Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Qinyi She April 6, 2017

Statistical Assessment of the Impact of Coronary Geometry on their Functionality

by

Qinyi She

Alessandro Veneziani
Adviser

Department of Mathematics and Computer Science

Alessandro Veneziani

Adviser

Bree Ettinger

Committee Member

Yunxiao Chen

Committee Member

2017

Statistical Assessment of the Impact of Coronary Geometry on their Functionality

By

Qinyi She

Alessandro Veneziani

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Mathematics and Computer Science

2017

Abstract

Statistical Assessment of the Impact of Coronary Geometry on their Functionality
By Qinyi She

Coronary artery disease has become a prevalent cause of death globally. Studies have shown that the hemodynamics within arteries are strongly correlated with vascular geometry. This paper combines numerical analysis and statistical analysis to investigate the impact of curvatures of coronary arteries' centerline on their functionality, which is measured by Fractional Flow Reserve (FFR), on a data set of 12 patients affected by coronary artery disease. The project involves a pipeline of algorithms to process the data including image segmentation that reconstructs vascular network from image data, centerline computation that calculates the spatial coordinates of vessel centerline, free knot regression spline for curvature estimation, and correlation analysis that applies a random effects model. A negative correlation was found between estimated curvatures and FFR, though the significance level is not high enough. This insignificance in the correlation might be caused by the small data set we processed and the lack of variability elimination during preprocessing.

Statistical Assessment of the Impact of Coronary Geometry on Their Functionality

By

Qinyi She

Alessandro Veneziani

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Mathematics and Computer Science

2017

Acknowledgements

I would first like to thank my thesis advisor Dr. Alessandro Veneziani for the patient and support in overcoming numerous obstacles throughout the entire project. He provided his generous help to allow me to work in an area of my true interest and steer me in the right direction.

I would also like to thank Dr. Laura Sangalli for sharing her work in the free knot regression spline algorithm with me. The statistical analysis would not have been possible without her work.

I would also like to acknowledge Dr. Yunxiao Chen for posing questions and answering my questions for the statistical analysis in this project.

I wish to express my sincere thanks to Dr. Bree Ettinger for giving me great advice and encouragement as my academic advisor for the past three years.

Table of Contents

List of Figures

1 Introduction

In recent years, the prevalence of coronary artery disease has raised awareness in exploring the method to prevent and diagnose the disease. According to a study, coronary artery disease is the most common cause of death globally in 2013, resulting in 8.14 millions deaths (GBD 2013 Mortality and Causes of Death Collaborators 2015). The increasing cases of death due to coronary artery disease motivate the studies in coronary geometries in order to understand the pathology and hence be able to identify the disease at its early stage.

Understanding the physiological formation of coronary artery disease is essential to the disease identification. Coronary artery disease is a chronic disease that occurs when part of coronary arteries develops atherosclerosis and swollen with calcium and fatty deposits (Thiriet and Parker 2009). These deposits gradually forms a plaque. Because the plaque limits the blood flow to heart, coronary artery disease causes myocardial cell to die from lack of oxygen, which leads to severe heart muscle damage and induces myocardial infarction or heart failure.

Although groups of studies have associated the occurrence of coronary artery disease with some risk factors such as smoking, obesity, and diabetes, these risk factors are more applicable to disease prevention instead of diagnosis purposes (Torpy, Burke, and Glass 2009). Finding an efficient and accurate method to identify coronary artery disease has become a growing need. The physiology of coronary artery disease provides evidence that the disease increases the myocardial blood flow and the pressure on vessel walls. Since coronary autoregulation maintains the blood flow at a rather stable level, a hyperemic stimuli is needed to observe a difference in blood flow and thus induce useful information. Starting from 1990s, several drugs have been introduced to allow rapid change of blood flow (Johnson, Kirkeeide, and Gould 2015). Fractional Flow Reserve (FFR), for example, is a popular diagnostic tool that was first proposed in 1979 (Grüntzig, Senning, and Siegenthaler 1979). It measures the pressure differences across coronary artery stenosis to determine the functionality of the artery. FFR is defined as

$$FFR = \frac{P_d}{P_a}$$

where P_d is the pressure distal to the lesion and P_a is the pressure proximal to the lesion, with both measured during maximal hyperemia (Pijls et al. 1996). Generally, a FFR below 0.8 is a strong predictor that narrowing is causing ischemia, which justifies a surgical procedure for treating coronary artery disease.

A variety of studies have shown that the hemodynamics of arteries are strongly correlated with vascular geometry (for example Friedman et al. 1983, Caro 2009, Nerem 1984). Arteries can be described as hollow cylindrical pipes, and are characterized by features such as bends, branching, and bifurcations. Particularly, an adimensional index called *Dean number* has been used a lot to quantify fluid stability in a circular pipe. With R representing the radius of the pipe, R_{curv} being the radius of curvature of the pipe's centerline, and μ being the viscosity of the fluid, ρ being the density, and U_0 being the fluid's mean velocity, the Dean number is

$$D = \frac{2\rho U_0 R}{\mu}\left(\frac{R}{R_{curv}}\right)^{1/2}$$

The index explains the reason that estimating the vessel curvature accurately is fundamental to understanding the local hemodynamics of coronary artery. Although FFR is efficient in determining the functional severity of ischemia, the measurement procedure requires an angiogram procedure that inserts a thin guide wire into the artery to measure the flow and pressure in blood vessel after a hyperemic agent, as displayed in Figure 1. The procedure is expensive and the hyperemic agent might bring potential influences on patients. Because curvature is an important entity to understand the hemodynamics inside coronary artery, FFR and curvature might be correlated. If a statistically significant correlation can be found between FFR and curvature, taking an intravascular ultrasound will be able to provide enough information for preliminary diagnosis,

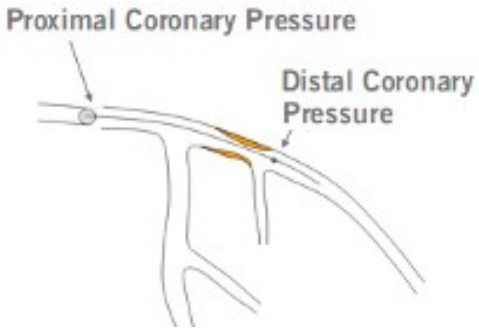


Figure 1: Illustration of a FFR measurement (Telecor n.d.).

making the diagnostic procedure much simpler and safe.

This paper, in particular, mimics the steps of ANEURISK project on coronary arteries for a preliminary proof of concepts. The ANEURISK project is a joint research program that involves researchers from several scientific fields to investigate the role of vessel morphology, blood fluid-dynamics, and biomechanical properties of vascular wall on the pathogenesis of cerebral aneurysms (Sangalli, Secchi, Vantini, and Veneziani 2009). The project collected the image data of 65 patients and then performed the following operations for each patient (see Figure 2 for an example that illustrates the first four steps of this flow of work):

1. The three-dimensional geometric representation of the carotid arteries of interest was reconstructed with level set approach that is implemented by Vascular Modeling Toolkit (VMTK). Figure 2(a)(b) shows the input and output of image segmentation.

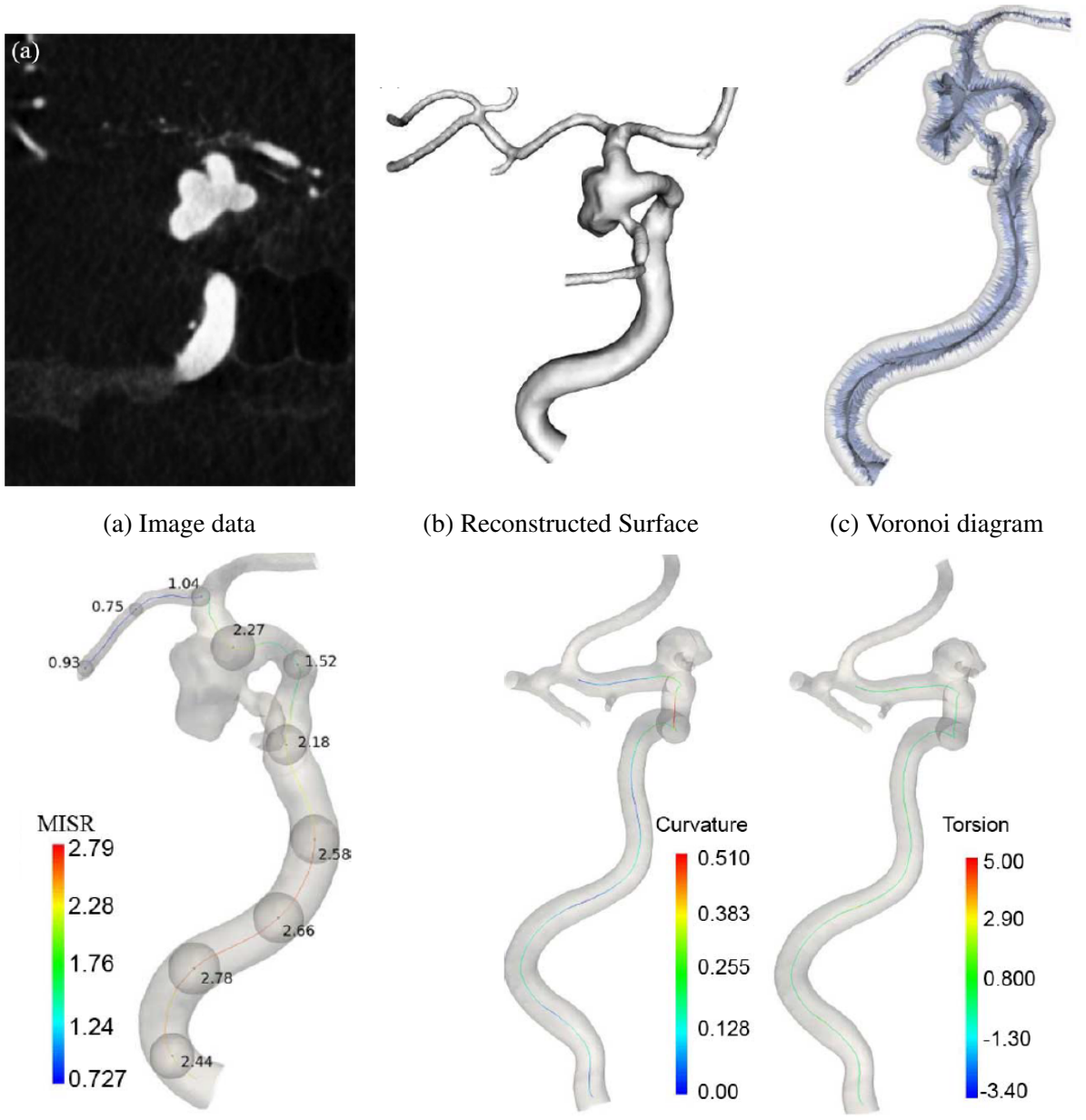
2. The centerline of the reconstructed vascular network was extracted using the concept of embedded Voronoi diagram that is also implemented by VMTK. Figure 2(b) exemplifies the application of Voronoi diagram. Figure 2(c)(d) illustrated the process from setting up the Voronoi diagram to extracting centerline and the radius of associated maximal inscribed sphere.

3. The estimated centerlines underwent a process of registration in order to remove the other variability such as size of arteries and age of patient.

4. An modified free knot regression spline algorithm fit the centerline to a spline function and then estimate its first, second, and third derivatives to compute the curvatures and torsions of the centerline. Figure 2(d) displays the estimated curvatures and torsions of the centerline color-coded with the values.

5. The estimated curvatures and torsions were then studied in computational fluid-dynamics (CFD) modeling.

6. Primary component analysis (PCA) was performed on the curvatures and torsions data and



(a) Image data (b) Reconstructed Surface (c) Voronoi diagram

(d) Extracted centerline (e) Estimated curvatures and torsions

Figure 2: Illustration of steps that were carried out in the ANEURISK project (Piccinelli et al. 2009). (a) The image data obtained from 3-D rotational angiography (RA); (b) Reconstructed surface model using level set approach; (c) Internal Voronoi diagram of the reconstructed model; (d) The centerline computed with Voronoi diagram that is color-coded with the corresponding radius of the maximal inscribed sphere; (e) Color-coded estimation of curvatures and torsions achieved from free knot regression spline.

quadratic discriminant analysis (QDA) was carried out to cluster the geometric features into the occurrence of ruptures.

Because this project is only a proof of concept, some steps that were performed in ANEURISK project are skipped in this project. This paper focuses on determining the existence of the correlation between curvatures of centerlines with FFR using a pipeline of methodologies and tools including image segmentation, centerline computation, free knot regression spline algorithm, and random effects model. In addition, we would evaluate the performance of free knot regression spline algorithm, which was developed to solve curvature estimation problem for inner carotid artery, by comparing the estimates with the result achieved from central differences.

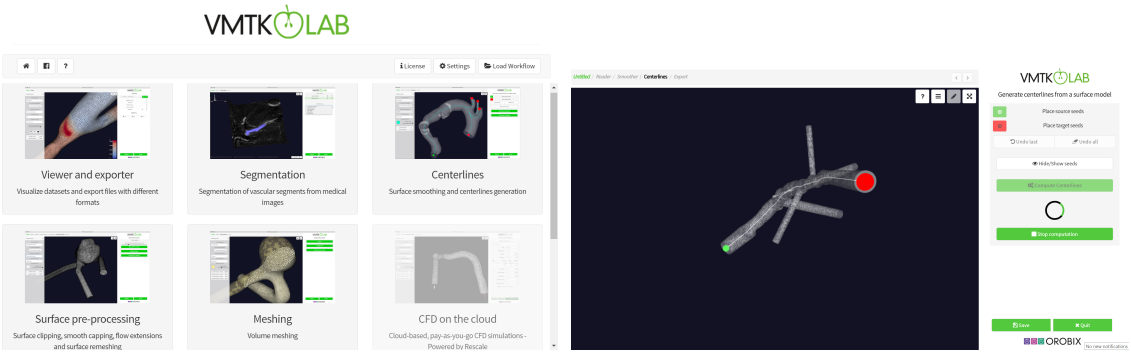
The analysis for this project is carried out in three steps. The first step of the analysis is covered in **Section 2**, which mainly discusses the reconstruction of the three dimensional geometries from imaging data and centerline computation with surface smoothing using VMTK. After retrieving the centerline from image data, the next step is related to curvature estimation by estimating the derivatives along each coordinate. In **Section 3**, the free knot regression spline algorithm is introduced and applied to the centerline in order to estimate smoothed first and second derivatives. The estimated curvatures of the coronary arteries are then derived from the estimated derivatives. After introducing the algorithms that are applied to this project, we describe how the entire process is carried out for the data we have in **Section 4**. The third step includes the correlation analysis between FFR and the estimated curvatures by fitting a random effects model, which is covered in **Section 5**. Lastly, results and discussions are drawn in **Section 6**.

2 Image Reconstruction and Processing

After gaining patients' data from intravascular ultrasound (IVUS) procedure, image segmentation is performed to reconstruct the geometric characteristics of coronary arteries. Then, the centerline of each artery is computed using the surface coordinates constructed by image segmentation. These two processing steps are both achieved by VMTK, which is an open source framework that processes for image segmentation, geometric characterization, mesh generation and computational hemodynamics specifically developed for the analysis of vascular structures (Antiga et al. 2008).

The project utilizes an interface called VMTK Lab that integrated within VMTK but is much easier to understand for front-end users. The implementation of VMTK Lab follows a two-layers approach, where the algorithms for each module are implemented with C++ while the algorithms that dynamically assembled the processing pipelines use Python (Antiga et al. 2008). Figure 3 is a screen shot of the interface VMTK Lab.

The general idea of the two methodologies, image segmentation and centerline computation, are discussed below.



(a) Menu of VMTK Lab (b) Module for centerline computation

Figure 3: Screen shot of the interface VMTK Lab that integrated within VMTK

2.1 Image Segmentation

IVUS is a clinical imaging technique that retrieves information about the anatomy of blood vessels. This methodology provides images from inside of blood vessels, visualizing the inner

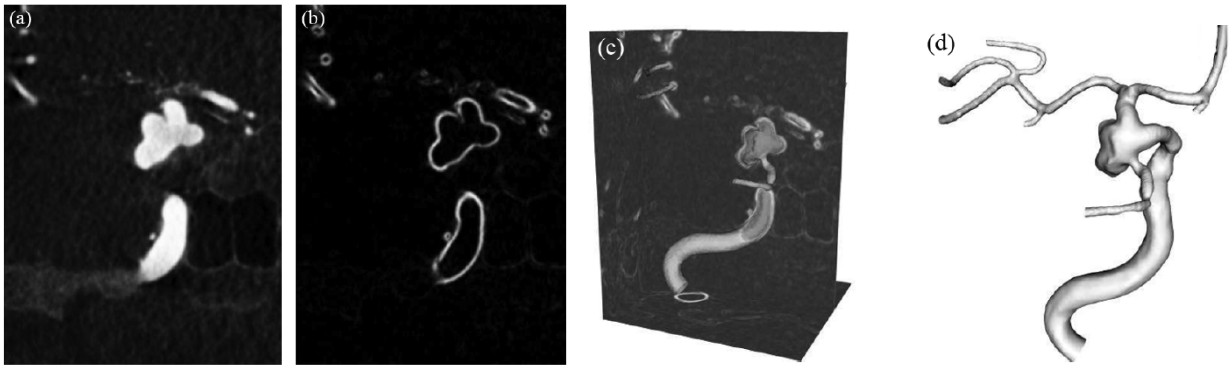


Figure 4: Example of the reconstruction with level set method. (a) Image data of an internal carotid artery bearing an aneurysm; (b) Gradient magnitude of image intensity; (c) Location of the level set model with respect to the ridges of the gradient intensity magnitude; (d) The reconstructed three-dimensional surface model.

wall of the vessels (Garcìa-Garcìa et al. 2011). IVUS is widely applied to observe the clinical significance of coronary artery disease because it provides insights to the anatomy of the plaques. The result of IVUS is represented as a 3D array of grayscale intensities $I(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^3$, where lighter pixels show presence of flowing blood and darker pixels show absence of flowing blood (Garcìa-Garcìa et al. 2011).

After retrieving the set of information from IVUS, image segmentation is performed. The goal is to reconstruct a geometric representation of the vascular network from image data and to identify the surface corresponding to the boundary between the lumen and the vascular wall (Antiga et al. 2008). Assuming that the lumen boundary exists at the locations of maximal intensity exiting lumen, the boundary is found where the magnitude of the gradient of image intensity $I(\mathbf{x})$ has a ridge. Figure 4(a) and (b) shows the identification of lumen boundaries.

The image segmentation employs implicit deformable models, which are deformable surfaces that are flexible with the topology of the segmenting vessels and are described using the level set technique. The segmentation process utilizes a less automatic identification process by focusing more on operating through the seed points chosen from the imaging data (Antiga et al. 2008). The surface of the vascular model is represented as the zero-level isosurface of $\Phi(x, t)$ that satisfies the

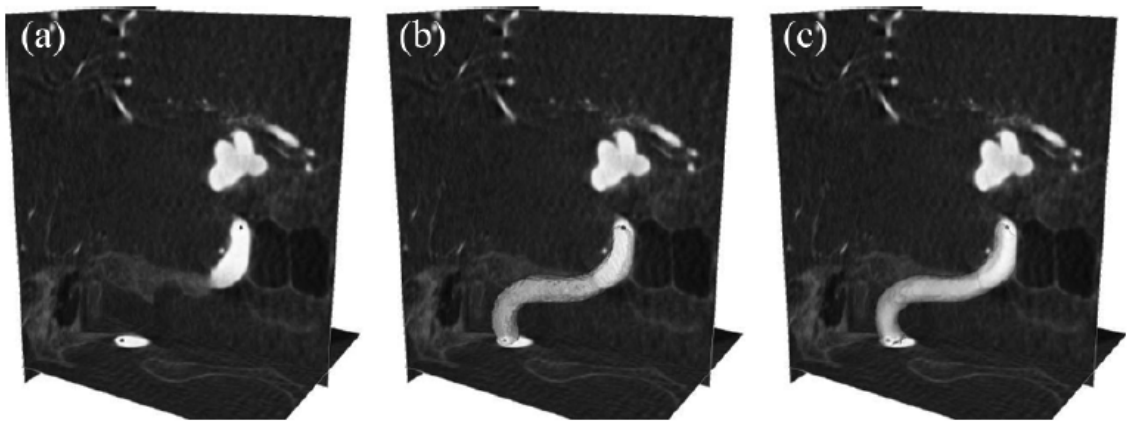


Figure 5: Level set initialization using the Colliding Fronts algorithm (Piccinelli et al. 2009). (a) Positioning of two seeds on the 3-D image for the identification of the vascular branch of interest; (b) Initial surface generated with the Colliding Fronts algorithm; (c) The final level set model of the selected vascular tract.

partial differential equation

$$\frac{\partial \Phi}{\partial t} = w_2 \nabla \cdot \left(\frac{\nabla \Phi}{|\nabla \Phi|} \right) \nabla \Phi - w_3 \nabla \cdot |\nabla I| \nabla \Phi$$

where w_2 and w_3 are user-defined parameters. The equation attracts the surface of the vascular model to the ridges of the gradient magnitude as shown in Figure 4(c).

Appropriate initial and boundary conditions needs to be found to yield more accurate results. The optimal initial condition is when the zero-level set of Φ is very close to the target surface, which is initialized using Colliding Fronts algorithm. Figure 5 demonstrates the initialization steps for a single vascular segment by selecting two seeds for the vascular branch identification. The algorithm maintains the robustness to the localization of the lumen boundary and also provides flexible choice of segmented regions at the same time(Piccinelli et al. 2009). The details of Colliding Fronts algorithm are not discussed here because it is out of range of this paper.

Because VMTK utilizes implicit representation of the surface using the level set approach, different tracts achieved from level set model can be merged successively to obtain a complete vascular network.

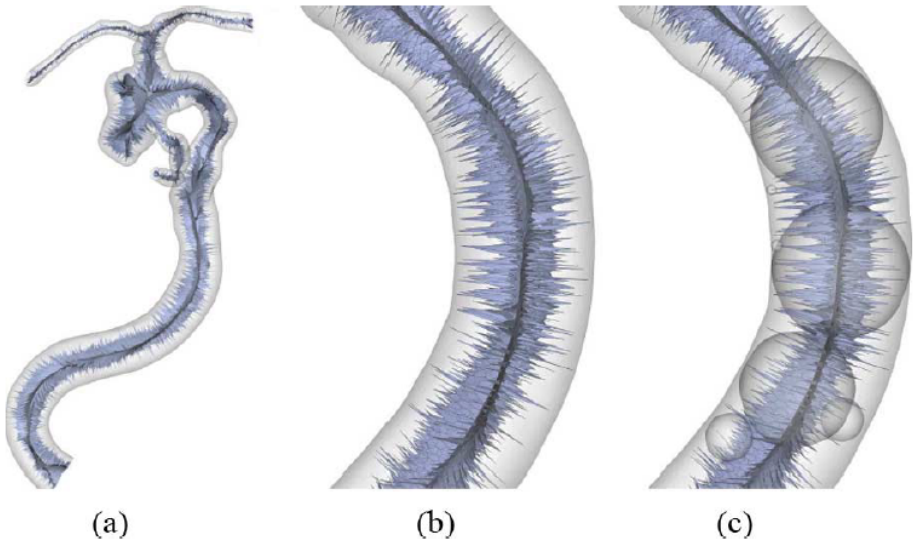


Figure 6: Demonstration of embedded Voronoi diagram (Piccinelli et al. 2009). (a) Internal Voronoi diagram of the 3-D model of an internal carotid artery bearing an aneurysm. (b) and (c) Details of the Voronoi diagram as a nonmanifold surface composed of convex polygons, whose vertices are the centers of maximal inscribed spheres; the larger spheres are associated with basic features of the geometry and the smaller spheres capture the small-scale details of the model surface.

2.2 Centerline Extraction

As explained in Section 1, the centerline of coronary artery is an important entity for exploring the geometric feature of a tubular structure. It is defined as the line between two sections of the lumen, whose minimal distance to the boundary surface is maximized (Piccinelli et al. 2009). Although coronary arteries exhibit nontubular structure at points such as bifurcation, the vascular network bifurcations can be automatically identified in VMTK framework because the centerline intersects the tube surface of another centerline at bifurcation points.

Essentially, centerline computation can be regarded as a functional minimization problem. Let s be the curvilinear abscissa along a curve $\mathbf{c}(s)$ connecting two points that corresponding to s_0 and s_1, we look for the centerline $\mathbf{c}(s)$ that minimizes the function

$$E(\mathbf{c}(s)) \equiv \int_{s_0}^{s_1} G(\mathbf{c}(s))ds$$

where $\mathbf{c}(s) \in \mathbb{R}^3$ and $G(\cdot)$ represent a cost-density function.

Selecting the function G relies on the concept of medial axis and its discrete form, Voronoi diagram. In 3D, the medial axis refers to the locus of centers of spheres maximally bounded within an object. The envelope of all maximal inscribed spheres is equivalent to the boundary surface of the object itself, which allows us to compute the centerline of the object. The medial axis is approximated using embedded Voronoi diagram, which is defined as the portion of the Voronoi diagram that is lying inside the surface. Voronoi diagram is computed using the union of the boundaries of Voronoi regions. Let $\Omega \in \mathbb{R}^3$ be a volume with boundary $\partial\Omega$, P be a set of points on $\partial\Omega$, and p be a point in P, the Voronoi region $V(p)$ is defined as the region whose points are closer to p than any other points in P, which is expressed as

$$V(p) = \{x \in \mathbb{R}^3 : ||p - x|| < ||q - x||, \forall q \in P\}$$

The removal of the non-internal part of the Voronoi diagram results in the embedded Voronoi diagram. Fig. 6(a) and (b) illustrates the internal (or also referred as embedded) Voronoi diagram for a portion of cerebral vasculature. The embedded Voronoi diagram is formed by the polygons whose vertices are the centers of the maximal inscribed sphere. The radius of the maximal inscribed sphere $R_M(\mathbf{x})$ captures basic geometry features for large values and collects small-scale details for small values. Then, the centerline equals the minimal action path on top of the embedded Voronoi diagram, where the action field G is the inverse of $R_M(\mathbf{x})$. Thus, VMTK first solve the Eikonal equation

$$|\nabla T(\mathbf{x})| = G(\mathbf{x}) = R_M(\mathbf{x})^{-1}$$

where $R_M(\mathbf{x})$ also represents the speed of wave traveling on the Voronoi diagram, and $T(\mathbf{x})$ represents the arrival time of the traveling wave. Figure 7 shows the embedded Voronoi diagram and the solutions to the Eikonal equation. The centerline is then generated by tracing the steepest path of T from each target point, which is equivalent to solving the trajectory equation

$$\frac{d\mathbf{c}}{ds} = -\nabla T$$

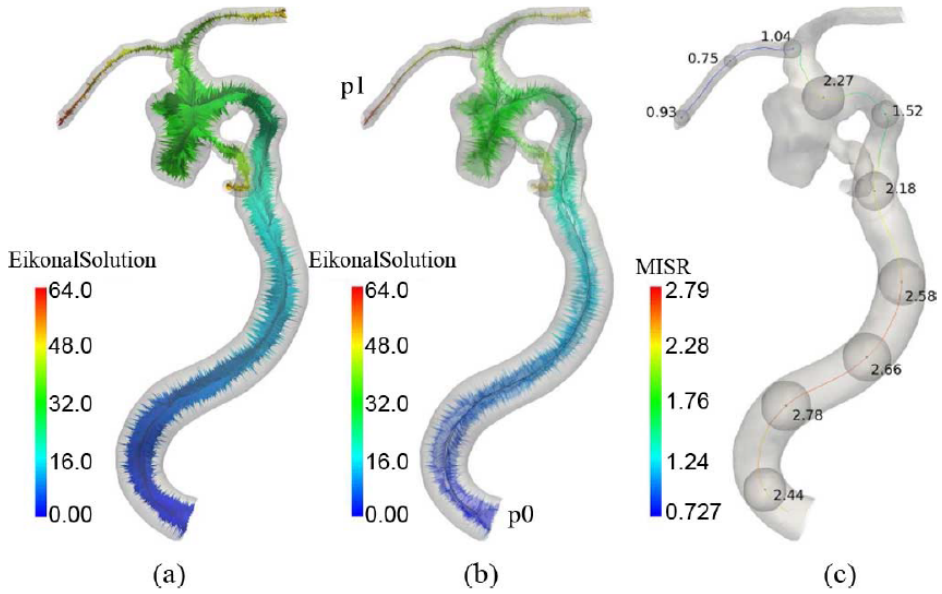


Figure 7: Illustartion of embedded Voronoi diagram with solutions to Eikonal equation (Piccinelli et al. 2009). (a) Voronoi diagram color-coded with the solution of the Eikonal equation. (b) Computation of the centerline path between the endpoints. (c) Final centerline: each point $\mathbf{x}$ is associated with $R_M(\mathbf{x})$.

where s is the curvilinear abscissa. This method results in centerlines that maximizes the minimal distance to the boundary.

3 1D Free Knot Regression Spline

Free knot regression spline algorithm has been a useful tool that applies to many numerical analysis problems. It is widely preferred because the algorithm implements an optimal knots search process that minimizes the sum of squared errors and penalizes the dimensions at the same time. In 2009, the members of ANEURISK project are the first to apply this method to analyze internal carotid artery (Sangalli, Secchi, Vantini, and Veneziani 2009). Following that, the algorithm was used to analyze thoracic aorta in 2016 (Volonghi et al. 2016). The successful application of the free knot spline algorithm to these two problems motivates the idea to apply this algorithm to coronary arteries.

The process of the application is explained below. Section 3.1 reviews the general idea of regression spline and section 3.2 introduces the knot search algorithm used in this free knot regression spline.

3.1 Regression Spline

The general idea of free knot regression spline is to find optimal knots that minimizes the sum of squared error while progressing the regression spline instead of using evenly spaced knots (Sangalli, Secchi, Vantini, Veneziani, et al. 2007). After computing the centerline for each coronary artery with VMTK as described in Section 2, the centerline is fitted to a model using free knot regression spline and then the first and second derivatives for each coordinate, which are essential to the accurate estimation of curvature, are estimated from the model.

Consider an interval $[a, b]$ with subintervals split by $k_1, k_2, \cdots, k_{n_k}$ where $a < k_1 < k_2 < \cdots < k_{n_k} < b$, an order-$m$ spline over this interval will give a piecewise polynomial of degree $m - 1$ with continuous derivatives of order $m - 2$ at the knots. The spline identifies parameters for $n_k + 1$ polynomials of degree m and for $m-1$ continuous constraints at each knot. Hence, the vector space formed by the set of all order-m spline with knot vector $\mathbf{k} = (k_1, k_2, \cdots, k_{n_k})$ has a dimension of $m(n_k + 1) + n_k(m - 1) = m + n_k$. Using the b-spline basis system $b_{r,m}^{[k]}(s) : r = 1, \cdots, m + n_k$,

the order-m spline over $[a, b]$ can be expressed as the expansion in bases

$$f(s) = \sum_{r=1}^{m+n_k} \lambda_r b_{r,m}^{[k]}(s) \quad \text{where} \quad \lambda = (\lambda_1, \cdots, \lambda_{m+n_k})^T$$

Let vector $\mathbf{v} = [a, k_1, k_2, \cdots, k_{m+n_k}, b, \cdots, b]^T$ has $2m+n_k$ entries, the first and second derivatives of the order-m spline can be expressed as $f'(s) = \sum_{r=1}^{m+n_k-1} \lambda_r^{[1]} b_{r,m-1}^{[k]}(s)$ and $f''(s) = \sum_{r=1}^{m+n_k-2} \lambda_r^{[2]} b_{r,m-2}^{[k]}(s)$ where

$$\lambda_r^{[1]} = \frac{\lambda_{r+1} - \lambda_r}{v_{m+r} - v_{r+1}} \quad \text{for} \quad r = 1, \cdots, m + n_k - 1$$

and

$$\lambda_r^{[2]} = \frac{\lambda_{r+1}^{[1]} - \lambda_r^{[1]}}{v_{m+r} - v_{r+2}} \quad \text{for} \quad r = 1, \cdots, m + n_k - 2$$

The regression spline function $\hat{f}(s)$ is the spline that minimizes the sum of squared errors

$$SSE(\lambda) = \sum_{j=1}^{n} (w_j - \sum_{r=1}^{m+n_k} \lambda_r b_{r,m}^{[k]}(s_j))^2$$

where w_j represents the true data which can be expressed as $w_j = f(s_j) + e_j$ with e_j following identical and independent distribution, $E[e_j] = 0$ and $Var[e_j] = \sigma^2$.

The vector λ that minimizes the sum of squared errors is calculated by solving the equation that sets the derivative of $SSE(\lambda)$ to zero. Let $\{B_m^{[k]}\}_{j,r} = b_{r,m}^{[k]}(s_j)$, the estimated λ can be written as

$$\hat{\lambda} = ({B_m^{[k]}}^T B_m^{[k]})^{-1} {B_m^{[k]}}^T \mathbf{w}$$

with

$$\hat{\mathbf{w}} = B_m^{[k]} ({B_m^{[k]}}^T B_m^{[k]})^{-1} {B_m^{[k]}}^T \mathbf{w} = H\mathbf{w}$$

where $H = B_m^{[k]} ({B_m^{[k]}}^T B_m^{[k]})^{-1} {B_m^{[k]}}^T$ being an orthogonal projection matrix with $tr(H) = m + n_k$.

In order to obtain smooth estimates of the first and second derivatives, order-5 spline is usually

chosen to make sure that the algorithm can use a cubic piecewise polynomial to estimate the second derivatives.

3.2 Optimal Knots Search

The free knot regression spline algorithm utilizes an algorithm that updates the total number and positions of knots to improve the quality of spline estimation. The optimal knots are selected by minimizing an error function that penalizes the dimension of the fitted model. In this case, the algorithm uses the *Stein's unbiased risk estimate* (Stein 1981)

$$pSSE(\lambda, \mathbf{k}) = \sum_{j=1}^{n}(w_j - \sum_{r=1}^{m+n_k} \lambda_r b_{r,m}^{[k]}(s_j))^2 + C\hat{\sigma}^2(m + n_k)$$

which can be rewritten in matrix form as

$$pSSE(\lambda, \mathbf{k}) = (\mathbf{w} - B_m^{[k]})^T(\mathbf{w} - B_m^{[k]}) + C\hat{\sigma}^2(m + n_k)$$

where C is the penalization constant, and $\hat{\sigma}^2$ is an estimate of errors variance.

The accuracy of the estimation depends on the choice of the penalization constant C, where a higher value of C gives a more loose model with fewer knots and higher sum of squared error. The penalization term is fixed as indicated by the classical bias/variance tradeoff. Hence, when the penalization increases, the average squared error increases while the degrees of freedom decreases (Sangalli, Secchi, Vantini, Veneziani, et al. 2007).

The algorithm of searching for the optimal knots is based on the algorithm developed by Zhou and Shen (Zhou and Shen 2001). This algorithm includes knot relocation moves and thus significantly improved the stepwise forward/backward knot selection procedures which suffered from knot confounding problems originally. For computational convenience, we search for optimal knots along the curvilinear abscissa s. Using the grid values $s_1, \cdots, s_n$ will not affect the performance of the knot searching algorithm because the grids are very fine.

We start with a set of initial knots which are evenly spaced along the grid. Then, the algorithm alternates between the two steps below.

Knot Addition Let $\mathbf{k} = (k_1, \cdots, k_{n_k})$ be the current knot vector, and λ be the corresponding estimate of the coefficient. Set $k_0 = a$ and $k_{n_k+1} = b$. We check each subinterval $[k_r, k_{r+1}]$ for $r = 0, 1, \cdots, n_k$ for possible addition of a knot. For each subinterval, a knot s_j with $k_r < s_j < k_{r+1}$ is added to the current knot vector $\mathbf{k}$ that creates a new knot vector k_j^* and a new estimate $\hat{\lambda}^* = (B_m^{[k]^T} B_m^{[k]})^{-1} B_m^{[k]^T} \mathbf{w}$ if

$$pSSE(\lambda^*, \mathbf{k}_j^*) < pSSE(\lambda, \mathbf{k}) \quad \text{and}$$

$$s_j = argmin_{s_l : k_r < s_l < k_{r+1}} pSSE(\lambda^*, \mathbf{k}_j^*)$$

In addition, if $\mathbf{k}$ is not the initial knot vector, the subinterval $[k_r, k_{r+1}]$ is checked for possible knot addition only if at least one among the neighbor knots, $k_{r-2}, k_{r-1}, \cdots, k_{r+3}$, has been added in the previous iteration.

Knot Relocation/Deletion Let $\mathbf{k} = (k_1, \cdots, k_{n_k})$ be the current knot vector. For $r = 1, \cdots, n_k$, if k_r is the neighbor of a knot added in the preceding step, then the knot is checked for potential relocation or deletion. The knot k_r is removed from the knot vector, which results in a new knot vector k_j^* and a new estimate $\hat{\lambda}^* = (B_m^{[k]^T} B_m^{[k]})^{-1} B_m^{[k]^T} \mathbf{w}$ if

$$pSSE(\lambda^*, \mathbf{k}_j^*) < pSSE(\lambda, \mathbf{k})$$

The knot is relocated to a new position s_j with $k_r < s_j < k_{r+1}$ and creates a new knot vector k_j^* and a new estimate $\hat{\lambda}^* = (B_m^{[k]^T} B_m^{[k]})^{-1} B_m^{[k]^T} \mathbf{w}$ only if

$$pSSE(\lambda^*, \mathbf{k}_j^*) < pSSE(\lambda, \mathbf{k}) \quad \text{and}$$

$$s_j = argmin_{s_l : k_r < s_l < k_{r+1}} pSSE(\lambda^*, \mathbf{k}_j^*)$$

When the knot addition does not add any additional knots for the first time, the algorithm searches all knots for relocation and deletion, sets the resulting knot vector as the initial knot vector, and then proceeds to another loop. When the knot addition fails to add any new knots for the second time, the algorithm searches all knots for relocation or deletion and then terminates.

The free knot regression spline algorithm, including the process of searching for optimal knots, have been coded in R (Sangalli, Secchi, Vantini, Veneziani, et al. 2007).

4 Data and Model

The dataset used in this project is based on the three dimensional angiographies of 12 patients that were accepted to a clinical trial from 2007 to 2010 [1]. All of the patients were experiencing coronary artery disease with different severity. Among the 12 patients' data we processed, one of them is examined to have a FFR less than 0.8, meaning that a surgical procedure is needed, while 11 patients have a FFR greater than 0.8, meaning that only conservative treatment is needed at this point. None of these patients have other severe disease that affects their cardiovascular system. Each patient underwent an angiography procedure to image the cardiovascular system using IVUS, which is processed later.

For each patient, two steps are applied as data preparation. The first step involves image reconstruction. The imaging data was processed using image segmentation implemented by VMTK to reconstruct the geometric representation of the coronary artery. This process utilizes implicit deformable models and level set method, as introduced in Section 2.1. After the reconstruction of the geometric structure for the vascular network, the second step is to compute the three-dimensional spatial coordinates of the centerline. The algorithm of centerline computation relies on the concepts of medial axis, Voronoi diagram, as well as wave propagation, which is explained in Section 2.2. The result from each step of the data preparation is exemplified in Figure 8.

After the centerline reconstruction, every patient i in the dataset $(i = 1, \cdots, 12)$ is represented as a set of points in $\mathbb{R}$, $\{(x_{ij}, y_{ij}, z_{ij}) : j = 1, \cdots, n_i\}$ where x, y, z are the spatial coordinates. The points are associated with a curvilinear abscissa set $\{(s_{ij} : j = 1, \cdots, n_i\}$ which measures the approximate distance to the terminal of coronary artery. The value of s_{ij} is defined by

$$s_{ij} - s_{ij-1} = -\sqrt{(x_{ij} - x_{ij-1})^2 + (y_{ij} - y_{ij-1})^2 + (z_{ij} - z_{ij-1})^2} \quad \text{for} \quad j = 2, \cdots, n_i$$

Figure 9 shows the plot of reconstructed three spatial coordinates of centerline for a patient 1 versus the set s_{1j}, for $j = 1, \cdots, 256$. The number of data points for each patient n_i ranges from

[1] The author of this paper does not have any access to the personal information of the patients. All the analyses in this paper are directly processed on the reconstructed three-dimensional vascular network.

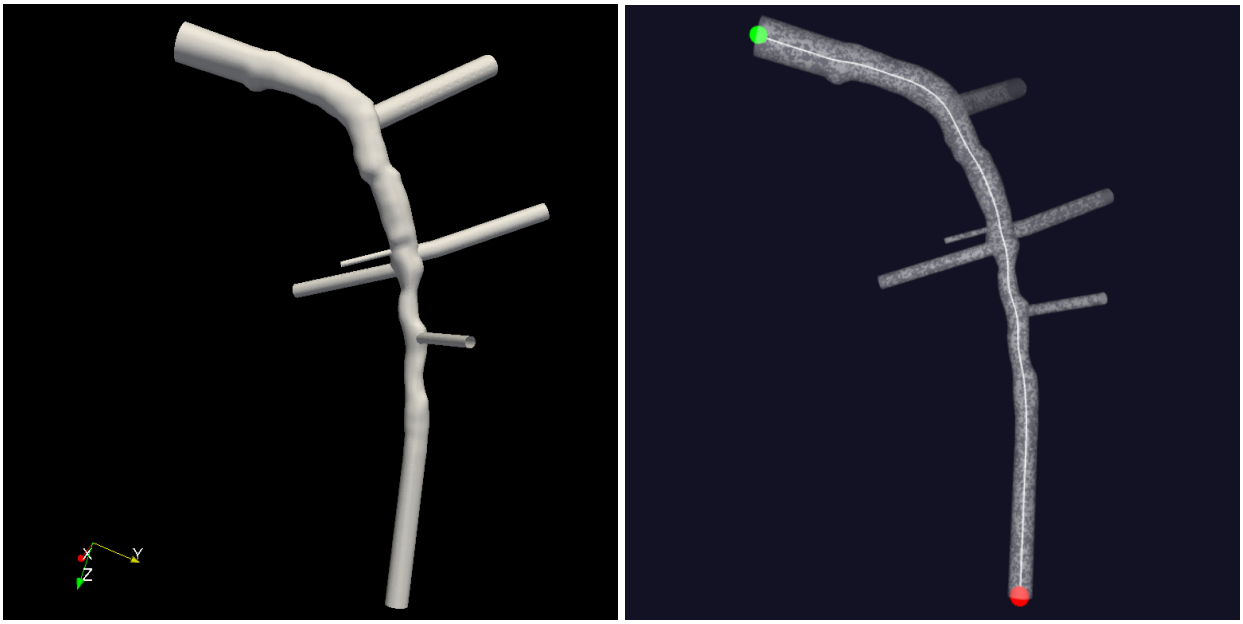


Figure 8: **Left:** Example of a reconstructed vessel using image segmentation, which is also the input for centerline computation; **Right:** Example of the resulting centerline with the transparent grid representing the surface of the reconstructed coronary artery, the white line being the centerline, and the green and red points representing the source and target seeds respectively.

204 to 385, and it has a strong correlation with the length $|s_{in_i} - s_{i1}|$ with correlation coefficient 0.845. This means that the grid density of the 12 reconstructed centerline is almost the same.

For each patient i, we represent the true centerline as $\mathbf{c}(s)$ and the centerline computed as

$$(x_j, y_j, z_j) = \mathbf{c}(s_j) + \mathbf{e}_j \quad \text{for} \quad j = 1, \cdots, n$$

where $\mathbf{e}_j = (e_j^{[x]}, e_j^{[y]}, e_j^{[z]})$ follows an independent and identical distribution with $E[\mathbf{e}_j] = 0$ and $Var[\mathbf{e}_j] = \sigma^2\mathbf{I}$. The variance of noise is assumed to be the same for each patient for the fact that the same set of tools are used to compute the centerline and hence the systematic error should equal. Hence, the curvature of the centerline is defined by

$$curv_c(s) = \frac{|\mathbf{c}'(s) \times \mathbf{c}''(s)|}{|\mathbf{c}'(s)|^3} =$$

$$\frac{\sqrt{(x'(s)y''(s) - x''(s)y'(s))^2 + (y'(s)z''(s) - y''(s)z'(s))^2 + (z'(s)x''(s) - z''(s)x'(s))^2}}{(x'(s)^2 + y'(s)^2 + z'(s)^2)^{3/2}}$$

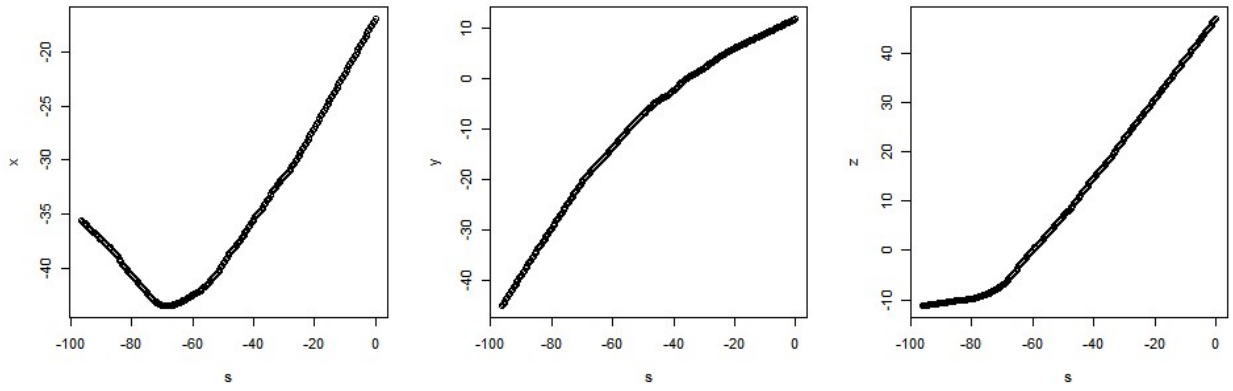


Figure 9: Example of reconstructed three spatial coordinates $x(s_1)$, $y(s_1)$, and $z(s_1)$ of the centerline for patient 1 versus the curvilinear abscissa s_{1j} for $j = 1, \cdots, 256$.

where $\mathbf{c}'(s)$ and $\mathbf{c}''(s)$ represent the first and second derivatives of the centerline, respectively. The classical approach is to use central differences to estimate the derivatives. The estimate of first derivatives for three dimensions using first central differences is given by

$$x'(s_j) = \frac{x_{j+1} - x_{j-1}}{s_{j+1} - s_{j-1}} \qquad y'(s_j) = \frac{y_{j+1} - y_{j-1}}{s_{j+1} - s_{j-1}} \qquad z'(s_j) = \frac{z_{j+1} - z_{j-1}}{s_{j+1} - s_{j-1}}$$

for $j = 2, \cdots, n-1$. Similarly, second derivatives estimated with second central differences equal

$$x''(s_j) = 2\frac{\frac{x_{j+1}-x_j}{s_{j+1}-s_j} - \frac{x_j-x_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}} \qquad y''(s_j) = 2\frac{\frac{y_{j+1}-y_j}{s_{j+1}-s_j} - \frac{y_j-y_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}} \qquad z''(s_j) = 2\frac{\frac{z_{j+1}-z_j}{s_{j+1}-s_j} - \frac{z_j-z_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}}$$

Although central differences are easy to compute, this method creates too much noise. Hence, free knot regression spline algorithm, which is covered in Section 3, is proposed here to fit the centerline to a piecewise polynomial and then to get a smoothed derivatives estimation. Instead of using evenly spaced knots as regression spline, this algorithm improves the accuracy of estimation by selecting an optimal knot vector with flexible number and positions that minimized a penalized sum of squared error.

5 Correlation Analysis

Because the geometric structure of arteries has been proved to be highly correlated with the hemodynamics within the blood vessel, our hypothesis is that estimating the curvature of coronary arteries might be a good enough indicator for preliminary diagnosis of coronary artery disease. More specifically, because of the physiological formation of plaques in coronary arteries, we expect patients with more severe coronary artery disease to have a higher curvature. So, our hypothesis is that there is a negative correlation between curvature and FFR. If there is a statistically significant correlation between curvature and FFR, then the estimated curvature can be used as a diagnostic tool for coronary artery disease. Comparing with the most popular clinical indicator of coronary artery disease, the FFR index, which requires an invasive procedure and a drug injection, estimating the curvature of coronary arteries is much easier operatively.

Due to the fine grid we used, the geometric structure of a point would be very similar to its neighbor. Hence, we cannot ignore the correlation within curvature estimated at each point for a patient. We need to consider both individual-specific random effects and group-specific random effects. To take into account of these two effects, a random effects model is applied to conduct the correlation analysis.

A random effects model is a hierarchical linear model that assumes the data is drawn from different population. The model is generally written as

$$Y = X\beta + Z\gamma + e$$

where β represents the fixed effect that we are estimating, γ represents the random effects with $E[\gamma] = 0$, $Cov(\gamma) = D$ and $Cov(\gamma, e) = 0$, and e represents the noise in the model with $Cov(e) = R$ (Christensen 2002). A least square estimate of β is

$$\hat{\beta} = \left[X'(I - M_Z)X\right]^{-} X'(I - M_Z)Y$$

where $M_Z = Z(Z'Z)^{-}Z'$. The estimate is BLUE.

Random effects model considers the individual specific effects that are uncorrelated with the independent random variable. In the problem, using a random effects model removes the effect of the internal correlation within the curvatures of a patient that is uncorrelated with FFR, and only considers the correlation between each patients' estimated curvatures and the values of FFR.

Since each patient i only has one FFR index but n_i curvatures estimated along the abscissa parameter s, a new data frame is created to match the dimensions of curvatures and FFR. We rearrange the data into three columns, where the first column represents the patients' ID, the second column represents the estimated curvature and the third column is the corresponding FFR. Each patient i takes n_i rows where first column is a vector filled with value i, the second column records the estimated curvatures of patient i, and the third column is also a vector but filled with FFR_i. The resulting data frame groups the data with same patient ID as a different population. Then, the data is fitted to the following model to estimate the fixed effect β (Bates et al. 2015),

$$curv_{ij} = \beta_1 * FFR_i + \beta_0 + \alpha_i + e_{ij} \quad \text{for} \quad i = 1, \cdots, 12 \quad \text{and} \quad j = 1, \cdots, n_i$$

where β represents the fixed effect, α_i represent the patient-specific random effect which is the same for each patient i with $E[\alpha_i] = 0$ and $Var[\alpha_i] = \sigma_\alpha^2$, and e is the noise in the model that follows identical and independent distribution with $E[e_{ij}] = 0$ and $Var[e_{ij}] = \sigma_e^2$.

The algorithm that finds the parameters of fitted random effects model is implemented in R package *lme4*. After the estimated fixed effect β is calculated, we look at the model to determine whether the correlation between the curvature and FFR is statistically significant.

6 Results and Discussion

After reconstructing the centerline from the imaging data with VMTK, the 12 centerlines are fitted to a spline function using the 1D free knot regression spline algorithm of order $m = 5$ and penalization constant $C = 4$. An order-5 spline is chosen to guarantee smooth estimates till the second derivatives. $C = 4$ is chosen as the penalization constant because this value allows the free knot spline to capture the features of centerlines while avoid overfitting (Sangalli, Secchi, Vantini, Veneziani, et al. 2007).

Since our data is 3D, the idea is to fit each coordinate of the centerline to a spline function using this algorithm, calculate the first and second derivative from the estimated spline function, and eventually determine the curvature of centerline along the abscissa parameter s.

Figure 10 shows an example of the estimation result for patient 1. The estimates of $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$ as well as their first and second derivatives are plotted versus the curvilinear abscissa s. The vertical lines in the top line show the positions of the optimal knots on s. The estimates of $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$ are superimposed to the original plot of $x(s), y(s)$, and $z(s)$ in black; and the estimates of the first and second derivatives of the three spatial coordinates, which are displayed in middle and bottom rows of Figure 10 respectively, are superimposed to the estimations achieved by first and second central differences that are plotted in black. Figure 11 shows the corresponding estimated curvatures of patient 1, superimposed to the estimated curvatures using central differences.

In the example, the free knot regression spline algorithm gives a really good fit for $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$, since the original curves are almost covered by the estimated values. Also, the algorithm produces more smoothed estimates for first and second derivatives of three spatial coordinates as well as the curvatures of centerlines. The smoothed estimates better detect significant peaks and troughs, which are extremely important for understanding the interactions between curvature landmarks and hemodynamics (Sangalli, Secchi, Vantini, Veneziani, et al. 2007). In addition, according to the comparative simulations carried out by Gervini, free knot regression spline algorithm has a advantage over other smoothing methods in detecting peaks and troughs (Gervini 2006).

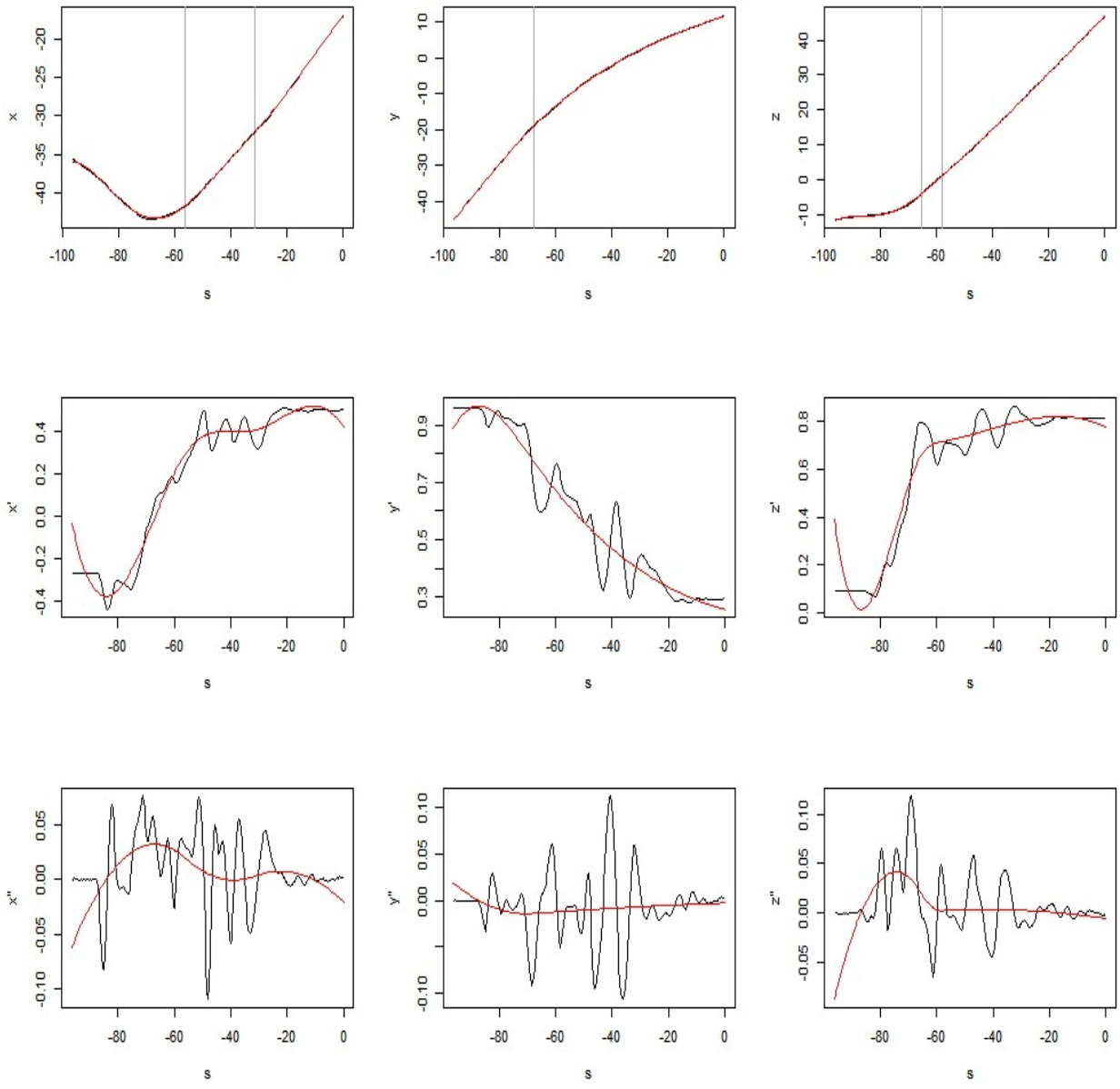


Figure 10: Example of the estimation result of patient 1. **Top:** Fitted curves $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$ in red versus abscissa parameter s with vertical lines representing the positions of optimal knots, superimposed to the plot of original $x(s), y(s)$, and $z(s)$ in black; **Middle:** First derivatives of $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$ superimposed to the first central differences in black; **Bottom:** Second derivatives of $\hat{x}(s), \hat{y}(s)$, and $\hat{z}(s)$ superimposed to the second central differences in black.

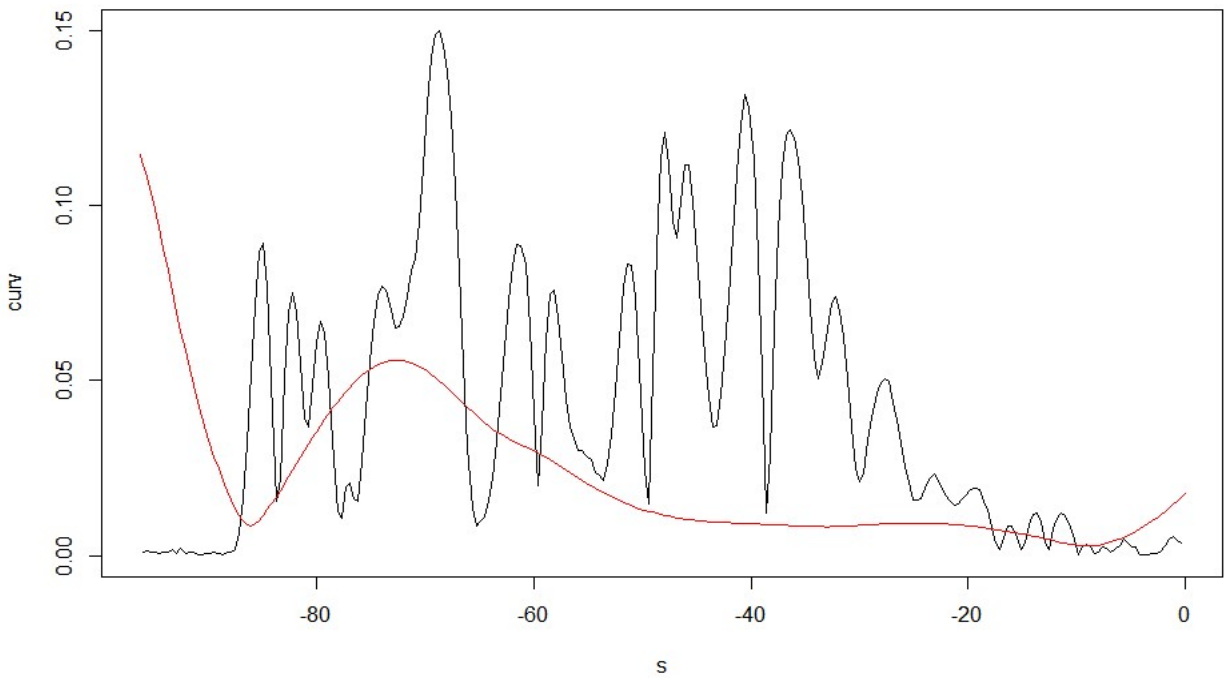


Figure 11: Example of the estimated curvatures of patient 1 superimposed to the estimated curvatures achieved by central differences in black.

After estimating the curvature of centerlines, we evaluate the correlation between curvature and FFR. For preliminary analysis, we first look at the relationship between mean curvature of each patient and corresponding FFR. The correlation between mean curvatures and FFR equals -0.234, which exhibits a negative correlation as we expected, but not significant enough. Figure 12 is a scatterplot of FFR versus corresponding mean curvatures. The red line in Figure 12 is the result of the fitted linear regression model

$$FFR_i = -1.341 * mean(curvatures)_i + 0.941$$

Although the coefficient also shows a negative correlation, the associated p value of the coefficient equals 0.456, suggesting the estimated coefficient is not statistically significant.

To take in to account both individual specific effect and group specific effect when evaluating the statistical significance of the correlation, we fit the data to a random effects model. As explained in Section 5, the curvatures of 12 patients and FFR are integrated into a new data frame with 3483

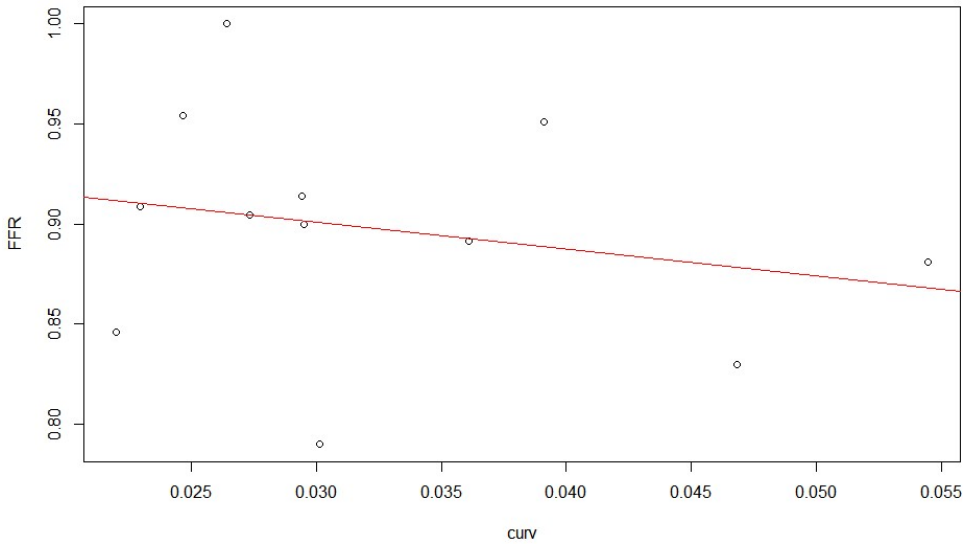


Figure 12: Scatterplot of FFR versus the mean curvatures of the patient. The overlaid red line is the result of fitted linear regression model $FFR = -1.341 * \bar{curvature} + 0.941$.

rows, which represents the total number of curvatures estimated for 12 patients, and 3 columns including patients' ID $i = 1, \cdots, 12$, curvatures, and the FFR value corresponding to each patient. Then, the data frame is fitted to a random effects model and the fixed effect β, which is the slope in the model, equals -0.04056 with standard error 0.05362. The test statistics of the slope gives a p-value 0.22 (t value = -0.756 with df = 3841). Although the result did indicate that there is a negative correlation between curvatures and FFR, the high p-value indicates that in this dataset the correlation is not significant.

Although the result rejects our hypothesis that there is a significant correlation between curvatures and FFR, there are some limitations in this project that might lead to this result. First of all, the amount of patients' data available is very limited. With only twelve patients, especially only one of them has a FFR index less than 0.8, we can expect that noises have a greater influence on the fitted random effects model. Also, due to the fact that each patient only have one FFR but different curvature values on hundreds of points along coronary artery, we should expect the correlation between each curvature and FFR become less significant.

To improve the methodology of this project and to infer more useful information, we might want to apply this pipeline of methodologies to a larger data set to reduce the influence of noises.

We will also need to process the data more carefully by, for example, applying registration filters to remove other variability such as age. Also, we might apply to our data set with a PCA for dimension reduction and a QDA for clustering as ANEURISK did.

In addition, the torsion of the centerline may also be a field of study in the future. Torsion measures how the curve is twisting out of the plane of curvature, which is estimated using third derivatives.

$$\tau(s) = \frac{|\mathbf{c}'(s) \times \mathbf{c}''(s)| \cdot \mathbf{c}'''(s)}{||\mathbf{c}'(s) \times \mathbf{c}''(s)||^2}$$

$$= \frac{\sqrt{x'''(y'z'' - y''z') + y'''(x'z'' - x''z') + z'''(x'y'' - x''y')}}{(y'z'' - y''z')^2 + (x'z'' - x''z')^2 + (x'y'' - x''y')^2}$$

where **c** represents the centerline function of coronary artery. The hemodynamics is influenced not only by vessel radius and curvatures, but also by its torsion, although in a still unclear way (Sangalli, Secchi, Vantini, Veneziani, et al. 2007). Hence, integrating the estimated torsions into our model might improve the predictability of the model. Another potential fields of study is to determine the significance of correlation of torsion and FFR and thus understand deeper about the interaction between geometric characteristics and hemodynamics as well as functionality of coronary arteries.

References

Antiga, Luca et al. “An image-based modeling framework for patient-specific computational hemodynamics”. In: *Medical & Biological Engineering & Computing* 46.11 (2008), p. 1097. ISSN: 1741-0444. DOI: 10.1007/s11517-008-0420-1.

Bates, Douglas et al. “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.i01 (2015). URL: https://ideas.repec.org/a/jss/jstsof/v067i01.html.

Caro, C.G. “Discovery of the Role of Wall Shear in Atherosclerosis”. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 29.2 (2009), pp. 158–161. ISSN: 1079-5642. DOI: 10.1161/ATVBAHA.108.166736. eprint: http://atvb.ahajournals.org/content/29/2/158.full.pdf. URL: http://atvb.ahajournals.org/content/29/2/158.

Christensen, Ronald. “Plane Answers to Complex Questions: The Theory of Linear Models”. In: Springer Texts in Statistics (2002).

Friedman, Morton H. et al. “Arterial geometry affects hemodynamics”. In: *Atherosclerosis* 46.2 (1983), pp. 225–231. ISSN: 0021-9150. DOI: http://dx.doi.org/10.1016/0021-9150(83)90113-2. URL: http://www.sciencedirect.com/science/article/pii/0021915083901132.

Garcìa-Garcìa, Hector M. et al. “IVUS-based imaging modalities for tissue characterization: similarities and differences”. In: *The International Journal of Cardiovascular Imaging* 27.2 (2011), pp. 215–224. ISSN: 1573-0743. DOI: 10.1007/s10554-010-9789-7.

GBD 2013 Mortality and Causes of Death Collaborators. “Global, Regional, and National Age-Sex Specific All-Cause and Cause-Specific Mortality for 240 Causes of Death, 1990-2013: A Systematic Analysis for the Global Burden of Disease Study 2013”. In: *Lancet* 385.9963 (2015), pp. 117–171. DOI: 10.1016/S0140-6736(14)61682-2.

Gervini, Daniel. “Free-knot spline smoothing for functional data”. In: *Journal of the Royal Statistical Society Series B* 68.4 (2006), pp. 671–687. URL: https://ideas.repec.org/a/bla/jorssb/v68y2006i4p671-687.html.

Grüntzig, Andreas R., Åke Senning, and Walter E. Siegenthaler. “Nonoperative Dilatation of Coronary-Artery Stenosis”. In: *New England Journal of Medicine* 301.2 (1979). PMID: 449946, pp. 61–68. DOI: 10.1056/NEJM197907123010201.

Johnson, Nils P., Richard L. Kirkeeide, and K. Lance Gould. “History and Development of Coronary Flow Reserve and Fractional Flow Reserve for Clinical Applications”. In: *Interventional Cardiology Clinics* 4.4 (Oct. 2015), pp. 397–410. ISSN: 2211-7458. DOI: 10.1016/j.iccl.2015.06.001.

Nerem, RM. “Atherogenesis: hemodynamics, vascular geometry, and the endothelium”. In: *Biorheology* 21.4 (1984), pp. 565–569. ISSN: 0006-355X. URL: http://europepmc.org/abstract/MED/6487767.

Piccinelli, Marina et al. “A framework for geometric analysis of vascular structures: application to cerebral aneurysms”. In: *IEEE transactions on medical imaging* 28.8 (2009), pp. 1141–1155.

Pijls, Nico H.J. et al. “Measurement of Fractional Flow Reserve to Assess the Functional Severity of Coronary-Artery Stenoses”. In: *New England Journal of Medicine* 334.26 (1996). PMID: 8637515, pp. 1703–1708. DOI: 10.1056/NEJM199606273342604. eprint: http://dx.doi.org/10.1056/NEJM199606273342604. URL: http://dx.doi.org/10.1056/NEJM199606273342604.

Sangalli, Laura M., Piercesare Secchi, Simone Vantini, and Alessandro Veneziani. “A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery”. In: *Journal of the American Statistical Association* 104.485 (2009), pp. 37–48. DOI: 10.1198/jasa.2009.0002.

Sangalli, Laura M., Piercesare Secchi, Simone Vantini, Alessandro Veneziani, et al. “Efficient estimation of 3-dimensional centerlines of inner carotid arteries and their curvature functions by free knot regression splines”. In: (2007).

Stein, Charles M. “Estimation of the Mean of a Multivariate Normal Distribution”. In: *The Annals of Statistics* 9.6 (1981), pp. 1135–1151. ISSN: 00905364. URL: http://www.jstor.org/stable/2240405.

Telecor, Biomedica. “Fast and Easy FFR – Outstanding Handling Performance”. In: (). URL: http://www.telecor.com.br/materiais_hospitalares_radianalyzer_eua.asp.

Thiriet, Marc and Kim H. Parker. “Physiology and pathology of the cardiovascular system: A physical perspective”. In: (2009). Ed. by Luca Formaggia, Alfio Quarteroni, and Alessandro Veneziani, pp. 1–45. DOI: 10.1007/978-88-470-1152-6_1.

Torpy, Janet M, Alison E Burke, and Richard M Glass. “Coronary heart disease risk factors”. In: *Jama* 302.21 (2009), pp. 2388–2388.

Volonghi, Paola et al. “Automatic extraction of three-dimensional thoracic aorta geometric model from phase contrast MRI for morphometric and hemodynamic characterization”. In: *Magnetic Resonance in Medicine* 75.2 (2016), pp. 873–882. ISSN: 1522-2594. DOI: 10.1002/mrm.25630. URL: http://dx.doi.org/10.1002/mrm.25630.

Zhou, Shanggang and Xiaotong Shen. “Spatially Adaptive Regression Splines and Accurate Knot Selection Schemes”. In: *Journal of the American Statistical Association* 96.453 (2001), pp. 247–259. DOI: 10.1198/016214501750332820.