Math. Model. Nat. Phenom. 19 (2024) 20
https://doi.org/10.1051/mmnp/2024017

Mathematical Modelling of Natural Phenomena
www.mmnp-journal.org

SYNTHETIC PULSE WAVE DATASET FOR ANALYSIS OF VASCULAR AGEING IN ELDERLY PATIENTS

Artem Rogov[1,2], Timur Gamilov[1,2,3,*], Anna Bragina[1,4], Magomed Abdullaev[1], Natalia Druzhinina[1,4], Yuliya Rodionova[1,4], Rustam Shikhmagomedov[4], Maksim Tyulin[4] and Valeriy Podzolkov[4]

Abstract. This paper presents a methodology to generate synthetic pulse wave database. Each virtual subject is generated with the help of one-dimensional hemodynamics model of systemic circulation with lumped model of the left heart. This paper describes and compares two parameter optimization methods: unscented Kalman filter and Bayesian optimization. As a case study, an experiment is conducted to predict cardio-ankle vascular index (CAVI) values for real individuals with a machine learning algorithm trained on a synthetic population. The average error of 6.5% is achieved.

Mathematics Subject Classification. 92C50, 90C31, 92C35, 76-10, 76Z05.

Received June 4, 2024. Accepted August 28, 2024.

1. Introduction

Development of novel diagnostic techniques is a costly and time-consuming endeavor. One of the ways to reduce the costs is utilizing synthetic patient data and virtual populations [1, 2]. Expensive *in vitro* and *in vivo* trials can be preceded by tests on synthetically generated virtual subjects. There are a number of virtual population datasets suitable for various tasks: anatomy studies and meshing [3], pharmacology [4], pulse wave analysis [5]. These datasets can be used to test preliminary hypotheses, train various machine learning methods or they can serve as an educational tool. The effectiveness of virtual population depends on how well it represents a certain group of patients. Mathematical models used to generate synthetic subjects should be thoroughly tested and their parameters should be carefully adjusted.

The selection of parameters for mathematical models is a common problem [6–8]. The growing popularity of extensive synthetic databases of virtual patients further actualizes the problems of parameter selection in

Keywords and phrases: Virtual population, blood flow model, CAVI, parameter optimization, pulse wave database.

[1] World-Class Research Center "Digital Biodesign and Personalized Healthcare", Sechenov First Moscow StateMedical University, 19991 Moscow, Russia.
[2] Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Russia.
[3] Department of Mathematical Modelling of Processes and Materials, Sirius University of Science and Technology, 354340 Sochi, Russia.
[4] I.M. Sechenov First Moscow State Medical University (Sechenov University), 19991, Moscow, Russia.

* Corresponding author: gamilov_t_m@staff.sechenov.ru

biomedical modeling nowadays. The development of reliable algorithms for selecting global parameters of the model will facilitate the creation of such databases and streamline the work of researchers.

In this work, we generate a pulse wave dataset of healthy virtual patients with the age between 60 and 70 years old. The dataset is generated with the help of one-dimensional model of systemic circulation with a lumped parameter heart model with valve dynamics. Parameters of the model are optimized with help of the Unscented Kalman Filter (UKF) and Bayesian optimization to reach desired values of systolic blood pressure (SBP), diastolic blood pressure (DBP) and stroke volume (SV). Pros and cons of these methods were evaluated. Bayesian optimisation is a classical ML algorithm for finding the global maximum of some multidimensional function. It is used when the properties of the function are not known and its computation is costly [9, 10]. The UKF was originally presented as a method for nonlinear estimation [11], but it is shown that this approach can be applied to parameter identification task in large-dimensional systems [12]. Large systems of partial derivative equations are suitable for such parameter identification method. That's why this or similar approach is used in blood flow modelling [7, 13, 14].

We validate resulting virtual patient dataset of pulse waves by comparing pulse waves with another synthetic database presented in [5]. We also compare Cardio-Ankle Vascular Index (CAVI) of synthetic population and CAVI of real patient cohort. CAVI is used in practice as a marker of arterial stiffness and vascular ageing [15, 16]. We also train machine learning algorithm with the help of virtual population to identify patient's CAVI. After that we use trained algorithm to predict CAVI of real patients and compare it with actual measurements.

This paper is organized as follows. The model of a systemic circulation with lumped parameter left heart is described in Section 2.1. Then, we describe parameter optimization task and optimization methods in Section 2.2. The results are given in Section 3. Section 4 presents the discussion of the results and approaches.

2. Methods

2.1. One-dimensional hemodynamic model with lumped parameter heart

2.1.1. One-dimensional hemodynamic model

The blood flow in large systemic arteries is represented by one-dimensional model of unsteady flow of viscous incompressible Newtonian fluid in network of connected elastic tubes [17, 18]. For every separate vessel with index k, the laws of conservation of mass and momentum must be fulfilled:

$$\frac{\partial S_k}{\partial t} + \frac{\partial (S_k u_k)}{\partial x} = 0 \tag{2.1}$$

$$\frac{\partial u_k}{\partial t} + \frac{\partial (u_k^2/2 + p_k/\rho)}{\partial x} = \psi_k \tag{2.2}$$

where x – coordinate over the vessel; t – time; $\rho = 1.04\,\mathrm{g/cm^3}$ – constant density of blood; S_k – cross-sectional area over the vessel; u_k – cross-sectional averaged flow velocity; p_k – cross-sectional averaged blood pressure; ψ_k – friction force per unit mass of blood:

$$\psi_k = -\frac{8\pi\mu u_k}{\rho S_k} \tag{2.3}$$

where μ – dynamic viscosity of blood.

The elastic properties of the vessel wall are given by the tube law. This equation establishes the relationship between the cross-sectional area of the vessel and the blood pressure:

$$p_k(S_k) = \rho c_k^2 f(S_k) \tag{2.4}$$

where c_k – velocity of small disturbance propagation in the vessel wall, which can be interpreted as PWV (pulse wave velocity). $f(S_k)$ is monotone function:

$$f(S_k) = \begin{cases} exp(S_k/S_k^0 - 1) - 1, \ S_k > S_k^0, \\ ln(S_k/S_k^0), \ S_k \leq S_k^0 \end{cases} \tag{2.5}$$

The initial conditions can be defined for predetermined flow Q_0 as:

$$S_k(0,x) = S_k^0, \ u_k(0,x) = \frac{Q_0}{S_k^0} \tag{2.6}$$

In this work $Q_0 = 0$, except for the vessels connected to the heart, where boundary conditions set as:

$$u_{in}(t,0)S_{in}(t,0) = Q_{heart}(t) \tag{2.7}$$

$Q_{heart}(t)$ is determined from the heart model, which is described below.

The following boundary conditions are set in the region of the vessel conjunction. Firstly, mass conversation law must be satisfied:

$$\sum_{k=k_1,\dots,k_m} \varepsilon S_k(t,\widetilde{x}_k)u_k(t,\widetilde{x}_k) = 0 \tag{2.8}$$

where k_i – indexes of vessels connected with the junction; m – number of vessel connected with the junction; $\varepsilon = 1$, $\widetilde{x}_k = L_k$ for incoming vessels (L_k – length of the vessel); $\varepsilon = -1$, $\widetilde{x}_k = 0$ for outgoing vessels. Moreover, for all possible pairs of vessels in the junction, Bernoulli's equation must be fulfilled, thus ensuring complete continuity of pressure.

$$p_i(S_i(t,\widetilde{x}_i)) + \frac{\rho u_i^2(t,\widetilde{x}_i)}{2} = p_j(S_j(t,\widetilde{x}_j)) + \frac{\rho u_j^2(t,\widetilde{x}_j)}{2} \tag{2.9}$$

The outflow boundary conditions are placed under the assumption that the systemic arteries connect to the Windkessel compartments, which replace the remaining arteries:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{1}{R_1}\left(\frac{\mathrm{d}p_k(S_k(t,L_k))}{\mathrm{d}t} - \frac{\mathrm{d}P}{\mathrm{d}t}\right) \tag{2.10}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{Q}{C}\left(1 + \frac{R_1}{R_2}\right) - \frac{p_k(S_k(t,L_k)) - p}{R_2 C} \tag{2.11}$$

$$Q = u_k(t,L_k)S_k(t,L_k) \tag{2.12}$$

where R_1, R_2, C are parameters of Windkessel compartments; $p = 7mmHg$ for all compartments; p – pressure in compartment.

The structure of systemic arteries and terminal Windkessel elements are taken from [19].

All length, diameter of vessels, PWV and Windkessel compartment parameters for the base model are set according to [18]. Individual parameters are set with the help of optimization algorithms described below.

2.1.2. Lumped parameter model of left heart with valve dynamics

In this section we present a system of differential equations describing the dynamics of the heart chambers: left atrium (la) and left ventricle (lv). They connect the pulmonary veins and the aortic root. Also, there is a description of dynamics of mitral valve (mv) and aortic valve (av).

The model of the heart is one of the main features that distinguishes this work from other synthetic pulse wave databases. In other approaches [5] cardiac output is given as predefined function instead of a heart model. The presence of an additional model adds computational complexity, but allows modeling the effects directly related to heart diseases, *e.g.*, valve regurgitation.

For each of the chambers, equation describing the change in volume during the cardiac cycle is:

$$I_k \frac{\mathrm{d}^2 V_k}{\mathrm{d}t^2} + R_k P_k \frac{\mathrm{d}V_k}{\mathrm{d}t} + E_k(t)(V_k - V_k^0) + P_k^0 = P_k, \ k = la, lv \tag{2.13}$$

where k – indexes of chambers; I_k – coefficient of inertia; R_k – coefficient of hydraulic resistance; P_k – pressure averaged over the chamber volume; P_0 and V_0 – reference values of pressure and volume. The function $E_k(t)$ is used express the variable elasticity of cardiac tissue during contraction [20]:

$$E(t) = E^d + \frac{E^s - E^d}{2} e(t), \ 0 \le e(t) \le 1 \tag{2.14}$$

where $e(t)$ corresponds to the periodic activation potential. For the left ventricle the function is:

$$e_{lv}(t) = \begin{cases} 0.5(1 - \cos(\frac{t}{T_{s1}}\pi)) & 0 \le t < T_{s1} \\ 0.5(1 - \cos(\frac{t - T_{s2}}{T_{s1} - T_{s2}}\pi)) & T_{s1} \le t < T_{s2} \\ 0 & T_{s2} \le t \le T \end{cases} \tag{2.15}$$

For left atrium it is:

$$e_{la}(t) = \begin{cases} 0 & 0 \le t < T_{pb} \\ 0.5(1 - \cos(\frac{t - T_{pb}}{T_{pw}} 2\pi)) & T_{pb} \le t < T \end{cases} \tag{2.16}$$

T is the duration of cardiac cycle, T_{s1} is the time to systolic peak, T_{s1} is the duration of the systole (LVET), T_{pb} is the duration of left atrium contraction.

The mass conversation law for left atrium and left ventricle can be presented as:

$$\begin{aligned} \frac{\mathrm{d}V_{la}}{\mathrm{d}t} &= Q_{plv} - Q_{mv} \\ \frac{\mathrm{d}V_{lv}}{\mathrm{d}t} &= Q_{mv} - Q_{av} \end{aligned} \tag{2.17}$$

where Q_{plv} – flow from pulmonary veins; Q_{mv} and Q_{av} – flow through mitral valve and aortic valve respectively. Here Q_{av} is equivalent to Q_{heart} from (2.7), which links the heart and blood flow models together. To determine the relationship between flow and pressure drop equations from work [21] are used. They take into account the physiological non-stationarity of blood flow:

$$\Delta P_k = L_k \frac{\mathrm{d}Q_k}{\mathrm{d}t} + \alpha_k Q_k + \beta_k Q_k |Q_k| \tag{2.18}$$

where P_k – pressure drop for each flow (Q_plv, Q_mv, Q_av); L_k – constant coefficient; $\alpha_k = 0$ in this work, because viscous resistance is low; β_k – function, which depends on area of the valve, according to [22, 23]:

$$\begin{gathered}\beta_{mv} = \frac{\rho}{2B_{mv}}(\frac{1}{A_{mv}} - \frac{1}{A_{ref}})^2 \\ \beta_{av} = \frac{\rho}{2B_{av}}(\frac{1}{A_{av}} - \frac{1}{A_{aorta}})^2 \\ \beta_{plv} = const\end{gathered} \tag{2.19}$$

where B_k – constant parameters; A_{mv}, A_{av} – valve angle dependent cross-section area; A_{aorta} – aortic root cross-sectional area, which is derived from the hemodynamic equations; $A_{ref} = 5\,\mathrm{cm}^2$.

The equations describing the dynamics of opening-closing angles of the heart valves are essentially Newton's second law. In this work we use equations in the following form:

$$\begin{aligned}\frac{\mathrm{d}^2\theta_{mv}}{\mathrm{d}t^2} &= (P_{la} - P_{lv})K^p_{mv}\cos\theta_{mv} - K^f_{mv}\frac{\mathrm{d}\theta_{mv}}{\mathrm{d}t} - F_{mv}(\theta_{mv}) \\ \frac{\mathrm{d}^2\theta_{av}}{\mathrm{d}t^2} &= (P_{lv} - P_{av})K^p_{av}\cos\theta_{av} - K^f_{av}\frac{\mathrm{d}\theta_{av}}{\mathrm{d}t} - F_{av}(\theta_{av})\end{aligned} \tag{2.20}$$

These equations correspond to the work [24] with the addition of the third term. Each term in the equation is responsible for a specific force acting on the valve: the pressure force that drives the valve, the friction force, and the force that prevents the valve from over-opening ($F_k(\theta_k)$).

$$F(\theta) = \begin{cases} 0 & \theta_{min} < \theta < \theta_{max} \\ \mathrm{e}^{1000(\theta - \theta_{max})} - 1 & \theta > \theta_{max} \\ -(\mathrm{e}^{1000(\theta_{min} - \theta)} - 1) & \theta_{min} < \theta \end{cases} \tag{2.21}$$

It is necessary to add a boundary condition in the form of constant pressure in the pulmonary veins:

$$P_{plv} = 13\ mmHg. \tag{2.22}$$

2.2. Parameter optimization

2.2.1. Optimization problem

An optimization problem for previously described model of systemic circle with left heart can be formulated as follows. The ultimate goal is to select a set of model's parameters to obtain concrete values of three commonly used medical indicators: systolic blood pressure in brachial artery (SBP), diastolic blood pressure in brachial artery (DBP) and stroke volume of left ventricle (SV). These indicators were chosen because they are widely used in clinical practice, can be easily measured and are illustrative markers of cardiovascular conditions [25]. There are other important cardiovascular characteristics that should be identified: heart rate, aorta diameter, left ventricular ejection time (LVET). However, these characteristics can be prescribed as model's parameters directly.

The parameters that are used to achieve the required SBP, DBP and SV are:

- R_{la}, R_{lv} – coefficients of hydraulic resistance of left atrium and left ventricle from (2.13)
- I_{la}, I_{lv} – coefficients of inertia of left atrium and left ventricle from (2.13)
- K^p_{mv}, K^p_{av} – coefficients of pressure force of mitral and aortic valve from (2.20)
- K^f_{mv}, K^f_{av} – coefficients of friction force of mitral and aortic valve from (2.20)
- Total resistance of Windkessel compartments; R_1, R_2 – resistances of a single element from (2.10), (2.11)
- Total compliance of Windkessel compartments; C – compliance of a single element from (2.11)

- p_{inf} from (2.11)
- Pulmonary vein pressure (2.22)
- Rescaling coefficient of all PWV c_k from (2.4)

This choice of parameters is due to the fact that it allows to change the properties of the heart chambers, valves, vessels, terminal vessels and the value of venous pressure. Thus, it is possible to influence practically any component of the above described system of systemic circle and left heart.

It is worth noting that the question of which parameters to choose was decided in the process of numerous tests. The most important criterion as to whether to keep the parameter was how significantly it changes in the process of selection from its initial value.

The necessary component for optimization problem definition is the formulation of the numerical function, the minimum or maximum of which the algorithm will search for. In this work such target function formulated as follows:

$$error(SBP, DBP, SV) = max\left(\frac{|SBP - SBP_{target}|}{SBP_{target}}, \frac{|DBP - DBP_{target}|}{DBP_{target}}, \frac{|SV - SV_{target}|}{SV_{target}}\right) \tag{2.23}$$

In order to generate a virtual population of different ages, base models were introduced for each decade from 25 years to 75 years, similarly to [5]. For each decade from 25 years to 75 years diameter, length and base values of PWV have been recalculated. Length and diameter of aorta and some other large vessels were defined based on experimental data from [26]. PWV values for each patient were re-scaled according to data from virtual database from [5]. Age-specfic heart rate (HR) is predefined parameter for every virtual patient and was set according to data from [5], where extensive analyses of medical data were done.

In this work, we focus on patients between 60 and 70 years old. We use a baseline set of parameters that correspond to typical cardiovascular characteristics (SBP, DBP, SV, HR, aortic diameter, LVET) of average 65 year old. We vary these characteristics and solve optimization problem to achieve desired values of SBP, DBP and SV.

We present a general description of two methods, which were used to solve previously defined optimization problem.

2.2.2. The unscented Kalman filter

There are different types of Kalman Filter methods, all of them are used for nonlinear optimization problems. In this work we will focus on the Unscented Kalman Filter (UKF), because of its significant advantages over analogues. For example, the UKF uses exact nonlinear system, not linearized version. It captures distribution of current state, which assumes to be Gaussian random variable, using minimal set of sigma-points. After propagation through nonlinear system, UKF captures posterior covariance and mean with third order (Taylor expansion) for any non-linearity [11].

Detailed explanation of the mathematical background can be found in [11, 12]. UKF can be used for parameter estimation in dynamical time-discrete models from available measurements of its state [7, 12, 14]. After consideration of general approach for these type of systems, we propose a simplification of this algorithm, which is more suitable for our purpose. Let a dynamic time-discretized system be represented by:

$$X_{n+1} = \widetilde{F}(X_n, \Theta_n), \tag{2.24}$$

where X_n is vector, which represents state of the system at n-th iteration, Θ_n is a set of parameters for estimation, $\widetilde{F}$ is operator of propagation of the system on next step.

$$Z_n = \widetilde{H}(X_n) + \epsilon, \tag{2.25}$$

where Z_n is vector of measured variables at some time instants, $\widetilde{H}$ is operator, which transforms state of the system to concrete measured variables, ϵ is a vector of Gaussian noise. Algorithm of any Kalman filter can be represented as prediction-correction scheme for state (X_n, Θ_n), where prediction step calculates X^-_{n+1} and Θ^-_{n+1}:

$$\begin{aligned} X^-_{n+1} &= \widetilde{F}(X^+_n, \Theta_n) \\ \Theta^-_{n+1} &= \Theta_n \end{aligned} \tag{2.26}$$

And correction, where the difference between the measurement and the one calculated according to the new state of the system is taken into account.

$$\begin{aligned} X^+_{n+1} &= X^-_{n+1} + K_X(Z_{n+1} - \widetilde{H}(X^-_{n+1})) \\ \Theta^+_{n+1} &= \Theta^+_{n+1} + K_\Theta(Z_{n+1} - \widetilde{H}(X^-_{n+1})) \end{aligned} \tag{2.27}$$

K_X and K_Θ are Kalman matrices defined in a way to minimize difference between actual measurement and observed state.

In case of our specific parameter estimation problem, the state vector of the system $X_n = [SBP, DBP, SV]^T_n$ defined only in a steady state, *i.e.* after calculation of several cardiac cycles. The vector of measurements Z_n in our case is defined as constant vector Z of required values of SBP, DBP, SV and doesn't change with iterations. Consequently, subscript of all variables should be seen not as parameter at certain time step, but as an iteration number of the optimization process. Version of exact algorithm with concrete details can be found in [14].

2.2.3. Bayesian optimization

Bayesian optimization is a machine-learning-based optimization algorithm, which solves task:

$$max\, f(\boldsymbol{x}), \quad \text{where} \quad \boldsymbol{x} \in A$$

Usually target function and domain A satisfy a number of conditions for the most successful application of method [9].

- $A \subset R^d$, where $d < 21$.
- A is simple set, for example multi-dimensional square or simplex.
- Calculation of $f(\boldsymbol{x})$ is time-consuming. It might be that there is an ability to perform only few hundred iterations.
- Structure of function is completely unknown, which means that there is no information about linearity, convexity or other specific characteristics. So no specially constructed methods can be used.
- There is no information about first and second derivative of function. That is why gradient descent or Newton's method are impossible to use.
- Goal is to find a global extremum, not local.

Generally, algorithm consist of two main components:

1. Gaussian process regression, which provides posterior probability distribution and describe potential values of objective function.
2. Acquisition function, which provides point where to sample next based on current distribution. This function should be easy to calculate. In this work it is always expected improvement.

The method works as follows. In concrete iteration there are n points $x_1, ..., x_n$ and values of target function $f(x_1), ..., f(x_n)$. Gaussian process regression makes evaluation of posterior probability distribution of values of functions, then acquisition function finds x_{n+1} for next sample. After calculation of target function occurs

TABLE 1. Statistics for target age group.

Age, years	65.47 ± 2.88
Males, %	39.73%
HR, bpm	78.85 ± 10.37
SBP, mmHg	137.07 ± 19.29
DBP, mmHg	82.07 ± 10.46
RCAVI, m/s	8.90 ± 1.50
LCAVI, m/s	9.24 ± 2.08

$f(x_{n+1})$ cycle ends and new evaluation of posterior distribution needed. The stopping criterion is usually defined by an amount of possible iterations. Returned solution might be largest $f(x)$ or point with the largest mean in distribution.

For a previously optimization problem Bayesian optimization could be a suitable method, because time-consuming calculations of stable target values (SBP, DBP, SV). Structure of function, which depends on parameters of heart model, vascular network model and specialties in geometry of vascular network, remains completely unknown in terms of features of function or its derivatives. In order to use Bayesian optimization, domain of the function should be properly appointed. In this work we use multi-dimensional analogue of square, *i.e.* every parameter has its fixed boundaries and can be selected from a preset interval.

This optimization algorithm is generally used in machine learning for tuning hyperparameters in neural networks. However, it can also be used for parameter optimization in numerical solution of systems of differential equations [27]. For more detailed explanation of algorithm and its modifications see [9, 10].

2.3. Overview of patient's dataset

The data required for the validation of the virtual population was collected by staff of the University Clinical Hospital No. 4 of the I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University). The study was able to collect data on 600 patients, with an average age of 36.0 ± 18.3 years. The study was approved by the institutional ethics committee and complied with the Declaration of Helsinki and good clinical practice guidelines. All patients provided informed consent prior to participation. They were administered a questionnaire to assess risk factors, as well as a detailed anthropometric study with evaluation of obesity markers and some data from blood biochemical analysis.

RCAVI and LCAVI were measured for all patients. The CAVI index is directly proportional to the stiffness of the arterial wall. It is calculated based on the classical method of determining vascular wall stiffness – pulse wave velocity (PWV). The sphygmomanometric method on the VaSera-1000 FUCUDA DENSHI device was employed to evaluate CAVI for each patient. The difference between RCAVI and LVACI is that LCAVI is measured with the help of pressure cuffs on left hand and ankle and RCAVI – on the right hand and ankle.

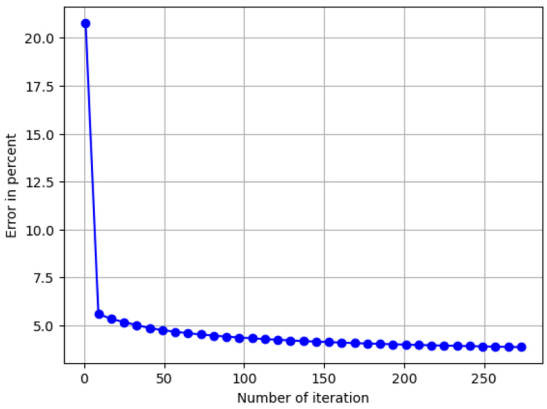
For further comparison of physiological parameters between the virtual and synthetic populations, only patients within the age range of 60 to 70 years (73 patients) were included in the analysis. The main characteristics with their standard deviations are contained in the Table 1.

3. RESULTS

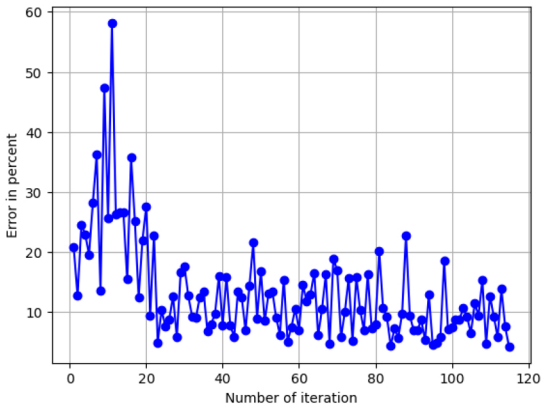
3.1. Comparison of parameter evaluation methods

A number of test cases were calculated to compare speed and quality of the two optimization algorithms described above. The base values of SBP, DBP, SV, HR, vessels' diameters and LVET for 25, 55, 75 year old subjects were taken from the database of the article [5]. The parameters for these cases were then optimized using both algorithms.

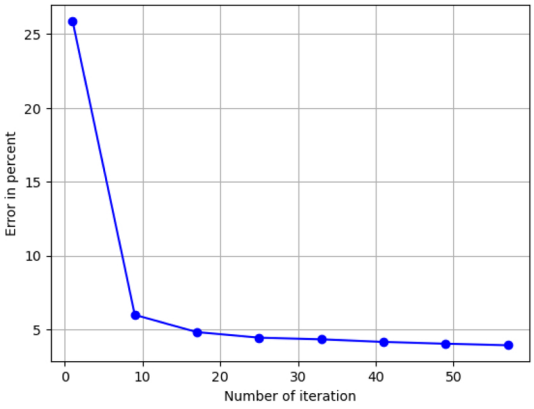
The plots of convergence rates are presented on Figure 1. On x-axis there is a number of iterations, on y-axis – error in percents. It should be noticed, that unscented Kalman filter converges uniformly unlike bayesian



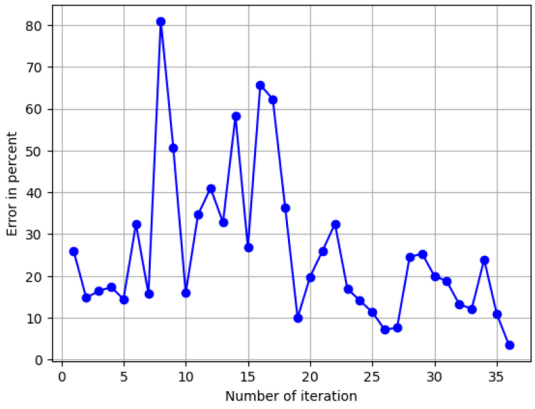
25 years, UKF



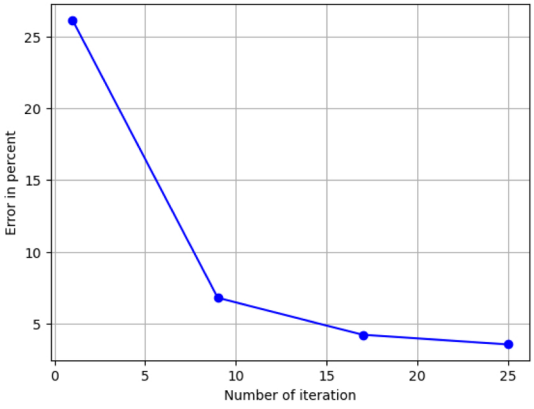
25 years, bayesian optimization



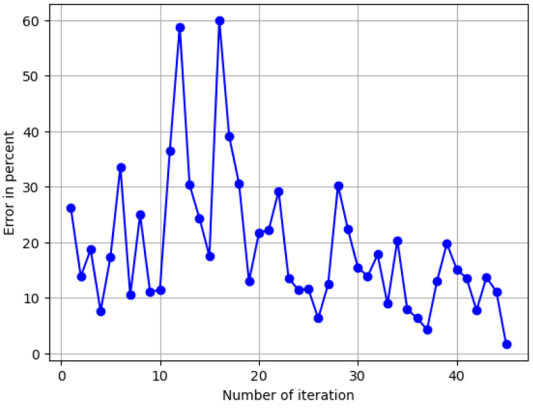
55 years, UKF



55 years, bayesian optimization



75 years, UKF



75 years, bayesian optimization

FIGURE 1. Convergence of UKF and Bayesian optimization.

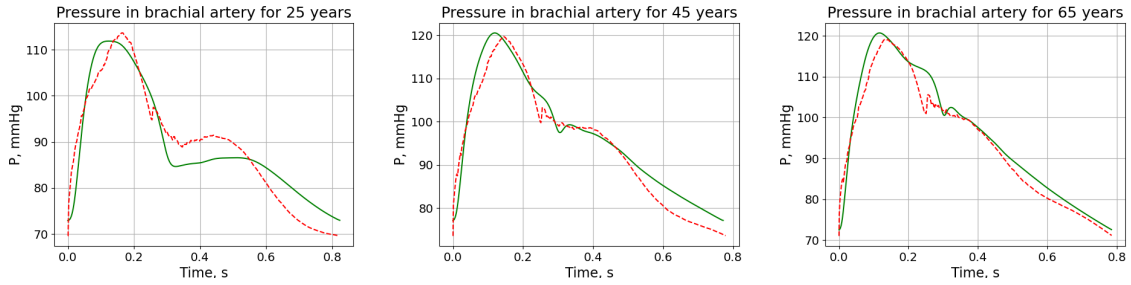


FIGURE 2. Comparison of brachial artery pulse waves calculated using UKF (continuous line) with the database from the [5] (dotted line).

optimization. This can be explained by the essence of the bayesian optimization algorithm. During the process of parameter adjusment this algorithm checks a lot of different points to extract more data for effective Gaussian process regression.

In case of 75 years UKF achieves the required accuracy of 3% or less faster than the Bayesian optimization, in case for 55 and 25 years Bayesian optimization performs better.

This result can be explained by the fact that for UKF the initial approximation is critical. If it is close enough to the required result, it will converge quickly.

However, if we set the values of the required parameters to be extremely far from baseline values, then Bayesian optimization is more effective. For the 25 year old subject, it was the significant difference from the initial approximation that led to a faster rate of convergence of the Bayesian optimization.

It should also be noted that it is possible to increase the efficiency of Bayesian optimization by giving not one point of the initial approximation, but several. And it is not necessary that it should be data related to the object with specific parameters of vessels and even age. Such data can be considered as noisy. Either the algorithm itself can work with them, or we can consider the use of special modifications for noisy data. It means that the more synthetic subjects we have the more effective the Bayesian optimization.

3.2. Comparison with other synthetic database

The initial comparison of the efficacy of the selected method for constructing a comprehensive database was conducted with a database from [5].

The mean values of SBP, DBP, and SV for individuals aged 25, 45, and 65 years were extracted from [5]. Subsequently, the UKF was employed to identify parameters that corresponded to these targets. A comparison of pulse waves in the brachial artery is presented in Figure 2. It should be noted that the curves are quite similar in shape, with systolic and diastolic pressures different by no more than 5%.

3.3. Generation of virtual population between 60 and 70 years old

A baseline model for a 65 year old healthy person was used. The following parameters were varied for direct generation: SBP, DBP, SV, HR, diameters of large vessels, left ventricular ejection time (LVET). Data on average values of the indicators and their variations are taken from [5]. We vary the mean values of the parameters by adding and subtracting the standard deviation. As a result, each parameter has three possible values: mean, mean + standard deviation, mean − standard deviation. All values presented in Table 2. We obtain $3^6 = 729$ virtual patients for a given age group.

The parameters were optimized using the computational cluster of Sechenov University (24 CPU nodes, 52 cores each). The iterative UKF described above was used to fit the parameters since it performed better for elderly virtual patients. Iterations were stopped when the error function (2.23) is below 0.05 (5%).

TABLE 2. Average values of the indicators and their variations. Diameters vary simultaneously.

SBP, mmHg	120.2 ± 8.3
DBP, mmHg	72.3 ± 6.6
SV, mmHg	55.8 ± 10.4
HR, bpm	76.3 ± 9.0
RCAVI, m/s	8.90 ± 1.50
Diameter of ascending aorta, mm	41.4 ± 3.0
Diameter of descending thoracic aorta, mm	27.6 ± 2.0
Diameter of abdominal aorta, mm	16.3 ± 1.2

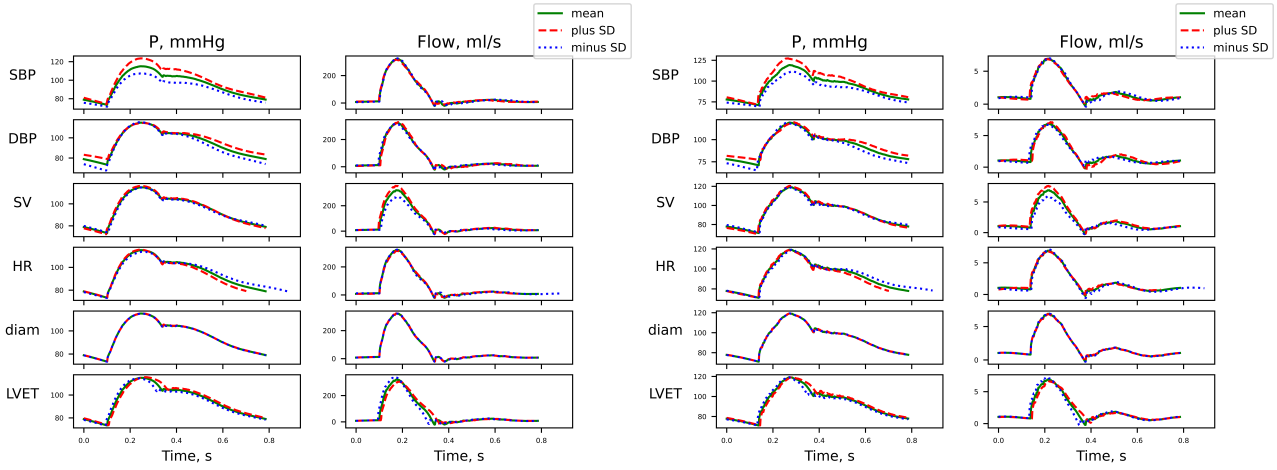


FIGURE 3. Pulse waves in aortic arch (left) and brachial artery (right). Blood pressure and flow in are shown for isolated variation of model characteristics. The characteristic on the left is the one that underwent variations.

Figure 3 depicts the pulse waves in brachial artery and aortic arch. On the left are the parameters that were varied during parameter selection. We can see that variation of arteries' diameters had the lowest effect on the shape of the pulse waves.

3.4. CAVI evaluation

3.4.1. Comparison on CAVI in synthetic and real patients

We calculated LCAVI for each virtual patient. The calculation was performed according to the methodology [28, 29]. In essence, to calculate CAVI it is necessary to know the distance from the aortic valve to the tibial artery and the time offset between the onset of pulse wave rise in the brachial artery and the tibial artery. It should be noted that height did not vary in any way when the database was obtained. Therefore, we consider the distance from the aortic valve to the tibial artery to be fixed. To obtain it, the lengths of the head, neck and the distance from the neck to the location of the aortic valve were subtracted from the height of 165 cm.

As a result, the obtained mean value of LCAVI for 65 years old virtual subjects – 9.28 m/s with standard deviation 1.11 $\frac{m}{s}$. Mean LCAVI value for real patients of the same age group is 9.24 $\frac{m}{s}$ with standard deviation 2.07 $\frac{m}{s}$. We can see that mean LCAVI values for virtual subjects are very close to mean LCAVI values of real patients. This is despite the fact that CAVI (or PWV) was not used as one of the characteristics for

Table 3. Comparison of ML methods.

ML method	Mean Deviation from the measured CAVI, %
Linear regression	16.71%
Decision tree	6.47%
Random forest	8.26%
Support vector	10.80%
Feedforward neural network	30.70%

optimization. Standard deviations for real patients are larger than standard deviations for synthetic subjects. This might be due to the fact that the height of synthetic subjects is the same, which affects variations in CAVI values.

3.4.2. Application of virtual population to predict CAVI value in real patients

As a test for virtual population the following experiment was set. From real data we found 10 patient of target age group for which the following indicators were measured: SBP, DBP, SV, blood flow velocity in the common carotid artery, LCAVI. There are studies that show a correlation between CAVI and blood flow velocity in the carotid artery [30]. Therefore, we decided consider it as one of the input parameters.

This set of measurement was extracted for every virtual subject in synthetic population. Machine learning methods were then tested to predict the value of LCAVI from the set of indicators mentioned above. The following methods were used: linear regression, decision tree, random forest, support vector machine, feedforward neural network.

Training was done entirely on synthetic data, while testing was done on real data. The best result was demonstrated by regression tree. Comparison of different methods shown in Table 3. Hyperparameters for regression tree was chosen with help of grid search cross validation with the help of sklearn Python library. Resulting hyperparameters: maximum depth of the tree – 10; minimum number of samples required to split an internal node – 2; minimum number of samples required to be at a leaf node – 1. Average deviation from the desired LCAVI value is 6.5%, maximum deviation is 18%, root mean square error is 0.25 $\frac{m}{s}$. A Bland-–Altman plot for this case presented in Figure 4.

Resulting average deviation is 6.5%. For comparison, deviations during repeated CAVI measurements on the the same patient are around 3% [31]. The difference between typical healthy CAVI values of 55 year old and 65 year is around 6% and the difference between 65 year old and 75 year old CAVI is 15% [29]. Our approach requires more patients to test machine learning algorithms, but achieved precision already allows us to distinguish healthy CAVI values from pathological ones.

4. Discussion

In this study, a virtual population of 729 healthy subjects aged 65 was generated based on a one-dimensional systemic circulation model with a lumped parameter heart model. We used UKF to adjust parameters of the model. UKF was also compared with Bayesian optimization. The advantages and disadvantages of each approach were noted. It should be noted that UKF provides a more gradual and precise approach to achieving the desired accuracy, although it relies upon the initial starting point. In contrast, Bayesian optimization is less reliant on the starting point. In this particular case UKF was more effective, as the majority of the requisite targets are situated in close proximity to the initial baseline model. Additionally, the specific intricacies associated with the implementation and execution of the parameter selection algorithm on a high-performance cluster made UKF the most suitable option. Other age groups might require Bayesian optimization that proved to be more

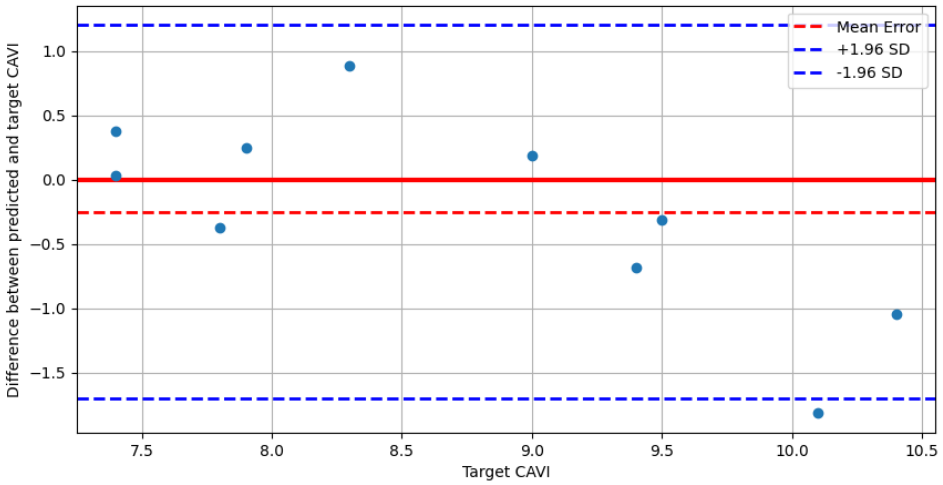


FIGURE 4. Bland–Altman plot for regression tree with 95% confidence interval represented by dotted lines.

effective for 25 years old baseline model. UKF demonstrated faster convergence for older virtual patients with and it was used for our CAVI case study.

The virtual population was validated using real patient data. The comparisons of CAVI demonstrated satisfactory results, which indirectly indicate the quality of the obtained data. Moreover, an experiment was conducted to predict the CAVI value of real patients from the sample after training the decision tree on the synthetic set. The results of this experiment were encouraging, with an average deviation of 6.5%. We should note that only 10 patient of the desired age group underwent blood flow velocity measurements. A larger cohort of patients is required for further testing.

In the future, it is planned to generate virtual populations of other age groups using similar approach and also to validate it with the help of real data. Heart model allows to simulate populations with various pathologies (mitral, aortic insufficiency) and study the effects of these pathologies on peripheral pulse waves. This could find use in wearable pulse wave detectors. Another crucial task for the future is to identify additional applications of such a database in clinical trials, in preliminary testing of the predictive power of machine learning algorithms and in the educational process.

One of the limitations of this approach is the lack of a single solution for parameter optimization problem. Nevertheless, the precise nature of the solution is of lesser importance than the degree of correspondence between the indirect physiological parameters of a virtual patient and those of real people. Moreover, the current methodology does not specify the sex of the virtual subjects, as the primary focus is on the hemodynamic model of systemic circulation and parameter optimization methods. However, incorporating sex-specific differences could be considered as future work. This could enhance both accuracy and applicability of the synthetic pulse wave database. Furthermore, in certain countries, the utilisation of such databases is subject to regulatory constraints, which may restrict their applicability in clinical trials. Our synthetic population was validated with the help of patients in a single hospital. Ideally, a larger cohort of patients from various countries should be studied. Virtual population should be tuned to the population it will be used for.

It should also be noted that the model under consideration does not incorporate anthropometric data (height, weight). Consequently, when calculating CAVI on the virtual population, an average height was assumed. Furthermore, the parameters were selected for healthy individuals with average basic indices, which does not account for the individual characteristics of some patients.

Funding

This work was financed by the Ministry of Science and Higher Education of the Russian Federation within the framework of state support for the creation and development of World-Class Research Centers "Digital biodesign and personalized healthcare" no. 075-15-2022-304.

Author contributions

Artem Rogov: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. Timur Gamilov: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Project administration. Anna Bragina: Writing – review & editing, Conceptualization, Project administration, Resources (patient data acquisition). Magomed Abdullaev: Visualization, Formal analysis (statistical analysis). Natalia Druzhinina: Writing – original draft, Conceptualization, Methodology, Validation, Resources (patient data acquisition – blood flow velocity) . Yuliya Rodionova: Validation, Resources (patient data acquisition – CAVI), Data Curation. Rustam Shikhmagomedov: Resources (patient data acquisition – CAVI), Data Curation. Maksim Tyulin: Resources (patient data acquisition, data preparation), Investigation. Valeriy Podzolkov: Conceptualization, Supervision, Resources.

References

[1] I. Ghebrehiwet, N. Zaki, R. Damseh, Rafat and M. Mohamad, Revolutionizing personalized medicine with generative AI: a systematic review. *Artif. Intell. Rev.* **57** (2024) 128.

[2] M. Craig, J.L. Gevertz, I. Kareva and K.P. Wilkie, A practical guide for the generation of model-based virtual clinical trials. *Front. Syst. Biol.* **3** (2023) 5287–5303.

[3] M.-C. Gosselin, E. Neufeld, H. Moser, E. Huber, S. Farcito, L. Gerber, M. Jedensjö, I. Hilber, F. Di Gennaro, B. Lloyd, E. Cherubini, D. Szczerba, W. Kainz and N. Kuster, Development of a new generation of high-resolution anatomical models for medical device evaluation: the Virtual Population 3.0. *Phys. Med. Biol.* **59** (2014) 5287–5303.

[4] R.J. Allen, T.R. Rieger and C.J. Musante, Efficient generation and selection of virtual populations in quantitative systems pharmacology models. *CPT Pharmacometrics Syst. Pharmacol.* **5** (2016) 140–146.

[5] P.H. Charlton, J. Mariscal Harana, S. Vennin, Y. Li, P. Chowienczyk and J. Alastruey, Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes. *Am. J. Phys. Heart Circ. Physiol.* **317** (2019) H1062–H1085.

[6] M.A. Remli, S. Deris, M.S. Mohamad, S. Omatu and J.M. Corchado, An enhanced scatter search with combined opposition-based learning for parameter estimation in large-scale kinetic models of biochemical systems. *Eng. Appl. Artif. Intell.* **62** (2017) 164–180.

[7] H. Saxton, T. Schenkel, I. Halliday and X. Xu, Personalised parameter estimation of the cardiovascular system: Leveraging data assimilation and sensitivity analysis. *J. Computat. Sci.* **74** (2023) 102158.

[8] D. Nolte and C. Bertoglio, Inverse problems in blood flow modeling: a review. *Int. J. Numer. Method Biomed. Eng.* **38** (2022) e3613.

[9] P.I. Frazier, A Tutorial on Bayesian Optimization. (2018) arXiv:1807.02811 [stat.ML].

[10] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams and N. de Freitas, Taking the human out of the loop: a review of bayesian optimization. *Proc. IEEE* **104** (2016) 148–175.

[11] E.A. Wan and R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation. *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. Lake Louise, AB, Canada, (2000) 153–158.

[12] P. Moireau and D. Chapelle, Reduced-order unscented Kalman filtering with application to parameter identification in large-dimensional systems. *ESAIM Control Optim. Calc. Var.* **17** (2011) 380–405.

[13] R. Lal, B. Mohammadi and F. Nicoud, Data assimilation for identification of cardiovascular network characteristics. *Int. J. Numer. Methods Biomed. Eng.* **33** (2017) e2824.

[14] A. Caiazzo, F. Caforio, G. Montecinos, L.O. Muller, P.J. Blanco and E.F. Toro, Assessment of reduced-order unscented Kalman filter for parameter identification in 1-dimensional blood flow models using experimental data. *Int. J. Numer. Methods Biomed. Eng.* **33** (2017) e2843.

[15] A. Giani, R. Micciolo, E. Zoico, G. Mazzali, M. Zamboni and F. Fantin, Cardio-ankle vascular index and aging: differences between CAVI and CAVI0. *J. Clin. Med.* **12** (2023) 6726.

[16] J. Ibata, H. Sasaki, T. Kakimoto, S. Matsuno, M. Nakatani, M. Kobayashi, K. Tatsumi, Y. Nakano, H. Wakasaki, H. Furuta, M. Nishi and K. Nanjo, Cardio-ankle vascular index measures arterial wall stiffness independent of blood pressure. *Diabetes Res. Clin. Pract.* **80** (2008) 265–270.

[17] T. Gamilov and S. Simakov, Blood flow under mechanical stimulations. *Adv. Intell. Syst. Comput.* **1028** (2020) 143–150.

[18] S. Simakov, A. Timofeev, T. Gamilov, P. Kopylov, D. Telyshev and Y. Vassilevski, Analysis of the impact of left ventricular assist devices on the systemic circulation. *Russ. J. Numer. Anal. Math. Model.* **35** (2020) 295–314.

[19] E. Boileau, P. Nithiarasu, P.J. Blanco, L.O. Müller, F.E. Fossan, L.R. Hellevik, W.P. Donders, W. Huberts, M. Willemet and J. Alastruey, A benchmark study of numerical schemes for one-dimensional arterial blood flow modelling. *Int. J. Numer. Method Biomed. Eng.* **31** (2015) e02722.

[20] T. Korakianitis and Y. Shi, Numerical simulation of cardiovascular dynamics with healthy and diseased heart valves. *J. Biomechanics* **39** (2006) 1964–1982.

[21] Y. Sun, B.J. Sjöberg, P. Ask, D. Loyd and B. Wranne, Mathematical model that characterizes transmitral and pulmonary venous flow velocity patterns. *Am. J. Physiol.* **268** (1995) H476–H489.

[22] D.F. Young and F.Y. Tsai, Flow characteristics in models of arterial stenoses. I. Steady flow. *J. Biomechanics* **6** (1973) 395–410.

[23] D.F. Young and F.Y. Tsai, Flow characteristics in models of arterial stenoses. II. Unsteady flow. *J. Biomechanics* **6** (1973) 547–559.

[24] T. Korakianitis and Y. Shi, A concentrated parameter model for the human cardiovascular system including heart valve dynamics and atrioventricular interaction. *Med. Eng. Phys.* **28** (2006) 613–628.

[25] J. Stamler, R. Stamler and J.D. Neaton, Blood pressure, systolic and diastolic, and cardiovascular risks: US population data. *Arch. Intern. Med.* **153** (1993) 598–615

[26] S.S. Hickson, M. Butlin, M. Graves, V. Taviani, A.P. Avolio, C.M. McEniery and I.B. Wilkinson, The relationship of age with regional aortic stiffness and diameter. *JACC Cardiovasc. Imaging* **3** (2010) 1247–1255.

[27] A. Borowska, H. Gao, A. Lazarus and D. Husmeier, Bayesian optimisation for efficient parameter inference in a cardiac mechanics model of the left ventricle, *Int. J. Numer. Method Biomed. Eng.* **38** (2022) e3593.

[28] V.A. Milyagin, I.V. Milyagina, M.A. Purygina and T.A. Osipenkova, Method of Volume Sphygmography on VaSera VS-1500 N Device. Methodical Recommendations. Smolensk (2014) 30.

[29] Yu.A. Vasyuk, S.V. Ivanova, E.L. Shkolnik, Yu.V. Kotovskaya, V.A. Milyagin, V.E. Oleynikov, Ya.A. Orlova, A.N. Sumin, A.A. Baranov, S.A. Boytsov, A.S. Galyavich, Zh.D. Kobalava, O.V. Kozhevnikova, A.O. Konradi, Yu.M. Lopatin, V.Yu. Mareev, D.S. Novikova, R.G. Oganov, A.N. Rogoza, O.P. Rotar, N.V. Sergatskaya and V.V. Skibitsky, Consensus of Russian experts on the evaluation of arterial stiffness in clinical practice. *Cardiovasc. Ther. Prev.* **15** (2016) 4–19.

[30] T. Okura, S. Watanabe, M. Kurata, S. Manabe, M. Koresawa, J. Irita, D. Enomoto, K. Miyoshi, T. Fukuoka and J. Higaki, Relationship between cardio-ankle vascular index (CAVI) and carotid atherosclerosis in patients with essential hypertension. *Hypertens. Res.* **30** (2007) 335–340.

[31] T. Kumagai, T. Kasai, M. Kato, R. Naito, K.I. Maeno, S. Kasagi, F. Kawana, S. Ishiwata and K. Narui, Establishment of the cardio-ankle vascular index in patients with obstructive sleep apnea. *Chest* **136** (2009) 779–786.