Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 297 (2015) 167–190
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Impact of geometric uncertainty on hemodynamic simulations using machine learning

Sethuraman Sankaran[a,*], Leo Grady[b], Charles A. Taylor[c,d]

[a] HeartFlow Inc., 1400, Seaport Blvd, Building B, Redwood City, CA 94063, USA
[b] R& D, HeartFlow Inc., 1400, Seaport Blvd, Building B, Redwood City, CA 94063, USA
[c] CTO, HeartFlow Inc., 1400, Seaport Blvd, Building B, Redwood City, CA 94063, USA
[d] Stanford University, 443 Via Ortega, Stanford, CA 94305, USA

Received 3 February 2015; received in revised form 27 August 2015; accepted 30 August 2015
Available online 15 September 2015

Abstract

In the cardiovascular system, blood flow rates, blood velocities and blood pressures can be modeled using the Navier–Stokes equations. Inputs to the system are typically uncertain, such as (a) the geometry of the arterial tree, (b) clinically measured blood pressure and viscosity, (c) boundary resistances, among others. Due to a large number of such parameters, efficient quantification of uncertainty in solution fields in this multi-parameter space is challenging. We use an adaptive stochastic collocation method to quantify the impact of uncertainty in geometry in patient-specific models. We develop a novel subdivision method to define the stochastic space of geometries. To accelerate convergence and make the problem tractable, we use a machine learning approach to approximate the simulation-based solution. Towards this, a reduced order model of the Navier–Stokes equations is developed using a segmental resistance analog boundary conditions (ratio of pressure to flow). Using an offline database of pre-computed solutions, we compute a map (rule) from the features to solution fields. We achieve significant speed-up (of a few orders of magnitude) by approximating the simulation-based solution using a machine learning predictor. A bootstrap aggregated decision tree was found to be the best predictor among many candidate regressors (correlation coefficient of training set was 0.94). We demonstrate stochastic space convergence using the adaptive stochastic collocation method, and also show robustness to the choice of geometry parameterization. The sensitivities to geometry obtained using machine learning had a correlation coefficient of 0.92 with the values obtained using finite element simulations. Segments with significant disease in the larger arteries had the highest sensitivities. Terminal segments are more sensitive to dilation and proximal healthy segments are more sensitive to erosion. Sensitivity to geometry is highest when geometric resistance is comparable to net downstream resistance.

Keywords: Sensitivity analysis; Hemodynamics; Machine learning; Stochastic collocation; Geometric uncertainty

* Corresponding author. Tel.: +1 607 2274154; fax: +1 650 368 2564.
E-mail addresses: ssankaran@heartflow.com (S. Sankaran), lgrady@heartflow.com (L. Grady), ctaylor@heartflow.com, taylorca@stanford.edu (C.A. Taylor).

http://dx.doi.org/10.1016/j.cma.2015.08.014
0045-7825/

1. Introduction

Computer simulations are becoming increasingly powerful for understanding mechanisms of disease, diagnosing disease and devising treatment strategies in the human circulatory system [1–3]. Such simulations can be chiefly classified as hemodynamic simulations to compute spatio-temporal evolution of blood flow rate and pressure, fluid–structure interaction to capture wall motion and wall stresses [4], and growth and remodeling simulations [5–7] to predict long term evolution of arterial properties in health and disease. There has been a significant increase in the clinical applicability of computational tools, owing to sophistication in image acquisition, better understanding of boundary conditions, interaction between wall motion and fluid flow, and arterial remodeling mechanisms [8]. Such tools have shown significant promise in assessing the functional significance of coronary artery disease [9–12], prediction of atherosclerosis [13], aneurysm growth [6,7], failure mechanism in bypass grafts [14], outcome of stenting, outcome of pediatric surgeries [15–17], etc. In this paper, we restrict attention to hemodynamic simulations.

Uncertainties that arise in hemodynamic simulations include (a) flow rate and pressure at inlets/outlets to the model, (b) lumped parameter boundary conditions such as resistances and capacitances, (c) clinical variables such as blood viscosity and density, and (d) uncertainty in reconstructed lumen geometry. Imaging data using techniques such as magnetic resonance imaging (MRI) or computed tomography (CT) are used to reconstruct the arterial lumen geometry [18], but can be noisy due to motion and registration artifacts, blooming artifacts, motion of the arteries during the cardiac cycle, etc [19,20]. Assumptions such as constant blood viscosity [21], approximation of micro-vessels using lumped parameter boundary conditions, and population averaged empirical laws give rise to more sources of uncertainties. Here, we restrict discussion to the impact of geometry in quantifying uncertainties in hemodynamic simulations of the coronary artery. Impact of clinical parameters such as flow rate and boundary conditions have been explored earlier [22]. We will focus attention on blood flow simulations in human coronary arteries and derived quantity of significant utility in diagnosing the severity of coronary artery disease, the fractional flow reserve (FFR). FFR is defined as the ratio of blood flow rate under conditions of maximal hyperemia (reduced myocardial bed resistance) at a given location to the hypothetical value if no disease were present in the coronary artery. Under modest assumptions, the FFR can be shown to be equal to the ratio of local coronary artery blood pressure to aortic blood pressure under maximal hyperemic conditions.

Clinically, FFR is measured in the cardiac catheterization laboratory using a pressure wire during the intravenous administration of adenosine to elicit maximal hyperemic response [23]. Measurement of FFR has emerged as the gold-standard for determining which lesions in the coronary arteries are flow-limiting and should be stented and which patients should be treated medically [24–26]. Recent developments in patient-specific CFD modeling have enabled the computation of FFR noninvasively from CT data [9], referred to as FFR_{CT}. Data from three multicenter clinical trials indicates that this technology significantly improves the noninvasive assessment of coronary artery disease [10–12]. Thus, the assessment of the sensitivity of patient-specific coronary artery blood flow simulations is of significant interest as these tools are currently being used for clinical decision-making. However, there is still scope for improvement of FFR_{CT} and understanding of sources of error compared to invasive measurements. For example, geometric sensitivity information can aid in identifying regions of the patient-specific model that require extra attention during review, which is the motivation for the present work.

In the past decade, various methods have been developed to quantify uncertainty in partial differential equations (PDEs). Some of these have been used in hemodynamic simulations [14,22] to quantify the impact of uncertainties in inlet blood flow rate, lumped parameter resistances and capacitances, and simple geometric parameters such as angle of anastomosis in bypass grafts and representative stenotic radius. Recently, Steinman and co-workers performed computational simulations to evaluate sensitivities in quantities such as wall shear stress and oscillatory shear index to variations in blood rheology, secondary flows, etc. in human subjects [19,20]. These were some of the first studies showing the relationship between fluctuations in input parameters and output quantities for patient-specific cardiovascular simulations. However, a comprehensive assessment of diverse sources of uncertainties including clinical and geometrical variables on patient-specific models has not been performed to-date owing to the following challenges — (a) depending on the size of the arterial tree, it is computationally intractable to quantify the impact of a large number of uncertain variables, and (b) parameterizing and defining shape-space for patient-specific geometries has not yet been performed. Quantifying uncertainties can help us compute the coefficient of variation, sensitivity as well as the probability distribution of quantities of interest. These, in turn, can impact diagnostic capability as well as prediction of disease progression. For instance, a cutoff of 0.8 in FFR_{CT} is used to determine the treatment protocol.

Uncertainty quantification (UQ) tools can help quantify confidence of diagnosis using FFR_{CT} that is calculated through simulation. Solving stochastic differential equations under geometric uncertainty has not been explored for complex and/or patient-specific geometries. This is the first time uncertainty with respect to patient-specific vascular geometry is performed. Novel developments include (a) a subdivision strategy to define a stochastic space of patient-specific geometries, (b) a machine learning algorithm to accelerate solution to Navier–Stokes equations using a large database of patient-specific solutions, and (c) a reduced order model used as a feature in the machine learning algorithm that simplifies the solution to the Navier–Stokes equations by solving flowrate and pressure loss only using lumen areas.

The goal of this paper is to present an efficient framework to quantify impact of uncertainty in geometry on FFR_{CT}. To achieve this goal, we first reduce the infinite dimensional space of surface geometries to a discretized finite dimensional subset using branch locations as separators. Subsequently, we develop a machine learning based surrogate to the governing equations to make the stochastic problem computationally tractable and accelerate convergence. The stochastic collocation method [27,28] is used to model different sources of uncertainties. Using this approach, simulations are performed at specific collocation points in the stochastic space [27,28]. This technique combines the exponential convergence rates of the GPCE scheme [29–32] with the decoupled nature of Monte-Carlo techniques. It is non-intrusive and can be used with legacy codes or in situations where source code is not available. The Smolyak algorithm helps accelerate the convergence of stochastic collocation method in higher dimensions. It is to be noted that the machine learning method can also be used with other stochastic methods such as probabilistic collocation method [33], simplex collocation method [34] and hybrid stochastic projection method [35].

In spite of the efficiency of the Smolyak algorithm, inclusion of geometric uncertainties could result in hundreds of quadrature points. Even with recent improvements in computational methods [4,36] and processor speeds, it could take days to evaluate sensitivity throughout the arterial geometry. Hence, we use a surrogate model using a machine learning approach to accelerate convergence. Simpler surrogate models have been proposed earlier which do not parameterize variables globally, implying that only one data point is available per simulation and do not provide any spatial resolution. In contrast, we use a descriptive feature set so as to utilize simulation results from every location in the coronary tree. Other alternatives to our machine learning approach are to use reduced-order models, wherein the 3D Navier–Stokes equations are reduced to centerline vessel trees, and mass and energy conservation equations are imposed. These methods are powerful if the reduced order equations could adequately describe behavior of the system. If not, they lack sufficient free parameters to incorporate patient-specific data to predict simulation outcomes. Machine learning does not explicitly satisfy conservation laws, but has the advantage of being significantly faster and more descriptive. Whenever possible, we use features motivated from conservation laws such as using reduced order models. We also implement a disease burden score to capture deviation of arterial geometry from a theoretical healthy radius.

We limit the developments of this paper to uncertainty in lumen radius, since that is hypothesized to be the primary variable driving pressure loss near diseased regions. Further, we assume continuity of FFR_{CT} as well as its derivatives in the stochastic space. The validation of the perturbation model is restricted to deformation maps with fixed bifurcation locations. Further, we assume rigid walls, but methods for fluid–structure interaction [37,38] may also be used with the framework developed in this paper. Other numerical methods such as iso-geometric analysis may also be used [39].

The paper is organized as follows. In Section 2, we discuss some mathematical formalism and background of the techniques developed. We also describe the method used to define uncertainty in geometry. In Section 3, we discuss accuracy of the machine learning model, and subsequently show geometric sensitivity for different patient-specific geometries. In Section 4, we discuss the results and implications of the findings. The nomenclature of symbols used are described in Table 1.

2. Methods

2.1. Image acquisition and model construction

Computed tomography angiography (CTA) images encompassing all coronary arteries and a portion of the ascending aorta for all patients are first extracted. Subsequently, a centerline branching tree is extracted which passes through the center of the aorta, splits into the left and right coronary arteries, which further branches out to the rest of

Table 1
A nomenclature of the symbols used in the paper.

Nomenclature			
ρ	Density	u	Velocity
μ	Dynamic viscosity	f	Body forces
p	Pressure	$P_{\mathtt{aorta}}$	Pressure in aorta
$\mathrm{FFR}_{\mathtt{CT}}$	FFR calculated from CT	P_c	Mean pressure in coronary artery
Ω	Patient-specific geometry	Ω^*	Family of geometries
C_k	Centerline point k	$A(C_k)$	Lumen area at C_k
S_m	Segment m composed of set of centerlines	$A(S_m)$	Lumen area of centerline set S_m
$\Delta\Omega_i$	Perturbation of $i\mathtt{th}$ segment	$\Delta A(S_i)$	Perturbation of $i\mathtt{th}$ segment
$\boldsymbol{\xi}$	Stochastic space	ξ_k^m	$k\mathtt{th}$ collocation point of $m\mathtt{th}$ segment
u_0	Magnitude of uncertainty	α	Magnitude of area perturbation
σ	Correlation length	$d(.,.)$	Euclidean distance
$\mathbf{c}$	Co-ordinate of centerline point	$\mathbf{c_0}$	Co-ordinate of mid-centerline point of segment
Δp	Perturbation of surface node	c_x	Projection of surface node on centerline
$\mathcal{L}$	Lagrange polynomials	$\mathcal{U}$	Uniform distribution
$u_{0,mean}$	Mean uncertainty	$u_{0,std}$	Standard deviation in uncertainty
$u_{\mathtt{max}}$	Maximum uncertainty	κ	Health index
r or r_x	Lumen radius	$r_{\mathtt{healthy}}$	Theoretical healthy radius
$\mathcal{N}$	Gaussian function	w_x	Kernel weighting functions
$I(.,.)$	Indicator function	$S(.,.)$	Sigmoidal function
α	Functional form for perturbation in area	$d_{\mathtt{offset}}$	Distance to nearest upstream bifurcation
$x_{\mathtt{ostium}}$	Co-ordinate of upstream ostium	$x_{\mathtt{up}}$	Co-ordinate of upstream bifurcation
ΔP	Pressure drop	$q_{\mathtt{cor}}$	Total coronary flow rate
$m_{\mathtt{myo}}$	Mass of myocardium	q_{dil}	Dilation coefficient
$R_{\mathtt{eff}}$	Effective downstream boundary resistance	q_a	Aortic flow rate coefficient
R_o	Resistance of outlet o	$n_{\mathtt{down}}$	Number of downstream outlets for a given point
$P_{\mathtt{diastole}}$	Blood pressure at diastole	$P_{\mathtt{systole}}$	Blood pressure at systole
R_{6geom}	Net geometric resistance	$r_{\mathtt{avg}}$	Average radius in a vessel segment
L	Length of vessel segment	$P_{\mathtt{recovery}}$	Pressure recovery factor
$\mathcal{G}$	Information gain	$\mathcal{H}$	Entropy function
$Q(x)$	Flow-rate at a centerline point	$R_{\mathtt{geom}}$	Geometric resistance of a section
$r_{\mathtt{pre}}$	Radius before stenosis onset	$r_{\mathtt{post}}$	Radius after stenosis onset
F_k	$k\mathtt{th}$ Feature in the machine learning algorithm	p_k	Probability of feature k

the coronary artery vasculature. The centerline is described by a finite number of points $N_{\mathtt{CL}}$. Following this, lumen segments are calculated at each centerline point (using automated and manual thresholding) which are lofted to form the three dimensional geometry.

2.2. Deterministic blood flow simulations

Blood flow in the cardiovascular system can be modeled using the Navier–Stokes equations which are given by:

$$\begin{aligned} \rho\left(u_{,t}(x,t)+(u\cdot\nabla)u(x,t)\right) &= -\nabla p(x,t)+\mu\nabla^2 u(x,t)+f \quad \forall x\in\Omega \\ \nabla\cdot u(x,t) &= 0, \end{aligned} \tag{1}$$

where f represents all body forces, ρ denotes density, μ denotes dynamic viscosity, u denotes velocity, p denotes pressure, and Ω represents the patient-specific problem geometry. Finite element simulations have emerged as a powerful and robust tool to solve these equations in complex patient-specific geometries [3,36]. The three dimensional geometry is first discretized into a mesh with tetrahedral elements. A stabilized finite element technique using the generalized-α method [3,40] is employed. Linear basis elements are used within each element. Walls are assumed to be rigid, and Newtonian constitutive behavior of the fluid is assumed, with viscosity of blood being 0.04 g/cm.s. and density being 1.06 g/cm^3. At the aortic inlet, a velocity profile (Dirichlet) boundary condition is prescribed, and the outlets are modeled using a resistance condition which couples blood pressure and velocity at outlet nodes [36]. A parabolic profile was prescribed at the model inlet across the finite element mesh nodes that constitute the inlet surface. Fractional flow reserve is calculated as $\mathrm{FFR}_{\mathtt{CT}}(x) = \frac{P_c(x)}{P_{\mathtt{aorta}}}$ where $P_{\mathtt{aorta}}$ is the mean aortic pressure and

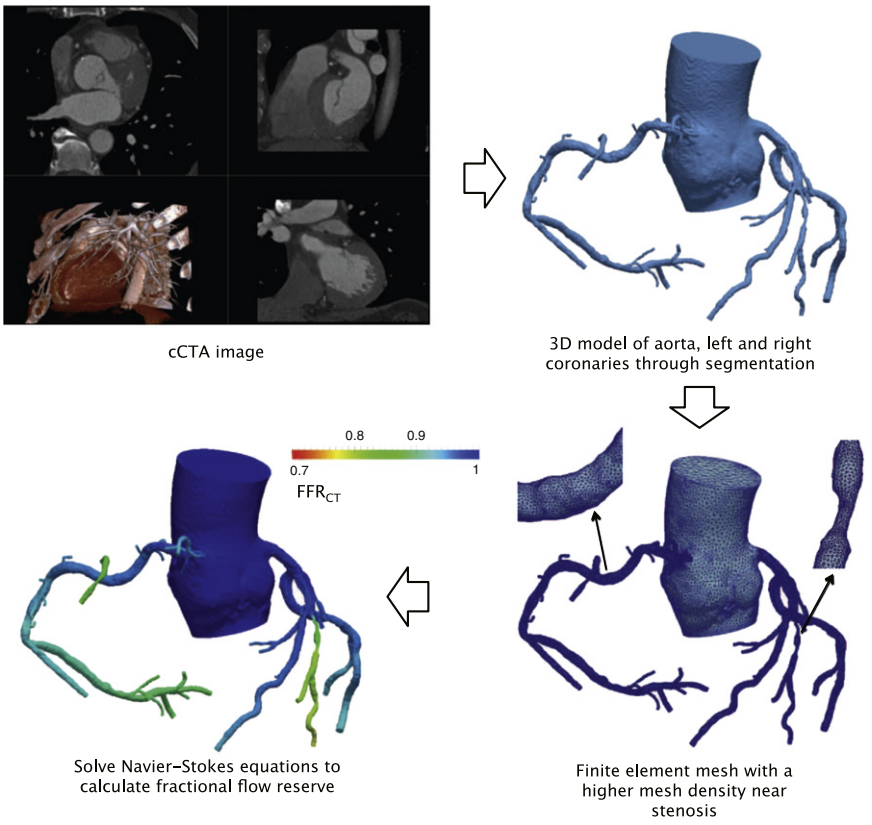

Fig. 1. An overview of the process of calculating $\mathrm{FFR}_{\mathrm{CT}}$ from cCTA images (top left) a coronary computed tomography angiography (cCTA) image encompassing all the arteries of interest, (top right) a three dimensional computational model of the aorta, left and right coronary arteries, (bottom right) a finite element tetrahedral mesh of the reconstructed model and (bottom left) $\mathrm{FFR}_{\mathrm{CT}}$ map calculated from the solution of Navier–Stokes equations.

$P_c(x)$ is the mean pressure in the coronary artery [9]. The various steps in the process of calculating $\mathrm{FFR}_{\mathrm{CT}}$ from geometry are shown in Fig. 1.

2.3. Modeling uncertainty in geometry

Since the true patient-specific geometry is unknown, the goal is to solve the equations in a family of geometries, Ω^*, such that the true geometry lies within this family. Each geometry in Ω^* is associated with a probability, and we are interested in a multitude of ensemble properties of blood flow and pressure within Ω^*. We represent stochastic nature of an entity using the symbol $\boldsymbol{\xi}$, where a probability distribution function is associated with $\boldsymbol{\xi}$.

First, we assume that the arterial geometry is represented by cross-sectional lumen area, denoted $A(.)$, at distinct centerline points (fictional points at the center of artery), denoted by C_i corresponding to the $i\mathtt{th}$ centerline point. Hence, the discrete geometry Ω is approximated first as

$$\Omega \sim \left[A(C_1), A(C_2) \cdots A(C_{N_{\mathrm{CL}}})\right], \tag{2}$$

where N_{CL} is the number of centerline points. Subsequently, centerline points which fall between (i) two branching points, (ii) a branch and an outlet, or (iii) a branch and an inlet, are grouped together. Hence, the geometry could be approximated as

$$\begin{aligned}\Omega &\sim \left[A(C_i) : i \in S_1, A(C_i) : i \in S_2 \cdots A(C_i) : i \in S_{N_{\mathrm{S}}}\right] \\ &\equiv \left[A(S_1), A(S_2), \ldots, A(S_{N_S})\right],\end{aligned} \tag{3}$$

where each S_i consists of all the centerline points which fall in group i as described earlier, and N_S is the number of segments. N_S and N_{CL} are patient-specific and can vary widely across patients.

The allowable space of perturbations is now defined segment-wise. Hence, we can define the allowable space of geometries as

$$\Omega + \Delta\Omega_i(\boldsymbol{\xi}) = \left[A(S_1), A(S_2), \ldots, A(S_i) + \Delta A(S_i, \boldsymbol{\xi}), \ldots, A(S_{N_S})\right],$$

where the perturbation in area of the ith segment is defined as

$$\Delta A(S_i, \boldsymbol{\xi}) \equiv [\Delta A(C_k, \boldsymbol{\xi}) : k \in S_i] \equiv \alpha(\mathbf{c_{k,i}}, \mathbf{c_{i,o}}, \sigma; \boldsymbol{\xi}),$$

where $\mathbf{c_{k,i}}$ is the co-ordinate of the kth centerline point in ith segment, $\mathbf{c_{i,o}}$ is the co-ordinate of center of the segment S_i (i goes from 1 to N_S) and σ is a correlation length defined so that the perturbation at the boundary of the segments is zero.

We explore two forms of the function α. The first is defined as a spatial Gaussian function centered on the center of the segment,

$$\alpha(\mathbf{c_{k,i}}, \mathbf{c_{i,o}}, \sigma; \boldsymbol{\xi}) = u_{0,i}(\boldsymbol{\xi}) \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{d(\mathbf{c_{k,i}}, \mathbf{c_{i,o}})}{2\sigma_i^2}} \tag{4}$$

where $u_{0,i}(\boldsymbol{\xi})$ is the magnitude of uncertainty, $d(.,.)$ is the euclidean distance and the correlation length is set to be 1/6th the length of the segment so that the segment ends are at a distance 3σ from center with negligible perturbation. The other option is to define a uniform perturbation such that

$$\alpha(\mathbf{c_{k,i}}, \mathbf{c_{i,o}}, \sigma; \boldsymbol{\xi}) = u_{0,i}(\boldsymbol{\xi}). \tag{5}$$

The latter does not ensure C^0 continuity of the surface at bifurcation locations, and hence cannot be used with 3D simulations. However, the uniform distribution can be used with the machine learning method and ensures that points close to bifurcations are treated the same as points away from the bifurcations. Hence, we use the former definition (Gaussian) to validate the model by comparing sensitivities, but use the uniform distribution to predict and report sensitivity values.

We can also extend the definition above to define perturbations of a point in the arterial wall. These perturbations will be defined as a perturbation ($\boldsymbol{\Delta}\mathbf{p}$) in the co-ordinate of a surface point $\mathbf{x_i}$, defined as

$$\boldsymbol{\Delta}\mathbf{p}(\mathbf{x_i}) = \alpha(\mathbf{c_{x_i}}, \mathbf{c_{i,o}}, \sigma; \boldsymbol{\xi}) \left(\mathbf{x_i} - \mathbf{c_{x_i}}\right),$$

where $\mathbf{c_{x_i}}$ is the projection of $\mathbf{x_i}$ onto the centerline, such that $\left(\mathbf{c_{x_i}} - \mathbf{x_i}\right) \cdot \mathbf{t_{x_i}} = 0$, where $\mathbf{t_{x_i}}$ is a tangent line along the centerline. To calculate $\mathbf{t_{x_i}}$, we first obtain the closest centerline point to $\mathbf{x_i}$, say c_1. Both the adjacent centerline points are also calculated, say c_2 (succeeding) and c_3 (preceding). Then the projection is calculated by calculating optimal $\lambda = [0, 1]$ where $\mathbf{c_{x_i}} = c_1 + \lambda(c_2 - c_1)$. If $\lambda \neq [0, 1]$, then we calculate optimal $\lambda = [0, 1]$ where $\mathbf{c_{x_i}} = c_3 + \lambda(c_1 - c_3)$. A quadratic equation in λ with a unique real solution results when we substitute $\mathbf{c_{x_i}}$ in the projection equation. These perturbations are used only for the validation problem to ensure continuity of the perturbed geometry. Fig. 3 shows a schematic of the geometry representation.

2.4. Stochastic blood flow simulations

The stochastic Navier–Stokes equations are given by similar partial differential equations as Eq. (1), where the velocity and pressure at each spatial and time point is a random field, due to uncertainty in the problem geometry. The equations are represented by:

$$\begin{aligned} \rho\left(u_{,t}(x, t, \boldsymbol{\xi}) + (u \cdot \nabla)u(x, t, \boldsymbol{\xi})\right) &= -\nabla p(x, t, \boldsymbol{\xi}) + \mu\nabla^2 u(x, t, \boldsymbol{\xi}) + f \quad \forall x \in \Omega^*(\boldsymbol{\xi}) \\ \nabla \cdot u(x, t, \boldsymbol{\xi}) &= 0. \end{aligned} \tag{6}$$

To compute stochastic ensemble properties and sensitivity of the solution to geometry, we first parameterize and define a stochastic space that encompasses possible patient geometries. We then compute the quadrature points where simulations will be performed, by sampling and interpolating the stochastic space using the adaptive Smolyak quadrature (collocation) method. To accelerate convergence, we use a machine learning predictor instead of 3D

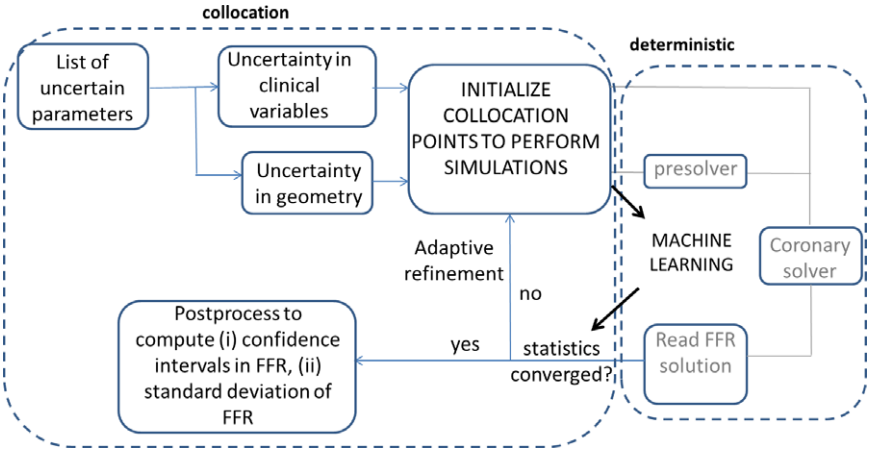

Fig. 2. Schematic of the algorithm that couples adaptive collocation with a machine learning algorithm.

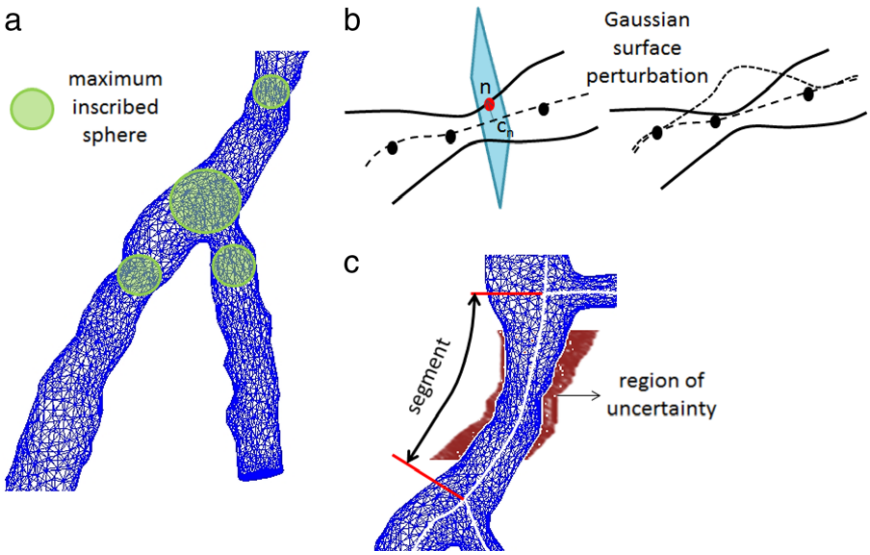

Fig. 3. Schematic of geometric representation of the computational model, (a) radius at each centerline point, calculated using maximum inscribed spheres, is used to represent geometry (b) perturbations on geometry are defined on a plane normal to the centerline and fixed at the end of segments and (c) pictorial representation of a family of geometries around the reconstructed geometry.

Navier–Stokes equations to calculate $\mathrm{FFR_{CT}}$. This entails defining attributes that affects $\mathrm{FFR_{CT}}$. Finally, we can evaluate probability distribution function of $\mathrm{FFR_{CT}}$ and confidence intervals in FFRct from $p(x, t, \boldsymbol{\xi})$. In Fig. 2, we illustrate how the different components involved in quantifying geometric uncertainty are coupled together. Further details and a discrete version of the problem are described in the next section.

2.4.1. Adaptive stochastic collocation framework

In the stochastic collocation technique [27,28], instead of dealing with probability density functions (PDFs) directly, we perform computations in a stochastic space. A finite dimensional stochastic space is described by its truncated descriptor (random vector) $\boldsymbol{\xi}$ so that

$$\boldsymbol{\xi} = \left[\xi^1, \xi^2, \ldots, \xi^N\right]$$

where the dimensionality N of the stochastic support space is problem dependent, superscripts denote dimension, and ξ^i represents either uniform or normally distributed random variables which are mapped to a parameter of interest. Any construct on the stochastic space has an unique PDF associated with it.

In the collocation method, the stochastic space is approximated using mutually orthogonal interpolating functions. To represent a function $g(x,t,\boldsymbol{\xi})$ at any point in the stochastic space, it is written as $g(x,t,\boldsymbol{\xi}) = \sum_i g(x,t,\boldsymbol{\xi}_i)\mathcal{L}_i(\boldsymbol{\xi})$ where $\mathcal{L}_i(.)$ are orthogonal interpolating polynomials, superscript denotes dimension, and subscript denotes the ith collocation point, hence

$$\boldsymbol{\xi}_i = \left[\xi_i^1, \xi_i^2, \ldots, \xi_i^N\right] : \mathcal{P} \longrightarrow \mathcal{R}^N.$$

These interpolating polynomials have the property that $\mathcal{L}_i(\boldsymbol{\xi}_j) = \delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$ and zero otherwise. This property is necessary to decouple the Navier–Stokes equations in stochastic space, i.e., each simulation will correspond to a specific stochastic collocation point. Another possible choice is to use piecewise linear interpolates, but their derivatives are discontinuous at the collocation points and hence converge slowly. The stochastic space can then be queried at any point to compute PDFs. This method is specifically designed for uncertainty quantification in large-scale simulations, such as CFD simulations in complex geometries.

By substituting the Lagrange interpolation formula into Eq. (1) and imposing the residuals to be 0 at $\boldsymbol{\xi} = \boldsymbol{\xi}_i$, we have the discretized stochastic Navier–Stokes equations given by

$$\begin{aligned}
&\rho\left(u(x,t,\boldsymbol{\xi}_i)_{,t} + (u(x,t,\boldsymbol{\xi}_i)\cdot\nabla)u(x,t,\boldsymbol{\xi}_i)\right) = -\nabla p(x,t,\boldsymbol{\xi}_i) + \mu\nabla^2 u(x,t,\boldsymbol{\xi}_i) + f \\
&\quad \forall i = 1,2,\ldots,M \ \forall x \in \Omega^i \\
&\nabla\cdot u(x,t,\boldsymbol{\xi}_i) = 0 \quad \forall i = 1,2,\ldots,M
\end{aligned} \tag{7}$$

where the stochastic blood velocity, $u(x,t,\boldsymbol{\xi}) = \sum_{i=1}^{M} u_i(x,t,\boldsymbol{\xi}_i)\mathcal{L}_i(\boldsymbol{\xi})$, stochastic blood pressure, $p(x,t,\boldsymbol{\xi}) = \sum_{i=1}^{M} p_i(x,t,\boldsymbol{\xi}_i)\mathcal{L}_i(\boldsymbol{\xi})$, and $\Omega^i = \Omega^*(\boldsymbol{\xi}_i)$ is the perturbed geometry at collocation point corresponding to index i. Each $\Omega^*(\boldsymbol{\xi}_i)$ is translated to a specific geometry by using the perturbation functions defined in Eqs. (4) and (5). The coefficient u_0 in Eq. (5) is defined as a uniform function in the stochastic space, i.e. $u_0 \sim \mathcal{U}[0, u_{\mathtt{max}}]$, where $u_{\mathtt{max}}$ is the maximum uncertainty defined usually as a percentage of the radius. Hence, the coefficient is modeled as $u_{0,j}(\boldsymbol{\xi}_j) = u_{\mathtt{max}}/2 + u_{\mathtt{max}}/2(2\boldsymbol{\xi}_j - 1)$. Similarly, for a Gauss distribution, the inverse error function ($\mathtt{erf}^{-1}$) is used as $u_{0,j}(\boldsymbol{\xi}_j) = u_{0,mean} + u_{0,std}\,\mathtt{erf}^{-1}(0.99 * (2\boldsymbol{\xi}_j - 1))$ where $u_{0,mean}$ and $u_{0,std}$ are the mean and standard deviation of the coefficient respectively.

The sparse-grid Smolyak algorithm [41,42] is used to compute the collocation points in multidimensional random space. It has been shown that the sparse grid works best when the 1D quadrature rules are nested [42]. In this paper, we employ 1D nested Chebyshev collocation nodes. It has been reported that using 1D Gauss quadrature rules results in a higher number of function evaluations than using sparse grids for the same level. In a Smolyak sparse grid, the depth of interpolation [41] defines the number of simulations to be run and error indicators are computed for each depth of interpolation. The depth of interpolation allows flexibility in implementation because it can be adjusted according to computational expense. The choice of sparse grid depends on the interpolation scheme that is being employed. We showed recently that Lagrange interpolates converge very quickly in several cardiovascular blood-flow examples [22].

The Smolyak algorithm does not take into account any information about the function itself [27,28]. Dimensional adaptivity has been proposed in previous work [42] in which a dimension is refined based on its error indicator. Yet, the refinement of each dimension itself is uniform, and hence adaptivity over a specific region in the stochastic space is not performed. Hence, it does not account for steep function variations, and reduces to a conventional sparse grid for any symmetric problem. We recently developed a method to adaptively choose collocation points for Lagrange interpolating polynomials [14,22]. The essence of the algorithm is to split collocation points into a frozen set and an active set. A neighborhood is associated with each collocation point. The neighbors are estimated by traversing along one dimension at a time. This neighborhood consists of at most two collocation points in each dimension. Collocation points at the boundary of the N-dimensional hypercube will have one neighbor in at least one dimension. A point is considered frozen if its error indicator is zero or if it lies inside a frozen patch. We start with a coarse level and all the collocation points are chosen to be active initially. The algorithm terminates when all collocation points are frozen (refer Algorithm 1). Details of the algorithm are provided in [22] and Algorithm 1 described key steps. A schematic of the algorithm is provided in Fig. 4.

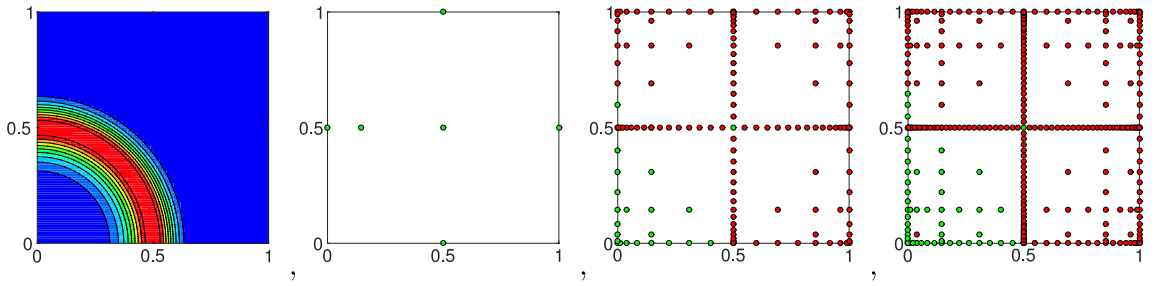

Fig. 4. The figure shows (from left to right) an analytical function being represented using collocation, $f(\xi_1, \xi_2) = exp(-(\xi_1^2 + \xi_2^2 - 0.25) \times 100)$, the initial collocation grid with a depth of interpolation 1, the adaptive collocation grid at an intermediate iteration (depth = 5), and the adaptive collocation grid at a depth of 7. The frozen set $\mathcal{F}$ is depicted using red circles and the rest of the collocation points are depicted using green circles. The figure shows how the active collocation points are restricted to the neighborhood of the region with steep gradients in the function.

Algorithm 1 Algorithm for adaptive stochastic collocation adapted from [22]. $\mathcal{F}$ denotes a set of frozen collocation points, $\mathcal{C}_d$ denotes stochastic collocation points at depth of interpolation (denoted by d), $\mathcal{M}$ denotes neighborhood of a collocation point (denoted by c), the suffix "adapt" denotes adaptive stochastic collocation points, $\mathcal{E}$ denotes interpolation error and ϵ denotes threshold.

Set $d = 1$
Set $\mathcal{F} = \emptyset$
Perform simulations at $\mathcal{C}_1$ and $\mathcal{C}_2$
while $\exists c$ such that $\mathcal{E}(c) < \epsilon$ **do**
  Evaluate function using Lagrange interpolation at $\mathcal{C}_{\texttt{adapt},d+1} \equiv (\mathcal{C}_{d+1} - \mathcal{C}_d)$
  for $\forall c \in \mathcal{C}_{\texttt{adapt},d+1}$ **do**
    if $\exists m \in \mathcal{M}(c) \notin \mathcal{F}$ **then**
      Perform simulation at c
      Calculate $\mathcal{E}(c) = f(c) - \mathcal{I}f(c)$
      if $\mathcal{E}(c) < \epsilon$ **then**
        Update $\mathcal{F} = \mathcal{F} \cup c$
      end if
    else
      Interpolate function value at c
      $\mathcal{F} = \mathcal{F} \cup c$
    end if
  end for
  Set $d = d + 1$
end while

2.5. Accelerate convergence using machine learning predictor

Though adaptive collocation is computationally much faster compared to Monte-Carlo methods, this method still needs numerous blood flow simulations to reliably evaluate statistics of the quantity of interest. Instead of solving 3D Navier–Stokes equations at each quadrature point, we evaluate a surrogate model which approximates the solution obtained using the 3D equations. Surrogate methods use the results of simulations performed off-line, and substitutes the original 3D equations with a simpler input–output relationship that is very inexpensive to evaluate. Such methods have a "training" and "testing/prediction" mode. While this general concept has been explored earlier using techniques such as stochastic response surface [14], reduced order modeling(ROM), proper orthogonal decomposition (POD), the machine learning methods we use here have the following advantages: (a) they do not rely on solutions to ODE's, hence they are significantly faster to evaluate than ROM or POD methods, (b) they are not necessarily interpolatory, hence they provide more flexibility than response surface methods. However, they rely on adequate training data so that the conservative laws (momentum and mass) are "built-in" and the space of patient geometries is adequately represented.

Machine learning techniques encompass a broad class of built-in methods that map input features to quantities of interest. Hence, the general steps involved in the machine learning algorithm are (i) define relevant features (ii) construct training dataset (iii) compute best regressor and (iv) test performance of regressor on test set. We describe the steps in this order.

2.5.1. Defining problem-specific features

Choosing problem and application-specific features is an important step in the machine learning algorithm. Each feature should be chosen to capture some influence on the predicted variable. These features must be easy to compute given a geometry, boundary conditions, and clinical parameters (i.e., they themselves cannot be the solution of PDE or ODE's), but also sufficient to give a reasonable estimate of the solution. Since it is not possible to know the sufficiency of a feature set *a-priori*, we break it down to multiple steps. We first pick a set of features encompassing geometric, clinical and analytical model based features, such as analytical solutions for pipe flow parameterized by pipe radii, flow rate, viscosity, and length of the pipe. These features are chosen based on their relevance to prediction of $\mathrm{FFR}_{\mathrm{CT}}$. We test if a level of desired accuracy is reached, and if not, investigate all the failed cases to enrich the feature space. We describe each of these factors and steps below.

Geometric features: We use a combination of local, upstream, downstream, and global geometric measures as geometric features associated with a given point. The local geometric features used are lumen area and lumen diameter. "Local" refers to features that depends on the centerline point under consideration, whereas "global" refers to features that depends on the rest of the coronary tree. The general geometric features used are

- Local, upstream and downstream radius: Since pressure drop depends inversely on lumen radius, and flow rate increases with higher lumen radius (with a fixed pressure boundary condition), local, upstream and downstream neighboring radius are included as features. The energy loss through the current point also depends on the local radius.
- Distance to nearest bifurcations: Since there are energy losses associated with flow-split, stagnation region, flow stasis and recirculating regions near bifurcations, it is important to model the distance to nearest upstream and downstream bifurcation. Distance from nearest upstream branch is important to distinguish between branching losses and losses away from branch.
- Minimum, maximum and mean area of downstream outlets: These are important to model the effect of boundary resistances, from which we post-process to calculate the net downstream resistance by solving the circuits in parallel.
- Minimum upstream and downstream diameter: These are flow limiting through the current point, and hence important to include. If both of these are very small, the net path resistance is lower. If one of these is small, the net path resistance depends on the other paths between ostium and outlet through this point. Hence, maximum, minimum and average upstream and downstream diameters are also included.
- Pressure recovery factor: Sometimes, the location downstream of a disease has a region of increasing area where the kinetic energy at stenosis is converted into pressure energy. We call this a pressure recovery zone, and include a pressure recovery factor which is the ratio of areas downstream of stenosis, to the area at stenosis.

Since the definition of radius at bifurcations is ambiguous, we use the radius of the maximum inscribed sphere as representative of lumen radius. We ensured that all of these features contributed to the final decision tree, and are crucial to the machine learning algorithm.

In addition to these, a health index score ($\kappa^i(x)$) is calculated which is defined as

$$\kappa^i(x) = \frac{r(x)}{r^i_{\mathtt{healthy}}(x)}$$

where $r^i_{\mathtt{healthy}}(x)$ is the theoretical healthy radius of the lumen, $i = 1$–15 denoting different methods of calculating disease burden score, and $r(x)$ is the radius of the maximum inscribed sphere within the lumen. A disease burden score ($\gamma^i(x)$) is derived from the health index score as $\gamma^i(x) = 1 - \kappa^i(x)$ if $\kappa^i(x) \leq 1$ and 0 otherwise.

We calculate $r^i_{\mathtt{healthy}}$ using global regression of the radii calculated in a subset of the vessel [43]. The idea is to use global interpolation such that the healthy radii can be reasonably inferred. Stenosed regions are typically characterized by a u-shape in the radius curve. However, since diseases could either be sharp and abrupt (acute) or long (diffuse),

and since radii naturally have a sharp decrease at bifurcations (dictated by Murray's law), we need a family of global regressors to infer the health index score [43]. We use three different regressors here [44], which are explained in Appendix A.

Patient-specific features: The blood viscosity, derived from hematocrit(hct) as $\mu = \frac{\mu_p}{(1-\mathtt{hct}/100)^{2.5}}$ is used as a feature where μ_p is the viscosity of plasma ($\mu = \mu_p = 0.0011$Pa.s. when hematocrit is zero). The height (h), weight(w), systolic and diastolic blood pressures, and myocardial mass ($m_{\mathtt{myo}}$) of the patient are used as additional patient-specific features. Derived patient-specific features used are Body Surface Area (BSA) = $\sqrt{\frac{\mathtt{hw}}{3600}}$ where height is in cm and weight is in kg, inlet aortic flow rate $Q = a_Q \mathtt{BSA}^{1.15}$ in $\mathtt{cm}^3/\mathtt{s}$, and coronary flow rate $q_{\mathtt{cor}} = a_{\mathtt{dil}} m_{\mathtt{myo}}^{0.75}$ in $\mathtt{cm}^3/\mathtt{s}$, where $a_{\mathtt{Q}} = 1/60$ is a normalizing constant and $a_{\mathtt{dil}} = 0.33$ is the dilation factor.

Hemodynamic features: Using the downstream resistances and coronary flow rate, we approximate the flow rate through each segment of the model. We describe the method for calculating net effective resistance and pressure loss using reduced order model in Appendix B. The pressure loss model is based on the health index score, $\kappa^i(x)$.

2.6. Algorithm

The algorithm to calculate optimal rule set or regressor depends on the choice of regressor. We explored linear regressor, univariate decision trees and multivariate decision trees. The optimal linear regressor can be calculated using a standard least-squares fit algorithm. To calculate optimal decision trees (T), we first define information gain ($\mathcal{G}$) associated with a feature, defined as:

$$\mathcal{G}(\mathrm{FFR}_{\mathtt{CT}}, F_k, \eta) = \mathcal{H}(\mathrm{FFR}_{\mathtt{CT}}) - \mathcal{H}_{F_k}(\mathrm{FFR}_{\mathtt{CT}}, \eta)$$

where $\mathcal{H}$ is the entropy function and F_k is a specific feature. The entropy function for the training data is given by

$$\mathcal{H}(\mathrm{FFR}_{\mathtt{CT}}) = -\sum_i p_i \mathtt{log}_2(p_i)$$

where p_i is the probability of FFR = FFR_i (FFR is split into discrete intervals of size 0.01), and the entropy function associated with a specific feature is given by

$$\mathcal{H}_{F_k}(\mathrm{FFR}_{\mathtt{CT}}, \eta) = -\sum_{i|F_k<\eta} q_i \mathtt{log}_2(q_i) - \sum_{i|F_k\geq\eta} r_i \mathtt{log}_2(r_i)$$

where q_i is the probability of FFR = FFR_i for the subset $F_k < \eta$, and r_i is the probability of FFR = FFR_i for the subset $F_k \geq \eta$. The feature and η that maximizes gain is chosen as the decision variable and cutoff respectively. This step is recursively repeated till there are no more features to make decision on. During each step, the information gain is calculated on the sub-tree under consideration. We use a greedy divide-and-conquer algorithm [45] to calculate the optimal decision tree, T. T takes a feature vector $\mathbf{f}$ as input and outputs the corresponding $\mathrm{FFR}_{\mathrm{ML}}$. The tree depth is imposed by having a minimum number of allowable elements in the sets $F_k < \eta$ and $F_k > \eta$. Algorithm 2 describes the set-up of the algorithm, and the reader is referred to [45] for further details.

2.7. Implementation

For each patient, we traverse through the coronary tree to calculate all the geometric features and effective boundary conditions described above. Two full sweeps of the coronary tree are needed, once from the root to the leaves to calculate all upstream features and once from the leaves to the root to calculate all downstream features. Local features and patient-specific features are also assigned in this step. Two additional sweeps starting from the root are needed, one to calculate the flow rate and another to calculate pressure drop, since they involve a combination of upstream and downstream indices. Algorithm 2 summarizes the basic steps involved. First, in the training mode, a function is called to compute values of different features. The "down_propagate" function involves traversing all the centerline points from the root (ostium) to the terminal leaves. The "up_propagate" function involves traversing all the centerline points from the leaves to the root. Both of these functions are recursive. All the feature values in Table 2 can be calculated using these steps. Derived parameters such as flow-rates and effective resistances are calculated from these, and are also used as input features in the machine learning algorithm.

Table 2
Independent features used in the machine learning algorithm, and their minimum, maximum, mean and standard deviation in the training set. Distances are in mm, areas in mm^2 and volume in mm^3 unless otherwise specified. Systolic and diastolic pressures are in millimeters of mercury, myocardial mass is in grams, and pressure drop and resistances are in $g/mm/s^2$ and $g/mm^4/s$ respectively.

Feature name	Minimum	Maximum	Mean	Standard deviation
Number of downstream bifurcations	0	28	4.243	3.645
Total downstream volume	0.034	2514	328	333
Average downstream diameter	0.54	2.77	1.68	0.30
Minimum downstream diameter	0.05	0.94	0.23	0.12
Distance to minimum downstream diameter	0.08	185	41.68	30.03
Area of nearest downstream bifurcation	0.25	22.07	4.25	2.54
Distance to nearest downstream bifurcation	0	102.85	11.14	11.47
Number of downstream outlets	1	29	5.243	3.645
Total area of downstream outlets	0.633	63.07	10.135	7.09
Inlet area	0.761	939.29	285.79	303.80
Lumen area	0.02	24.26	4.67	3.00
Mean outlet resistance	6.73	346.24	41.45	24.43
Estimated flow	0.006	1.02	0.202	0.134
Estimated pressure	9511	16000	12397	1416.4
Estimated FFR	0.4	1	0.931	0.08
Systolic pressure	100	170	127.1	15.1
Diastolic pressure	55	100	75.92	10.26
Height (cm)	148	181	166.6	7.14
Weight (kg)	47	96	69.3	9.88
Myocardial mass	64	282	123.45	35.44
Number of upstream bifurcations	0	17	4.23	2.64
Total upstream volume	0.187	11307.7	748.98	1458.72
Average upstream diameter	1.03	17.64	3.33	1.11
Minimum upstream diameter	0.151	5.484	1.79	0.74
Distance to minimum upstream diameter	0	130.99	18.58	22.54
Area of nearest upstream bifurcation	0.251	939.29	42.56	145.37
Distance to nearest upstream bifurcation	0	102.32	12.21	11.94
Distance to ostia	11.53	200.43	74.41	35.87
is stenotic	0	1	0.99	0.05
Net geometric resistance	0	11.42	0.35	0.56
Geometric resistance	0	0.763	0.003	0.01
Pressure drop estimate	−91.68	8794.41	1057.92	1086.68
Worst upstream disease burden	0.12	1.00	0.684	0.159
Pressure recovery factor	1.00	2.00	1.01	0.062

Given all the patients, we split them into two sets, a training and testing set. These attributes are aggregated and written to a training database, and analyzed using *Weka* [45]. The method to calculate optimal decision (REP) tree is described in the previous section. To achieve better results in diseased segments, we perform regression on a "square" transformation of FFR_{CT}.

3. Results

3.1. Convergence analysis

3.1.1. Mesh convergence

To ensure convergence, we performed a mesh-independence study on 20 patients. We gradually refined the mesh density and compared the FFR_{CT}. Our reference solution was calculated using an average of 40 million elements. We found that the maximum error in FFR_{CT} fields was less than 2% when average number of elements was around 4 million. Spatially, the mesh density depends on the cross-sectional area, and hence depends on disease severity.

Algorithm 2 Algorithm for calculating decision tree regressor

```
function COMPUTE_FEATURES_AND_VALUES(mode, (optional)T)
    for i = 1 to number_of_ostium do
        down_propagate(ostium(i), mode)
    end for
    for i = 1 to number_of_outlets do
        up_propagate(outlet(i), mode)
    end for
    function DOWN_PROPAGATE(current_node, mode)
        calculate and update upstream properties
        if is_branch(current_node) then
            down_propagate(daughter_1(current_node))
            down_propagate(daughter_2(current_node))
        else
            if !is_outlet(current_node) then
                down_propagate(daughter(current_node))
            end if
        end if
        if mode == TrainingMode then
            Set FFR from simulation
        else
            Calculate FFR_ML = T(f)
        end if
    end function
    function UP_PROPAGATE(current_node, mode)
        assign and update downstream properties
        if !is_ostium(current_node) then
            up_propagate(parent(current_node))
        end if
        if mode == TrainingMode then
            Set FFR_ML from simulation
        else
            Calculate FFR_ML = T(f)
        end if
    end function
end function
Training Mode
compute_features_and_values(TrainingMode)
Compute optimal regressor
while ( do depth of T < t_d)
    for ∀ f_i ∈ F do
        if G(FFR_CT, f_i, η) > G* then
            G* = G(FFR_CT, f_i, η)
            f* = f_i
            η* = arg max_η G(FFR_CT, f_i, η)
        end if
    end for
    Add f* and η* to the terminal leaf of T
end while
return T
Prediction Mode
compute_features_and_values(PredictionMode, T)
```

3.1.2. Stochastic space convergence

We perform a convergence study in the stochastic space by comparing Monte-Carlo simulations to stochastic collocation and adaptive stochastic collocation method. Since the number of independent stochastic dimensions for calculating geometric sensitivity is of the order of 50, and to make it feasible to compare the different methods, we perform a four-dimensional uncertainty analysis using the following parameters — (a) aortic blood pressure (boundary condition at the inlet), (b) blood viscosity, (c) fraction of flow into the coronary determined by two independent parameters, myocardial mass and a scaling constant. The objective here is to illustrate convergence of $\mathrm{FFR_{CT}}$ with respect to four simulation parameters, and compare with Monte-Carlo and tensor product grids. We define the relative error in statistics κ as

$$\kappa = \frac{|\mu - \mu^*|}{\mu^*} + \frac{|\sigma - \sigma^*|}{\sigma^*} + \frac{|\mathtt{CI}_{90} - \mathtt{CI}^*_{90}|}{\mathtt{CI}^*_{90}}$$

where μ is the mean FFR, σ is the standard deviation of FFR, CI_{90} is the right 90% confidence interval limit of FFR (averaged over the entire model), and the superscript * denotes the corresponding values obtained using converged Monte-Carlo simulations ($\sim$600). A comparison of number of simulations using various methods is shown in Fig. 5.

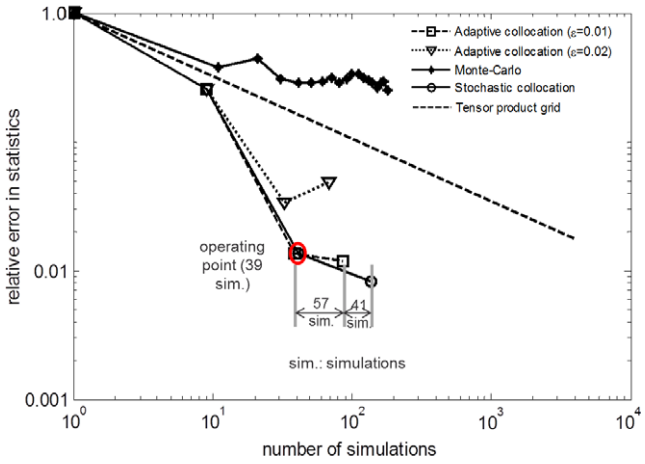

Fig. 5. Comparison of convergence using stochastic collocation, tensor product grid, Monte-Carlo and adaptive stochastic collocation methods to quantify uncertainty in a four stochastic dimensional patient-specific cardiovascular blood flow simulation problem. The tensor product grid line is representative because even coarse grids can run up to thousands of simulations.

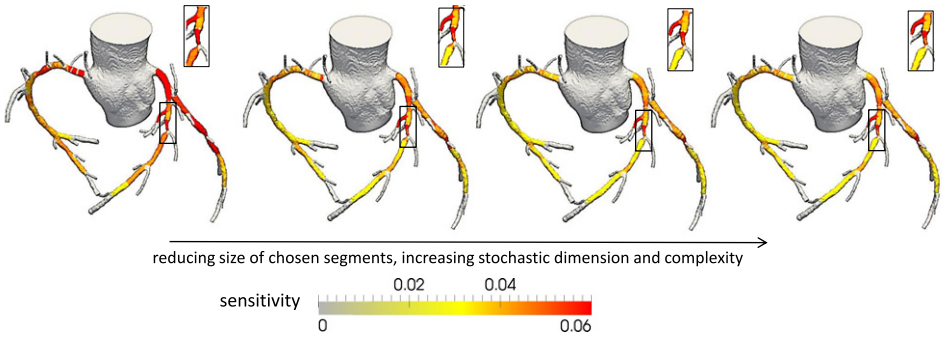

Fig. 6. Dependence of the sensitivity of FFR_{CT} on problem discretization. Figure on the left shows sensitivity field calculated by using segmental sensitivities, and we progressively show sensitivity fields calculated by aggregating centerline points of a certain length, irrespective of the location of bifurcation. The number of independent segments used are (from left to right) 55, 150, 300 and 600. We are able to obtain higher spatial resolution by choosing segments by aggregating centerline points, with a higher computational burden. However, the segmental sensitivity captures all the regions of high sensitivity (with poor spatial resolution).

In the figure, we extrapolate results from tensor product interpolation, since the number of simulation points needed for a level 2 is $5^4 = 625$. The figure demonstrates that adaptive stochastic collocation method offers an attractive tradeoff between number of simulations and convergence. A tolerance parameter of 0.01 ($\epsilon = 0.01$) was chosen.

3.2. Solution dependence on geometry parameterization

Here, we test the dependence on the choice of geometric parameterization on the calculated sensitivity fields. To show the impact of the chosen geometric parameterization, we perform four analyses on a patient-specific dataset — (a) the geometry is split based on branching locations, and (b) the geometry is split based on the aggregates of centerline points, aggregate sizes corresponding to 20, 10 and 5 respectively. The average number of centerline points between two bifurcations is typically of the order of hundreds, hence the chosen geometric split provides finer levels of discretization. The number of centerline points in Eq. (2), N_{CL}, was ~3000. Hence, the number of splits in Eq. (3), N_S, increased from 55, to 150, 300, and 600 corresponding to aggregate sizes of 20, 10 and 5 respectively. Fig. 6 shows the sensitivity field for these four levels of parameterization. The sensitivity analysis based on branching location splits is an upper-bound on the sensitivity values obtained with finer spatial discretization. The latter helps focus on a smaller region at a higher computational expense. We choose the segmental sensitivity for the rest of the paper, but depending on the application, a finer split can be used.

Table 3
Summary of performance of different machine learning regressors to predict full simulation FFR_{CT}, over four combination of training and test sets (DT: decision tree, DTR: decision tree with a linear regression rule at leaves, Def: DeFACTO dataset, Disc: DiscoverFlow dataset). We selected the DT regression rule trained on Def as the final machine learning regressor.

Regression rule	Training set	Test set	ρ	Mean abs. error	RMS error
DT	Disc	Def	0.909	0.031	0.049
DTR	Disc	Def	0.723	0.035	0.103
DT	Def	Disc	0.943	0.026	0.040
DTR	Def	Disc	0.924	0.030	0.046

3.3. Verification of machine learning model

We split the verification of machine learning model into two steps — (i) we compare the FFR_{CT} calculated at different points in the centerline vessel tree using 3D simulations and ML, and (ii) we compare the effect of perturbation of randomly chosen segments using 3D simulations and ML.

3.3.1. Comparison of FFR_{CT}: machine learning versus 3D solution

We use the spatially averaged scalar field, FFR_{CT}, over a cross-section from 3D simulations calculated at different centerline points as our target variable. Our datasets comprised of two clinical trials — the DiscoverFlow trial [11] and the DeFACTO trial [12]. We trained the machine learning algorithm using either DiscoverFlow data (90 patients) or DeFACTO data (240 patients), and the optimal rules were calculated using Weka software. Results of the performance are summarized in Table 3. To avoid over-fitting on the training data, and to show the efficacy of chosen features and the regressor, we show results for different combinations in Table 3. Results were of similar magnitude when the patients were split into 66%–34% in each of the sets (e.g. separating DeFACTO data into a training set of 158 and testing on the rest yielded correlation coefficient of 0.943, mean absolute error of 0.026 and RMS error of 0.040). The regression rules shown are: (i) DT, which is a decision tree and (ii) DTR, which is a decision tree but the leaves are allowed to have a linear regression output instead of a decision value. The correlation coefficient is represented by ρ. All the results were obtained using bootstrap aggregating [46] over 10 decision trees.

Performance of the ML regressor on the test set was similar between using DeFACTO data as test set and DiscoverFlow as training set, as shown in Table 3, or vice versa. Most cases have correlation coefficient >0.9 with a mean absolute error of approximately 0.03. Due to the interpolatory nature of DTR rules at the leaves, the predicted FFR_{CT} value could take negative values and hence can result in poor performance with smaller training datasets (as evidenced in Table 3). Hence, we picked the decision tree regressor. This helps reduce the variance and avoid over-fitting to the training set. Poiseuille approximation and lumen narrowing scores featured closer to the root in the final decision tree, whereas boundary resistance and pressure recovery featured closer to the leaves. Hence, the latter can be considered corrective sub-trees to the FFR_{CT} predicted by Poiseuille equation.

3.3.2. Comparison of sensitivity: machine learning versus 3D solution

Since 3D simulations are relatively time consuming, calculating sensitivities by multiple runs of 3D simulations are computationally burdensome and not feasible in a reasonable time frame. Hence, for the purpose of comparison and verifying the sensitivity fields, we compare the solutions of perturbation analysis. Sensitivities $\sigma_{FFR_{CT}}$ are defined as

$$\sigma_{FFR_{CT}} = \sqrt{\sum_{i=1}^{N} w_i FFR_{CT_i}^2 - \mu_{FFR_{CT}}^2},$$

where $\mu_{FFR_{CT}} = \sum_{i=1}^{N} w_i FFR_{CT_i}$ is the mean FFR_{CT}, N is the number of collocation points, and $w_i = \int \mathcal{L}_i(\xi)\, p(\xi) d\xi$ are the weights obtained by integrating Lagrange polynomial ($\mathcal{L}_i(\xi)$) corresponding to collocation point i. Sensitivities can be computed from multiple perturbation solutions (ΔFFR_{CT}). An example of a single perturbation study is

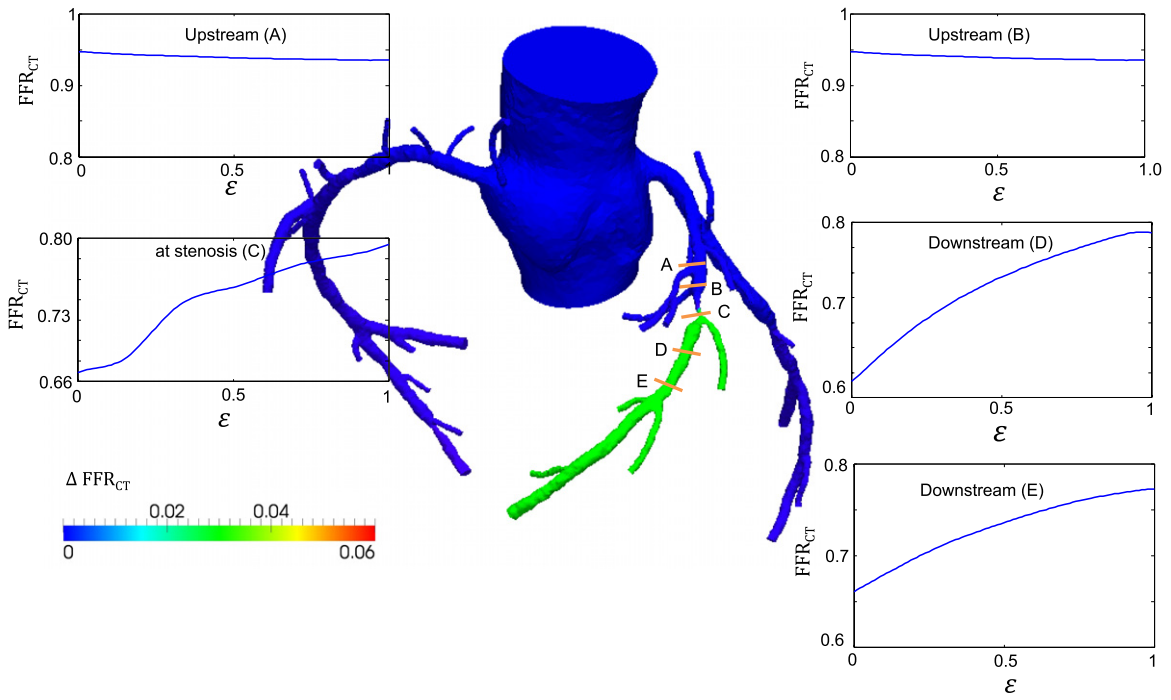

Fig. 7. Difference in $\mathrm{FFR_{CT}}$ due to uncertainty in the segment containing the stenosis. The figure also shows how $\mathrm{FFR_{CT}}$ varies in the stochastic space at two locations upstream and downstream of the stenosis. $\mathrm{FFR_{CT}}$ upstream shows a drop because of higher blood flow when the stenosis resistance is reduced, but downstream shows higher $\mathrm{FFR_{CT}}$ due to lower pressure drop at stenosis.

illustrated in Fig. 7. The sensitivity can be rewritten as

$$\sigma_{\mathrm{FFR_{CT}}} = \sqrt{\sum_{i=2}^{N} \tilde{w}_i \, \Delta \mathrm{FFR_{CT}}_i^2 - \mu^2_{\mathrm{FFR_{CT}}}},$$

where $\tilde{w}_i$ are re-adjusted weights. Comparison of $\Delta \mathrm{FFR_{CT}}_i$ calculated at randomly chosen collocation points and coronary segments using machine learning and 3D analysis is shown in Fig. 8. The 3D perturbation analysis is achieved by first perturbing a surface node as described in Eq. (4) in Section 2.1. A perturbation magnitude of 20% is used. These new surface nodes are used as input to the machine learning algorithm, where the new centerline areas are calculated using radii of maximum inscribed spheres. The correlation coefficient between predicted and actual values in Fig. 8 is 0.92 over 20 patients and 30 segments. The mean absolute error and RMS error were 0.014 and 0.018 respectively.

Fig. 9 compares the time taken for a single simulation of N–S equations as well as sensitivity equations for a problem of steady flow in a patient-specific coronary artery model. For a single simulation, the machine learning algorithm computes the solution in a few seconds on a single processor workstation compared to over a half hour on 90 cores of a server using 3D simulations at a minimal cost in accuracy. For the sensitivity problem, it is infeasible to use the 3D solver in routine clinical use which could take a few days even using 90 cores on server, while our solution produces a sensitivity calculation in a few minutes.

3.4. Geometric sensitivity analysis

Here, we compute sensitivity to geometry by using repeated evaluations of the machine learning based prediction, where the space of possible lumen geometries is explored using the stochastic collocation method. A segment-wise uniform perturbation model described earlier is used. The $\mathrm{FFR_{CT}}$ solution corresponding to different collocation points is evaluated first, after which the standard deviation in $\mathrm{FFR_{CT}}$, $\sigma_{\mathrm{FFR_{CT}}}$ is calculated at all spatial locations. We are interested in the impact of a segment on the global $\mathrm{FFR_{CT}}$, hence we associate the maximum value of $\sigma_{\mathrm{FFR_{CT}}}$ across

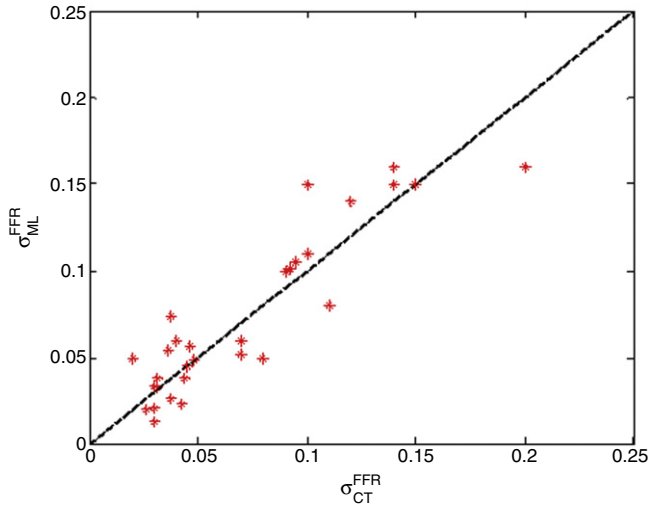

Fig. 8. Correlation between sensitivities obtained using machine learning and through perturbation of surface mesh in the 3D model. Correlation coefficient is 0.92.

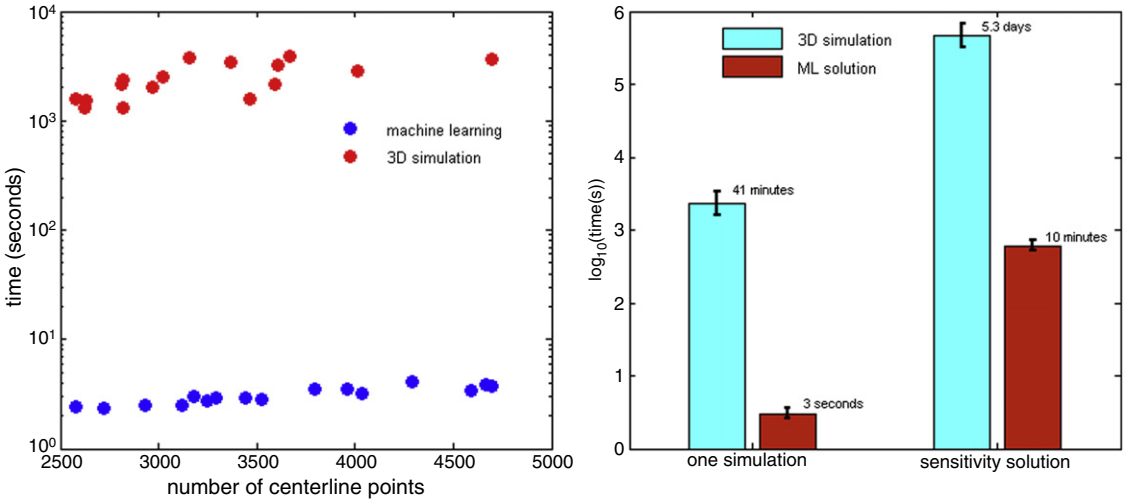

Fig. 9. The figure shows (left) comparison of time taken to perform a single simulation using Navier–Stokes equations in 3D and machine learning method and (right) bar plot comparing time for the sensitivity problem using 3D simulation and machine learning, the latter reducing from more than 5 days using 90 cores to 10 min in a single core machine. Note that for both figures, time is on log scale.

the coronary tree with a chosen segment. This value indicates the maximum impact of changing the cross sectional area of all the lumens in the given segment, which could be either upstream or downstream of the segment. In general, we observe that maximum changes occur downstream of the segment, though the upstream FFR_{CT} is also affected due to the different flow rate. Fig. 10 shows sensitivities of different segments in the coronary tree, indicating diseased segments tend to have higher sensitivities. However, it is not necessary that all highly sensitive regions are diseased.

Sensitivity fields computed on five patients from the test set are shown in Fig. 11. The plots show that sensitivity information can help identify and localize segments which have the most impact on predicted FFR_{CT} in each of the patients. Note that even segments with no or minimal disease could be sensitive, if they are critical to blood transport. An example of critical blood transport vessel is the ostial segment, which shows high sensitivity in most situations.

In general, sensitivity varies non-linearly with respect to magnitude of perturbation. A positive perturbation of a segment does not necessarily have the same impact magnitude as a negative perturbation. The effect of positive perturbation is significant if a vessel is healthy or moderately diseased (akin to dilation through stenting), but this is

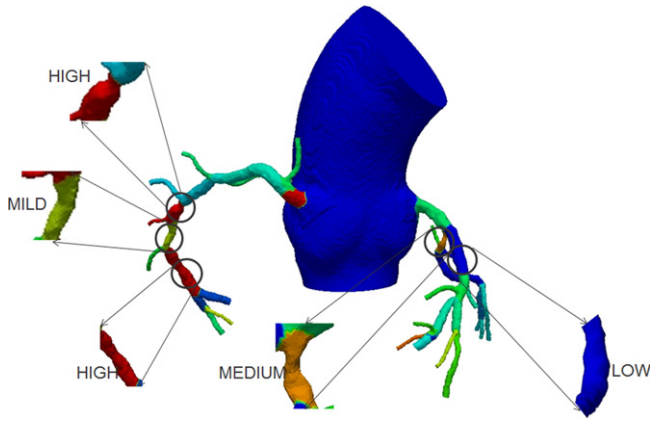

Fig. 10. Figure shows different segments ranked according to their worst case impact on FFR_{CT} in a patient-specific model. Representative segments for high, medium, and low impact on FFR_{CT} are magnified.

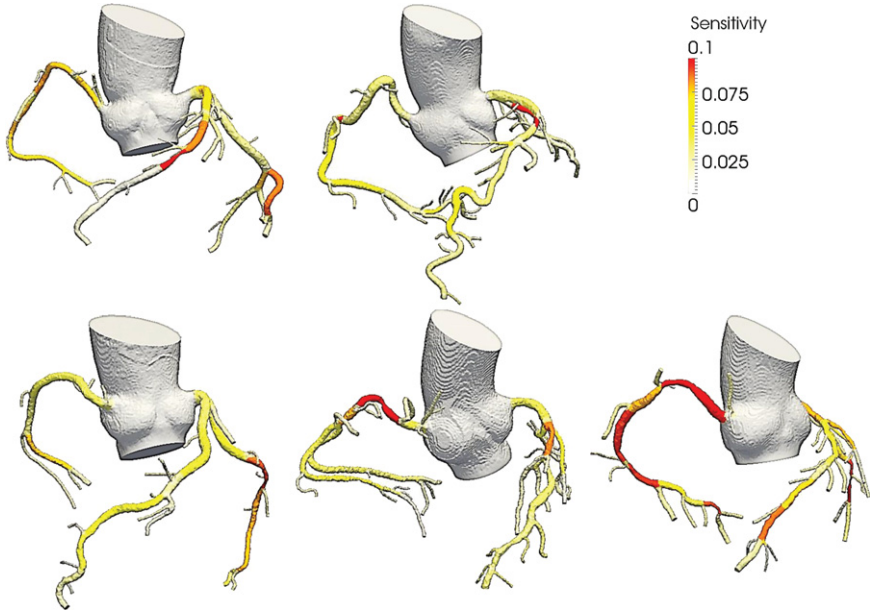

Fig. 11. Sensitivity values for five patients obtained using a 20% uniform increase in segmental area. High sensitivities are observed for diseased regions, segments off ostium that are critical to transport and boundary segments that control the boundary resistances.

not necessarily true for a healthy segment. Similarly, the effect of negative perturbation is significant if a vessel is healthy and it is critical to blood transport in coronary arteries. Contrast in the effect of segmental dilation and erosion on sensitivity for a patient is shown in Fig. 12. Terminal and small vessels show a higher sensitivity to dilation whereas all vessels critical to blood transport show a higher sensitivity to erosion. This is especially useful in long segments with focal lesions.

4. Discussion

Sensitivity information can provide insight into the fidelity of computed results, modeling assumptions and provide pointers to clinical information that could potentially be useful in improving accuracy of the computed fields. Due to a combination of large computational effort involved in patient-specific CFD simulations and a large number of uncertain parameters, existing methods are inadequate to compute sensitivities to geometry. We coupled an adaptive collocation method with a machine learning surrogate for CFD to accelerate convergence, and hence evaluate the

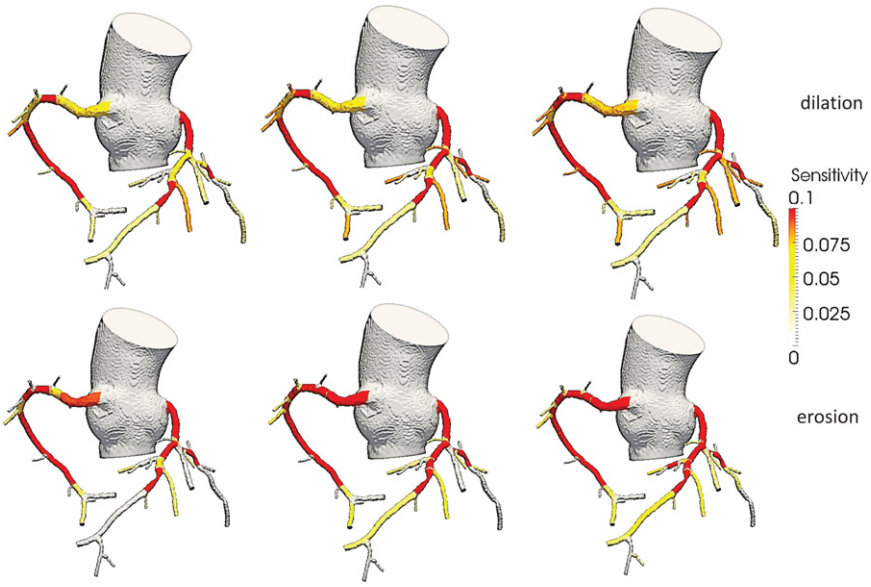

Fig. 12. Comparison of eroded and dilated sensitivity values corresponding to two extreme quadrature points on either side of mean (from left) 20%, 40%, and 60%, showing that sensitivity values are non-linear and asymmetric, with proximal vessels being more sensitive to erosion and diseased and distal vessels of small caliber being more sensitive to dilation.

sensitivity information in almost real time. We demonstrated that adaptive stochastic collocation method has a much faster convergence rate compared to traditional stochastic collocation and Monte-Carlo methods. While Monte-Carlo methods make no fundamental assumption about the stochastic space representation of the fundamental variables and might still be preferred for very high stochastic dimension (of the order of hundreds), collocation methods are able to achieve much faster convergence rates using sparse-grid quadrature methods combined with assumption of higher-order continuity. Functional adaptivity helps in improving the computational efficiency further. We demonstrated the above on a problem with four stochastic dimensions and believe that the trend is indicative of behavior in higher dimensions.

In the context of blood flow simulations in coronary arteries, the impact of change in geometry of a given coronary segment has two basic effects — (a) impact on blood flow through that segment and rest of the coronary tree, and (b) impact on pressure drop based on geometric resistance of the segment and (a). Blood flow rate through the entire model depends on lumped boundary resistance and geometric resistance of the model. The latter is usually negligible, except for diseased segments and vessels of small caliber. A representative plot of normalized radius versus flow rate is shown in Fig. 13. When the vessel is completely occluded, the flow rate is zero and segmental resistance is infinity. Gradually, as the radius is increased, the flow rate shows a non-linear increase (due to non-linear decrease in segmental resistance in serial with constant boundary resistance). Slowly, the lumped boundary resistance starts to dominate the segmental resistance. Similarly, if the lumen radius is large, segmental resistance is almost zero and hence flow is dominated by effective downstream boundary resistance. If the radius is gradually decreased from this value, the resistance increases and approaches the boundary resistance. Hence, segments with large radius which are critical to blood transport are more sensitive to erosion, and diseased and segments far from aorta are more sensitive to dilation. These phenomena were observed in the examples presented earlier.

We observed that diseased segments and arteries critical to blood transport show high sensitivity. The method presented here describes a computationally feasible method to obtain quantitative information about the impact of geometry on the solutions obtained from the Navier–Stokes equations. The proposed method takes only a few minutes to compute sensitivity information, and is able to encode all the geometrically significant information. There is a small loss in accuracy, but significant gain in computational time. Further, we are able to rank the different segments of the geometry as well as the clinical parameters.

The ideas presented in this paper are applicable to other hemodynamic or CFD systems. Other stochastic methods such as GPCE or simplex collocation can also be used instead of the adaptive stochastic collocation method used

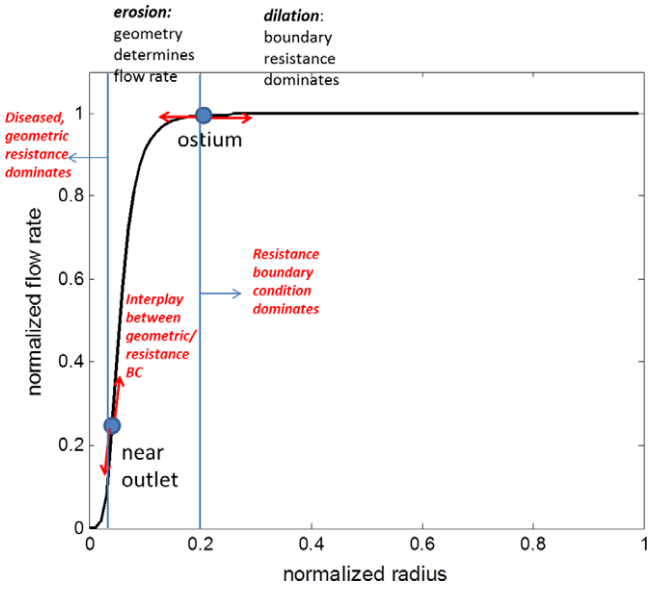

Fig. 13. Empirical relation between representative segmental radius between 0 and 1, and global flow rate (normalized to its maximum value) for a tube geometry with boundary resistance in series. The figure shows interaction between local geometry and boundary resistance, where patient-specific analysis is required.

here. The features used for the surrogate machine learning predictor can be used in other bio-fluids simulations such as cerebral flow, air flow in the lungs, or even simulations such as simulating flow through porous media and thermal transport. We expect more data to be needed to learn vector fields such as velocity or wall stress, however the results of the machine learning method show promise to be extended to other systems. We trained the machine learning algorithm for time averaged FFR_{CT}. It would be interesting to study if the same features will be sufficient for performing machine learning on three-dimensional velocity fields or if additional features are required. It is not clear if turbulence in transient velocity data can be captured using the features described here. Yet, we showed that with sufficient data, results from steady state low Reynolds number Navier–Stokes equation can be reasonably captured. The main potential limitation of this approach is that the stochastic space of geometries does not encompass the true geometry. However, due to our conservative approach in defining the stochastic space of geometries (uniform random variable), the true geometry is not captured only in the case when the initial lumen segmentation has a large error from the true segmentation. To mitigate this error, in our geometric modeling process (which is a controlled manufacturing process regulated by the FDA), analysts go through sections of the model to ensure that the initial lumen segmentation is as accurate as possible against the image data. We also only captured uncertainty in the lumen area, since minimum lumen diameter is the most clinically relevant measure that is hypothesized to drive large pressure changes. However, uncertainty in surface profiles also needs to be investigated, along with the development of a geometry parameterization tool for surfaces (e.g. NURBS). The latter can also help define deformation maps at bifurcation locations for the 3D geometry. Future work could also include defining uncertainty based on local image quality. For example, locations of the coronary tree close to artifacts like blooming could have higher uncertainty. This can also be used to quantify if uncertainty in two regions are correlated. We also need to validate the sensitivities using clinical data. We can achieve this by calculating, say 95% confidence intervals and ensuring that 95% of measurement points lies within this. This, however, requires many patient-specific data and should be investigated in the future.

Appendix A. Healthy radius estimate

Calculation of estimate of healthy radius is performed using three different kernel regressors, with five kernel sizes for each, making a total of 15 features, as described below.

- global kernel fit, defined for each path from the root (ostium) to the leaves, where the healthy radius is given by

$$r^{i}_{\texttt{healthy}}(x) = \frac{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) w_{x',i} r_{x'}}{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) w_{x',i}}$$

  where $w_{x,i} = \mathcal{N}(r_x|r_{x,\texttt{max},i}, \nu_{\texttt{max},i})$, C_l and C_u are lower and upper centerline indices in the current path, i goes from 1 to 5, and w_x are weighting functions.
- segmental fit defined for each segment between branches, where the healthy radius is given by

$$r^{i+5}_{\texttt{healthy}}(x) = \frac{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) I(x', x) w_{x',i} r_{x'}}{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) I(x', x) w_{x',i}}$$

  where $I(x', x) = 1$ if there are no bifurcations between x' and x, i goes from 1 to 5, and $I(x', x) = 0$ otherwise.
- anisotropic kernel fit which is defined for each path from the root to the leaves, but weighted with a sigmoidal function centered at the nearest ostium designed to minimize the effect of sharp radius variation at the branch.

$$r^{i+10}_{\texttt{healthy}}(x) = \frac{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) S(x', x, i) w_{x',i} r_{x'}}{\sum_{x'=C_l}^{C_u} \mathcal{N}(x'|x, \nu_{x,i}) S(x', x, i) w_{x',i}}$$

  where the sigmoidal function, S is given by:

$$S(x', x, i) = \frac{1}{1 + 3e^{-k_i d_{\texttt{offset}}(x', x)}}$$

  and

$$d_{\texttt{offset}}(x', x) = d(x', x_{\texttt{ostium}}) - d(x, x_{\texttt{ostium}}) - d(x, x_{\texttt{up}}),$$

  where $x_{\texttt{up}}$ is the location of the nearest upstream branch to x.

Five paired parameter set for $i = 1, 2, 3, 4, 5$ were chosen for each of these regressors, making a total of 15 health index scores. These are given by $\nu_{x,i} = 6 * (1 + (i - 3) * 0.4)$, $\nu_{\texttt{max},i} = 200 * (1 + (i - 3) * 0.4)$, $r_{x,\texttt{max},i} = 0.25 * (1 + (i - 3) * 0.4)$ and $k_i = 0.1 * (1 + (i - 3) * 0.4)$. Accurate description of disease is important to model the appropriate pressure drop across a diseased vessel which directly impacts FFR_{CT}. Anisotropic kernel fit is better suited for vessels with numerous branches where radius has step changes numerous times. Global kernel fit could over-predict healthy radius for terminal vessels, and hence predict disease even in the absence of one. The different kernels are required to capture different disease lengths from focal (very small disease length) to diffuse (very large segment length which could go across multiple vessels). However, global kernel fit performs better for the main coronary arteries.

Appendix B. Reduced order model

A reduced order model is used as a feature for the machine learning algorithm. First, the net downstream boundary condition for a segment is given by

$$\frac{1}{R_{\texttt{eff}}(x)} = \sum_{i \in n_{\texttt{down}}(x)} \frac{1}{R_i}$$

where $n_{\texttt{down}}(x)$ is the number of downstream outlets at location x, and R_i is the downstream resistance of each outlet. To calculate the resistance for each outlet, we assume that the resistance is inversely proportional to the area,

hence

$$\frac{1}{R_{\mathtt{eff}}} = \sum_{i \in n_{\mathtt{down}}(x)} \frac{A_i}{k},$$

where k can be found by substituting for $R_{\mathtt{eff}} = \frac{P_{\mathtt{aorta}}}{q_{\mathtt{cor}}}$ at the ostium ($x = x_{\mathtt{ostium}}$), $P_{\mathtt{aorta}} = (2P_{\mathtt{diastole}} + P_{\mathtt{systole}})/3$. Three flow rate features are used, $Q_1 = \frac{P_{\mathtt{aorta}}}{R_{\mathtt{eff}}}$, $Q_2(x) = q_{\mathtt{cor}} \frac{R_{\mathtt{eff}}(x_{\mathtt{root}})}{R_{\mathtt{eff}}(x)}$ and $Q_3(x) = q_{\mathtt{cor}} \frac{R_{\mathtt{eff}}(x_{\mathtt{root}}) + R_{\mathtt{geom}}(x_{\mathtt{root}})}{R_{\mathtt{eff}}(x) + R_{\mathtt{geom}}(x)}$ where $R_{\mathtt{geom}}$ is the net geometric resistance downstream of a given point, obtained by cumulating the geometric resistance of all segments between centerline points. The geometric resistance is obtained using Poiseuille's law as $R_{\mathtt{geom}} = \frac{8\mu L}{\pi r_{\mathtt{avg}}^4}$.

Second, we model energy loss along the model using three components — a pressure loss model, a stenotic index, and a pressure recovery feature. Stenotic index is defined which is 1 if the average health index is less than 0.5. Pressure loss feature depends on the stenotic index. If stenotic index is 0, then Poiseuille equation is used to model pressure loss as

$$\Delta P_i = \frac{8\mu L Q_i}{\pi r_{\mathtt{avg}}^4}, \quad i = 1, 2, 3.$$

If stenotic index is 1, then a modified Poiseuille equation is used to model higher pressure loss in diseased locations as

$$\Delta P_i = \frac{8\mu L Q_i}{\pi r_{\mathtt{avg}}^3}, \quad i = 1, 2, 3.$$

Hence, three pressure loss indices are computed for each of the flow rates calculated in the previous section. An estimated $\mathrm{FFR_{CT}}$ is calculated corresponding to each of these pressures. It is hypothesized that diseased regions might have a different power law relation between pressure and flow-rate. Instead of including additional features to model such possibilities, we include the log of the flow rate and log of the pressure drop as features.

We also introduce a pressure recovery factor, based on observations in a few patients, where pressure can recover distal to a disease if the radius returns back to higher than the original healthy value and remains there. We define a cutoff of 20% (chosen empirically), i.e. if the radius increases to more than 20% of its pre-stenotic healthy value, then we define the pressure recovery factor to be $P_{\mathtt{recovery}} = \frac{r_{\mathtt{post}}}{r_{\mathtt{pre}}}$, where $r_{\mathtt{post}}$ is the post-stenotic radius and $r_{\mathtt{pre}}$ is the pre-stenotic radius.

References

[1] C.A. Taylor, M.T. Draney, J.P. Ku, D. Parker, B.N. Steele, K. Wang, C.K. Zarins, Predictive medicine: Computational techniques in therapeutic decision-making, Comput. Aided Surg. 4 (5) (1999) 231–247.

[2] C.A. Taylor, M.T. Draney, Experimental and computational methods in cardiovascular fluid mechanics, Annu. Rev. Fluid Mech. 36 (2004) 197–231.

[3] C.A. Taylor, T.J.R. Hughes, C.K. Zarins, Finite element modeling of blood flow in arteries, Comput. Methods Appl. Mech. Engrg. 158 (1–2) (1998) 155–196.

[4] C.A. Figueroa, S. Baek, C.A. Taylor, J.D. Humphrey, A computational framework for fluid-solid-growth modeling in cardiovascular simulations, Comput. Methods Appl. Mech. Engrg. 198 (2008) 3583–3602.

[5] J.D. Humphrey, K. Rajagopal, A constrained mixture model for growth and remodeling of soft tissues, Math. Models Methods Appl. Sci. 12 (3) (2002) 407–430.

[6] M. Kroon, G.A. Holzapfel, A model for saccular cerebral aneurysm growth by collagen fibre remodeling, J. Theoret. Biol. 247 (2007) 775–787.

[7] P.N. Watton, N.A. Hill, M. Heil, A mathematical model for the growth of the abdominal aortic aneurysm, Biomech. Model. Mech. Biol. 3 (2004) 98–113.

[8] C.A. Taylor, J.D. Humphrey, Open problems in computational vascular biomechanics: Hemodynamics and arterial wall mechanics, Comput. Methods Appl. Mech. Engrg. 198 (45–46) (2009) 3514–3523.

[9] C.A. Taylor, T.A. Fonte, J.K. Min, Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: scientific basis, J. Am. Coll. Cardiol. 61 (22) (2013) 2233–2241.

[10] B.L. Norgaard, J. Leipsic, S. Gaur, S. Seneviratne, B.S. Ko, H. Ito, J.M. Jensen, L. Mauri, B. De Bruyne, H. Bezerra, K. Osawa, M. Marwan, C. Naber, A. Erglis, S.J. Park, E.H. Christiansen, A. Kaltoft, J.F. Lassen, H.E. Btker, S. Achenbach, Diagnostic performance of non-invasive fractional flow reserve derived from coronary ct angiography in suspected coronary artery disease: The NXT trial, J. Am. Coll. Cardiol. 63 (12) (2014) 1145–1155.

[11] J.K. Min, J. Leipsic, M.J. Pencina, D.S. Berman, B.K. Koo, C. ven Mieghem, A. Erglis, F.Y. Lin, A.M. Dunning, P. Apruzzese, M.J. Budoff, J.H. Cole, F.A. Jaffer, M.B. Leon, J. Malpeso, G.B. Mancini, S.J. Park, R.S. Schwarts, L.J. Shaw, L. Mauri, Diagnostic accuracy of fractional flow reserve from anatomic CT angiography, J. Am. Med. Assoc. 308 (12) (2012) 1237–1245.
[12] B.K. Koo, A. Erglis, J.H. Doh, D.V. Daniels, S. Jegere, H.S. Kim, A. Dunning, T. DeFrance, A. Lansky, J. Leipsic, J.K. Min, Diagnosis of ischemia-causing coronary stenoses by noninvasive fractional flow reserve computed from coronary computed tomographic angiograms results from the prospective multicenter DISCOVER-FLOW (diagnosis of ischemia-causing stenoses obtained via noninvasive fractional flow reserve) study, J. Am. Coll. Cardiol. 58 (19) (2011) 1989–1997.
[13] D.P. Giddens, C.K. Zarins, S. Glagov, The role of fluid mechanics in localization and detection of atherosclerosis, J. Biomech. Eng. 115 (1993) 588–594.
[14] S. Sankaran, A.L. Marsden, The impact of uncertainty on shape optimization of idealized bypass graft models in unsteady flow, Phys. Fluids 22 (2010) 121902.
[15] A.L. Marsden, A.D. Bernstein, V.M. Reddy, S. Shadden, R. Spilker, F.P. Chan, C.A. Taylor, J.A. Feinstein, Evaluation of a novel y-shaped extracardiac Fontan baffle using computational fluid dynamics, J. Thorac. Cardiovasc. Surg. 137 (394–403) (2009).
[16] W. Yang, J.A. Feinstein, A. Marsden, Constrained optimization of an idealized y-shaped baffle for the Fontan surgery at rest and exercise, Comput. Methods Appl. Mech. Engrg. 199 (33–36) (2010) 2135–2149.
[17] A.L. Marsden, A.J. Bernstein, R.L. Spilker, F.P. Chan, C.A. Taylor, J.A. Feinstein, Large differences in efficiency among Fontan patients demonstrated in patient specific models of blood flow simulations, Circulation 116 (16) (2007) 480–480.
[18] N. Wilson, K. Wang, R. Dutton, C.A. Taylor, A software framework for creating patient specific geometric models from medical imaging data for simulation based medical planning of vascular surgery, Lecture Notes Comput. Sci. 2208 (2001) 449–456.
[19] J.A. Moore, D.A. Steinman, C.R. Ethier, Computational blood flow modelling: Errors associated with reconstructing finite element models from magnetic resonance images, J. Biomech. 31 (1998) 1–12.
[20] J.A. Moore, D.A. Steinman, D.W. Holdsworth, C.R. Ethier, Accuracy of computational hemodynamics in complex arterial geometries reconstructed from magnetic resonance imaging, Ann. Biomed. Eng. 27 (1999) 32–41.
[21] S.W. Lee, D.A. Steinman, On the relative importance of rheology for image-based CFD models of the carotid bifurcation, J. Biomech. Eng. 129 (2) (2007) 273–278.
[22] S. Sankaran, A.L. Marsden, A stochastic collocation method for uncertainty quantification and propagation in cardiovascular simulations, J. Biomech. Eng. 133 (2011) 031001–1.
[23] N.H. Pijls, B. De Bruyne, K. Peels, et al., Measurement of fractional flow reserve to assess the functional severity of coronary-artery stenoses, N. Engl. J. Med. 334 (1996) 1703–1708.
[24] N.H. Pijls, P. van Schaardenburgh, G. Manoharan, et al., Percutaneous coronary intervention of functionally nonsignificant stenosis: 5-year follow-up of the DEFER study, J. Am. Coll. Cardiol. 49 (2007) 2105–2111.
[25] P.A. Tonino, B. De Bruyne, N.H. Pijls, et al., Fractional flow reserve versus angiography for guiding percutaneous coronary intervention, N. Engl. J. Med. 360 (2009) 213–224.
[26] B. De Bruyne, N.H. Pijls, B. Kalesan, et al., Fractional flow reserve-guided pci versus medical therapy in stable coronary disease, N. Engl. J. Med. 367 (2012) 991–1001.
[27] D. Xiu, J.S. Hesthaven, High order collocation methods for the differential equation with random inputs, J. Sci. Comput. 27 (2005) 1118–1139.
[28] I. Babuska, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. ICES report, 2005, pp. 05–47.
[29] P.G. Constantine, M.S. Eldred, E.T. Phipps, Sparse pseudospectral approximation method, Comput. Methods Appl. Mech. Engrg. 229 (232) (2012) 1–12.
[30] H.N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, Annu. Rev. Fluid Mech. 41 (2009) 35–52.
[31] D. Xiu, G.E. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, Comput. Methods Appl. Mech. Engrg. 191 (2002) 4927–4948.
[32] A. Doostan, R.G. Ghanem, J. Red-Horse, Stochastic model reduction for chaos representations, Comput. Methods Appl. Mech. Engrg. 196 (37) (2007) 3951–3966.
[33] G.J.A. Loeven, J.A.S. Witteveen, H. Bijl, Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties, in: Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno (NV), United States, 2007, pp. 2007–317.
[34] J. Witteveen, G. Iaccarino, Refinement criteria for simplex stochastic collocation with local extremum diminishing robustness, SIAM J. Sci. Comput. 34 (3) (2012) A1522–A1543.
[35] G. Iaccarino, P. Constantine, Large eddy simulations of flow around a cylinder with uncertain wall heating, in: 47th AIAA Aerospace Sciences Meeting, Orlando, Florida, AIAA Paper No. 2009-975, 2009.
[36] H.J. Kim, I.E. Vignon-Clementel, J.S. Coogan, C.A. Figueroa, K.E. Jansen, C.A. Taylor, Patient-specific modeling of blood flow and pressure in human coronary arteries, Ann. Biomed. Eng. 38 (10) (2010) 3195–3209.
[37] Y. Bazilevs, V.M. Calo, T.J.R. Hughes, Y. Zhang, Isogeometric fluid–structure interaction: theory, algorithms, and computations, Comput. Mech. 43 (1) (2008) 3–37.
[38] V. Kalro, T.E. Tezduyar, A parallel 3D computational method for fluid structure interactions in parachute systems, Comput. Methods Appl. Mech. Engrg. 190 (3) (2000) 321–332.
[39] T.J.R. Hughes, A.C. John, Y. Bazilevs, Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement, Comput. Methods Appl. Mech. Eng. 194 (39) (2005) 4135–4195.
[40] K.E. Jansen, C.H. Whiting, G.M. Hulbert, A generalized-alpha method for integrating the filtered Navier Stokes equations with a stabilized finite element method, Comput. Methods Appl. Mech. Engrg. 190 (3–4) (2000) 305–319.
[41] A. Klimke, Uncertainty modeling using fuzzy arithmetic and sparse grids (Ph.D. thesis), Universitt Stuttgart, Shaker Verlag, Aachen, 2006.

[42] T. Gerstner, M. Griebel, Numerical integration using sparse grids, Numer. Algorithms 18 (3) (1998) 209–232.
[43] R. Shahzad, et al., Detection and quantification of coronary artery stenoses on CTA, Int. J. Cardiovasc. Imaging 29 (2013) 1–13.
[44] S. Sankaran, L. Grady, C.A. Taylor, Fast computation of hemodynamic sensitivity to lumen segmentation uncertainty, IEEE Trans. Med. Imaging (2015) in press.
[45] M. Hall, et al., The WEKA data mining software: An update, SIGKDD Explorations 11 (1) (2009).
[46] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.