

MIT Open Access Articles

Enhancing physiologic simulations using supervised learning on coarse mesh solutions

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Kolandaivelu, Kumaran, Caroline C. O'Brien, Tarek Shazly, Elazer R. Edelman, and Vijaya B. Kolachalama. "Enhancing Physiologic Simulations Using Supervised Learning on Coarse Mesh Solutions." *Journal of The Royal Society Interface* 12, no. 104 (February 4, 2015): 20141073–20141073.

As Published: <http://dx.doi.org/10.1098/rsif.2014.1073>

Publisher: Royal Society Publishing

Persistent URL: <http://hdl.handle.net/1721.1/102946>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution



INTERFACE

rsif.royalsocietypublishing.org

Research



Cite this article: Kolandaivelu K, O'Brien CC, Shazly T, Edelman ER, Kolachalama VB. 2015 Enhancing physiologic simulations using supervised learning on coarse mesh solutions.

J. R. Soc. Interface **12**: 20141073.

<http://dx.doi.org/10.1098/rsif.2014.1073>

Received: 25 September 2014

Accepted: 15 January 2015

Subject Areas:

bioengineering, biomedical engineering, mathematical physics

Keywords:

computational modelling, machine learning, Gaussian process, nearest neighbours, drug-coated balloons, drug-eluting stents

Author for correspondence:

Vijaya B. Kolachalama

e-mail: vkolachalama@draper.com

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.1073> or via <http://rsif.royalsocietypublishing.org>.

THE ROYAL SOCIETY
PUBLISHING

Enhancing physiologic simulations using supervised learning on coarse mesh solutions

Kumaran Kolandaivelu^{1,3}, Caroline C. O'Brien¹, Tarek Shazly⁴, Elazer R. Edelman^{1,2,3} and Vijaya B. Kolachalama^{1,5}

¹Institute for Medical Engineering and Science, and ²David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Cardiovascular Division, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA

⁴College of Engineering and Computing, University of South Carolina, Columbia, SC 29208, USA

⁵Charles Stark Draper Laboratory, 555 Technology Square, Cambridge, MA 02139, USA

Computational modelling of physical and biochemical processes has emerged as a means of evaluating medical devices, offering new insights that explain current performance, inform future designs and even enable personalized use. Yet resource limitations force one to compromise with reduced order computational models and idealized assumptions that yield either qualitative descriptions or approximate, quantitative solutions to problems of interest. Considering endovascular drug delivery as an exemplary scenario, we used a supervised machine learning framework to process data generated from low fidelity coarse meshes and predict high fidelity solutions on refined mesh configurations. We considered two models simulating drug delivery to the arterial wall: (i) two-dimensional drug-coated balloons and (ii) three-dimensional drug-eluting stents. Simulations were performed on computational mesh configurations of increasing density. Supervised learners based on Gaussian process modelling were constructed from combinations of coarse mesh setting solutions of drug concentrations and nearest neighbourhood distance information as inputs, and higher fidelity mesh solutions as outputs. These learners were then used as computationally inexpensive surrogates to extend predictions using low fidelity information to higher levels of mesh refinement. The cross-validated, supervised learner-based predictions improved fidelity as compared with computational simulations performed at coarse level meshes—a result consistent across all outputs and computational models considered. Supervised learning on coarse mesh solutions can augment traditional physics-based modelling of complex physiologic phenomena. By obtaining efficient solutions at a fraction of the computational cost, this framework has the potential to transform how modelling approaches can be applied in the evaluation of medical technologies and their real-time administration in an increasingly personalized fashion.

1. Introduction

Emerging applications in clinical medicine and medical device design are just beginning to embrace the potential of computational modelling [1,2]. In the face of increasing interest by medical device companies, regulatory agencies [3] and clinicians [4,5], it is critical that we find approaches that yield accurate information quickly—whether it is in the optimization of device design over a wide parameter space [6–9], to better understand device and implant performance in patient-specific environments [10–16], or even to perform virtual feasibility studies for procedural planning (i.e. implantation of stents, endovascular grafts, heart valves, bioprosthetics, etc.) [17–20]. Particularly in clinical applications, speed and efficiency are paramount as results must be provided

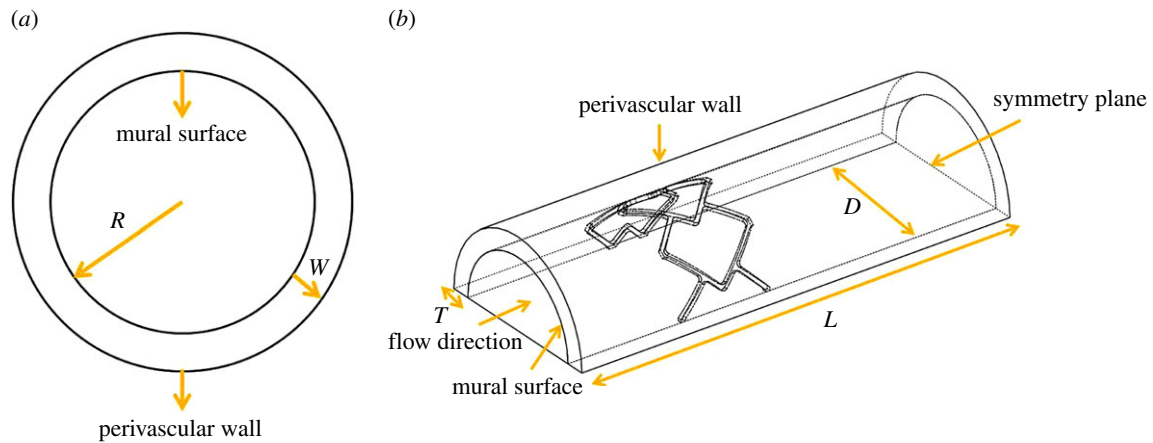


Figure 1. Schematics of the two computational models used for the studies. (a) A two-dimensional model of an expanded drug-coated balloon is shown that transfers drug to the tissue-lumen (or mural) surface. $R = 3$ mm indicates the lumen radius and $W = 0.5$ mm is the arterial wall thickness. (b) A three-dimensional model of a single cell, delta-winged stent deployed in an arterial vessel is shown with drug releasing from the stent surface. Here, the lumen diameter was defined as $D = 3.6$ mm, arterial wall thickness as $T = 0.35$ mm, domain length as $L = 11.36$ mm. The upstream section length relative to the stent position was defined as 3 mm and the downstream section length as 6 mm. (Online version in colour.)

cost-effectively and often in urgent settings (i.e. within minutes). Historically, efficiency and practicality have trumped accuracy due to resource limitations associated with computational hardware and software as well as access to relevant information. The last decades have witnessed a remarkable explosion of medical knowledge (both multi-scale and multi-physics) as well as growth in our ability to access and store data in real-time, patient-specific formats [21]. Yet processing this knowledge has outstripped the exponential growth in high-performance computing and even the most promising, clinically relevant applications to date remain fundamentally limited by computational delays [22]. To fully leverage the possibilities promised by simulation, it is critical that we find methods that better handle information without loss of efficiency or accuracy.

Cardiovascular applications are among the most important and widely studied cases of medical simulation. Devices such as drug-eluting stents [23,24] and balloon catheters [25–28] are used millions of times each year to help manage coronary heart disease, the leading cause of global mortality [29]. In this disease, blockages build up in the blood vessels supplying the heart and must be reopened to prevent untoward cardiovascular events. Both disease progression as well as the success and failure of such devices depend on local physical and biochemical processes coupled within complex physiological environments. While local delivery of drugs helps modulate device response, the drugs themselves are highly toxic (often chemotherapeutics) characterized by a narrow therapeutic window—delivery of too little compromises effectiveness; delivery of too much compromises safety [30,31]. Designing and achieving optimal device performance requires that we understand the confluence of fluid flow, drug transport and tissue reactivity coupled within relevant, often patient-specific environments.

We sought to determine whether supervised machine learning can be leveraged to estimate higher fidelity data from lower fidelity computational datasets, considering two computational mesh-based simulations of distinct computational complexities. The first is a two-dimensional time-dependent model of drug release, transport and reversible tissue binding from a drug-coated balloon (figure 1a); the

second a steady-state simulation of blood flow coupled with species transport and reversible tissue binding in a three-dimensional model of a drug-eluting stent deployed in an arterial vessel (figure 1b). Using Gaussian process modelling (GPM) [32], we constructed supervised learners that used lower resolution mesh results to predict solutions of higher mesh density. Computational simulations were performed on a low fidelity, mesh configuration (baseline) comprised of variably sized elements that captured underlying physics with reasonable accuracy as well as on mesh configurations of increasing density (Refinements 1, 2 and 3). Nearest neighbour information of a reference mesh node for a low fidelity configuration (baseline) along with low fidelity solutions at these nearest neighbours and at the reference mesh node defined the input feature vector for the supervised learner. With the true solution of the reference mesh node at the intermediate mesh setting (Refinement 1) serving as scalar output, GPM was used to construct computationally inexpensive surrogates to predict higher fidelity solutions (Refinement 2); these predictions were cross-validated using the true solutions obtained from the highest mesh density (Refinement 3). Our results demonstrate the potential of applying supervised machine learning techniques to approximate physics-based solutions both cost-effectively and accurately, providing a paradigm for emerging applications in medicine.

2. Material and methods

2.1. Model of drug-coated balloon therapy

Drug-coated balloons have recently emerged as viable therapeutic options for treating obstructive arterial disease [25–28]. With this technology, short-term transfer of therapeutic drugs to the arterial wall can be achieved without the requirement of permanent indwelling delivery systems. Transient arterial tissue distribution from an expanded drug-coated balloon was modelled as a two-dimensional, time-dependent continuum transport problem. The computational domain constituted an arterial cross-section with radius ($R = 3$ mm) and wall thickness ($W = 0.5$ mm) (figure 1a). Free drug was allowed to diffuse with

a constant diffusivity (D_w) and reversibly bind to tissue sites according to the reaction–diffusion equation

$$\frac{\partial C}{\partial t} = D_w \nabla^2 C - k_a C(B_M - B) + k_d B, \quad (2.1)$$

where C and B denote the local concentrations of free and bound drug in the arterial wall, respectively. $D_w = 1.712 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ is the apparent net diffusivity, $B_M = 0.356 \text{ mmol l}^{-1}$ is the net tissue binding capacity [33], and k_a and k_d are the association and dissociation rate constants, respectively, for the model drug (zotarolimus). k_a and k_d were computed as $k_a = D_w D_a / B_M W^2$ and $k_d = k_a k_d$, where $D_a = 50\,000$ is the Damköhler number and $k_d = 0.0326 \text{ mmol l}^{-1}$ is the equilibrium dissociation constant [33]. Equation (2.1) was solved subject to zero initial free and bound drug concentrations within the tissue, a perfect sink condition at the adventitial surface and a flux boundary condition at the mural surface defined as

$$J_b(t) = \begin{cases} \frac{k_1 A_1 e^{-k_1 t}}{Z_{MW}} & t \leq t_0 \\ 0 & t > t_0, \end{cases} \quad (2.2)$$

where $J_b(t)$ is the flux approximating the releasable portion of zotarolimus from the balloon during inflation, $t_0 = 30 \text{ s}$ is the balloon inflation time, $A_1 = 23.95 \text{ kg m}^{-3}$ and $k_1 = 0.009208 \text{ s}^{-1}$ are empirical constants estimated from bench-top release kinetics experiments [33], and $Z_{MW} = 966.21 \text{ g mol}^{-1}$ is the molecular weight of zotarolimus. A zero concentration condition was applied on the perivascular side of the arterial wall for the free drug. For the bound drug, both the lumen-tissue (or mural) and the perivascular aspects of the arterial wall were assigned a zero flux boundary condition. Time-dependent simulations (COMSOL 4.3a, Comsol Inc.) were performed on the computational domain that was meshed using the Delaunay triangulation scheme. Three model configurations were then constructed by sequentially refining the entire domain using longest edge refinement technique, where the longest edge of each mesh element is bisected at each of the three levels, defined as Refinements 1, 2 and 3 (true solution), respectively (table 1a). The Direct (SPOOLES) method was used to solve the system of equations with a nested dissection pre-ordering algorithm and a backward differentiation formula method was used for time stepping with relative and scaled absolute tolerances assigned at 1×10^{-9} and 1×10^{-6} , respectively. Simulations were performed until the minimum damping and tolerance factors reached 1×10^{-9} and 1×10^{-6} , respectively. Arterial tissue distributions of free and bound drug for three model configurations of increasing mesh density were extracted at 1 h from balloon inflation for supervised learning.

2.2. Model of stent-based drug therapy

Unlike drug-coated balloons, the metallic drug-eluting stents remain permanently implanted at the lesion site but can provide sustained release of therapeutic drugs. Owing to their remarkable clinical success, they are considered as the primary choice for treating coronary artery disease [23,24]. A three-dimensional computational model of a drug-eluting stent deployed in a non-bifurcating arterial vessel was constructed (SolidWorks, Dassault Systèmes) [34,35]. The diameter of the arterial vessel was defined as $D = 3.6 \text{ mm}$, arterial wall thickness as $T = 0.35 \text{ mm}$ and vessel length as $L = 11.36 \text{ mm}$. A fully apposed single cell, delta-wing shaped slotted tube design was used for the stent with diameter 3.5 mm , and the intrinsic strut shape modelled as a square with dimensions $10^{-4} \times 10^{-4} \text{ m}^2$. The effects of pulsatile flow were approximated using a steady-state flow equivalent within this artery–stent combination, as this approximation was used to produce a physiologically realistic assessment of the impact of flow on arterial drug distribution [36]. In the arterial lumen,

Table 1. Mesh configurations of increasing density for (a) two-dimensional time-dependent model simulating drug-coated balloon therapy and (b) three-dimensional steady-state model of drug-eluting stent therapy. Mesh refinement was performed by bisecting the longest edge of each element. The increase in computational time (rounded to the tenths place) of the true solution scales with mesh density. Simulations were performed on a dual-core Intel Xeon X5687 @ 3.6 GHz processor with an installed memory of 72 GB.

mesh setting	cell count	degrees of freedom	simulation time (hours)
<i>(a)</i>			
baseline	5302	5890	0.8
Refinement 1	11 326	12 500	2.5
Refinement 2	23 861	25 042	3.8
Refinement 3	49 829	52 178	5.8
<i>(b)</i>			
baseline	83 397	211 981	0.3
Refinement 1	333 445	848 238	2.6
Refinement 2	1 077 508	2 709 490	65.7

the continuity and momentum equations

$$\nabla \cdot \mathbf{v}_f = 0 \quad (2.3)$$

and

$$\rho_f (\mathbf{v}_f \cdot \nabla \mathbf{v}_f) = -\nabla P + \mu_f \nabla^2 \mathbf{v}_f, \quad (2.4)$$

respectively were solved where \mathbf{v}_f , $\rho_f = 1060 \text{ kg m}^{-3}$, P and $\mu_f = 3.5 \times 10^{-3} \text{ Pa} \cdot \text{s}$ are, respectively, the velocity, density, pressure and the viscosity of flowing blood. The arterial wall was assumed to be a porous medium where the continuity equation

$$\nabla \cdot \mathbf{v}_t = 0 \quad (2.5)$$

was solved, where \mathbf{v}_t is the interstitial fluid velocity. The momentum equation

$$\frac{\rho_t}{\epsilon} \left((\mathbf{v}_t \cdot \nabla) \frac{\mathbf{v}_t}{\epsilon} \right) = -\nabla P + \frac{\mu_t}{\epsilon} \nabla^2 \mathbf{v}_t - \left(\frac{\mu_t}{\epsilon K} \right) \mathbf{v}_t \quad (2.6)$$

was assumed to follow Darcy's Law, where $K = 1.43 \times 10^{-18} \text{ m}^2$ is Darcy's wall permeability, $\epsilon = 0.43$ is wall porosity, $\rho_t = 1000 \text{ kg m}^{-3}$ is fluid density and $\mu_t = 8.9 \times 10^{-4} \text{ Pa} \cdot \text{s}$ fluid viscosity within the arterial wall [37]. A constant velocity profile was prescribed at the luminal inlet. At the outlet, a zero pressure boundary condition was set. No-slip boundary conditions were imposed on the strut-blood and mural interfaces. The inlet condition fixed at a constant Reynolds number $\Re = \langle \mathbf{v}_f \rangle D / ((\mu_t / \rho_t)) \approx 282$ was based on mean blood flow and diameter measurements obtained from the human left anterior descending coronary artery [37].

Drug transport in the lumen was modelled as an advection–diffusion process defined as

$$\mathbf{v}_f \cdot \nabla C_f = D_f \nabla^2 C_f, \quad (2.7)$$

where C_f denotes lumen drug concentration and $D_f = 3.89 \times 10^{-11} \text{ m}^2 \text{ s}^{-1}$ is the model drug (paclitaxel) diffusivity in the lumen [37,38]. Drug transport in the arterial wall follows an advection–diffusion–reaction model as follows:

$$\mathbf{v}_t \cdot \nabla C_t = D_t \nabla^2 C_t - k_a C_t (B_M - B) + k_d B, \quad (2.8)$$

where $D_t = 3.65 \times 10^{-12} \text{ m}^2 \text{ s}$ is paclitaxel diffusivity in the arterial wall [37]. C_t and B denote the local concentrations of free and bound drug in the arterial wall, respectively, $B_M = 1.3 \text{ mM}$ is the net tissue binding capacity [33], and k_a and k_d are the association and dissociation rate constants, respectively. Flux continuities for

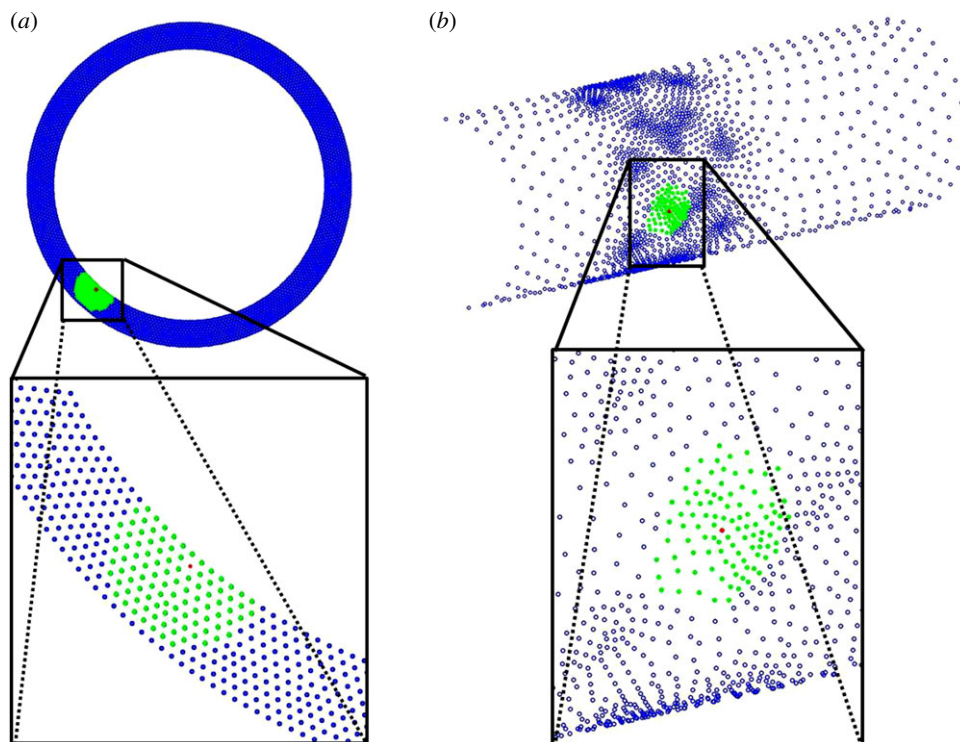


Figure 2. Schematic representing a finite number of nearest neighbours (denoted in green) that were extracted for a specific node (denoted in red) on the computational domains for the model simulating drug-coated balloon therapy (a) and the one simulating drug-eluting stent therapy (b). Computational mesh nodes were denoted in blue.

drug transport were maintained at the mural interface. k_a and k_d were computed as $k_a = D_i Da / B_M T^2$ and $k_d = k_a K_d$, where $Da = 2700$ is the Damköhler number and $k_d = 0.136$ mM is the equilibrium dissociation constant for paclitaxel [33].

A finite-element solver (COMSOL 4.3a, Comsol Inc.) was used to perform the coupled flow and drug transport simulations. The three-dimensional computational domain was discretized using tetrahedral control volumes with a linear shape function into an initial mesh (baseline). The mesh density was highest near the strut and the mural surface and then decreased towards the centreline of the lumen. Sequential mesh refinement using the longest edge bisection method was then performed to create cases with refined mesh densities (table 1b). An open boundary condition for drug concentration was applied at the inlet and the outlet of the arterial lumen. An impermeable boundary condition was established at the perivascular aspects of the model vessel. Stent drug release was simulated using a Dirichlet boundary condition of unit concentration. For the bound drug, the intramural and the perivascular aspects of the arterial wall, tissue inlet and tissue outlet were all assigned a zero flux boundary condition. Symmetry boundary conditions were exploited on the model geometry as the stent location was far from the lumen centreline. The physiologic rates of luminal blood flow coupled with transport of low diffusive drugs generate high Peclet numbers within the computational domain. An upwind Petrov–Galerkin streamline diffusion formulation was therefore enforced with full residual to stabilize the governing equations. Simulations were performed for computational configurations of increasing mesh densities and until there was a 1×10^{-9} reduction in the mass transport residuals for each case. The solutions and mesh information for all cases were subsequently processed for supervised learning.

2.3. Supervised learning framework

Supervised learning is a branch of machine learning where functions are inferred from labelled training data comprising a set of variables called inputs and their corresponding outputs. The outputs can be qualitative/categorical or quantitative in nature, with

the corresponding learning problem denoted as ‘classification’ or ‘regression’, respectively [39,40]. In this paper, a construct for the training data defined by an input vector $\mathbf{x} \in \mathbb{R}^p$ trained with a scalar output $y(\mathbf{x})$ was used for supervised learning. Numerical solutions on the mesh configurations of increasing density followed by computation of the pairwise Euclidean distances for each node’s nearest neighbours (figure 2a,b) within each computational mesh configuration were obtained [41]. During model training, each node (x_i) and the nearest neighbour (x_j) have a pairwise distance

$$d_{ij} = \sqrt{(x_i - x_j)(x_i - x_j)^T}, \quad (2.9)$$

for the baseline mesh configuration. Note that $j \in [1, 2, \dots, J]$, where J denotes the number of nearest neighbours. These pairwise distances along with the corresponding output for the baseline mesh configuration

$$\eta_j^0 = [y^0(x_1), y^0(x_2), \dots, y^0(x_j)], \quad (2.10)$$

were included as part of the input feature space $\mathbf{x}_i = \{\mathbf{d}_{ij}, y^0(x_i), \eta_j^0\}$, where $\mathbf{d}_{ij} = [d_{i1}, d_{i2}, \dots, d_{ij}]$ with $y^1(x_i)$ as the output of interest for training the supervised learner. A standard cross-validation procedure (k -fold, $k = 5$) was used to determine the fidelity of the trained supervised model. This procedure was repeated over s shuffles ($s = 10$) to demonstrate consistency within model training. Average values of the root mean square error (RMSE) over the shuffles were computed to compare model predictions with the true physics-based solution. The prediction phase then involves determining $\hat{y}^2(x_i)$ using the trained model and the test data $\mathbf{x}_i = \{\mathbf{d}_{ij}, y^1(x_i), \eta_j^1\}$, where $\eta_j^1 = [y^1(x_1), y^1(x_2), \dots, y^1(x_j)]$. Using the same modelling construct, two other neighbourhood distance metrics ‘mahalanobis’ [42] and ‘cityblock’ were explored and their respective RMSE values evaluated (see the electronic supplementary material for more details).

2.4. Gaussian process modelling

A Bayesian approach to GPM was used to train the model for a given set of l input vectors $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^l] \in \mathbb{R}^{q \times l}$, where $q = 2J + 1$, with the corresponding output values $\mathbf{y} = [y^1, y^2, \dots, y^l] \in \mathbb{R}^l$ assumed to be available. Using the learned model, output $y(\mathbf{x})$ can be predicted for a new point \mathbf{x} [43,44]. The Gaussian process model can be compactly written as

$$Y(\mathbf{x}) = \beta + Z(\mathbf{x}), \quad (2.11)$$

where β is an unknown hyperparameter and $Z(\mathbf{x})$ is a Gaussian stochastic process with zero-mean and covariance

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \sigma_z^2 R(\mathbf{x}, \mathbf{x}'), \quad (2.12)$$

where $R(\mathbf{x}, \mathbf{x}')$ is a correlation function that can be tuned to the data and σ_z^2 is the process variance. A commonly used choice of correlation function is the stationary family which obeys the *product correlation rule* [45].

$$R(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^p \exp(-\theta_j |\mathbf{x}_j - \mathbf{x}'_j|^{p_j}), \quad (2.13)$$

where $\theta_j \geq 0$ and $0 < p_j \leq 2$ are the hyperparameters. We select $p_j = 2$ such that the underlying function being modelled is smooth and infinitely differentiable.

In the Bayesian approach to data modelling, two levels of inferences are present. The first level is to infer the parameters given the data and defined using the Bayes' theorem as

$$P(\theta_j, \beta, p_j | \mathbf{X}, \mathbf{y}) = \frac{P(\mathbf{X}, \mathbf{y} | \theta_j, \beta, p_j) P(\theta_j, \beta, p_j)}{P(\mathbf{X}, \mathbf{y})}, \quad (2.14)$$

where $P(\theta_j, \beta, p_j | \mathbf{X}, \mathbf{y})$ is the posterior probability of the parameters, $P(\mathbf{X}, \mathbf{y} | \theta_j, \beta, p_j)$ is the likelihood, $P(\theta_j, \beta, p_j)$ is the prior (assumed to be Gaussian) information about the parameters and $P(\mathbf{X}, \mathbf{y})$ is a normalizing constant called the evidence. Note that Gaussian processes are straightforward ways of defining prior distributions for regression and classification problems [46]. Once the hyperparameters are estimated, the second level of inferencing uses these values to estimate $y(\mathbf{x})$ for a new feature vector \mathbf{x} . Because of the prior, the observed outputs are realizations of a Gaussian random field, and therefore, we see that the posterior distribution ($P(y(\mathbf{x}) | \mathbf{X}, \mathbf{y}, \theta_j, \beta, p_j)$) of $y(\mathbf{x})$ is also Gaussian [47], i.e. $Y(\mathbf{x}) | \mathbf{y} \sim \mathcal{N}(\hat{Y}(\mathbf{x}), \sigma_z^2 C(\mathbf{x}, \mathbf{x}'))$. The posterior mean and covariance can be computed as

$$\hat{Y}(\mathbf{x}) = \beta + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\beta) \quad (2.15)$$

and

$$C(\mathbf{x}, \mathbf{x}') = \sigma_z^2 (R(\mathbf{x}, \mathbf{x}') - \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}')), \quad (2.16)$$

where $\mathbf{R} \in \mathbb{R}^{l \times l}$ is the correlation matrix computed using the training points; the ij th element of this matrix is computed as $\mathbf{R}_{ij} = R(\mathbf{x}^i, \mathbf{x}^j)$. $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}^1), R(\mathbf{x}, \mathbf{x}^2), \dots, R(\mathbf{x}, \mathbf{x}^l)] \in \mathbb{R}^l$ is the correlation between the new point \mathbf{x} and the training points, and $\mathbf{1} = [1, 1, \dots, 1] \in \mathbb{R}^l$ [43,47].

Maximum-likelihood estimation (MLE) was used to compute the hyperparameters $\boldsymbol{\theta} = \{\theta_j, j = 1, 2, \dots, q, \beta$, and σ_z^2 defined in equation (2.12). After dropping the constant terms that do not depend on the hyperparameters, the log-likelihood function becomes

$$L(\boldsymbol{\theta}, \beta, \sigma_z^2) = -\frac{1}{2} \left[l \ln \sigma_z^2 + \ln |\mathbf{R}| + \frac{1}{\sigma_z^2} (\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\beta) \right]. \quad (2.17)$$

Given the maximum-likelihood estimate of $\boldsymbol{\theta}$, the parameters β and σ_z^2 were estimated as

$$\hat{\beta} = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y} \quad (2.18)$$

and

$$\hat{\sigma}_z^2 = \frac{1}{l} (\mathbf{y} - \mathbf{1}\hat{\beta})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\beta}). \quad (2.19)$$

The log-likelihood function can be rewritten using equations (2.17)–(2.19) such that the elements of $\boldsymbol{\theta}$ are the only unknown hyperparameters. These hyperparameters were estimated using the DIRECT global optimization algorithm [48]. Refer to the electronic supplementary material for more details.

3. Results

We used GPM to determine whether the data-driven modelling framework can reduce computational time while maintaining accurate solutions to problems of interest. For the case of drug-coated balloon therapy, two-dimensional physics-based models simulating drug transport into the arterial tissue followed by reversible binding predicted arterial distribution patterns of free and bound drug for four mesh configurations (table 1a). GPM-based predictions for free and bound arterial drug concentrations qualitatively resembled the true solutions obtained on the two-dimensional time-dependent model with the highest mesh density (figure 3a and electronic supplementary material, figure S1a). More specifically, the supervised learner trained using information at the baseline configuration and predicted using Refinement 1 settings generated results that were closer to the true solution (defined by Refinement 3) than the respective true solution at Refinement 2 (figure 3b and electronic supplementary material, figure S1b). Note that for these cases, the 'cityblock' metric was used to compute 50 nearest neighbour distances for the supervised learner-derived predictions of arterial bound and free drug concentrations, respectively.

3.1. Distance metric selection and number of nearest neighbours

The accuracy of models trained using supervised learning demonstrated a strong dependence on the metric used for measuring neighbourhood distances and the number of neighbours used for approximating coarse mesh solutions. In the case of the two-dimensional model, when free tissue drug was considered as the output of interest, RMSE between the true simulation at baseline and Refinement 3 (RMSE₀₃) was 1.1345×10^{-4} while the RMSE between the true simulation at Refinements 1 and 3 (RMSE₁₃) was 1.1104×10^{-4} (figure 4a). When bound drug was used as the output, RMSE₀₃ was 1.2×10^{-3} while RMSE₁₃ was 1.1×10^{-3} (figure 4b). These values served as references to validate the performance of the Gaussian process model. Regardless of whether free or bound drug was used as the output, the 'cityblock' metric served as the best metric for distance measurement as the corresponding RMSE values computed across any number of nearest neighbours was consistently lower than RMSE₁₃. This result indicates that supervised learner-based predictions derived using coarser mesh solutions can serve as more accurate representations of true physics than their computational simulation counterparts, and perhaps closer to the order of the true simulation obtained using meshes with higher density. On the other hand, when metrics such as 'euclidean' or 'mahalanobis' were used to compute neighbourhood distances, RMSE values of the model were not always lower than RMSE₁₃. Note that

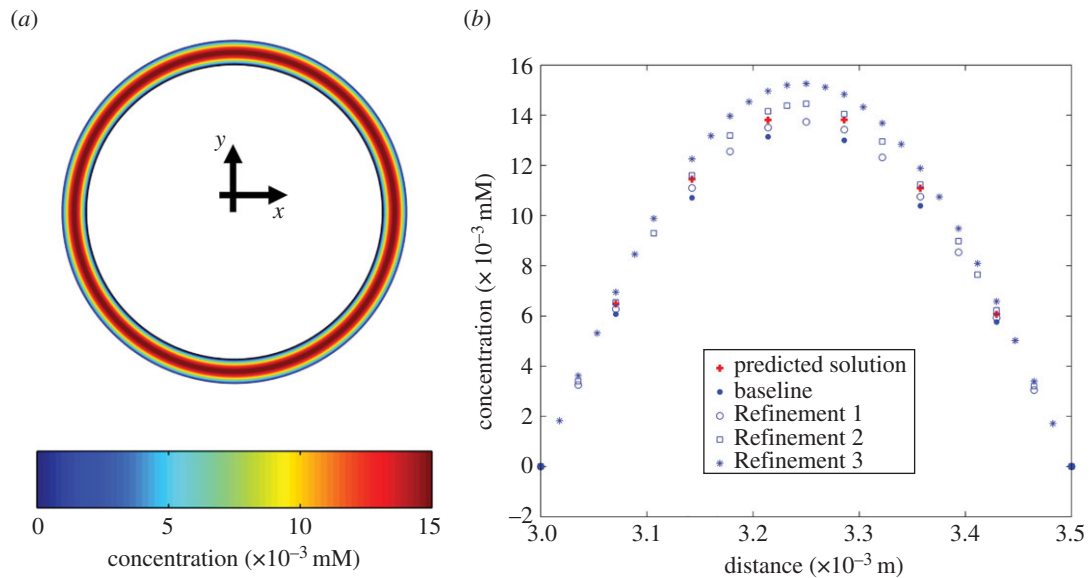


Figure 3. Physics-based solutions for the two-dimensional model of drug-coated balloon delivery and model-based predictions at 1 h after balloon inflation. (a) True solution for the bound drug was computed on the mesh setting with the highest density (Refinement 3 in table 1a). (b) Arterial tissue drug concentration for bound drug for solutions based on four mesh configurations and the model-predicted solution were plotted as a function of cross-sectional depth.

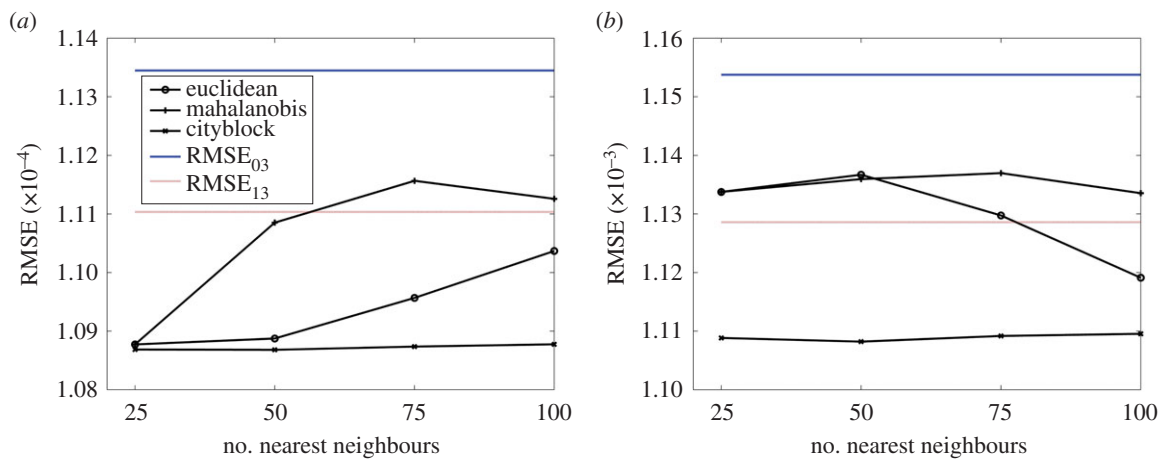


Figure 4. Model accuracy as measured by RMSE varies with distance metric selection and the number of nearest neighbours. RMSE values were computed for three distance metrics and four different values for the number of nearest neighbours. Both (a) free and (b) bound drug distribution were used as the outputs of interest. $RMSE_{03}$ denotes the RMSE value computed between the solution obtained using the mesh with the lowest density (baseline) and the solution using the mesh with the highest density (Refinement 3). $RMSE_{13}$ denotes the RMSE value computed between the solutions obtained with the Refinements 1 and 3 cases. (Online version in colour.)

while the ‘euclidean’ metric is well known in the literature as an ordinary distance between two points (see §2.3), ‘mahalanobis’ distance takes both the distance measurement and the directionality into account, and ‘cityblock’ distance is simply defined using the \mathcal{L}_1 norm. Overall, our findings point to the importance of carefully choosing and validating the distance metric for computing neighbourhood distances as well as the number of nearest neighbours used to compute model performance in order to obtain closer approximations of true physics via the framework of supervised learning on coarse mesh solutions. Using the same modelling construct, several other neighbourhood distance metrics including ‘chebyshev’ [49,50] and ‘minkowski’ and their variations can also be explored.

3.2. Extension to memory-intensive simulations

Having validated the utility of combining the GPM framework to learn from the two-dimensional physiologic simulation data, we extended the framework to physics-

based models with increased physiologic and computational complexities (figures 1b and 2b). Simulations were performed on the three-dimensional model of drug delivery from a stent deployed in an arterial vessel and the effects of steady luminal flow considered as a coupled phenomenon with drug transport and reversible drug binding to arterial tissue (table 1b). In contrast to the simulations for the two-dimensional model of drug-coated balloon therapy, the true simulation-based solutions of mural drug distribution for the three-dimensional model of stent-based drug delivery demonstrated an apparent qualitative difference in drug distribution pattern as a function of computational mesh density (figure 5). Also, validation of the supervised learner for the three-dimensional model in the same way as the two-dimensional model was not feasible as our computational resources precluded simulation of the Refinement 3 case (table 1b). Therefore, we computed the RMSE values for the supervised learner relative to Refinements 1 and 2 cases, respectively. Models were generated with free and bound drug as outputs using ‘mahalanobis’ metric for

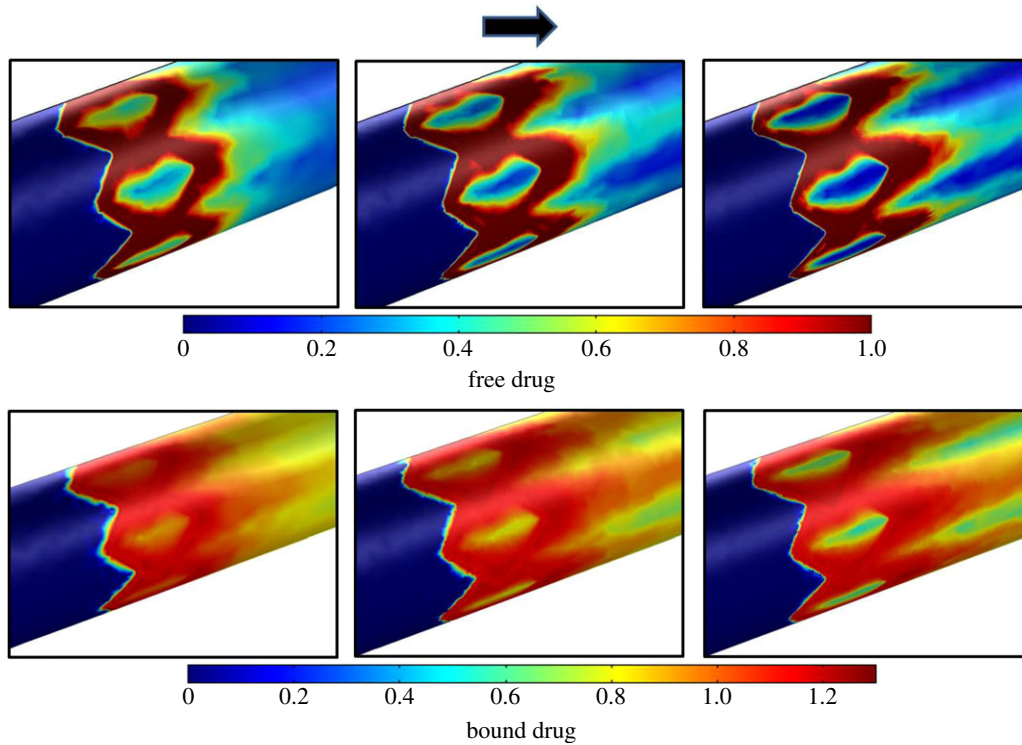


Figure 5. Free and bound arterial drug distribution patterns as predicted by the three-dimensional model of stent-based drug delivery demonstrate qualitative refinement in solution with increased mesh density. The arrow points to the direction of increasing mesh density. Mural drug concentration is normalized to the source and table 1*b* has more details on the three mesh configurations.

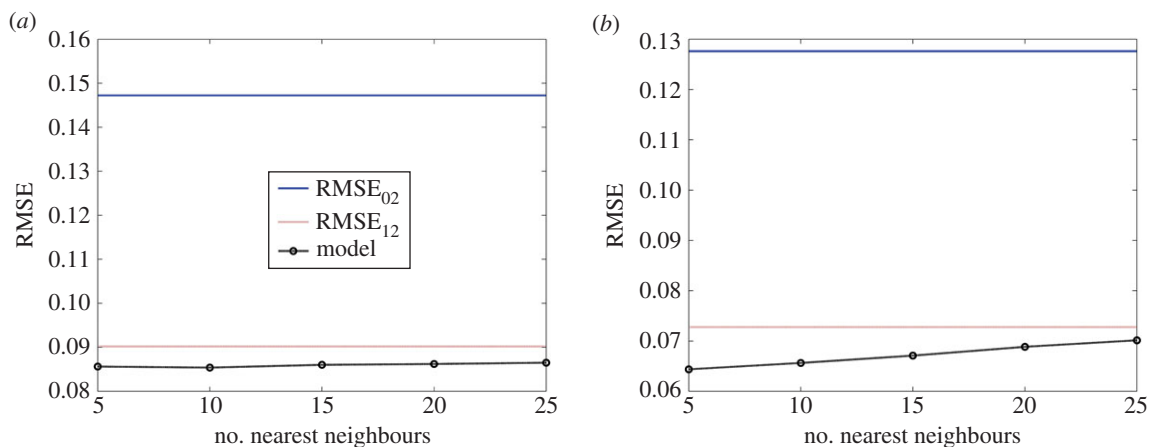


Figure 6. Supervised learner-based predictions for the problem of stent-based drug delivery. Both (a) free mural and (b) bound tissue drug distribution were used as the outputs of interest. Model accuracy as measured by RMSE is shown to vary with the number of nearest neighbours used to construct the supervised learner. (Online version in colour.)

nearest neighbour distance measurement and as a function of the number of nearest neighbours (figure 6*a,b*). These trends indicate that the model predictions were improved in comparison to that of the true simulation outputs for the Refinement 1 case, but as a function of the number of nearest neighbours and the distance metric.

4. Discussion

Computational modelling is being used increasingly in medical device applications, both in the pre-clinical and regulatory domains as well as in emerging clinical applications directly related to patient care. As our ability to characterize patient- and device-specific features in real-time and with micrometre-level precision has improved, such models promise to yield mechanistic and data-driven predictions that will

enable personalized care [21]. Yet the computational time required to perform multi-scale high fidelity simulations can be enormous, forcing adoption of high-end computational architecture or reduced fidelity approaches with simplifying assumptions and recognized inaccuracies (e.g. one-dimensional blood flow and lumped parameter models [51], reduced basis methods [52,53], data-driven clinical risk scores [54], etc.). As an alternative, we now demonstrate that use of machine learning on low fidelity, coarse mesh solutions can augment fidelity at a fraction of the computational cost of traditional simulation.

The need to generate high fidelity insight from low fidelity information is not unique to medicine, but rather characteristic of any sufficiently advanced field. Indeed, GPM forms the basis for *Kriging*, an approach that has been widely applied to problems of fidelity optimization in other disciplines including geostatistics as well as aerospace and

automobile engineering [43,45,55–57]. Our goal was to leverage this paradigm and demonstrate its utility within a medical context, and in particular that of cardiovascular devices, where the potential for simulation is widely appreciated and accepted, yet its broad reaching impact remains largely anticipated and bottlenecked by modelling inefficiencies. To this effect, by combining nearest neighbour information from low fidelity meshes in a GPM framework, we demonstrate the potential of fused machine learning/physics-based simulation approaches.

4.1. Supervised learning for different physics-based models

Using supervised learning, we examined endovascular drug delivery in two contexts, each with unique modelling complexities. In the balloon delivery model, time-dependent simulations predicted intramural free drug diffusion soon after delivery as well as reversible binding of drug and tissue retention after 1 h from the onset of balloon inflation (figure 3*a*). Extraction of data at such long time points naturally extends simulation times. For stent-based delivery, luminal flow transports mural drug from the deployment site to downstream regions in a manner such that pattern of drug imprinted on the mural surface tracks with stent design (figure 5). Here, drugs with low diffusivity in a highly convective environment lead to extremely high Peclet numbers which in turn can result in large instabilities in the numerical solver. Moreover, stent implantation intervenes with the blood flow milieu by introducing perturbations into the boundary layer, causing drug-rich recirculating pools proximal and distal to stent struts [34,36,37,58]. The former aspect can be resolved using highly dense mesh configurations; the latter captured reasonably well with boundary layer meshes. While the two-dimensional model requires high-end computer processors for fast computation due to the model's time-dependent nature, the three-dimensional model is memory intensive and requires large RAM. Each complexity prolongs simulation times, prompting us to explore alternative means by which to obtain fast and accurate solutions to problems of interest.

4.2. Role of Gaussian processes

Our approach of combining physics-based computational modelling with data-driven machine learning demonstrates a method to efficiently model complex, multi-scale/multi-physics physiological problems. We specifically selected Gaussian processes because they offer a practical and probabilistic approach to learning and model predictions of real-world problems [32]. However, if this modelling framework is used on an input feature space with several data points and feature vectors, computational time for MLE (equation (2.17)) becomes high. To circumvent part of this problem, we computed the Cholesky decomposition of \mathbf{R} . This factorization technique produces an upper triangular matrix (\mathbf{A}) from the diagonal and the upper triangle of the matrix \mathbf{R} , satisfying the relation $\mathbf{A}^T\mathbf{A} = \mathbf{R}$, which then allows for efficient matrix inversion. This allowed the posterior mean to be computed using a vector–vector product, i.e. $\hat{\mathbf{Y}}(\mathbf{x}) = \beta + \mathbf{r}(\mathbf{x})^T\mathbf{w}$, where $\mathbf{w} = \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta})$. Nevertheless, the combination of physics-based simulations generated on the coarse mesh configurations followed by construction of a supervised learner consumed significantly less time than

that was required for the highest mesh density simulation (table 1*a,b*). Moreover, time consumed for predicting the output on new nodes using the trained supervised learner is almost negligible. This result highlights the power of harnessing GPM to efficiently approximate data generated from computationally intensive physiologic simulations.

4.3. Mesh refinement and nearest neighbour approaches

The concept of increasing mesh density in a successive manner to obtain a solution with better accuracy is core to physics-based computational modelling and mandatory for validating any computational simulation. Yet increased mesh density increases simulation time. Therefore, alternative means by which to obtain fast, more accurate solutions using coarse mesh solutions and without performing high density mesh simulations seems very attractive. We demonstrated that supervised machine learning using Gaussian process modelling on nearest neighbours serves as an appropriate framework to achieve this task.

Nearest neighbour approaches are widely used in the pattern recognition field where one typically solves an optimization problem to identify closest points in a dataset [39,40]. Foundational to such applications is the assumption that some degree of similarity arises purely by virtue of proximity. By assuming that a computational solution space is more or less continuous and void of extreme transitions, we hypothesized that a group comprising solutions of drug concentrations from a node's nearest neighbours carries with it useful information that can facilitate capture of underlying physics more accurately and in the direction of refined mesh solutions. In our study, we observed a strong dependence of the supervised learner's predictive accuracies as a function of the number of nearest neighbours. Interestingly, a higher number of neighbours was found to be optimal for the two-dimensional model (figure 4*a,b*) as compared with the three-dimensional setting (figure 6*a,b*), likely reflective of a trade-off between added information content and excessive smoothing. In order to implement this framework, one must therefore appreciate the interdependence between the distance metric, number of nearest neighbours and the algorithm for supervised learning. While appropriate combinations can augment efficiency of computationally intensive simulations, inappropriate ones may have an opposite effect.

4.4. Study limitations and future directions

We used drug-coated balloon and drug-eluting stent therapies as exemplary medical devices, where modelling solutions require that multiple physics be coupled to gain relevant design and performance insights. Since these device simulations are computationally expensive, we sought alternative means by which to obtain fast and efficient solutions. While we made some assumptions and approximations, such as idealized arterial geometries, these did not detract from the overall scope of the study. Going forward, we hope to apply these learning approaches to examine the impact of pulsatile blood flow [36], other drug, device and procedural settings [35,58–62], disease-induced changes in the arterial wall [63], and patient-specific geometries extracted using image reconstruction to obtain physiologically realistic solutions. As we continue to develop such applications, a number of other approaches can be readily

considered. For example, while we present GPM as a useful learning framework, other supervised learning techniques such as artificial neural networks and support vector machines could be implemented in a similar fashion, serving as efficient approximations to standard physics-based simulations [39,40,64,65]. Furthermore, several other methods related to fidelity optimization and applications to data mining exist within the literature [66–69], and could be leveraged in conjunction with the presented framework. Also, when standard linear dimensionality reduction techniques such as principal component analysis (PCA) and nonlinear dimensionality reduction techniques such as kernel PCA are used to pre-condition the input feature space prior to supervised learning [70], computational time for model training can be further reduced. Pseudocodes are provided in the electronic supplementary material to assist those interested to implement our supervised learning schema for specific applications.

5. Conclusion

Emerging applications in medicine and medical device design require that we find ways to provide high fidelity

insight efficiently. Fusion of machine learning with standard physics-based computational models is perfectly suited to meet this challenge. In the context of endovascular device-based drug therapies, we demonstrate the potential of fidelity augmentation, wherein high fidelity inferences were drawn quickly from low fidelity information. Use of such a framework may help computational tools overcome limiting bottlenecks, allowing them to be realized in real-world medical applications characterized by inherent modelling complexities with a time critical nature.

Acknowledgements. K.K. and V.B.K. thank Mr Kenneth Vandevordt from the Charles Stark Draper Laboratory for sharing his expertise on model geometry generation, and gratefully acknowledge the insights provided by the Institute for Medical Engineering and Science Clinical Research Center (IMES/CRC) at the Massachusetts Institute of Technology regarding the clinical applicability of this work.

Funding statement. This work was supported by funding from the Charles Stark Draper Laboratory to V.B.K. (CSDL-29414-005, CSDL-29889-001, CSDL-30716-005 & CSDL-30736-003), from the American Heart Association to K.K. (12FTF12080241) and from the United States Department of Health and Human Services-National Institutes of Health to T.S. (P2ORR016461).

References

- Min JK *et al.* 2012 Diagnostic accuracy of fractional flow reserve from anatomic CT angiography. *JAMA* **308**, 1237–1245. (doi:10.1001/2012.jama.11274)
- Taylor CA, Fonte TA, Min JK. 2013 Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: scientific basis. *J. Am. Coll. Cardiol.* **61**, 2233–2241. (doi:10.1016/j.jacc.2012.11.083)
- US Food and Drug Administration. 2009 The critical path initiative. Report on key achievements in 2009. See <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/UCM221651.pdf>.
- Wentzel JJ, Krams R, Schuurbiens JC, Oomen JA, Kloet J, van Der Giessen WJ, Serruys PW, Slager CJ. 2001 Relationship between neointimal thickness and shear stress after Wallstent implantation in human coronary arteries. *Circulation* **103**, 1740–1745. (doi:10.1161/01.CIR.103.13.1740)
- Koskinas KC, Chatzizisis YS, Antoniadis AP, Giannoglou GD. 2012 Role of endothelial shear stress in stent restenosis and thrombosis: pathophysiologic mechanisms and implications for clinical translation. *J. Am. Coll. Cardiol.* **59**, 1337–1349. (doi:10.1016/j.jacc.2011.10.903)
- Timmins LH, Moreno MR, Meyer CA, Criscione JC, Rachev A, Moore Jr JE. 2007 Stented artery biomechanics and device design optimization. *Med. Biol. Eng. Comput.* **45**, 505–513. (doi:10.1007/s11517-007-0180-3)
- Pant S, Limbert G, Curzen NP, Bressloff NW. 2011 Multiobjective design optimisation of coronary stents. *Biomaterials* **32**, 7755–7773. (doi:10.1016/j.biomaterials.2011.07.059)
- Pant S, Bressloff NW, Limbert G. 2012 Geometry parameterization and multidisciplinary constrained optimization of coronary stents. *Biomech. Model. Mechanobiol.* **11**, 61–82. (doi:10.1007/s10237-011-0293-3)
- Gundert TJ, Marsden AL, Yang W, LaDisa Jr JF. 2012 Optimization of cardiovascular stent design using computational fluid dynamics. *J. Biomech. Eng.* **134**, 011002. (doi:10.1115/1.4005542)
- Gijsen FJ *et al.* 2008 Simulation of stent deployment in a realistic human coronary artery. *Biomed. Eng. Online* **7**, 23. (doi:10.1186/1475-925X-7-23)
- Mortier P, Holzapfel GA, De Beule M, Van Loo D, Taeymans Y, Segers P, Verdonck P, Verheghe B. 2010 A novel simulation strategy for stent insertion and deployment in curved coronary bifurcations: comparison of three drug-eluting stents. *Ann. Biomed. Eng.* **38**, 88–99. (doi:10.1007/s10439-009-9836-5)
- Bryan R, Nair PB, Taylor M. 2012 Influence of femur size and morphology on load transfer in the resurfaced femoral head: a large scale, multi-subject finite element study. *J. Biomech.* **45**, 1952–1958. (doi:10.1016/j.jbiomech.2012.05.015)
- Morlacchi S, Colleoni SG, Cárdenes R, Chiastra C, Diez JL, Larrabide I, Migliavacca F. 2013 Patient-specific simulations of stenting procedures in coronary bifurcations: two clinical cases. *Med. Eng. Phys.* **35**, 1272–1281. (doi:10.1016/j.medengphys.2013.01.007)
- Chiastra C, Morlacchi S, Gallo D, Morbiducci U, Cárdenes R, Larrabide I, Migliavacca F. 2013 Computational fluid dynamic simulations of image-based stented coronary bifurcation models. *J. R. Soc. Interface* **10**, 20130193. (doi:10.1098/rsif.2013.0193)
- Ragkousis GE, Curzen N, Bressloff NW. 2014 Simulation of longitudinal stent deformation in a patient-specific coronary artery. *Med. Eng. Phys.* **36**, 467–476. (doi:10.1016/j.medengphys.2014.02.004)
- Morganti S, Conti M, Aiello M, Valentini A, Mazzola A, Reali A, Auricchio F. 2014 Simulation of transcatheter aortic valve implantation through patient-specific finite element analysis: two clinical cases. *J. Biomech.* **47**, 2547–2555. (doi:10.1016/j.jbiomech.2014.06.007)
- Xiong G, Taylor CA. 2010 Virtual stent grafting in personalized surgical planning for treatment of aortic aneurysms using image-based computational fluid dynamics. *Med. Image Comput. Comput. Assist. Interv.* **13**, 375–382.
- Auricchio F, Conti M, Morganti S, Totaro P. 2011 A computational tool to support pre-operative planning of stentless aortic valve implant. *Med. Eng. Phys.* **33**, 1183–1192. (doi:10.1016/j.medengphys.2011.05.006)
- Audigier C *et al.* 2013 Lattice Boltzmann method for fast patient-specific simulation of liver tumor ablation from CT images. *Med. Image Comput. Assist. Interv.* **16**, 323–330.
- Kim KH *et al.* 2014 A novel noninvasive technology for treatment planning using virtual coronary stenting and computed tomography-derived computed fractional flow reserve. *JACC Cardiovasc. Interv.* **7**, 72–78. (doi:10.1016/j.jcin.2013.05.024)
- Kolandaivelu K, Leiden BB, Edelman ER. 2014 Predicting response to endovascular therapies: dissecting the roles of local lesion complexity, systemic comorbidity, and clinical uncertainty. *J. Biomech.* **47**, 908–921. (doi:10.1016/j.jbiomech.2014.01.006)
- Berman DS, Stoeber RA, Dey D. 2014 Combined anatomy and physiology on coronary computed tomography angiography: a step or two in the right direction. *J. Am. Coll. Cardiol.* **63**, 1913–1915. (doi:10.1016/j.jacc.2014.02.559)

23. Woods TC, Marks AR. 2004 Drug-eluting stents. *Annu. Rev. Med.* **55**, 169–178. (doi:10.1146/annurev.med.55.091902.105243)
24. Stefanini GG, Holmes Jr DR. 2013 Drug-eluting coronary artery stents. *N Engl. J. Med.* **368**, 254–265. (doi:10.1056/NEJMra1210816)
25. Waksman R, Pakala R. 2009 Drug-eluting balloon: the comeback kid? *Circ. Cardiovasc. Interv.* **2**, 352–358. (doi:10.1161/CIRCINTERVENTIONS.109.873703)
26. Gray WA, Granada JF. 2010 Drug-coated balloons for the prevention of vascular restenosis. *Circulation* **121**, 2672–2680. (doi:10.1161/CIRCULATIONAHA.110.936922)
27. Byrne RA, Joner M, Alfonso F, Kastrati A. 2014 Drug-coated balloon therapy in coronary and peripheral artery disease. *Nat. Rev. Cardiol.* **11**, 13–23. (doi:10.1038/nrcardio.2013.165)
28. Sarode K, Spelber DA, Bhatt DL, Mohammad A, Prasad A, Brilakis ES, Banerjee S. 2014 Drug delivering technology for endovascular management of infrainguinal peripheral artery disease. *JACC Cardiovasc. Interv.* **7**, 827–839. (doi:10.1016/j.jcin.2014.05.008)
29. World Health Organization. 2014 *World Health Statistics*.
30. Hwang CW, Wu D, Edelman ER. 2001 Physiological transport forces govern drug distribution for stent-based delivery. *Circulation* **104**, 822–829. (doi:10.1161/hc3101.092214)
31. Wessely R, Schömlig A, Kastrati A. 2006 Sirolimus and paclitaxel on polymer-based drug-eluting stents: similar but different. *J. Am. Coll. Cardiol.* **47**, 708–714. (doi:10.1016/j.jacc.2005.09.047)
32. Rasmussen CE, Williams CKI. 2006 *Gaussian processes for machine learning*. Cambridge, MA: The MIT press.
33. Kolachalama VB *et al.* 2013 Mechanisms of tissue uptake and retention in zotarolimus-coated balloon therapy. *Circulation* **127**, 2047–2055. (doi:10.1161/CIRCULATIONAHA.113.002051)
34. Kolachalama VB, Levine EG, Edelman ER. 2009 Luminal flow amplifies stent-based drug deposition in arterial bifurcations. *PLoS ONE* **4**, e8105. (doi:10.1371/journal.pone.0008105)
35. Shazly T, Kolachalama VB, Ferdous J, Oberhauser JP, Hossainy S, Edelman ER. 2012 Assessment of material by-product fate from bioresorbable vascular scaffolds. *Ann. Biomed. Eng.* **40**, 955–965. (doi:10.1007/s10439-011-0445-8)
36. O'Brien CC, Kolachalama VB, Barber TJ, Simmons A, Edelman ER. 2013 Impact of flow pulsatility on arterial drug distribution in stent-based therapy. *J. Control Release* **168**, 115–124. (doi:10.1016/j.jconrel.2013.03.014)
37. Kolachalama VB, Tzafriri AR, Arifin DY, Edelman ER. 2009 Luminal flow patterns dictate arterial drug deposition in stent-based delivery. *J. Control Release* **133**, 24–30. (doi:10.1016/j.jconrel.2008.09.075)
38. Lovich MA, Creel C, Hong K, Hwang CW, Edelman ER. 2001 Carrier proteins determine local pharmacokinetics and arterial distribution of paclitaxel. *J. Pharm. Sci.* **90**, 1324–1335. (doi:10.1002/jps.1085)
39. Bishop CM. 2007 *Pattern recognition and machine learning*. Berlin, Germany: Springer.
40. Hastie T, Tibshirani R, Friedman J. 2009 *The elements of statistical learning: data mining, inference and prediction*. Berlin, Germany: Springer.
41. Deza E, Deza MM. 2009 *Encyclopedia of distances*. Berlin, Germany: Springer.
42. Mahalanobis PC. 1936 On the generalised distance in statistics. *Proc. Natl Inst. Sci. India* **2**, 49–55.
43. Keane AJ, Nair PB. 2005 *Computational approaches for aerospace design. The pursuit of excellence*. New York, NY: John Wiley & Sons Ltd.
44. Kolachalama VB, Bressloff NW, Nair PB. 2007 Mining data from hemodynamic simulations via Bayesian emulation. *Biomed. Eng. Online* **6**, 47. (doi:10.1186/1475-925X-6-47)
45. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. 1989 Design and analysis of computer experiments. *Stat. Sci.* **4**, 409–435. (doi:10.1214/ss/1177012413)
46. Neal RM. 1998 Regression and classification using Gaussian process priors. In *Bayesian statistics 6* (eds JM Bernardo, JO Berger, AP Dawid, AFM Smith), pp. 475–501. Oxford, UK: Oxford University Press.
47. Currin C, Mitchell T, Morris M, Ylvisaker D. 1991 Bayesian prediction of deterministic functions with applications to the design and analysis of computer experiments. *J. Am. Stat. Assoc.* **86**, 953–963. (doi:10.1080/01621459.1991.10475138)
48. Jones DR, Perttunen CD, Stuckman BE. 1993 Lipschitzian optimization without the Lipschitz constant. *J. Optimiz. Theory App.* **79**, 157–181. (doi:10.1007/BF00941892)
49. Cantrell CD. 2000 *Modern mathematical methods for physicists and engineers*. Cambridge, UK: Cambridge University Press.
50. Abello JM, Pardalos PM, Resende MGC. 2002 *Handbook of massive data sets*. Berlin, Germany: Springer.
51. Shi Y, Lawford P, Hose R. 2011 Review of zero-D and 1-D models of blood flow in the cardiovascular system. *Biomed. Eng. Online* **10**, 33. (doi:10.1186/1475-925X-10-33)
52. Quarteroni A, Rozza G. 2007 Numerical solution of parametrized Navier–Stokes equations by reduced basis methods. *Numer. Methods Part D E* **23**, 923–948. (doi:10.1002/num.20249)
53. Manzoni A, Quarteroni A, Rozza G. 2011 Shape optimization for viscous flows by reduced basis methods and free-form deformation. *Int. J. Numer. Method Fluids* **70**, 646–670. (doi:10.1002/flid.2712)
54. Sianos G *et al.* 2005 The SYNTAX Score: an angiographic tool grading the complexity of coronary artery disease. *EuroIntervention* **1**, 219–227.
55. Matheron G. 1963 Principles of geostatistics. *Econ. Geol.* **58**, 1246–1266. (doi:10.2113/gsecongeo.58.8.1246)
56. Alexandrov NM, Lewis RM, Gumbert CR, Green LL, Newman PA. 2001 Approximation and model management in aerodynamic optimization with variable-fidelity models. *J. Aircraft* **38**, 1093–1101. (doi:10.2514/2.2877)
57. Gano SE, Renaud JE. 2005 Hybrid variable fidelity optimization by using a Kriging-based scaling function. *AAAA J.* **43**, 2422–2430. (doi:10.2514/1.12466)
58. Kolandaivelu K *et al.* 2011 Stent thrombogenicity early in high-risk interventional settings is driven by stent design and deployment and protected by polymer-drug coatings. *Circulation* **123**, 1400–1409. (doi:10.1161/CIRCULATIONAHA.110.003210)
59. Levin AD, Vukmirovic N, Hwang CW, Edelman ER. 2004 Specific binding to intracellular proteins determines arterial transport properties for rapamycin and paclitaxel. *Proc. Natl Acad. Sci. USA* **101**, 9463–9467. (doi:10.1073/pnas.0400918101)
60. Kolachalama VB, Bressloff NW, Nair PB, Shearman CP. 2008 Predictive haemodynamics in a one-dimensional human carotid artery bifurcation. Part II: application to graft design. *IEEE Trans. Biomed. Eng.* **55**, 1176–1184. (doi:10.1109/TBME.2007.912398)
61. Onuma Y, Serruys PW. 2011 Bioresorbable scaffold: the advent of a new era in percutaneous coronary and peripheral revascularization? *Circulation* **123**, 779–797. (doi:10.1161/CIRCULATIONAHA.110.971606)
62. Ferdous J, Kolachalama VB, Shazly T. 2013 Impact of polymer structure and composition on fully resorbable endovascular scaffold performance. *Acta Biomater.* **9**, 6052–6061. (doi:10.1016/j.actbio.2012.12.011)
63. Tzafriri AR, Vukmirovic N, Kolachalama VB, Astafieva I, Edelman ER. 2010 Lesion complexity determines arterial drug distribution after local drug delivery. *J. Control Release* **142**, 332–338. (doi:10.1016/j.jconrel.2009.11.007)
64. McCullagh P, Nelder JA. 1999 *Generalized linear models*. London, UK: Chapman & Hall.
65. Dobson AJ. 2001 *An introduction to generalized linear models*. London, UK: Chapman & Hall.
66. Booker AJ, Dennis Jr JE, Frank PD, Serafini DB, Torczon V, Trosset MW. 1999 A rigorous framework for optimization of expensive functions by surrogates. *Struct. Optimiz.* **17**, 1–13. (doi:10.1007/BF01197708)
67. Garcke J, Griebel M, Thess M. 2001 Data mining with sparse grids. *Computing* **67**, 225–253. (doi:10.1007/s006070170007)
68. Shirazi AS, Mammen von S, Jacob C. 2013 Abstraction of agent interaction processes. *Simul-T Soc. Mod. Sim.* **89**, 524–538.
69. Shirazi AS, Davison T, Mammen von S, Denzinger J, Jacob C. 2014 Adaptive agent abstractions to speed up spatial agent-based simulations. *Simul. Model Pract. Theory* **40**, 144–160. (doi:10.1016/j.simpat.2013.09.001)
70. Lee JA, Verleysen M. 2007 *Nonlinear dimensionality reduction*. Berlin, Germany: Springer.