



This is a repository copy of *Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/115191/>

Version: Accepted Version

---

**Article:**

Melis, A. [orcid.org/0000-0002-8261-0421](http://orcid.org/0000-0002-8261-0421), Clayton, R.H. [orcid.org/0000-0002-8438-7518](http://orcid.org/0000-0002-8438-7518) and Marzo, A. [orcid.org/0000-0002-6702-7932](http://orcid.org/0000-0002-6702-7932) (2017) Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators. *International Journal for Numerical Methods in Biomedical Engineering*. ISSN 2040-7939

<https://doi.org/10.1002/cnm.2882>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Bayesian sensitivity analysis of a 1D vascular model with Gaussian process emulators

A. Melis<sup>1,2</sup>, R.H. Clayton<sup>1,3</sup>, A. Marzo<sup>1,2\*</sup>

<sup>1</sup>*INSIGNEO Institute for in silico Medicine, The University of Sheffield, UK*

<sup>2</sup>*Department of Mechanical Engineering, The University of Sheffield, UK*

<sup>3</sup>*Department of Computer Science, The University of Sheffield, UK*

## SUMMARY

One-dimensional models of the cardiovascular system can capture the physics of pulse waves, but involve many parameters. Since these may vary among individuals, patient-specific models are difficult to construct. Sensitivity analysis can be used to rank model parameters by their effect on outputs, and to quantify how uncertainty in parameters influences output uncertainty. This type of analysis is often conducted with a Monte Carlo method, where large numbers of model runs are used to assess input-output relations. The aim of this study was to demonstrate the computational efficiency of variance based sensitivity analysis of 1D vascular models using Gaussian process emulators, compared to a standard Monte Carlo approach. The methodology was tested on four vascular networks of increasing complexity to analyse its scalability. The computational time needed to perform the sensitivity analysis with an emulator was reduced by the 99.96% compared to a Monte Carlo approach. Despite the reduced computational time, sensitivity indices obtained using the two approaches were comparable. The scalability study showed that the number of mechanistic simulations needed to train a Gaussian process for sensitivity analysis was of the order  $\mathcal{O}(d)$ , rather than  $\mathcal{O}(d \times 10^3)$  needed for Monte Carlo analysis (where  $d$  is the number of parameters in the model). The efficiency of this approach, combined with capacity to estimate the impact of uncertain parameters on model outputs, will enable development of patient-specific models of the vascular system, and has the potential to produce results with clinical relevance.

Received ...

\*Correspondence to: The University of Sheffield, Pam Liversidge Building, Room F39, Mappin Street, Sheffield, S1 3JD, UK. E-mail: a.marzo@sheffield.ac.uk

Contract/grant sponsor: EPSRC; contract/grant number: EP/K037145/1

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as Int. J. Numer. Meth. Biomed. Engng., e02882. doi: 10.1002/cnm.2882

KEY WORDS: 1D vascular model; Gaussian process; emulator; sensitivity analysis; Sobol

## 1. INTRODUCTION

The cardiovascular system is a complex network of elastic vessels through which blood is pumped by contraction of the heart. This pulsatile regime and vessel elasticity cause pressure to propagate along the arterial circulation as waves. Mechanical discontinuities caused by bifurcation, bends or cardiovascular pathology, e.g. vessel stenoses, cause pressure waves to be reflected in all directions. As a result, pressure waves measured at a specific location can be seen as the result of a superimposition of incident and reflected waves. The analysis of this superposition mechanism allows for the study of mechanical properties upstream and downstream of the measurement point and can be a rich source of diagnostic information about the system through which these waves propagate. However, given the vascular system complexity, it is at present difficult to ascribe a particular waveform feature to a specific trait of the arterial circulation [1, 2, 3].

Numerical models of the vascular tree are ultimately aimed at a better understanding of the circulation. Depending on the model dimensionality and on the target application, different levels of complexity can be reached [4]. Three-dimensional (3D) models are spatially accurate and can predict blood flow in complex geometries, e.g. around heart valves, near aneurysms, or arterial bifurcations [5, 6]. However, these models rely on an accurate description of the system geometry and of the boundary conditions. Hence, when non-invasive measurements are difficult to be obtained, 3D simulations cannot be performed. One-dimensional (1D) models are capable of simulating the physics of pulse wave propagation and reflection. Many of these models are based on a reduction of three-dimensional Navier-Stokes equations and on a constitutive equation linking trans-mural pressure to arterial wall displacement [7, 8, 9]. Each vessel of the arterial tree is represented by a straight elastic tube whose parameters can be fixed or varied along the tube length [10].

One-dimensional models are less computationally expensive than three dimensional ones and require a less precise description of the vasculature geometry. The drawbacks reside in the lack of accuracy in areas where the flow is not developed, e.g. in proximity of valve outlets and bifurcations, where secondary and recirculatory flows may originate. 1D models have

been used to study and reproduce the behaviour of the entire systemic circulation [11, 12], the coronaries [13], the cerebral vasculature [14], and the pulmonary circulation [15].

Although faster to solve than full three-dimensional simulations, 1D models still require a large number of parameters to be specified for each vessel in the system. In a patient-specific scenario, the total number of parameters to be measured becomes easily infeasible and clinically non-justifiable, e.g. an arterial tree made of 103 segments requires about 500 parameters [12]. This issue can be overcome by identifying those parameters that will have a more significant effect on the output of interest. By ranking the inputs, the parameters worth to be accurately measured could be identified and their uncertainty reduced. Whereas, parameters whose variation has less effect on the outcome of interest can be fixed to their typical reference values. The process of parameter ranking and fixing consists in performing a model sensitivity analysis [16].

The state of the art technique for parameter ranking and fixing is the computation of sensitivity indices [17]. These indices assess the sensitivity of outputs to variation in individual inputs or combinations of inputs. First-order sensitivity indices measure the proportion of output variance that can be accounted for by the variance of one individual input. The variation of a single output due to the combined variation of multiple inputs is measured by higher-order sensitivity indices. The sum of all indices concerning a specific input is called total sensitivity index. First-order indices can be used to rank inputs and to decide which has the strongest influence on the model outputs. Eventually, by highlighting input collaborations, total sensitivity indices assess the inputs or model parameters that can be fixed.

The Monte Carlo method is a simple approach to compute sensitivity indices. Several thousand simulator (i.e. the mechanistic model) runs are done. For each simulator run a different set of inputs are randomly drawn from a distribution of point, which covers the entire input space. Ideally, the distribution contains infinite points, hence by obtaining a result for each point, the model global behaviour would be known. In practice, an infinite distribution cannot be achieved, and Monte Carlo sampling requires a number of runs of the order of  $\mathcal{O}(d \times 10^6)$ , where  $d$  is the number of input parameters. Saltelli et al. [18, 19] introduced a Monte-Carlo-based technique to calculate sensitivity indices which requires a number of runs of order  $\mathcal{O}(d \times 10^3)$ , but this can still be an impractical number of runs for most applications.

When considering a large number of parameters, the computational time needed for the  $d \times 10^3$  simulations becomes prohibitively high. This analysis can be made more efficient by introducing a fast-running approximation of the mechanistic model, i.e. an emulator [16].

Emulators are well-known in both the applied mathematics and in the statistics community [25]. The former mostly uses a tool called polynomial chaos expansions (PCE), whereas the latter principally uses Gaussian process (GP) emulators. Both tools aim to infer the simulator global behaviour starting from observed simulator runs. The main advantage of GP over the PCE technique resides in the availability of uncertainty information. This characteristic directly descends from the probabilistic nature of a GP, which allows embedding of uncertainty and explicit treatment of model parameters as uncertain quantities. This is essential because it enables the impact of missing, uncertain, or noisy measurements on model outputs to be quantified which can become relevant in a clinical setting. For a detailed analysis of the differences between the two techniques see [25, 21]. PCE has been successfully used for sensitivity analysis in the cardiovascular field, see for example [22, 23, 24, 20]. Outside of the cardiovascular field, GP are a widely used emulation technique [26], but to our knowledge GP has never been used to predict 1D blood flow model outcomes for sensitivity analysis purposes.

The aim of this work was therefore to compare sensitivity analysis of a 1D cardiovascular model based on GP emulators with the traditional approach based on Monte Carlo sampling. We have also shown how the GP properties scale with the vascular network complexity. Four patient-generic arterial networks of increasing size were employed to demonstrate the benefits of using GP emulators for cardiovascular applications. Results were validated on the outcomes of a Monte Carlo analysis.

## 2. METHODOLOGY

### 2.1. Vascular model

The vascular model was based on the reduced 1D form of the general continuity and Navier-Stokes equations for incompressible flows within narrow straight elastic tubes [7, 9]. A constitutive equation describing elastic wall behaviour was used to close the equation system. The equations governing the problem of flow through and an elastic vessel are:

$$\begin{cases} \frac{\partial A}{\partial t} + \frac{\partial (Au)}{\partial x} = 0, \\ \frac{\partial (Au)}{\partial t} + \frac{\partial (Au^2)}{\partial x} + \frac{A}{\rho} \frac{\partial P}{\partial x} = -8\pi \frac{\mu}{\rho} u, \\ P = P_{ext} + \beta \left[ \left( \frac{A}{A_0} \right)^{1/2} - 1 \right], \quad \beta = \sqrt{\frac{\pi}{A_0}} \frac{Eh_0}{1 - \sigma^2}. \end{cases} \quad (1)$$

where  $t$  is time,  $x$  is the longitudinal space coordinate,  $\rho$  is the blood density,  $\mu$  is the blood dynamic viscosity,  $A_0$  is the reference cross-sectional area (i.e. when  $P - P_{ext} = 0$ ),  $h_0$  is the reference wall thickness,  $E$  is the wall Young's modulus,  $\sigma = 1/2$  is the wall Poisson's ratio, and  $P_{ext}$  is the external pressure. The numerical solution of (1) was achieved by means of a finite-volume scheme [27, 28, 29]. The numerical mesh was independently set for each vessel by using at least 5 elements or, where possible,  $\Delta x = 1$  mm. To guarantee numerical stability, time steps were adaptively computed at each iteration in all vessels depending on the maximum local wave speed  $c_{max}$  and  $\Delta x$  as

$$\Delta t = C_{cfl} \frac{\Delta x}{c_{max}}, \quad (2)$$

where the Courant number was set to  $C_{cfl} = 0.9$  and the  $\Delta t$  set for the entire system was the smallest computed for all vessels in the system.

A boundary condition was applied to the inlet of the root vessel as a flow time-function. Outlet boundaries of peripheral bifurcating vessels were coupled to three-element windkessel models [30]; at capillary level, the last boundary condition was assigned by assuming the arterial-venous interface relative pressure equal to zero. In order to avoid artificial wave reflections induced by discontinuities caused by time-dependent changes in diameters at the bifurcation outlets, windkessel impedances  $\mathcal{Z}$  were calculated at each time step to match daughter vessel outlet impedance.

In order to show how the computational time, the accuracy, and the convergence of the proposed methodology scales with the mechanistic model complexity, four vascular networks were analysed (Figure 1). These represent the iliac bifurcation (8 arteries), the ascending and the upper thoracic aorta (7 arteries), the thoracic aorta and the right arm (15 arteries), and a more complete cardiovascular system (61 arteries). Geometrical values (lumen radius  $R_0$  and vessel length  $\ell$ ), material properties (wall Young's modulus  $E$ ), inlet flow time function (Figure 2(a)), and windkessel parameters (peripheral resistance  $\mathcal{R}$  and peripheral compliance  $\mathcal{C}$ ) were based on data published in [31, 32, 33]. Uncertainty domains for all the parameters were set as in [1] (Table I). Analysis of the results was conducted by extracting two outputs:

the pressure waveforms were computed at each node of the system, and minimum and maximum values at middle point of the root vessel (i.e. the vessel to which the inlet boundary condition is applied) were recorded (Figure 2(b)).

## 2.2. Gaussian process for regression

Gaussian process theory for both regression and classification is well known and established in the machine learning community. In this context, only GP for regression is presented. A complete and exhaustive mathematical description of this method can be found in the literature, see for example [26, 34, 35, 36].

Let us consider a set of  $N$   $d$ -dimensional input vectors  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , and  $N$  corresponding observations  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ . Inputs are mapped into outputs by an unknown function  $f(\mathbf{X})$ . The goal of regression is to learn function  $f$  from dataset  $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\}$ . Since the input data can be uncertain, it is desirable to obtain a set of likely interpolating functions rather than a single function  $f(\mathbf{X})$ .

A GP describes a probability distribution over functions,  $p(f)$ . This conveys the prior knowledge about the interpolating functions before observing data  $\mathcal{D}$ . By definition, the prior of GP has a multivariate Gaussian distribution, mathematically described by

$$p(f) = \mathcal{N}(\mathbf{m}, \Sigma), \quad (3)$$

where  $\mathbf{m} = m(\mathbf{x})$  is the mean vector computed with the mean function  $m$ , and  $\Sigma$  is a covariance matrix. The mean is often set to zero; this is because datasets can be simply pre-processed to have zero mean before training the statistical model. The covariance matrix is built through a kernel function  $k(\mathbf{x}, \mathbf{x}')$  as

$$\Sigma_{ij} = k(\mathbf{x}_i, \mathbf{x}_j). \quad (4)$$

The kernel choice depends on the problem. Complex behaviour may require a combination of kernel functions. Kernel function building starts from widely used kernel functions like the squared exponential

$$k_{SE}(\mathbf{x}, \mathbf{x}') = \theta_1^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\theta_2^2}\right), \quad (5)$$

or the Matérn class of functions

$$k_{M\nu}(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\|\right)^\nu I_\nu\left(\frac{\sqrt{2\nu}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\|\right), \quad (6)$$

where  $\theta_i$  ( $i = 1, \dots, 3$ ) are hyper-parameters,  $\Gamma(\cdot)$  is the Gamma function,  $I_\nu$  is a modified Bessel function, and  $\nu$  is a positive parameter which controls the smoothness of the kernel function [37]. Of particular interest are the cases  $\nu = 3/2$  and  $\nu = 5/2$  for which the kernel function can be simplified as

$$k_{M^{3/2}}(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{3}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\|\right) \exp\left(-\frac{\sqrt{3}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\|\right), \quad (7)$$

$$k_{M^{5/2}}(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{5}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\| + \frac{5}{3\theta_3^2} \|\mathbf{x} - \mathbf{x}'\|^2\right) \exp\left(-\frac{\sqrt{5}}{\theta_3} \|\mathbf{x} - \mathbf{x}'\|\right). \quad (8)$$

Hyper-parameters  $\theta_i$  govern kernel properties, and their selection is part of the optimisation process. The GP optimises hyper-parameters from training data. This is done by applying Bayes' Theorem,

$$p(f|\mathcal{D}) = \frac{p(\mathcal{D}|f)p(f)}{p(\mathcal{D})}, \quad (9)$$

which evaluates the uncertainty in  $f$  after  $\mathcal{D}$  has been observed. The posterior probability  $p(f|\mathcal{D})$ , is obtained by taking into account prior knowledge about  $f$  before seeing  $\mathcal{D}$ ,  $p(f)$ , and by expressing how likely it is to observe  $\mathcal{D}$  for different  $f$ . The effect of the observed data is expressed by the likelihood function  $p(\mathcal{D}|f)$ . The denominator term is a normalisation constant. The likelihood function is used as a cost function and it is maximised with respect to hyper-parameters  $\theta_i$ . This, in turn, maximises the posterior probability of  $f$  describing  $\mathcal{D}$ .

Let us now introduce a new set of test inputs  $\mathbf{X}_T$  and corresponding outputs  $\mathbf{Y}_T$ . The predictive distribution is assumed to have a multivariate Gaussian distribution

$$p(\mathbf{Y}_T|\mathbf{Y}) = \mathcal{N}(\mathbf{m}_T, \mathbf{K}_T), \quad (10)$$

with mean and covariance given by

$$\begin{aligned} \mathbf{m}_T &= \Sigma_{NT}^T \Sigma^{-1} \mathbf{Y}, \\ \mathbf{K}_T &= \Sigma_T - \Sigma_{NT}^T \Sigma^{-1} \Sigma_{NT}, \end{aligned} \quad (11)$$

where  $\Sigma_{NT} = k(\mathbf{x}_n, \mathbf{x}_{N+T})$  for  $n = 1, \dots, N$ . Predictions are made by sampling the distribution (10) at points from  $\mathbf{X}_T$ . Note that the predictive distribution does not have a zero mean function, and that its covariance matrix is computed by using the optimised kernel function learned from the training set  $\mathcal{D}$ .



The training cost has  $\mathcal{O}(N^3)$  complexity, due to the covariance matrix inversion. Nevertheless, this is a one-time operation, and the GP obtained can be used to predict an output at any input point within the input space.

### 2.3. Gaussian process emulator verification

In order to assess the GP prediction error, a Monte Carlo (MC) analysis was performed on the system (1). This consisted of running  $d \times 10^3$  simulations by sampling the  $d$ -dimensional input space. The  $d \times 10^3$  inputs and the  $d \times 10^3 \times 2$  outputs constituted the GP design data  $\mathcal{D}$ . A portion of the dataset was randomly sampled and saved for diagnostic purposes. The GP model was trained on the remaining part. The MC analysis was performed for the networks *I*, *II*, and *III*, but not for the complete model *IV*, as the computational time required would have been prohibitive and estimated around 9 years and 4 months (see Figure 5).

The GP model was implemented by using the GPy library [38]. In order to avoid numerical problems due to bad conditioning of the covariance matrix, training inputs and training outputs were normalised dimension-wise, i.e. each dimension in the input space was separately normalised. The kernel function to compute the covariance matrix was obtained as the sum of a squared exponential kernel and a Matérn  $5/2$  kernel. Hyper-parameter optimisation was conducted by minimising the log-likelihood through the conjugate gradient descent method.

The training sample was varied in size to assess the relation between the prediction error and the sample size. Three diagnostics [39] were used to verify GP model predictions:

1. *Graphical analysis.* Normalised GP model (emulator) predictions were compared with normalised vascular model (simulator) outputs (Figure 3a-c). **The normalised values were computed by first subtracting their mean, by then dividing by their standard deviation, and then combined all together.** Points lying on the dashed line of equality indicate a good agreement between emulator and simulator.
2. *Standardised prediction error.* The differences between simulator outputs ( $\mathbf{Y}$ ) and emulator mean predictions ( $\mathbf{m}[\zeta(\mathbf{X})|\mathbf{Y}]$ ) were quantified as

$$D(\mathbf{Y}) = \frac{\mathbf{Y} - \mathbf{m}[\zeta(\mathbf{X})|\mathbf{Y}]}{\sqrt{\Sigma[\zeta(\mathbf{X})|\mathbf{Y}]}}. \quad (12)$$

The emulator is claimed to be able to represent properly the simulator when its error distribution is normally distributed (Figure 3d-f). The normal distribution having

mean and standard deviation computed from the error distribution is plotted over the actual error distribution. A visual inspection of the error distribution confirms the similarity between the two distributions.

3. *Mean average prediction error (MAPE)*. The MAPE was computed between emulator predictions and simulator outputs for each output of interest (Figure 4) for a different input sample size. In case IV, for which the complete MC analysis would require (Figure 5(c)) a computational time of 9.3 years, the MAPE was computed on a reduced dataset of 1000 simulator runs.

#### 2.4. Sensitivity analysis

Sobol's sensitivity indices are briefly introduced. More mathematical background, is given in [17, 19, 22].

If we consider a model  $y = f(x_1, \dots, x_d)$ , the model total variance reads

$$V(y) = \int_{\mathcal{S}} \left( f(x_1, \dots, x_d) - y_0 \right)^2 d\mathcal{S}, \quad (13)$$

where the term  $y_0$  represents the model mean value and  $\mathcal{S}$  is the input hyperspace. Analysis of variance (ANOVA) decomposition allows the variance (13) to be rewritten as the sum

$$V(y) = \sum_{i=1}^d V_i + \sum_{1 \leq i < j \leq d} V_{ij} + \dots + \sum_{1 \leq i < j < \dots < k \leq d-1} V_{ij\dots k} + V_{12\dots d}, \quad (14)$$

where the partial variances are defined as

$$V_i = \int_{\mathcal{S}_i} f_i^2(x_i) d\mathcal{S}_i, \quad (15)$$

$$V_{ij} = \int_{\mathcal{S}_i} \int_{\mathcal{S}_j} f_{ij}^2(x_i, x_j) d\mathcal{S}_i d\mathcal{S}_j, \quad (16)$$

$$V_{ij\dots k} = \int_{\mathcal{S}_i} \int_{\mathcal{S}_j} \dots \int_{\mathcal{S}_k} f_{ij\dots k}^2(x_i, x_j, \dots, k) d\mathcal{S}_i d\mathcal{S}_j \dots d\mathcal{S}_k, \quad (17)$$

where  $\mathcal{S}_i$  is the domain of  $x_i$ . The partial summands  $f_i$ ,  $f_{ij}$ , and  $f_{ij\dots k}$  are univocally defined as

$$f_i(x_i) = \int_{\mathcal{S}_i^-} f(x_i, \dots, x_d) d\mathcal{S}_i^- - y_0, \quad (18)$$

$$f_{ij}(x_i, x_j) = \int_{\mathcal{S}_{ij}^-} f(x_i, \dots, x_d) d\mathcal{S}_{ij}^- - f_i(x_i) - f_j(x_j) - y_0, \quad (19)$$

$$f_{ij\dots k}(x_i, x_j, \dots, x_k) = \int_{\mathcal{S}_{ij\dots k}^-} f(x_i, \dots, x_d) d\mathcal{S}_{ij\dots k}^- - f_i(x_i) - f_j(x_j) - \dots - f_k(x_k) - y_0, \quad (20)$$

where  $\mathcal{S}_i^-$  is the input hyperspace except the domain of  $x_i$ , i.e.  $\mathcal{S} = \mathcal{S}_i \cup \mathcal{S}_i^-$ .

The partial variances in (14), normalised with the total variance  $V(y)$ , return sensitivity indices. First-order sensitivity indices  $S_i$  measure the amount of variance that can be attributed to variance on each input. The variance in the output due to interaction between variances of two inputs is given by second-order sensitivity indices  $S_{ij}$ . The sum of all the indices regarding input  $i$  is called the total sensitivity index  $T_i$ . The influence of input  $i$  on the output  $y$  due to various input interactions is measured by the difference between  $T_i$  and  $S_i$ . By construction, sensitivity indices are smaller than 1 and can be interpreted as proportions.

The sensitivity analysis for parameter fixing and prioritisation was performed following a three step strategy [16]:

- I The mechanistic model was run with a small set of input values spanning evenly the entire parameter space. Simulation inputs and outputs constituted the dataset  $\mathcal{T}$  on which the GP emulators were trained.  $\mathcal{T}$  was designed through orthogonal Latin hypercube sampling method to ensure an even coverage of the input space. In order to avoid an ill-conditioned covariance matrix, inputs and outputs were normalised.
- II The trained GP was then used to predict outcomes for a bigger set of inputs with size of order  $\mathcal{O}(d \times 10^3)$ .
- III Sobol's sensitivity indices were computed by means of ANOVA decomposition. Inputs were ranked accordingly to first-order indices. The largest first-order indices indicated those inputs which mainly affect the outcome. Any differences between first-order and total sensitivity indices indicated that the outcome variance could be ascribed to covariance of more than one input.

### 3. RESULTS AND DISCUSSION

The time spent to run a single simulation with the mechanistic solver,  $t_s$ , for each network is reported in Figure 5a. The trained GP emulators were used to predict results for all the  $d \times 10^3$  input points needed for the sensitivity analysis (Figure 5b). These input points were chosen to explore thoroughly the input hyperspace. The sensitivity analysis computational running times,  $t_{SA}$ , of both the numerical vascular model and the GP emulator are reported in Figure 5c. The computational time in the MC analysis added up to 2.5 days for the single bifurcation, and it was estimated to scale up to about 9 years in the case of the complete vascular model. The GP computational time for both training and prediction phases increased as the number of vessels in the network increased. The  $t_{SA}$  was always four orders of magnitude smaller with respect to the MC approach. By coupling the numerical model with the GP regression model, the bulk of the computational time was taken by numerical simulations on the dataset used for training (Table II), e.g. 14 hours for the 61 arteries model. Predictions for the  $d \times 10^3$  datasets were made in 0.42 seconds by the GP emulator. Therefore, the sensitivity analysis on the complete model was performed in 14 hours.

The GP prediction error decreased as the number of points in the training sample increased (Figure 4). The number of training points needed to score a MAPE lower than 1% was always lower than the number of points needed to perform the MC analysis (Table II). These results indicated that the number of points needed to train the GP is of the order  $\mathcal{O}(d)$ .

In the case of the complete model, the ensemble of input points and predicted outputs was used to compute first-order Sobol' sensitivity indices (Figure 6a-e). To ease the analysis, the indices were subdivided in five sets depending on the vessel location as in [2]. The vessel length and the peripheral compliance ( $\ell$  and  $\mathcal{C}$ , respectively) scored low first-order sensitivity indices (less than 0.02, Figure 6b,e). Upper limbs vessel radii and Young's moduli slightly affected ( $S_{R_0,E} \sim 0.05$ ) the maximum pressure in the ascending aorta whereas they had no effect on the minimum pressure (Figure 6a,c). Aorta vessel radii and Young's moduli were the topological parameters to have the most effect on the maximum pressure ( $S_{R_0,E} \sim 0.1$ ). The peripheral resistance ( $\mathcal{R}$ ) variation at organ vessels affected the variation of the maximum and minimum pressure ( $S_{\mathcal{R}} \sim 0.25$ ), whereas the upper part of the network (upper limbs and neck vessels) has a slightly lower effect ( $0.05 < S_{\mathcal{R}} < 0.15$ ) on both minimum and maximum pressure at the ascending aorta (Figure 6d). In order to show the sensitivity indices spatial distribution, the first order indices relating the maximum pressure variation to the variation

of the Young's modulus were converted to luminance values and plotted in a heat map (Figure 6f).

The choice of different inlet boundary conditions for the networks investigated did not seem to affect the main outcome of the study, as computational time reduced linearly with the number of vessels for all networks regardless of the conditions set at their inlets (Figure 5(c)). In addition, the same prediction accuracy, as measured by the MAPE (Figure 4), was achieved for a sample size proportional to the number of vessels rather than to the type of flow time-function used. Nonetheless, the inlet boundary condition is a source of uncertainty due to its large variability between individuals. In future studies aimed at finding clinical bio-mechanical markers, the inlet flow function will be included as a GP input.

#### 4. CONCLUSIONS

One-dimensional models of the cardiovascular system provide an accurate description of the physics of wave transmission in blood and can be used to provide realistic or patient-specific pulse and flow rate waveforms. Their mathematical description relies on the specification of a large number of parameters, which are often not readily available as typical or patient-specific values. Many of these cannot be specified as a constant value either, as they will vary within the physiological envelope of an individual. In this context, ranking and fixing of parameters through sensitivity analysis has been previously proposed as a way to focus on the most influential model inputs or simply to quantify input uncertainty on variables of interest. However, these operations may require many simulations, resulting in large or infeasible computational time.

A time efficient approach to sensitivity analysis is proposed in this paper. A reduced number of model simulations were used to train a GP regression model. This emulator was able to mimic the vascular numerical model with a percentage error lower than 1% when compared to the actual model runs. The emulator running time was also much shorter than the simulator: the GP prediction phase for network IV took 0.19s as opposed to an average of 197s required for a single run of the deterministic model. More in general, the adoption of a GP emulator in the sensitivity analysis framework allowed for a minimum reduction of computational time by 99.96% compared to Monte Carlo analysis.

This framework scalability was tested by developing four vascular models of increasing complexity, i.e. starting from a single bifurcation, the number of vessels was increased up to

61 in the case of a complete vascular model. In all the four cases, by introducing the GP, the simulator runs needed for the sensitivity analysis is reduced from  $d \times 10^3$  to  $\mathcal{O}(d)$ .

The analysis of sensitivity indices allowed us to identify the location in the network of model parameters affecting maximum and minimum pressures in the ascending aorta. In particular, the minimum pressure was affected by changes in the peripheral resistance of organ arteries. The maximum pressure was sensitive to changes in the aorta Young's modulus as well as in the upper limbs arteries.

The introduction of a GP regression model as an output generator for a mechanistic model is a novel approach in the cardiovascular research community. The conclusions drawn from sensitivity analysis are not novel, but they confirm that the developed framework is sound, and it is capable of capturing the intrinsic non linear behaviour of flows through a vascular network. Running times were drastically reduced when using the emulator approach, which allowed a thorough sensitivity analysis with comparable accuracy to a much more time consuming approach. The study of model sensitivity indices gave an insight into how the inputs interact and could be used to study how input uncertainty propagates through to the outputs. The same approach has the potential to improve efficiency in the analysis of more complex and complete models of the cardiovascular systems.

#### ACKNOWLEDGEMENT

We gratefully acknowledge funding from the UK Engineering and Physical Sciences Research Council (Grant Number EP/K037145/1).

#### REFERENCES

1. Huberts W, de Jonge C, van der Linden WP, Inda MA, Tordoir JH, van de Vosse FN, Bosboom EM. A sensitivity analysis of a personalized pulse wave propagation model for arteriovenous fistula surgery. Part A: Identification of most influential model parameters. *Medical engineering & physics* 2013;**35**(6):810-26, DOI: 10.1016/j.medengphy.2012.08.012.
2. Eck VG, Feinberg J, Langtangen HP, Hellevik LR. Stochastic sensitivity analysis for timing and amplitude of pressure waves in the arterial system. *International journal for numerical methods in biomedical engineering* 2015; **31**(4), DOI: 10.1002/cnm.2711.
3. Willemet M, Alastruey J. Arterial pressure and flow wave analysis using time-domain 1-D hemodynamics. *Annals of biomedical engineering* 2015;**43**(1):190-206, DOI: 10.1007/s10439-014-1087-4.
4. Shi Y, Lawford P, Hose R. Review of zero-D and 1-D models of blood flow in the cardiovascular system. *BioMedical Engineering OnLine* 2011; **10**(1):33, DOI: 10.1186/1475-925X-10-33.

5. Steinman DA, Taylor CA. Flow imaging and computing: large artery hemodynamics. *Annals of biomedical engineering* 2005;**33**(12):1704-1709, DOI: 10.1007/s10439-005-8772-2.
6. Sankaran S, Moghadam ME, Kahn AM, Tseng EE, Guccione JM, Marsden AL. Patient-specific multiscale modeling of blood flow for coronary artery bypass graft surgery. *Annals of biomedical engineering* 2012;**40**(10):2228-2242, DOI: 10.1007/s10439-012-0579-3.
7. Čanić S, Kim EH. Mathematical analysis of the quasilinear effects in a hyperbolic model blood flow through compliant axi-symmetric vessels. *Mathematical Methods in the Applied Sciences* 2003; **26**(14):1161-1186, DOI: 10.1002/mma.407.
8. Čanić S, Hartley CJ, Rosenstrauch D, Tambača J, Guidoboni G, Mikić A. Blood flow in compliant arteries: an effective viscoelastic reduced model, numerics, and experimental validation. *Annals of Biomedical Engineering* 2006; **34**(4):575-592, DOI: 10.1007/s10439-005-9074-4.
9. Formaggia L, Quarteroni A, Veneziani A. (eds.) *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*. Vol. 1. Springer Science & Business Media, 2010, DOI: 10.1007/978-88-470-1152-6.
10. Sherwin SJ, Formaggia L, Peiró J, Franke V. Computational modelling of 1D blood flow with variable mechanical properties and its application to the simulation of wave propagation in the human arterial system. *International Journal for Numerical Methods in Fluids* 2003; **43**(6):673-700, DOI: 10.1002/flid.543.
11. Azer K, Peskin CS. A one-dimensional model of blood flow in arteries with friction and convection based on the Womersley velocity profile. *Cardiovascular Engineering* 2007; **7**(2):51-73, DOI: 10.1007/s10558-007-9031-y.
12. Reymond P, Merenda F, Perren F, Rüfenacht D, Stergiopoulos N. Validation of a one-dimensional model of the systemic arterial tree. *American Journal of Physiology-Heart and Circulatory Physiology* 2009; **297**(1):H208-H222, DOI: 10.1152/ajpheart.00037.2009.
13. Huo Y, Kassab GS. A hybrid one-dimensional/Womersley model of pulsatile blood flow in the entire coronary arterial tree. *American Journal of Physiology-Heart and Circulatory Physiology* 2007; **292**(6):H2623-H2633, DOI: 10.1152/ajpheart.00987.2006.
14. Alastruey J, Parker KH, Peiró J, Byrd SM, Sherwin SJ. Modelling the circle of Willis to assess the effects of anatomical variations and occlusions on cerebral flows. *Journal of biomechanics* 2007; **40**(8):1794-1805, DOI: 10.1016/j.jbiomech.2006.07.008.
15. Lungu A, Wild JM, Capener D, Kiely DG, Swift AJ, Hose DR. MRI model-based non-invasive differential diagnosis in pulmonary hypertension. *Journal of biomechanics* 2014; **47**(12):2941-2947, DOI: 10.1016/j.jbiomech.2014.07.024.
16. Santner TJ, Williams BJ, Notz WI. *The design and analysis of computer experiments*. Springer Science & Business Media, 2013, DOI: 10.1007/978-1-4757-3799-8.
17. Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation* 2001; **55**(1):271-280, DOI: 10.1016/S0378-4754(00)00270-6.
18. Saltelli A. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 2002; **145**(2):280-297, DOI: 10.1016/S0010-4655(02)00280-1.
19. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 2010; **181**(2):259-270, DOI: 10.1016/j.cpc.2009.09.018.
20. Olsen, C.H., Tran, H., Ottesen, J.T., Mehlsen, J. and Olufsen, M.S. Challenges in practical computation of global sensitivities with application to a baroreceptor reflex model. *PLOS Computational Biology (online)* 2015.

21. Hussain MF, Barton RR, Joshi SB. Metamodeling: radial basis functions, versus polynomials. *European Journal of Operational Research* 2002; **138**(1):142-154, DOI: 10.1016/S0377-2217(01)00076-5.
22. Huberts W, Donders WP, Delhaas T, Vosse van de FN. Applicability of the polynomial chaos expansion method for personalization of a cardiovascular pulse wave propagation model. *International journal for numerical methods in biomedical engineering* 2014; **30**(12):1679-1704, DOI: 10.1002/cnm.2695.
23. Donders WP, Huberts W, Vosse van de FN, Delhaas T. Personalization of models with many model parameters: an efficient sensitivity analysis approach. *International journal for numerical methods in biomedical engineering* 2015; **31**(10), DOI: 10.1002/cnm.2727.
24. Ellwein, L.M., Tran, H.T., Zapata, C., Novak, V. and Olufsen, M.S. Sensitivity analysis and model assessment: mathematical models for arterial blood flow and blood pressure. *Cardiovascular Engineering* 2008;**8**(2):94-108, DOI: 10.1007/s10558-007-9047-3.
25. O'Hagan A. Polynomial chaos: A tutorial and critique from a statistician's perspective. *SIAM/ASA Journal of Uncertainty Quantification* 2013; **201**-20.
26. Rasmussen CE, Williams CKI. *Gaussian processes for machine learning*. The MIT Press, 2006.
27. Toro EF. A fast Riemann solver with constant covolume applied to the random choice method. *International Journal for Numerical Methods in Fluids* 1989; **9**(9):1145-1164, DOI: 10.1002/fld.1650090908.
28. Toro EF. *Shock-capturing methods for free-surface shallow flows*. Wiley, 2001.
29. Toro EF. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Springer Science & Business Media, 2013, DOI: 10.1007/b79761.
30. Fernández MÁ, Milisic V, Quarteroni A. Analysis of a geometrical multiscale blood flow model based on the coupling of ODEs and hyperbolic PDEs. *Multiscale Modeling & Simulation* 2005; **4**(1):215-236, DOI: 10.1137/030602010.
31. Xiao N, Alastruey J, Alberto Figueroa C. A systematic comparison between 1D and 3D hemodynamics in compliant arterial models. *International journal for numerical methods in biomedical engineering* 2014;**30**(2):204-31, DOI: 10.1002/cnm.2598.
32. Brown AG, Shi Y, Marzo A, Staicu C, Valverde I, Beerbaum P, Lawford PV, Hose DR. Accuracy vs. computational time: translating aortic simulations to the clinic. *Journal of biomechanics* 2012;**45**(3):516-23, DOI:10.1016/j.jbiomech.2011.11.041.
33. Boileau E, Nithiarasu P, Blanco PJ, MÃijller LO, Fossan FE, Hellevik LR, Donders WP, Huberts W, Willemet M, Alastruey J. A benchmark study of numerical schemes for one-dimensional arterial blood flow modelling. *International journal for numerical methods in biomedical engineering* 2015;**31**(10), DOI:10.1002/cnm.2732.
34. Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2004; **66**(3):751-769, DOI: 10.1111/j.1467-9868.2004.05304.x.
35. Bishop CM. *Pattern Recognition and Machine Learning*. Springer, 2006.
36. O'Hagan A. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety* 2006; **91**(10):1290-1300, DOI: 10.1016/j.res.2005.11.025.
37. Genton MG. Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research* 2001;**2**:299-312.
38. The GPy authors, *GPy: A Gaussian process framework in python*. <http://github.com/SheffieldML/GPy>, 2012-2016.



39. Bastos LS, O'Hagan A. Diagnostics for Gaussian process emulators. *Technometrics* 2009; **51**(4):425-438, DOI: 10.1198/TECH.2009.08019.

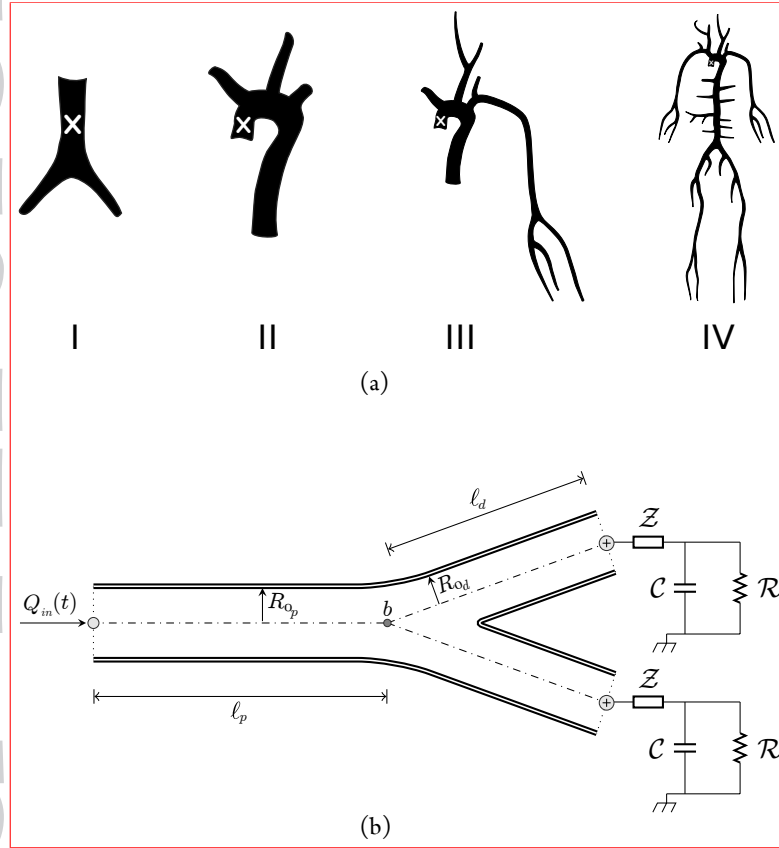


Figure 1. (a) Diagrams of the four vascular **networks** used in the scalability study: (I) iliac bifurcation [31], (II) ascending and thoracic aorta [32], (III) thoracic aorta and right arm [32, 33], (IV) complete model of the main 61 arteries adapted from [33]. **The deterministic model outputs were extracted at the middle point of the root vessel of each network (white × marker).** (b) Symmetric iliac bifurcation detailed model scheme. For each vessel in any of the four networks (I – IV), the length  $\ell$ , lumen radius  $R_0$ , and wall Young's modulus  $E$  are defined. At the outlets, the windkessel model requires the peripheral resistance  $\mathcal{R}$  and the peripheral compliance  $\mathcal{C}$  to be defined.

Table I. Simulation parameters uncertainty domain [1]:  $\ell$  vessel length,  $R_0$  lumen radius,  $E$  wall Young's modulus,  $\mathcal{R}$  peripheral resistance, and  $\mathcal{C}$  peripheral resistance.

$\ell$ (%)	$R_0$ (%)	$E$ (%)	$\mathcal{R}$ (%)	$\mathcal{C}$ (%)
-10, 10	-10, 10	-20, 20	-25, 25	-50, 50

Table II. Gaussian process and Monte Carlo training sample size ( $N_{GP}$ ,  $N_{MC}$ ) for the four vascular **networks**. The time needed by the GP method to perform the sensitivity analysis predictions is reported as percent of the time taken by the MC,  $t_{SAMC}$ %. For each network, its complexity is reported in terms of the number of input parameters  $d$ .

	$d$	$N_{GP}$	$N_{MC}$	$t_{SAMC}$ (%)
I	8	17	8000	0.216
II	29	34	29000	0.120
III	61	47	61000	0.075
IV	245	500	245000	0.040

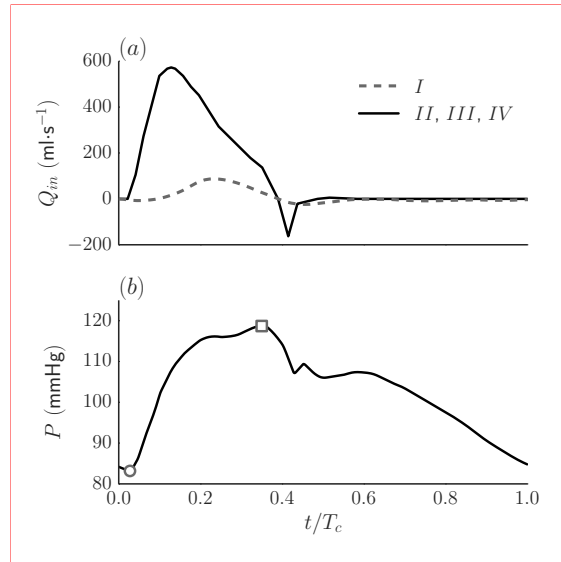


Figure 2. (a) Typical inlet flow time functions for the four networks, taken from [33]. (b) Example of computed pressure waveform at the middle point of the root vessel of network IV. The waveforms are reported in function of the time normalised with respect to the cardiac cycle period  $T_c$ . The minimum (circle) and the maximum (square) pressure values are the waveform features taken as outputs of the deterministic model.

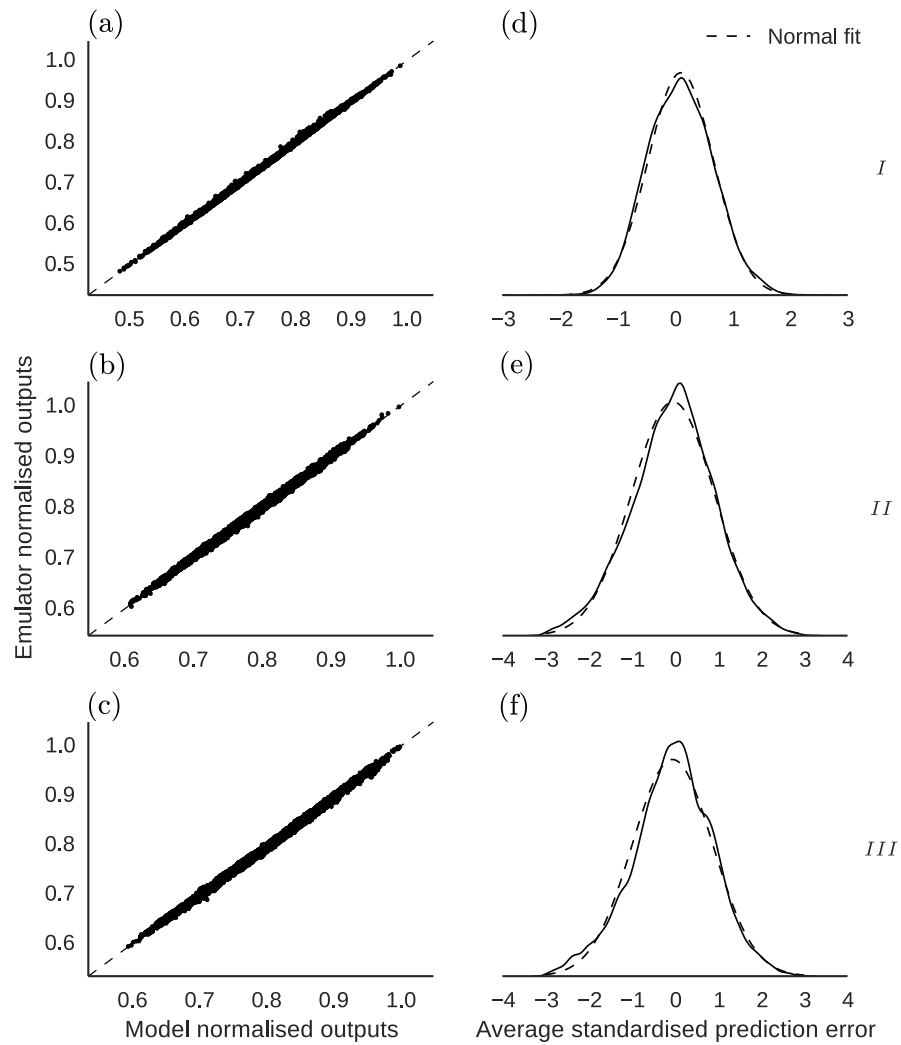


Figure 3. Gaussian process validation diagnostics for the first three vascular networks (*I – III*). (a-c) Graphical comparison between emulator predictions and simulator outputs; points lying on the dashed line of equality indicate a good agreement between the two techniques. (d-f) A properly trained emulator is expected to have standardised prediction errors normally distributed.

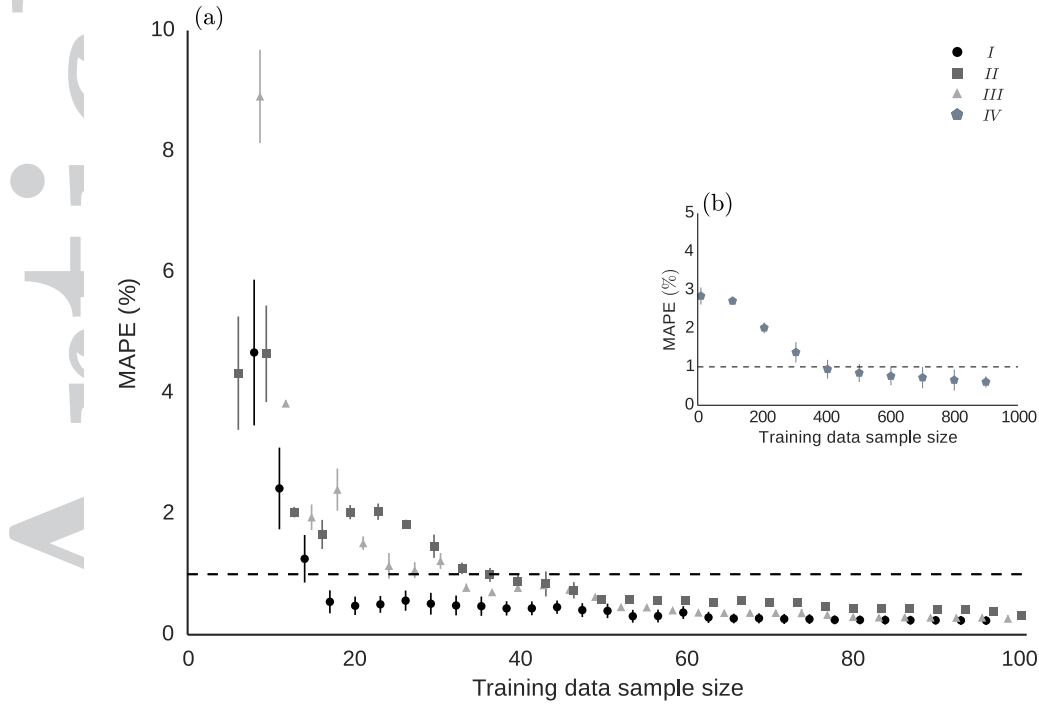


Figure 4. Mean average prediction error (MAPE) between emulator predictions and simulator outputs for each input of interest on the first three networks (a) and for the complete model (b). The MAPE decreases as the number of points in the training sample increases.

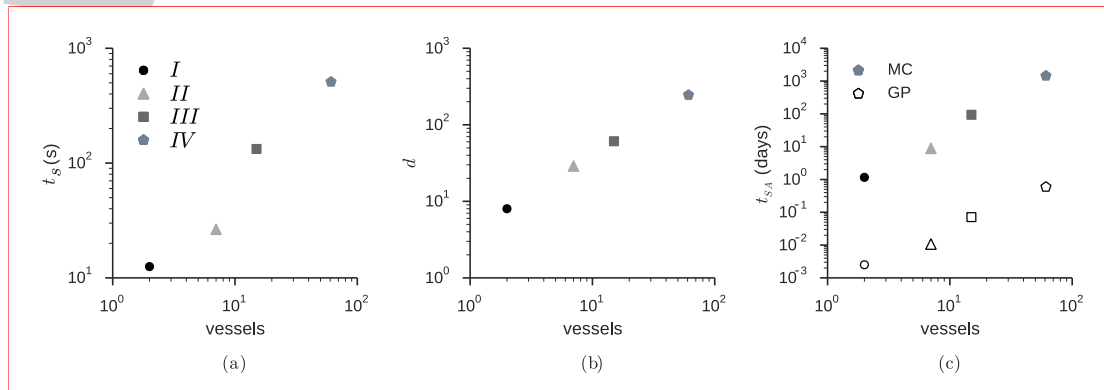


Figure 5. (a) Computational time required by the mechanistic solver to compute the flow solution for each of the four networks ( $I - IV$ ). The computational time increases as more vessels are added to the network. (b) Number of input parameters in the vascular model to be studied with the sensitivity analysis as the number of vessels increases. (c) Total computational time required to compute all the  $d \times 10^3$  simulation for the sensitivity analysis by both the Monte Carlo (MC, filled markers) and the Gaussian process (GP, empty markers) methods.

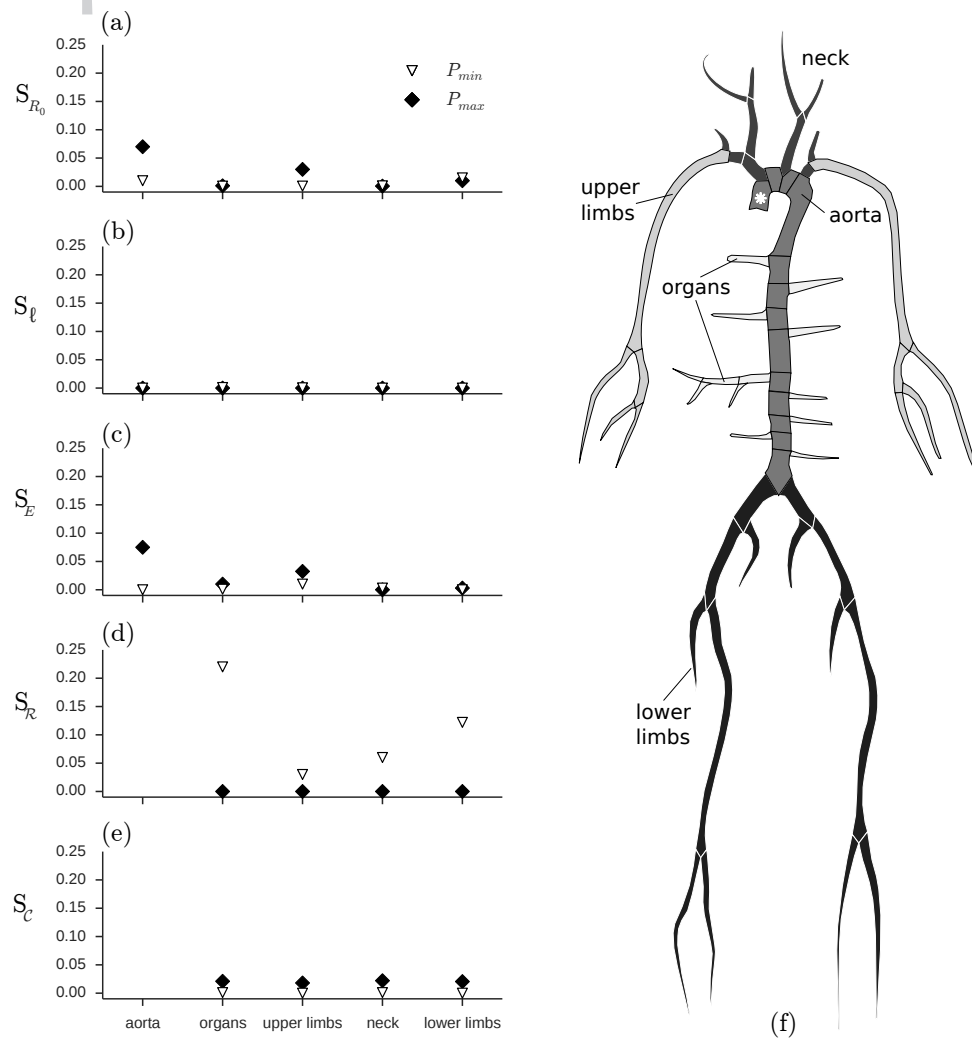


Figure 6. (a-e) First-order sensitivity indices ( $S_i$ ) relating the variation of minimum and maximum pressure (the outputs of the deterministic model,  $P_{min}$  and  $P_{max}$ , respectively) at the ascending aorta location (\* marker in (f)) with the variation of the lumen radius  $R_0$ , vessel length  $\ell$ , wall Young's modulus  $E$ , peripheral resistance  $\mathcal{R}$ , and peripheral compliance  $\mathcal{C}$  for five groups of arteries. (f) Diagram of the 61 vessels vascular network. The five groups of arteries used for the sensitivity analysis are indicated by different colours.