Statistical parameter estimation and signal classification in cardiovascular diagnosis

S. Bernhard[1], K. Al Zoukra[1] & C. Schütte[1]
[1]*Fachbereich Mathematik, Freie Universität Berlin, Germany*

Abstract

Medical technology has seen impressive success in the past decades, generating novel clinical data at an unexpected rate. Even though numerous physiological models have been developed, their clinical application is limited. The major reason for this lies in the difficulty of finding and interpreting the model parameters, because most problems are ill-posed and do not have unique solutions. On the one hand the reason for this lies in the information deficit of the data, which is the result of finite measurement precision and contamination by artifacts and noise and on the other hand on data mining procedures that cannot sufficiently treat the statistical nature of the data. Within this work we introduce a population based parameter estimation method that is able to reveal structural parameters that can be used for patient-specific modeling. In contrast to traditional approaches this method produces a distribution of physiologically interpretable models defined by patient-specific parameters and model states. On the basis of these models we identify disease specific classes that correspond to clinical diagnoses, which enable a probabilistic assessment of human health condition on the basis of a broad patient population. In an ongoing work this technique is used to identify arterial stenosis and aneurisms from anomalous patterns in parameter space. We think that the information-based approach will provide a useful link between mathematical models and clinical diagnoses and that it will become a constituent in medicine in near future.
Keywords: statistical cardiovascular system model, cardiovascular system identification, multi-channel measurement, state-space model, parameter estimation, Bayesian signal classification, patient-specific diagnosis.

WIT Transactions on Biomedicine and Health, Vol 15, © 2011 WIT Press
www.witpress.com, ISSN 1743-3525 (on-line)
doi:10.2495/EHR110401

1 Introduction

Due to the high morbidity and mortality arising with cardiovascular disorders, an efficient and highly specific diagnosis is of paramount importance for the patient. There exist a variety of different diagnostic approaches that consider several factors and symptoms leading to a diagnosis by an exclusion principle. However, the symptoms are not always well-defined and differ from patient to patient, in other words - not every patient has distinct symptoms that allow a clear disease specific assignment. This demands for a high degree of expertise from the physician and sometimes makes the diagnosis tedious and even misleading diagnoses are the result. Therefore, the attempt of using physiological models combined with knowledge and experience of experts collected in databases to support the diagnosis process by computational methods seems a reasonable approach. The solution to a non-trivial classification task like this is valuable to improve diagnosis.

Typically traditional signal processing techniques are used to extract therapeutically relevant information like for e.g. the heart rate, oxygen saturation and the cardiac output from clinical data. Even though the information content within the data has grown continuously, the number of reliable procedures for feature extraction had not.

On the one hand the difficulty of information extraction of richer data sets into improved therapies lies in the deterministic view immanent in current mathematical models of physiological processes [1]. On the other hand there is a lack of sufficient therapeutic strategies that can handle improved diagnostic information. The former problem is a result of limited statistical data integration into the model, the latter problem is dependent on the time required to transfer the results into novel therapeutic strategies.

According to studies regarding the needs of clinical applications [2-4], we combine physical and physiological aspects of pathological conditions in the cardiovascular system with patient-specific simulations that are based on non-invasively accessible data [1].

Within this study we use two basic approaches to describe the cardiovascular system dynamics on a statistical basis: (i) measurement-based parameter estimation and (ii) model-based prediction and classification methods.

In the first approach, the desired parameters are estimated from physiological measurements using statistical inference techniques. In the current experimental setup the data is either obtained by a series of non-invasive multi-channel measurements from a specific sub-population of patients (e.g. healthy/diseased) or generated by invasive measurements from a fluid-dynamical cardiovascular simulator that models normal and pathological flow conditions. This is a very expensive approach, since it requires a distinct subset of specific patients or the construction of a fluid-dynamical system that can be used to simulate realistic conditions in the vasculature. Furthermore the measurements have to be acquired according to a standard operating procedure (SOP) and analyzed in a predefined statistically setting.

In contrast to this, the second model-based approach is inexpensive and easier to perform, since the system dynamics is obtained by computational methods. However, the model-based approach has some limitations in representing the actual physical state of a real system, since the underlying processes are complex and the number of model parameters is large. Moreover, the choice of parameters in most physiological models is crucial for interpretation and prediction. Even though the sensitivity analysis approach allows to determine specific system features and to identify the critical parameters to which the system is most sensitive, parameter determination is difficult. Not to mention that the interpretation of estimated parameters can only occur on a statistical basis that is based on a broad patient population.

Within this work we combine two approaches, statistical inference methods that are suitable to determine model parameters and states observed in a broad patient population, and Bayesian classification to extract hidden information like for e.g. classes of diseased cardiovascular states that describe the health condition of specific subgroups. This combined approach leads to a novel statistical interpretation of cardiovascular system that optimally uses information of specific sub-populations for diagnostic purposes. The technique provides the potential to develop highly individualized therapeutic strategies – a benefit for the patient.

2 Statistical model of the cardiovascular system

In contrast to traditional cardiovascular system models we are interested in models that include an important feature: randomness. Randomness is characterized by a non-deterministic behavior that can be described by probability theory and the concept of random variables. Randomized processes are described by either time discrete or continuous functions (e.g. probability density functions and distribution functions). For most real world examples state space models are sufficient to describe the underlying dynamics. Having constructed such a model, the time discrete behavior of the system can be simulated and desired measures can be evaluated.

2.1 Complexity reduction

Within the model building process the complexity of the cardiovascular system is a critical problem that requires a tradeoff between accuracy of representing the true dynamic behavior and the ability to solve the model equations in reasonable amount of time.

There exist a variety of approaches to model and solve complex systems. Within this work we follow a very pragmatic decomposition/aggregation approach described in [1]. Here the complexity problem is treated by the construction of sub-models (channels) through the definition of specific interface conditions. The basic idea is to decompose the complex structure of the vasculature into a set of simpler sub-models being solved separately. The solutions are then combined to obtain an aggregate solution for the actual model.

Of particular interest are simple lumped-parameter models (also known as zero dimensional or Windkessel models) that describe the transport of blood distribution within the vasculature [5]. These models where developed to provide answers to important questions in cardiovascular physiology that have abscond intuitive understanding [1]. However until now only very basic parameters can be estimated for individual patients, not to mention that there is no reasonable chance for a population-based interpretation.

In the following example we build a sub-model for a structure of the carotis bifurcation given in [6] and derive the corresponding state-space model.

2.2 Windkessel model for cardiovascular fluid flow

Dynamical systems are generally described by ordinary differential equations (ODEs) in canonical form. It has been shown, that lumped parameter models are reasonable approximations to describe the fluid flow in most elements of the cardiovascular system. Following [5, 7, 8] each segment of the arterial system can be modeled by a 3-element Windkessel electrical circuit analogue (see Figure 1).

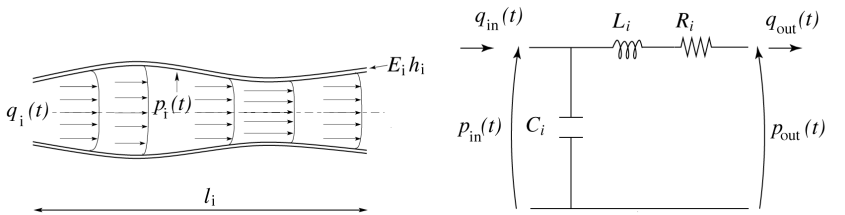

Figure 1: Arterial vessel segment (left) and corresponding three-element Windkessel electric analogue (right).

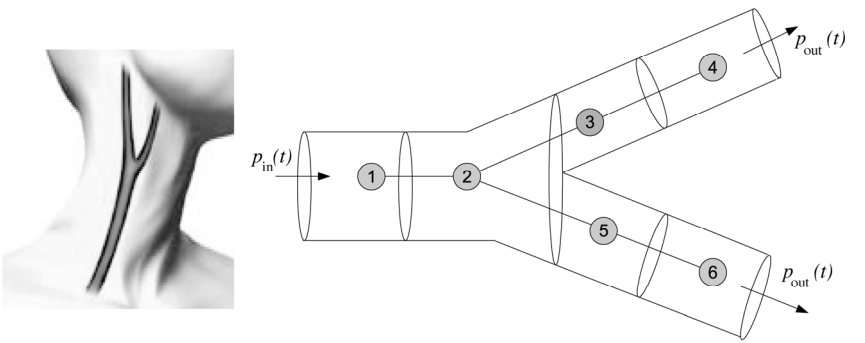

Figure 2: Human carotis bifurcation and the corresponding simplified network structure for N_s=6.

In the example model discussed here, every arterial segment $i = 1, \ldots, N_s$ is represented by a electrical circuit consisting of a resistance and inductance in

series and a capacitor in parallel. The analogy relates electric current and voltage to arterial blood flow q and pressure p, respectively. The electrical resistances R_i correspond to the viscos flow resistance, the inductances L_i account for the blood inertia forces and the arterial compliances, i.e. the elasticity of the vessel walls, are described by electrical capacitors C_i. The peripheral resistances Z_j for the number of terminating ends N_t of the network account for the viscous flow resistance and compliance in the microcirculation. Within the example of the carotis bifurcation shown in fig. 2, the number of segments is chosen to be $N_s = 6$ and the number of terminals is $N_t = 2$.

In this fundamental form the cardiovascular system dynamics can be represented by a set of $n = 2N_s$ coupled ordinary differential equations of first order that depend on the unknown dynamical parameters λ.

$$\begin{aligned} \dot{x} &= f(\mathrm{t}, x, \lambda) \qquad x \in R^n \qquad \mathrm{t} \in [T_0, T_0 + \mathrm{T}] \\ x(T_0) &= x_0 \end{aligned} \tag{1}$$

The solution can be given analytically in terms of exponential functions and sine waves, if the right-hand side function f is linear in x. Generally the initial values are also unknown parameters, so that the vector of unknown parameters θ is:

$$\theta = (\lambda, x_0) \in R^{\lambda+n} \tag{2}$$

In analogy to Kirchhoff's current and voltage law we obtain a system of n coupled ordinary differential equations for the pressure and flow:

$$\begin{aligned} \dot{q}_1 &= \frac{p_{in} - R_1 q_1 - p_1}{L_1} & \dot{q}_4 &= \frac{p_3 - R_4 q_4 - p_4}{L_4} \\ \dot{p}_1 &= \frac{q_1 - q_2}{C_1} & \dot{p}_4 &= \frac{q_4 - (p_4 - p_{out})/Z_1}{C_4} \\ \dot{q}_2 &= \frac{p_1 - R_2 q_2 - p_2}{L_2} & \dot{q}_5 &= \frac{p_2 - R_5 q_5 - p_5}{L_5} \\ \dot{p}_2 &= \frac{q_2 - q_3 - q_5}{C_2} & \dot{p}_5 &= \frac{q_5 - q_6}{C_5} \\ \dot{q}_3 &= \frac{p_2 - R_3 q_3 - p_3}{L_3} & \dot{q}_6 &= \frac{p_5 - R_6 q_6 - p_6}{L_6} \\ \dot{p}_3 &= \frac{q_3 - q_4}{C_3} & \dot{p}_6 &= \frac{q_6 - (p_6 - p_{out})/Z_2}{C_6} \end{aligned} \tag{3}$$

According to [7], the parameters of the electric analogue circuit are determined from the structural and physiological parameters. Assuming Hagen-Poiseuille flow the electrical parameters become:

$$R_i = \frac{8\nu l_i}{\pi r_i^4}, \quad \begin{matrix} L_i \\ = \end{matrix} \frac{\rho l_i}{\pi r_i^2}, \quad \begin{matrix} C_i \\ = \end{matrix} \frac{2\pi r_i^2 l_i}{E_i h_i}. \tag{4}$$

Here every vessel segment i is specified by its length l_i, its radius r_i, the wall thickness h_i and the Youngs modulus E_i. The blood is characterized by the density and viscosity of $\rho = 1050\ kg/m^3$ and $\nu = 4 * 10^{-4}\ Pa\ s$, respectively.

Having defined the parameters $\lambda = (R_i, C_i, L_i, Z_j), \forall i, j$ and the pressure at the inlet $p_{in}(t)$ and outlet p_{out}, the system can be numerically solved for the nodal pressures p_i and flows q_i.

2.3 State space model

The state space representation is a useful notation to describe the dynamics in arterial networks. Besides the explicit description of the measurement process, statistical processes, like the measurement noise, can be described in a very simple and efficient way. In state space form the dynamical system is written in terms of input and observation vectors and the state space variables. It is expressed as a first order differential state equation and the observation equation:

$$\dot{x}_t = \mathbf{A}x_{t-1} + \mathbf{B}\mathrm{u}_t + \mathrm{w}_t \quad (5\ a)$$

$$y_t = \mathbf{C}x_t + \mathbf{D}\mathrm{u}_t + \mathrm{v}_t \quad (5\ b)$$

Here x_t is the vector of state space variables, u_t the input vector and y_t the observation vector. The dynamics of the system is described by the state dynamics matrix $\boldsymbol{A} \in \boldsymbol{M}\ (n \times n)$. The input matrix $\boldsymbol{B} \in \boldsymbol{M}\ (n \times i)$ specifies the time dependency of the in- and outflow as boundary values, the observation matrix $\boldsymbol{C} \in \boldsymbol{M}\ (m \times n)$ defines the observation locations, with the number of observations, *m,* and the input to observation matrix $\boldsymbol{D} \in \boldsymbol{M}\ (m \times i)$ quantifies the influence of the input vectors to the observation vectors (see fig. 3). In the current setting we use synthetic data as measurement data, so we neglect the noise matrix w_t and v_t in the state and observation equation respectively. The noise terms can be included with minimal effort if real data is available.

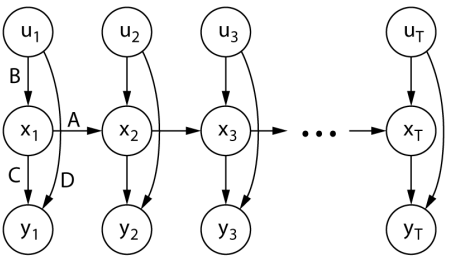

Figure 3: Time discrete state space system with matrix relations between the state, input and observation vectors.

The state equation (5 a) relates the state vector *x* at a time *t* to the unknown parameters θ, while the observation equation (5 b) relates the state vector to the measured data *y*. The state vector x_t contains the flow and pressure functions at

all network locations, whereas the observation vector y_t contains the flow and pressure at selected nodes *i,* to which measurement time series are available.

We assume that the input pressure p_{in} is a given function of time and that the output pressure p_{out} is known. In the hemodynamic system output pressure corresponds to the mean venous pressure, which is almost constant and has mean values of about *15 mmHg.* For these two inputs $a = 2$, the input vector is:

$$\mathrm{u(t)}:=\begin{pmatrix}\mathrm{p}_{in}(\mathrm{t})\\ \mathrm{p}_{\mathrm{out}}(\mathrm{t})\end{pmatrix} \qquad \mathrm{u} \in R^{a}. \tag{6}$$

For m available observations like for e.g. p_i and q_i at nodes 3 and 5 in the network, the observation vector is

$$y(t):=\begin{pmatrix}q_3(t)\\ p_3(t)\\ q_5(t)\\ p_5(t)\end{pmatrix} \qquad y \in R^{m}. \tag{7}$$

Defining the state vector $x:=(\mathrm{q_1},\mathrm{p_1},\mathrm{q_2},\mathrm{p_2},\mathrm{q_3},\mathrm{p_3},\mathrm{q_4},\mathrm{p_4},\mathrm{q_5},\mathrm{p_5},\mathrm{q_6},\mathrm{p_6})^T$ leads to a states space system for the carotis bifurcation described in the previous section that is denoted by

$$A=\begin{pmatrix}
-\frac{R_1}{L_1} & -\frac{1}{L_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\frac{1}{C_1} & 0 & -\frac{1}{C_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & \frac{1}{L_2} & -\frac{R_2}{L_2} & -\frac{1}{L_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \frac{1}{C_2} & 0 & -\frac{1}{C_2} & 0 & 0 & 0 & -\frac{1}{C_2} & 0 & 0 & 0\\
0 & 0 & 0 & \frac{1}{L_3} & -\frac{R_3}{L_3} & -\frac{1}{L_3} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & \frac{1}{C_3} & 0 & -\frac{1}{C_3} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & \frac{1}{L_4} & -\frac{R_4}{L_4} & -\frac{1}{L_4} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_4} & -\frac{1}{Z_1C_4} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \frac{1}{L_5} & 0 & 0 & 0 & 0 & -\frac{R_5}{L_5} & -\frac{1}{L_5} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_5} & 0 & -\frac{1}{C_5} & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_6} & -\frac{R_6}{L_6} & -\frac{1}{L_6}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_6} & -\frac{1}{Z_2C_6}
\end{pmatrix},$$

$$B=\begin{pmatrix}
\frac{1}{L_1} & 0\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & \frac{1}{Z_1C_4}\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & \frac{1}{Z_2C_6}
\end{pmatrix},$$

$$C = \begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix},$$

and

$$D = \begin{pmatrix} 0 & 0\\ 0 & 0 \end{pmatrix}.$$

The time discrete state space equations, the parameters θ, and the input vector u_t describe the time evolution of the system in form of the state space variables x_t. In practical applications most of the system parameters are unknown. Further more the vector of state variables of the cardiovascular system cannot be measured directly. In the following we propose a parameter estimation method based on the measurements defined in the observation vector y_t. Due to the fact that the measurements are incomplete ($m \neq N_s$) the inverse problem is ill-posed.

3 Parameter estimation in the cardiovascular system

There are several methods to estimate unknown system parameters from time discrete measurements. We assume that we have measured the observation vector $y_i(t)$ for discrete times, $t = 0, \dots, T$, by a multi-channel measurement at locations i in the vascular network. The aim is to estimate the parameters and hidden signals from the measurements, in other words, we seek a solution to the hemodynamic inverse problem that was proposed to have infinite number of solutions [9].

A well-established approach to determine the parameters is the maximum likelihood estimator. It is defined as the vector that minimizes the measurement likelihood L, given θ:

$$\hat{\theta} = arg \max_{\theta} L(y|\theta). \tag{8}$$

In other words, without assumptions about the parameter to estimate, one chooses the value that makes the output most likely. This maximization problem can be solved even if the data is high dimensional, incomplete and noisy [10]. Due to the fact that the ODE model is not based on a probability model we may assume that the data is normally distributed around the deterministic solution, so that the likelihood is defined in terms of a least square distance. Then the distance function of squared residuals between the measured data and the model trajectory is equivalent to the maximum-likelihood problem. In other words, minimizing the distance function

$$\Xi(d, y) = \sum_{i=1}^{m} \sum_{k=1}^{K} (d_i(t_k) - y_i(t_k, \theta))^2, \tag{9}$$

where d_i are data points for the locations i, and $y_i(t_k, \theta)$ is the solution of the dynamical system at times $t_1, \dots, t_K$ is equivalent to maximizing the likelihood function

$$L(d|\theta) = \prod_{i=1}^{m} \prod_{k=1}^{K} \frac{e^{-(1/2)\left(d_i(t_k) - y_i(t_k,\theta)\right)^T \Sigma^{-1} \left(d_i(t_k) - y_i(t_k,\theta)\right)}}{(2\pi)^{m/2} |\Sigma|^{1/2}}, \tag{10}$$

where the scalar matrix Σ is a diagonal matrix with equal diagonal entries [11]. The maximization of the measurement likelihood may be obtained by two different approaches: application of (i) the expectation maximization (EM) algorithm or (ii) approximate Bayesian computation (ABC) techniques.

In (i) the likelihood function is maximized by the EM algorithm, which iteratively increases the likelihood function. This maximum typically is the global one, if some good initial estimate $\theta(0)$ is available. The initial estimate can gained from a classical parameter estimation procedure (see section 3.1).

In (ii) the likelihood function is maximized by finding a sufficient approximation $\mathbb{P}(\theta|y^*)$ to the posterior probability distribution $\mathbb{P}(\theta|d)$ in Bayes' formula

$$\mathbb{P}(\theta|d) = \frac{\mathbb{P}[d|\theta]\mathbb{P}[\theta]}{\sum_W \mathbb{P}[d|\theta]\mathbb{P}[\theta]}. \tag{11}$$

The ABC algorithm is sample based, i.e. it generates simulation data $y^*(\theta^*)$ for parameter vectors θ^* drawn from the prior probability distribution $\mathbb{P}[\theta]$. The parameters θ^* are accepted if a distance measure $\delta(d, y^*) \leq \varepsilon$ is sufficiently small, then $\mathbb{P}(\theta|d) \cong \mathbb{P}(\theta|y^*)$ and the measurement likelihood consequently is $\mathbb{P}[d|\theta] \cong \mathbb{P}[y^*|\theta]$.

3.1 Solution of the optimization problem

To obtain realistic initial parameters we formulate an optimization problem that includes additional knowledge about the parameters as equality or inequality constraints. In the cardiovascular system all parameters are non-negative and smaller than a parameter specific upper bound, i.e. $\theta_l \leq \theta \leq \theta_h$. Additionally we decompose the optimization problem into s optimization sub-problems for θ_s, and apply the transfer function relations between the interfaces of the sub-models as constraints.

Within this nonlinear optimization problem we seek the vector of parameters θ_s for each sub-model such that:

$$\begin{aligned} &\text{Minimize} && \Xi(d_s, y_s) \\ &\text{subject to:} && \\ & && \theta_l \leq \theta_s \leq \theta_h \end{aligned} \tag{12}$$

Due to the nonlinearity iterative algorithms must be used to find a solution. For efficient optimization, at least first derivatives with respect to the parameters (sensitivities) should be provided. According to the large number of variables we decided to use two constraint optimization algorithms: (i) a weighted variant of Levenberg-Marquardt nonlinear least squares algorithm using parameter sensitivities to control the step size (SENSOP) [12] and (ii) a non-linear steepest-descent algorithm (NLSD). The details of the optimization results are discussed in our previous work [1].

4 Bayesian classification as concept in cardiovascular diagnosis

Bayesian signal classification is concerned with two tasks: Firstly the identification of specific sub-populations on the basis of training data sets containing observations whose sub-population is known a priori and secondly the determination of the affiliations of observations where the identity of the sub-population is unknown. The use of Bayesian classifiers in cardiovascular diagnoses is discussed in [1] in more detail. Basically the classifier learns the signal distribution of instances of specific diseases in a sub-population to determine the classification probability of unknown data sets.

In order to outline the classification problem, we start with the assumption that we have a sequence of measurements $d_1, \ldots, d_M$ for M different patients that in some appropriate sense form a sampling of a specific patient population. Classification now means to find out which of the M patients belong to certain classes $\mathcal{C}_1, \ldots, \mathcal{C}_W$. We further define the class of healthy patients by $\mathcal{C}_0$, and the classes of patients that have specific diseases by $\mathcal{C}_w$, in the following referred to as class w. The classification probability of a patient b in class w is thus the conditional probability of being in class w given the sequence of observed signals, $\mathbb{P}(b \in \mathcal{C}_w | d_{1:M})$. Using Bayes' formula, this probability can be computed from

$$\mathbb{P}(b \in \mathcal{C}_w | d_{1:M}) = \frac{\mathbb{P}[d_{1:M} | b \in \mathcal{C}_w]\mathbb{P}[b \in \mathcal{C}_w]}{\sum_W \mathbb{P}[d_{1:M} | b \in \mathcal{C}_w]\mathbb{P}[b \in \mathcal{C}_w]}. \tag{13}$$

Here $\mathbb{P}(b \in \mathcal{C}_w)$ denotes the prior probability of patients of class w and $\mathbb{P}[d_{1:M} | b \in \mathcal{C}_w]$ denotes the probability of measuring the signal from a patient of class w. While the former probability is a classical prior, the latter probability has to be estimated algorithmically from the sequence of observations $d_{1:M}$. Thus the classification algorithm must perform two different estimations simultaneously:

1. Density estimation: Estimate the probability $\mathbb{P}[d_{1:M} | b \in \mathcal{C}_w]$ that a certain signal is observed from a patient of class w.
2. Classification: Find the hidden information whether the signal of a patient, d_b, belongs to class w.

The combination of these two tasks in the sense of a joint likelihood optimization again leads to the EM algorithm. In other words, in every step of the EM iteration the density is estimated before the classification probabilities are evaluated. The iteration again converges if we choose appropriate initial values and results in the optimal densities and classifications based on the available observation. Consequently, the accuracy of the results increases when data is reintegrated.

The aim of the above classification algorithm is to classify measured data into classes with common properties – i.e. with a relation to specific diseases. These classes are then used to classify unknown data measured at patients with unknown diagnosis by means of fuzzy probabilities. As obvious from the above description the Bayesian classification method comprises two steps:

Firstly, in the learning phase, the measurement data and known relationships to cardiovascular diseases (training data) are used to train the priors and densities needed in the EM algorithm. The probability distributions over these training datasets are learned from examples verified by the gold standard (valid diagnosis), thus allowing the generation of new relationships that describe disease specific classes. The gold standard relationships can include any properties of time series related or unrelated to a particular disease. In general, these gold standards are formed by data obtained for a sub-population of patients with known diagnosis.

Secondly, in the prediction phase, the classification probabilities are predicted by the classification algorithm based on newly acquired signals (testing data) without available diagnosis. The prediction is generally based on a network indicating how likely the observation of measurements fits to a specific class. If one considers this network as a connection matrix, it is just a collection of fuzzy like measures, each representing a probability of functional relationship between the measurement and the class. According to the classification probabilities the procedure provides a diagnostic hint about the existence of afore characterized diseases. In other words, the algorithm sets up a series of hypothesis, that are based on the prior information of a sub-population obtained in clinical observations, to classify the health condition of the patient.

The algorithmic classification procedure is as follows:

1. Use training data $d_{1:M}^{\dagger}$ to determine optimal parameters θ, priors and density estimation via the EM or ABC algorithm.
2. Classify testing data $d_{1:M}^{\ddagger}$ via (fuzzy) classification probabilities $\mathbb{P}^{M}(d_{1:M}^{\dagger})$.
3. Integrate testing data into training data set and re-optimize parameters.

In order to realize this approach for cardiovascular diseases we will have to train the algorithm on a significantly large population of patients, which is the next challenge we will have to face. Then the statistical classification becomes a method that allows us to identify cardiovascular diseases in an early state that are followed by therapeutic intervention convenient for individual patients. In contrast to other methods that determine a set of selected parameters with pretended relevance for diagnosis, the classification method automatically selects and quantifies all relevant parameters to prove a series of proposed diseases in the fashion of differential diagnosis.

5 Conclusion and outlook

Within this work we have outlined the solution of the constraint hemodynamic inverse problem. The proposed statistical inference approach provides various advantages including quantitative parameter estimates, determination of confidence intervals and error estimates given incomplete and noisy measurement data. Further more the classification algorithm quantifies all disease specific model parameters in terms of classes in the fashion of differential diagnosis. These techniques are proposed to be the basis in patient

specific diagnoses, because they provide a statistical framework for the description of the cardiovascular system allowing improved therapeutic interventions for individual patients.

Although the interdisciplinary challenges involved in the ongoing project are daunting, it is important to recognize the potential gains for cardiovascular diagnosis. However, up to now the progress has been inhibited by the lack of a broad data basis of non-invasive hemodynamic measurements, advanced inverse modeling tools and databases for large-scale data integration and classification. Nevertheless we are sure that the new modeling techniques will find several applications in cardiovascular medicine. The progress will depend on the level of support from funding and industry and the interest of clinicians. There are signs suggesting strong interest from all areas.

References

[1] Bernhard, S., Al Zoukra, A. and Schütte, C., From non-invasive hemodynamic measurements towards patient-specific cardiovascular diagnosis (Chapter) to appear in: *Quality Assurance in Healthcare Service Delivery, Nursing and Personalized Medicine: Technologies and Processes*, ed. A. Lazakidou & A. Daskalaki. IGI-Global, 2011.

[2] Capova, K., Blazek, V., Cap, I., & Buckuliakova, L., Physiological Fluid Systems Modelling For Non-Invasive Investigation. *Advances in Electrical and Electronical Engineering,* pp. 38-42, 2002.

[3] Xiao, X., Ozawa, E. T., Huang, Y. & Kamm, R. D., Model-Based Assessment of Cardiovascular Health from Noninvasive Measurements. *Annals of Biomedical Engineering,* **30**, pp. 612-623, 2002.

[4] Wessel, N., Voss, A., Malberg, H., Ziehmann, Ch., Voss, H. U., Schirdewan, A., Meyerfeldt, U., & Kurths, J., Nonlinear analysis of complex phenomena in cardiological data. *Herzschrittmachertherapie und Elektrophysiologie,* **11(3)**, pp. 159-173, 2000.

[5] Olufsen, M. S., & Nadim, A., On Deriving Lumped Models for Blood Flow and Pressure in the Systemic Arteries. *Math. Biosc. & Eng.,* **1(1)**, pp. 61-80, 2004.

[6] Neil, W., Parametric geometry exploration of the human carotid artery bifurcation, *Journal of Biomechanics* **40**, pp. 2484-2491, 2007.

[7] Alastruey, J., Parker, K. H., Peiró, J., & Sherwin, S. J., Lumped Parameter Models for 1-D Blood Flow Simulations: Effect on Pulse Waves and Parameter Estimation. *Communications in Computational Physics,* **4(2)**, pp. 317-336, 2008.

[8] Quateroni, A., Ragni, S., & Veneziani, A.,Coupling between lumped and distributed models for blood flow problems. *Computational Visual Sciences,* **4(2)**, pp. 111-124, 2001.

[9] Quick, C. M., Young, W. L., & Noordergraaf, A., Infinite number of solutions to the hemodynamic inverse problem. *American Journal of Physiology - Heart and Circulatory Physiology,* **280(4)**, pp. 1472-1479, 2001.

[10] Hartley, H. O., Maximum likelihood estimation from incomplete data. *Biometrics*, **14**, pp. 174-194, 1958.
[11] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M.P.H., Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface,* **6**, pp. 187-202, 2009.
[12] Chan I.S., Goldstein A.A., & Bassingthwaighte J.B., SENSOP: A derivative-free solver for non-linear least squares with sensitivity scaling. *Ann Biomed Eng.*, **21**, pp. 621-631, 1993.