Bayesian Uncertainty Quantification with Multi-Fidelity Data and Gaussian Processes for Impedance Cardiography of Aortic Dissection

Sascha Ranftl [1,*] **, Gian Marco Melito** [2] **, Vahid Badeli** [3] **, Alice Reinbacher-Köstinger** [3] **, Katrin Ellermann** [2] **and Wolfgang von der Linden** [1,*]

[1] Institute of Theoretical Physics-Computational Physics, Graz University of Technology, 8010 Graz, Austria
[2] Institute of Mechanics, Graz University of Technology, 8010 Graz, Austria; gmelito@tugraz.at (G.M.M.); ellermann@tugraz.at (K.E.)
[3] Institute of Fundamentals and Theory in Electrical Engineering, Graz University of Technology, 8010 Graz, Austria; vahid.badeli@tugraz.at (V.B.); alice.koestinger@tugraz.at (A.R.-K.)
*   Correspondence: ranftl@tugraz.at (S.R.); vonderlinden@tugraz.at (W.v.d.L.)

Abstract: In 2000, Kennedy and O'Hagan proposed a model for uncertainty quantification that combines data of several levels of sophistication, fidelity, quality, or accuracy, e.g., a coarse and a fine mesh in finite-element simulations. They assumed each level to be describable by a Gaussian process, and used low-fidelity simulations to improve inference on costly high-fidelity simulations. Departing from there, we move away from the common non-Bayesian practice of optimization and marginalize the parameters instead. Thus, we avoid the awkward logical dilemma of having to choose parameters and of neglecting that choice's uncertainty. We propagate the parameter uncertainties by averaging the predictions and the prediction uncertainties over all the possible parameters. This is done analytically for all but the nonlinear or inseparable kernel function parameters. What is left is a low-dimensional and feasible numerical integral depending on the choice of kernels, thus allowing for a fully Bayesian treatment. By quantifying the uncertainties of the parameters themselves too, we show that "learning" or optimising those parameters has little meaning when data is little and, thus, justify all our mathematical efforts. The recent hype about machine learning has long spilled over to computational engineering but fails to acknowledge that machine learning is a big data problem and that, in computational engineering, we usually face a little data problem. We devise the fully Bayesian uncertainty quantification method in a notation following the tradition of E.T. Jaynes and find that generalization to an arbitrary number of levels of fidelity and parallelisation becomes rather easy. We scrutinize the method with mock data and demonstrate its advantages in its natural application where high-fidelity data is little but low-fidelity data is not. We then apply the method to quantify the uncertainties in finite element simulations of impedance cardiography of aortic dissection. Aortic dissection is a cardiovascular disease that frequently requires immediate surgical treatment and, thus, a fast diagnosis before. While traditional medical imaging techniques such as computed tomography, magnetic resonance tomography, or echocardiography certainly do the job, Impedance cardiography too is a clinical standard tool and promises to allow earlier diagnoses as well as to detect patients that otherwise go under the radar for too long.

Keywords: uncertainty quantification; multi fidelity; Gaussian processes; probability theory; Bayes; impedance cardiography; aortic dissection

1. Introduction

While Uncertainty Quantification (UQ) has become a term on its own in the computational engineering community, Bayesian Probability Theory is not widely spread yet. A comprehensive collection of reviews on the various methods and aspects of UQ from the point of view of the computational engineering and applied mathematics community can be found in Reference [1]. In References [2–4], a statistician's perspective is discussed. The computational effort for performing UQ with brute force is typically prohibitively large; thus, surrogate models such as Polynomial Chaos Expansion (PCE) [5–8] or Gaussian Process (GP) regression [9–12] are used, the latter of which has had its renaissance recently from within the machine learning community.

This work is inspired by the article of Kennedy and O'Hagan in 2000 [13]. They performed UQ by making use of a computer simulation with different levels of "fidelity", "sophistication", "accuracy", or "quality". In other words, a cheap, simplified simulation serves as a surrogate. We will refer to this approach as Multi-Fidelity scheme (MuFi). The idea of MuFi [14], and MuFi with GPs specifically [15], recently found increasing attention again. In contrast to previously reported MuFi-GPs , we do not learn the parameters and subsequently neglect the parameter uncertainties but explicitly incorporate them in a rigorous manner. We find that this is tractable analytically for all parameters but especially for the ones that occur nonlinearly or inseparably in the GPs covariance.

While UQ in general has arrived fully in the biomedical engineering community [16], the Bayesian approach has not. Biehler et al. [17] were, to the best knowledge of the authors, the first to apply a Bayesian MuFi Scheme in the context of computational biomechanical UQ. We apply our method to quantify the uncertainties in finite element simulations [18] of Impedance Cardiography (ICG) [19] of Aortic Dissection (AD) [20]. The aorta is the largest blood vessel in the human body. In aortic dissection, blood fluid dynamics force open a tear in a weakened aortic wall, dilate it, and fill the wall itself with blood. This deforms the geometry of the aorta and, obviously, affects blood circulation unfavourably (p. 459, [21]). Aortic Dissection is highly dangerous and likely lethal if untreated. Thus, a fast response and, hence, a fast diagnosis are key to the treatment of patients. For diagnosis, physicians use a variety of imaging techniques such as Magnetic Resonance Tomography (MRT), Computed Tomography (CT), and Echocardiography [22]. Echocardiography performed by a trained cardiologist is comparably cheap and fast, yet sound wave propagation might be hindered, e.g., by the rib cage or body fat. In CT and MRT, the radiation fully penetrates the body. Still, they require a trained radiologist and long measurement times and pose radiation risks and high costs. Most importantly, these examinations are not performed without a specific reason.

Alternatively, impedance cardiography is rather cheap and simple and, more importantly, available in any clinic and many medical practices. In ICG, one places a pair of electrodes on the thorax (upper body), injects a defined low-amplitude, alternates electric current into the body, and measures the voltage drop. The specific resistance of blood is much lower than that of muscle, fat, or bone [23]. Since electric current seeks the path of least resistance, the current propagates through the aorta rather than through, e.g., the spine. Thus, if the local blood volume changes due to aortic dissection, the impedance signal changes as well. Impedance cardiography could therefore complement existing clinical procedures and could detect aortic dissection when medical imaging is not performed, be it due to the absence of suspicion or to the unavailablity of the device itself. We find a number of parameters which are well defined but usually neither known precisely nor accessible in the clinical setting, e.g., the size of the aortic dissection. A clinical trial is extremely difficult, and we resort to a theoretical investigation instead, in which we account for the uncertainties as well.

In Section 2, we develop a Bayesian uncertainty quantification model based on Gaussian processes using multi-fidelity data. We scrutinize the method with mock data in Section 3 and show that learning regression parameters has little meaning when data is little. In Section 4, we apply our method to finite element simulations of impedance cardiography of aortic dissection and show that low-fidelity data can indeed decrease high-fidelity uncertainties.

2. Bayesian Multi-Fidelity Scheme

2.1. Statistical Model

Let $\mathcal{C}$ be the conditional complex. Let $t = 1, ..., N_t$ denote the ranked levels of fidelity of a simulation, with level N_t being the highest fidelity. $z_t(x_t)$ is a vector of simulation results of fidelity-level t given input vector x_t. We assume that $z_t(x_t)$ is a realisation of a Gaussian Process (GP) $z_t(x)$ with a Markov property of order 1, meaning that level t depends on level $t - 1$ only via the following recursive relationship:

$$z_1(x) = \delta_1(x) \tag{1a}$$

$$z_t(x) = \rho_{t-1} z_{t-1}(x) + \delta_t(x) \ \ \forall t \geq 2 \tag{1b}$$

$$\Rightarrow z_{N_t}(x) = \sum_{t=1}^{N_t} \delta_t(x) \prod_{l=t}^{N_t-1} \rho_l \tag{1c}$$

with a "difference-GP" $\delta_t(x)$ and a proportionality constant ρ_{t-1}. Further, we assume that all information about a level is contained in the data corresponding to the same pivot point at that level and its previous level. Formally, that is $Cov\Big(z_t(x), z_{t-1}(x') \mid z_{t-1}(x)\Big) = 0$. The difference-GP $\delta_t(x)$ shall be defined by the covariance matrix $\sigma_t^2 K_t(x, x')$ and the mean function $h_t(x)\boldsymbol{\beta}_t$. $h_t(x)$ is a matrix of regression functions $h_t^{(k)}$ evaluated at $x = (x^{(1)}, ..., x^{(j)}, ..., x^{(N_x)})^T$ with size $N_x \times N_{\boldsymbol{\beta}_t}$, where N_x is the length of input vector x and $N_{\boldsymbol{\beta}_t}$ is the expansion power, i.e., number of regression functions, at level t. $\boldsymbol{\beta}_t = (\beta_t^{(1)}, ..., \beta_t^{(N_{\beta_t})})^T$ are the coefficients of level t's regression functions, and $\boldsymbol{\alpha}_t$ is the set of hyperparameters parametrizing the kernel function k_t. Formally, this is

$$p\Big[\delta_t(x)|\mathcal{C}\Big] = \mathcal{N}(h_t(x)\boldsymbol{\beta}_t, \sigma_t^2 K_t(x, x)) \tag{2a}$$

$$\big[h_t(x)\big]_{jk} = h_t^{(k)}(x^{(j)}) \tag{2b}$$

$$\big[h_t(x)\boldsymbol{\beta}_t\big]_j = \sum_{k=1}^{m} h_t^{(k)}(x^{(j)})\beta_t^{(k)} \tag{2c}$$

$$\big[K_t(x, x)\big]_{ij} = k_t(x^{(i)}, x^{(j)}; \boldsymbol{\alpha}_t) \tag{2d}$$

At this point, neither have we chosen the basis functions $h_t^{(k)}$ building the mean function nor have we chosen the kernel functions k_t building the covariance matrices. Let us subsume the parameters as $\boldsymbol{\theta} = \{\theta_t\}$, $\theta_t = \{\boldsymbol{\beta}_t, \sigma_t, \rho_{t-1}, \boldsymbol{\alpha}_t\}$, $\boldsymbol{\beta} = \{\boldsymbol{\beta}_t\}$, $\boldsymbol{\beta}_t = \{\beta_t^{(k)}\}$, $\sigma = \{\sigma_t\}$, $\rho = \{\rho_t\}$, and $\alpha = \{\alpha_t\}$ with $t = 1, 2, ..., N_t$. The data shall be $D = \{D_t\}$ with $D_t = \big\{\big(x_t, z_t(x_t), z_{t-1}(x_t)\big)\big\}$, which comprises the input vector at level t, namely x_t, and its corresponding computer code outputs at level t, namely $z_t(x_t)$, and at the previous level $t - 1$, namely $z_{t-1}(x_t)$. Further, we require a nested design of input vectors, i.e., $x_{N_t} \subseteq x_{N_t-1} \subseteq ... x_t \subseteq x_{t-1} ... \subseteq x_1$. We want to draw conclusions from the predictive posterior probability of $z_{N_t}(x)$ at a set of points x,

$$p\Big[z_{N_t}(x)|D\Big] = \int \prod_{t=1}^{N_t} d\delta_t(x) \int d\boldsymbol{\theta} \ p\Big[z_{N_t}(x)|\{\delta_t(x)\}_{t=1}^{N_t}, \boldsymbol{\theta}, D\Big] p\Big[\{\delta_t(x)\}_{t=1}^{N_t}|\boldsymbol{\theta}, D\Big] p\Big[\boldsymbol{\theta}|D\Big]$$

This thing is quite unhandy. We will instead just deal with its moments only, namely the posterior mean $\langle z_{N_t}(x) \rangle$ and the posterior covariance $Cov(z_{N_t}(x)) = \langle z_{N_t}(x) z_{N_t}(x)^T \rangle - \langle z_{N_t}(x) \rangle \langle z_{N_t}(x) \rangle^T$,

where the diagonal of the posterior covariance is the uncertainty band of the prediction. The moments of $z_{N_t}(x)$ follow from

$$
\begin{aligned}
\left\langle f(z_{N_t}(x)) \right\rangle &= \int dz_{N_t}(x) f(z_{N_t}(x)) p\left[z_{N_t}(x)|D\right] \\
&= \int \prod_{t=1}^{N_t} d\delta_t(x) \int \prod_{t=1}^{N_t} d\theta_t f\left(\sum_{t=1}^{N_t} \delta_t(x) \prod_{l=t}^{N_t-1} \rho_l\right) \prod_{t=1}^{N_t} p\left[\delta_t(x)|\theta_t, D_t\right] p\left[\theta_t|D_t\right] \quad (3)
\end{aligned}
$$

We have used the fact that $p\left[z_{N_t}(x)|\{\delta_t(x)\}_{t=1}^{N_t}, \boldsymbol{\theta}, D\right]$ reduces to Dirac's delta-distribution since $z_{N_t}(x)$ is uniquely determined by Equation (1) and the knowledge of all difference-GPs, $\{\delta_t(x)\}_{t=1}^{N_t}$. Thus, integration with respect to $z_{N_t}(x)$ is merely a replacement of $z_{N_t}(x) \to \sum_{t=1}^{N_t} \delta_t(x) \prod_{l=t}^{N_t-1} \rho_l$. Per construction, we can factor $p\left[\{\delta_t(x)\}_{t=1}^{N_t}|\boldsymbol{\theta}, D\right] p\left[\boldsymbol{\theta}|D\right] = \prod_t p\left[\delta_t(x)|\theta_t, D_t\right] p\left[\theta_t|D_t\right]$. Since $\delta_t(x)$ is assumed to obey a GP, the prior probability of $\delta_t(x_t)$ is multivariate normal. If the likelihood $p\left[D_t|\delta_t(x), \theta_t\right]$ is Gaussian, then the posterior probability of $\delta_t(x_t)$, $p\left[\delta_t(x_t)|\theta_t, D_t\right]$, is multivariate normal as well. Integration with respect to $\delta_t(x)$ yields thus the standard result of the posterior mean value (see, e.g., Reference [10]) and results in a replacement of the GP with its posterior mean.

2.2. Prediction and Its Uncertainty

To compute posterior mean and covariance, we are thus left with integration with respect to the hyperparameters. This is done analytically with all parameters (β, σ, ρ) but the parameters of the kernel function, α, since those usually occur nonlinearly in the kernel function. We are then left with numerical integration of expectations and covariances, both conditioned on α. We assume flat priors for β and ρ and Jeffreys' prior for σ, i.e., $p\left[\sigma_t|\mathcal{C}\right] = \frac{1}{\sigma_t}$. The prior of α needs to be chosen only after the covariance kernel has been chosen. Let $\langle \cdot \mid \alpha_t \rangle$ denote the expectation value conditioned on α. Here, for ease of notation, we will instead write $\langle \cdot \rangle$ only. The technicalities shall be detailed in the Appendix A , and the result is as follows:

$$
\left\langle z_{N_t}(x) \right\rangle = \int \sum_{t=1}^{N_t} \left\langle \delta_t(x) \right\rangle \prod_{l=t}^{N_t-1} \left\langle \rho_l \right\rangle p\left[\alpha|D\right] d\alpha
$$

$$
\left\langle z_{N_t}(x)(z_{N_t}(x))^T \right\rangle = \int \sum_{t=1}^{N_t} \left(\left\langle \sigma_t^2 \right\rangle \Sigma_t(x,x) + \left\langle \delta_t(x) \right\rangle \left\langle \delta_t(x) \right\rangle^T\right) \prod_{l=t}^{N_t-1} \left\langle \rho_l^2 \right\rangle p\left[\alpha|D\right] d\alpha
$$

$$
\left\langle \delta_t(x) \right\rangle = h_t(x)\left\langle \boldsymbol{\beta}_t \right\rangle + K_t(x,x_t)K_t(x_t,x_t)^{-1}\left(\delta_t(x_t) - h_t(x_t)\left\langle \boldsymbol{\beta}_t \right\rangle\right)
$$

$$
\Sigma_t = \left(K_t(x,x) - K_t(x,x_t)K_t(x_t,x_t)^{-1}K_t(x_t,x)\right)
$$

$$
p\left[\alpha|D\right] = \frac{2^{-N_t}}{\sqrt{\Phi_1}}\frac{\Gamma(\gamma_1)}{\Gamma(\gamma_1 - \frac{1}{2})} \prod_{t=1}^{N_t} a_t^{-\frac{1}{2}} \left(\pi\Phi_t\right)^{-\gamma_t + \frac{1}{2}} \Gamma\left(\gamma_t - \frac{1}{2}\right) \left[\frac{|K_t(x_t,x_t)|}{|A_t|}\right]^{-\frac{1}{2}} p\left[\alpha|\mathcal{C}\right] \quad (4a)
$$

where $\delta_t(x_t)$ is determined from the data and Equation (2), $\Gamma(\cdot)$ is the complete Gamma-function, $|\cdot|$ is the matrix determinant, and the conditional expectations of the hyperparameters are

$$
\left\langle \boldsymbol{\beta}_t \right\rangle = A_t B_t\left(z_t(x_t) - \left\langle \rho_{t-1} \right\rangle z_{t-1}(x_t)\right) \qquad\qquad \left\langle \rho_{t-1} \right\rangle = \frac{b_t}{a_t}
$$

$$
\left\langle \sigma_t^2 \right\rangle = \frac{\Phi_t}{2\gamma_t - 3 + 3\delta_{t1}} \qquad\qquad\qquad\qquad \left\langle \rho_{t-1}^2 \right\rangle = \frac{\left\langle \sigma_t^2 \right\rangle}{a_t} + \left(\frac{b_t}{a_t}\right)^2 \quad (4b)
$$

where $\delta_{t1} = 1$ for $t = 1$ and 0 otherwise and the abbreviations are

$$\gamma_t = \frac{N_{x_t} - N_{\beta_t}}{2} \qquad\qquad \Phi_t = c_t - \frac{b_t^2}{a_t}$$

$$C_t = K_t(x_t, x_t)^{-1} - B_t^T A_t B_t \qquad\qquad c_t = (z_t(x_t))^T C_t z_t(x_t)$$

$$B_t = (h_t(x_t))^T K_t(x_t, x_t)^{-1} \qquad\qquad b_t = (z_t(x_t))^T C_t z_{t-1}(x_t)$$

$$A_t = \left((h_t(x_t))^T K_t(x_t, x_t)^{-1}(h_t(x_t))\right)^{-1} \qquad\qquad a_t = (z_{t-1}(x_t))^T C_t z_{t-1}(x_t) \qquad \text{(4c)}$$

N_{x_t} is the number of pivot points in input vector x_t, and N_{β_t} is the expansion order of level t's mean function. For $t = 1$, we need to define $\rho_0 = 0$, $a_1 = 1$, and $b_1 = 0$. Quite importantly, we find following the requirement:

$$N_{\beta_1} < N_{x_1} - 2$$

$$N_{\beta_t} < N_{x_t} - 3 \quad \forall t \geq 2 \qquad\qquad \text{(4d)}$$

since otherwise the second moments of σ_t are not defined. The numerical evaluation of this result merely involves a couple of matrix operations. The only input is the data, regression functions, and covariance matrices. No parameters need to be tuned.

3. Algorithm and Mock Data Scrutiny

We test the method for the special case of two levels. Then, posterior mean and covariance are simply

$$\langle z_2(x) \rangle = \int \langle \delta_1(x) \rangle p\left[\alpha_1|D_1\right] d\alpha_1 + \int \langle \rho_1 \rangle \langle \delta_2(x) \rangle p\left[\alpha_2|D_2\right] d\alpha_2$$

$$\langle z_2(x)(z_2(x))^T \rangle = \int \left(\langle \sigma_1^2 \rangle \Sigma_1(x, x) + \langle \delta_1(x) \rangle \langle \delta_1(x) \rangle^T\right) p\left[\alpha_1|D_1\right] d\alpha_1$$

$$+ \int \langle \rho_1^2 \rangle \left(\langle \sigma_2^2 \rangle \Sigma_2(x, x) + \langle \delta_2(x) \rangle \langle \delta_2(x) \rangle^T\right) p\left[\alpha_2|D_2\right] d\alpha_2$$

We generate mock data according to Equations (1) and (2) and compare the data analysis results to the underlying truth. We have chosen as mean function bases the Legendre polynomials up to orders 10 and 4 for Levels 1 and 2, respectively. This is convenient since this basis is both orthogonal as well as normalized on $[-1, 1]$ already and the map onto the desired domain is trivial. The covariance kernel was chosen to be the squared exponential kernel, where α_1 and α_2 were defined as the inverse of the correlation length squared. This choice was inspired by the typical form of signals encountered in impedance cardiography, about which we will talk more in Section 4. The data set was one sample drawn per level; see Figure 2 . We chose Jeffreys' prior for α_t. The integration bounds can be read from Figure 1, and an integration grid of 100×100 equally sized volumes turned out to be well converged. As an intermediate result, we compare the multi-fidelity estimates to the true parameter values in Table 1. The posterior probabilities of $\beta_t, \rho_t, \sigma_t$ are Gaussian. Since α_1 and α_2 are not and thus cannot be reasonably well described with mean and variance only, we additionally show their posterior probabilities in Figure 1. The predictions and prediction uncertainties are compared to the true mean in Figure 2. We find that both the parameter estimates as well as the predictions statistically match the truth within their uncertainties. The mean function parameter uncertainties (β) clearly illustrate that learning parameters by optimization has little meaning if there is little data. As the data set grows big, the posterior will contract to the maximum likelihood solution. Still, and luckily, the prediction uncertainties are kept low because the mean function parameter uncertainties do not appear in the

prediction uncertainties directly. We emphasize that our proposed method naturally is applied to little data problems.

Since on level 2 we only have 11 data points, the expansion order of the mean function is limited to a maximum of 7 according to Equation (4d). Unsurprisingly, the solution rapidly worsens as we approach this constraint and entirely breaks down as we reach it because the posteriors become nonconclusive. This is exactly where we find the strength of our multi-fidelity approach. We can choose a high-order mean function on a level where data is abundant and a low-order mean function on a level which we are actually interested in but where data is scarce. The trick is thus actually that the difference of the levels can be modeled by a low-order mean function.
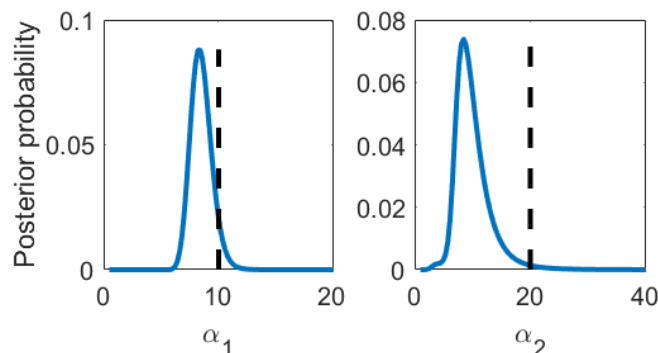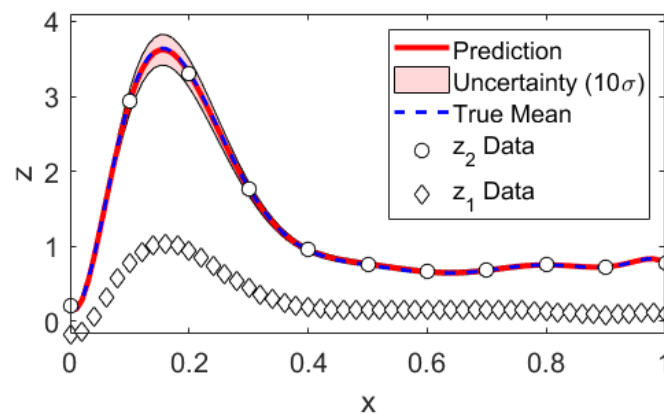

Figure 1. Mock data analysis: Posterior probability density functions of the nonlinear kernel parameters α_1 and α_2. Black dashed line: True value

Table 1. Mock data analysis: Comparison of the hyperparameter estimates with their true values.

Hyperparameter	Estimate (Multi-Fidelity)	Truth
$\beta_1^{(1)}$	0.30 ± 0.07	0.32
$\beta_1^{(2)}$	-0.30 ± 0.09	-0.40
$\beta_1^{(3)}$	0.02 ± 0.08	0.1
$\beta_1^{(4)}$	0.34 ± 0.06	0.35
$\beta_1^{(5)}$	-0.50 ± 0.04	-0.51
$\beta_1^{(6)}$	0.35 ± 0.02	0.33
$\beta_1^{(7)}$	-0.033 ± 0.008	-0.034
$\beta_1^{(8)}$	-0.146 ± 0.005	-0.142
$\beta_1^{(9)}$	0.1745 ± 0.0008	0.1750
$\beta_1^{(10)}$	-0.1031 ± 0.0004	-0.1034
$\beta_2^{(1)}$	0.42 ± 0.59	$0.\dot{3}$
$\beta_2^{(2)}$	-0.17 ± 0.55	0.15
$\beta_2^{(3)}$	0.11 ± 0.17	$0.01\dot{6}$
$\beta_2^{(4)}$	0.03 ± 0.10	0
σ_1	0.133 ± 0.096	0.1
σ_2	0.47 ± 0.58	0.01
ρ_1	2.97 ± 0.02	3
α_1	8.6 ± 0.9	10.1
α_2	10 ± 3	20.1

Figure 2. Mock data analysis: Prediction. Note that the uncertainties have been multiplied by a factor of 10 for illustrative purposes.

For the sake of completeness, we report the converged log-evidence to be 430 ± 10. In real-life applications of the method, one should and could compare different choices of mean function expansions and covariance kernel functions by Bayesian model comparison [24], i.e., compute each choice's evidence. Let $N_{\boldsymbol{\alpha}_t}$ be the number of hyperparameters in kernel function k_t. Since $p\left[\boldsymbol{\alpha}|D\right]$ factorizes, the integrals are $N_{\boldsymbol{\alpha}_t}$-dimensional each rather than one single integral of dimension $\prod_t N_{\boldsymbol{\alpha}_t}$, making the computation of the evidence relatively easy. In our case, numerical Riemann integration was good enough. When choosing more sophisticated kernel functions with more hyperparameters, one might need to use statistical integration methods such nested sampling [25], which conveniently and automatically yields the evidence as well.

For the sake of numerical stability, it is advisable to rescale the data. Further, one might want to improve the condition numbers of the prior covariance matrices, $\sigma_t^2 K_t(x, x)$, by adding a small term proportional to the identity matrix, where the proportionality constant should be several orders of magnitude smaller than σ_t^2.

Finally, we would like to point out that Equation (4) suggests trivial parallelisation of the code levels. This is not easily recognizable in the presentation of Kennedy and O'Hagan [3] but was found by Le Gratiet and Garnier [26] already.

The algorithm, implemented in Matlab (R2019a), shall be available on https://github.com/Sranf/Bayesian-MuFi-GP.git.

4. Application to Finite Element Simulations of Impedance Cardiography of Aortic Dissection

In this section, we quantify the uncertainties of real simulation data. We compare the uncertainties of our Bayesian MuFi-GP with normal Bayesian GP neglecting the additional LoFi data, that is, the special case of $N_t = 1$ in Equation (4). We have described the physical and physiological model in our previous work [20,27] but shall restate a brief summary here for the reader's convenience.

We solve Laplace's equation:

$$\nabla \cdot \left((\sigma + i\omega\varepsilon)\nabla V \right) = 0 \,,$$

with finite elements on a geometry depicted in Figure 3, where V is the electric potential, σ is the electrical conductivity (not to be confused with the regression parameter in the GP kernel), ω is the angular current frequency, ε is the permittivity, and i is the imaginary unit. We had von-Neumann boundary conditions, where $V = const.$ on the top electrode and $V = 0$ on the bottom electrode, where the surface integral of the current was held constant at 4 mA and air was assumed to be perfectly insulating. The current had a frequency of 100 kHz. We considered one cardiac cycle that spans one second. The dynamics were modelled via a time-dependent radius of the aorta and its

dissection, which arises from pressure waves in a pulsatile flow. Further, the blood conductivity is parametrized in time via its dependence on flow velocity. In the dissected aorta, we assume flow to be stagnant [20,28]. The voltage drop is then measured from just below the top electrode to just above the bottom electrode. The impedance is then the ratio of voltage over current, and the admittance is the inverse of the impedance. We used Comsol Multiphysics for the modelling [29].
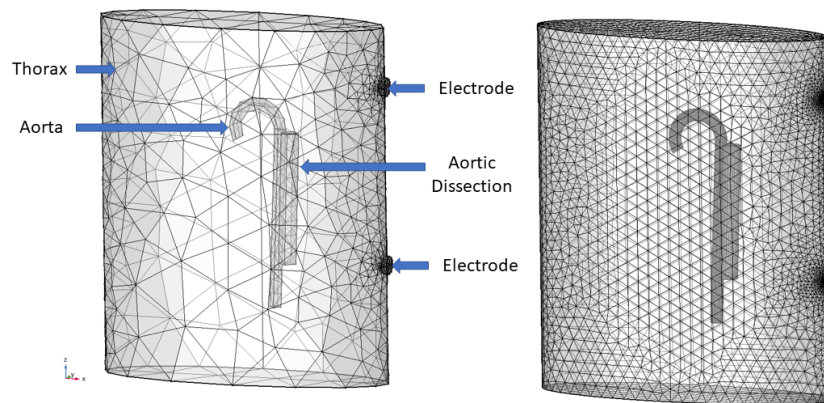

Figure 3. Right: Mesh-converged HiFi model with 100,000–550,000 degrees of freedom. Left: LoFi model with 9000–15,000 with labels of the geometrical objects. Adapted from Reference [27]

For the uncertainty quantification, we chose to expand the mean functions in Legendre polynomials up to order 8 and 2 for HiFi and LoFi, respectively. We choose the squared exponential kernel for both covariance matrices. In principle, one should compute the evidence for a number of plausible choices and choose the one with the most evidence. For the mean function, that would most simply be different expansion orders, while the covariance kernel could be taylored to the PDE at hand to enforce physical behaviour, as suggested in References [30,31].

For uncertain parameters, we consider the radius of the dissected aorta and perform a number of simulations with sensible values within the physiological and physical range, i.e., 1.0–24.0 mm. The LoFi data set consisted of 24 time series, each with 21 pivot points in time. The HiFi data set consisted of 3 time series (5 mm, 11 mm, and 18 mm), each with 11 pivot points in time.

In Figure 4, we show the posterior of the kernel parameters, which turns out to be quite conclusive. In Figure 5, we plot the HiFi predictions and uncertainties which are enhanced by LoFi data and compare them to HiFi predictions and uncertainties which are not enhanced by LoFi data as well as to the test data set.
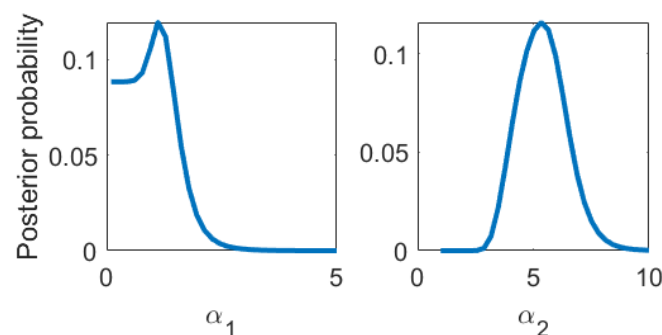
Figure 4. Posterior probability of the nonlinear kernel parameters.
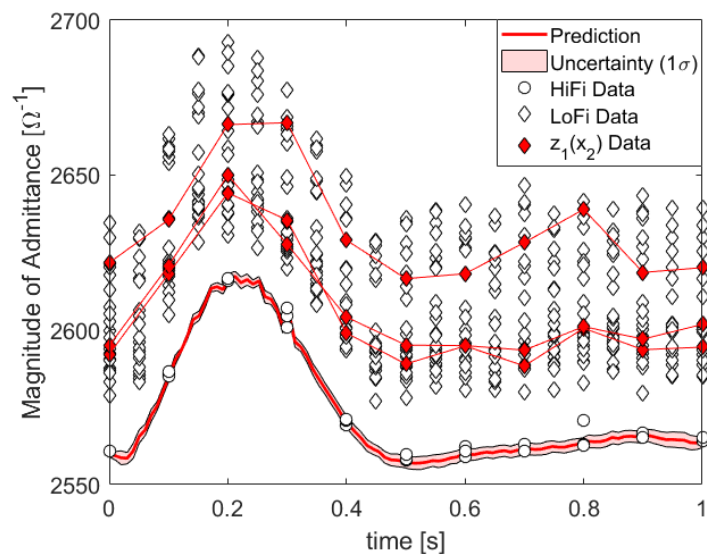
Figure 5. Data, prediction, and prediction uncertainty of the absolute value of the admittance, i.e., the inverse impedance in units of inverse Ohm: $z_1(x_2)$ denotes LoFi data at the same pivot points as HiFi data.

5. Conclusions

We devised a fully Bayesian multi-level Gaussian process model to improve uncertainty quantification of expensive and little high-fidelity simulation data by augmenting the data set with low-fidelity simulations. Our proposed method is rigorous and logically consistent, no ad hoc assumptions have been made, and the user is spared the embarrassment of having to tune any parameters. The method was scrutinized with mock data and shown to work with as little data as where simple Bayesian Gaussian process regression is not conclusive at all. We applied the method to finite element simulations of impedance cardiography of aortic dissection and quantified the uncertainty due to the unknown size of the aortic dissection. By using meshes of both high fidelity (defined by mesh convergence) and low fidelity, we reduced the uncertainty significantly. We have thus further shown that uncertainties due to geometrical parameters can be described with Gaussian processes on each level of fidelity. With a coarsened mesh, the result is qualitativelybut not quantitatively similar. Usually, that is not good enough and the low-fidelity data is entirely useless to the engineer. Here we show that this is not necessarily true in the context of uncertainty quantification. Ultimately, we want to diagnose aortic dissection from impedance cardiography signals, i.e., in the parlance of probability theory, we need to compare the evidences of healthy and diseased aortae. Unambiguous judgement will most likely, if at all, be possible only with several electrodes at once.

Appendix A. Mathematical Proofs

We start from Equation (3) and want to compute Equation (4) from it. We reintroduce the notation of conditional expectations, where $\langle Q \mid P \rangle$ is the expectation of Q given some P, i.e., the integration is done with respect to all parameters but P.

$$\boldsymbol{p}\Big[\delta_t(x_t)|x_t,\theta_t\Big] = \mathcal{N}\Big[\delta_t(x_t)|\bar{\delta}_t(x_t),K_t(x_t,x_t)\Big]$$

$$\bar{\delta}_t(x) = h_t(x)\boldsymbol{\beta}_t + K_t(x,x_t)K_t(x_t,x_t)^{-1}\big(\delta_t(x_t) - h_t(x_t)\boldsymbol{\beta}_t\big)$$

$$\Rightarrow \boldsymbol{p}\Big[\boldsymbol{\theta}|D\Big] = \prod_t \boldsymbol{p}\Big[\theta_t|D_t\Big]$$

$$\boldsymbol{p}\Big[D_t|\theta_t\Big] = \mathcal{N}\Big[\delta_t(x_t)|\bar{\delta}_t(x_t),K_t(x_t,x_t)\Big]$$

Thus, in the moments of $z_{N_t}(x)$, by integration with respect to $\delta_t(x)$, we can replace $\delta_t(x)$ by its posterior mean:

$$\Big\langle z_{N_t}(x) \Big\rangle = \int d\boldsymbol{\theta} \Big(\sum_{t=1}^{N_t} \bar{\delta}_t(x) \prod_{l=t}^{N_t-1} \rho_l \Big) \boldsymbol{p}\Big[\boldsymbol{\theta}|D\Big] d\boldsymbol{\theta}$$

$$\Big\langle z_{N_t}(x)z_{N_t}(x)^T \Big\rangle = \int d\boldsymbol{\theta} \Big(\sum_{t=1}^{N_t} \bar{\delta}_t(x) \prod_{l=t}^{N_t-1} \rho_l \Big) \Big(\sum_{t=1}^{N_t} \bar{\delta}_t(x) \prod_{l=t}^{N_t-1} \rho_l \Big)^T \boldsymbol{p}\Big[\boldsymbol{\theta}|D\Big] d\boldsymbol{\theta}$$

Appendix A.1. Parameter Posterior and Parameter Estimates

We need the posterior of the parameters:

$$\mathcal{N}\Big[\delta_t(x_t)|\bar{\delta}_t t, K_t(x_t,x_t)\Big] = \frac{1}{Z_t} e^{-\frac{\psi_t}{2\sigma_t^2}}$$

$$\psi_t = (\delta_t(x_t) - h_t(x_t)\boldsymbol{\beta}_t)^T K_t(x_t,x_t)^{-1}(\delta_t(x_t) - h_t(x_t)\boldsymbol{\beta}_t)$$

$$Z_t = (2\pi\sigma_t^2)^{N_{x_t}/2}|K_t(x_t,x_t)|^{1/2} \tag{A1a}$$

The exponent of this Gaussian is a quadratic form in $\boldsymbol{\beta}_t$, and we rewrite it as

$$-\frac{\psi_t}{2\sigma_t^2} = -\frac{1}{2\sigma_t^2}(\boldsymbol{\beta}_t - \langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t, \rho_{t-1} \rangle)^T A_t^{-1}(\boldsymbol{\beta}_t - \langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t, \rho_{t-1} \rangle) - \frac{1}{2\sigma_t^2}\delta_t(x_t)^T C_t \delta_t(x_t)$$

$$\langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t, \rho_{t-1} \rangle = A_t B_t(z_t(x_t) - \rho_{t-1} z_t(x_{t-1}))$$

$$C_t = K_t(x_t,x_t)^{-1} - B_t^T A_t B_t$$

$$B_t = (h_t(x_t))^T K_t(x_t,x_t)^{-1}$$

$$A_t = \Big((h_t(x_t))^T K_t(x_t,x_t)^{-1}(h_t(x_t))\Big)^{-1}$$

From this form, we can read the (conditional) expectation and covariance of $\boldsymbol{\beta}_t$. We can now integrate with respect to $\boldsymbol{\beta}_t$. We assume a flat prior for $\boldsymbol{\beta}_t$:

$$\boldsymbol{p}\Big[\boldsymbol{\alpha}_t, \sigma_t, \rho_{t-1}|D_t\Big] = \int \boldsymbol{p}\Big[\boldsymbol{\alpha}_t, \sigma_t, \rho_{t-1}, \boldsymbol{\beta}_t|D_t\Big] d\boldsymbol{\beta}_t$$

$$= \exp\Big(-\frac{1}{2\sigma_t^2}\delta_t(x_t)^T C_t \delta_t(x_t)\Big) \Big[\frac{|K_t(x_t,x_t)|}{|A_t|}\Big]^{-1/2}(2\pi\sigma_t^2)^{-\gamma_t}$$

with $\gamma_t = \frac{N_{x_t} - N_{\beta_t}}{2}$, N_{x_t} as the number of pivot points in input vector x_t, and N_{β_t} as the number of basis functions of level t. Apparently, the expectation of $\boldsymbol{\beta}$ is independent of σ and the covariance of $\boldsymbol{\beta}$

is independent of ρ. We will next tend to integrating with respect to ρ_{t-1}. We find ρ_{t-1} in $\delta_t(x_t)$ only and thus as a quadratic form again:

$$-\frac{1}{2\sigma_t^2}\delta_t(x_t)^T C_t \delta_t(x_t) = -\frac{1}{2\sigma_t^2}(c_t - \frac{b_t^2}{a_t}) - \frac{1}{2\sigma_t^2}a_t(\rho_{t-1} - \langle \rho_{t-1} \mid \alpha_t, \sigma_t \rangle)^2$$

$$\langle \rho_{t-1} \mid \alpha_t, \sigma_t \rangle = \frac{b_t}{a_t}$$

$$var(\rho_{t-1} \mid \alpha_t, \sigma_t) = \frac{\sigma_t^2}{a_t}$$

$$c_t = (z_t(x_t))^T C_t z_t(x_t)$$

$$b_t = (z_t(x_t))^T C_t z_{t-1}(x_t)$$

$$a_t = (z_{t-1}(x_t))^T C_t z_{t-1}(x_t)$$

We can now integrate with respect to ρ_{t-1}. We assume a flat prior:

$$p\left[\alpha_t, \sigma_t \mid D_t\right] = \int p\left[\alpha_t, \sigma_t, \rho_{t-1} \mid D_t\right] d\rho_{t-1}$$

$$= \exp\left(-\frac{1}{2\sigma_t^2}\left(c_t - \frac{b_t^2}{a_t}\right)\right)\left[\frac{|K_t(x_t, x_t)|}{|A_t|}\right]^{-1/2}(2\pi\sigma_t^2)^{-\gamma_t}\sqrt{\frac{2\pi\sigma_t^2}{a_t}}$$

The last random variable we can treat analytically is σ_t. With Jeffrey's prior $p\left[\sigma_t \mid \mathcal{C}\right] = \frac{1}{\sigma_t}$, we have

$$p\left[\alpha_t \mid D_t\right] = \int p\left[\alpha_t, \sigma_t \mid D_t\right] d\sigma_t$$

$$= \int \Lambda_t \exp\left(-\frac{\Phi_t}{2\sigma_t^2}\right)(\sigma_t^2)^{-\gamma_t+\frac{1}{2}}\frac{d\sigma_t}{\sigma_t}$$

$$\Phi_t = c_t - \frac{b_t^2}{a_t}$$

$$\Lambda_t = \left[\frac{|K_t(x_t, x_t)|}{|A_t|}\right]^{-1/2}(2\pi)^{-\gamma_t}\sqrt{\frac{2\pi}{a_t}}$$

With the substitution $v = \frac{\Phi_t}{2\sigma_t^2}$, we find this to be a Γ-integral:

$$p\left[\alpha_t \mid D_t\right] = \frac{\Lambda_t}{2}\left(\frac{\Phi_t}{2}\right)^{-\gamma_t+\frac{1}{2}}\underbrace{\int e^{-v}v^{(\gamma_t-\frac{1}{2})-1}dv}_{=\Gamma(\gamma_t-\frac{1}{2})}$$

The moments of σ_t are Γ-integrals as well, and we find

$$\frac{\langle \sigma_t^\nu \mid \alpha_t \rangle}{\langle \sigma_t^0 \mid \alpha_t \rangle} = \sqrt{\Phi_t}^\nu \frac{\Gamma(\gamma_t - \frac{1}{2} - \frac{\nu}{2})}{\Gamma(\gamma_t - \frac{1}{2})}$$

For the above Γ-integral and the σ-moments to exist, we require $\gamma_t - \frac{1}{2} - \frac{\nu}{2} > 0$, i.e., Equation (4d). Note that, in the case of $t = 1$, we have no integration with respect to ρ_0 and thus find one power of $\sqrt{\sigma_1}$ less in the above Γ-integral, i.e., we need to substitute $\gamma_1 - \frac{1}{2} \to \gamma_1$. The constraint is accordingly weakened for $t = 1$, which is due to the absence of the single parameter ρ_0. For most choices of the covariance kernel, we cannot go further analytically.

Now, we have everything we need to marginalize σ_t in the conditional expectations of ρ_{t-1}.

$$\left\langle \rho_{t-1}^v \mid \boldsymbol{\alpha}_t \right\rangle = \int \left\langle \rho_{t-1}^v \mid \boldsymbol{\alpha}_t, \sigma_t \right\rangle p\left[\boldsymbol{\alpha}_t, \sigma_t \mid D_t \right] d\sigma_t$$

which is now relatively easy.

$$\left\langle \rho_{t-1}^1 \mid \boldsymbol{\alpha}_t \right\rangle = \frac{b_t}{a_t}$$

$$\left\langle \rho_{t-1}^2 \mid \boldsymbol{\alpha}_t \right\rangle = \frac{\left\langle \sigma_t^2 \mid \boldsymbol{\alpha}_t \right\rangle}{a_t} + \left(\frac{b_t}{a_t} \right)^2$$

The expected mean function parameters are simply

$$\left\langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t \right\rangle = \int \left\langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t, \rho_{t-1}, \sigma_t \right\rangle p\left[\boldsymbol{\alpha}_t, \rho_{t-1}, \sigma_t \mid D_t \right] d\rho_{t-1} d\sigma_t$$

$$= A_t B_t(z_t(x_t) - \left\langle \rho_{t-1} \mid \boldsymbol{\alpha}_t \right\rangle z_t(x_{t-1}))$$

The variance we read from the quadratic form in $\boldsymbol{\beta}_t$ is as follows:

$$var(\beta_t^{(k)} \mid \boldsymbol{\alpha}_t) = \left\langle \sigma_t^2 \mid \boldsymbol{\alpha}_t \right\rangle \cdot (A_t)_{kk}$$

with $(A_t)_{kk}$ as the kth diagonal element of A_t.

Appendix A.2. Predictive Mean

We rewrite $\bar{\delta}_t$ such that

$$\bar{\delta}_t(x) = \underbrace{K_t(x, x_t) K_t(x_t, x_t)^{-1}(z_t(x_t) - \rho_{t-1} z_t(x_{t-1}))}_{u_t} + \underbrace{\left(h_t(x) - K_t(x, x_t) K_t(x_t, x_t)^{-1} h_t(x_t) \right)}_{g_t} \boldsymbol{\beta}_t$$

The first term, u_t, assumes a rather simple form and only depends on ρ_{t-1} and $\boldsymbol{\alpha}$. Since we found $p\left[\boldsymbol{\rho}, \boldsymbol{\alpha} \mid D \right] = \prod_k p\left[\rho_k, \boldsymbol{\alpha}_t \mid D_k \right]$ previously, and each ρ_k occurs linearly. We find

$$\left\langle \prod_k \rho_k \Big| \boldsymbol{\alpha} \right\rangle = \prod_k \left\langle \rho_k \mid \boldsymbol{\alpha}_t \right\rangle$$

$$\Rightarrow \left\langle u_t \prod_{l=t}^{N_t-1} \rho_l \Big| \boldsymbol{\alpha} \right\rangle = K_t(x, x_t) K_t(x_t, x_t)^{-1} \left(z_t(x_t) - \left\langle \rho_{t-1} \mid \boldsymbol{\alpha}_t \right\rangle z_t(x_{t-1}) \right) \prod_{l=t}^{N_t-1} \left\langle \rho_l \mid \boldsymbol{\alpha}_t \right\rangle$$

The second term is rather trivial.

$$\left\langle g_t \boldsymbol{\beta}_t \prod_{l=t}^{N_t-1} \rho_l \mid \boldsymbol{\alpha}_t \right\rangle = g_t \left\langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t \right\rangle \prod_{l=t}^{N_t-1} \left\langle \rho_l \mid \boldsymbol{\alpha}_t \right\rangle$$

We have thus shown the predictive mean in Equation (4).

Appendix A.3. Predictive Covariance

We easily find

$$\left(\sum_{t=1}^{N_t} \delta_t(x)(x) \prod_{l=t}^{N_t-1} \rho_l \right) \left(\sum_{t=1}^{N_t} \delta_t(x)(x) \prod_{l=t}^{N_t-1} \rho_l \right)^T = \sum_{tt'}^{N_t} \delta_t(x) \delta_{t'}(x)^T \prod_{ll'}^{N_t-1} \rho_l \rho_{l'}$$

Due to the assumption of independence of the levels or each levels' parameters, formally $\delta_t(x) \perp z_{t-1}(x)$, the cross-terms mixing different levels vanish in the expectation value and it suffices to reduce

the double sum to a single sum. Further, we can factor the expectation of the product of ρ_l and the expectation of the sum.

$$\left\langle \sum_{tt'}^{N_t} \delta_t(x)\delta_{t'}(x)^T \prod_{ll'}^{N_t-1} \rho_l\rho_{l'} \right\rangle = \left\langle \sum_{t=1}^{N_t} \delta_t(x)\delta_t(x)^T \right\rangle \prod_{l=t}^{N_t-1} \left\langle \rho_l^2 \right\rangle$$

The first factor on the right-hand side can be expressed via the GP's posterior covariance and mean:

$$\sum_{t=1}^{N_t} \left\langle \delta_t(x)\delta_t(x)^T \right\rangle = \sum_{t=1}^{N_t} \Sigma_t(x, x) + \left\langle \delta_t(x) \right\rangle \left\langle \delta_t(x) \right\rangle^T$$

where

$$\Sigma_t(x, x) = \sigma_t^2 \left(K_t(x, x) - K_t(x, x_t)K_t(x_t, x_t)^{-1}K_t(x_t, x) \right)$$

is the GPs' posterior covariance and the predictive mean we already know as follows:

$$\left\langle \delta_t(x) \mid \boldsymbol{\alpha}_t \right\rangle = \left\langle u_t \mid \boldsymbol{\alpha}_t \right\rangle + g_t \left\langle \boldsymbol{\beta}_t \mid \boldsymbol{\alpha}_t \right\rangle$$

We have thus shown everything we claimed.

References

1. Ghanem, R.G.; Owhadi, H.; Higdon, D. *Handbook of Uncertainty Quantification*; Springer: Basel, Switzerland, 2017.
2. Haylock, R. Bayesian Inference about Outputs of Computationally Expensive Algorithms with Uncertainty on the Inputs. Ph.D. Thesis, Universtity of Nottingham, Nottingham, UK, May 1997.
3. O'Hagan, A.; Kennedy, M.C.; Oakley, J.E. Uncertainty Analysis and Other Inference Tools for Complex Computer Codes. 1999. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.446&rep=rep1&type=pdf (accessed on 26 December 2019).
4. O'Hagan, A. Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 1290–1300. [CrossRef]
5. Wiener, N. The Homogeneous Chaos. *Am. J. Math.* **1938**, *60*, 897–936. [CrossRef]
6. Ghanem, R.G.; Spanos, P.D. *Stochastic Finite Elements: A Spectral Approach*; Springer: Basel, Switzerland, 1991.
7. Xiu, D.; Karniadakis, G.E. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **2005**, *27*, 1118–1139. [CrossRef]
8. O'Hagan, A. Polynomial Chaos: A Tutorial and Critique from a Statistician's Perspective. Available online: http://tonyohagan.co.uk/academic/pdf/Polynomial-chaos.pdf (accessed on 25 June 2019).
9. O'Hagan, A. Curve Fitting and Optimal Design for Prediction. *J. R. Stat. Soc. Ser. B Methodol.* **1978**, *40*, 1–42. [CrossRef]
10. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
11. Bishop, C. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, UK, 1996.
12. MacKay, D.J. *Information Theory, Inference, and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
13. Kennedy, M.C.; O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **2000**, *87*, 1–13, doi:10.1093/biomet/87.1.1. [CrossRef]
14. Koutsourelakis, P.S. Accurate Uncertainty Quantification using inaccurate Computational Models. *SIAM J. Sci. Comput.* **2009**, *31*, 3274–3300, doi:10.1137/080733565. [CrossRef]
15. Perdikaris, P.; Raissi, M.; Damianou, A.; Lawrence, N.D.; Karniadakis, G.E. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2017**, *473*, 20160751, doi:10.1098/rspa.2016.0751. [CrossRef] [PubMed]

16. Eck, V.G.; Donders, W.P.; Sturdy, J.; Feinberg, J.; Delhaas, T.; Hellevik, L.R.; Huberts, W. A guide to uncertainty quantification and sensitivity analysis for cardiovascular applications. *Int. J. Numer. Methods Biomed. Eng.* **2016**, *32*, e02755, doi:10.1002/cnm.2755. [CrossRef] [PubMed]

17. Biehler, J.; Gee, M.W.; Wall, W.A. Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme. *Biomech. Model. Mechanobiol.* **2015**, *14*, 489–513, doi:10.1007/s10237-014-0618-0. [CrossRef] [PubMed]

18. Zienkiewicz, O.C.; Taylor, R.L.; Zhu, J.Z. *The Finite Element Method: Its Basis and Fundamentals*; Elsevier: Amsterdam, The Netherlands, 1967.

19. Miller, J.C.; Horvath, S.M. Impedance Cardiography. *Psychophysiology* **1978**, *15*, 80–91, doi:10.1111/j.1469-8986.1978.tb01340.x. [CrossRef] [PubMed]

20. Reinbacher-Köstinger, A.; Badeli, V.; Biro, O.; Magele, C. Numerical Simulation of Conductivity Changes in the Human Thorax Caused by Aortic Dissection. *IEEE Trans. Magn.* **2019**, *55*, 1–4, doi:10.1109/tmag.2019.2895418. [CrossRef]

21. Humphrey, J.D. *Cardiovascular Solid Mechanics*; Springer: Basel, Switzerland, 2002; p. 758.

22. Khan, I.A.; Nair, C.K. Clinical, diagnostic, and management perspectives of aortic dissection. *Chest* **2002**, *122*, 311–328, doi:10.1378/chest.122.1.311. [CrossRef] [PubMed]

23. Gabriel, C.; Gabriel, S.; Corthout, E.C. The dielectric properties of biological tissues: I. Literature survey. *Phys. Med. Biol.* **1996**, *41*, 2231–2249. [CrossRef] [PubMed]

24. Sivia, D.; Skilling, J. *Data Analysis: A Bayesian Tutorial*; Oxford University Press: Oxford, UK, 2006.

25. Skilling, J. Nested sampling for general Bayesian computation. *Bayesian Anal.* **2006**, *1*, 833–859, doi:10.1214/06-BA127. [CrossRef]

26. Le Gratiet, L.; Garnier, J. Recursive Co-Kriging Model for Design of Computer Experiments With Multiple Levels of Fidelity. *Int. J. Uncertain. Quantif.* **2014**, *4*, 365–386, doi:10.1615/Int.J.UncertaintyQuantification.2014006914. [CrossRef]

27. Ranftl, S.; Melito, G.M.; Badeli, V.; Reinbacher-Köstinger, A.; Ellermann, K.; von der Linden, W. On the Diagnosis of Aortic Dissection with Impedance Cardiography: A Bayesian Feasibility Study Framework with Multi-Fidelity Simulation Data. *Proceedings* **2019**, *33*, 24. [CrossRef]

28. Alastruey, J.; Xiao, N.; Fok, H.; Schaeffter, T.; Figueroa, C.A. On the impact of modelling assumptions in multi-scale, subject-specific models of aortic haemodynamics. *J. R. Soc. Interface* **2016**, *13*, doi:10.1098/rsif.2016.0073. [CrossRef] [PubMed]

29. Comsol A.B., Stockholm, S. Comsol Multiphysics Version 5.4. Available online: http://www.comsol.com (accessed on 4 June 2019).

30. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Inferring solutions of differential equations using noisy multi-fidelity data. *J. Comput. Phys.* **2017**, *335*, 736–746, doi:10.1016/j.jcp.2017.01.060. [CrossRef]

31. Albert, C.G. Gaussian processes for data fulfilling linear differential equations. *Proceedings* **2019**, *33*, 5. [CrossRef]