptoms among the six patients showed gait disturbances in four cases and cognitive deficits in all patients while only two

had urinary incontinence (Table 1). Five patients improved in their clinical symptoms, individually evaluated between 19 and 96 days after the shunt operation, while the sixth patient did not probably due to the presence of dementia.

3.2. Strain Level at Ventricular Wall, CC and CS Fiber Tracts.

The increased strain levels in the six patients are presented for preoperative (Figure 2) and postoperative stages (Figure 3). Before and after the operation all patients had a substantial dilatation of the lateral ventricles in various degrees although in less degree after the treatment. The ventricular dilatation caused increased strain levels not only in the ventricular walls ranging from 99% to 333%, but also in the brain tissue surrounding the ventricles in different degrees. As a consequence, the increased strain levels found in the CC fiber tracts ranged from 58% to 139% and in the CS fiber tracts from 29% to 74% among the patients (Figure 7).

After the ventriculoperitoneal shunt had been installed, the lateral ventricles decreased although not to an extent as could be expected after such an operation. Thus, the postoperative strain levels of the ventricular walls decreased in all six patients ranging from 9% to 60% based on the preoperative values (Figure 7).

In a similar way, although decreased after the shunt operation, the strain levels in the CC and CS fiber tracts were still on a surprisingly high level ranging from 9% to 56% for CC fiber tracts and from 5% to 20% in the CS fiber tracts (Figure 7). The strain level data showed normal distribution, and therefore, a paired sample t -test was used to compare the CC and CS fiber strain levels. The results were evaluated by a 1-tailed t -test at 95% confidence level. It was found that the strain levels in CC fiber tracts are significantly higher than those of CS fiber tracts for both pre- ($P < 0.05$) and postcraniotomy period ($P < 0.05$). This indicates a more stretched scenario for the CC fiber tracts.

The trend was that those patients with the highest ventricular dilatation and increased strain levels also showed the highest strain levels in the CC and CS fiber tracts both before and after treatment. A similar trend was found between CC and CS fiber tracts. Also, since the CC fiber tracts are closer to the expanded ventricles, it seems logical that the increased strain levels are higher in these fibers compared to those found in the CS fiber tracts which are anatomically located further away from the ventricles.

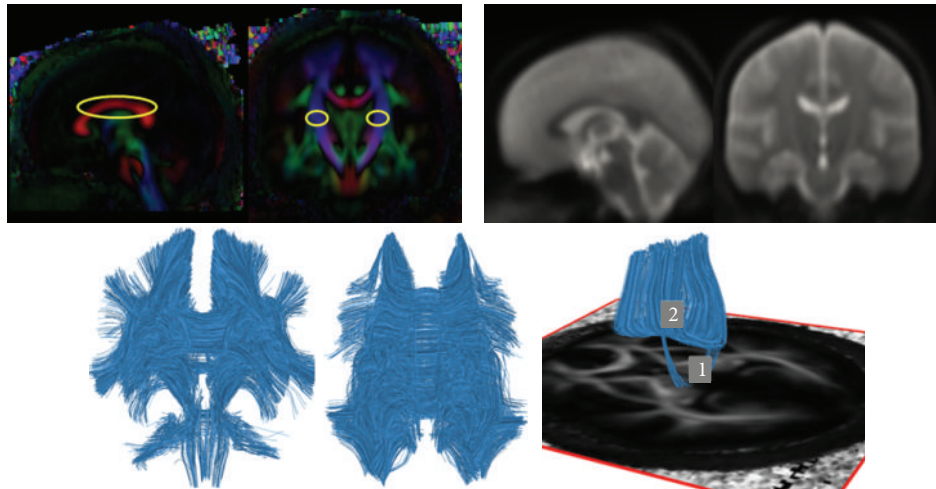


FIGURE 1: Upper row: color-coded FA map calculated from the ICBM DTI-81 atlas (left). Sagittal and coronal sections of T2-weighted images of the atlas (right). Lower row: extracted fiber tracts from the ICBM DTI-81 atlas in the whole brain. Rear view (left) and top view of the extracted fiber tracts in the whole brain (middle). Note the isolated CC and CS tracts (right).

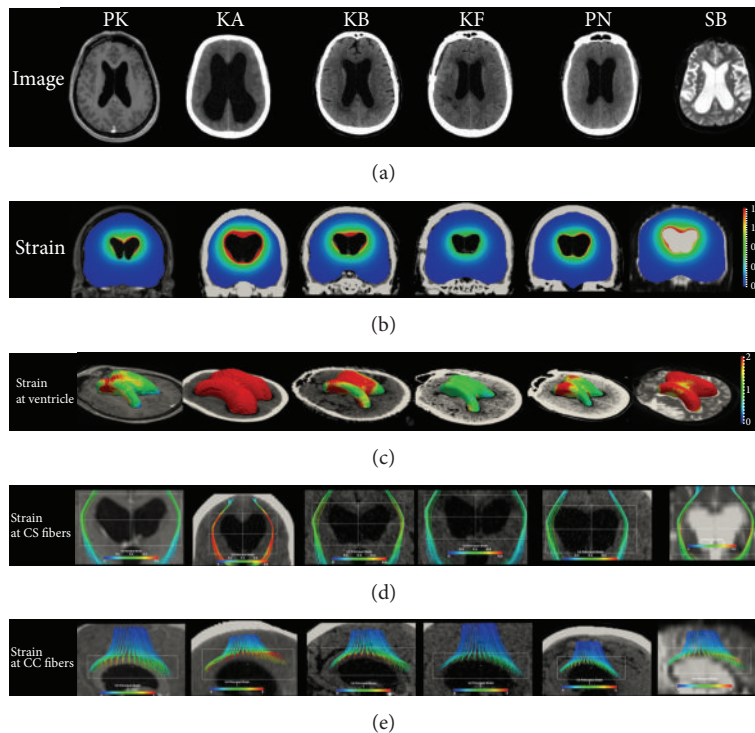


FIGURE 2: Overview of the 6 patients at their preoperative stages. Each row from top to bottom represents the medical images used for nonlinear image registration from which the strain level was quantified in the entire brain, and the calculated strain levels are illustrated at coronal sections and the ventricular walls. At bottom the strain level at CS and CC fiber tracts is presented.

The individual follow-up time ranged from 19 to 96 days. The increased strain levels should be expected to decrease more in those patients with longer follow-up times compared to those with a short follow-up time. However, there was no such pattern found among the patients in this study.

The result is exemplified with one of the six patients (patient PK) which is presented in more detail (Figure 4). As defined previously, the dilatation of the lateral ventricles

causes distortion of most of the surrounding axonal fiber tracts and where the CC and CS fiber tracts are addressed here. Thus, during the hydrocephalus development the CS fiber tracts were displaced outwards by the arrows representing the displacement vector obtained from image registration (Figure 4, upper row). The maximum displacement evaluated was 10.7 mm (Figure 4, upper row: left). The closer to the lateral ventricles, the larger the displacement found in the CS

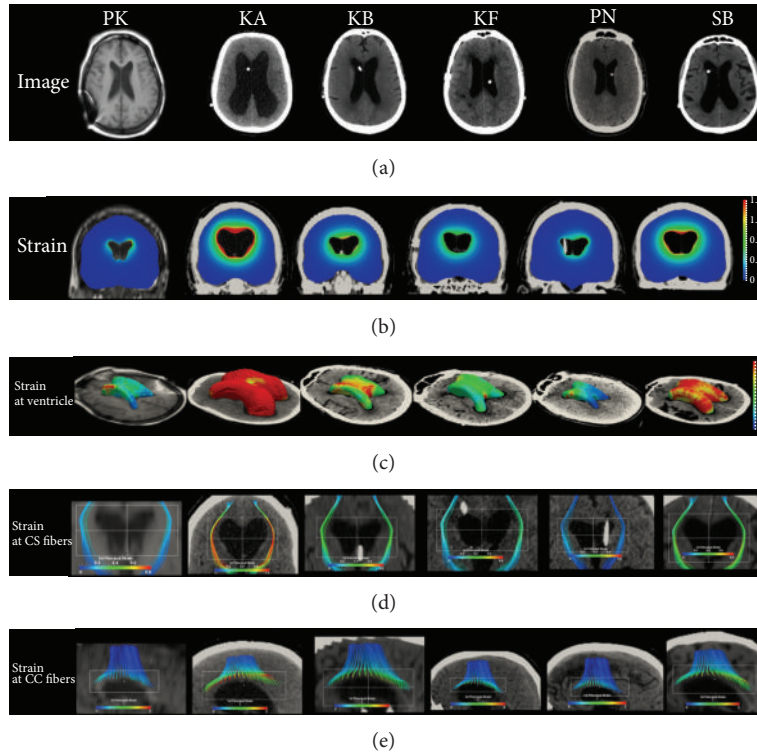


FIGURE 3: Overview of the 6 patients at their postoperative stages. Each row from top to bottom represents the medical images used for nonlinear image registration from which the strain level was quantified in the entire brain, and the calculated strain levels are illustrated at coronal sections and the ventricular walls. At bottom the strain level at CS and CC fiber tracts is presented.

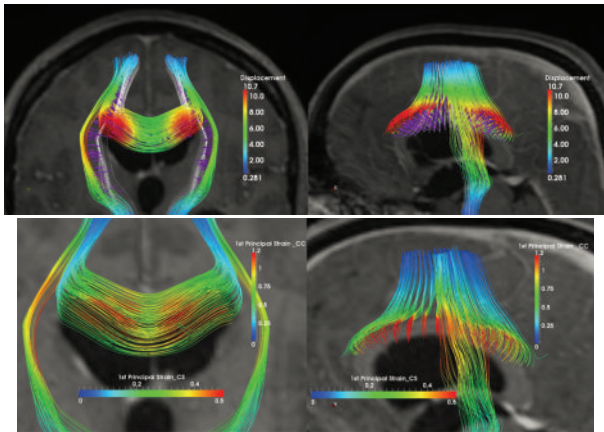


FIGURE 4: Upperrow: displacement of the CC and CS fiber tracts during hydrocephalus development overlaid with the pre-operative brain image at coronal cross-section (left) and sagittal cross-section (right). The axonal tracts with white color represent the tracts at the healthy stage, the purple arrow shows the displacement occurred during hydrocephalus development with the arrow length in proportion to the displacement magnitude, and the colored fiber tracts represent the distorted tracts at postoperative stage with color scale legend indicating magnitude of displacement. Lower row: the 1st principal strain levels at the CS and the CC fiber tracts. Note that the color scale for the CS and CC fiber tracts is different in order to have a better illustration.

fiber tracts. This resulted into compression of the CS fiber tracts close to the lateral ventricles. For axonal CC fiber tracts, the displacement field showed a more dispersed pattern with a maximum value up to 15.4 mm located at the anterior horn of lateral ventricle (Figure 4, upper row: right).

The strain levels for both CS and CC fiber tracts (Figure 4, lower row) showed that the CS fiber tracts presented a lower value compared to that found in the CC fiber tracts. Thus, the maximum strain level increase in the CS fiber tracts was 0.5, or 50%, while that in the CC fiber tracts was as high as 1.2 or 120%.

When comparing the average strain levels found in the ventricles, the CC and CS fibers, the trend seemed clear that the longer distance from the ventricles the less the strain levels both before and after the neurosurgical procedure (Figure 7). Hence, the difference was the least in fiber tracts in the periphery of the brain tissue for both pre- and postoperative stages.

To evaluate the strain level at brain tissue with further distance from the lateral ventricles, the original ventricular walls were scaled outwards until it reached the cranial boundary and which was taken as the maximum distance for evaluation. Between the original normal ventricles and the maximum dilated ventricles, ten different-sized lateral ventricles were created by equally interpolating the two extreme cases. The obtained different-sized ventricles are

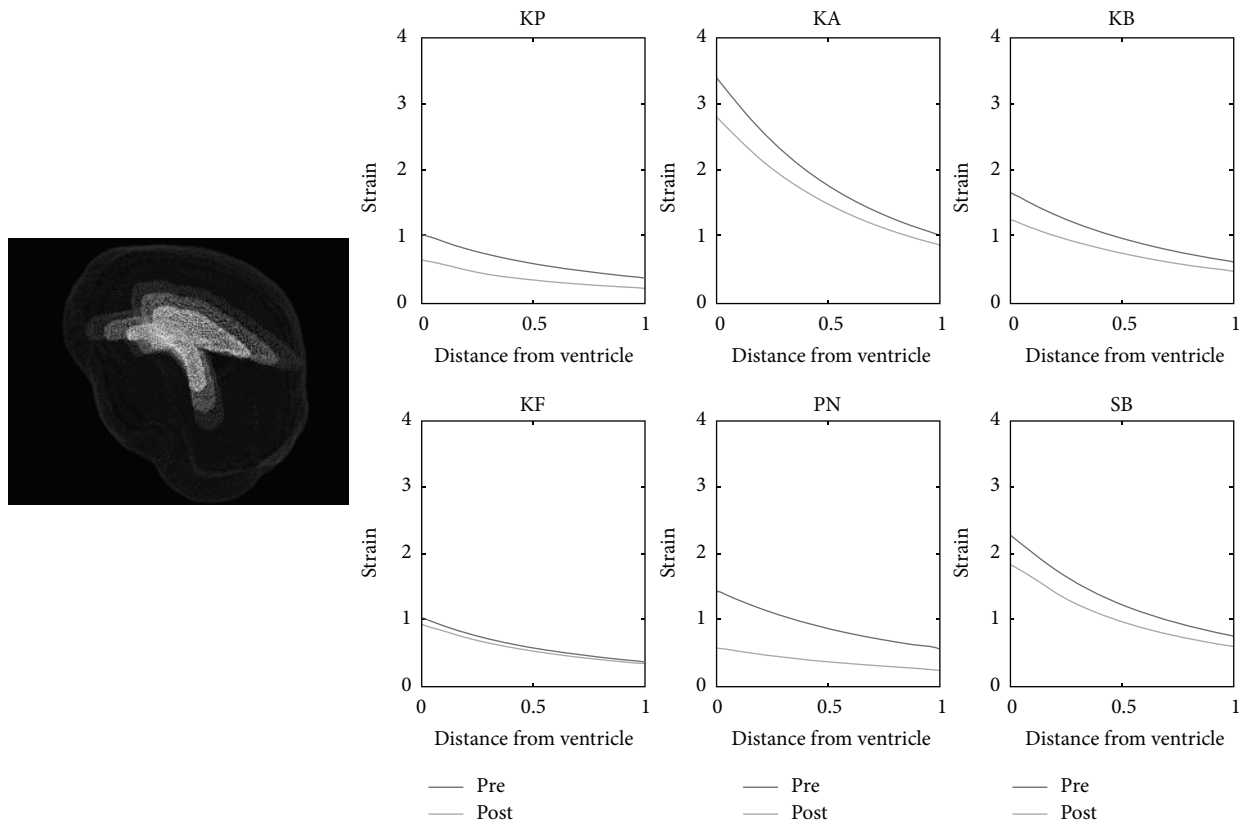


FIGURE 5: Relationship between the distance from the lateral ventricles and the mean strain level of brain tissue. The picture on the left illustrates the interpolated different-sized ventricles which were used for average strain level calculation at different distances. A decreased strain level was found after the neurosurgical procedure in all of the 6 patients in various degrees.

illustrated in Figure 5, left with only three ventricles shown aiming at a better illustration. The obtained, interpolated ventricles were then used to resample the strain levels in the entire brain and based on which the average strain levels at different distances to the ventricles were calculated. As the brain sizes are different across the patients in the present study, the distance from the ventricles was normalized for each patient by dividing the brain size to the largest distance in that particular patient. This yielded distances range from 0, that is, brain tissue adjacent to the ventricles, to 1, that is, the outermost brain tissue, which were binned in an increment of 0.1. A profile of the strain levels in brain tissue as a function distance from the ventricles could then be plotted. Thus, the relationship between the distance from the ventricles and the average strain levels showed a logarithmic profile in all patients both at the pre- and postoperative periods (Figure 5). The evaluation showed decreasing strain levels with increasing distances from the ventricles for each patient although in different degrees.

An earlier study has shown a logarithmic correlation between the regional cerebral blood flow (rCBF) and the distance from the lateral ventricles in hydrocephalic patients [1]. Thus, the displacement and increased strain levels presented in this study may have a substantial impact on the rCBF among these patients. By combining the rCBF values from that earlier study at a distance from the lateral ventricles

together with the increased strain levels found in this study at the same distance, the relationship between CBF and strain levels was plotted (Figure 6) showing a linear correlation between the increased strain level in this study and the rCBF ($r = 0.9905$, $P < 0.0001$) from the earlier study [1]. Thus, as illustrated at one point in the figure, by reducing the increased strain level from preoperative 176% to postoperative 130% shown on the horizontal axis, the rCBF increased from about 6.5 to 11.4 mL/100 mg/min (Δ rCBF in Figure 6).

4. Discussions

All six patients had increased displacement as well as profound increased strain levels in the surrounding brain tissue both before and, although less pronounced, after the neurosurgical procedure. Also, all patients were defined with cognitive disturbances. Incontinence was defined in four of the patients and without any correlation to displacement or strain level. However, the neurosurgical procedure resulted in a definite reduction of both displacement and strain level. In those two cases with the lowest increase of displacement and strain level there was no gait ataxia. This study, to the best of our knowledge, is the first to provide a quantitative view regarding the stretching of ventricular wall, CC and CS tracts in hydrocephalic patients. Further, a close correlation was found to exist between the increased strain level and the rCBF.

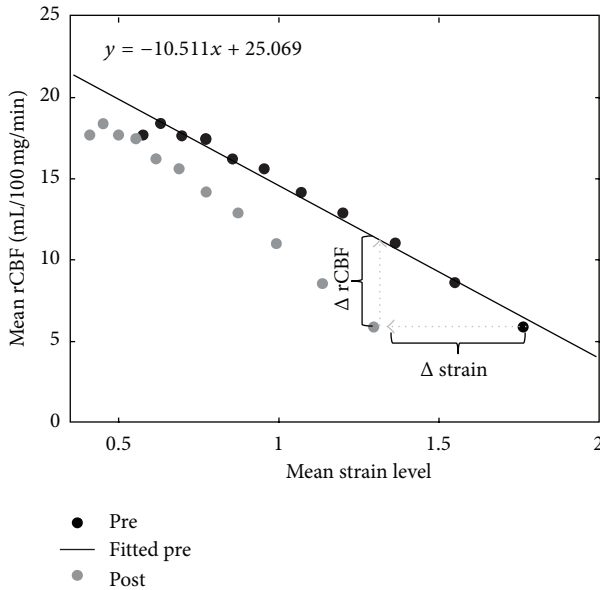


FIGURE 6: A linear relationship ($r = 0.9905$, $P < 0.0001$) was found to exist between the preoperative strain levels (x -axis where 1 = 100% strain level increase) and the regional cerebral blood flow (rCBF) (black dots, data from a previous publication [1]). One point was illustrated to show improved rCBF due to shunt operation (light dots, Δ rCBF).

The presented results provide direct evidence that Hakim triad could to some extent be explained by the increased strain levels.

Those patients with the largest expansion of the ventricles showed the most profound strain level increase in the fiber tracts of CC and CS, while those with smaller ventricular increase presented a more modest strain level increase. This may indicate that patients with larger ventricles should be expected to present a worse outcome. However, this was not the case in this study. Instead, it is suggested that larger displacement of the ventricles will influence the surrounding brain tissue more distant compared to those patients with less expanded ventricles. Also, the larger the ventricle expansion the more severe the strain level in CC fibers tracts, and this could tentatively interfere negatively with connections between the two hemispheres thereby prolonging the clinical improvement for such patients. Since the CS fiber tracts are anatomically located further away from the ventricles, they had a less pronounced strain level increase which was to be expected.

Experimental studies of axonal stretching have shown a clear correlation between the strain level and axonal function. An increased strain level of 5% was found to alter neuronal function, while 10% may cause cell death in rapid axonal stretching models [29]. Bain and Meaney (2000) demonstrated in an animal model that a strain level of approximately 21% will elicit electrophysiological changes, while a strain of approximately 34% will cause morphological signs of damage to the white matter [30]. Under slow loading rates, however, axons could tolerate much higher strain levels and with stretch of up to twice of their original length

(a strain of 100%) with no evidence of damage [31]. Using a model of sciatic nerve stretch, Fowler et al. reported that even minimal tension, if maintained for a significant amount of time, may result in loss of neuronal function [32]. Based on these, it should be expected that an increased strain level in both CC and CS fiber tracts in hydrocephalus patients found in the present study very well interferes with the normal conductivity in the axons, hence also interferes with the patients' rehabilitation. When the strain level increase reaches a critical point or threshold, it may very well initiate a biochemical response which may further alter the CC and CS fiber tracts. From a clinical aspect, the sustained increased displacement and increased strain levels of axons may also result in gait disturbances due to the potential interference with the long corticospinal fiber tracts connecting motor and sensory areas with the spinal cord as well as those in the corpus callosum connecting the electrophysiological functions of both hemispheres with each other [33].

The displacement and strain level in hydrocephalus patients belong to a static or semistatic increase thereby allowing the fiber tracts to accept the changes during a longer period. This may favor the rehabilitation even if the increased strain level is substantial due to the very low deformation rate [31]. However, of special interest are elderly people who already have an atrophied brain tissue and which may suffer more than those in employment age. Many of the hydrocephalus patients improve after ventriculoperitoneal shunt treatment. However, there is no treatment for those, who do not improve due to a lack of explanation of the causes. It was suggested that the increased strain levels may cause a disturbance in the electrophysiological function in these fiber tracts, and consequently, a tailored matrix of conductive, organic bioelectrodes could potentially be implanted in the area of corpus callosum fiber tracts thereby counterbalancing the electrophysiological dysfunction as proposed in a previous study [33].

In the present study all patients had indeed a reduction in the strain level postoperatively. However, the reduced strain level was still on a surprisingly high level and could therefore have a negative influence on the anatomy and histology of the nervous tissue together with a sustained alteration in the conductivity of the axons in the white matter. It seems that the reduction of rCBF clearly correlated with that of increased strain levels. Even a small reduction in rCBF may influence the rehabilitation especially among the vulnerable elderly patients and which should be taken into consideration.

5. Conclusions

A numerical method based on nonlinear image registration was used to quantify the stretching of ventricular wall, corpus callosum, and corticospinal axonal fiber tracts in six patients with hydrocephalus both before and after shunt operation. These data provide new insight into the mechanical cascade of events due to tissue stretching, thereby to provide us with more knowledge into understanding of the role of brain tissue and axonal stretching in some of the hydrocephalus clinical symptoms. Moreover, a linear correlation was found

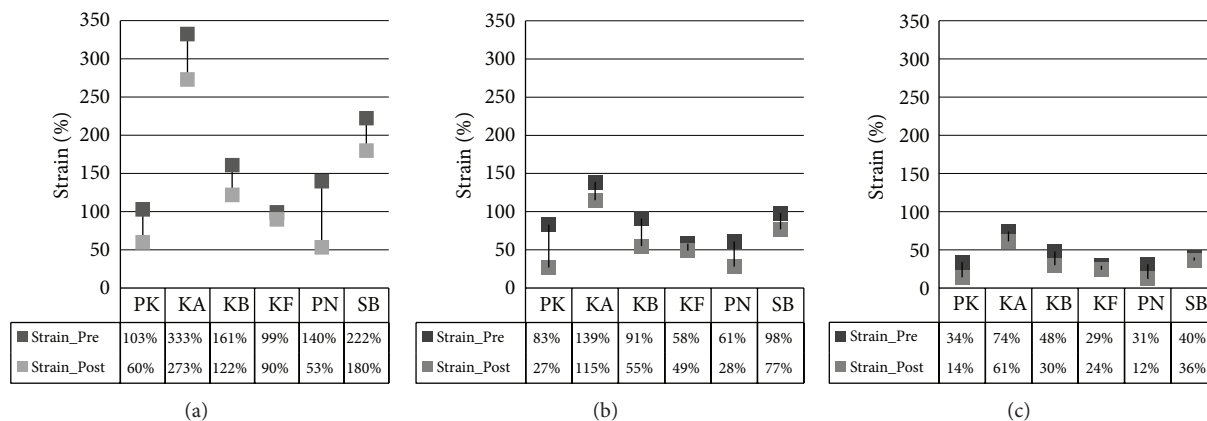


FIGURE 7: Diagram of average strain levels at the lateral ventricle (a) CC fiber (b) and CS fiber (c) for all the 6 patients. The horizontal axis is the patient's *identification* (id), and the vertical axis shows the strain levels for the corresponding patient at precraniotomy stage (Strain_Pre with darker gray color) and postcraniotomy stage (Strain_Post, with lighter gray color). The exact strain level values are presented in the lower table.

to exist between strain level and the rCBF. The combination of increased displacement of the ventricular walls, sustained increased strain levels, and derangement in electrophysiology activity in the white matter together with a reduced rCBF may, under certain circumstances, has a profound impact on the outcome following hydrocephalus.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgment

The present study was supported by research funds from the Royal Institute of Technology, Stockholm, Sweden.

References

- [1] S. Momjian, B. K. Owler, Z. Czosnyka, M. Czosnyka, A. Pena, and J. D. Pickard, "Pattern of white matter regional cerebral blood flow and autoregulation in normal pressure hydrocephalus," *Brain*, vol. 127, no. 5, pp. 965–972, 2004.
- [2] B. K. Owler and J. D. Pickard, "Normal pressure hydrocephalus and cerebral blood flow: a review," *Acta Neurologica Scandinavica*, vol. 104, no. 6, pp. 325–342, 2001.
- [3] J. M. Miller and J. P. McAllister, "Reduction of astrogliosis and microgliosis by cerebrospinal fluid shunting in experimental hydrocephalus," *Cerebrospinal Fluid Research*, vol. 4, article 5, 2007.
- [4] M. Mataró, M. Matarín, M. A. Poca et al., "Functional and magnetic resonance imaging correlates of corpus callosum in normal pressure hydrocephalus before and after shunting," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 78, no. 4, pp. 395–398, 2007.
- [5] S. Rörich, B. U. Meyer, C. Woiciechowsky, and R. Lehmann, "Callosal and corticospinal tract function in patients with hydrocephalus: a morphometric and transcranial magnetic stimulation study," *Journal of Neurology*, vol. 245, no. 5, pp. 280–288, 1998.
- [6] E. Hattingen, A. Jurcoane, J. Melber et al., "Diffusion tensor imaging in patients with adult chronic idiopathic hydrocephalus," *Neurosurgery*, vol. 66, no. 5, pp. 917–924, 2010.
- [7] T. Hattori, T. Yuasa, S. Aoki et al., "Altered microstructure in corticospinal tract in idiopathic normal pressure hydrocephalus: comparison with Alzheimer disease and Parkinson disease with dementia," *The American Journal of Neuroradiology*, vol. 32, no. 9, pp. 1681–1687, 2011.
- [8] A. Marmarou, M. Bergsneider, P. Klinge, N. Relkin, and P. M. L. Black, "INPH guidelines, part III: the value of supplemental prognostic tests for the preoperative assessment of idiopathic normal-pressure hydrocephalus," *Neurosurgery*, vol. 57, no. 3, p. S2, 2005.
- [9] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [10] H. von Holst, X. Li, and S. Kleiven, "Increased strain levels and water content in brain tissue after decompressive craniotomy," *Acta Neurochirurgica*, vol. 154, no. 9, pp. 1583–1593, 2012.
- [11] X. Li, *Finite Element and Neuroimaging Techniques to Improve Decision-Making in Clinical Neuroscience*, Royal Institute of Technology (KTH), 2012.
- [12] T. Vercauteren, X. Pennec, E. Malis, A. Perchant, and N. Ayache, "Insight into efficient image registration techniques and the demons algorithm," in *Information Processing in Medical Imaging*, pp. 495–506, Springer, New York, NY, USA, 2007.
- [13] P. Cachier, E. Bardenet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based nonrigid registration: the PASHA algorithm," *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 272–298, 2003.
- [14] J. P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.
- [15] P. Cachier, X. Pennec, and N. Ayache, *Fast Non Rigid Matching by Gradient Descent: Study and Improvements of the "Demons" Algorithm*, Rapport de Recherche-Institut National de Recherche en Informatique et en Automatique, 1999.
- [16] G. A. Holzapfel, *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*, John Wiley & Sons, West Sussex, UK, 2000.

- [17] J. Gee, T. Sundaram, I. Hasegawa, H. Uematsu, and H. Hatabu, "Characterization of regional pulmonary mechanics from serial magnetic resonance imaging data," *Academic Radiology*, vol. 10, no. 10, pp. 1147–1152, 2003.
- [18] P. J. Basser and D. K. Jones, "Diffusion-tensor MRI: theory, experimental design and data analysis—a technical review," *NMR in Biomedicine*, vol. 15, no. 7-8, pp. 456–467, 2002.
- [19] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance B*, vol. 103, no. 3, pp. 247–254, 1994.
- [20] T. L. Chenevert, "Principles of diffusion-weighted imaging (DW-MRI) as applied to body imaging," in *Diffusion-Weighted MR Imaging*, pp. 3–17, 2010.
- [21] P. J. Basser and C. Pierpaoli, "A simplified method to measure the diffusion tensor from seven MR images," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 928–934, 1998.
- [22] D. Le Bihan, J. F. Mangin, C. Poupon et al., "Diffusion tensor imaging: concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534–546, 2001.
- [23] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical Journal*, vol. 66, no. 1, pp. 259–267, 1994.
- [24] P. J. Basser, "Inferring microstructural features and the physiological state of tissues from diffusion-weighted images," *NMR in biomedicine*, vol. 8, no. 7-8, pp. 333–344, 1995.
- [25] S. Mori and J. Zhang, "Principles of diffusion tensor imaging and its applications to basic neuroscience research," *Neuron*, vol. 51, no. 5, pp. 527–539, 2006.
- [26] S. Mori, B. J. Crain, V. Chacko, and P. van Zijl, "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging," *Annals of Neurology*, vol. 45, no. 2, pp. 265–269, 1999.
- [27] S. Mori, K. Oishi, H. Jiang et al., "Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template," *NeuroImage*, vol. 40, no. 2, pp. 570–582, 2008.
- [28] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using DT-MRI data," *Magnetic Resonance in Medicine*, vol. 44, no. 4, pp. 625–632, 2000.
- [29] Z. Yu and B. Morrison, "Experimental mild traumatic brain injury induces functional alteration of the developing hippocampus," *Journal of Neurophysiology*, vol. 103, no. 1, pp. 499–510, 2010.
- [30] A. C. Bain and D. F. Meaney, "Tissue-level thresholds for axonal damage in an experimental model of central nervous system white matter injury," *Journal of Biomechanical Engineering*, vol. 122, no. 6, pp. 615–622, 2000.
- [31] M. D. Tang-Schomer, A. R. Patel, P. W. Baas, and D. H. Smith, "Mechanical breaking of microtubules in axons during dynamic stretch injury underlies delayed elasticity, microtubule disassembly, and axon degeneration," *FASEB Journal*, vol. 24, no. 5, pp. 1401–1410, 2010.
- [32] S. S. Fowler, J. P. Leonetti, J. C. Banich, J. M. Lee, R. Wurster, and M. R. I. Young, "Duration of neuronal stretch correlates with functional loss," *Otolaryngology—Head and Neck Surgery*, vol. 124, no. 6, pp. 641–644, 2001.
- [33] H. von Holst, "Organic bioelectrodes in clinical neurosurgery," *Biochimica et Biophysica Acta*, 2012.

Research Article

Mathematical Modelling of Biomechanical Interactions between Backpack and Bearer during Load Carriage

Lei Ren,¹ David Howard,² and Richard K. Jones²

¹ School of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester M60 1QD, UK

² Centre for Rehabilitation and Human Performance Research, University of Salford, Salford M5 4WT, UK

Correspondence should be addressed to Lei Ren; lei.ren@manchester.ac.uk

Received 11 March 2013; Accepted 27 March 2013

Academic Editor: Chang-Hwan Im

Copyright © 2013 Lei Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a three-dimensional mathematical model of the biomechanical interactions between backpack and bearer during load carriage. The model considers both the coupled pack motions, which follow the torso, and also the longitudinal compliance and damping in the backpack suspension. The pack interaction forces and moments, acting on the bearer, are determined from kinematic relationships, equations of motion, and a dynamic pack suspension model. The parameters of the pack suspension model were identified from test data obtained using a load carriage test rig. Output from the load carriage mathematical model has been compared with measurement data during human gait and conclusions drawn with regard to the validity of the proposed approach.

1. Introduction

Backpacks are common devices for increasing human load carriage performance but when heavily loaded may lead to excessive joint loadings, muscle fatigue, or even injury [1–5]. Most studies of personal load carriage systems have focused on physiological and biomechanical aspects [6–10]. Biomechanical studies have concentrated on experimental gait analysis, including the effects of load carriage on electromyographic activity [3, 8, 11], gait and posture [6, 7, 10], and ground reactions [6, 8, 9]. Almost all of these studies consider the backpack to be part of the trunk segment, with no relative motion between the two. Little is known about the pack interaction forces and moments acting on the bearer's trunk.

During load carriage, the pack interaction forces exerted on the torso relate directly to perceived discomfort, fatigue, and the risk of injury, for example, rucksack palsy and back problems [12, 13]. A better understanding of these interaction forces would help to improve the design of future load carriage systems. Unfortunately, in contrast to the contact pressure distribution [14–16], the interaction forces and moments between pack and torso cannot be measured

directly; however mathematical modelling and simulation offer an alternative approach.

Very little has been published on the modelling of load carriage biomechanics. While examining the effects of pack load on back muscle EMG, Bobet and Norman [3] considered the pack to be part of the trunk segment for inverse dynamics purposes. Recently, Pelot et al. [17] developed a two-dimensional backpack model to investigate the forces at the shoulder and hip belts. However, their model assumes no relative movement between pack and torso and can only be used to evaluate static forces. To properly model backpack dynamics, it is necessary to include the relative motions between pack and torso, which is capable of representing the dynamic impacts on the bearer rather than static interactions only.

This paper proposes a mathematical load carriage model, which represents the three-dimensional biomechanical interactions between pack and torso. All pack motions are directly coupled to those of the bearer's trunk, with the exception of longitudinal motion up and down the bearer's back, which depends on a pack suspension model, the parameters of which were identified from test data obtained using a load carriage test rig [18]. The validity of the proposed

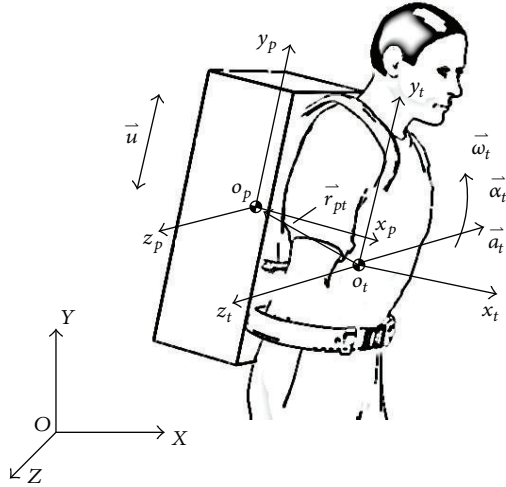


FIGURE 1: The 3D pack and trunk motions during load carriage and the three coordinate systems. $OXYZ$ is the global (inertial) coordinate system, $o_t x_t y_t z_t$ is the trunk coordinate system, and $o_p x_p y_p z_p$ is the backpack coordinate system.

approach is investigated and conclusions drawn with regard to future developments in the modelling of load carriage biomechanics.

2. Methods

2.1. Pack-Trunk Kinematics. Throughout this study, the typical backpack system with two shoulder straps and a waist belt is addressed. If deformations of the backpack are neglected, the pack can be considered to be a rigid body (Figure 1). Due to the actions of the shoulder straps and hip belt, holding the pack against the bearer's back, pack rotations relative to the trunk in the sagittal and transverse planes, and translation perpendicular to the back are all constrained. For simplicity, relative pack rotation in the frontal plane and lateral pack translation are also considered to be negligible. In other words, pack motion relative to the trunk only occurs in a direction which coincides with the trunk's longitudinal axis. So, the backpack is modelled as a rigid body that slides up and down the bearer's back while all other degrees of freedom follow the motions of the trunk.

A global coordinate system $OXYZ$, a trunk system $o_t x_t y_t z_t$, and a backpack system $o_p x_p y_p z_p$ are defined (Figure 1). The global system's x -axis lies in the sagittal plane and points in the direction of forward progress, the y -axis also lies in the sagittal plane and points upwards, and the z -axis lies in the frontal plane and points to the right. The trunk coordinate system $o_t x_t y_t z_t$ is attached to the bearer's trunk with its origin located at the trunk centre of mass (CoM). The y_t -axis is parallel to the trunk longitudinal axis (midpoint between C7 and T8 spinous process to the midpoint between xiphoid process and jugular notch) and points upwards. The x_t -axis is perpendicular to the bearer's back and points in the anterior direction. The z_t -axis is perpendicular to both x_t -axis and y_t -axis and points to the right. The local pack

coordinate system $o_p x_p y_p z_p$ is attached to the backpack with its origin at the pack CoM, and its axes are parallel to those of the trunk coordinate system. In the analysis below, we have followed the common practice of not expressing vector equations in any particular coordinate frame. Where vector equations have been expanded, to derive the individual components, the resulting scalar equations are expressed in a particular coordinate frame, which is stated.

Based on the rigid body pack model proposed above, the pack's kinematic relationship with the trunk can be described by

$$\vec{a}_p = \vec{a}_t + \vec{a}_{pt}^n + \vec{a}_{pt}^{\tau} + \vec{a}_r + \vec{a}_r^c, \quad (1a)$$

$$\vec{\omega}_p = \vec{\omega}_t, \quad (1b)$$

$$\vec{\alpha}_p = \vec{\alpha}_t, \quad (1c)$$

where \vec{a}_p is the absolute acceleration of the pack CoM; \vec{a}_t is the absolute acceleration of the trunk CoM; \vec{a}_{pt}^n and \vec{a}_{pt}^{τ} are the centripetal and tangential acceleration vectors, relative to the trunk CoM, of a point fixed in the trunk coordinate frame and instantaneously coincident with the pack CoM; \vec{a}_r is the relative pack acceleration vector along the back; and \vec{a}_r^c is the Coriolis acceleration vector.

The various acceleration terms in (1a) are given by

$$\vec{a}_{pt}^{\tau} = \vec{\alpha}_p \times \vec{r}_{pt}, \quad (2a)$$

$$\vec{a}_{pt}^n = \vec{\omega}_t \times (\vec{\omega}_t \times \vec{r}_{pt}), \quad (2b)$$

$$\vec{a}_r = \ddot{\vec{u}}, \quad (2c)$$

$$\vec{a}_r^c = 2\vec{\omega}_p \times \dot{\vec{u}}, \quad (2d)$$

where $\vec{r}_{pt} = (\vec{d} + \vec{u})$ is the position vector of the pack CoM relative to the trunk CoM, \vec{u} is the pack translation relative to the trunk, and \vec{d} is the position of the pack CoM, relative to the trunk CoM, in the unloaded condition.

Substituting (2a), (2b), (2c), and (2d) into (1a) and considering each component in the backpack coordinate system, the absolute translational accelerations of the pack CoM in all three directions can be derived as

$$a_{px} = a_{tx} + \alpha_{ty} d_z - \alpha_{tz} (d_y + u) + \omega_{tx} \omega_{ty} (d_y + u) - (\omega_{ty}^2 + \omega_{tz}^2) d_x + \omega_{tx} \omega_{tz} d_z - 2\omega_{tz} \dot{u}, \quad (3a)$$

$$a_{py} = a_{ty} + \alpha_{tz} d_x - \alpha_{tx} d_z + \omega_{ty} \omega_{tz} d_z - (\omega_{tx}^2 + \omega_{tz}^2) (d_y + u) + \omega_{tx} \omega_{ty} d_x + \ddot{u}, \quad (3b)$$

$$a_{pz} = a_{tz} + \alpha_{tx} (d_y + u) - \alpha_{ty} d_x + \omega_{tx} \omega_{tz} d_x - (\omega_{tx}^2 + \omega_{ty}^2) d_z + \omega_{ty} \omega_{tz} (d_y + u) + 2\omega_{tx} \dot{u}. \quad (3c)$$

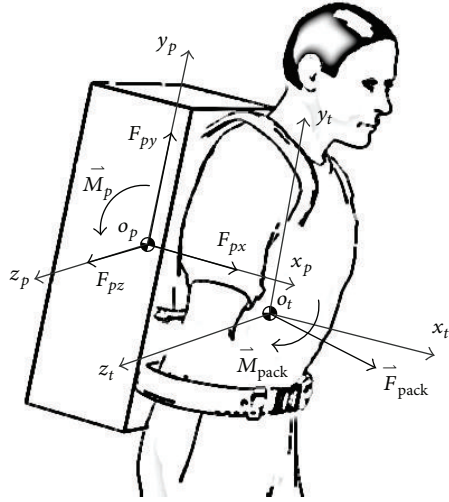


FIGURE 2: The human-pack interaction force \vec{F}_p (F_{px} , F_{py} , F_{pz}) and moment \vec{M}_p acting at the pack mass centre, and the associated reactions \vec{F}_{pack} and \vec{M}_{pack} acting at the trunk mass centre.

From (3a), (3b), and (3c), it can be seen that the backpack's absolute accelerations are not only influenced by trunk translation but also strongly coupled with trunk rotation and relative pack motion.

2.2. Backpack's Equation of Motion. If the backpack is regarded as a separate object and isolated from the torso (Figure 2), the pack interaction forces and moments can be assessed. In this case, the interaction between the pack and the body is described by the resultant force and moment (\vec{F}_p and \vec{M}_p) acting on the pack and at its CoM. This equivalent force system represents the net effects of the various forces at the pack-human interface as they act on the pack.

Applying the Newton-Euler equations leads to

$$m_p \vec{a}_p = \vec{F}_p + m_p \vec{g}, \quad (4a)$$

$$J_p \vec{\alpha}_p + \vec{\omega}_p \times (J_p \vec{\omega}_p) = \vec{M}_p. \quad (4b)$$

Equation (4a) can be expressed in the backpack coordinate system as

$$F_{px} = m_p a_{px} + m_p g (\sin \psi \cos \phi + \cos \psi \cos \theta \sin \phi), \quad (5a)$$

$$F_{py} = m_p a_{py} + m_p g (\sin \psi \sin \phi + \cos \psi \cos \theta \cos \phi), \quad (5b)$$

$$F_{pz} = m_p a_{pz} - m_p g \cos \psi \sin \theta, \quad (5c)$$

where the orientation of the backpack with respect to the global coordinate system $OXYZ$ is defined by Euler angles (ψ, θ, ϕ) in ZXZ sequence.



FIGURE 3: The dynamic load carriage test rig. The hydraulic ram drives the mannequin up and down at different frequencies and amplitudes. Accelerometers measure the motion of the backpack and the mannequin, and a load cell measures the dynamic force propelling the mannequin.

2.3. Backpack Suspension Model. In this study, a general non-linear pack suspension model is employed, which relates the pack interaction force along the trunk longitudinal axis to the relative pack motion. The relationship can be written as

$$F_{py} = F_{pe} + F_{pd} + F_{pi}, \quad (6)$$

where F_{pe} is the elastic component of the interaction force, F_{pd} is the damping component, and F_{pi} is an inertial coupling term. Cubic polynomials are used to describe these non-linear properties [19] as follows:

$$\begin{aligned} F_{pe} &= a_3 u^3 + \text{sign}(u) a_2 u^2 + a_1 u, \\ F_{pd} &= b_3 \dot{u}^3 + \text{sign}(\dot{u}) b_2 \dot{u}^2 + b_1 \dot{u}, \\ F_{pi} &= c_3 \ddot{u}^3 + \text{sign}(\ddot{u}) c_2 \ddot{u}^2 + c_1 \ddot{u}, \end{aligned} \quad (7)$$

where $a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3$ are model parameters, which are constant for a given type of pack and specified working conditions (i.e., pack load, load distribution, strap and belt tensions, etc.).

For a particular backpack, the parameters of the suspension model can be identified from dynamic test data, obtained using the hydraulically driven load carriage test rig shown in Figure 3 [19]. The mannequin was covered in a neoprene-like material to mimic the soft tissues. The hydraulic ram drives the mannequin up and down with different frequency and amplitude inputs, which allows non-linear frequency response testing over a range of frequencies and amplitudes.

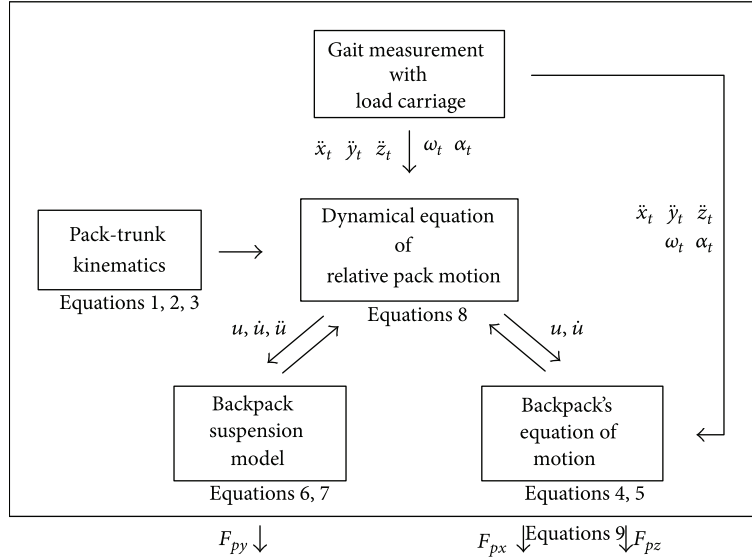


FIGURE 4: The schematic diagram for the pack-bearer interaction force calculation based on pack-trunk kinematics, backpack's equation of motion, and backpack suspension model.

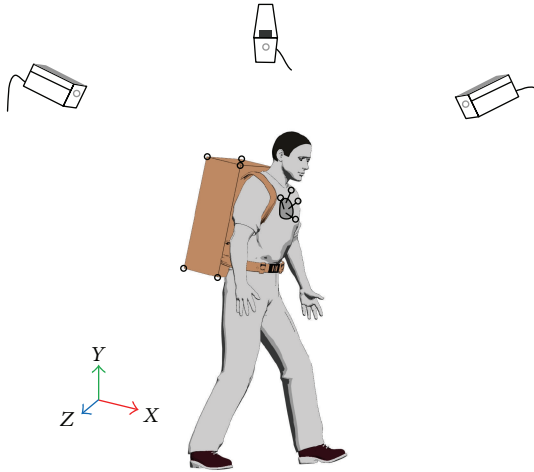


FIGURE 5: The 3D gait measurement setup for capturing the trunk and pack motions during load carriage. A specially designed plastic plate carrying four markers, firmly attached to the thorax, was used to capture the position and orientation of the trunk. Eight markers were attached to the corners of the pack to capture its position and orientation.

The harmonic analysis method was used to identify the model parameters [18].

2.4. Pack-Bearer Interaction Forces during Walking. Using (1a), (1b), and (1c) to (7) and measured trunk motion data, the pack interaction forces and moments can be derived as follows. Substituting the expression for acceleration a_{py} (3b) into the pack's equation of motion (5b), and using the backpack suspension model (6) and (7) to substitute for F_{py} , leads to the following non-linear second-order differential

equation, which describes the dynamic pack behaviour along the bearer's back:

$$\begin{aligned}
 & c_3 \ddot{u}^3 + c_2 \text{sign}(\ddot{u}) \ddot{u}^2 + (c_1 + m_p) \ddot{u} + b_3 \dot{u}^3 + b_2 \text{sign}(\dot{u}) \dot{u}^2 \\
 & + b_1 \dot{u} + a_3 u^3 + a_2 \text{sign}(u) u^2 + (a_1 - m_p (\omega_x^2 + \omega_z^2)) u \\
 & + m_p (a_{ty} + \alpha_z d_x - \alpha_x d_z + \omega_y \omega_z d_z \\
 & - (\omega_x^2 + \omega_z^2) d_y + \omega_x \omega_y d_x \\
 & + g (\sin \psi \sin \phi + \cos \psi \cos \theta \cos \phi)) = 0.
 \end{aligned} \tag{8}$$

As this differential equation is of a highly non-linear form and some of the terms are time varying, a numerical solution is required (4th-order Runge-Kutta). Because the numerical integration time step is normally smaller than the gait measurement interval, cubic interpolation is used to provide trunk motion data at the necessary frequency. The initial values of the relative pack displacement and velocity were set to zero, and the numerical integration algorithm executed until a steady-state cyclic pack motion was achieved. Steady state was defined as the initial and final values of the state variables for one gait cycle being equal within an acceptable tolerance ($1.0e - 6$ was used in this study).

Thus, given the trunk motion data (a_{ty} , ψ , θ , ϕ , ω_x , ω_y , ω_z , α_x , α_z), the relative pack motion can be derived (u , \dot{u} , and \ddot{u}). Then the pack force along the bearer's back F_{py} can be obtained by substituting the calculated relative motion (u , \dot{u} , and \ddot{u}) into (6) and (7).

The expressions for the normal pack force F_{px} and lateral pack force F_{pz} are obtained by substituting the expressions

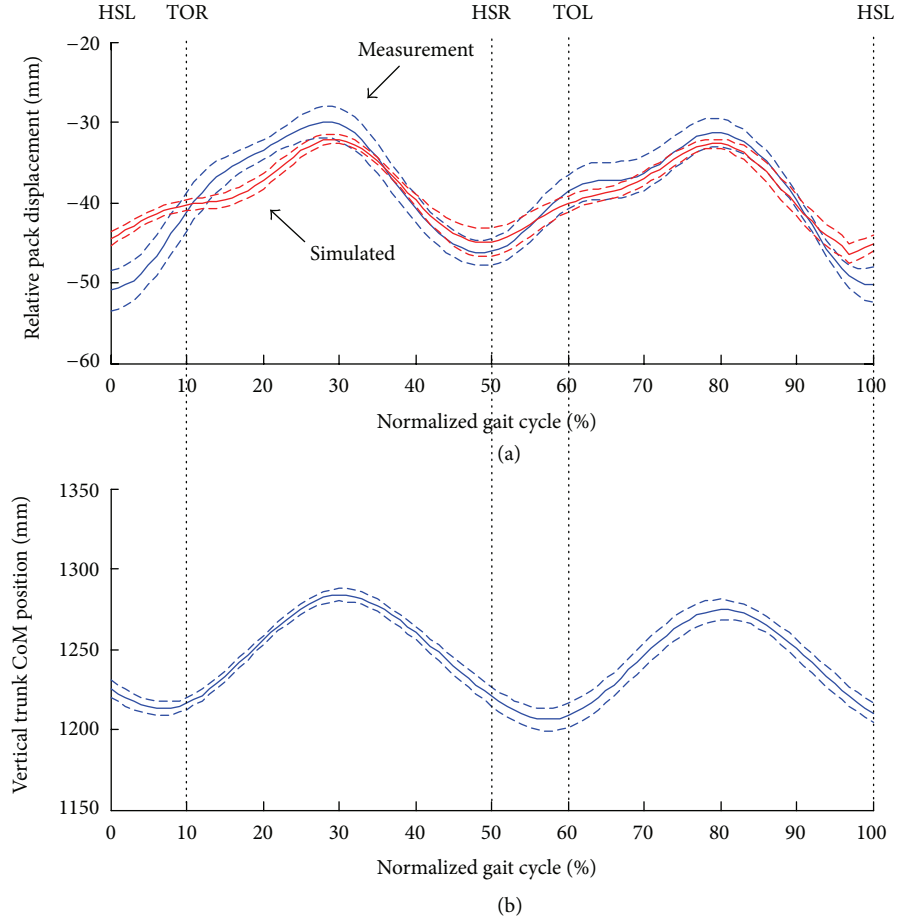


FIGURE 6: The simulated relative pack displacement along the back (mean (solid red line) \pm one standard deviation (dashed red lines), (a)), compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines), (a)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm), and measured vertical trunk CoM displacement (mean (solid blue line) \pm one standard deviation (dashed blue lines), (b)). The average walking speed was 1.63 ± 0.03 m/s, and the pack load was 11.5 kg. The stance phase for the left limb is from 0 to 60%, and the swing phase is from 60% to 100%. The double support phase is from 0 to 10% and from 50 to 60%.

for accelerations a_{px} and a_{pz} ((3a) and (3c)) into the pack's equations of motion ((5a) and (5c)), which leads to

$$\begin{aligned}
 F_{px} &= m_p \left(a_{tx} + \alpha_y d_z - \alpha_z (d_y + u) + \omega_x \omega_y (d_y + u) \right. \\
 &\quad \left. - (\omega_y^2 + \omega_z^2) d_x + \omega_x \omega_z d_z - 2\omega_z \dot{u} \right. \\
 &\quad \left. + g (\sin \psi \cos \phi + \cos \psi \cos \theta \sin \phi) \right), \\
 F_{pz} &= m_p \left(a_{tz} + \alpha_x (d_y + u) - \alpha_y d_x + \omega_x \omega_z d_x \right. \\
 &\quad \left. - (\omega_x^2 + \omega_y^2) d_z + \omega_y \omega_z (d_y + u) \right. \\
 &\quad \left. + 2\omega_x \dot{u} - g \cos \psi \sin \theta \right).
 \end{aligned} \tag{9}$$

Thus, using (6), (7), and (9), the three components of the pack force, acting on the pack CoM, can be determined from the trunk motion data and the calculated relative pack motion. The resultant pack moment, acting about the pack

CoM, can be calculated directly from the trunk rotations, given that the pack rotates with the trunk (see (1b), (1c), and (4b)).

The pack interaction force and moment can also be represented by an equivalent force system (\vec{F}_{pack} , \vec{M}_{pack}) acting on the trunk CoM (Figure 2). According to Newton's third law, the interaction force and moment acting at the trunk CoM can be expressed as

$$\begin{aligned}
 \vec{M}_{\text{pack}} &= \vec{F}_p \times \vec{r}_{tp} - \vec{M}_p, \\
 \vec{F}_{\text{pack}} &= -\vec{F}_p.
 \end{aligned} \tag{10}$$

The pack force and moment can be used to assess the mechanical loads imposed on the human body during load carriage. Moreover, when input into a three-dimensional human gait model, joint loads and mechanical energy expenditure can also be evaluated.

The proposed dynamic load carriage model has been implemented in the MATLAB programming environment (see Figure 4 for a schematic diagram). Measured 3D trunk

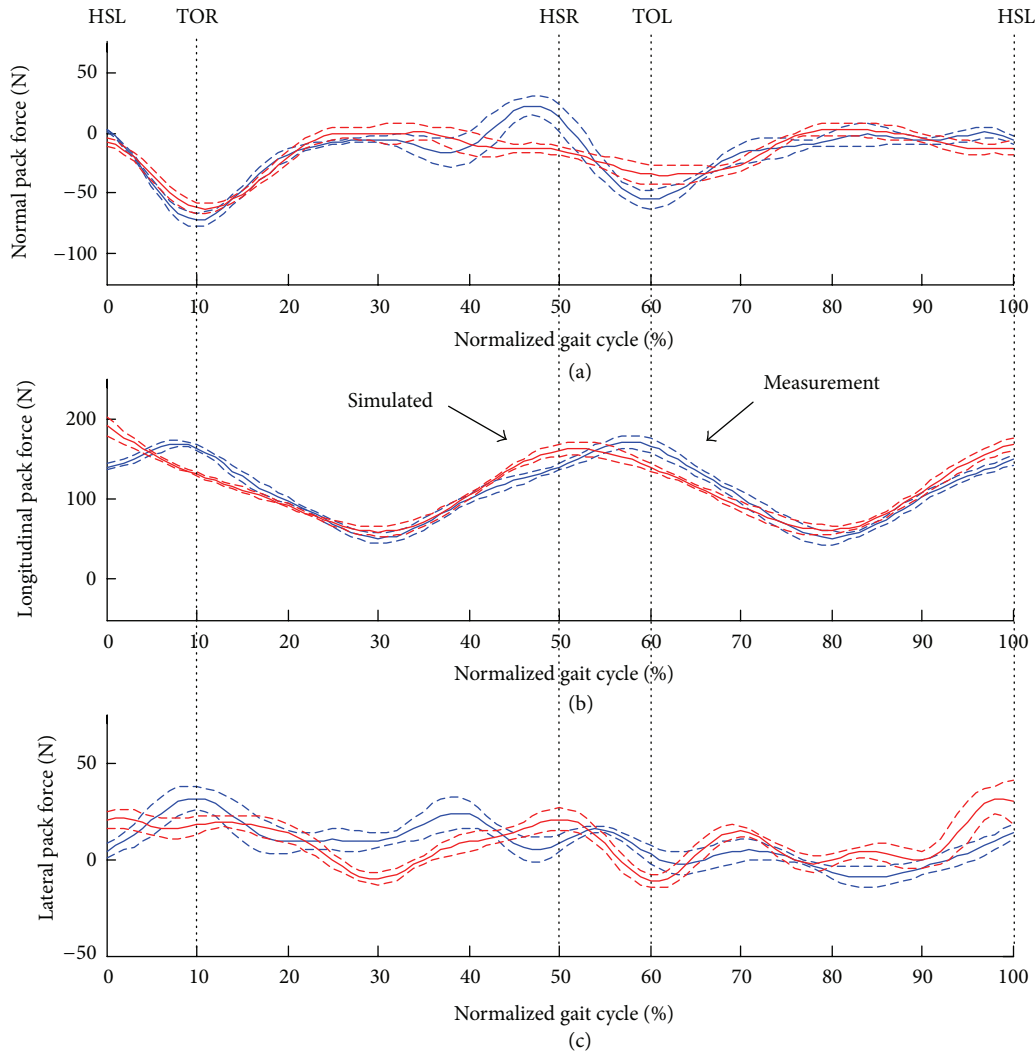


FIGURE 7: The simulated pack interaction forces in the pack coordinate system mean ((solid red line) \pm one standard deviation (dashed red lines): (a) longitudinal pack force, (b) normal pack force (c) and lateral pack force, compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm). The average walking speed was 1.63 ± 0.03 m/s, and the pack load was 11.5 kg. The stance phase for the left limb is from 0 to 60%, and the swing phase is from 60% to 100%. The double support phase is from 0 to 10% and from 50 to 60%.

rotations and translations over a complete gait cycle, with a loaded backpack, were used as input data.

2.5. Experimental Model Assessment. To assess the validity of the proposed modelling approach, three-dimensional gait measurement was conducted to capture trunk and pack motions whilst carrying a backpack. Two healthy male subjects (age: 30 ± 2 yrs, weight: 75 ± 3.2 kg) were selected from a population of postgraduate students. Prior to participation, the subjects provided informed consent in accordance with the policies of Salford University's Ethical Advisory Committee. The subjects walked in trainers along a walkway, inside a gait laboratory, while motion data was collected at 100 Hz using a 6-camera Qualisys motion analysis system (Qualisys AB, Svedalén, Sweden). The subjects walked at two different speeds, normal and fast, with two backpack loads, 11.5 kg

and 23.0 kg (see Figure 5). Each experimental condition was measured six times to ensure that a repeatable data set for a complete walking cycle was obtained.

To capture trunk motions, a specially designed plastic plate, carrying four reflective markers, was firmly attached to the thorax [20]. Four anatomical landmarks (C7 and T8 spinous processes, xiphoid process, and jugular notch) were used to define the trunk coordinate system. Before the walking trials, a static calibration procedure was used to locate these anatomical landmarks based on the CAST methodology (calibrated anatomical system technique) [21]. The position of the trunk CoM was estimated from the anatomical landmarks using anthropometric data from the literature [22].

To minimize the deformation of the backpack, the test pack contained a specially designed plastic foam block with

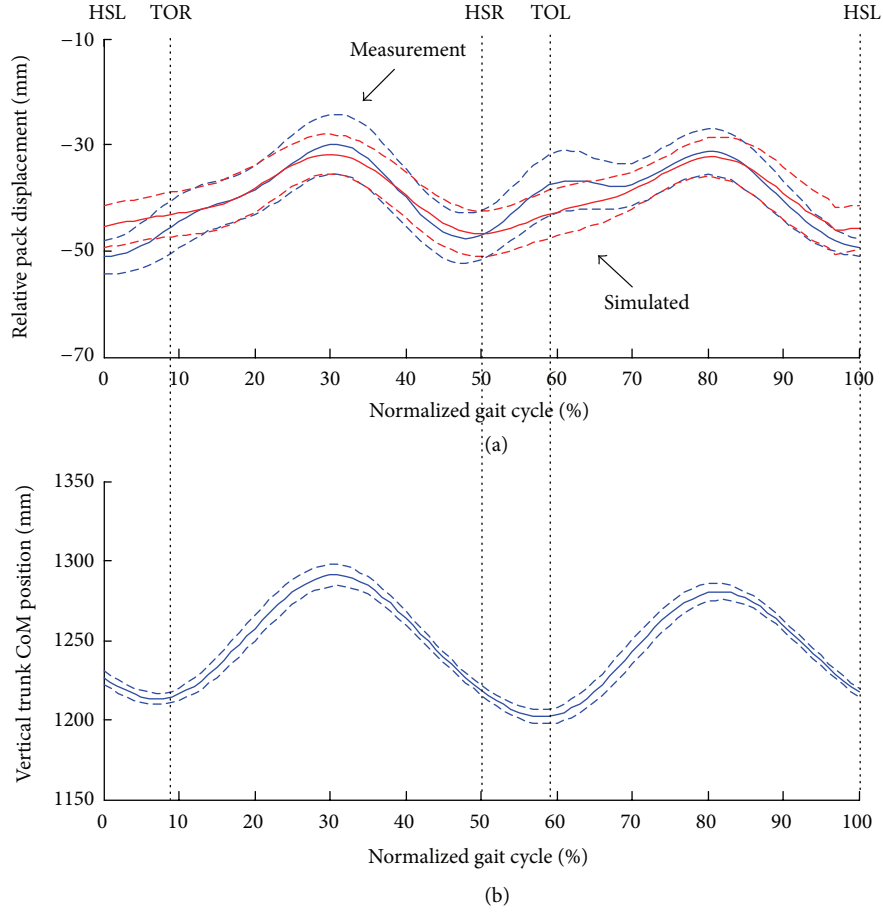


FIGURE 8: The simulated relative pack displacement along the back (mean (solid red line) \pm one standard deviation (dashed red lines), (a)), compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines), (a)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm), and measured vertical trunk CoM displacement (mean (solid blue line) \pm one standard deviation (dashed blue lines), (b)). The average walking speed was 1.93 ± 0.03 m/s, and the pack load was 11.5 kg. The stance phase for the left limb is from 0 to 59%, and the swing phase is from 59% to 100%. The double support phase is from 0 to 9% and from 50 to 59%.

cylindrical cavities to allow metal bars to be firmly inserted. By putting different metal bars in different cavities, different pack loads and load distributions were possible. During walking trials, eight reflective markers were attached to the corners of the backpack, and the geometric centre of the pack was assumed to be equidistant from these markers. The pack's CoM position and inertia matrix were then calculated based on this geometric centre and the load distribution within the pack.

The raw 3D marker data were processed using a custom-written software package, SMAS (Salford Motion Analysis Software), which has been developed for 3D kinematic and kinetic analysis of general biomechanical multibody systems [20, 23]. The data were filtered using a low pass 4th-order Butterworth digital filter with a cut-off frequency of 15 Hz. The orientations of the backpack and trunk segments were derived using an optimal estimation algorithm [24]. The pack and trunk angular velocities, angular accelerations, and CoM linear accelerations were calculated using a finite difference method.

Based on these measurement data, the interaction force and moment acting at the pack CoM were estimated as follows:

$$\begin{aligned}\vec{F}_{pm} &= m_p (\vec{a}_{pm} - \vec{g}), \\ \vec{M}_{pm} &= J_p \vec{\alpha}_{pm} + \vec{\omega}_{pm} \times (J_p \vec{\omega}_{pm}).\end{aligned}\quad (11)$$

These data for the whole gait cycle were then used to validate the 3D load carriage model.

3. Results

Experimental and simulated data for one subject walking at two different speeds, normal (1.6 m/s) and fast (1.9 m/s), with two backpack loads, 11.5 kg and 23.0 kg, are presented in Figures 6 to 11. Results are shown for repeated experimental trials and the corresponding simulations.

Figures 6 and 7 show the vertical trunk CoM motion, the simulated relative pack displacement, and the pack

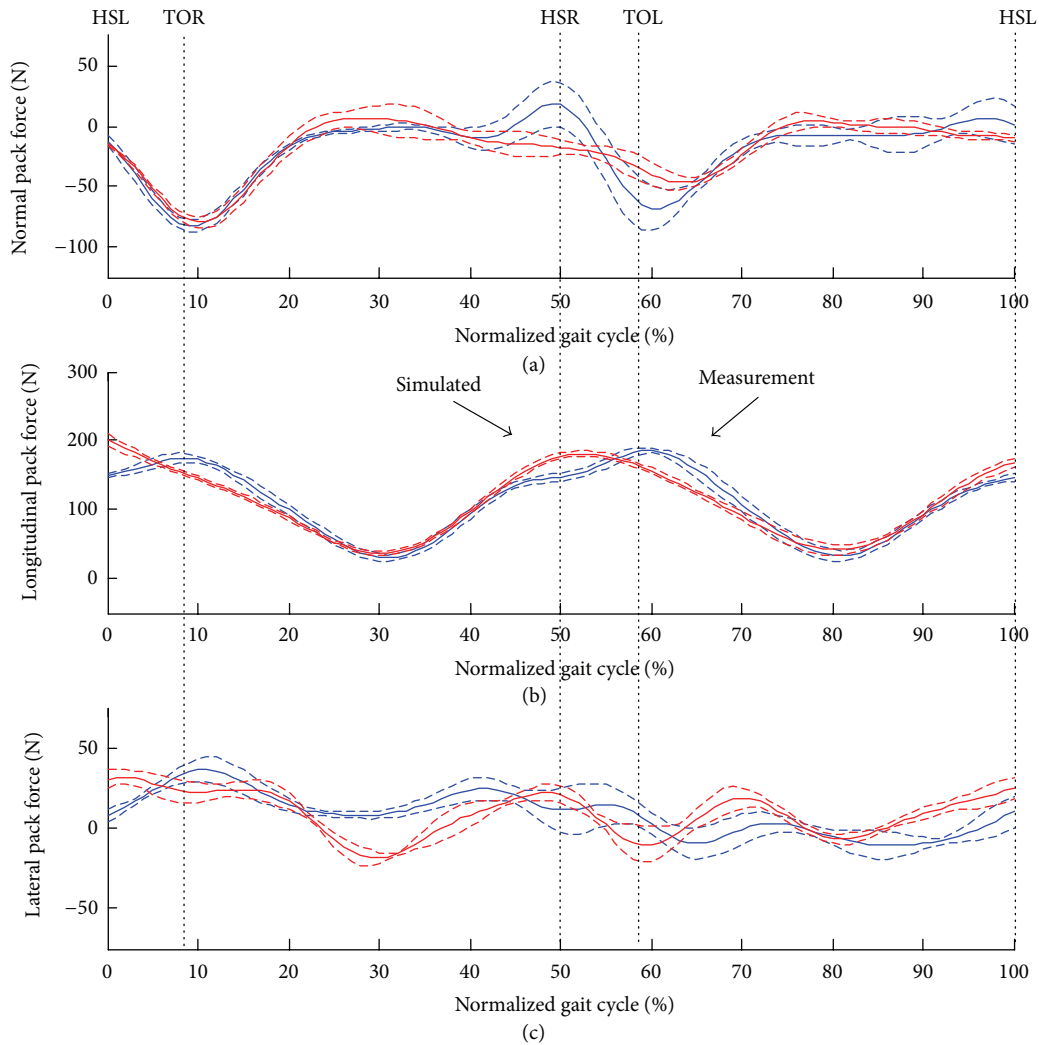


FIGURE 9: The simulated pack interaction forces in the pack coordinate system (mean (solid red line) \pm one standard deviation (dashed red lines)): (a) longitudinal pack force, (b) normal pack force (c), and lateral pack force, compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm). The average walking speed was 1.93 ± 0.03 m/s, and the pack load was 11.5 kg. The stance phase for the left limb is from 0 to 59%, and the swing phase is from 59% to 100%. The double support phase is from 0 to 9% and from 50 to 59%.

interaction forces, over a complete gait cycle, for normal walking with an 11.5 kg load. Figures 8 and 9 show the same data for fast walking and an 11.5 kg load. Figures 10 and 11 show the same data for normal walking with a 23 kg load.

4. Discussion

A 3D load carriage model has been proposed, which simulates the dynamic interactions between pack and bearer. This considers pack motions that are directly coupled to the bearer's trunk movement and also the longitudinal compliance and damping in the backpack suspension. The 3D pack interaction forces and moments are determined from kinematic relationships, equations of motion, and a dynamic backpack suspension model. The parameters of the pack suspension model were identified from test data collected using a load carriage test rig [18, 19]. Output from the load

carriage simulation model was compared with measurement data obtained from a motion analysis system at two different pack loads, and during both normal and fast walking.

Referring to Figures 6(a), 8(a), and 10(a), the simulated relative pack displacements (longitudinally along the bearer's back) show reasonably good agreement with the measured data at both pack loads, and at both walking speeds. They reproduce a sinusoidal fluctuation about 20 mm with a very similar pattern to vertical displacement of trunk CoM (Figures 6(b), 8(b), and 10(b)), which is consistent with previous load carriage measurements [10]. However, the results appear to contain an additional higher-frequency harmonic that is not present in the measurement data, especially noticeable at the normal walking speed.

Referring to Figures 7(b), 9(b), and 11(b), the simulated longitudinal pack interaction forces also show reasonably good agreement with the measured data in terms of

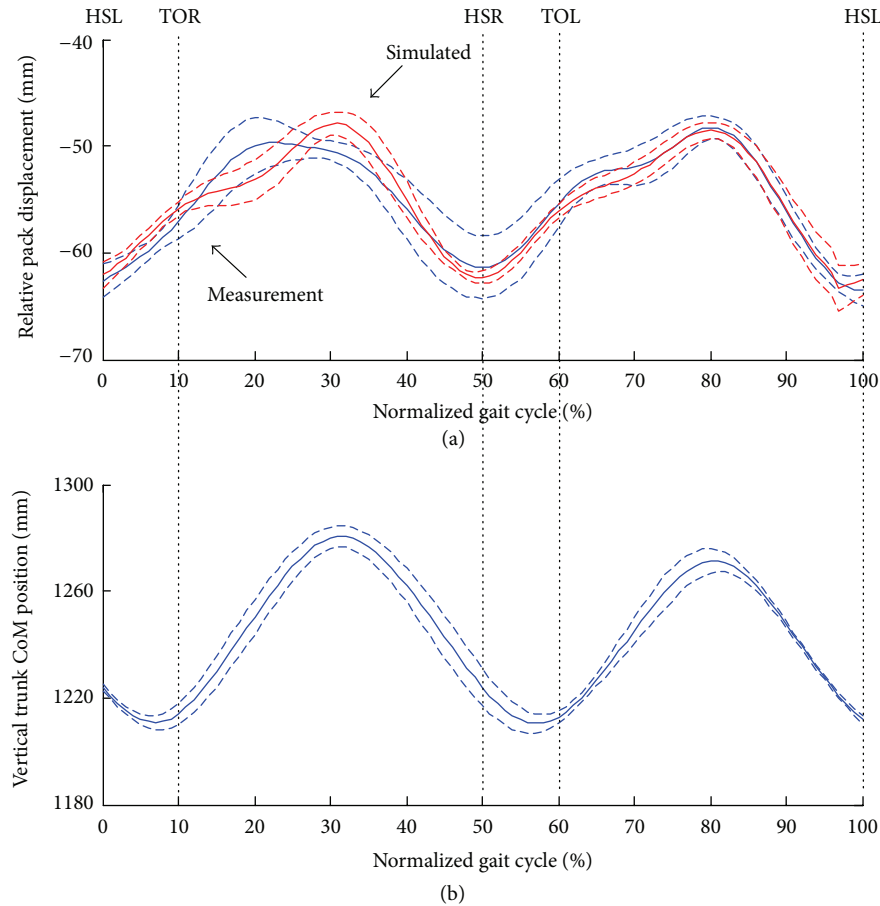


FIGURE 10: The simulated relative pack displacement along the back (mean (solid red line) \pm one standard deviation (dashed red lines), (a)), compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines), (a)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm), and measured vertical trunk CoM displacement (mean (solid blue line) \pm one standard deviation (dashed blue lines), (b)). The average walking speed was 1.62 ± 0.04 m/s, and the pack load was 23.0 kg. The stance phase for the left limb is from 0 to 60.3%, and the swing phase is from 60.3% to 100%. The double support phase is from 0 to 10.3% and from 50 to 60.3%.

amplitudes and general trends. The largest pack forces occur during double stance because of the high accelerations at heel strike and push-off. As one would expect, the pack forces reach their minima in midstance when the vertical accelerations are negative. Both peak and mean values of the longitudinal pack forces increase significantly with increased pack load, which is consistent with the results of previous experimental load carriage studies [10]. However, the simulated longitudinal pack force also exhibits an additional harmonic which is not present in the measured data.

It is likely that the additional harmonic seen in the longitudinal results is due to the limitations of the load carriage test-rig [18, 19]. Despite the fact that the mannequin was covered in a neoprene-like material to mimic the soft tissues, it is unlikely that there was sufficient compliance; resulting in a suspension model with a high natural frequency. Furthermore, neither the mannequin nor the simulation model can alter torso posture in response to changes in pack interaction forces. In practise, the bearer may dynamically adjust their posture to damp the pack motion and thereby decrease high-frequency fluctuations in pack force.

Furthermore, the test rig only generates vertical displacement of the mannequin, rather than the true 3D trunk motion seen during walking. Hence, the pack suspension model is limited to one degree of freedom and the suspension model parameters do not properly account for coupling with other motions.

Although the pack model neglects all relative pack motions except for longitudinal translation up and down the back, the simulated normal pack forces (Figures 7(a), 9(a), and 11(a)) were in reasonable agreement with the experimental data with the exception of unexplained discrepancies around 40% to 60% of the gait cycle. In percentage terms, there are major discrepancies between the model predictions and the measurement data in the lateral pack forces (Figures 7(c), 9(c), and 11(c)). However, these forces are of small amplitude and the experimental data exhibit significant variability.

In conclusion, whilst promising results have been achieved, the proposed methodology still needs some improvements especially the technique to identify pack suspension parameters [18, 19]. Although the load carriage

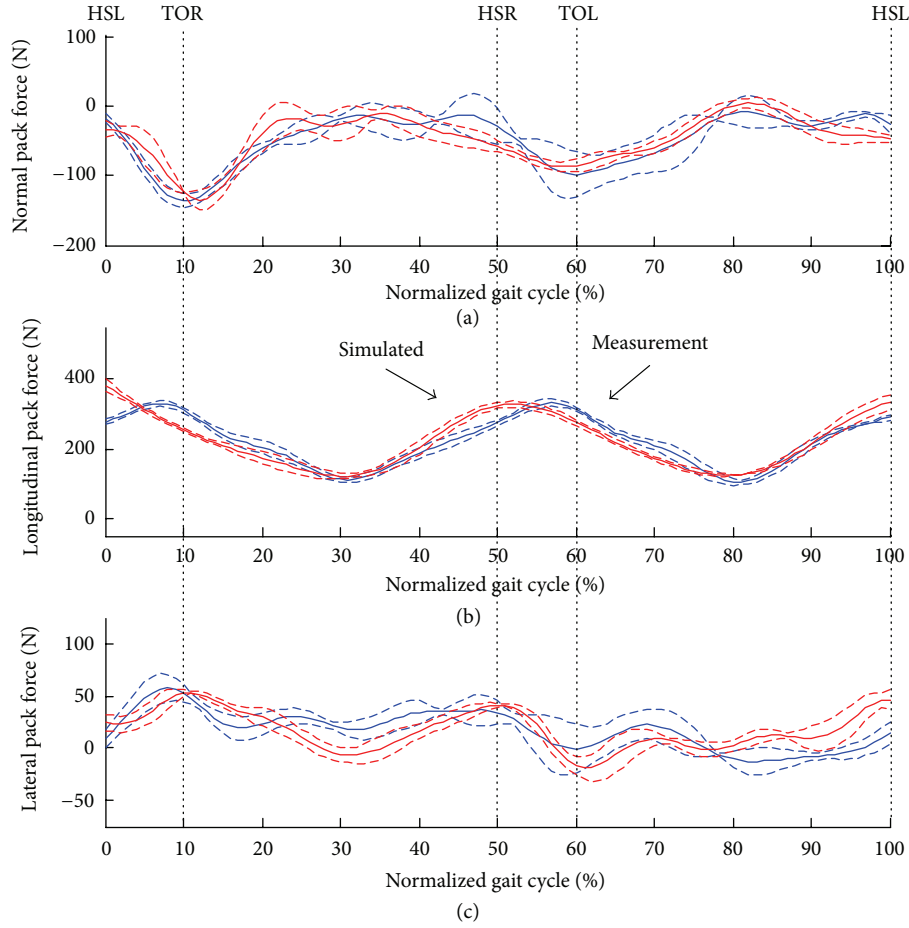


FIGURE 11: The simulated pack interaction forces in the pack coordinate system (mean (solid red line) \pm one standard deviation (dashed red lines)): (a) longitudinal pack force, (b) normal pack force (c), and lateral pack force, compared with measured data (mean (solid blue line) \pm one standard deviation (dashed blue lines)), for six repeated trials for one subject (age: 28 yrs, weight: 71.8 kg, height: 178 cm). The average walking speed was 1.62 ± 0.04 m/s, and the pack load was 23.0 kg. The stance phase for the left limb is from 0 to 60.3%, and the swing phase is from 60.3% to 100%. The double support phase is from 0 to 10.3% and from 50 to 60.3%.

test-rig has the advantage that full non-linear frequency response testing can be conducted, it does not adequately represent the effective compliance of the torso. Therefore, we suggest that future work should focus on techniques for identifying backpack suspension models by combining both load carriage test-rig data and in vivo data captured in the gait laboratory. This would have the advantage that the small but significant rotations (in the plane of the back) and lateral translations of the pack, relative to the torso, could be taken into account. Including these relative motions, in addition to the longitudinal translation, may further improve the accuracy of the predicted pack forces.

Nomenclature

\vec{a}_p : Linear acceleration vector of the pack CoM with respect to the global frame
 \vec{a}_t : Linear acceleration vector of the trunk CoM w.r.t. the global frame
 \vec{a}_{pt}^n : Centripetal acceleration vector of the pack CoM w.r.t. the trunk CoM

\vec{a}_{pt}^τ : Tangential acceleration vector of the pack CoM w.r.t. the trunk CoM
 \vec{a}_r : Relative pack acceleration vector along the back
 \vec{a}_r^c : Coriolis acceleration vector induced by relative pack motion
 $\vec{\omega}_p$: Angular velocity of the backpack
 $\vec{\alpha}_p$: Angular acceleration of the backpack
 $\vec{\omega}_t$: Angular velocity of the trunk
 $\vec{\alpha}_t$: Angular acceleration of the trunk
 \vec{u} : Relative pack translation w.r.t. the trunk
 \vec{d} : Position vector of the pack CoM w.r.t. trunk CoM in unloaded condition
 \vec{r}_{pt} : Position vector of the pack CoM relative to the trunk CoM with load
 \vec{u} : Relative velocity vector of the pack w.r.t. the trunk
 \vec{u} : Relative acceleration vector of pack w.r.t. the trunk

a_{px}, a_{py}, a_{pz} :	Linear accelerations of pack CoM in pack coordinate system
$\alpha_{px}, \alpha_{py}, \alpha_{pz}$:	Angular accelerations of pack in pack coordinate system
a_{tx}, a_{ty}, a_{tz} :	Linear accelerations of trunk CoM in pack coordinate system
$\omega_{tx}, \omega_{ty}, \omega_{tz}$:	Angular velocities of trunk in pack coordinate system
$\alpha_{tx}, \alpha_{ty}, \alpha_{tz}$:	Angular accelerations of trunk in pack coordinate system
d_x, d_y, d_z :	relative positions of pack CoM w.r.t. trunk CoM in pack coordinate system
\vec{F}_p :	Resultant pack force vector acting at the pack CoM
F_{px}, F_{py}, F_{pz} :	Pack force components in pack coordinate system
\vec{M}_p :	Resultant pack moment vector acting at the pack CoM
m_p :	Mass of the backpack
\vec{g} :	Gravitational acceleration vector
J_p :	Inertial matrix of the backpack about the pack CoM
ψ, θ, ϕ :	Backpack Euler angles w.r.t. the global coordinate system in ZXZ sequence
F_{pe} :	Elastic component of the longitudinal pack force
F_{pd} :	Damping component of the longitudinal pack force
F_{pi} :	Inertial coupling component of the longitudinal pack force
a_1, a_2, a_3 :	Elastic parameters of the pack suspension model
b_1, b_2, b_3 :	Damping parameters of the pack suspension model
c_1, c_2, c_3 :	Inertial parameters of the pack suspension model
\vec{F}_{pack} :	Resultant pack force vector acting at the trunk CoM
\vec{M}_{pack} :	Resultant pack moment vector acting about the trunk CoM
\vec{F}_{pm} :	Estimated pack force from the measurement data
\vec{M}_{pm} :	Estimated pack moment from the measurement data
\vec{a}_{pm} :	Estimated linear acceleration of the pack CoM from the measurement data
$\vec{\omega}_{pm}$:	Estimated angular velocity of the pack from the measurement data
$\vec{\alpha}_{pm}$:	Estimated angular acceleration of the pack from the measurement data.

Conflict of Interests

There is no conflict of interests.

Acknowledgments

This work has been supported by the UK Ministry of Defence and also the UK EPSRC from Grant no. EP/I033602/1. The assistance and support of Will Tutton (UK MoD) has been particularly valuable.

References

- [1] W. S. Myles and P. L. Saunders, "The physiological cost of carrying light and heavy loads," *European Journal of Applied Physiology and Occupational Physiology*, vol. 42, no. 2, pp. 125–131, 1979.
- [2] B. H. Jones, "Overuse injuries of the lower extremities associated with marching, jogging, and running: a review," *Military Medicine*, vol. 148, no. 10, pp. 783–787, 1983.
- [3] J. Bobet and R. W. Norman, "Effects of load placement on back muscle activity in load carriage," *European Journal of Applied Physiology and Occupational Physiology*, vol. 53, no. 1, pp. 71–75, 1984.
- [4] J. Knapik, E. Harman, and K. Reynolds, "Load carriage using packs: a review of physiological, biomechanical and medical aspects," *Applied Ergonomics*, vol. 27, no. 3, pp. 207–216, 1996.
- [5] J. Knapik, "Physiological, biomechanical and medical aspects of soldier load carriage," in *Proceedings of NATO RTO Meeting 56: Innovations in Load Carriage System Design and Evaluation*, Kingston, Canada, 2000, keynote paper 1.
- [6] H. Kinoshita, "Effects of different loads and carrying systems on selected biomechanical parameters describing walking gait," *Ergonomics*, vol. 28, no. 9, pp. 1347–1362, 1985.
- [7] P. E. Martin and R. C. Nelson, "The effect of carried loads on the walking patterns of men and women," *Ergonomics*, vol. 29, no. 10, pp. 1191–1202, 1986.
- [8] E. Harman, K. H. Han, P. Frykman, M. Johnson, F. Russell, and M. Rosenstein, "The effects on gait timing, kinetics and muscle activity of various loads carried on the back," *Medicine and Science in Sports and Exercise*, vol. 24, p. S129, 1992.
- [9] D. C. Tilbury-Davis and R. H. Hooper, "The kinetic and kinematic effects of increasing load carriage upon the lower limb," *Human Movement Science*, vol. 18, no. 5, pp. 693–700, 1999.
- [10] M. LaFiandra, R. C. Wagenaar, K. G. Holt, and J. P. Obusek, "How do load carriage and walking speed influence trunk coordination and stride parameters?" *Journal of Biomechanics*, vol. 36, no. 1, pp. 87–95, 2003.
- [11] G. M. U. Ghori and R. G. Luckwill, "Responses of the lower limb to load carrying in walking man," *European Journal of Applied Physiology and Occupational Physiology*, vol. 54, no. 2, pp. 145–150, 1985.
- [12] J. Knapik, K. Reynolds, J. Staab, J. A. Vogel, and B. Jones, "Injuries associated with strenuous road marching," *Military Medicine*, vol. 157, no. 2, pp. 64–67, 1992.
- [13] W. J. Wilson, "Brachial plexus palsy in basic trainees," *Military Medicine*, vol. 152, no. 10, pp. 519–522, 1987.
- [14] M. Holewijn, "Physiological strain due to load carrying," *European Journal of Applied Physiology and Occupational Physiology*, vol. 61, no. 3-4, pp. 237–245, 1990.
- [15] M. Holewijn and T. Meeuwssen, "Physiological strain during load carrying: effects of mass and type of backpack," in *Proceedings of NATO RTO Meeting 56: Innovations in Load Carriage System Design and Evaluation*, Kingston, Canada, 2000, paper 1.

- [16] J. Martin and R. Hooper, "Military load carriage: a novel method of interface pressure analysis," in *Proceedings of NATO RTO Meeting 56: Innovations in Load Carriage System Design and Evaluation*, Kingston, Canada, 2000, paper 22.
- [17] R. P. Pelot, A. Rigby, J. M. Stevenson, and J. T. Bryant, "A static biomechanical load carriage model," in *Proceedings of NATO RTO Meeting 56: Innovations in Load Carriage System Design and Evaluation*, Kingston, Canada, 2000, paper 25.
- [18] M. Gretton, *Identification of the dynamic suspension characteristics of rucksacks [Ph.D. thesis]*, The University of Salford, Manchester, UK, 2004.
- [19] M. Gretton and D. Howard, "Identifying and modeling the dynamics behaviour of load carriage systems," in *Proceedings of NATO RTO Meeting 56: Innovations in Load Carriage System Design and Evaluation*, pp. 27–29, Kingston, Canada, 2000.
- [20] L. Ren, R. K. Jones, and D. Howard, "Whole body inverse dynamics over a complete gait cycle based only on measured kinematics," *Journal of Biomechanics*, vol. 41, no. 12, pp. 2750–2759, 2008.
- [21] A. Cappozzo, F. Catani, U. Della Croce, and A. Leardini, "Position and orientation in space of bones during movement: anatomical frame definition and determination," *Clinical Biomechanics*, vol. 10, no. 4, pp. 171–178, 1995.
- [22] P. de Leva, "Adjustments to zatsiorsky-seluyanov's segment inertia parameters," *Journal of Biomechanics*, vol. 29, no. 9, pp. 1223–1230, 1996.
- [23] L. Ren, R. Jones, and D. Howard, "Generalized approach to three-dimensional marker-based motion analysis of biomechanical multi-segment systems," in *Proceedings of the International Society of Biomechanics XXth Congress*, Cleveland, Ohio, USA, 2005.
- [24] J. H. Challis, "A procedure for determining rigid body transformation parameters," *Journal of Biomechanics*, vol. 28, no. 6, pp. 733–737, 1995.

Research Article

Forming Mechanism and Correction of CT Image Artifacts Caused by the Errors of Three System Parameters

Ming Chen and Gang Li

College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Gang Li; ligangccm@163.com

Received 27 January 2013; Accepted 30 March 2013

Academic Editor: Hang Joon Jo

Copyright © 2013 M. Chen and G. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We know that three system parameters, a center of X-ray source, an isocenter, and a center of linear detectors, are very difficult to be calibrated in industrial CT system. So there are often the offset of an isocenter and the deflection of linear detectors. When still using the FBP (filtered backprojection) algorithm under this condition, CT image artifacts will happen and then can seriously affect test results. In this paper, we give the appearances and forming mechanism of these artifacts and propose the reconstruction algorithm including a deflection angle of linear detectors. The numerical experiments with simulated data have validated that our propose algorithm can correct CT images artifacts without data rebinning.

1. Introduction

We usually adopt the FBP (filtered backprojection) algorithm in industrial CT. This algorithm requires two necessary conditions [1, 2]: (i) the isoray (an imaginary ray that connects a center of X-ray source with an isocenter) is perpendicular to linear detectors; (ii) an intersection point where the isoray and linear detectors cross is the center of linear detectors. However, a center of X-ray source, an isocenter, and a center of linear detectors are difficult to be calibrated in industrial CT. When still using the FBP algorithm under the errors of three system parameters, the image artifacts will happen and then can seriously affect test results, especially for these reconstructed points away from the center of CT images.

When there happens the offset of an isocenter, people usually take the projection point of the isocenter on linear detectors as the center of linear detectors and then translate the projections. For parallel beam, this translation can recalibrate the isocenter [3]. However for fan beam, this translation is impossible unless fan beam projections are rebinned as parallel beam projections [4]. For measuring and correcting the CT system parameters, there are some methods proposed. Gullberg et al. [5] proposed the method to correct the isocenter for fan beam; however, the involved parameters are

difficult to obtain. Sun et al. [6] assumed that the plane of four small balls is perpendicular to the turn table and then measured cone-beam CT system parameters by use of projection data under one angle. For micro-CT, Patel et al. [7] proposed the autocalibration method without model measurement and measured and corrected some system parameters. For cone-beam CT, Chen et al. [8] estimated some parameters by obtaining the barycenter under the condition that the plane detector is parallel to the axis of rotation.

The remainder of this paper is organized as follows. In Section 2, we introduce the FBP algorithm for fan beam and point out its necessary conditions. In Section 3, we give the appearances of three image artifacts caused by the offset of an isocenter and the deflection of linear detectors. In Section 4, we analyze the forming mechanism of three artifacts. In Section 5, we propose the FBP algorithm including a deflection angle of linear detectors. Finally, numerical experiments and conclusions are presented in Section 6.

2. The FBP Algorithm for Fan Beam

For convenience of the formula derivation in Section 5, we introduce the FBP algorithm for equal-spaced fan beam in this section.

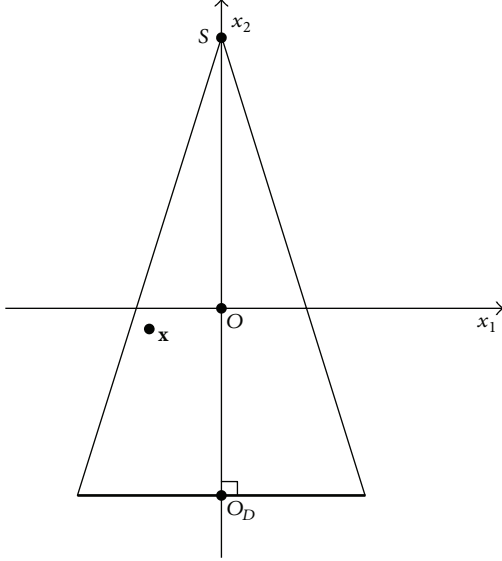


FIGURE 1: A simple geometric relationship of an equal-spaced fan beam.

A simple geometric relationship with no errors of CT system parameters is shown in Figure 1. We define a right-handed coordinate system Ox_1x_2 , where the origin O is an isocenter, x_1 axis is parallel to linear detectors (the bold line in Figure 1), and x_2 axis is parallel to the isoray. Let R_1 denote the distance from X-ray source to x_1 axis (if the isoray is perpendicular to linear detectors, R_1 is also the distance between X-ray source S and the isocenter O), and let R_2 denote the distance between S and the center O_D of linear detectors. Let $p(\beta, s)$ denote the equal-spaced fan beam projection data, where β is the angle of the isoray formed with the x_2 axis and s is a sample on the imaginary detectors which are through the isocenter O and parallel to the actual linear detectors. Making use of the FBP reconstruction algorithm, the image function, $f(\mathbf{x}) = f(x_1, x_2)$, can be shown to be

$$f(\mathbf{x}) = \int_0^{2\pi} \frac{R_1^2}{(R_1 - \mathbf{x} \cdot \beta^\perp)^2} \times \left\{ \frac{R_1}{\sqrt{R_1^2 + s^2}} p(\beta, s) * h(s) \right\} \Big|_{s=s_0} d\beta, \quad (1)$$

where s_0 is the projection address of a reconstructed point \mathbf{x} on the imaginary detectors, $\beta = (\cos \beta, \sin \beta)$, $\beta^\perp = (-\sin \beta, \cos \beta)$, $s_0 = (R_1 \mathbf{x} \cdot \beta) / (R_1 - \mathbf{x} \cdot \beta^\perp)$, and $h(s) = \int_{-\infty}^{+\infty} |\omega| e^{i2\pi\omega s} d\omega$ is a filter function.

According to a geometric relationship in Figure 1, the FBP algorithm requires two necessary conditions: (i) the isoray is perpendicular to linear detectors; (ii) an intersection point where the isoray and linear detectors cross is the center of linear detectors.

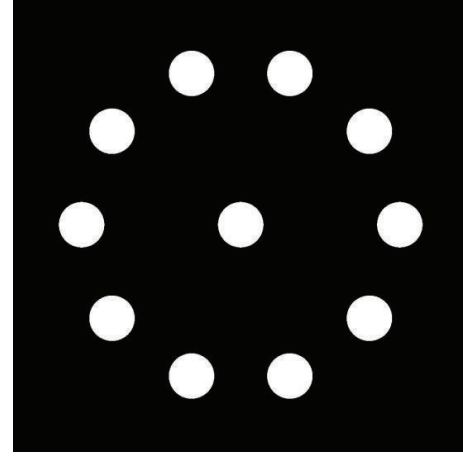


FIGURE 2: Phantom.

3. Appearances of Three Image Artifacts

We give a test phantom, which is comprised of eleven circles, and ten circles are distributed evenly among the center of the phantom, as shown in Figure 2.

The offset of an isocenter may cause CT image artifacts [9–12]. This offset is divided into two cases: along linear detectors or along the direction which is perpendicular to linear detectors. The latter is equivalent to error of R_1 . When R_1 is much larger than the field of view (FOV), we can still reconstruct a satisfying CT image, even if R_1 remains some errors [13]. For this reason, we only consider the offset of an isocenter along linear detectors.

We give a simple geometric relationship of CT scanning system with the offset of an isocenter in Figure 3, where O_1 is the isocenter. Let x_2 axis denote the direction which is through the center O_D of linear detectors and perpendicular to linear detectors. Let x_1 axis denote the direction which is through O_1 and perpendicular to x_2 axis. Let O denote the intersection point where the x_1 axis and the x_2 axis cross. Let γ denote the angle contained by the line OS and the line O_1S .

We perform numerical experiments with the simulated data to show the appearance of image artifacts caused by the offset of an isocenter. CT scanning system parameters are as follows: the distance from X-ray source S to x_1 axis $R_1 = 550.000$ mm, the distance from X-ray source S to linear detectors $R_2 = 905.000$ mm, linear detectors are composed of 1024 cells, with the size of each cell 0.4 mm. We assume that there is the offset of an isocenter $|OO_1| = 5.0$ mm. Each detector takes 720 projections in 2π . The image matrix is 1024×1024 . For the phantom in Figure 2, we reconstruct CT images using the FBP formula (1), as shown in Figure 4, where the artifacts nonuniformly spread to all directions. And the reconstructed points away from the center of CT images are comparatively worse.

For this offset of an isocenter, we may obtain the projection point of the isocenter on linear detectors by many experiments and then translate the projection data. The reconstructed images from the translated projection data are

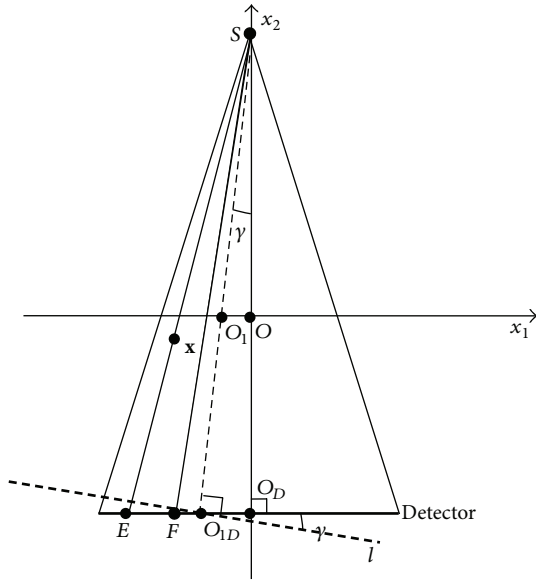


FIGURE 3: A simple geometric relationship of CT scanning system with the offset of an isocenter.

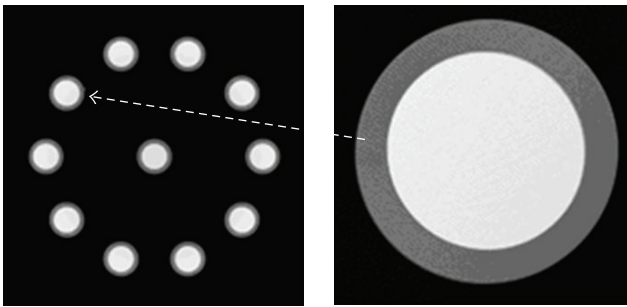


FIGURE 4: CT image artifacts caused by the offset of an isocenter.

shown in Figure 5, where the artifacts obviously reduce. In fact, an isoray is not perpendicular to linear detectors when the offset of an isocenter happens. So there still exist some image artifacts caused by the deflection of linear detectors in Figure 5. That is, linear detectors deflect to the dotted line from x_1 axis in Figure 3.

We also give a simple geometric relationship of CT scanning system with the deflection of linear detectors in Figure 6, where linear detectors deflect to the heavy continuous line from x_1 axis. Let γ denote the clockwise deflection angle. Making use of the same previous parameters, we can calculate $\gamma = 0.52^\circ$. We can reconstruct CT images using the FBP formula (1) from the projection data with the deflection of linear detectors, as shown in Figure 5, which is exactly same with the reconstructed image from the translated projections data with the offset of isocenter.

Similarly, when the offset of an isocenter and the deflection of linear detectors simultaneously happen, we also obtain CT image artifacts, as shown in Figure 7, where the isocenter offset is 2.0 mm and $\gamma = 0.52^\circ$, and the other parameters are same as above mentioned.

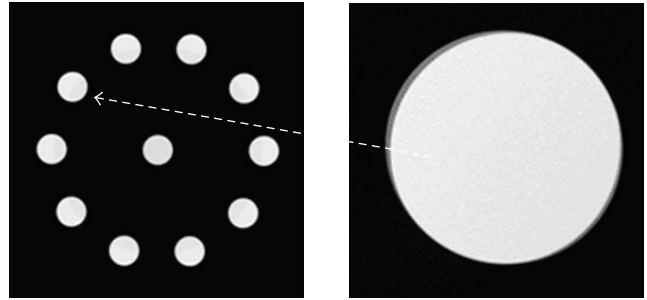


FIGURE 5: Reconstruction images from the translated projection data, or CT image artifact caused by the deflection of linear detectors.

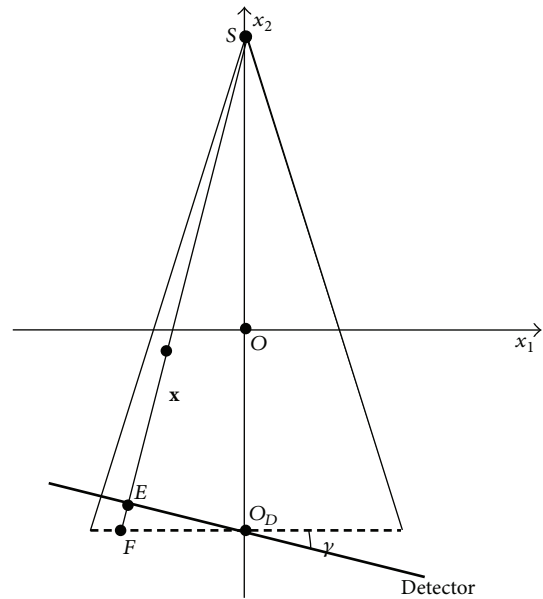


FIGURE 6: A simple geometric relationship of CT scanning system with the deflection of linear detectors.

4. Forming Mechanism of Three Image Artifacts

We give the forming mechanism of three image artifacts in this section. For a reconstructed point \mathbf{x} , we analyze a reconstruction process of \mathbf{x} and give a minimum bias expression under every projection angle.

For ease of the following analysis, let a polar coordinate (r, θ) denote \mathbf{x} , and let s_0 denote its projection address on linear detectors. If there is no error in CT system, we can calculate $s_0 = R_2 \times r \cos(\beta - \theta) / (R_1 + r \sin(\beta - \theta))$. From Figure 3, the projection point of \mathbf{x} is a point E on linear detectors, and a projection address is $s = O_{1D}E$, where O_{1D} is the projection point of the isocenter O_1 on linear detectors. However, we still take O as an isocenter in image reconstruction when using the FBP formula (1). So the other point F is regarded as the projection point of \mathbf{x} where $O_D F = O_{1D}E$. Under this condition, \mathbf{x} will be reconstructed on the line SF . Now, we draw a vertical line which is through \mathbf{x}

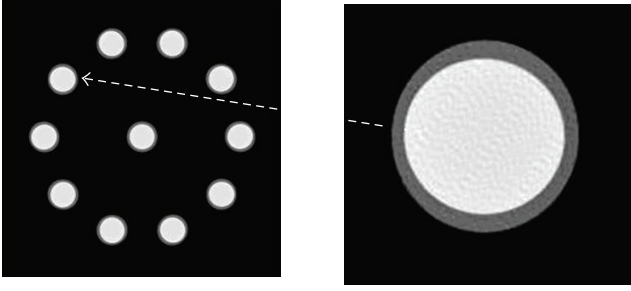


FIGURE 7: CT image artifacts caused by the offset of an isocenter and the deflection of linear detectors.

and perpendicular to the line SF , and let $M(\beta)$ denote the intersection point. The trajectory of $M(\beta)$ can approximately describe the reconstruction result of \mathbf{x} when β ranges from 0 to 2π . Now, firstly we calculate the distance $R(\beta)$ between \mathbf{x} and $M(\beta)$ as follows:

$$R(\beta) = \left| |OS|^2 \times r \cos(\beta - \theta) - s \times |O_D S| \times |OS| \right. \\ \left. - s \times |O_D S| \times r \sin(\beta - \theta) \right| \\ \times \left(\sqrt{s^2 \times |O_D S|^2 + |OS|^4} \right)^{-1}, \quad (2)$$

where $s = s_0 - (|O_D S| \times |OO_1| / |OS|)$.

So, we can obtain a coordinate of $M(\beta)$ as follows:

$$M(\beta) = \left(r \cos \theta - R(\beta) \times \cos \left(\beta + \tan^{-1} \frac{s \times |O_D S|}{|OS|^2} \right), \right. \\ \left. r \sin \theta - R(\beta) \times \sin \left(\beta + \tan^{-1} \frac{s \times |O_D S|}{|OS|^2} \right) \right). \quad (3)$$

We choose a reconstructed point $\mathbf{x}_0 = (90, \pi/4)$ and assume that the offset $|OO_1|$ of an isocenter is 0.744 mm, that is, 2.15 pixel. According to formula (3), we may draw the trajectory of $M(\beta)$ by Mathematica, where β ranges from 0 to 2π , as shown in Figure 8(a). The reconstruction image of \mathbf{x}_0 using the FBP formula (1) is shown in Figure 8(b), which explain the artifacts in Figure 4.

Similarly, for the linear detectors deflection, we may calculate the same previous expressions (2) and (3) of $R(\beta)$ and $M(\beta)$, where $s = s_0 \sin \alpha / \sin(\gamma + \alpha)$, $\alpha = \tan^{-1}(s_0 / |OS|)$. We choose $\gamma = 0.52^\circ$. The trajectory of $M(\beta)$ and the reconstruction image of \mathbf{x}_0 are as shown in Figure 9, which explain the artifacts in Figure 5.

Similarly, for the offset of an isocenter and the deflection of linear detectors, we also calculate the previous expressions (2) and (3) of $R(\beta)$ and $M(\beta)$, where $s = (s_0 \times |OS| - |O_D S| \times |OO_1|) \times \sin \alpha / |OS| \times \sin(\gamma + \alpha)$, $\alpha = \tan^{-1}(s_0 / |OS|)$. We choose the offset of an isocenter $|OO_1| = 0.5$ mm and $\gamma = 0.6^\circ$. The trajectory of $M(\beta)$ and the reconstruction image of \mathbf{x}_0 are as shown in Figure 10, which explain the artifacts in Figure 7.

5. Derivation of FBP Formula Including a Deflection Angle of Linear Detectors

In this section, we describe a new coordinate system and derive the FBP formula including a deflection angle of linear detectors, where the offset of an isocenter is attributed to the deflection of linear detector.

Referring to Figure 11, we establish the coordinate system Ox_1x_2 , where the origin O is the isocenter, x_2 axis is parallel to the isoray and points to X-ray source S , and x_1 axis and x_2 axis form right-handed coordinate system. Let φ denote the angle contained by the x_1 axis and linear detectors. Obviously, x_2 axis is not perpendicular to linear detectors, and there is a deflection of linear detectors and no offset of an isocenter in this system.

For convenience of derivation, let the polar coordinate $f(r, \theta)$ denote the image function. Let O' denote the projection point of the isocenter O on linear detectors, $R_1 = |OS|$ and $R_2 = |O'S|$. We use the imaginary detectors in formula derivation. Let d, q , and s denote three projection points of the reconstructed point $\mathbf{x} = (r, \theta)$ on linear detectors, the imaginary detectors, and x_1 axis, respectively. Let $p(d, \beta)$, $p_1(q, \beta)$, and $p_2(s, \beta)$ denote the corresponding projection data. For a reconstructed point $\mathbf{x}_0 = (r_0, \theta_0)$, and let d_0, q_0 , and s_0 denote three projection points corresponding to \mathbf{x}_0 , respectively.

From Figure 11, we can obtain the relationship between s_0 and q_0, s , and q as follows:

$$s_0 = \frac{R_1 q_0 \cos \varphi}{R_1 - q_0 \sin \varphi}, \quad (4)$$

$$s = \frac{R_1 q \cos \varphi}{R_1 - q \sin \varphi}. \quad (5)$$

Now, we rewrite the FBP formula (1) as follows:

$$f(r_0, \theta_0) = \frac{1}{2} \int_0^{2\pi} \frac{R_1^2}{(R_1 - r_0 \sin(\theta_0 - \beta))^2} \\ \times \int_{-\infty}^{\infty} \frac{R_1}{\sqrt{R_1^2 + s^2}} p_2(s, \beta) h(s_0 - s) ds d\beta, \quad (6)$$

where $h(s) = \int_{-\infty}^{\infty} |\omega| e^{i2\pi\omega s} d\omega$, $s_0 = R_1 r_0 \cos(\theta_0 - \beta) / (R_1 - r_0 \sin(\theta_0 - \beta))$.

From formula (4), (5), and (6), we may obtain

$$q_0 = \frac{R_1 r_0 \cos(\theta_0 - \beta)}{R_1 \cos \varphi - r_0 \sin(\theta_0 - \beta - \varphi)}. \quad (7)$$

From formula (5) and $h(s)$, we can obtain

$$\frac{ds}{dq} = \frac{R_1^2 \cos \varphi}{(R_1 - q \sin \varphi)^2}, \quad (8)$$

$$h(s_0 - s) = \frac{1}{C^2} h(q_0 - q),$$

where $C = R_1^2 \cos \varphi / (R_1 - q_0 \sin \varphi)(R_1 - q \sin \varphi)$.

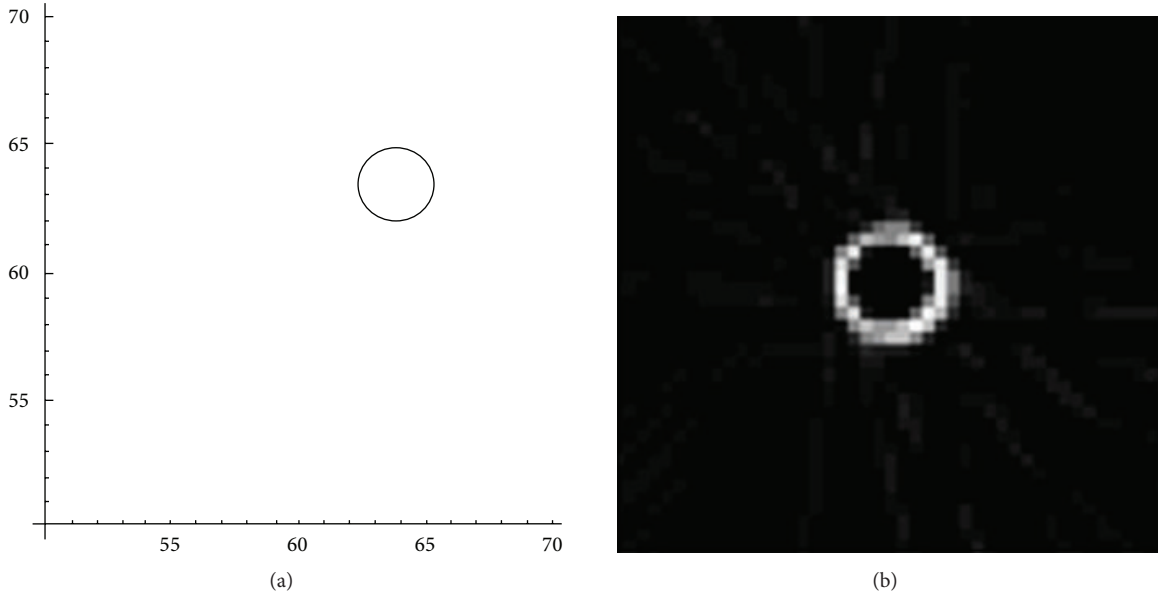


FIGURE 8: Analysis of artifacts caused by the offset of an isocenter: (a) the trajectory of $M(\beta)$; (b) the reconstruction image.

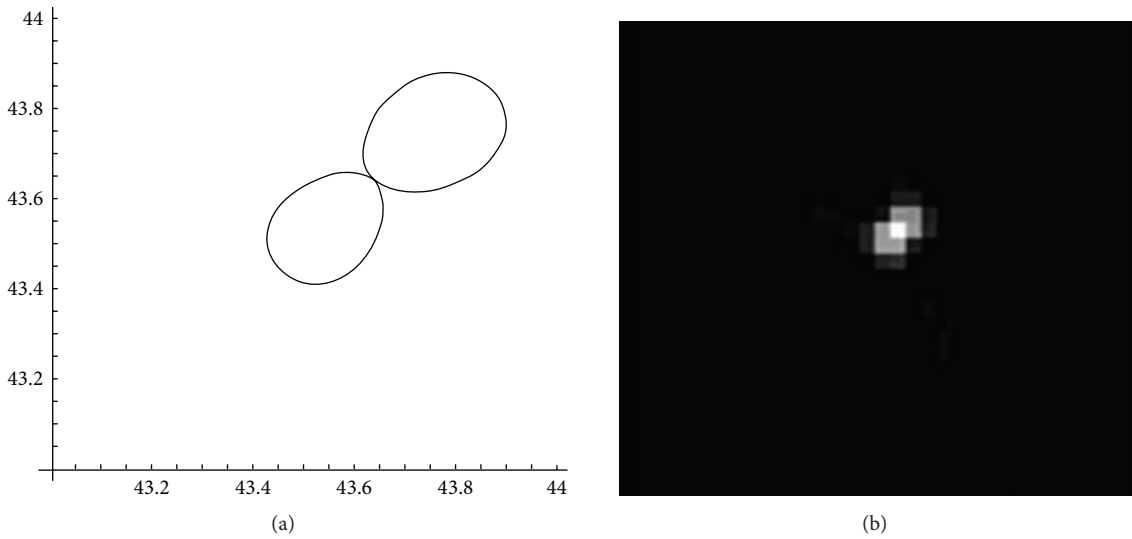


FIGURE 9: Analysis of artifacts caused by the deflection of linear detectors: (a) the trajectory of $M(\beta)$; (b) the reconstruction image.

Finally, we substitute formulae (5) and (8) into (6) and obtain after simplifying

$$\begin{aligned}
 f(r_0, \theta_0) = & \frac{1}{2} \int_0^{2\pi} \frac{(R_1 - q_0 \sin \varphi)^2}{(R_1 - r_0 \sin(\theta_0 - \beta))^2 \cos \varphi} \\
 & \times \int_{-\infty}^{\infty} \frac{R_1 - q \sin \varphi}{\sqrt{R_1^2 + q^2 - 2R_1 q \sin \varphi}} \\
 & \times p_1(q, \beta) h(q_0 - q) dq d\beta,
 \end{aligned} \tag{9}$$

where $p_1(q, \beta) = p(R_2 q / R_1, \beta)$.

The proposed previous formula can directly reconstruct CT image without data rebinning. The formula includes three parameters R_1 , R_2 , and φ , which are unknown, independence from the inspected objects, and identified by CT system. For obtaining three parameters, we have designed the model with a dense matter such as iron or steel, by a row of mutual parallel width and of the slit spacing formed. By super precise scanning for the model in 2π , we could make use of the geometric relationship of these slit spacing projection and estimate three parameters. But, this method is very sensitive to a deflection angle of linear detectors φ . We can improve measurement precision by averaging the testing values of repeated measurements.

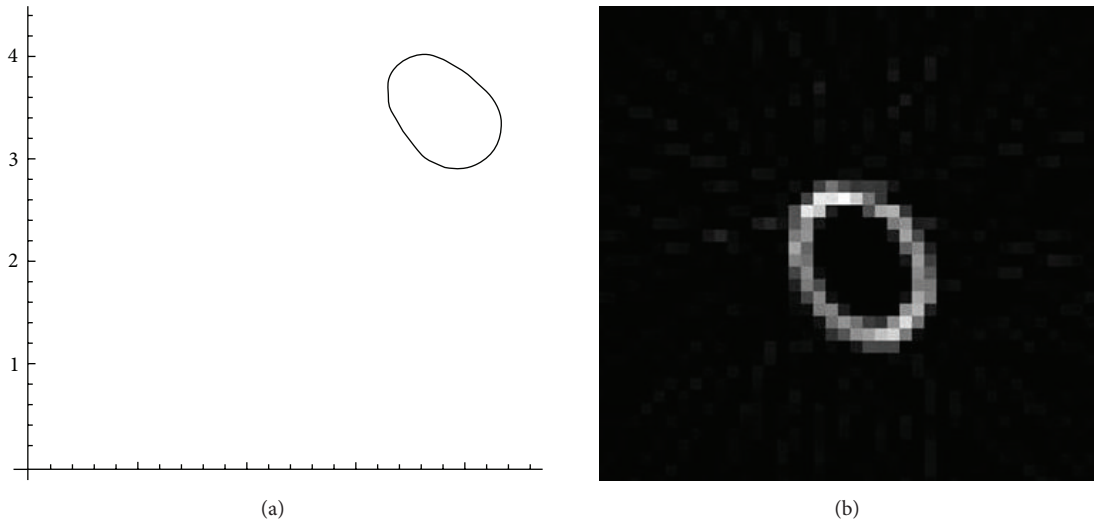


FIGURE 10: Analysis of artifacts caused by the offset of an isocenter and the deflection of linear detectors: (a) the trajectory of $M(\beta)$; (b) the reconstruction image.

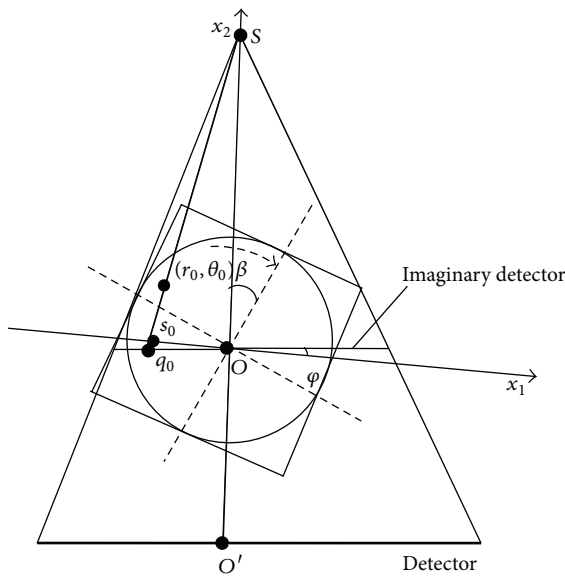


FIGURE 11: A geometric relationship of FBP formula derivation including a deflection angle of linear detectors.

6. Numerical Simulation Experiment and Conclusion

In this section we perform numerical experiments with simulated data to demonstrate our formula (9). We choose the phantom in Figure 2 and the system parameters in Figure 4. We can estimate $R_1 = 550.023$ mm, $R_2 = 905.014$ mm, and $\varphi = 0.52^\circ$ in the formula (9). The reconstruction results are shown in Figure 12 using the formula (9). Obviously, the results validate our formula, which can correct the image

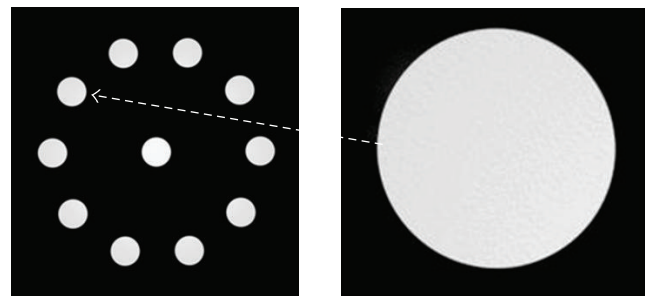


FIGURE 12: Reconstruction images using the FBP formula (9) including a deflection angle of linear detectors.

artifacts caused by the offset of an isocenter and the deflection of linear detectors.

We have given the appearances of three image artifacts caused by the offset of an isocenter and the deflection of linear detectors and analyzed the forming mechanism, which can provide reference for three artifacts identification. The correction method of the image artifacts is also proposed. Our FBP algorithm including a deflection angle of linear detectors can effectively correct three artifacts in CT images.

Acknowledgments

This work was supported in part by three Grants from the National Natural Science Foundation of China (61201430, 61002041, and 61201431), International Scientific and Technological Cooperation Program of Shenzhen (Grant JC201105190923A), China Postdoctoral Science Foundation and Shandong Province Postdoctoral Innovation Foundation.

References

- [1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, NY, USA, 1988.
- [2] B. K. P. Horn, "Fan-beam reconstruction methods," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1616–1623, 1979.
- [3] M. Dennis, R. Waggener, W. McDavid, W. Payne, and V. Sank, "Processing X-ray transmission data in CT scanning," *Optical Engineering*, vol. 16, no. 2, pp. 6–10, 1977.
- [4] P. Dreike and D. P. Boyd, "Convolution reconstruction of fan beam projections," *Computer Graphics and Image Processing*, vol. 5, no. 4, pp. 459–469, 1976.
- [5] G. T. Gullberg, C. R. Crawford, and B. M. W. Tsui, "Reconstruction algorithm for fan beam with a displaced center-of-rotation," *IEEE Transactions on Medical Imaging*, vol. MI-5, no. 1, pp. 23–29, 1986.
- [6] Y. Sun, Y. Hou, and J. Hu, "Reduction of artifacts induced by misaligned geometry in cone-beam CT," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 8, pp. 1461–1471, 2007.
- [7] V. Patel, R. N. Chityala, K. R. Hoffmann et al., "Self-calibration of a cone-beam micro-CT system," *Medical Physics*, vol. 36, no. 1, pp. 48–58, 2009.
- [8] L. Chen, Z. Wu, X. Liu, and M. Yao, "Analytical geometric parameter calibration algorithm for cone-beam CT," *Journal of Tsinghua University*, vol. 50, no. 3, pp. 418–421, 2010.
- [9] J. Li, R. J. Jaszczak, K. L. Greer, and R. E. Coleman, "A filtered backprojection algorithm for pinhole SPECT with a displaced centre of rotation," *Physics in Medicine and Biology*, vol. 39, no. 1, pp. 165–176, 1994.
- [10] J. Li, R. J. Jaszczak, H. Wang, G. T. Gullberg, K. L. Greer, and R. E. Coleman, "A cone beam SPECT reconstruction algorithm with a displaced center of rotation," *Medical Physics*, vol. 21, no. 1, pp. 145–152, 1994.
- [11] H. Wang, M. F. Smith, C. D. Stone, and R. J. Jaszczak, "Astigmatic single photon emission computed tomography imaging with a displaced center of rotation," *Medical Physics*, vol. 25, no. 8, pp. 1493–1501, 1998.
- [12] Z. B. Wang, "Effect of center deviation on CT reconstruction images," *Acta Armamentarii*, vol. 22, no. 3, pp. 323–326, 2001.
- [13] H. N. Lu, M. Yang, and L. Zhang, "A study on the reconstruction bias originating from error of focal distance of x-ray source," *Acta Armamentarii*, vol. 24, no. 1, pp. 65–67, 2003.