

Efficient Probabilistic Model Personalization Integrating Uncertainty on Data and Parameters: Application to Eikonal-Diffusion Models in Cardiac Electrophysiology

Ender Konukoglu^{a,*}, Jatin Relan^b, Ulas Cilingir^c, Bjoern H. Menze^d, Phani Chinchapatnam^e, Amir Jadidi^f, Hubert Cochet^f, M el eze Hocini^f, Herv e Delingette^b, Pierre Jais^f, Michel Ha issaguerre^f, Nicholas Ayache^b, Maxime Sermesant^{b,e,**}

^aMicrosoft Research Cambridge, UK

^bINRIA, Asclepios team, France

^cUniversity of Cambridge, Engineering Department, UK

^dMIT, CSAIL, USA

^eKing's College London, St Thomas' Hospital, UK

^fBordeaux University Hospital, France

Abstract

Biophysical models are increasingly used for medical applications at the organ scale. However, model predictions are rarely associated with a confidence measure although there are important sources of uncertainty in computational physiology methods. For instance, the sparsity and noise of the clinical data used to adjust the model parameters (personalization), and the difficulty in modeling accurately soft tissue physiology. The recent theoretical progresses in stochastic models make their use computationally tractable, but there is still a challenge in estimating patient-specific parameters with such models. In this work we propose an efficient Bayesian inference method for model personalization using polynomial chaos and compressed sensing. This method makes Bayesian inference feasible in real 3D modeling problems. We demonstrate our method on cardiac electrophysiology. We first present validation results on synthetic data, then we apply the proposed method to clinical data. We demonstrate how this can help in quantifying the impact of the data characteristics on the personalization (and thus prediction) results. Described method can be beneficial for the clinical use of personalized models as it explicitly takes into account the uncertainties on the data and the model parameters while still enabling simulations that can be used to optimize treatment. Such uncertainty handling can be pivotal for the proper use of modeling as a clinical tool, because there is a crucial requirement to know the confidence one can have in personalized models.

Keywords: Probabilistic Inverse problems, Bayesian Inference, PDE models, Polynomial Chaos, Spectral Representation, Cardiac Electrophysiology, Model Personalization, Eikonal Models, Compressed Sensing

1. Introduction

Modeling of physiological processes and pathologies is important for understanding the underlying phenomena and also for designing better clinical systems for diagnosis and therapy planning. The research community has focused on the mathematical models for a long time aiming to create more realistic and multi-scale models, Murray (2002). Recently, thanks to the advances in clinical imaging and mapping, modeling research broadened its interest and started to focus on a different type of problem, *model personalization*. Model personalization is the adaptation of a generic model to a specific patient based on the available clinical data. This is crucial both for transferring mathematical models into clinical practice and for validating models with patient data. In this article we concentrate on the estimation of patient specific parameters.

*Principal corresponding author. Address: 7 JJ Thomson Avenue, Cambridge, CB3 0FB, UK

**Corresponding author. Address: 2004 Route des Lucioles, 06902, Sophia Antipolis, France

Email addresses: ender.konukoglu@live.com (Ender Konukoglu), maxime.sermesant@inria.fr (Maxime Sermesant)

Parameter estimation from clinical data is a challenging task due to the nature of the clinical observations and the complexity of the mathematical models, even in simplified cases, He and Keyes (2007). The challenge mainly arises due to the following reasons:

- i. *Sparsity* of the data: The temporal (frequency) and spatial (resolution) sampling of the acquired clinical observations are much lower than the corresponding scales of the underlying processes. For example, during a cycle of a healthy heart the electrical wave passes through all the cells in the cardiac tissue. The cardiac mappings on the other hand, samples at best a few hundred points at 1 Hz on the boundary of the heart.
- ii. *Uncertainty* on the data: The measurements obtained from clinical data are often gathered through invasive and/or indirect procedures. For example, cardiac electrophysiology mapping can either be achieved by catheterization or indirectly derived from body surface potentials. In the first case the contact points are on a moving heart and can slide due the beating of the heart and the breathing motion. Furthermore, the matching between the measurement points and the computational mesh used in the simulations - which is extracted from medical images - is ambiguous. Both the motion and the ambiguous matching are sources of uncertainty on the measured values.
- iii. *Complexity* of the model: The mathematical models describe complex dynamics resulting in nonlinear formulations. The interactions between the model parameters, the model variables and the observations are therefore not trivial. Furthermore, models often include large number of parameters that must be estimated. The parameter estimation problems for such models are in general hard and often results in non-unique solutions in the sense that observations could have resulted via a range of possible parameter sets instead of a single one.
- iv. *Assumptions* of the model: Modeling is by definition a simplification, thus the degree of realism of the model compared to the observations is limited by the assumptions that make the model valid. Finding the right level of details in a model in order to be able to both adjust it to the data and achieve meaningful predictions is not simple.

These challenges undermine the uniqueness of the parameter estimation problem. Under these settings one can expect multiple parameter sets satisfying the observations with the same accuracy. Methods for estimating patient-specific parameters should take these challenges into account in providing not only point estimates for the parameters but also ranges of possible parameters and confidence margins.

Most physiological models used today are continuum models based on systems of partial differential equations (PDEs), Murray (2002). Parameter estimation problems are well studied for PDE systems, Tarantola (2005). Approaches can be coarsely classified as deterministic or probabilistic, Kaipio and Somersalo (2005). Deterministic approaches are based on minimizing a cost function that measures the difference between the observations and the model solutions. The solution giving the minimum value then serves as the estimate. The most common techniques involve variational (e.g. adjoint model) or sequential (e.g. Kalman filter) methods, Tarantola (2005); Voss and Timmer (2004). These methods are very popular due to the tractability of the estimation and the well studied control theory. On the other hand, it is well known that these methods are prone to local minima and the sensitivity analysis can become cumbersome in complex models where analytical derivations are not possible. In the case of sparse and uncertain observations these drawbacks become important issues.

Probabilistic or statistical methods are based on the theories of Bayesian inference and numerical schemes to generate the posterior distributions for the parameters. The problem of parameter estimation is cast as a probabilistic model defining a joint probability distribution between observations, parameters and variables. The estimation task is then defined as computing the posterior distribution for the parameters given the observations. These distributions provide much richer information than the point estimates obtained in deterministic methods. They contain information about how well all possible sets of model parameters fit the observations. They can be used to obtain point estimates (e.g. Maximum-A-Posteriori (MAP) estimates) as well as confidences on these estimates. Another advantage of probabilistic methods is that they are less sensitive to local minima since they estimate the whole distribution.

The main disadvantage of the probabilistic methods is their computational cost. In general posterior distributions cannot be computed analytically and must be approximated numerically. The conventional numerical method for this task is sampling through Markov-Chain Monte-Carlo (MCMC) methods, Kaipio and Somersalo (2005). MCMC methods is based on simulating the model with many number of different parameter sets and approximating the posterior distribution using these samples. Obtaining an accurate estimate using probabilistic methods requires a large number of sample simulations. When a single model simulation can take a considerable amount of time on

a high performance computer, as it can be the case in organ-scale biophysical models, running thousands of them becomes impractical.

Recent advances in uncertainty quantification (UQ) and numerical methods for stochastic PDEs provide solutions for reducing the computational cost of Monte-Carlo simulations of PDE systems, Ghanem and Doostan (2006); Xiu and Karniadakis (2002). These methods are based on polynomial chaos (PC) expansions of random variables and spectral representations of the equations. Through spectral approximations, the model simulations can be executed using only linear combinations of spectral basis functions rather than solving the PDE. This yields faster Monte-Carlo simulations and faster uncertainty quantification, which have significant impact on probabilistic methods for sensitivity analysis and parameter estimation, Ghanem and Doostan (2006); Xiu and Karniadakis (2002); Marzouk et al. (2007); Marzouk and Najm (2009); Marzouk and Xiu (2009); Ma and Zabarar (2009); Doostan et al. (2009); Xiu (2007, 2009).

In this article we propose and demonstrate an efficient Bayesian inference method for the parameter estimation problems involving large (> 10) number of parameters. In order to demonstrate the interactions between uncertainty on the data and the model parameters we focus on the parameter estimation problem in Eikonal-Diffusion (ED) type models for cardiac electrophysiology (EP). We formulate the Bayesian inference problem for the ED model and demonstrate the inference method. In Section 2 we briefly discuss the cardiac electrophysiology model used in this work. Section 3 explains the probabilistic formulation and the Bayesian inference problem. In Section 4 we explain the inference algorithm (mathematical details are given as appendices). The proposed methodology is demonstrated both on synthetic and clinical data. Sections 5 and 6 present these results. Finally we conclude this article with discussions.

2. Eikonal-Diffusion Model for Cardiac Electrophysiology

Among the different models describing the electrical activity in the cardiac tissue, Fink et al. (2011); Clayton et al. (2011), here we chose to focus on the Eikonal-Diffusion (ED) model Colli Franzone et al. (1990), as it allows a compact formulation while still including important nonlinearities. Moreover, by using the anisotropic Fast-Marching Method, Konukoglu et al. (2007), Eikonal models can be solved rapidly which makes them suitable for clinical applications, Chinchapatnam et al. (2008), and treatment simulations, Pernod et al. (2011).

2.1. The Eikonal-Diffusion Model

The ED model describes the depolarization time $T(\mathbf{x})$ in the cardiac tissue $\mathbf{x} \in \Omega \setminus \Omega_E$, through the nonlinear PDE

$$c_0 D(\mathbf{x}) \left(\sqrt{\nabla T^t \mathbf{M}(\mathbf{x}) \nabla T} \right) - \nabla \cdot (D(\mathbf{x}) \mathbf{M}(\mathbf{x}) \nabla T) = \tau, \quad (1)$$

$$T(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega_E, \quad (2)$$

where the boundary conditions depict the initialization of the excitation (electric wave) at location Ω_E , *the onset location* (which is a set of points in the discrete setting). In this equation the superscript t denotes the transposition, ∇ is the gradient operator, c_0 is the dimensionless constant related to the cell membrane and τ is the cell membrane time constant. $D(\mathbf{x})$ is the square of the local tissue space constant and is related to the conductivity of the tissue. The tensor quantity $\mathbf{M}(\mathbf{x})$ incorporates the local fiber orientation to represent the anisotropy in conduction velocity of the cardiac tissue (see Tomlinson et al. (2002) for further details on the Eikonal-Diffusion model parameters).

The parameters of interest in this equation are the conductivity $D(\mathbf{x})$ and the onset location Ω_E . $D(\mathbf{x})$ is a spatially varying conductivity function with specific structure. Its value in the cardiac myocardium volume is taken as a global parameter. But on the endocardium surface, it varies spatially due to the existence of the Purkinje network extremities. This network creates important fast conduction pathways but cannot be observed using medical imaging. The goal of the inverse problem is to find the global myocardium conductivity and the locally varying conductivity on the endocardium. We represent $D(\mathbf{x})$ with a radial basis function approximation as often done in the literature, Nobile et al. (2008); Ma and Zabarar (2009).

$$D(\mathbf{x}) = \begin{cases} D_0 & \text{if } \mathbf{x} \in \Omega_{myo}, \\ \frac{1}{Z(\mathbf{x})} \sum_{m=1}^M D_m \exp\left(-\left(\frac{\|\mathbf{x}-\mathbf{x}_m\|^2}{\sigma^2}\right)\right) & \text{if } \mathbf{x} \in \Omega_{endo}, \end{cases} \quad (3)$$

where Ω_{myo} is the myocardial tissue, Ω_{endo} is the endocardial surface, $Z(\mathbf{x}) = \sum_{m=1}^M \exp(-(\|\mathbf{x}-\mathbf{x}_m\|^2/\sigma^2))$ is the normalization term, D_0 is the muscle tissue conductivity, \mathbf{x}_m are pre-defined radial basis centers, $\|\cdot\|$ is the geodesic distance on the endocardium surface and D_m 's are scalar weights. Based on this representation the inverse problem for $D(\mathbf{x})$ is defined as estimating the vector $\bar{D} = [D_0, \dots, D_M]$. Alternative representations can also be used, Marzouk and Najm (2009); Marzouk and Xiu (2009). Ideally one would represent $D(\mathbf{x})$ as a different value at each discretization point in the computation domain. This however, yields a high dimensional inverse problem creating a high computational cost, Kaipio and Somersalo (2005).

We model the onset location Ω_E as a point on the septum wall on the endocardium. Although a more realistic model would formulate this as a set of different junctions we currently approximate this with a single point and endocardial areas of fast conduction. Based on this assumption we represent the location of the Ω_E with two spatial coordinates on the septum of the endocardial surface x_E, y_E .

2.2. The Observation Model

The clinically available observations are depolarization times on the left ventricular endocardium (from catheter-based mapping) and depolarization times on the epicardium, from catheters or computed from the inverse problem of electrocardiography. These observations are mathematically represented as $\tilde{T}(\mathbf{x}) = T(\mathbf{x}) + \epsilon(\mathbf{x})$ for $\{\mathbf{x}_i\}_1^N \subset \partial\Omega$ on the boundary of the heart, where $\epsilon(\mathbf{x})$ is the observation noise and $\{\tilde{T}(\mathbf{x}_i)\}_1^N$ forms the observation vector. The observation noise approximates the following sources of uncertainty:

- During the acquisition procedure the catheters position changes due to the beating of the heart and the breathing motion.
- The position of each observation point is projected onto the computation domain Ω which is extracted from medical images taken either before or after the acquisition. Therefore, these points do not necessarily correspond to each other.
- The depolarization times are estimated from the electrical potentials measured by the catheter. This estimation is also prone to noise, Steinhaus (1989).

An in depth analysis of each of these sources of error would be very beneficial however, this is beyond this presented work. Here we approximate the contribution of all of these sources with the observation noise $\epsilon(\mathbf{x})$.

3. Probabilistic Modeling of Parameter Estimation for the ED Model

Different methods have been proposed to adjust cardiac electrophysiology models, including for instance genetic algorithms for the fast conduction system, Camara et al. (2010), Maximum A Posteriori state estimation, Wang et al. (2011), or in similar conditions using a deterministic approach and a trust-region minimizations, Chinchapatnam et al. (2008); Relan et al. (2011). Up to the best of our knowledge none of these approaches use a full probabilistic treatment of the problem.

The probabilistic approach taken here formulates each model parameter as a random variable. The distributions for these random variables reflect the possible ranges of the parameters as well as their expected value. Under this randomness the relationship between the model parameters and the model variables T can be written as a joint probability distribution. This distribution can be decomposed based on the conditional dependencies of the PDE, i.e. different parameters leading to different solutions. The joint distribution for the ED model described above can be written as

$$p(\tilde{T}, T, D(\mathbf{x}), c_0, \tau, \mathbf{M}(\mathbf{x}), \Omega_E) = p(T|D(\mathbf{x}), c_0, \tau, \mathbf{M}(\mathbf{x}), \Omega_E)p(\tilde{T}|T)p(D(\mathbf{x}), c_0, \tau, \mathbf{M}(\mathbf{x}))p(\Omega_E) \quad (4)$$

where the $\tilde{T} = \{\tilde{T}(\mathbf{x}_1), \tilde{T}(\mathbf{x}_2), \dots, \tilde{T}(\mathbf{x}_N)\}$ is the set of observations, $p(\Omega_E)$ is the prior distribution for the onset location and $p(D(\mathbf{x}), c_0, \tau, \mathbf{M}(\mathbf{x}))$ is the prior distribution for the model parameters. On the right hand side of this equation, the first term describes the probability that for a given parameter set the simulated depolarization times will be T , the second term describes the probability of observing \tilde{T} when the simulation result is given as T and the last two terms are the prior distributions for the parameters. Here we only focus on $D(\mathbf{x})$ and Ω_E therefore we can drop the other parameters in the formulation. Furthermore, we assume that the observation noise $\epsilon(\mathbf{x})$ is an independent random variable for each point - a common assumption for the probabilistic models, Kaipio and Somersalo (2005). Considering the independent nature of contact mappings this assumption can be justified for the catheter based observations.

This assumption allows us to state that for a given location \mathbf{x} if $T(\mathbf{x})$ is given then $\tilde{T}(\mathbf{x})$ does not depend on the observations taken from other locations \mathbf{y} (however, we would also like to note that this independence assumption could also be changed without greatly affecting the overall method.). Based on this conditional independence and using the approximation given in Eq. (3) the formulation becomes

$$p(\tilde{T}, T, \bar{D}, x_E, y_E) = \prod_{i=1}^N \left\{ p(\tilde{T}(\mathbf{x}_i) | T(\mathbf{x}_i)) p(T(\mathbf{x}_i) | \bar{D}, x_E, y_E) \right\} p(\bar{D}) p(x_E, y_E). \quad (5)$$

In the case of parameter estimation for PDEs, given a boundary condition and a set of parameters the forward simulation often provides a unique solution. This is also the case for the ED model. This is integrated as an additional simplification: $p(T(\mathbf{x}) | \bar{D}, x_E, y_E) = \delta_{T(\mathbf{x}) | \bar{D}, x_E, y_E}$, where we use $T(\mathbf{x} | \bar{D}, x_E, y_E)$ to represent the solution of the model Eq. (1) given a specific parameter set and δ is the Dirac's delta function. We can then compute the posterior distribution for the parameters, $p(\bar{D}, x_E, y_E, | \tilde{T})$, given an observation vector:

$$p(\bar{D}, x_E, y_E | \tilde{T}) = \frac{\prod_{i=1}^N p(\tilde{T}(\mathbf{x}_i) | T(\mathbf{x}_i | \bar{D}, x_E, y_E)) p(\bar{D}) p(x_E, y_E)}{\int_{\bar{D}, x_E, y_E} \prod_{i=1}^N p(\tilde{T}(\mathbf{x}_i) | T(\mathbf{x}_i | \bar{D}, x_E, y_E)) p(\bar{D}) p(x_E, y_E) d\bar{D} dx_E dy_E}, \quad (6)$$

where the integral in the denominator of the posterior distribution is over all the possible parameter sets. One can see the advantage of probabilistic methods in this formulation. The posterior distribution $p(\bar{D}, x_E, y_E, | \tilde{T})$ not only provides us with the best possible parameter set that fits the observation, but also contains information on how all other possible parameter sets fit the observations.

In principle the prior distributions for the conductivity values and the onset locations can be obtained through populations statistics. In order to take into account our lack of knowledge of such statistics – which up to the best of our knowledge do not exist yet – here the prior distributions are set to uniform distributions, namely the uninformative priors

$$\begin{aligned} D_m &\sim \mathcal{U}(D_m^a, D_m^b), \quad m = 0, \dots, M \\ x_E &\sim \mathcal{U}(x_E^a, x_E^b), \quad y_E \sim \mathcal{U}(y_E^a, y_E^b). \end{aligned} \quad (7)$$

The values for the upper and lower limits of the distributions will be specified in the experiments section.

4. Fast Bayesian Inference

Bayesian inference corresponds to computing the posterior distribution $p(\bar{D}, x_E, y_E, | \tilde{T})$ for a given observation set. Equation (6) admits an analytical solution only if the term $T(\mathbf{x} | \bar{D}, x_E, y_E)$ can be written analytically. For most models, as is the case for this one, solutions cannot be written analytically. They are found by numerically solving the PDE systems. The common approach to compute $p(\bar{D}, x_E, y_E | \tilde{T})$ is to use sampling methods such as MCMC, Kaipio and Somersalo (2005). MCMC methods evaluate the numerator in Equation 6 sequentially at many random points and approximate the posterior distributions using these samples (we would like to note here that the independence assumption on $\epsilon(\mathbf{x})$ only determines the method of computation of this numerator and it can easily be interchanged with another observation model). One can immediately see the bottleneck here: evaluation of the model, $T(\mathbf{x} | \bar{D}, x_E, y_E)$. Simulating models often takes time and this is the reason of the high computational cost of probabilistic methods.

Spectral representations based on polynomial chaos expansions efficiently approximate $T(\mathbf{x} | \bar{D}, x_E, y_E)$ and yield substantial speed ups, Ghanem and Doostan (2006); Xiu (2009); Marzouk and Najm (2009). It has already been used for the sensitivity analysis of the direct problem of electrocardiography at the macroscopic scale, where number of parameters is small and the underlying equation is linear, Geneser et al. (2008). The spectral approximation for the probabilistic formulation of the ED model given above can be given as

$$T(\mathbf{x} | \bar{D}, x_E, y_E) \approx \sum_{i=0}^P T_i(\mathbf{x}) \Phi_i(\bar{D}, x_E, y_E), \quad (8)$$

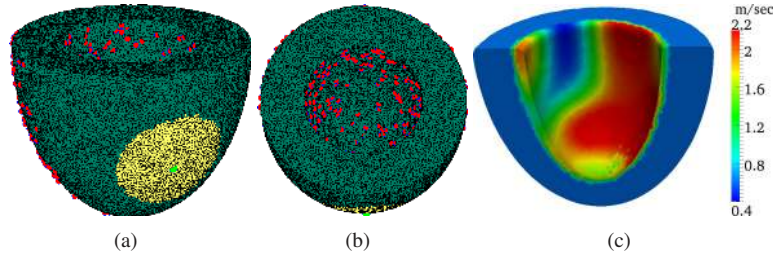


Figure 1: Geometric setting: (a) Simulation domain with observation points (in red). Green point: onset location for experiments 1. Yellow disc: region of unknown onset location in experiment 2, (b) Figure (a) shown in a different view point, (c) Ground truth varying conductivity. Muscle conductivity: $D_0 = 0.06 \text{ mm}^2$ (the conduction velocity $v = c_0 \sqrt{D_0}/\tau = 0.6 \text{ m/s}$). Endocardial conductivity set using Eq. (3) with $M = 17$.

where $T_i(\mathbf{x})$ are so-called the spectral basis, Φ_i are multivariate orthonormal Legendre polynomials of increasing order and P is the total number of spectral components involved in the approximation. The summation on the right hand side approximates the solution of the model equation. One can see that the computation of this summation can be executed much faster than numerically solving the model equation because it only involves addition of functions. Therefore, in the Bayesian inference this approximation can be used to compute the samples necessary to build the posterior distribution. For the ease of explanation we do not provide the mathematical details at this stage please refer to Appendix A for these details.

The computation of the spectral basis $T_i(\mathbf{x})$ for high number of parameters is in general a computationally expensive task. However, for the ED model, compressed sensing can be used to determine the basis in a fast manner. Such a technique has been demonstrated for the linear diffusion equation in Doostan et al. (2009). As a result, one can compute the basis functions easily and in an efficient manner (see Appendix B for a brief explanation). Using these two improvements together, we can then apply Bayesian inference on such non-linear 3D models in a tractable way. We will provide computational times in the experiments section to elaborate on this advantage.

5. Validation on Synthetic Data

In this section we demonstrate the proposed Bayesian inference method on synthetic data where the ground truth for the local cardiac muscle conductivity is known. Using the synthetic problem we aim to demonstrate the influence of the data uncertainty on the confidences of the estimated parameters. We use an idealised left ventricle geometry (see Fig. 1). The computation domain is discretized as a Cartesian grid in 3D, its size is $15 \times 15 \times 10 \text{ cm}^3$ and the resolution of each voxel is $1 \times 1 \times 1 \text{ mm}^3$. The ED model Eq. (1) is solved within the volume enclosed by the mesh shown in Fig. 1a (mesh is only shown for visualization purposes and the computations were done on a Cartesian grid for this synthetic study) using a Fast-Marching approach. The parameters – which are assumed to be known and constant for the presented experiment – are $c_0 = 2.5$, $\tau = 1$ and \mathbf{M} is here an isotropic tensor.

5.1. Problem Setting

We generated a ground truth conductivity function using Eq. (3) and simulated data similar to what can be acquired in a clinical environment: a set of endocardial depolarization times as measured with a catheter, Dong et al. (2006); Richmond et al. (2008), and a set of epicardial depolarization times, as measured either by a catheter, Sacher et al. (2010) or from body surface potential mapping, Ramanathan et al. (2004). Using the simulated data we infer the conductivity function under different settings. The presented experiments encompass some of the main difficulties of this inverse problem: the spatial variation of conductivity on the endocardium from very low values (scars) to very high values (Purkinje extremities), and the uncertainty on the onset location.

In order to exclude the onset from the data, we simulate a Left Bundle Branch Block, therefore we initiate the electrical onset on the right ventricular septum (green point in Fig 1a). In the first experiment this location is known, and in the second experiment we relax this assumption and perform a joint inference of both the onset location and the conductivity function. The yellow region shown in Fig. 1a denotes possible locations of the onset point for this second experiment.

The ground truth conductivity common to both of the presented experiments is shown in Fig. 1c. In this figure and all the following figures, we show the conduction velocity $v = c_0 \sqrt{\bar{D}}/\tau$ rather than the D value due to its clinical relevance. For completeness we will provide numerical values for both D and v . To model the scars and the existence of Purkinje extremities, the conductivity on the endocardium varies from very low values ($0.0016 \text{ mm}^2 \sim 0.1 \text{ m/s}$) to very high values ($0.9 \text{ mm}^2 \sim 2.4 \text{ m/s}$), as seen in Fig. 1c. We construct the conductivity function as given in Eq. (3). We used an 18 dimensional \bar{D} , with $M = 17$ radial basis functions defined on $\{\mathbf{x}_m\}$ which are equally spaced on the surface and the ground truth muscle conductivity is taken as $D_0 = 0.06 \text{ mm}^2 \sim 0.6 \text{ m/s}$.

Using the model Eq. (1) and the parameter settings described above, we simulated data at different observation points on the endocardium (200 points) and the epicardium (300 points). The epicardial data were assumed to be only on the opposite side of the onset point since the region around the onset would normally be located inside the right ventricle and not accessible. In Fig. 1a and b we show these observation points as red spheres on the surface. As explained in Section 2.2 the observed depolarization times can be uncertain due to several factors. In order to integrate this into our synthetic experiment we added Gaussian noise to the simulated depolarization times with variance of 25 ms (simulated depolarization times varied between 35 ms to 125 ms at the observation points).

5.2. Probabilistic Model Details and Computational Aspects

The parameters of the presented method are the limits of the uniform distributions given in Equation 7 and the number of spectral components P to be included in the spectral approximation. For the \bar{D} parameter the ranges for the synthetic problem are set as $(D_0^a, D_0^b) = (0.01, 0.09) \text{ mm}^2 \sim (0.25, 0.75) \text{ m/s}$ and $(D_m^a, D_m^b) = (0.001, 0.9) \text{ mm}^2 \sim (0.08, 2.4) \text{ m/s}$ for $m = 1, \dots, 17$. On the other hand for the onset location Ω_E the prior distribution is defined as the uniform distribution defined over the yellow region shown in Fig. 1a, which is a disc of 6 cm in diameter on the surface. Using this probabilistic setting we computed the spectral basis $\{T_i(\mathbf{x})\}_{i=0}^P$ on all the observation points. In our experiments we set $P = 10626$ containing all 20 dimensional multivariate Legendre polynomials of maximum order 4.

In terms of computational times the construction of the spectral basis is the most expensive part. However, as explained in Section 4 compressed sensing framework speeds up this construction. In the synthetic example the computation of the $\{T_i(\mathbf{x})\}_{i=0}^P$ took about 2.5 hours for all the 500 observation points.

5.3. Estimation Experiments

We present two different experiments using the setting explained above: 1) inference of \bar{D} using observations taken on endocardial and epicardial surfaces assuming the knowledge of Ω_E and 2) joint inference of \bar{D} and Ω_E . All the inferences are done using the formulation and the method explained in Sections 3 and 4. Once the spectral basis is constructed the inference for each of the experiments took *less than 5 minutes* for both of the experiments. The MCMC algorithm we used is described in Girolami and Calderhead (2011). The posterior distributions were approximated through generating 15000 samples where first 5000 samples were discarded as burn-out. As it is not feasible to visualize the high dimensional posterior distributions $p(\bar{D}|\tilde{T})$ and $p(\bar{D}, x_E, y_E|\tilde{T})$ in the experiments we present the mean of the posterior distribution and the marginal standard deviations on these estimates.

In Fig. 2a and b we show the mean of the posterior distribution of the conductivity and the posterior marginal standard deviation for \bar{D} obtained from the first experiment. We observe that the estimate is very similar to the ground truth both quantitatively and qualitatively (the RMS error between the simulated depolarization times using the posterior mean and the measurements is 1.7 msec). The standard deviation is low showing the higher confidence in the estimation. The posterior average and the standard deviation of muscle conductivity estimate in this case is $0.061 \pm 0.001 \text{ mm}^2 \sim 0.62 \pm 0.01 \text{ m/s}$. In Figure 2c we show the standard deviation of the estimation from the top view. Notice that the standard deviation is higher for the regions where there are no epicardial observation points. This is as expected since the sparsity of observation in that side of the synthetic left ventricle increases the uncertainty in the estimated parameters. Furthermore, we also notice that the standard deviation increases towards the top of the left ventricle where the number of observation points is smaller, in other words the observations are sparser.

For the second experiment we defined the onset location as a random parameter of the system as well. In the first experiment the onset location was centered in the allowable region. In order to show that the proposed methodology indeed works for the whole region we changed the ground truth onset location to the white sphere shown in Fig. 2f. In Fig. 2d we show the mean of the posterior distribution of the conductivity estimate of the conductivity. Observe

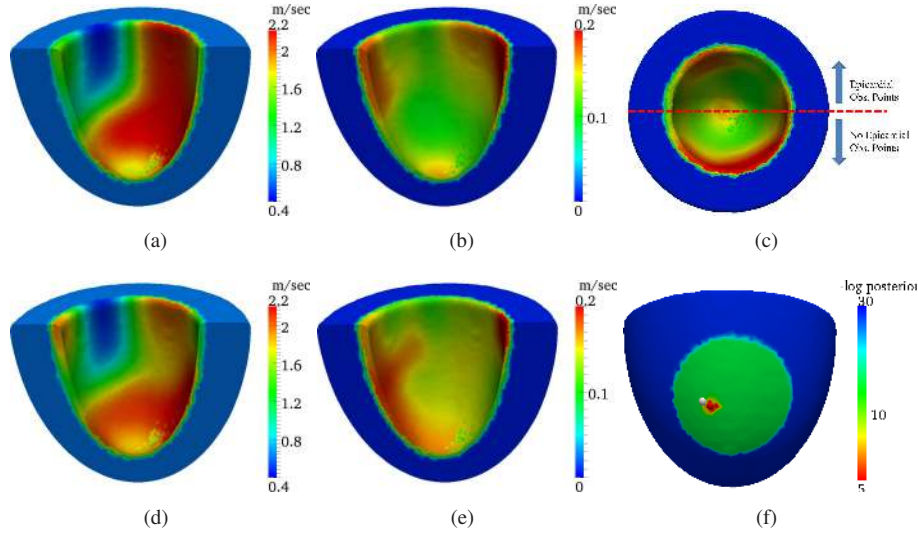


Figure 2: Mean estimate of the conductivity function (a) for experiment 1, and (d) for experiment 2. Standard deviation of the conductivity estimate (b) for experiment 1 and (e) for experiment 2. (c) Distribution of the marginal standard deviation (inverse of the confidence in estimation) with respect to the location of epicardial observation points. Notice that where there are more observations the standard deviation on the estimation is lower. (f) $-\log$ posterior distribution of the onset location given the observations. The white point denotes the ground truth and the red point denotes the mean of the posterior distribution.

that again the locations of the low and high conduction regions are well captured (the RMS error on the depolarization times is 3.5 msec). Moreover the standard deviation for this estimate shown in Fig. 2e demonstrates that again the difference between the ground truth and the estimate is captured within the standard deviation of the estimate. Comparing Figures 2b and e we notice that the standard deviation on the estimated conductivity function increases slightly. This is as expected since the posterior distributions for the parameters captures all possible parameter sets that fit the (noisy) observations. Therefore, by introducing the onset location in the parameter estimation we enlarge the possible parameter sets that would have a close fit with the observations. This increases the standard deviation leading to a decrease our confidence. The muscle conductivity in this case is estimated as $0.062 \pm 0.001 \text{ mm}^2 \sim 0.62 \pm 0.01 \text{ m/s}$. In Fig. 2f we show the mean of the posterior distribution of the onset location (small red sphere) and the negative log posterior distribution $-\log(p(\Omega_E|\tilde{T}))$. The mean of the posterior distribution well estimates the ground truth onset location (4.4 mm apart).

6. Evaluation on Clinical Data

In order to evaluate this method on clinical data, we used contact mapping of the myocardium, acquired during an electrophysiology study for ventricular tachycardia radio-frequency ablation. This study comprises both endocardial and epicardial mapping, which is rare in clinical routine but enables to precisely evaluate the impact of both measurements on the estimated parameters and its associated confidence. We used either the 20-polar 5-spline PentaRay I (Biosense-Webster, CA) or the 30-polar 5-spline PentaRay II (Biosense-Webster, CA) catheters for high density endo- and epicardial mapping. The PentaRay II catheter enables simultaneous multi-electrode contact mapping of up to 60 mm^2 myocardial surface, when fully expanded. This novel technique of high density map acquisition enables rapid high-density mapping. The mean duration for acquiring a high-density map is about 5-10 minutes. The EP measurements were processed to estimate the depolarization times, Steinhaus (1989) at 958 observation points (both endocardial and epicardial) which were then used to adjust the model.

Magnetic Resonance Images (MRI) of the patient were also acquired, both anatomical images and late-enhancement ones, in order to locate scar tissue (this patient has a large aneurysm in the left ventricle). The 3D anatomical image was segmented to generate the myocardial mesh and late-enhancement MRI was segmented and used to label scar tissue and peri-infarct area (border zone) in the mesh. The extra-cellular potential measurements obtained with the

catheters were then globally manually registered to this mesh, and projected on the mesh surface. In Fig. 3 we present the mesh extracted from the MR images and the observation points projected onto the mesh. The color of the points corresponds to the distance between the location of the observation point during the procedure and the mesh surface extracted from the MR image. Notice that these distances can be quite large and contribute to the uncertainty on the observation values.

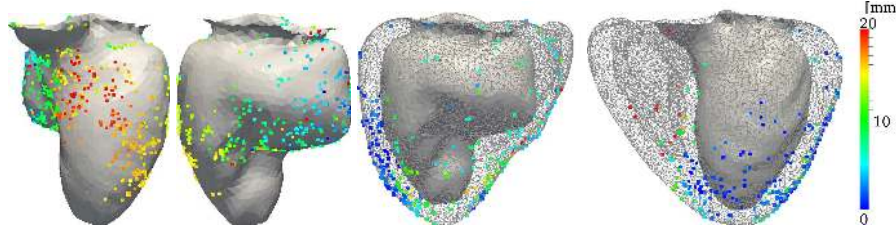


Figure 3: The computational mesh extracted from the MR images and the observation points. On the left two views of the endocardial surface is shown with the observation points plotted on top. On the right two views of the epicardial surface is also drawn (as a wire mesh). The observation points on the two surfaces are depicted in the images where the colors correspond to the *distance each point was projected to the mesh surface*.

6.1. Probabilistic Model Details and Computational Aspects

The parameters we focus on are: the onset location, the fast conduction system on the endocardium, and the myocardium / scar / border zone conductivities. As in the synthetic case, we determined 17 anatomical regions on the endocardial surface and use the formulation given in Equation 3 to describe the endocardial conductivity function. This (together with the myocardium/scar/border zone conductivities) in total yields 20 conductivity parameters to estimate. The conductivity parameters were again formulated as uniformly distributed random variables with ranges: i) myocardium and border zone - $(0.1, 1.6) \text{ mm}^2 \sim (0.25, 1) \text{ m/s}$, ii) scar - $(0.03, 0.4) \text{ mm}^2 \sim (0.15, 0.5) \text{ m/s}$ and iii) each endocardial parameter - $(0.1, 2.0) \text{ mm}^2 \sim (0.8, 3.5) \text{ m/s}$. In addition to these we formulated the uncertainty on the onset location as a uniform distribution on the septum (the yellow regions shown in Figures 4c, 5c and 6c) yielding two extra parameters, 22 parameters in total to estimate. For this 22 dimensional problem we used $P = 14950$ containing 22 dimensional multivariate Legendre polynomials of maximum order 4 similar to the synthetic examples shown before.

The ED model was solved on a tetrahedral mesh with an average element volume of 10 mm^3 . The parameters $\tau = 1 \text{ msec}$ and $c_0 = 2.5$ were used as in the synthetic experiments. The tensors \mathbf{M} in this case were generated using an analytical model of cardiac fiber orientations where the anisotropy ratio between the principal direction and the others is 0.15 (which corresponds to a conduction velocity three times faster along the fibre). The total time elapsed for the computation of $\{T_i(\mathbf{x})\}_{i=0}^P$ took 4.2 hours for all the observation points. In the phase of inference we used 30 000 samples (after the first 5000 samples that were discarded as burn-out) to compute the posterior distributions for the parameters under investigation. Each inference process took 15 minutes on average.

6.2. The Noise Model

Different sources of uncertainty acting on the cardiac mapping data explained above are: i) the displacement of the catheter contact points on the cardiac surface due to the motion of the heart, ii) the estimation of the depolarization times from the EP measurements and iii) the registration and the projection of the observation points onto the cardiac surface. All these sources affect the interaction between the model and the observed data and therefore, influence the parameter estimation problem. In depth understanding of these effects are crucial to model this interaction accurately. As a preliminary step in this work we used a simplified uncertainty model for the observations. The contributions of the first two sources are modeled as Gaussian noise with homogeneous variance over all the observation points, $\epsilon_1(\mathbf{x}) \sim \mathcal{N}(0, 400 \text{ msec}^2)$. The contribution of the third source is modeled as a Gaussian noise with inhomogeneous variance. It is modeled to depend on the distance between the mesh surface (extracted from the MR images) and the location of the observation points (computed during the catheterization). This projection distance is available for each mapping point, see Figure 3. The inhomogeneous variance is formulated in a very simple form as $\epsilon_2(\mathbf{x}) \sim$

$\mathcal{N}(0, \rho_{proj}(\mathbf{x})^2 msec^2)$ where $\rho_{proj}(\mathbf{x})$ is the projection distance for the observation point at \mathbf{x} . The overall noise model for each observation point is then given as the Gaussian noise $\epsilon(\mathbf{x}) = \epsilon_1(\mathbf{x}) + \epsilon_2(\mathbf{x}) \sim \mathcal{N}(0, 400 + \rho_{proj}(\mathbf{x})^2 msec^2)$.

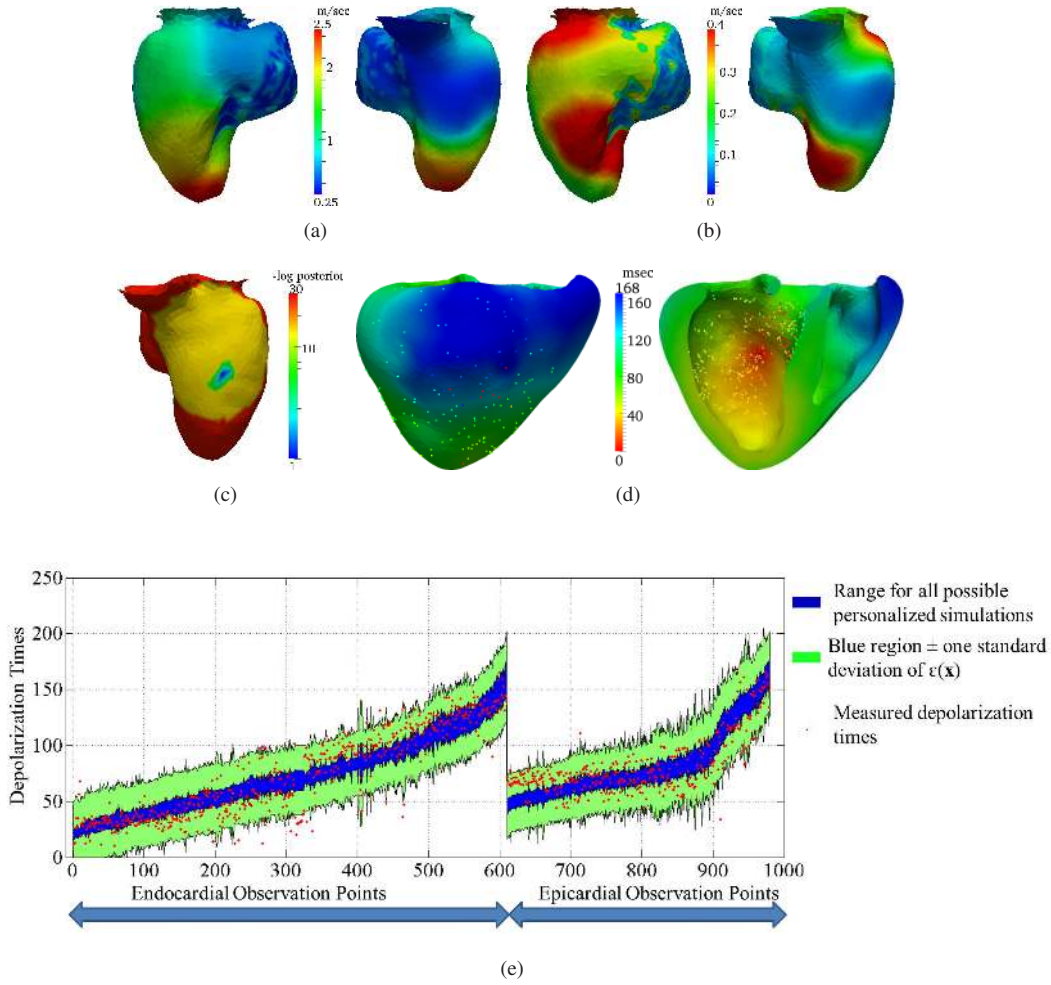


Figure 4: Estimation results for the experiment 1 - inference using all the measured points. (a) Mean estimates for the endocardial conductivity function shown from two views. (b) Marginal standard deviations on the estimated conductivity function. (c) $-\log(p(\Omega_E|\bar{T}))$ - the $-\log$ posterior for the onset location. (d) Depolarization time isochrones simulated with the mean estimates where color points represent the measurements and their depolarization times. (e) Comparison between the measured depolarization times and the probabilistic model fit. Range for all possible personalized simulations is computed through the posterior distributions for the parameters. The green region is the observation noise model added to the possible simulation range.

6.3. Estimation Experiments

Using the proposed Bayesian formulation and the inference method three different experiments were conducted. In all the experiments posterior distributions for the conductivity values and the onset location are estimated as explained in Section 6.1. In the first experiment all the 958 observation points (both endocardial and the epicardial) were used in the inference to compute the posterior distributions for the parameters. In the second experiment only the observation points on the epicardial surface (observation points shown in the right two most surfaces in Figure 3) were used in the inference. The computed posterior distribution for the parameters is then used to predict depolarization times for the points on the endocardial surface. These predictions are given as ranges of possible personalized model simulations based on the posterior distributions of the model parameters. The ranges are obtained using the

	Myocardium	Aneurysm	Border Zone
Experiment 1: Inference using all the observations	0.88 ± 0.05	0.31 ± 0.05	0.58 ± 0.08
Experiment 2: Inference using the epicardial observations	0.95 ± 0.03	0.24 ± 0.05	0.31 ± 0.09
Experiment 3: Inference using the endocardial observations	0.48 ± 0.14	0.41 ± 0.07	0.58 ± 0.10

Table 1: Estimated conductivities (mean and standard deviation in m/s) for myocardium, aneurysm and border zone for the three different experiments.

probabilistic framework and the fast sampling scheme presented in Sections 3 and 4. Comparisons with the measured values - which were not used during the estimation - is presented. In the third experiment, only the endocardial observation points were used in the inference and the predictions are done for the epicardial depolarization times. For ease of visualization, we present most of the results on surfaces, but everything was computed on the volumetric myocardial mesh.

Results on Experiment 1. Figure 4 and the first row of Table 1 present the results obtained for the first experiment. The posterior distribution for the endocardial conductivity function is presented by its mean (Fig. 4a) and the marginal standard deviation (Fig. 4b). The estimated volumetric conduction velocities are also given in Table 1, which are in the range of the values found in the literature. Notice that the standard deviations for the endocardial conductivities are higher in some regions. Comparing this with the locations of the observation points in the Figure 3, we see that the standard deviation is higher where the density of observation points is lower. This is intuitive as in the sense that our confidence on the estimated values is directly linked to the density of observation points. This shows that the presented probabilistic method well captures the interaction between the sparsity of the observations and the estimation of model parameters. Figure 4c shows that the posterior distribution for the onset location which is concentrated in a region. This is also in accordance with the density of the observation points.

Two different types of comparison between the model fit and the measured values are shown. In Figure 4d the isochrones on the mesh show the depolarization times obtained with the mean estimates of the parameters. On the same figure the observation points are shown where their color corresponds to the measured depolarization times. The match between the mean model (model using the mean estimates for the parameters) and the measured values is 16 *msec* in RMS error. The mean model however, does not show the rich information context of the probabilistic formulation. The link between the uncertainties in the estimated parameters and the model simulations yields a possible range of model fits. Figure 4e presents a comparison between the measured depolarization times and the possible range of model fits computed through the parameter posterior distributions. The blue region shows the range of all possible simulated (with personalized parameters) depolarization times, i.e. the range of all the possible values of $T(\mathbf{x}|\bar{D}, \Omega_E)$ computed using $p(\bar{D}, \Omega_E|\tilde{T})$. The green region shows the addition of the noise model on this possible range, i.e. \pm one standard deviation of the observation noise model $\epsilon(\mathbf{x})$ to the possible range for $T(\mathbf{x}|\bar{D}, \Omega_E)$. Notice that the width of the blue region is rather narrow showing confidence in the model fit. Furthermore, most of the observation points lie within the blue region while almost all of the remaining points lie within the green region.

Results on Experiment 2. The results for this experiment are presented in Figure 5 and the second row in Table 1. The surface meshes shown in (a) demonstrate that the mean of the posterior distribution for the parameters is structurally similar to the ones estimated in the first experiment with slight differences. The estimated volumetric conduction velocities given in Table 1 also show a similar trend. The standard deviations shown in Fig. 5b on the other hand are significantly higher compared to Fig. 4b. This suggests that only epicardial observations are not enough to confidently estimate endocardial conductivities. This is not surprising as the observations in this case are not directly taken from the endocardial surface. For most of the epicardial points the depolarization times are influenced by different regions on the endocardium as well as the myocardial conductivity. Therefore, we do not have a direct observations that can confidently estimate the endocardial conductivity function. On the other hand, we notice the low standard deviations for the volumetric conductivities given in Table 1 showing that the epicardial observations are enough to confidently estimate the volumetric conduction velocities. Regarding the numerical differences between the estimated values, we will come back to this point while presenting the results for Experiment 3.

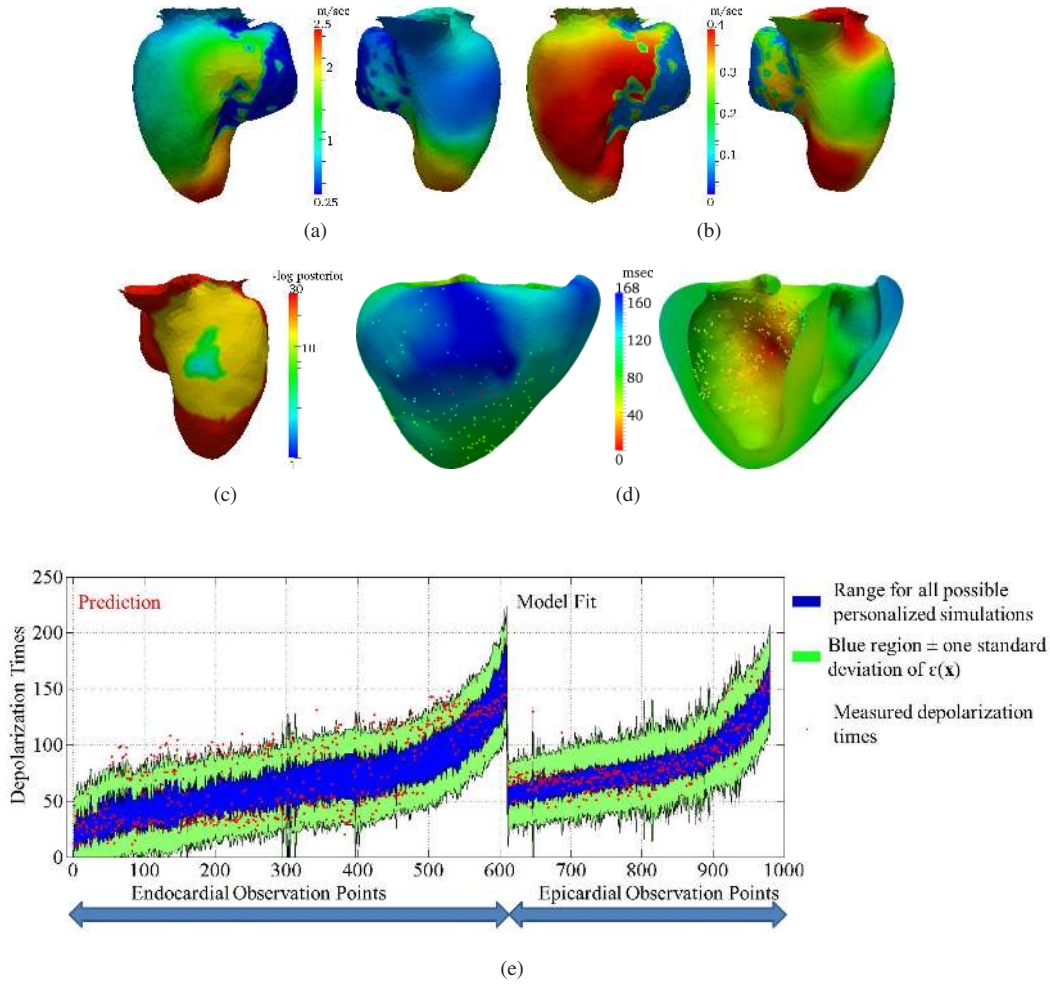


Figure 5: Estimation results for the experiment 2. (a) Mean estimates for the endocardial conductivity functions shown from two views. (b) Marginal standard deviations on the estimated conductivity functions. (c) $-\log(p(\Omega_E|\bar{T}))$ - the $-\log$ posterior for the onset location. (d) Depolarization time isochrones simulated with the mean estimates where color points represent the measurements and their depolarization times. (e) Comparison between the measured depolarization times and the probabilistic model fit and model prediction. The estimation uses only the epicardial observation points. The endocardial depolarization times are predicted.

Observing Fig. 5b one notices that the standard deviations are higher on one side of the endocardium surface. This is as expected since the projection distances of epicardial observation points shown in Fig. 3 are higher on the same side as well. Therefore, these points have higher uncertainty leading to lower confidences on the endocardial conductivity estimates on the corresponding side of the endocardium surface. This demonstrates that the presented method well captures the interaction between the uncertainty of the observations and the model parameter estimates.

Figure 5c shows that the posterior distribution for the onset location, which (almost) covers the one of Fig. 4c. However it is much more spread, also as expected. In Figure 5d the depolarization times simulated by the mean model are shown along with the observations. On the epicardium, the mean model matches the measured depolarization times with a RMS error of 12.9 msec. Using the estimated parameters, the depolarization times for the endocardium are also predicted. The mean model prediction matches the measured values on the endocardium with a RMS error of 26.1 msec.

In Figure 5e the comparison between the measured values and the possible ranges for the depolarization times simulated with the personalized model are shown. For the points that were used for the estimation - epicardial points

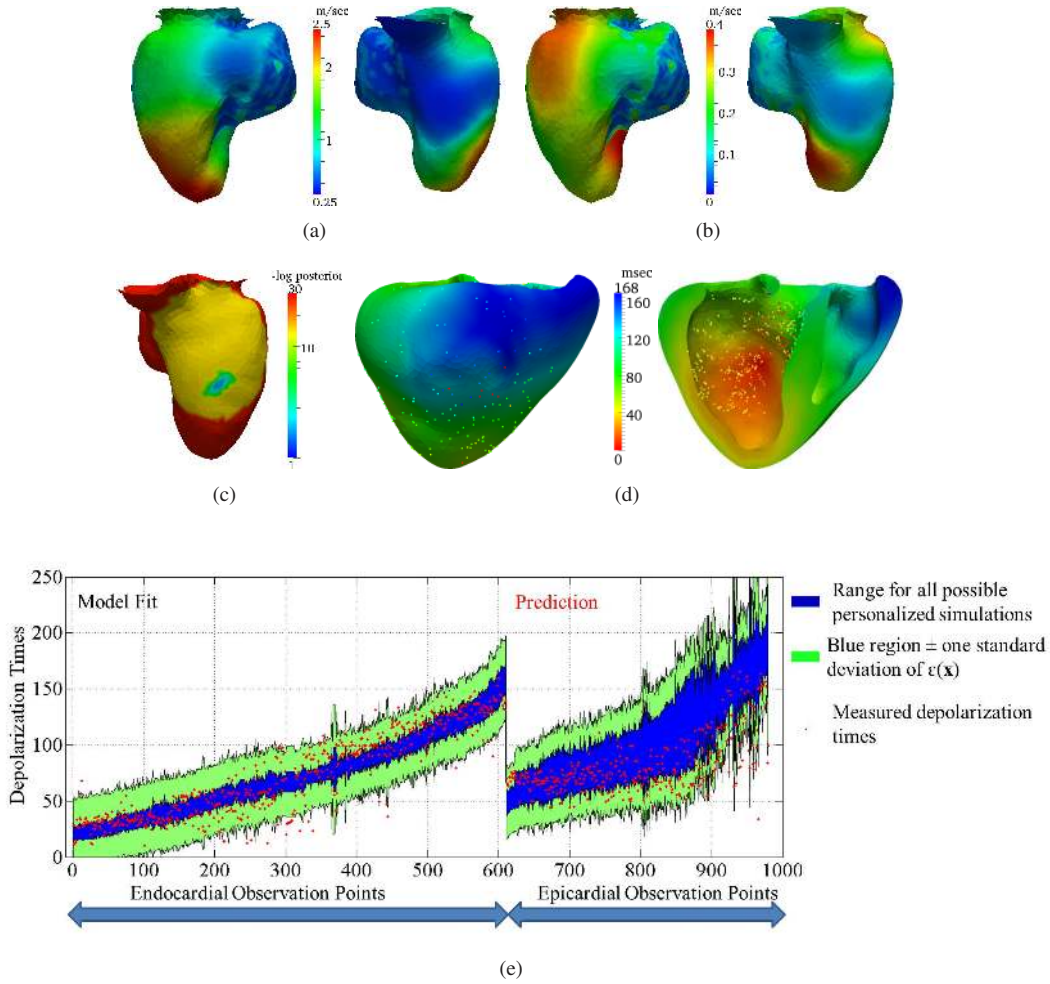


Figure 6: Estimation results for the experiment 3. (a) Mean estimates for the endocardial conductivity functions shown from two views. (b) Marginal standard deviations on the estimated conductivity functions. (c) $-\log(p(\Omega_E|\bar{T}))$ - the $-\log$ posterior for the onset location. (d) Depolarization time isochrones simulated with the mean estimates where color points represent the measurements and their depolarization times. (e) Comparison between the measured depolarization times and the probabilistic model fit and model prediction. The estimation uses only the endocardial observation points. The epicardial depolarization times are predicted.

- we see that the ranges are similar to the ones given in Fig. 4. The probabilistic model fit is similar in this case with a narrow range and almost all measurements lying within the blue region. The properties of the predicted endocardial depolarization times are however, significantly different. The range for the possible predictions is much wider reflecting the uncertainties on the estimated endocardial conductivities. The measured endocardial depolarization times remain within the model plus the observation noise region, however, the spread is much wider. Furthermore, this quantifies the importance of having observation points on both sides of the myocardium for the accuracy and confidence in the personalized model. We note that we used a different ordering of observation points while plotting Figure 4e and 5e to have a better visualization.

Results on Experiment 3. The results for this experiment are presented in Figure 6 and the third row in Table 1. We notice that the mean of the posterior distribution for the conductivity function and the volumetric conductivity values are again structurally similar to the ones estimated in the previous experiments. The numerical differences between the three experiments demonstrates that the observations taken on the endocardium and the epicardium slightly disagrees on the exact values of the conductivities (comparing Figure 6a with Figure 5a and second row of the Table 1 with

the third). This, we believe, is due to the discrepancy between the ED model and the underlying real dynamics. Two observations related to these discrepancies are: i) the mean estimates of Experiment 1 are - in loose terms - averaged between the mean estimates of Experiment 2 and 3, which is as expected and ii) the regions with high standard deviation shown in Figure 4b reflect the differences between the estimates given in Figures 5a and 6a demonstrating that the non-uniqueness of the model-observation interaction can be captured using the proposed method.

Observing Figure 6b we notice that the standard deviations are correlated to the density of the observation points as in the first experiment. Furthermore, as the inference is based on the observations taken only from the endocardial surface the standard deviation values shown in Figure 6b are lower than the other experiments. However, the high standard deviation on the myocardial conductivity given in Table 1 shows that the endocardial observations are not very informative about the muscle conductivity and one needs observations from the epicardium to confidently estimate this value.

The posterior distribution for the onset location (Figure 6c) shows similarities to the one obtained in the first experiment. In Figure 6d we show the depolarization times simulated by the mean model. The RMS model fit error between the mean model simulation and the endocardial measurements is 15.3 msec . The difference between the predicted epicardial depolarization times and the measured ones has an RMS error of 32.2 msec .

As in the previous cases in Figure 6e we show the possible ranges for the depolarization times simulated with the personalized model and the measured values. As expected we see that the range for possible simulated depolarization times is narrower for the endocardial points where the model is fit to the observations. On the other hand, the range of possible predictions of the depolarization times on the epicardium is much wider. Comparing Figures 5e and 6e we notice that the uncertainty on the predicted depolarization times in the latter case is higher. This result is due to the fact that the epicardial observations contain information on the whole system since these depolarization times are influenced by the endocardial conductivities as well as the volumetric ones. The endocardial observations on the other hand, contain much less information regarding the myocardial conductivity. Therefore, the predictions in this case have much a wider possible range, i.e. have much less confidence. We note that this separation of the information content is partly due to the conductivity model we use as given in Equation 3.

7. Discussion

The presented results show the capabilities of the proposed probabilistic method and the fast inference framework. However, there are some important factors that influence the results of the method:

- *Data Noise Model:* The selected noise model has an important impact on the results. Here we used some arbitrary values from our experience on such data, a more exhaustive process looking through the repeatability of the measures and the processing pipeline from the raw signal to the extracted depolarization time should be undertaken to properly define the noise model.
- *Model Parameter Variability:* The prior distributions used for the different model parameters also have an important impact on the results. Here uninformative priors have been used with large ranges. Narrower ranges or a different prior model would have altered the variations on the estimated parameters as well as the ranges for the possible model fits and predictions. Analysis of a set of clinical cases should help in improving the relevance of the prior distributions used here.
- *Model Error:* The realism of the selected model has also obviously an impact on the success of such personalization. Some strategies were proposed to explicitly include the model error in such joint probabilities, including for instance the approximation errors, Arridge et al. (2006). These methods can be integrated into the proposed framework.

Regarding the clinical applications, the presented probabilistic framework provides a confidence measure for the personalized model. This can help the clinician in weighing the predictions of a model with respect to the other clinically available information in the decision process. Moreover, it can also guide the data acquisition. Being able to quantify the contributions of both endo and epi surfaces is important. It can help in deciding if the additional information given by epicardial mapping - which in normal clinical practice would not be available - will be quantitatively significant or not. One application is also to move from invasive to non-invasive methods, namely from endocardial catheters to epicardial data estimated from body surface potentials, and there is then a need to know how confident one can be in the estimated parameters when using only epicardial data.

8. Conclusions

In this work we presented an efficient probabilistic framework for personalizing mathematical models. We demonstrated the proposed framework for the parameter estimation problem for cardiac electrophysiology. The probabilistic formulation describes the interactions between data uncertainties, parameter estimation and the model predictions. Following the formulation the Bayesian inference was possible through integrating spectral representations and compressed sensing techniques. In the experiments we showed that Bayesian inference for high dimensional parameters, i.e. locally varying conductivity, can be efficiently done *within few minutes*, given that the spectral basis is already computed. Moreover, we have shown that the spectral basis can be computed efficiently as well (within 5 hours). Considering the computational cost of conventional methods for probabilistic methods for PDEs, this is an important step in making stochastic personalization practical, and it already provides interesting results on clinical data. The interaction between the uncertainty on the observations and the uncertainty on the estimated parameters is a major phenomenon to consider when using biophysical models for personalized medicine. We demonstrated on clinical data how this interaction translates into the model parameters and predictions for patient data. Our future work is concentrated on analyzing the proposed methodology for different modeling problems (including reaction-diffusion models) and creating more realistic and yet tractable Purkinje network and myocardium interactions.

Appendix A. Polynomial Chaos and Spectral Representation

Polynomial chaos (PC) expansion is an orthonormal basis representation for random functions, Xiu and Karniadakis (2002). For the ease of explanation let us concentrate only on the vector \bar{D} where different components were modeled as random variables with uniform distributions. The PC expansion for the random vector \bar{D} is given as

$$\bar{D} \triangleq \bar{D}(\xi) = \sum_{i=0}^{\infty} \bar{d}_i \Phi_i(\xi) \approx \sum_{i=0}^P \bar{d}_i \Phi_i(\xi), \quad (\text{A.1})$$

where $\xi = [\xi_1, \dots, \xi_Q]$ is a random vector of independent components with uniform distributions, $\{\Phi_i\}_0^{\infty}$ is the orthonormal (multivariate) Legendre polynomial basis and $\{\bar{d}_i\}_0^{\infty}$ are deterministic vectors – which we call the **spectral basis** for the \bar{D} vector. The second part of Eq. (A.1) shows the computationally useful finite approximation of the PC expansion where the parameter P corresponds to number of components to include in the approximation. The orthonormality condition of the PC basis is defined as

$$\langle \Phi_i(\xi), \Phi_j(\xi) \rangle = \int_{\Sigma(\xi)} \Phi_i(\xi) \Phi_j(\xi) p(\xi) d\xi = \delta_{i,j}, \quad (\text{A.2})$$

$$\bar{d}_i = \int_{\Sigma(\xi)} \bar{D}(\xi) \Phi_i(\xi) p(\xi) d\xi, \quad i, j \in \mathbf{N}_0, \quad (\text{A.3})$$

where $\Sigma(\xi)$ is the range of the ξ vector, $p(\xi)$ is the distribution - a constant value in this case - and $\delta_{i,j}$ is Dirac's delta.

The definition of $\bar{D}(\xi)$ and the dimension of the vector ξ depend on the distribution of the vector itself and the random vector \bar{D} , Xiu and Karniadakis (2002). Focusing on the inverse problem defined in Section 3, we defined each component of the random vector $\bar{D} = [D_0, \dots, D_M]$ as uniformly distributed (noted \mathcal{U}) independent random variables. The PC expansion given in Equation A.1 can then be written as

$$\xi = [\xi_0, \dots, \xi_M], \quad \xi_m \sim \mathcal{U}(-1, 1), \quad \forall m \in \{0, \dots, M\} \Rightarrow D_m = D_m(\xi_m) = \frac{D_m^b - D_m^a}{2} \xi_m + \frac{D_m^b + D_m^a}{2}. \quad (\text{A.4})$$

The power of the spectral methods for PDEs is that the randomness of $T(\mathbf{x})$ – which is due to the randomness in parameters – can be represented using the same PC basis functions and the same random vector ξ as the parameters, Ghanem and Doostan (2006); Xiu (2009). We can write

$$T(\mathbf{x}) \triangleq T(\mathbf{x}, \xi) = \sum_{i=0}^{\infty} T_i(\mathbf{x}) \Phi_i(\xi) \approx \sum_{i=0}^P T_i(\mathbf{x}) \Phi_i(\xi), \quad (\text{A.5})$$

where $T_i(\mathbf{x})$ are the spectral basis functions for $T(\mathbf{x})$. This representation provides the computationally efficient link between \overline{D} and $T(\mathbf{x})$. Let \tilde{D} be a single instance of \overline{D} then there is a unique $\tilde{\xi}$ such that $\tilde{D} = \overline{D}(\tilde{\xi})$ due to the orthonormality of the spectral representation given in Equation A.2. As a result

$$T(\mathbf{x}|\tilde{D}) = T(\mathbf{x}, \tilde{\xi}) \approx \sum_{i=0}^P T_i(\mathbf{x})\Phi_i(\tilde{\xi}). \quad (\text{A.6})$$

As a result instead of solving Eq. (1) to compute $T(\mathbf{x}|\tilde{D})$ we can approximate it using the linear combination of Eq. (A.6). This provides substantial speed up yielding faster sampling.

The computationally expensive part in the given discussion is the estimation of the spectral basis functions $\{T_i(\mathbf{x})\}_{i=0}^P$. Two different methods that have been proposed in the literature are Galerkin projections, Xiu and Karniadakis (2002); Ghanem and Doostan (2006), and stochastic collocation methods, Ma and Zabaras (2009); Nobile et al. (2008). Both of these methods suffer from the curse of dimensionality, i.e. when the dimension of the parameter \overline{D} is high the construction of $T_i(\mathbf{x})$ becomes computationally impractical. Depending on the quality of the approximation, P , and the model equation, in the case of Galerkin projections one can end up solving a system of 2000 coupled PDEs, Xiu and Karniadakis (2002), and in the case of stochastic collocation methods a substantial number of model simulations, $O(10^5)$, might be needed, Ma and Zabaras (2009).

In this work we overcome the curse of dimensionality by integrating compressed sensing based sparse reconstructions of $\{T_i(\mathbf{x}_i)\}_{i=0}^P$ in the Bayesian inference framework. This integration provides substantial speed-ups in constructing the spectral basis and makes the Bayesian inference feasible.

Appendix B. Spectral Reconstructing using Compressed Sensing

Compressed sensing (CS) is a very active field of research in signal and image processing, Donoho (2006); Lustig et al. (2008). The main idea is that if a signal is sparse we need much less information to reconstruct it. This information, for example, can be in terms of number of samples taken in the frequency domain. The reason for this is that sparse signals contain only a small number of frequency components and therefore, a small number of random samples is enough for accurate reconstruction, Donoho (2006). In the context of PDEs this corresponds to the fact that in spectral expansion of the solution $T(\mathbf{x}, \xi)$, only a small number of components $T_i(\mathbf{x})$ are significantly larger than zero. Following the compressed sensing theory the significant components can be recovered using a small random set of $\hat{\xi} = \{\xi_n\}_{n=0}^K$ where $K \ll P$. This recovery is possible through the formulation

$$\begin{aligned} \arg \min_{\mathbf{T}(\mathbf{x})} \|\mathbf{T}(\mathbf{x})\|_1 \quad \text{subject to} \quad & \|T(\mathbf{x}, \hat{\xi}) - \Phi(\hat{\xi})\mathbf{T}(\mathbf{x})\|_2 < \delta, \\ & \mathbf{T}(\mathbf{x}) = [T_0(\mathbf{x}), \dots, T_P(\mathbf{x})]^t \\ & T(\mathbf{x}, \hat{\xi}) = [T(\mathbf{x}, \xi_0), \dots, T(\mathbf{x}, \xi_K)]^t, \forall \mathbf{x} \in \Omega \end{aligned} \quad (\text{B.1})$$

where $\Phi(\hat{\xi})$ is the matrix of $K \times P$ with $[\Phi(\hat{\xi})]_{ij} = \Phi_j(\xi_i)$, $\|\cdot\|_1$ is the L_1 norm, $\|\cdot\|_2$ is the L_2 norm and δ is the desired accuracy. This minimization problem is defined for each point \mathbf{x} separately. In the case of Bayesian inference problem we solve it for all the observation points $\{\mathbf{x}_i\}_1^N$. Equation (B.1) is a linear program and can be solved efficiently and accurately using various available tools, for example the one given in Tomioka and Sugiyama (2009).

In Doostan et al. (2009) the one dimensional linear diffusion equation has been analyzed. They have shown that the significant spectral coefficients obtained using the Eq. (B.1) were very similar to the exact coefficients. The advantage of this approach is that it requires a few hundred simulations to construct the significant spectral basis components while stochastic collocation based methods requires high number of simulations ($O(10^5)$) to reach the same accuracy. The amount of decrease in the computational burden depends on the sparsity of the underlying equation. Although, for the general equations the applicability of this method is not known, in the case of the Eikonal-Diffusion model the sparsity holds and can be shown experimentally.

Editor's Note

Please see also related communications in this issue by Aguado-Sierra et al. (2011) and Camara et al. (2011).

References

- Aguado-Sierra, J., Krishnamurthy, A., Villongco, C., Howard, E., Chuang, J., 2011. Patient-specific modeling of dyssynchronous heart failure: A case study. *Prog. Biophys. Mol. Biol.* (doi:10.1016/j.pbiomolbio.2011.06.014).
- Arridge, S. R., Kaipio, J. P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., Vauhkonen, M., 2006. Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems* 22 (1), 175.
- Camara, O., Pashaei, A., Sebastián, R., Frangi, A. F., 2010. Personalization of fast conduction purkinje system in eikonal-based electrophysiological models with optical mapping data. In: STACOM/CESC. Vol. 6364 of Lecture Notes in Computer Science. Springer, pp. 281–290.
- Camara, O., Sermesant, M., Lamata, P., Wang, L., Pop, M., 2011. Inter-model consistency and complementarity: Learning from ex-vivo imaging and electrophysiological data towards an integrated understanding of cardiac physiology. *Prog. Biophys. Mol. Biol.* (NUMBER), [note to publisher: please update before print].
- Chinchapatnam, P., Rhode, K., Ginks, M., Rinaldi, C., Lambiase, P., Razavi, R., Arridge, S., Sermesant, M., 2008. Model-based imaging of cardiac apparent conductivity and local conductivity velocity for diagnosis and planning of therapy. *IEEE TMI* 27 (11), 1631–42.
- Clayton, R., Bernus, O., Cherry, E., Dierckx, H., Fenton, F., Mirabella, L., Panfilov, A., Sachse, F., Seemann, G., Zhang, H., 2011. Models of cardiac tissue electrophysiology: Progress, challenges and open questions. *Prog Biophys Mol Biol* 104 (1–3), 22–48.
- Colli Franzone, P., Guerri, L., Rovida, S., 1990. Wavefront propagation in activation model of the anisotropic cardiac tissue. *J. Math. Biol.* 28 (2), 121–176.
- Dong, J., Calkins, H., Solomon, S., Lai, S., Dalal, D., Lardo, A., Brem, E., Preiss, A., Berger, R., Halperin, H., et al., 2006. Integrated electroanatomic mapping with three-dimensional computed tomographic images for real-time guided ablations. *Circulation* 113 (2), 186.
- Donoho, D., 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52 (4), 1289–1306.
- Doostan, A., Owhadi, H., Lashgari, A., Iaccarino, G., 2009. Non-adapted sparse approximation of PDEs with stochastic inputs. Center for Turbulence Research, Annual Research Briefs.
- Fink, M., Niederer, S. A., Cherry, E. M., Fenton, F. H., Koivumki, J. T., Seemann, G., Thul, R., Zhang, H., Sachse, F. B., Beard, D., Crampin, E. J., Smith, N. P., 2011. Cardiac cell modelling: Observations from the heart of the cardiac physiome project. *Prog Biophys Mol Biol* 104 (1-3), 2 – 21, cardiac Physiome project: Mathematical and Modelling Foundations.
- Geneser, S. E., Kirby, R. M., MacLeod, R. S., Jan 2008. Application of stochastic finite element methods to study the sensitivity of ECG forward modeling to organ conductivity. *IEEE Trans Biomed Eng* 55 (1), 31–40.
- Ghanem, R., Doostan, A., 2006. On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data. *J. Comp. Physics* 217, 63–81.
- Girolami, M., Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J.R. Statist. Soc. B.*
- He, Y., Keyes, D., 2007. Reconstructing parameters of the FitzHugh-Nagumo system from boundary potential measurements. *J. Comput. Neurosci.* 23, 251–264.
- Kaipio, J., Somersalo, E., 2005. *Statistical and Computational Inverse Problems*, 1st Edition. Springer.
- Konukoglu, E., Sermesant, M., Clatz, O., Peyrat, J. M., Delingette, H., Ayache, N., 2007. A recursive anisotropic fast marching approach to reaction diffusion equation: application to tumor growth modeling. *Inf Process Med Imaging* 20, 687–699.
- Lustig, M., Donoho, D., Santos, J. M., Pauly, J. M., March 2008. Compressed sensing in MRI. *IEEE Signal Processing Magazine* 72.
- Ma, X., Zabarbas, N., 2009. An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation. *Inverse Problems* 25.
- Marzouk, Y., Najm, H., Rahn, L., 2007. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics* 224, 560–586.
- Marzouk, Y., Xiu, D., 2009. A stochastic collocation approach to Bayesian inference in inverse problems. *Communications in computational physics* 6 (7), 826–847.
- Marzouk, Y. M., Najm, H. N., 2009. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics* 228, 1862–1902.
- Murray, J., 2002. *Mathematical Biology*, 2nd Edition. Springer-Verlag.
- Nobile, F., Tempone, R., Webster, C., 2008. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.* 46 (5), 2309–2345.
- Pernod, E., Sermesant, M., Konukoglu, E., Relan, J., Delingette, H., Ayache, N., 2011. A multi-front Eikonal model of cardiac electrophysiology for interactive simulation of radio-frequency ablation. *Computers and Graphics* 35, 431–40.
- Ramanathan, C., Ghanem, R. N., Jia, P., Ryu, K., Rudy, Y., Apr 2004. Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. *Nature Medicine* 10 (4), 422–428.
- Relan, J., Chinchapatnam, P., Sermesant, M., Rhode, K., Ginks, M., Delingette, H., Rinaldi, C. A., Razavi, R., Ayache, N., 2011. Coupled personalization of cardiac electrophysiology models for prediction of ischaemic ventricular tachycardia. *Journal of the Royal Society Interface Focus* 1 (3), 396–407.
- Richmond, L., Rajappan, K., Voth, E., Rangavajhala, V., Earley, M., Thomas, G., Harris, S., Sporton, S., Schilling, R., 2008. Validation of computed tomography image integration into the EnSite NavX mapping system to perform catheter ablation of atrial fibrillation. *Journal of Cardiovascular Electrophysiology* 19 (8), 821–827.
- Sacher, F., Roberts-Thomson, K., Maury, P., Tedrow, U., Nault, I., Steven, D., Hocini, M., Koplan, B., Leroux, L., Derval, N., Seiler, J., Wright, M., Epstein, L., Haissaguerre, M., Jais, P., Stevenson, W., 2010. Epicardial ventricular tachycardia ablation: A multicenter safety study. *Journal of the American College of Cardiology* 55 (21), 2366 – 2372.
- Steinhaus, B., 1989. Estimating cardiac transmembrane activation and recovery times from unipolar and bipolar extracellular electrograms: a simulation study. *Circulation Research* 64 (3).
- Tarantola, A., 2005. *Inverse Problem Theory*. SIAM.
- Tomioka, R., Sugiyama, M., 2009. Dual augmented Lagrangian method for efficient sparse reconstruction. *IEEE Signal Processing Letters* 16 (12), 1067–70.

- Tomlinson, K., Hunter, P., Pullan, A., 2002. A finite element method for an eikonal equation model of myocardial excitation wavefront propagation. *SIAM J. Appl. Math.* 63 (1), 324–350.
- Voss, H., Timmer, J., 2004. Nonlinear dynamical system identification from uncertain and indirect measurements. *Int. J. Bifurcation and Chaos* 14 (6), 1905–33.
- Wang, L., Wong, K. C., Zhang, H., Liu, H., Shi, P., Apr 2011. Noninvasive computational imaging of cardiac electrophysiology for 3-D infarct. *IEEE Trans Biomed Eng* 58 (4), 1033–1043.
- Xiu, D., 2007. Efficient collocational approach for parametric uncertainty analysis. *Communications in Computational Physics* 2 (2), 293–309.
- Xiu, D., 2009. Fast numerical methods for stochastic computations: A review. *Communications in Computational Physics* 5 (2-4), 242–272.
- Xiu, D., Karniadakis, G., 2002. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in applied Mechanics and Engineering* 191, 4927–4948.