Improved statistical emulation for a soft-tissue cardiac mechanical model

David Dalton[1], Dirk Husmeier[1]

[1] Department of Mathematics and Statistics, University of Glasgow, UK

E-mail for correspondence: d.dalton.1@research.gla.ac.uk

Abstract: This paper outlines a new approach to emulation based parameter inference in a cardiac mechanic model of the left ventricle (LV) of the heart that allows for prediction uncertainty to be accounted for. The emulation is performed using Gaussian processes, which are designed to build on the results of previous research in this area. This approach yields more accurate parameter estimates than previously reported in the literature.

Keywords: Cardiac Modelling; Holzapfel-Ogden Law; Statistical Emulation

1 Introduction

The Holzapfel-Ogden (HO) model (Holzapfel and Ogden, 2009) is a system of coupled partial differential equations that define the stretch-strain dependence of the inner tissue of the LV, known as the endocardium. The model depends on various material parameters, for example those which are related to the stiffness of the cardiac fibres. Interest is in determining these parameters for their potential to aid in the diagnosis of cardiac defects, however they can only be directly measured *in-vivo* by invasive procedures. An alternative, non-invasive approach to inferring the parameters which has potential for use in clinical decision support is to use magnetic resonance imaging (MRI). This is done by taking MRI scans of a subject's LV at end diastole, to determine the myocardium responses that are modelled by the HO law. The material properties can then be estimated as those values which minimise the discrepancy between the observed myocardium response, and the response predicted by the model. The diagnostic value of this approach has been shown in previous work (Gao *et al.* 2017). The problem, however, is that the HO cardio-mechanical equations describing

This paper was published as a part of the proceedings of the 35th International Workshop on Statistical Modelling (IWSM), Bilbao, Spain, 19–24 July 2020.

the kinematics of the heart have no closed form solutions. Instead, numerical procedures based on finite element discretisation are required, which typically take on the order of 15 minutes per evaluation on a high performance computer. Since solving for the material properties by iterative optimisation methods may require hundreds or thousands of such evaluations, the approach is rendered unsuitable as a real-time clinical decision support tool.
A number of methods exists which can be used to overcome this problem, one of which is *statistical emulation*.

2 Statistical Emulation

Statistical emulation involves approximating a computationally expensive model, referred to as a simulator, with a much cheaper surrogate model, known as an emulator. This is done by first choosing a set of points to cover the input region of interest, and then running the simulator $\mathbf{f}$ from each point. This creates a dataset of input-output pairs

$$\mathcal{D} = \{(\mathbf{x}_i, \mathbf{f}(\mathbf{x}_i))_{i=1}^N\} \tag{1}$$

on which the surrogate model $\hat{\mathbf{f}}$ is trained. While the creation of a dataset in this manner for the HO law is extremely computationally expensive, all simulations can be done in advance of clinical deployment. In clinic, $\hat{\mathbf{f}}$ can be used in place of $\mathbf{f}$ in the parameter optimisation routine, allowing estimates to be obtained in real time. The inputs required for the HO model are a LV geometry, $\mathcal{H}$, and a four dimensional parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^\mathsf{T}$, which we are interested in inferring. In this study, we considered a fixed LV geometry, and then used a Sobol sequence (Fang *et al.*, 2006) to generate $10,100$ parameter configurations within the physiologically realistic boundaries $(0.1, 5)^4$. The simulator was then run from each point, and 25 outputs were extracted: circumferential strains at $K = 24$ regions along the endocardial surface, and the LV volume, all measured at end-diastole.
Having created the dataset, the approach for constructing the surrogate model $\hat{\mathbf{f}}$ must be considered. Gaussian process regression is a Bayesian nonparametric approach that is commonly used for emulation (Kennedy and O'Hagan, 2001). A Gaussian process (GP) is a stochastic process where any finite collection of random variables from the process are Gaussian distributed. GPs can be used for regression to define a prior directly over a space of functions

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{2}$$

where the GP is completely specified by its mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$. Given a finite set of known training points and unknown test points, the GP marginalises to a multivariate Gaussian distribution, with mean and covariance found by evaluation of m and k at

the given points. Standard rules for conditional Gaussian distributions can then be used to find the posterior distribution of the test points, given the training points. For further details, the reader is directed to (Rasmussen and Williams, 2006).

One drawback of the GP modelling framework is that training and prediction times grow with the size of the dataset under consideration. Local Gaussian process regression (Gramacy and Apley, 2015) is an approach which can alleviate this complexity. With local GPs, a prediction is made at a given point using a GP trained on only the k nearest neighbours of the point in the training data. Initial work on the HO simulation data we analyse in this paper demonstrated the effectiveness of an emulator comprised of 25 independent local GPs, one for each dimension of the simulator output (Davies *et al.*, 2019). The problem with the local GP approach in our application context is that, as we adjust the input parameter values during the optimisation routine, the local neighbourhood will also change. This in turn means that the emulator will need to be refit at each iteration. Further work however has shown that a multivariate-output local GP emulator trained on the nearest neighbours of a test point in *output space* can accurately model the HO law (Noe *et al.*, 2019). The advantage of considering neighbours in output space is that this neighbourhood does not change during the parameter optimisation, meaning that the emulator only has to be fit once.

Given the above results, in this paper we consider an emulator made up of 25 independent local GPs trained on $k = 200$ neighbours in output space. The GPs were fit with linear mean functions and with squared exponential kernel function, where the length scales for each input dimension were allowed to vary. This is known as an ARD (Automatic Relevance Determination) prior in the Machine Learning community (Rasmussen and Williams, 2006). Although the data under consideration is deterministic, a small nugget term (10^{-6}) was added for reasons of numerical stability.

3 Maximum Likelihood Parameter Inference

Our objective is to find the optimal parameter estimates which minimise the loss between measured data, and the values predicted by the emulator. In what follows, we denote the measured quantities, after non-dimensionalisation, by

$$\mathbf{y} = (y_0, y_1, \ldots, y_K)^\mathsf{T} \tag{3}$$

where y_0 is the non-dimensionalised LV volume, and $y_1, \ldots, y_{24}$ are the non-dimensionalised circumferential strains. The corresponding outputs from the GP emulator, which depend on the cardio-mechanic parameters $\boldsymbol{\theta}$ and the LV geometry, $\mathcal{H}$, are denoted:

$$\hat{\mathbf{f}}(\boldsymbol{\theta}, \mathcal{H}) = \left(\hat{\mathrm{f}}_0(\boldsymbol{\theta}, \mathcal{H}), \hat{\mathrm{f}}_1(\boldsymbol{\theta}, \mathcal{H}), \ldots, \hat{\mathrm{f}}_K(\boldsymbol{\theta}, \mathcal{H})\right)^\mathsf{T} \tag{4}$$

In previous work, described in Section 2, the cardio-mechanic parameters $\boldsymbol{\theta}$ for a given LV geometry $\mathcal{H}$ were estimated by minimizing the L2 norm of the difference between $\mathbf{y}$ and $\hat{\mathbf{f}}$:

$$E(\boldsymbol{\theta}, \mathcal{H}) \;=\; ||\mathbf{y} - \hat{\mathbf{f}}(\boldsymbol{\theta}, \mathcal{H})||^2 \;=\; \sum_{i=0}^{K} \Big(y_i - \hat{\mathrm{f}}_i(\boldsymbol{\theta}, \mathcal{H})\Big)^2 \tag{5}$$

where each output $\hat{\mathrm{f}}_i(\boldsymbol{\theta}, \mathcal{H})$ was set to the corresponding posterior GP mean $\mu_i(\boldsymbol{\theta}, \mathcal{H})$. Under the assumption that the measurement noise is iid additive Gaussian with variance σ_m^2:

$$\mathbf{y} \;=\; \hat{\mathbf{f}}(\boldsymbol{\theta}, \mathcal{H}) + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_m^2 \mathbf{I}) \tag{6}$$

we obtain the log likelihood, conditional on the emulator output $\hat{\mathbf{f}}(\boldsymbol{\theta}, \mathcal{H})$:

$$\log p\Big(\mathbf{y}|\hat{\mathbf{f}}(\boldsymbol{\theta}, \mathcal{H})\Big) \;=\; \frac{-1}{2\sigma_m^2} \sum_{i=0}^{K} \Big(y_i - \hat{\mathrm{f}}_i(\boldsymbol{\theta}, \mathcal{H})\Big)^2 - \frac{K+1}{2} \log(2\pi\sigma_m^2) \tag{7}$$

Maximizing this conditional likelihood with respect to $\boldsymbol{\theta}$, for a given LV geometry $\mathcal{H}$, is equivalent to minimizing the original objective function (5), where again the emulator outputs are set to the posterior GP mean values. However, a disadvantage of this approach is that the uncertainty of the emulator, naturally predicted by the GP variance, is not taken into consideration. To rectify this, we can compute the marginal likelihood by integrating over the emulator outputs

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}, \sigma_m^2) = \int p(\mathbf{y}|\hat{\mathbf{f}}, \sigma_m^2) p(\hat{\mathbf{f}}|\boldsymbol{\theta}, \mathcal{H}) d\hat{\mathbf{f}} = \prod_{i=0}^{K} \int p(y_i|\hat{\mathrm{f}}_i, \sigma_m^2) p(\hat{\mathrm{f}}_i|\boldsymbol{\theta}, \mathcal{H}) d\hat{\mathrm{f}}_i \tag{8}$$

where conditional independence between the outputs has been assumed. The two probability distributions under the integral are given by

$$\begin{aligned} p(y_i|\hat{\mathrm{f}}_i, \sigma_m^2) &= \mathcal{N}(y_i|\hat{\mathrm{f}}_i, \sigma_m^2) \\ p(\hat{\mathrm{f}}_i|\boldsymbol{\theta}, \mathcal{H}) &= \mathcal{N}\Big(\hat{\mathrm{f}}_i|\mu_i(\boldsymbol{\theta}, \mathcal{H}), \sigma_i^2(\boldsymbol{\theta}, \mathcal{H})\Big) \end{aligned} \tag{9}$$

where $\mu_i(\boldsymbol{\theta}, \mathcal{H})$ is the mean of the i^{th} GP emulator, and $\sigma_i^2(\boldsymbol{\theta}, \mathcal{H})$ is its variance. The integral in (8) is a therefore a standard Gaussian integral with closed-form solution

$$p\left(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}, \sigma_m^2\right) = \prod_{i=0}^{K} \mathcal{N}\left(y_i|\mu_i(\boldsymbol{\theta}, \mathcal{H}), \sigma_m^2 + \sigma_i^2(\boldsymbol{\theta}, \mathcal{H})\right) \tag{10}$$

which gives

$$\log p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}, \sigma_m^2) = -\frac{1}{2} \sum_{i=0}^{K} \left\{ \frac{\Big(y_i - \mu_i(\boldsymbol{\theta}, \mathcal{H})\Big)^2}{[\sigma_m^2 + \sigma_i^2(\boldsymbol{\theta}, \mathcal{H})]} + \log\Big(2\pi[\sigma_m^2 + \sigma_i^2(\boldsymbol{\theta}, \mathcal{H})]\Big) \right\} \tag{11}$$

as a better objective function to optimise.

4 Results and Discussion

In the absence of large quantities of real patient data, we instead reserved the final 100 points of our simulated dataset as an independent test set on which to quantify the difference in parameter estimation accuracy when emulation uncertainty is accounted for. Using local GP emulators trained on the remaining simulation data, we used iterative optimisation methods to estimate each of the independent test parameter values with loss functions (5) and (11) respectively. By then evaluating the mean squared error (MSE) between the known true values and the estimated values, we obtain a list of 100 errors for each loss function. The median of these lists is displayed in Table 1, alongside the first and third quartiles.

TABLE 1. Test Set MSE (Parameter Space)

Emulator	Loss Function	25^{th} Perc.	Median	75^{th} Perc..
Local GP	Equation (5)	9.2×10^{-8}	3.1×10^{-7}	3.1×10^{-6}
Local GP	Equation (11)	7.1×10^{-8}	3.0×10^{-7}	2.0×10^{-6}

The results in Table 1 quantify the improvement in parameter estimation accuracy that can be achieved by accounting for emulation uncertainty. The gain in performance is slight, which may be due to the prediction variance for each output dimension being quite similar. Of note is that our parameter estimation accuracy has improved by more than one order of magnitude over the best results from the literature, particularly as a consequence of a decreased nugget term. This accuracy is visualised in Figure 1, which plots the 100 out of sample test parameter values, broken down into each dimension respectively, versus the corresponding values predicted when using loss function (11). We see extremely good agreement between the true and predicted values across the entire parameter space.

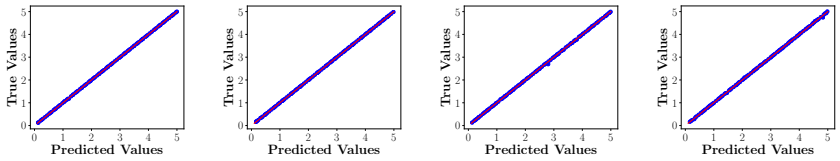

FIGURE 1. Plot of the true test set parameter values versus those predicted by the local GP emulator using loss function (11), when broken down into each dimension. Points lying on the red lines of unit slope indicate perfect prediction accuracy.

The limitation of the analysis presented here is that we have performed emulation for a fixed, known LV geometry $\mathcal{H}$. To be of clinical use however,

the emulator must be able to account for the unique LV geometry of a given patient. Therefore, the construction of emulators that can account for LV geometry variations will be the remit of further work.

Acknowledgments: This work was carried out as part of the SoftMech^{MP} project, funded by EPSRC, grant reference number EP/S030875/1. Dirk Husmeier was funded by the Royal Society of Edinburgh, grant reference number 62335.

References

Davies, V., Noè, U., Lazarus, A., Gao, H., Macdonald, B., Berry, C., Luo, X. and Husmeier, D. (2019). Fast parameter inference in a biomechanical model of the left ventricle by using statistical emulation *Journ. of the R. Stat. Soc., Series C (Appl. Statistics)*, **68**, Part 5, 1555–1576

Fang K., Li R., Sudjianto A. (2006). *Design and modeling for computer experiments*. London, UK: Chapman Hall/CRC

Gao, H., Li, W., Cai, L., Berry, C. and Luo, X. (2015). Parameter estimation in a Holzapfel–Ogden law for healthy myocardium. *J. Engng Math.*, **95**, 231–248.3)

Gramacy R.B., Apley D.W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Stat.*, **24**, 561 – 578. (doi:10.1080/ 10618600.2014.914442)

Holzapfel, G. A. and Ogden, R. W. (2009). Constitutive modelling of passive myocardium: a structurally based framework for material characterization *Phil. Trans. R. Soc. A,*, **367**, 3445–3475.

Kennedy M.C., O'Hagan A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. B (Statistical Methodology)*, **63**, 425– 464. (doi:10.1111/rssb. 2001.63.issue-3)

Noè, U., Lazarus, A., Gao, H., Davies, V., Macdonald, B., Mangion, K., Berry, C., Luo, X. and Husmeier, D. (2019). Gaussian process emulation to accelerate parameter estimation in a mechanical model of the left ventricle: a critical step towards clinical end-user relevance. *J. R. Soc. Interface*, **16**, 20190114. http://dx.doi.org/10.1098/rsif.2019.0114

Rasmussen C.E., Williams C.K.I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.