

RESEARCH ARTICLE

A physiologically realistic virtual patient database for the study of arterial haemodynamics

G. Jones¹ | J. Parr² | P. Nithiarasu¹ | S. Pant^{†1}¹College of Engineering, Swansea University, UK.²McLaren Technology Centre, Woking, UK.**Correspondence**[†]Email: Sanjay.Pant@swansea.ac.uk**Present Address**

College of Engineering, Swansea University, UK.

Summary

This study creates a physiologically realistic virtual patient database (VPD), representing the human arterial system, for the primary purpose of studying the effects of arterial disease on haemodynamics. A low dimensional representation of an anatomically detailed arterial network is outlined, and a physiologically realistic posterior distribution for its parameters constructed through the use of a Bayesian approach. This approach combines both physiological/geometrical constraints and the available measurements reported in the literature. A key contribution of this work is to present a framework for including all such available information for the creation of virtual patients (VPs). The Markov Chain Monte Carlo (MCMC) method is used to sample random VPs from this posterior distribution, and the pressure and flow-rate profiles associated with each VP computed through a physics based model of pulse wave propagation. This combination of the arterial network parameters (representing a virtual patient) and the haemodynamics waveforms of pressure and flow-rates at various locations (representing functional response and potential measurements that can be acquired in the virtual patient) makes up the VPD. While 75,000 VPs are sampled from the posterior distribution, 10,000 are discarded as the initial burn-in period of the MCMC sampler. A further 12,857 VPs are subsequently removed due to the presence of negative average flow-rate, reducing the VPD to 52,143. Due to undesirable behaviour observed in some VPs—asymmetric under- and over-damped pressure and flow-rate profiles in left and right sides of the arterial system—a filter is proposed to remove VPs showing such behaviour. Post application of the filter, the VPD has 28,868 subjects. It is shown that the methodology is appropriate by comparing the VPD statistics to those reported in literature across real populations. Generally, a good agreement between the two is found while respecting physiological/geometrical constraints. The pre-filter database is made available at <https://doi.org/10.5281/zenodo.4549764>.

KEYWORDS:

virtual patients, stenosis, aneurysm, pulse wave haemodynamics, MCMC, screening, virtual patient database

1 | INTRODUCTION

Arterial disease refers to any disease affecting the arterial system. Two of the most common forms of arterial disease are stenosis and aneurysm. These two forms of disease are estimated to affect between 1% and 20% of the population^{1,2,3,4}. Ruptured abdominal aortic aneurysms alone are estimated to be the cause of between 6,000 and 8,000 deaths a year within the United Kingdom⁵. Current methods for the detection of arterial disease are primarily based on imaging techniques which can be expensive and thus unsuitable for mass screening. A potential solution is offered by the recent emergence of data-driven Machine Learning (ML) algorithms which can be trained on easily acquirable pressure and flow-rate measurements. The key idea is that the presence of disease—both stenosis (a reduction in vessel area) and aneurysm (an increase in vessel area)—will create changes in such measurements, which may be detectable by the ML algorithms. To explore this possibility, however, a large database of both healthy and diseased subjects along with their respective measurements is necessary. This study presents a computational method for the creation of such a physiologically realistic database.

As opposed to acquiring measurements from a real population, which may be impractical, a virtual patient database (VPD) based on a well-known physics-based model is proposed. This idea of exploiting virtual patients (VPs) to gain meaningful information has been previously explored in the context of human arterial networks^{6,7,8}, and in the context of path clamp physiology⁹. Willemet et al.⁶ created a VPD of healthy subjects by varying seven parameters of a large 55-artery network to assess the relation between foot-to-foot pulse wave velocities and aortic stiffness. Chakshu et al.⁷ created a VPD of arterial networks by considering the empirical relationships between the network parameters and patient characteristics such as weight, height, and age. Jones et al.⁸, randomly varied nine parameters of a small network of the abdominal aortic bifurcation to create a VPD and subsequently applied ML algorithms to detect the presence of stenosis. The limitations of these studies are that either not all the parameters are varied in both healthy and diseased states, or *a priori* distributions are assumed for the parameters. Both of these are important considerations for the creation of a general purpose realistic VPD which captures the variability as close as possible to the real population. Furthermore, the VPD must account for known variation of the measurements reported in the literature along with the physical constraints, while making minimal assumptions. This study presents a probabilistic framework for creation of such a VPD based on a previously proposed, anatomically correct, 56-vessel and 77-segment arterial network^{10,11}. The proposed framework may also be useful for the novel and active research area of in-silico clinical trials¹², for which generation of realistic virtual patients is a key consideration. While not in the context of the subject area explored in this study, Bayesian techniques have been widely used in biomedical engineering for network modelling¹³, uncertainty quantification^{14,15}, parameter estimation & inference^{16,17,18,19,20,21,22,23,24,25}, optimisation²⁶, and optimal design²⁷.

This article is organised as follows. First, a brief description of the physics-based pulse wave propagation model is presented. This is followed by a description of the arterial network topology—vessels, locations of disease and potential measurements—and its reduction. Next, parameterisation of the arterial network—geometry, vessel properties, and boundary conditions—is presented, followed by the Bayesian framework for the VPD creation. Finally, the solution to the Bayesian problem through Markov Chain Monte Carlo method is presented, before discussing the results and statistics of the VPD.

2 | PULSE WAVE PROPAGATION MODEL

A validated one dimensional model of pulse wave propagation^{28,29} is adopted in this study. The model describes the arterial vessels as one-dimensional compliant tubes, i.e. all the properties are described by a single coordinate x along the length of the vessel. Blood flow is assumed to be Newtonian, laminar, incompressible, and axisymmetric. With the arterial cross-sectional area represented as A , and considering the average lumen pressure P and the average velocity U to be the primary variables, the mass and momentum conservation equations at any coordinate x can be, respectively, written as

$$\frac{\partial A}{\partial t} + \frac{\partial(UA)}{\partial x} = 0, \quad (1)$$

$$\frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} + \frac{1}{\rho} \frac{\partial P}{\partial x} = \frac{f}{\rho A}, \quad (2)$$

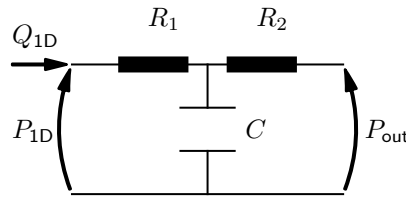


FIGURE 1 Configuration of a three-element Windkessel model. Q_{1D} and P_{1D} represent the volumetric flow-rate and pressure at the terminal boundary of the 1D system respectively.

where ρ represents blood density and f represents the frictional force per unit length. Depending on the assumption of velocity profile across the arterial cross section, f can be written as

$$f(x, t) = -2(\zeta + 2)\mu\pi U, \quad (3)$$

where ζ is a velocity-profile dependant constant, and μ represents dynamic viscosity of blood. Throughout this study, a parabolic profile assumption is used²⁹, giving $\zeta = 2$. The system of equations is closed by specification of the mechanical behaviour of the wall, commonly referred to as the “tube law”. A commonly used form²⁹ is

$$P - P_{ext} = P_d + \frac{\beta}{A_d}(\sqrt{A} - \sqrt{A_d}), \quad (4)$$

with

$$\beta = \frac{4}{3}Eh\sqrt{\pi}, \quad (5)$$

where P_{ext} represents the external pressure, P_d represents the diastolic blood pressure, A_d represents the diastolic area of the vessel, E represents the vessel wall’s Young’s modules, and h represents the vessel wall’s thickness. This system of equations has been widely employed, tested extensively, and validated against more complex three-dimensional models^{29,30,28,31,32,33}.

Since it is computationally impractical to model every single artery in the circulatory network, it is common to truncate the 1D network by only considering relatively large arteries which are of interest. The effect of all the vessels that come before and after the network are then incorporated as appropriate boundary conditions. Typically, at the inlet of the network a volumetric flow-rate is specified, and at the outlets Windkessel models³⁴ are employed. A three-element Windkessel model, shown in Figure 1, comprises of two resistors, R_1 and R_2 , which replicate the viscous resistance of large arteries and the micro-vascular system, respectively, and a capacitor C , which replicate the compliance of large arteries.

The 1D model is solved numerically, including interfacing conditions between vessels and boundary conditions, by the subdomain collocation (SDC) scheme^{35,36}, often referred as the enhanced trapezoidal rule method (ETM). This scheme is chosen for its computational efficiency. The SDC implementation is validated in-house against a discontinuous Galerkin (DCG) scheme²⁸, which in turn has been successfully validated against a 3D model of blood-flow through stenosed arterial vessels³⁷.

3 | ARTERIAL NETWORK AND BOUNDARY CONDITIONS

3.1 | Initial network

The arterial network used as a basis for the creation of VPs is based on a version of the ADAN network²⁹. This network contains 56 vessels within the human arterial system, described by 77 arterial vessel segments. The topology of this network is shown in Figure 2, and each vessel within the network is identified in the Supplementary Material Appendix A. The baseline properties of this network can be found in Section 4.6 and Tables XII and XIII of Boileau et al²⁹. This network is described by:

- **38 vessel segments** with constant reference radii (r_0) and vessel wall mechanical properties (β) along their lengths. The description of each of these segments requires specification of three parameters: the length L , reference radius r_0 , and mechanical property β of the segment.

- **39 vessel segments** with linearly varying r_0 and β along their lengths. The description of each of these segments requires specification of five parameters: the length L ; the reference radii r_0 at the proximal and the distal ends of the segment; and β at the proximal and distal ends.
- **31 Windkessel models.** Every Windkessel model at the outlets requires specification of 2 resistances and a compliance.
- **One inlet flow-rate profile.** In this study, the inlet flow-rate is parameterised by a 5th order Fourier Series (see Section 4.3.1 and⁸). This requires specification of 11 coefficients.

Thus, to describe a VP through direct specification of all the above parameters results in the total dimensionality of 413 per VP. This high dimensionality is problematic leading to increased complexity in creating VPs³⁸. In what follows, methods to reduce the dimensionality of VP description are presented. These are primarily related to either reducing the network or employing a parsimonious parameterisation. Since the primary purpose of the VPD is deployment of ML methods for screening of aneurysm and stenosis through easily acquirable measurements, it is important that the reduction in dimensionality does not compromise i) the locations where disease and measurements are possible, and ii) the precision and variability of measurements at such locations.

3.2 | Important locations

A review of literature is carried out to understand both where disease is likely to occur and where measurements can be obtained. The latter are restricted to locations at which continuous profiles can be recorded. While arterial disease can occur in a large number of vessels, for this study vessels with only high prevalence of stenosis or aneurysm are considered. In this study, unhealthy VPs (containing either a stenosis or an aneurysm) are not created. Instead, the purpose of understanding where disease is likely to occur is to allow for re-parameterisation of the network and its reduction without compromising the pressure and flow-rate profiles at these locations.

3.2.1 | Locations of measurements

Based on literature, the following locations where measurements can be taken are identified:

Pressure in the radial and common carotid arteries: Continuous non-invasive arterial blood pressure profiles can be obtained in the radial and common carotid arteries using applanation tonometry^{39,40}. Applanation tonometry is already clinically used⁴¹. The right and left radial and common carotid arteries are identified in Figure 2 as vessels 8 and 22; and 5 and 14, respectively.

Pressure in the brachial arteries: It is possible to estimate continuous blood pressure at the brachial arteries through reconstruction of finger arterial pressure⁴². Although the use of a model to estimate brachial pressure will introduce errors, these recreated brachial pressure profiles have been shown to meet the requirements set by the Association for the Advancement of Medical Instrumentation^{42,43}. The right and left brachial arteries are identified in Figure 2 as vessels 7c and 21c respectively.

Flow-rate in brachial, carotid, and femoral arteries: Using Doppler ultrasound techniques it is possible to measure blood velocity in both the upper and lower extremities. Doppler ultrasound has been shown to work on the brachial⁴⁴, common carotid⁴⁵, and femoral⁴⁶ arteries. The first and second, right and left femoral arterial segments are identified in Figure 2 as vessels 50b and 53a; and 56b and 59a respectively.

3.2.2 | Locations of disease

Four of the most common forms of arterial disease are subclavian artery stenosis (SAS), carotid artery stenosis (CAS), peripheral arterial disease (PAD), and abdominal aortic aneurysm (AAA).

SAS locations: The ADAN network splits the subclavian arteries into two segments. The right and left instances of the first and second subclavian artery segments are identified in Figure 2 as vessels 4 and 7a; and 18 and 21a respectively.

CAS locations: The carotid arteries consist of the common carotid, external carotid, and internal carotid segments. While the narrowing of an arterial vessel can occur within any of the three carotid segments⁴⁷, it is chosen to limit its occurrence to the common carotid arteries. The right and left instances of the common carotid arteries are identified in Figure 2 as vessels 5, and 14 respectively.

PAD locations: The third frequent form of stenosis is PAD, referring to the stenosis of any peripheral vessel. Isolating arterial vessels that are likely to experience stenosis at a high prevalence in patients with PAD is more difficult relative to SAS or CAS. Both SS and carotid artery stenosis have short and relatively easy to define lists of possible vessels they can affect. PAD, on the other hand, can cover a large range of different vessels depending on the definition prescribed. Studies into PAD primarily focus on the lower extremities^{48,49,50} and thus it is assumed that VPs with PAD experience change in these vessels. Based on⁵⁰ and⁴⁹, PAD in VPs is restricted to the common iliacs, external iliacs, first and second femoral segments, and the first popliteal segments. The right and left instances of each of these vessels are identified in Figure 2 as vessels 48 and 49; 50a and 56a; 50b and 56b; 53a and 59a; and 53b and 59b, respectively.

AAA locations: The ADAN network splits the abdominal aorta into five segments. These five segments are identified in Figure 2 as vessels 35, 41, 43, 45, and 47 respectively.

3.3 | Network reduction

In this section, the reduction of the network by removing vessels while preserving the disease and measurement locations is presented. It is important, however, to ensure that the removal of any vessels does not have a significant impact on the upstream pressure and flow-rate profiles. Network reduction can be performed by removing terminal vessels and merging them into the lumped parameter (Windkessel) boundary conditions⁵¹. This process is adopted in this study and is summarised in Figure 3. The following three groups of vessels eligible for removal are identified:

- **Group 1:** The first and second splenic segments; the left gastric; and the common hepatic (identified in Figure 2 as vessels 37 and 39; 38; and 36).
- **Group 2:** The common interosseous, the posterior interosseous, and the second ulnar segment (identified within Figure 2 as vessels 10a and 24a; 10b and 24b; and 11 and 25).
- **Group 3:** The second popliteal segments; the anterior tibial; the tibiofibular trunk; and the posterior tibial (identified within Figure 2 as vessels 55a; 54; 55b; and 55c on the right side and 61a; 60; 61b; and 61c on the left side).

Three possible reduced networks are proposed, each with the removal of a single group of vessels (group 1, 2, or 3 in isolation). The pressure and flow-rate profiles produced by each reduced network are compared against the full ADAN network at all the measurement locations (see Section 3.2). The error metrics proposed in⁵¹ are used. If the maximum error is less than 2%, the full group of vessels are omitted from the arterial network. Otherwise, the vessel segment at the proximal end of the group is re-introduced into the reduced network. This process is iteratively repeated until the maximum number of vessels that can be removed from each group individually are found. Finally, these vessels from the three groups are removed simultaneously.

It is found that the first and second splenic segments; left gastric; common hepatic; and the left and right posterior interosseous vessels (identified within Figure 2 as vessels 37 and 39; 38; 36; and 10b and 24b) can all be removed from the arterial network without introducing errors larger than 2% relative to the full ADAN network. The pressure and flow-rate profiles produced by the full and reduced networks are shown in Figure 4 and the errors are reported in Table 1. The reduced network is shown in Figure 5 and has a dimensionality of 389, a reduction of 24 from the full network. This network and its associated parameters—either directly taken from or computed based off of the ADAN network—forms the reference network for this study.

4 | NETWORK PARAMETERISATION

A parsimonious representation is sought for the aforementioned reference (reduced) network so that its dimensionality can be further reduced. The reference network has 389 parameters. The vessels in this network can be split into the following three categories, which are also indicated in Figure 5:

- **Category 1:** 33 vessel segments with varying β and r_0 , and continuous variation with respect to the prior and subsequent vessels.
- **Category 2:** six vessel segments with varying β and r_0 , and discontinuous variation with respect to the prior and subsequent vessels.

Location	Measurement	Average error [%]	Systolic error [%]	Diastolic error [%]
Right common carotid	Pressure	0.4380	0.114	0.316
	Flow-rate	0.250	-0.085	-0.271
Left common carotid	Pressure	0.438	0.157	0.318
	Flow-rate	0.246	-0.111	-0.186
Right brachial	Pressure	0.643	1.108	0.378
	Flow-rate	0.875	-1.223	-1.833
Left brachial	Pressure	0.644	1.179	0.380
	Flow-rate	0.889	-1.131	-1.788
Right radial	Pressure	0.654	1.236	0.383
Left radial	Pressure	0.657	1.231	0.386
Right femoral I	Flow-rate	0.256	-0.174	-0.114
Right femoral II	Flow-rate	0.365	0.515	-0.463
Left femoral I	Flow-rate	0.254	-0.166	-0.115
Left femoral II	Flow-rate	0.363	0.514	-0.463

TABLE 1 Percentage discrepancy between the measurements taken from the full arterial network and the reduced network produced by the removal of the first and second splenic segments; left gastric; common hepatic; and posterior interosseous. Percentage errors have been computed using the error metrics presented in⁵¹.

- **Category 3:** 32 vessel where β and r_0 are constant along their lengths.

For Categories 1 and 2 with varying intra-vessel property profile, the reference network assumes a linear variation²⁹. Thus any property $g(x)$ (reference radius r_0 or β) varies along length x as

$$g(x) = g_0 - (g_0 - g_L) \frac{x}{L}, \quad (6)$$

where g_0 and g_L represent the properties at the proximal end $x = 0$ and the distal end $x = L$, respectively, and whose values are prescribed from the reference arterial network²⁹. A re-parameterisation of these properties to reduce the total dimensionality is presented next. For all the aforementioned three categories of vessels, a unified parameterisation is proposed, which also results in reduction in the dimensionality associated with the description of Category 1 vessels. This parameterisation includes description of the mechanical properties (Section 4.1), geometric properties (Section 4.2), and boundary conditions (Section 4.3).

4.1 | Mechanical properties

As opposed to describing the property variations linearly and individually for every vessel segment, see Equation (6), it proposed to use exponential variations, which allows for a simultaneous description of many successive vessels, thus reducing the total dimensionality. The re-parameterisation of mechanical property β for the three categories is as follows:

Category 1: These vessels have continuous variation in properties between successive vessels. Thus, successive vessels can be lumped together into chains and a single parameterisation be adopted. The 33 vessel segments (Category 1 vessels described in Section 4 and labelled in Figure 5 as cat. 1), that meet this description can be lumped into the following five chains:

- **The aortic chain.** This chain of vessel segments includes the first to fourth aortic arch segments; the first to sixth thoracic aorta segments; and the first to fifth abdominal aorta segments.
- **The right and left arm chains.** These chains of vessels include the first and second subclavian segments; the axillary artery; and the brachial artery.

- **The right and left leg chains.** These chains of vessels include the external iliac; the first and second femoral segments; and the first and second popliteal segments.

The β profiles for an entire vessel chain can now be described using a single function as opposed to separate functions (and hence separate parameters) for each individual vessel. An appropriate choice for such intra-vessel variation is

$$\beta(x) = \beta_0 \exp(-\Omega_d x), \quad (7)$$

where $\beta(x)$ and β_0 represent the material properties of the chain at the spatial coordinate x and the proximal end of the chain respectively; and $\Omega_d \geq 0$ represents a decay parameter.

Category 2: These vessels do not form part of any chain but show an intra-vessel property variation. For consistency and uniformity, their variation is also described by the exponential re-parameterisation of Equation (7). These vessels are labelled in Figure 5 as cat. 2.

Category 3: These vessels are also not part of a chain and have constant intra-vessel properties. Again, for uniformity, the re-parameterisation of Equation (7) is used, with the added explicit condition that $\Omega_d = 0$. These vessels are labelled in Figure 5 as cat. 3.

With the above unified description for the three vessel categories, β properties can be hierarchically assigned. It should be re-iterated that the choice of exponential variations is adopted for parameter parsimony. Both linear and non-linear variations in vessel stiffness and radii have been proposed in the past, and recently methods to select the best models based on evidence provided by clinical data also been proposed¹⁴.

4.1.1 | Hierarchical property assignment

Proximal-distal coherence dictates that the β properties at the distal end of a vessel segment must be greater than or equal to the corresponding property at the proximal end of any subsequent daughter segments¹¹. To ensure this a hierarchical procedure is adopted. The β properties of the arterial network are initialised by explicitly assigning a value at the inlet of the first aortic arch segment. The property profile of the aortic chain, and so consequently the vessel wall mechanical property at the proximal and distal ends of each segment within the chain, is computed using Equation (7). The properties at the proximal end of any vessel segments branching from the aorta are computed by applying a scaling term to the corresponding property at the distal end of the parent aortic segment:

$$\beta_0^d = \beta_{L_p}^p \Omega_s, \quad (8)$$

where β_0^d represents the property at the proximal end of the daughter segment, $\beta_{L_p}^p$ represents the property at the distal end of the parent segment, and Ω_s represents a daughter-parent scaling term, where:

$$0 < \Omega_s \leq 1. \quad (9)$$

The intra-vessel or intra-chain property profiles of all daughter branches bifurcating from the aorta are computed using Equation (7). This process is sequentially repeated through each generation of the arterial network until all terminal boundaries are reached.

4.2 | Geometric properties

Here, the re-parameterisation of geometric properties—reference radius and length—is presented.

4.2.1 | Arterial vessel reference radius

Since the variation of reference radius needs the same requirements as those for the variation of β , the variation of r_0 is described exactly in the same manner as β , as presented in Section 4.1. Lumping arterial vessels into chains, where appropriate, and applying exponential functions to describe the properties β and r_0 of all the vessels reduces the number of dimensions from 389 in the reference network to 269.

4.2.2 | Arterial vessel length

Each arterial vessel requires specification of its length. To individually assign a length to every vessel requires 71 parameters, accounting for a large proportion of the 269 remaining dimensions. It is thus proposed to reduce the dimensionality of the network by applying a singular scaling term to the lengths of all the vessels. It is empirically found that behaviours, patterns, and overall variability observed in the pressure and flow-rate profiles when allowing for maximum freedom to the length of vessels, *i.e.* assigning independent lengths to each arterial vessel, are not lost when applying a singular vessel length scaling term to all the vessels. This analysis justifies the use of a single scaling term and can be found in the Supplementary Material Appendix B. With this assumption, the dimensionality reduces from 269 to 199.

4.3 | Boundary conditions

The boundary conditions consist of the inlet flow rate and the terminal lumped parameters. These are parameterised as follows:

4.3.1 | Inlet flow-rate

The volumetric inlet flow-rate to the network, vessel ‘a’ in Figure 5, is described using a Fourier series (FS)⁸:

$$Q_{\text{inflow}}(t) = \sum_{n=0}^N a_n \sin(n\omega t) + b_n \cos(n\omega t), \quad (10)$$

where a_n and b_n represent the n^{th} sine and cosine FS coefficients, respectively; $N = 5$ represents the truncation order; and $\omega = 2\pi/T$, with T as the time period of the cardiac cycle. Thus, the following 11 parameters are required to specify the inlet flow-rate: $\{a_0 = 0, b_0, a_1, b_1, \dots, a_5, b_5\}$. An alternative method to describe the inlet flow-rate profiles, that could be adopted in future studies, is to include a heart model that outputs inflow conditions for the network; see for example^{19,52,53,54}. In such a case, the heart-model parameters will replace the FS coefficients in the network parameterisation.

4.3.2 | Terminal lumped models

There are 29 terminal boundaries in the network. Each of these terminal boundaries is coupled to a Windkessel model. Each Windkessel model requires three parameters: two resistances, and a compliance. These Windkessel models therefore require specification of a total of 87 parameters.

The above description (presented in Sections 4.1, 4.2, and 4.3) has 199 parameters. These represent 22 decay parameters Ω_d , 78 daughter-parent scaling parameters Ω_s , 87 Windkessel parameters at the outlets, 11 FS coefficients describing the inlet flow-rate, and 1 length scaling term. These 199 parameters represent a reduction of approximately 48% with respect to the original description with 389 parameters. These parameters are denoted individually by Ω_i , for the i^{th} parameter, and collectively with the vector $\mathbf{\Omega} = [\Omega_1, \Omega_2, \dots, \Omega_{199}]$. The final step in parameterisation is to describe all these 199 parameters relative to the reference network, and is described next.

4.4 | Scaling with respect to the reference network

Instead of describing the networks by directly stating values of the 199 parameters, they are specified relative to the reference network.

$$\mathbf{\Omega} = \mathbf{\theta}^T \mathbf{\Omega}_{\text{ref}}, \quad (11)$$

where $\mathbf{\theta} = [\theta_1, \theta_2, \dots, \theta_{199}]^T$ represents the scaling vector in relation to the reference network and $\mathbf{\Omega}_{\text{ref}}$ represents the vector of reference network parameters. Random sampling from the distribution of $\mathbf{\theta}$ will result in the VPD. This is described in the next section.

5 | STATISTICAL MODELLING

The pulse-wave propagation model, denoted as \mathcal{M} , takes the parameters $\mathbf{\theta}$ as inputs and outputs the pressure and flow-rates, collectively denoted as the vector \mathbf{y} , at all the locations:

$$\mathbf{y} = \mathcal{M}(\mathbf{\theta}), \quad (12)$$

The specific measurements are denoted by $\boldsymbol{\tau}$, which are essentially transformations of \mathbf{y} corrupted by measurement noise \mathcal{E} :

$$\boldsymbol{\tau} = \mathcal{H}(\mathbf{y}) + \mathcal{E}, \quad (13)$$

where \mathcal{H} represents the observation operator. Commonly, the observation operator \mathcal{H} represents an identity transform of selected components of \mathbf{y} . The measurement error \mathcal{E} is typically assumed to be a zero-mean multivariate normal:

$$\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{error}}), \quad (14)$$

where $\mathbf{0}$ is a zero-vector and $\boldsymbol{\Sigma}_{\text{error}}$ represents the error covariance matrix (diagonal when errors are independent).

To create the VPD, random samples of $\boldsymbol{\theta}$ from their joint probability distribution $p(\boldsymbol{\theta})$ are required. This distribution $p(\boldsymbol{\theta})$ should satisfy two broad pieces of information: first, it should satisfy known physiological and geometric constraints, and second, it should be consistent with what is known about the distributions of measurements $\boldsymbol{\tau}$ from literature. The latter ensures that when the physics-based model is run with the parameters distributed as $p(\boldsymbol{\theta})$, the resulting distributions of the measurements are close to those reported in real populations.

5.1 | Bayesian formulation

A Bayesian formulation is adopted to find the posterior distribution of the parameters $\boldsymbol{\theta}$. This is suitable because the geometric and physiological constraints, in the form a prior, can be combined with the literature reported measurements, in the form of a likelihood, to result in a posterior distribution that considers both the pieces of available information. The Bayes' theorem allows such a combination naturally:

$$\underbrace{p(\boldsymbol{\theta} | \boldsymbol{\tau})}_{\text{posterior}} = \frac{\overbrace{p(\boldsymbol{\tau} | \boldsymbol{\theta})}^{\text{likelihood}} \overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{p(\boldsymbol{\tau})}_{\text{evidence}}}. \quad (15)$$

The prior distributions $p(\boldsymbol{\theta})$ are described in Section 5.2 and are chosen to be either uninformative or weakly informative. This avoids imposition of strong prior beliefs which may influence the posterior. Their primary purpose is to impose appropriate bounds on the supports of the distributions based on geometric and physiological constraints. For the computation of the likelihood, the literature reported measurements are first represented as multivariate Normal distribution with mean $\boldsymbol{\mu}_{\text{lit}}$ and covariance $\boldsymbol{\Sigma}_{\text{lit}}$. Then, in the context of the statistical model, Equations (12)–(15), it is assumed that $\boldsymbol{\tau}$ is measured to be equal to $\boldsymbol{\tau} = \boldsymbol{\mu}_{\text{lit}}$ and the error covariance is $\boldsymbol{\Sigma}_{\text{error}} = \boldsymbol{\Sigma}_{\text{lit}}$. The likelihood for individual measurements is described in Section 5.3. Finally, in Equation (15), the evidence term is a normalising constant, which is not of interest and need not be evaluated.

5.2 | Prior distributions

Note that the network parameterisation is relative to the reference network, see Equation (11), and hence the distributions are for the scaling terms $\boldsymbol{\theta}$. Since scarce information is available about the variations of the parameters, the prior distributions are primarily constructed based on the bounds that should be observed in the parameters. In future studies, another approach which may be employed to assign tighter prior distributions could be the use of empirical relationships between the parameters and weight, age, height, etc., see for example⁷. Such empirical and morphological relationships may also be used to reduce the number of independent parameters. The following three types of uninformative or weakly informative priors⁵⁵ are employed:

1. Bounded parameters: A uniform prior distribution is chosen for all the scaling parameters which are supported on a bounded interval. The daughter-parent ratio, see Equations (8), is the only example of such a parameter. For the i^{th} pair of vessels with reference parent-daughter ratio $\Omega_{i,\text{ref}}$, it can be seen from Equations (9) and (11) that the corresponding scaling term θ_i is bounded between 0 and $1/\Omega_{i,\text{ref}}$. The prior probability density function (PDF) is thus:

$$p(\theta_i) = \begin{cases} \Omega_{i,\text{ref}} & \text{if } \theta_i \in (0, 1/\Omega_{i,\text{ref}}] \\ 0 & \text{otherwise} \end{cases}, \quad (16)$$

and shown in the left plot of Figure 6.

2. Semi-bounded parameters: A log-normal prior distribution is chosen for all the scaling parameters with semi-infinite support. These parameters with a lower bound of zero and no upper bound are:

- initial values of radii r_0 and vessel wall mechanical properties β at the inlet of the aorta (see Sections 4.1, 4.1.1, and 4.2);
- the decay terms used within all exponential property profiles (see Sections 4.1, and 4.2);
- the length of arterial vessels (see Section 4.2.2); and
- the terminal boundary Windkessel model parameters (see Section 4.3.2)

The prior PDF of the i^{th} such parameter is thus described by the following log-normal distribution:

$$p(\theta_i) = \frac{1}{\theta_i \sigma_i \sqrt{2\pi}} \exp\left(-\frac{(\ln(\theta_i) - \mu_i)^2}{2\sigma_i^2}\right), \quad (17)$$

where μ_i and σ_i represent the mean and standard deviation of the underlying Normal distribution, i.e. $\ln(\theta_i) \sim \mathcal{N}(\mu_i, \sigma_i^2)$. This log-normal distribution is created with large variance, resulting in a weakly informative prior. For all the scaling parameters μ_i is set to 0.5 and σ_i^2 is set to 0.8, respectively. This PDF is shown in the middle plot of Figure 6.

3. Unbounded parameters: A normal prior distribution is chosen for all the scaling parameters with infinite support. The only such parameters are the inlet flow-rate FS coefficients (see Section 4.3.1). The prior PDF of the i^{th} FS coefficient with mean μ_i and standard deviation σ_i is thus described as:

$$p(\theta_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\theta_i - \mu_i}{\sigma_i}\right)^2\right). \quad (18)$$

While the FS scaling parameters do have infinite support, they are expected to be within physiological range and hence within several multiplications of the reference FS coefficients. Thus, $\mu_i = 1$ and a standard deviation $\sigma_i = 2$ is set. This PDF is shown in the right plot of Figure 6. It is assumed that all the priors are independent of each other, thus the joint prior PDF is:

$$p(\boldsymbol{\theta}) = \prod_{i=1}^N p(\theta_i), \quad (19)$$

where $N = 199$ is the total number of scaling parameters.

5.3 | Likelihood

The likelihood corresponds to the literature reported measurements. These measurements can be categorised as follows:

5.3.1 | Scalar pressure and flow-rate measurements

Pressure (diastolic and systolic) and flow-rate (average) measurements at locations (radial artery, ascending aorta, common carotid, and femoral artery) reported in literature are incorporated into the likelihood. Their statistics and sources are shown in Table 2. When multiple measurements are available for the same quantity, they are pooled together into a single mean and variance⁵⁶. For such a measurement τ_i that was measured (or pooled together) to be $\mu_{i,\text{dis}}$ with standard deviation $\sigma_{i,\text{dis}}$, the likelihood is:

$$p(\tau_i = \mu_{i,\text{dis}} | \boldsymbol{\theta}) = \frac{1}{\sigma_{i,\text{dis}} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\mu_{i,\text{dis}} - \mathcal{H}_i(\mathcal{M}(\boldsymbol{\theta}))}{\sigma_{i,\text{dis}}}\right)^2\right), \quad (20)$$

where \mathcal{H}_i represents the observation operator (see Section 5) that extracts the τ_i component of the model output. Denoting all such discrete measurements collectively with the vector $\boldsymbol{\tau}_{\text{dis}}$, the combined likelihood with the assumption of independence is

$$p(\boldsymbol{\tau}_{\text{dis}} = \boldsymbol{\mu}_{\text{dis}} | \boldsymbol{\theta}) = \prod_i p(\tau_i = \mu_{i,\text{dis}} | \boldsymbol{\theta}). \quad (21)$$

Measurement location	Measurement type	No. of patients	Measured mean (μ_{dis})	Measured std. (σ_{dis})	Sources
Radial artery	Diastolic pressure (mmHg)	71	65	12.84	57, 58
	Systolic pressure (mmHg)		123	22.75	
Ascending aorta	Diastolic pressure (mmHg)	69	65	7.14	57, 59
	Systolic pressure (mmHg)		103	8.24	
Common carotid artery	Diastolic pressure (mmHg)	134	75.58	6.01	60, 61, 62
	Systolic pressure (mmHg)		122.78	12.01	
Femoral artery (left and right)	Average flow-rate (ml/sec)	63	5.84	2.11	63 64

TABLE 2 The measurement of the discrete pressure and flow-rate taken from literature. When multiple source measurements are available for the same quantity, they are pooled together into a single mean μ_{dis} and variance σ_{dis} ⁵⁶.

5.3.2 | Time varying inlet flow-rate and cardiac output measurement

Some time-varying information about the behaviour of pressure and flow-rates over the cardiac cycle is necessary to obtain a physiologically realistic posterior. Statistics on any such time-varying behaviour is not reported in literature to the authors' knowledge. Thus, a time-varying pseudo-measurement is constructed by combining the reference (ADAN network²⁹) time varying inlet flow-rate $Q_{\text{ref}}(t)$ and a measurement of average cardiac output^{65,66}. This pseudo-measurement $Q_{\text{meas}}(t)$ is described by a Gaussian Random Field (GRF, also referred to as a Gaussian Process)^{67,68} with the following mean $\mu_{\text{pseudo}}(t)$ and standard deviation $\sigma_{\text{meas}}(t)$

$$\mu_{\text{meas}}(t) = Q_{\text{ref}}(t), \quad (22)$$

$$\sigma_{\text{meas}}(t) = \left(1 + \frac{Q_{\text{ref}}(t) - \min(Q_{\text{ref}})}{\max(Q_{\text{ref}}) - \min(Q_{\text{ref}})} \right) \delta, \quad (23)$$

where δ is a positive scaling parameter. The scaling of $\sigma_{\text{meas}}(t)$ is to ensure that i) the standard deviation is positive, and ii) the variance is proportional to the magnitude of the flow rate (i.e. the percentage variance with respect to the magnitude remains fixed). The covariance between the pseudo-measurements at any two times t_i and t_j is described through a periodic kernel⁶⁷

$$\text{cov}(Q_{\text{meas}}(t_i), Q_{\text{meas}}(t_j)) = \sigma_{\text{meas}}(t_1) \sigma_{\text{meas}}(t_2) \exp\left(\frac{-2 \sin^2(\pi|t_i - t_j|/T)}{(vT)^2}\right), \quad (24)$$

where T represents the cardiac period, and v represents the ratio of the correlation length to the cardiac period. Through a trial-and-error analysis, this ratio is set to $v = 0.5$ to produce a physiologically realistic variation in the inlet flow-rate profile. The scaling parameter δ is tuned to the average cardiac output (CO) measurement: a mean of 98 ml/sec and standard deviation of 35.55 ml/sec^{65,66}. Through a sweep of δ , an optimal value of $\delta = 59.09$ is found, when the GRF produces a mean CO of 97.8 ml/sec and a standard deviation of 34.84 ml/sec. This GRF for $Q_{\text{meas}}(t)$ is shown in Figure 7 with the mean and \pm one standard deviation profiles.

The GRF is evaluated at $k = 12$ time points, shown in Figure 7, for the computation of the likelihood. Denoting these evaluation time points as $\{t_1, t_2, \dots, t_k\}$, the vector of measurements at these times with $\mathbf{r}_{\text{inflow}}$, the mean GRF vector as $\boldsymbol{\mu}_{\text{inflow}} = [\mu_{\text{meas}}(t_1), \mu_{\text{meas}}(t_2), \dots, \mu_{\text{meas}}(t_k)]$ computed through Equation (22), and the vector of inlet flow rates produced by the network parameters through Equation (10) as $\mathbf{Q}_{\text{inflow}} = [Q_{\text{inflow}}(t_1), Q_{\text{inflow}}(t_2), \dots, Q_{\text{inflow}}(t_k)]$, the likelihood for flow-rate can be written as:

$$p(\mathbf{r}_{\text{inflow}} = \boldsymbol{\mu}_{\text{inflow}} | \boldsymbol{\theta}) = (2\pi)^{-\frac{k}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}\right), \quad (25)$$

where

$$\boldsymbol{\xi} = \boldsymbol{\mu}_{\text{inflow}} - \mathbf{Q}_{\text{inflow}}, \quad (26)$$

and $\boldsymbol{\Sigma}$ is the GRF covariance matrix whose i^{th} row and j^{th} column element $\Sigma_{i,j} = \text{cov}(Q_{\text{meas}}(t_i), Q_{\text{meas}}(t_j))$ can be computed from Equation (24).

5.3.3 | Vessel length measurement

As mentioned in Section 4.2.2, the length of the vessels are parameterised by a single scaling term relative to the reference network. Since statistics of any direct vessel length measurement are not available, it is assumed that the lengths of arterial vessels are directly proportional to the height of a subject. A study of 25,945 twins from eight countries reported the mean and standard deviation of the height of the full cohort to be 172.0 cm and 9.308 cm respectively⁶⁹. Since the reference arterial network has a patient of height 170 cm, the measurement data corresponds to a mean of $\mu_{\text{len}} = 1.0118$ and the standard deviation of $\sigma_{\text{len}} = 0.0548$ for the vessel length scaling term. Denoting this measurement as τ_{len} , the likelihood for the scaling parameters θ_{len} is thus

$$p(\tau_{\text{len}} = \mu_{\text{len}} | \theta) = \frac{1}{\sigma_{\text{len}} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\mu_{\text{len}} - \theta_{\text{len}}}{\sigma_{\text{len}}}\right)^2\right), \quad (27)$$

Assuming all the measurements to be independent, the combined likelihood in Equation (15) is:

$$p(\boldsymbol{\tau} | \boldsymbol{\theta}) = p(\boldsymbol{\tau}_{\text{dis}} = \boldsymbol{\mu}_{\text{dis}} | \boldsymbol{\theta}) \times p(\boldsymbol{\tau}_{\text{inflow}} = \boldsymbol{\mu}_{\text{inflow}} | \boldsymbol{\theta}) \times p(\tau_{\text{len}} = \mu_{\text{len}} | \boldsymbol{\theta}), \quad (28)$$

where the RHS terms can be computed from Equations (21), (25), and (27).

In summary, the prior distributions for the 199 network parameters are specified in Section 5.2. These priors will be modified through the likelihood—specified by 7 scalar measurements of pressure and flow-rate, 1 time-varying flow-rate profile evaluated at 12 time points in the cardiac cycle, and a measurement of vessel lengths—to yield the posterior distribution of the parameters through Equation (15). Random sampling from this posterior will result in the virtual patient database. The sampling procedure is presented next.

6 | SAMPLING FROM THE POSTERIOR DISTRIBUTION

With the prior and the likelihood specified, the posterior distribution is given by Equation (15), and can be evaluated at any given $\boldsymbol{\theta}$ up to a normalising constant (the evidence term in the denominator of the equation). Sampling from this analytically intractable posterior is achieved through the Markov-chain Monte Carlo method^{70,71}, which is a widely used method to sample from unnormalised distributions.

6.1 | Markov-chain Monte Carlo

The particular MCMC method employed is the Metropolis-Hastings algorithm⁷². Starting from an initial sample, a chain of samples is built sequentially. Denoting the k^{th} sample as $\boldsymbol{\theta}^{(k)}$, a candidate for the next sample is generated by a proposal distribution which depends only on $\boldsymbol{\theta}^{(k)}$. This candidate sample $\boldsymbol{\theta}^*$ is proposed by sampling from the following multivariate Normal distribution:

$$\boldsymbol{\theta}^* \sim \mathcal{N}(\boldsymbol{\theta}^{(k)}, \boldsymbol{\Sigma}_{\text{step}}), \quad (29)$$

where $\boldsymbol{\Sigma}_{\text{step}}$ is the covariance matrix, whose i^{th} row and j^{th} column element $\Sigma_{\text{step},ij}$ is given by

$$\Sigma_{\text{step},ij} = \begin{cases} 0 & i = j \\ \sigma_{\text{step},i}^2 & i = j \end{cases}, \quad (30)$$

where $\sigma_{\text{step},i}^2$ represents the marginal variance for the i^{th} element of $\boldsymbol{\theta}^*$. This proposed candidate $\boldsymbol{\theta}^*$ is either accepted or rejected based on the following ratio α of posterior probabilities at $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^{(k)}$

$$\alpha = \frac{p(\boldsymbol{\theta} = \boldsymbol{\theta}^* | \boldsymbol{\tau})}{p(\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)} | \boldsymbol{\tau})} \quad (31)$$

which, through the Bayes' rule of Equation (15) applied to both the numerator and the denominator, can be written as

$$\alpha = \frac{p(\boldsymbol{\tau} | \boldsymbol{\theta} = \boldsymbol{\theta}^*) p(\boldsymbol{\theta} = \boldsymbol{\theta}^*)}{p(\boldsymbol{\tau} | \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}) p(\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)})}. \quad (32)$$

The first term and the second term in both the numerator and the denominator are the likelihoods and priors at the candidate and current point, respectively, and can be computed by Equations (28) and (19). To accept or reject the proposed candidate α is compared to a random number γ drawn from a uniform distribution between 0 and 1. If $\alpha \geq \gamma$, the proposed candidate, $\boldsymbol{\theta}^*$ is

accepted and the chain progress with $\theta^{(k+1)} = \theta^*$. Otherwise, the proposed candidate is rejected and a new candidate is proposed through Equation (29). This process is repeated until the chain is sufficiently long. After discarding the first few thousands of samples from this chain, referred as the burn-in period⁷³, the remaining samples are independent samples from the posterior, and form the virtual patient database.

The acceptance rate, *i.e.* the proportion of candidate sample points that are accepted, is tuned by varying $\sigma_{\text{step},i}^2$. A high variance results in the Markov chain exploring the posterior density quickly, with a high proportion of the candidate points being rejected. A low variance results in slow movement of the chain moving but with a high proportion of candidate points being accepted. An optimum acceptance rate to balance these two opposing behaviours is stated in literature to be 0.234⁷⁴. It is empirically found that using a standard deviation of 0.0375, equivalent to 3.75% of the reference values, for the scaling terms applied to the inlet flow-rate FS coefficients and that of 0.025, equivalent to 2.5% of the reference values, for all other scaling terms results in an acceptance rate of approximately 0.2.

6.2 | Pre-fetching

The MCMC algorithm outlined above is inherently sequential—each subsequent sample depends on the previous sample and hence the chain grows only one sample at a time. Generating a long chain this way thus leads to very high computational run times, since computation of the likelihood requires running a pulse-wave propagation simulation at each step. To alleviate this issue and achieve some level of parallelisation, the concept of pre-fetching^{75,76} is used. The central idea is to think of MCMC accept-reject decisions in the form of a decision tree. Figure 8 shows such a decision tree with a depth of $\eta = 2$. This results in 2^η leaf-nodes at the end of the tree. A careful consideration of the decision tree (see Figure 8) shows that the only unique parameters in the entire decision tree correspond to those at the leaf-nodes. Thus, computational simulations for all the leaf-node parameters can be run simultaneously in-parallel. Then, η steps of the MCMC algorithm can be taken together without any computational overhead by walking the decision tree. A tree depth of $\eta = 4$ is used in this study. For the creation of 75,000 VPs, the estimated computational time without employing pre-fetching is 109 days (2,616 hours) on a machine with 2×Intel Xeon E5-2620 v4 CPUs (each cpu has eight cores running at 2.10 GHz) and using only a single core (sequential MCMC). When employing pre-fetching with $\eta = 4$ on an identical machine, the run times are approximately 28 days (672 hours) with the utilisation of 16 cores in parallel.

7 | RESULTS AND DISCUSSION

The VPD is created by generating samples through the MCMC algorithm. The results for the MCMC method and the VPD are presented and discussed here.

7.1 | Burn-in analysis and chain diagnostics

It is common for the initial iterations of an MCMC chain to be considered as the “burn-in” period, during which the chain converges to an equilibrium distribution. This burn-in period is, therefore, discarded from the final sampled posterior distribution. To estimate this burn-in, trace plots for every parameter at all the iterations of the MCMC are plotted. An example of a trace plot is shown for the scaling terms applied to the length of VPs arterial vessels in Figure 9, and all other trace plots are shown in the Supplementary Material. To aid visualisation, these plots are thinned out by plotting every 100th iteration of the chain. These trace plots are visually assessed to determine burn-in. Essentially, if there is an initial period where a clear migration of the chain is seen away from the sample used to start the chain, then the burn-in is selected to be until this initial migration is complete. These trace plots also show good movement of the chain around the distribution, referred to as the “mixing” of the chain. Figures in the Supplementary Material Appendix C show that most parameters are initialised in a region of high density and hence no net migration is observed—the samples oscillate around a central mean value. Some interesting behaviours in the trace plots, including features that deviate from such a general/desired behaviour, are:

- The scaling term applied to β at the inlet of the aortic chain are seen to centre around 5 (Figure 20-a in the Supplementary Material). The maximum scaling term applied is approximately 7.5.

- The scaling terms applied to the reference decay term of the left leg chain β profile initially oscillates around a value of approximately 1 for 25,000 iterations, before migrating and oscillating around a value of approximately 7 (Figure 20-j in the Supplementary Material). This may suggest that this distribution is multimodal.
- The scaling terms applied to the reference decay term of the brachiocephalic trunk β profile oscillates around a value of 1 for the first 40,000 iterations, before migrating and oscillating around a value of 9 (Figure 21b in the Supplementary Material). The scaling terms applied to the reference decay term of the brachiocephalic trunk r_0 profile shows no complimentary behaviour and remains within the region of 0–4 throughout. This may again suggest a multimodal distribution.
- A spike is seen in the scaling terms applied to the compliance of the right external carotid Windkessel model at approximately 55,000 iterations (Figure 35-e in the Supplementary Material). This behaviour is not seen within the scaling terms applied to the compliance of the left external carotid Windkessel model which remains within the region of 0–3 throughout.

Even though most of the parameters do not show a clear migration, an initial burn-in period of 10,000 iterations is chosen as a precautionary measure to minimise any effect the initial sample position may have. Once this burn-in period has been removed, the VPD contains 65,000 VPs. Along with the burn-in period, all VPs with negative average flow-rate within any arterial vessel are removed from the VPD. These VPs are removed from the VPD as it is physiologically unlikely for a patient to have negative average flow-rate in any vessel. Of the 65,000 post burn-in VPs, 12,857 are removed due to the presence of negative average flow-rate, reducing the VPD to 52,143. While low flow-rate is also unlikely in VPs, it is difficult to establish such a threshold for exclusion based on a minimum value of flow-rate. If such information is available, it can be included through the likelihood term or a post simulation filter (for example, see filter in Section 7.4).

For the assessment of convergence diagnostics, the measure of integrated auto-correlation time (IACT)^{22,77,78} for the parameters and the measurements is employed. It is found that most of the parameters and the measurements shown an IACT of less than 50, comparable to those reported in previous studies utilising 1D flow models²². For the parameter representing the scaling term for the vessel lengths, and for the measurement of diastolic pressure in the ascending aorta, plots of autocorrelation versus lag length^{22,77,78} are shown in Figure 10. For some parameters, even higher IACT (>200) was observed. While this does not invalidate the VPD, it does imply that the chain efficiency can be improved by either (i) reducing the dimensionality of the problem further; (ii) further tuning individual step-sizes in equation 30; or (iii) by utilising more efficient MCMC samplers; see, for example, the work by Paun and Husmeier²⁵. One reason for higher IACTs for some parameters may be the correlation between the parameters in the posterior. Across the 19,701 pair combinations, a histogram of the correlation coefficients is shown in Figure 11, showing that most pairs show low to moderate correlations $|r| < 5$, with few combinations showing more extreme correlations. Since the dimensionality is high, even moderate correlations can make the navigation of chain harder because of the curse of dimensionality⁷⁹.

7.2 | Posterior vs. literature reported measurements

The joint posterior distribution of the 199 parameters is difficult to visualise. Even if only pairwise slices of the joint distribution are considered, the number of pairs (19,701) is too high to be shown in this manuscript. To be consistent with the trace-plot presented in the previous section for the vessel length scaling parameter (θ_a), its posterior distribution is shown in Figure 12 (left). The remaining two plots (centre and right) show pairwise slices of the joint distribution of this parameter with two other parameters—the right arm chain’s scaling factor for the β daughter/parent ratio (θ_b) and the brachiocephalic trunk’s scaling factor for β decay (θ_c). Since the goal of this work is not to make inference about the parameters but to produce a VPD consisting of realistic measurements, the posterior distribution of the measurements within the VPD is important to be considered. As described in Section 5.1, the posterior distribution is a combination of the prior distribution and the likelihood term within Bayes’ theorem. Since the chosen prior distributions are weakly informative or uninformative (see Section 5.2), the distributions of measurements within the VPD should be close to the literature based measurements incorporated through the likelihood. These are not necessarily identical though, as the prior distribution corrects the posterior to account for the geometrical and physiological constraints. Furthermore, since the incorporated measurements are taken from several different sources, the posterior distribution resolves inconsistencies between such measurements, the physics of pulse-wave propagation, and the constraints imposed by the prior. For example, if the prior distribution corresponds to lower valued outputs compared to the literature measurements, the posterior distribution will underestimate the literature measurements, and vice versa. A comparison of the VPD pressure and flow-rate distributions to the literature based measurements (see Table 2 and Section 5.3.1) are shown in Figure

Measurement location	Measurement type	Source	Side	Measured mean (μ_{dis})	Measured std. (σ_{dis})
Radial artery	Diastolic pressure (mmHg)	Literature	Both	65	12.84
		Posterior	Right	67.91	6.76
			Left	67.69	6.29
	Systolic pressure (mmHg)	Literature	Both	123	22.75
		Posterior	Right	108.07	15.48
			Left	114.60	13.66
Ascending aorta	Diastolic pressure (mmHg)	Literature	-	65	7.14
		Posterior	-	75.52	3.69
	Systolic pressure (mmHg)	Literature	-	103	8.24
		Posterior	-	110.70	6.03
Common carotid artery	Diastolic pressure (mmHg)	Literature	Both	75.58	6.01
		Posterior	Right	74.43	3.72
			Left	74.80	3.75
	Systolic pressure (mmHg)	Literature	Both	122.78	12.01
		Posterior	Right	111.91	6.25
			Left	111.45	6.19
Femoral artery (left and right)	Average flow-rate (ml/sec)	Literature	Both	5.84	2.11
		Posterior	Right	1.30	0.77
			Left	1.08	0.81

TABLE 3 The measurement of the discrete pressure and flow-rate taken from literature and the corresponding posterior distributions. See Table 2 for details and sources of the literature measurements.

13 and Table 3. A similar comparison for the vessel length scaling term (see Section 5.3.3) is shown in Figure 12 (left). Finally, the statistics of the time varying inlet flow-rate profiles in the VPD are compared to the GRF (see Section 5.3.2) in Figure 15.

Generally, a good agreement between the scalar pressure and flow-rate measurements are seen in Figure 13 and Table 3 given that the posterior balances measurement inconsistencies, incomplete representation of reality by the model, and inconsistencies introduced by the geometrical/physiological constraints. This agreement enforces confidence in the overall approach. However, for the average flow-rates in the left and right femoral arteries, larger than expected differences are observed. This is likely due to the following reasons:

- A large inconsistency between the femoral flow-rates measurement and the other pressure measurements. Note that they are taken from different sources and hence not from the same population. Furthermore, there may be an inconsistency between the cardiac output measurement used for generating the pseudo-measurement for the time varying inlet flow-rate measurement (see section 5.3.2).
- The presence of more measurements for pressure as opposed to flow-rates. Since the likelihoods for all the measurements are weighted equally, it may be possible that the chain is influenced weakly by the few flow-rate measurements. Such an issue can be resolved in future studies by assigning variable weightings to the measurements.
- In the network the femoral arteries are split into the first (I) and second segments (II), shown in Figure 5 by ‘u’ and ‘v’, respectively. The precise location at which the literature based femoral flow-rate measurement has been taken is unknown. In this study, it is assumed that this measurement was acquired at the centre of the second segment ‘v’. Since the first segment ‘u’ bifurcates into ‘v’ and another segment *profunda femoris* (Figure 5), the flow-rate in ‘v’ is smaller than that

in ‘u’. Thus, it is possible that the measured flow-rate was in ‘u’. A comparison of VPD flow-rate in this vessel ‘u’ against the measurement is presented in Figure 14, and shows a significantly better agreement.

Figure 12 (left) shows a good agreement between the vessel length scaling term in the VPD and the literature based measurement. The mean and standard deviation of this term in the VPD are 0.0551 and 0.9819, respectively, which compare well to those reported in literature with values of 0.0548 and 1.0118, respectively.

Figure 15 shows a good agreement between the statistics of the time varying inlet flow-rate profile in the VPD and the pseudo-measurement constructed through the GRF (Section 5.3.2). The mean profile closely follows that of the GRF throughout the cardiac cycle, with maximum difference of approximately at 10% peak systole. Figure 15 also shows \pm one standard profiles for both the VPD flow-rate and the GRF pseudo-measurement. Throughout the cardiac cycle, the agreement and the trend is well captured in the VPD, with largest errors of approximately 10% magnitudes at peak systole.

7.3 | Evaluation of individual VPs

Random samples from the VPD are assessed to gain further insights on the VPs and the behaviour of pressure and flow-rate profiles. Pressure profiles are examined at the ascending aorta; right and left radial arteries; and right and left common carotid arteries. Flow-rate profiles are examined at the right and left second femoral segments. 15 VPs are randomly drawn from the VPD (excluding the burn-in), and the pressure and flow-rate profiles associated with each are shown at all examination locations in the Supplementary Material Appendix D. Five VPs of interest that show peculiar behaviour are extracted from the Supplementary Material and shown in Figure 16.

Figures in the Supplementary Material Appendix D show that pressure profiles in proximal vessels, *i.e.* the ascending aorta and two common iliacs, show more consistency and similarity to the reference profiles compared to the pressure and flow-rate profiles in distal vessels, *i.e.* the radial artery pressure and the femoral artery flow-rate. This suggests that the pseudo-measurement of the time varying inlet flow-rate profile is sufficient to impose sufficient control over the shape of pressure and flow-rate profiles in the proximal vessels. As the spatial distance from the inlet of the aorta increases, an increase in the variability of pressure and flow-rate profiles is observed. This is expected as the measurements that guide the posterior are only available at few locations. Thus, access to more measurements, distributed across the network in both space and time, will result in an even more realistic database as the likelihood at several such locations will guide the posterior.

Figure 16 shows a selection of undesirable behaviour in the VPD profiles. In the case of Patient-E, oscillatory behaviour is seen within all the pressure and flow-rate profiles. This is caused by the shape of the inlet flow-rate profile prescribed for this patient. While this behaviour is not very common in the database, if needed its occurrence can be further reduced by imposing stronger correlations between inlet flow-rates at two time points through the parameter ν in Equation (24) in the GRF.

Highly oscillatory pressure profiles are observed in the left radial artery, shown in Figure 16(c), of Patient-A and Patient-E. In the case of VP Patient-A there is no significant oscillatory behaviour in any other profile, but the pressure waveform in the right radial artery is featureless with no clear systolic and diastolic points. This apparent over- and under-damping of the pressure profiles in the radial arteries may suggest an imbalance in the compliances or resistances of the right and left arms. This hypothesis of left-right imbalance is further supported by Figures 16(f) and (g) when the femoral flow-rates are observed for Patient-I: the flow-rate in the right femoral artery shows a high-mean high-oscillation behaviour while in the left femoral artery it is low-mean with significantly lower oscillations. Similar behaviour is seen within Patient-J and Patient-O. In future studies, a symmetry metric that balances, while still allowing for some variability, the left and right side parameters for symmetric left and right side vessels may be implemented. In this study, it is chosen to filter out vessels that show a high left-right imbalance with an analysis that is described next.

7.4 | Filter for left-right imbalance

The left-right lower extremity imbalance is assessed by reducing the network (see Section 3.3 and Figure 3) up to the second bifurcation in the leg vessels, *i.e.* up to the end of vessels labelled ‘u’ in Figure 5. To assess the left-right imbalance, the ratio of femoral pulse (maximum flow-rate minus the minimum flow-rate) on the left and right sides is considered. It is found that this ratio shows highest correlation against the ratio of reduced network compliances on the left and right sides. A plot of the femoral pulse ratio against the compliance ratio is shown in Figure 17. The left-right imbalance is apparent in this figure which shows that the ratio of left-to-right femoral pulses varies from 10^{-4} and 10^4 across the VPD. To limit this, a filter on the left-to-right compliance ratio between 0.2–5 is introduced. These limits are shown as vertical lines in Figure 17 and constrains

the the femoral ratio between 10^{-2} and 10^2 across the dataset. With this filter, approximately 45% (23,275 patients) of the patients are discarded, leaving with 28,868 physiologically realistic patients in the VPD.

This VPD reflects both prior beliefs/constraints and measurements from the population. Thus, it is suitable to study population characteristics and behaviour. More importantly, disease can be virtually created in these patients and subsequent assessment of machine learning be made on the database; for example, see our recent work⁸⁰ which utilises such an approach. The pre-filter VPD is publicly made available at Zenodo repository^{†81}.

8 | CONCLUSIONS

A physiologically realistic virtual patient database is presented for the human arterial network. A methodology to create virtual patients guided by prior beliefs, geometrical/physiological constraints, and literature reported measurements is presented. Starting from a reference network describing the arterial network, the methodology includes i) network reduction without compromising relevant behaviour; ii) re-parameterisation to reduce dimensionality; iii) incorporation of geometrical and physiological constraints in the form of a prior; iv) incorporation of literature reported clinical measurements in the form of the likelihood; v) combination of the prior and likelihood to generate the posterior; and vi) sampling from the posterior with MCMC to create the VPD. This generic methodology, given a mathematical description of a biological system, can be adopted to create virtual patients for any biological system while accounting for all available information.

The key conclusion of this study is that despite computational challenges and dimensional complexity, the Bayesian approach to creating virtual patients is viable and yields a realistic database of patients consistent with the expected behaviour in terms of the range of outputs expected in a population. The study also shows that appropriate incorporation of priors and measurements (distributed both spatially and in time) is necessary for the generation of physiologically realistic virtual patients. In the absence of these, the virtual patients may be physiologically unrealistic while still statistically reproducing the literature measurements. In this study this phenomenon was observed as no likelihood or prior term incorporating left-right symmetry was included. While such issues can be fixed by incorporation of a filter, the end result is wasted effort in the generation of virtual patients which are subsequently discarded. Nevertheless, the Bayesian approach and the framework presented is an attractive framework that is capable of incorporating such expectations in the methodology so that such issues are avoided.

A realistic database of 28,868 patients representing human arterial network is presented and made available. This database is created with a view that it is useful in subsequent studies which exploit the advances in machine learning methods for the detection of arterial disease. This forms the primary area of future investigation with this database.

FUNDING

This work is supported by an EPSRC studentship ref. EP/N509553/1 and an EPSRC grant ref. EP/R010811/1.

References

1. Fowkes FGR, Rudan D, Rudan I, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *The Lancet* 2013; 382(9901): 1329–1340.
2. Shadman R, Criqui MH, Bundens WP, et al. Subclavian artery stenosis: prevalence, risk factors, and association with cardiovascular diseases. *Journal of the American College of Cardiology* 2004; 44(3): 618–623.
3. Mathiesen EB, Joakimsen O, Børnaa KH. Prevalence of and risk factors associated with carotid artery stenosis: the Tromsø Study. *Cerebrovascular diseases* 2001; 12(1): 44–51.
4. Li X, Zhao G, Zhang J, Duan Z, Xin S. Prevalence and trends of the abdominal aortic aneurysms epidemic in general population—a meta-analysis. *PLoS One* 2013; 8(12): e81260.

[†]Link to Zenodo repository: <https://doi.org/10.5281/zenodo.4549764>

5. Darwood R, Earnshaw JJ, Turton G, et al. Twenty-year review of abdominal aortic aneurysm screening in men in the county of Gloucestershire, United Kingdom. *Journal of vascular surgery* 2012; 56(1): 8–13.
6. Willemet M, Chowienczyk P, Alastruey J. A database of virtual healthy subjects to assess the accuracy of foot-to-foot pulse wave velocities for estimation of aortic stiffness. *American Journal of Physiology-Heart and Circulatory Physiology* 2015; 309(4): H663–H675.
7. Chakshu NK, Sazonov I, Nithiarasu P. Towards enabling a cardiovascular digital twin for human systemic circulation using inverse analysis. *Biomechanics and Modeling in Mechanobiology* 2020: 1–17.
8. Jones G, Parr J, Nithiarasu P, Pant S. A proof of concept study for machine learning application to stenosis detection. arXiv eprint 2102.07614 [cs.LG]; 2021.
9. Celik N, O'Brien F, Brennan S, et al. Deep-Channel uses deep neural networks to detect single-molecule events from patch-clamp data. *Communications Biology* 2020; 3(1): 1–10.
10. Blanco PJ, Watanabe SM, Dari EA, Passos MAR, Feijóo RA. Blood flow distribution in an anatomically detailed arterial network model: criteria and algorithms. *Biomechanics and modeling in mechanobiology* 2014; 13(6): 1303–1330.
11. Blanco PJ, Watanabe SM, Passos MAR, Lemos PA, Feijóo RA. An anatomically detailed arterial network model for one-dimensional computational hemodynamics. *IEEE Transactions on biomedical engineering* 2015; 62(2): 736–753.
12. Pappalardo F, Russo G, Tshinanu FM, Viceconti M. In silico clinical trials: concepts and early adoptions. *Briefings in Bioinformatics* 2019; 20(5): 1699–1708.
13. Becker AK, Dörr M, Felix SB, et al. From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach. *PLoS computational biology* 2021; 17(2): e1008735.
14. Paun LM, Colebank MJ, Olufsen MS, Hill NA, Husmeier D. Assessing model mismatch and model selection in a Bayesian uncertainty quantification analysis of a fluid-dynamics model of pulmonary blood circulation. *Journal of the Royal Society Interface* 2020; 17(173): 20200886.
15. Yin M, Yazdani A, Karniadakis GE. One-dimensional modeling of fractional flow reserve in coronary artery disease: Uncertainty quantification and Bayesian optimization. *Computer Methods in Applied Mechanics and Engineering* 2019; 353: 66–85.
16. Pant S, Fabrèges B, Gerbeau JF, Vignon-Clementel IE. A multiscale filtering-based parameter estimation method for patient-specific coarctation simulations in rest and exercise. In: Springer. ; 2013: 102–109.
17. Pant S, Fabrèges B, Gerbeau JF, Vignon-Clementel I. A methodological paradigm for patient-specific multi-scale CFD simulations: from clinical measurements to parameter estimates for individual analysis. *International journal for numerical methods in biomedical engineering* 2014; 30(12): 1614–1648.
18. Pant S, Lombardi D. An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Mathematical biosciences* 2015; 268: 66–79.
19. Pant S, Corsini C, Baker C, et al. Data assimilation and modelling of patient-specific single-ventricle physiology with and without valve regurgitation. *Journal of biomechanics* 2016; 49(11): 2162–2173.
20. Tran JS, Schiavazzi DE, Ramachandra AB, Kahn AM, Marsden AL. Automated tuning for parameter identification and uncertainty quantification in multi-scale coronary simulations. *Computers & fluids* 2017; 142: 128–138.
21. Schiavazzi DE, Baretta A, Pennati G, Hsia TY, Marsden AL. Patient-specific parameter estimation in single-ventricle lumped circulation models under uncertainty. *International journal for numerical methods in biomedical engineering* 2017; 33(3): e02799.
22. Paun LM, Qureshi MU, Colebank M, et al. MCMC methods for inference in a mathematical model of pulmonary circulation. *Statistica Neerlandica* 2018; 72(3): 306–338.

23. Paun LM, Colebank M, Qureshi MU, Olufsen M, Hill N, Husmeier D. MCMC with delayed acceptance using a surrogate model with an application to cardiovascular fluid dynamics. In: International ASET Inc. ; 2019: 28.1–28.8
24. Pant S. Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems. *Journal of The Royal Society Interface* 2018; 15(142): 20170871.
25. Paun LM, Husmeier D. Markov chain Monte Carlo with Gaussian processes for fast parameter estimation and uncertainty quantification in a 1D fluid-dynamics model of the pulmonary circulation. *International journal for numerical methods in biomedical engineering* 2020: e3421.
26. Perdikaris P, Karniadakis GE. Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *Journal of The Royal Society Interface* 2016; 13(118): 20151107.
27. Aggarwal A, Lombardi D, Pant S. An information-theoretic framework for optimal design: analysis of protocols for estimating soft tissue parameters in biaxial experiments. *Axioms* 2021.
28. Alastruey J, Parker KH, Sherwin SJ, others . Arterial pulse wave haemodynamics. In: Virtual PiE Led t/a BHR Group: Lisbon, Portugal. ; 2012: 401–442.
29. Boileau E, Nithiarasu P, Blanco PJ, et al. A benchmark study of numerical schemes for one-dimensional arterial blood flow modelling. *International journal for numerical methods in biomedical engineering* 2015; 31(10).
30. Formaggia L, Lamponi D, Quarteroni A. One-dimensional models for blood flow in arteries. *Journal of engineering mathematics* 2003; 47(3-4): 251–276.
31. Olufsen MS, Peskin CS, Kim WY, Pedersen EM, Nadim A, Larsen J. Numerical simulation and experimental validation of blood flow in arteries with structured-tree outflow conditions. *Annals of biomedical engineering* 2000; 28(11): 1281–1299.
32. Reymond P, Merenda F, Perren F, Rufenacht 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.
33. Matthys KS, Alastruey J, Peiró J, et al. Pulse wave propagation in a model human arterial network: assessment of 1-D numerical simulations against in vitro measurements. *Journal of biomechanics* 2007; 40(15): 3476–3486.
34. Westerhof N, Lankhaar JW, Westerhof BE. The arterial windkessel. *Medical & biological engineering & computing* 2009; 47(2): 131–141.
35. Carson JM. *Development of a cardiovascular and lymphatic network model during human pregnancy*. PhD thesis. Swansea University, Singleton Park, Sketty, Swansea SA2 8PP; 2019.
36. Carson J, Van Loon R. An implicit solver for 1D arterial network models. *International journal for numerical methods in biomedical engineering* 2017; 33(7): e2837.
37. Boileau E, Pant S, Roobottom C, et al. Estimating the accuracy of a reduced-order model for the calculation of fractional flow reserve (FFR). *International journal for numerical methods in biomedical engineering* 2018; 34(1): e2908.
38. Kuo FY, Sloan IH. Lifting the curse of dimensionality. *Notices of the AMS* 2005; 52(11): 1320–1328.
39. Adji A, Hirata K, O’rourke MF. Clinical use of indices determined non-invasively from the radial and carotid pressure waveforms. *Blood pressure monitoring* 2006; 11(4): 215–221.
40. O’rourke MF. Carotid Artery Tonometry: Pros and Cons. *American journal of hypertension* 2015; 29(3): 296–298.
41. Langwieser N, Prechtl L, Meidert AS, et al. Radial artery applanation tonometry for continuous noninvasive arterial blood pressure monitoring in the cardiac intensive care unit. *Clinical Research in Cardiology* 2015; 104(6): 518–524.
42. Guelen I, Westerhof BE, Sar v. dGL, et al. Validation of brachial artery pressure reconstruction from finger arterial pressure. *Journal of hypertension* 2008; 26(7): 1321–1327.

43. Guelen I, Westerhof BE, Sar v. dGL, et al. Finometer, finger pressure measurements with the possibility to reconstruct brachial pressure. *Blood pressure monitoring* 2003; 8(1): 27–30.
44. Byström S, Jensen B, Jensen-Urstad M, Lindblad L, Kilbom A. Ultrasound-Doppler technique for monitoring blood flow in the brachial artery compared with occlusion plethysmography of the forearm. *Scandinavian journal of clinical and laboratory investigation* 1998; 58(7): 569–576.
45. Oglat AA, Matjafri M, Suardi N, Oqlat MA, Abdelrahman MA, Oqlat AA. A review of medical doppler ultrasonography of blood flow in general and especially in common carotid artery. *Journal of medical ultrasound* 2018; 26(1): 3.
46. Radegran G. Ultrasound Doppler estimates of femoral artery blood flow during dynamic knee extensor exercise in humans. *Journal of Applied Physiology* 1997; 83(4): 1383–1388.
47. Dyken ML, Klatte E, Kolar OJ, Spurgeon C. Complete occlusion of common or internal carotid arteries: Clinical significance. *Archives of Neurology* 1974; 30(5): 343–346.
48. Kullo IJ, Rooke TW. Peripheral artery disease. *New England Journal of Medicine* 2016; 374(9): 861–871.
49. Aboyans V, Desormais I, Lacroix P, Salazar J, Criqui MH, Laskar M. The general prognosis of patients with peripheral arterial disease differs according to the disease localization. *Journal of the American College of Cardiology* 2010; 55(9): 898–903.
50. Chen Q, Smith CY, Bailey KR, Wennberg PW, Kullo IJ. Disease location is associated with survival in patients with peripheral arterial disease. *Journal of the American Heart Association* 2013; 2(5): e000304.
51. Epstein S, Willemet M, Chowienczyk PJ, Alastruey J. Reducing the number of parameters in 1D arterial blood flow modeling: less is more for patient-specific simulations. *American Journal of Physiology-Heart and Circulatory Physiology* 2015; 309(1): H222–H234.
52. Vosse V. dFN, Stergiopoulos N. Pulse wave propagation in the arterial tree. *Annual Review of Fluid Mechanics* 2011; 43: 467–499.
53. Pant S, Corsini C, Baker C, Hsia TY, Pennati G, Vignon-Clementel IE. Inverse problems in reduced order models of cardiovascular haemodynamics: aspects of data assimilation and heart rate variability. *Journal of The Royal Society Interface* 2017; 14(126): 20160513.
54. Arts T, Delhaas T, Bovendeerd P, Verbeek X, Prinzen FW. Adaptation to mechanical load determines shape and properties of heart and circulation: the CircAdapt model. *American Journal of Physiology-Heart and Circulatory Physiology* 2005; 288(4): H1943–H1954.
55. Bernardo JM, Smith AF. *Bayesian theory*. 405. John Wiley & Sons . 2009.
56. Rudmin JW. Calculating the exact pooled variance. *arXiv preprint arXiv:1007.1012* 2010.
57. Pauca AL, Wallenhaupt SL, Kon ND, Tucker WY. Does radial artery pressure accurately reflect aortic pressure?. *Chest* 1992; 102(4): 1193–1198.
58. Chen CH, Nevo E, Fetis B, et al. Estimation of central aortic pressure waveform by mathematical transformation of radial tonometry pressure: validation of generalized transfer function. *Circulation* 1997; 95(7): 1827–1836.
59. Murgu JP, Westerhof N, Giolma JP, Altobelli SA. Aortic input impedance in normal man: relationship to pressure wave forms.. *Circulation* 1980; 62(1): 105–116.
60. Sugawara M, Niki K, Furuhashi H, Ohnishi S, Suzuki S. Relationship between the pressure and diameter of the carotid artery in humans. *Heart and vessels* 2000; 15(1): 49–51.
61. Arndt JO, Klauske J, Mersch F. The diameter of the intact carotid artery in man and its change with pulse pressure. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* 1968; 301(3): 230–240.

62. Hansen F, Mangell P, Sonesson B, Länne T. Diameter and compliance in the human common carotid artery—variations with age and sex. *Ultrasound in medicine & biology* 1995; 21(1): 1–9.
63. Maroto A, Ginés P, Arroyo V, et al. Brachial and femoral artery blood flow in cirrhosis: relationship to kidney dysfunction. *Hepatology* 1993; 17(5): 788–793.
64. Lewis P, Psaila J, Davies W, McCarty K, Woodcock J. Measurement of volume flow in the human common femoral artery using a duplex ultrasound system. *Ultrasound in medicine & biology* 1986; 12(10): 777–784.
65. Ihlen H, Amlie J, Dale J, et al. Determination of cardiac output by Doppler echocardiography.. *Heart* 1984; 51(1): 54–60.
66. Chandraratna PA, Nanna M, McKay C, et al. Determination of cardiac output by transcutaneous continuous-wave ultrasonic Doppler computer. *The American journal of cardiology* 1984; 53(1): 234–237.
67. Rasmussen CE. Gaussian processes in machine learning. In: Springer. 2004 (pp. 63–71).
68. Ibragimov IA, Rozanov YA. *Gaussian random processes*. 9. Springer Science & Business Media . 2012.
69. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research and Human Genetics* 2003; 6(5): 399–408.
70. Geyer CJ. Practical markov chain monte carlo. *Statistical science* 1992: 473–483.
71. Gilks WR. M arkov Chain M onte C arlo. *Encyclopedia of biostatistics* 2005; 4.
72. Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. *The american statistician* 1995; 49(4): 327–335.
73. Raftery AE, Lewis SM. Implementing mcmc. *Markov chain Monte Carlo in practice* 1996: 115–130.
74. Roberts GO, Gelman A, Gilks WR, others . Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability* 1997; 7(1): 110–120.
75. Brockwell AE. Parallel Markov chain Monte Carlo simulation by pre-fetching. *Journal of Computational and Graphical Statistics* 2006; 15(1): 246–261.
76. Strid I. Efficient parallelisation of Metropolis–Hastings algorithms using a prefetching approach. *Computational Statistics & Data Analysis* 2010; 54(11): 2814–2835.
77. Christen JA, Fox C, others . A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis* 2010; 5(2): 263–281.
78. Roy V. Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application* 2020; 7: 387–412.
79. Bellman R. Dynamic programming. *Science* 1966; 153(3731): 34–37.
80. Jones G, Parr J, Nithiarasu P, Pant S. Machine learning for detection of stenoses and aneurysms: application in a physiologically realistic virtual patient database. arXiv eprint 2103.00599 [cs.LG]; 2021.
81. Jones G, Parr J, Nithiarasu P, Pant S. A physiologically realistic virtual patient database for the study of arterial haemodynamics [Data set] . *Zenodo* 2021. doi: 10.5281/zenodo.4549764



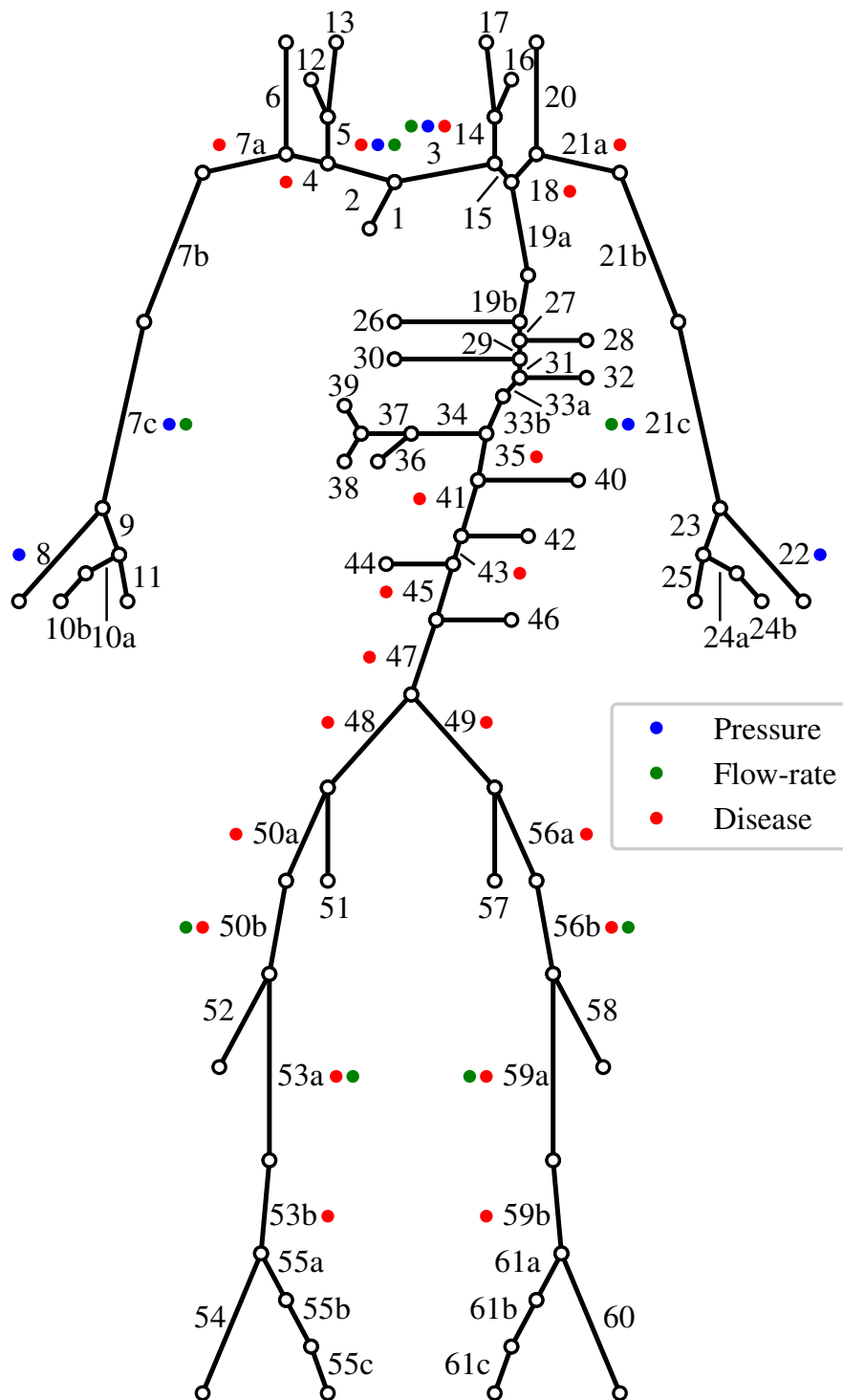


FIGURE 2 The connectivity of the reference arterial network, taken from²⁹. At the inlet (free end) of vessel 1, a volumetric flow-rate is specified and at all the outlets (free ends of the terminal vessels), a Windkessel model is specified. Locations at which pressure and flow-rates can be measured; and disease is likely to occur are also highlighted, see Section 3.2.

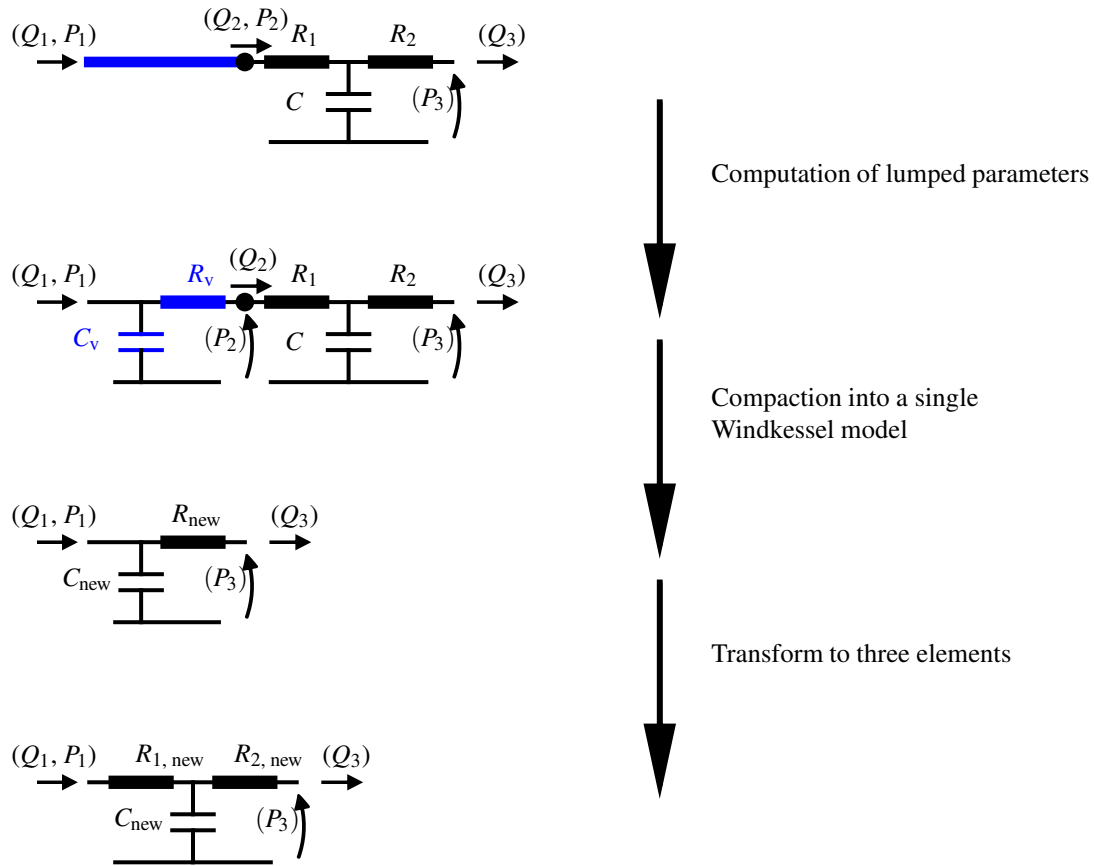


FIGURE 3 The process of incorporating peripheral 1D vessels into the 0D terminal boundary Windkessel model parameters⁵¹. Q_1 and P_1 represent the flow-rate and pressure at the proximal end of the vessel that is being removed; Q_2 and P_2 represent the flow-rate and pressure at the distal end of the vessel that is being removed; Q_3 and P_3 represent the the flow-rate and pressure at the outlet of the system; R_1 , R_2 and C represent the resistances and compliance of the original terminal Windkessel model; R_v , and C_v represent the viscous resistance and compliance of the vessel being removed; R_{new} and C_{new} represent the resistance and compliance of the new 2 element Windkessel model after the incorporation of the 1D vessel; and $R_{1,\text{new}}$, $R_{2,\text{new}}$, and C_{new} represent the resistances and compliance of the final 3 element Windkessel model. Details on computation of R_v and C_v and then R_{new} and C_{new} can be found in⁵¹. $R_{1,\text{new}}$ is assumed to be equal to the characteristic impedance of the connecting vessel (see⁵¹), and $R_{2,\text{new}} = R_{\text{new}} - R_{1,\text{new}}$.

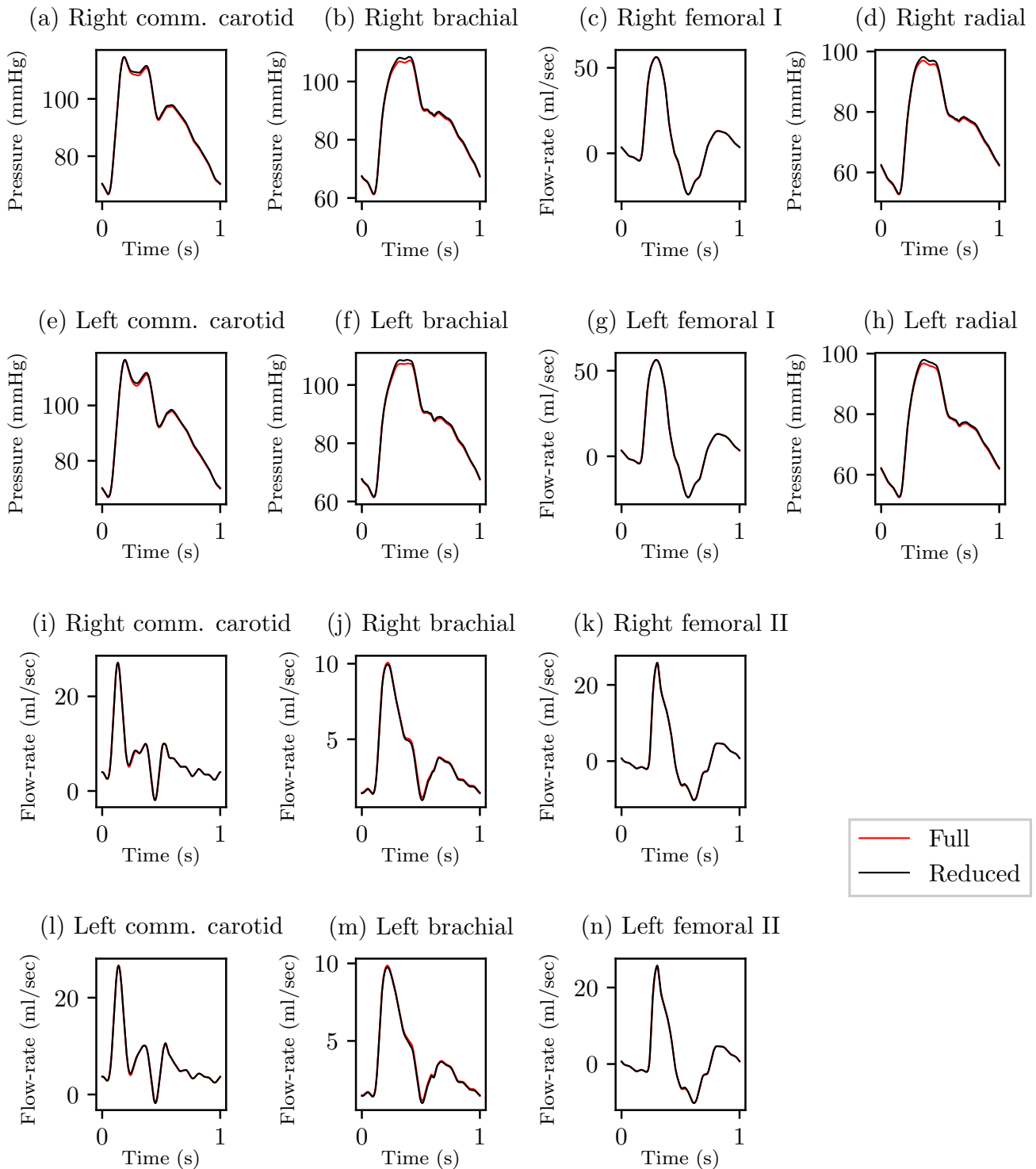


FIGURE 4 Comparison of pressure or flow-rate profiles at all the measurement locations taken from the full ADAN network and the reduced network produced by the removal of the first and second splenic segments, left gastric, common hepatic, and posterior interosseous.

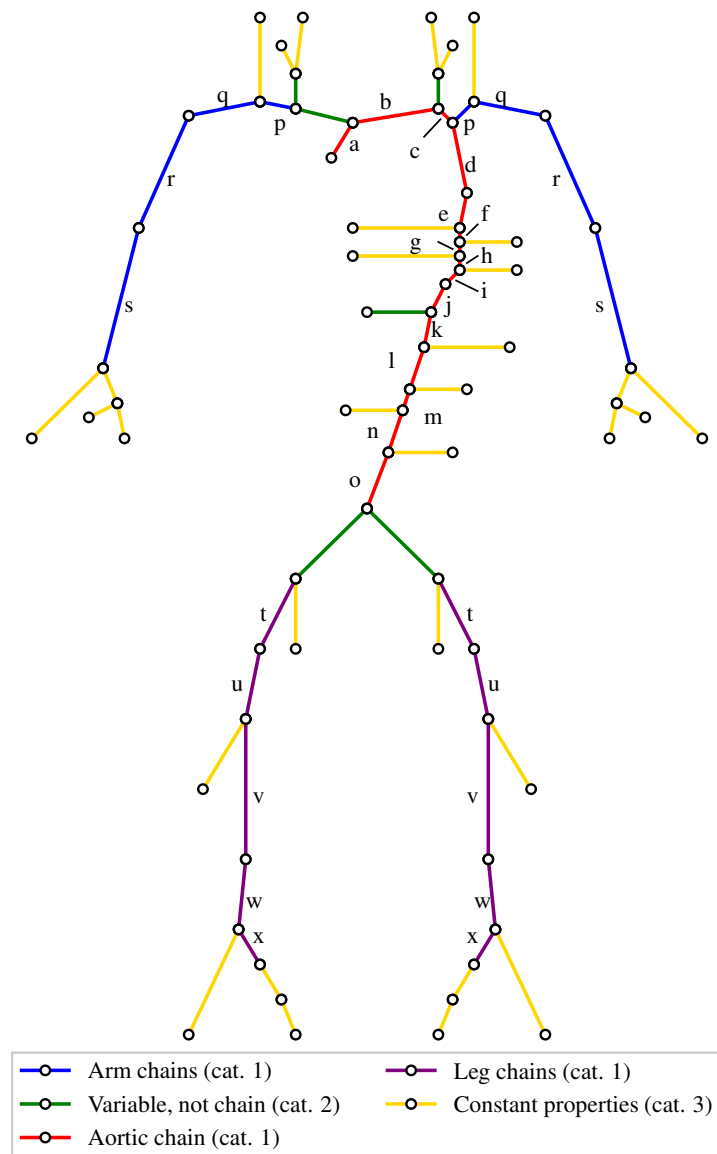


FIGURE 5 The reduced network and grouping of vessels. The aortic chain is shown in red and consists of: the first to fourth aortic arch segments denoted by ‘a’ through to ‘d’, the first to sixth thoracic aorta segments denoted by ‘e’ through to ‘j’, and the first to fifth abdominal aorta segments denoted by ‘k’ through to ‘o’. The right and left arm chains are shown in blue and consist of: the first and second subclavian segments denoted by ‘p’ and ‘q’ respectively, the axillary artery denoted by ‘r’, and the brachial artery denoted by ‘s’. The right and left leg chains are shown in purple and consist of: the external iliac denoted by ‘t’, the first and second femoral segments denoted by ‘u’ and ‘v’ respectively, and the first and second popliteal segments denoted by ‘w’ and ‘x’ respectively. Vessels which have variable properties but are not part of a chain are shown in green. Finally, vessels with constant properties are shown in yellow.

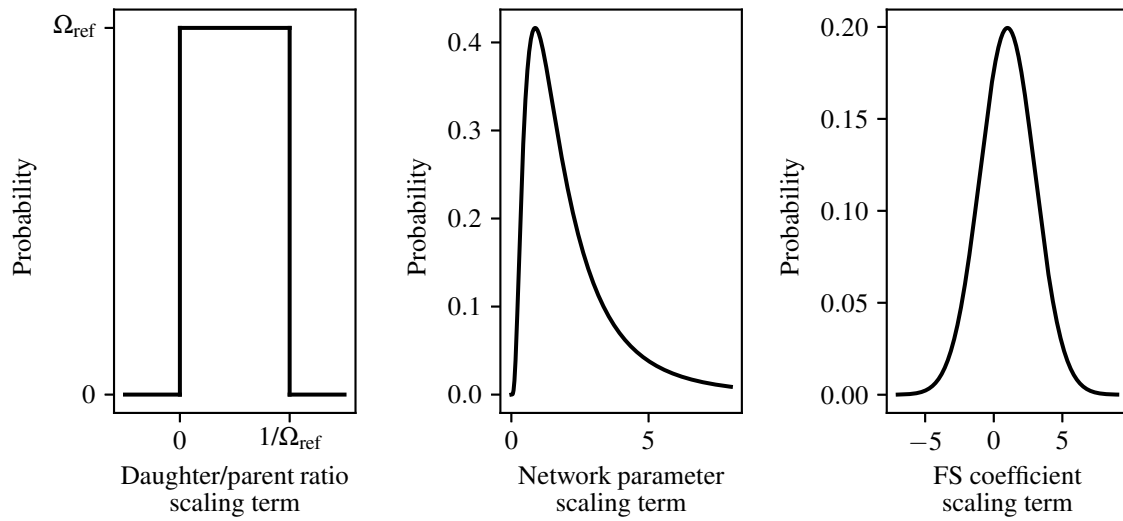


FIGURE 6 The three type of prior distributions used for all arterial network scaling parameters.

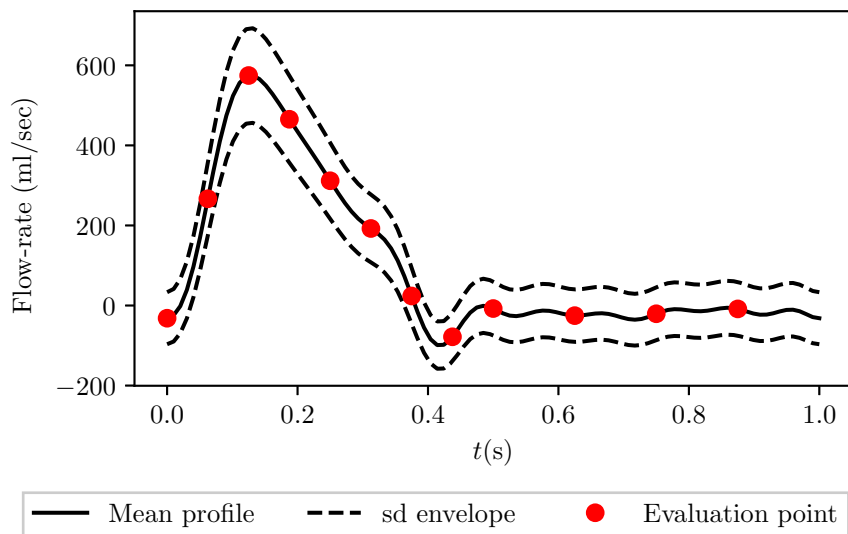


FIGURE 7 The measured time domain inlet flow-rate profile taken from the ADAN network is shown in solid black. The standard deviation (sd) envelope computed using Equation (23) are shown with dashed lines and the discrete time points at which the likelihood is evaluated is shown in red circles.

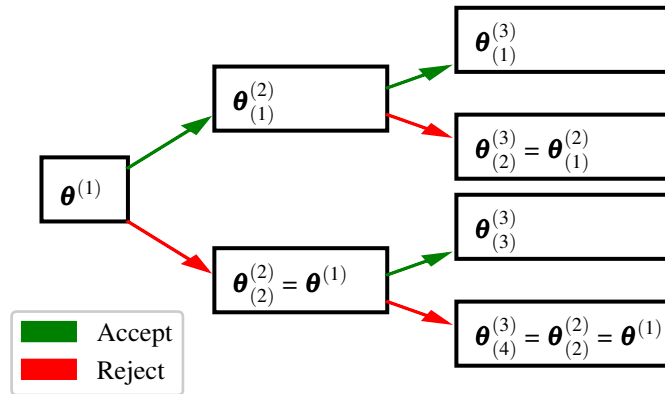
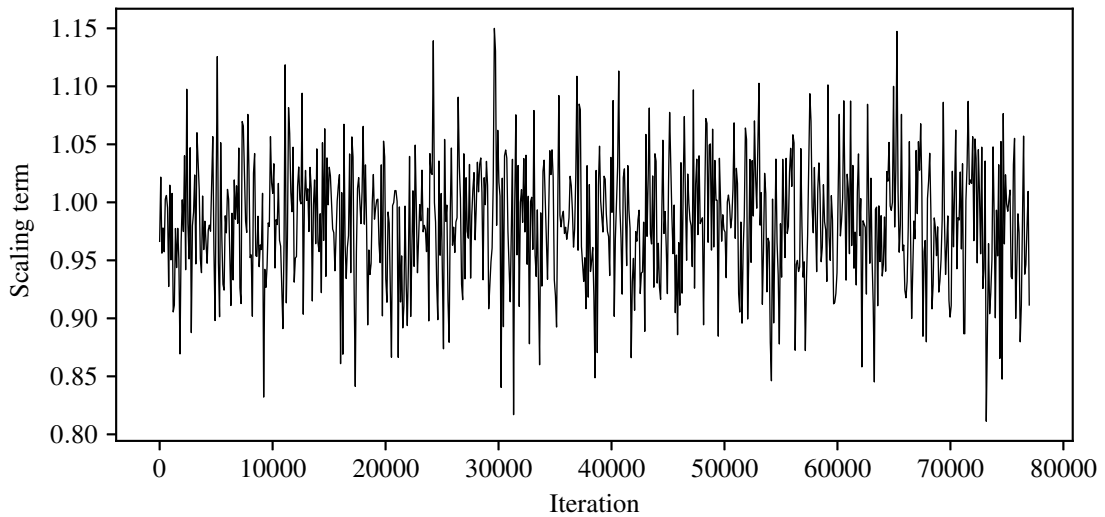


FIGURE 8 An example of an MCMC decision tree across two iterations of a chain. $\theta_{(j)}^{(i)}$ represents the j^{th} possible set (candidates) of the arterial network parameter scaling terms at the i^{th} step of the chain. Green branches denote the path of the chain taken if a candidate is accepted, while red branches show the path of the chain if a candidate is rejected.



e

FIGURE 9 MCMC trace plot of the scaling term applied to the length of the arterial vessels at every 100^{th} iteration. In this case, not net migration of the chain is observed.

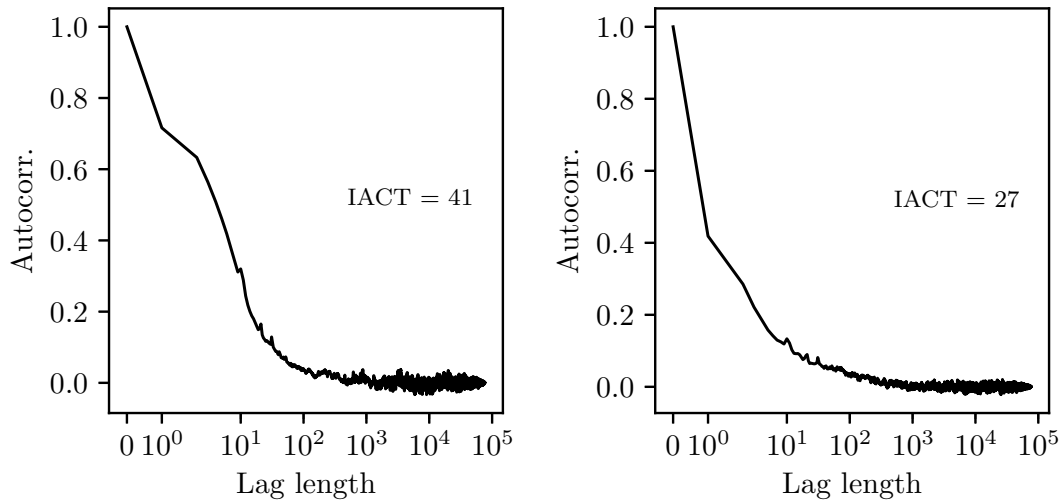


FIGURE 10 Autocorrelation as a function of lag-length in the MCMC chain for the vessel length scaling parameter (left), and for the diastolic pressure in the ascending aorta (right). IACT denotes the integrated autocorrelation time^{77,22,78}.

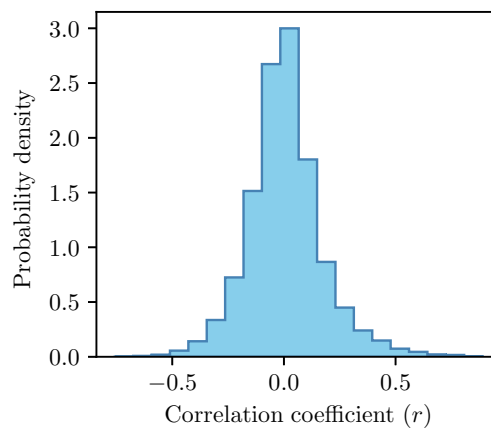


FIGURE 11 Histogram of correlation coefficient (r) between the 19,701 pairs of parameters in the posterior.

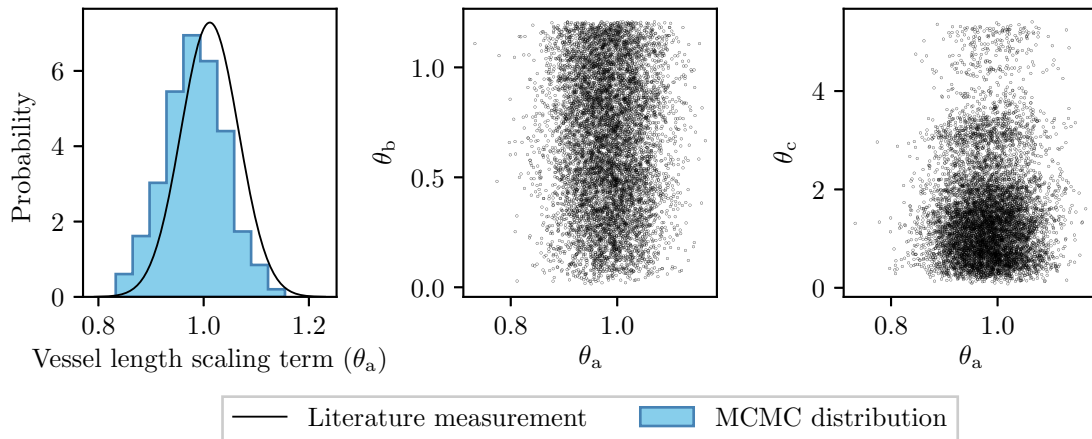


FIGURE 12 Posterior marginal distribution for the vessel length scaling term, denoted as θ_a (left); joint distribution between θ_a and the right arm chain's scaling factor for the β daughter/parent ratio denoted as θ_b (mid); and joint distribution between θ_a and the brachiocephalic trunk's scaling factor for β decay denoted as θ_c (right). While θ_a and θ_b show a largely uniform joint distribution, θ_a and θ_c show a potentially multimodal distribution with local peaks at $\theta_c \approx 1.0$ and $\theta_c \approx 3.2$.

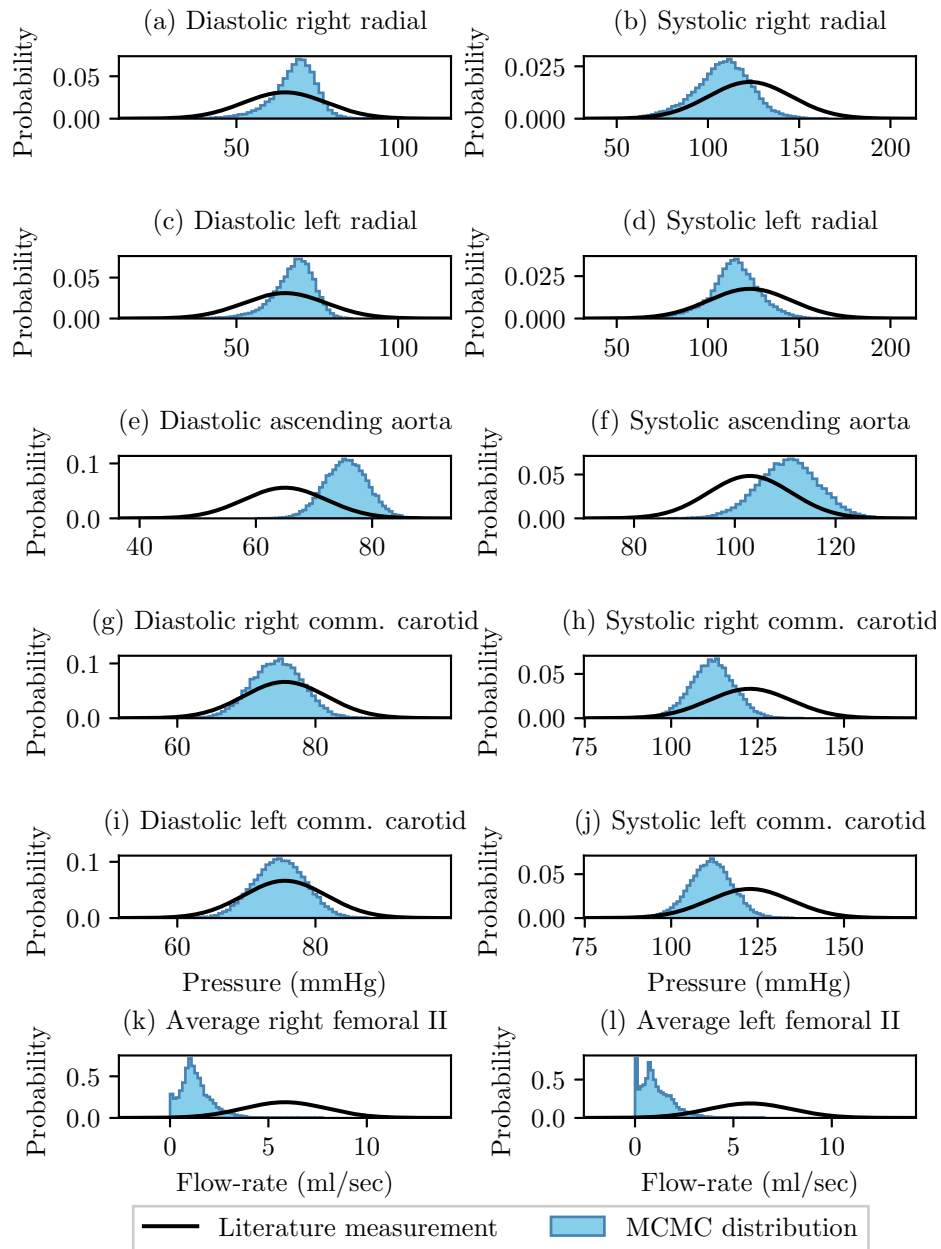


FIGURE 13 Histograms of the MCMC distributions of the pressure and flow-rate measurements at all measurement locations. The literature based measurements and associated error distribution at each location are overlaid in black (see Table 2). Diastolic and systolic pressure in the right radial artery shown in (a) and (b) respectively; the diastolic and systolic pressure in the left radial artery are shown in (c) and (d) respectively; the diastolic and systolic pressure in the ascending aorta are shown in (e) and (f) respectively; the diastolic and systolic pressure in the right common carotid artery are shown in (g) and (h) respectively; the diastolic and systolic pressure in the left common carotid artery are shown in (i) and (j) respectively; and the average flow-rate in the second segments of the right and left femoral artery are shown in (k) and (l) respectively.

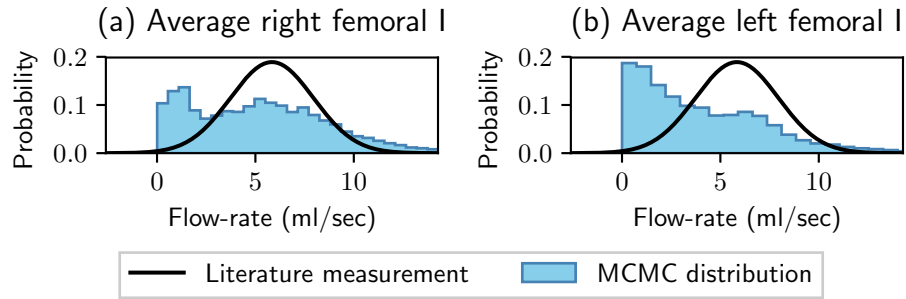


FIGURE 14 Histograms of the MCMC distributions of the average flow-rate in the first segment of the right femoral artery (c), and the first segment of the left femoral artery (d). The literature based measurement of average flow-rate in the femoral arteries and the associated error distribution is overlaid in black (see Table 2).

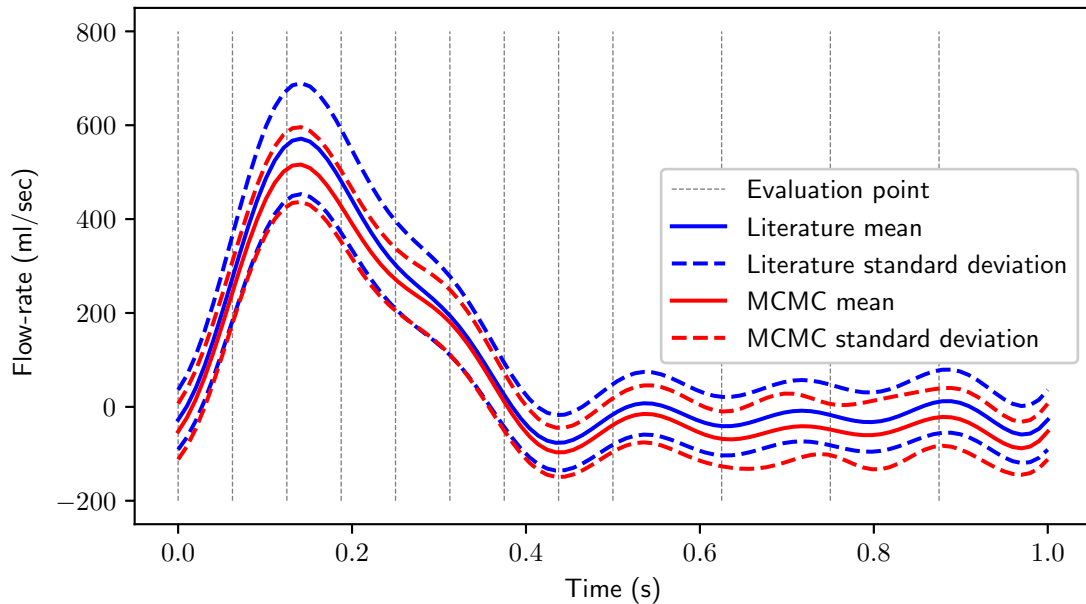


FIGURE 15 Comparison of the empirical distribution of the time varying inlet flow-rate profiles between the MCMC samples and the pseudo-measurement created through a GRF (see Section 5.3.2). All standard deviation curves in dashed lines depict mean ± 1 standard deviation.

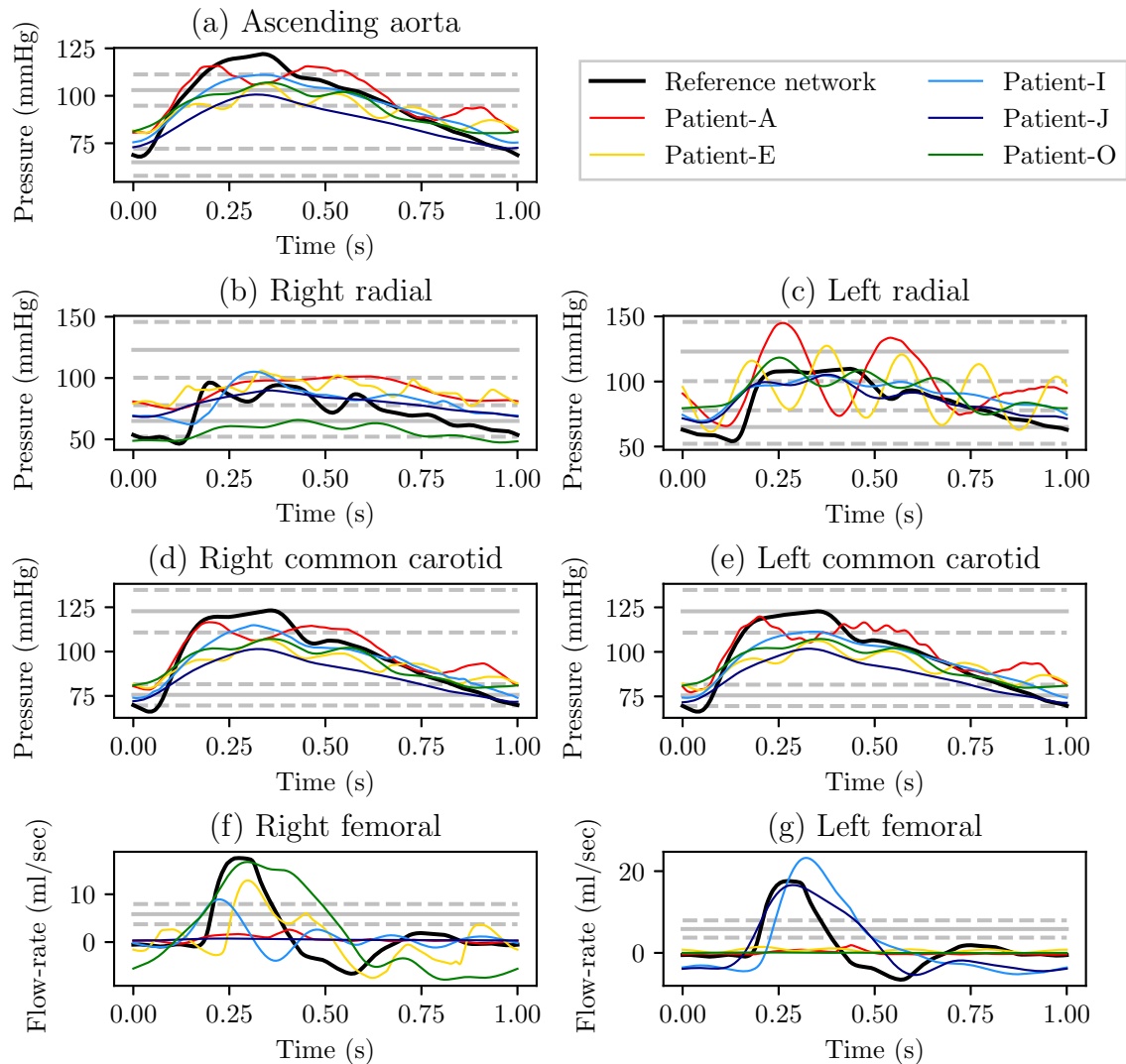


FIGURE 16 Pressure profiles in the ascending aorta (a), pressure profiles in the right radial artery (b), pressure profiles in the left radial artery (c), pressure profiles in the right common carotid artery (d), pressure profiles in the left common carotid artery (e), flow-rate profiles in the right second femoral artery (f), and flow-rate profiles in the left second femoral artery (g). In each figure the profiles taken from the reference network are shown in black; and the literature reported measurements and associated variances are shown by the solid (mean) and dashed grey lines (mean \pm 1 standard deviation) respectively.

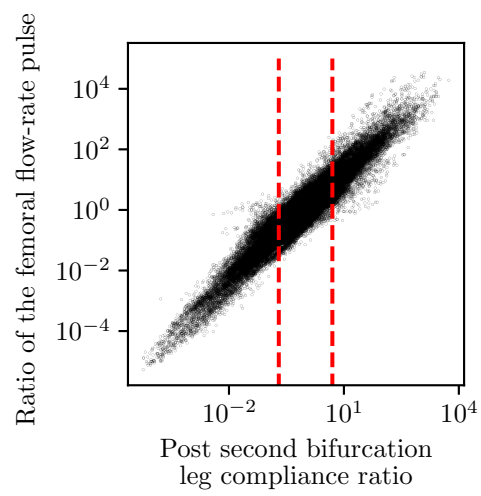


FIGURE 17 Left-to-right femoral pulse ratio against the left-to-right compliance ratio of reduced network (up to vessel ‘u’ in Figure 5). The vertical red dashed lines indicate the range of the filter imposed to remove VPs with extreme left-right asymmetry.