Gaussian Process Regression for Computer Model Emulation

Jaeik Jeon*
Advisor: Tim Waite†

School of Mathematics
The University of Manchester

July 24, 2019

Abstract

We consider a full Bayesian approach for the estimation of the parameters of a Gaussian process and its application to Computer model emulation. Gu et al. (2017) suggests certain parameterizations of the covariance structure that have certain advantages when finding the posterior modes. However it turns out that such parametrization may not be appropriate for sampling from the posterior of range parameters using the Metropolis-Hasting algorithm, especially when there are inert inputs. In this paper, we present one possible parametrization of correlation structure for the full Bayesian approach, and the method is applied to an emulator of Multiphase Flow with Interphase eXchanges (MFIX) which is a computational fluid dynamics model of a bubbling fluidized bed.

1 Introduction

Various scientific processes described by mathematical models typically require computer code to implement and they are often time-consuming. In order to use such models to predict

*E-mail: jaeik@futureworkers.net
†E-mail: timothy.waite@manchester.ac.uk

the behaviour of the processes, calibration of computer models may be required (Kennedy & O'Hagan, 2001). Here calibration refers to an estimation of the parameters that cannot be observable contained in the computer model. However this estimating process typically requires a large number of computer model evaluations, which is infeasible due to the large computational cost. Hence we require a fast surrogate model, known as an emulator, that approximates the computer model. In this paper we discuss how to construct such an emulator using Gaussian processes.

Gaussian process can be used as a prior for the output of a computationally expensive deterministic computer model (Oakley, 1999). The term "prior" refers to a prior distribution in a Bayesian context for which we update our prior information on the model with given datas. The prior information will be updated using the Bayes' rule to get the posterior distribution i.e.

$$f(\boldsymbol{\theta}|\text{data}) \propto f(\boldsymbol{\theta}) f(\text{data}|\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta}$ denotes the parameters in the Gaussian process model. The updated information of the model will be reported using Bayesian predictive inference which marginalizes out all the parameters i.e.

$$p(x_{new}|\mathbf{x}) = \int_{\boldsymbol{\theta}\in\Theta} p(x_{new}|\boldsymbol{\theta}) p(\theta|\mathbf{x}) d\boldsymbol{\theta} \tag{2}$$

in order to take uncertainties involved in the parameters into account. When the integration is analytically intractable, either the Metropolis-Hasting algorithm for the full Bayesian approach, or the empirical Bayes method which is done by simply plugging-in a single estimate will be used. There will be further discussion about these two approaches in Section 4.

Section 2 provides step-by-step explanations to the formulation of the predictive distribution in (2). Section 3 applies the method to Multiphase Flow with Interphase eXchanges (MFIX), which is a computational fluid dynamics model of a bubbling fluidized bed. Section 4 discusses the adequacy of full Bayesian approach over empirical Bayes. The notation in this paper follows Gu (2017).

2 Gaussian Process Regression

2.1 Gaussian Process

A Gaussian process is a collection of random variables, $\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$, any finite number of which have a joint Gaussian distribution. $\mathcal{X}$ is some index set, and in this paper, $\mathcal{X}$ is $\mathbb{R}^n$.

We consider a real-valued Gaussian process $y(\cdot) \sim \mathcal{GP}(\mathbf{h}(\mathbf{x})\boldsymbol{\theta} = \sum_{t=1}^{q} h_t(\mathbf{x})\theta_t, \sigma^2 c(\mathbf{x}, \mathbf{x}'))$ on a p-dimensional input domain $\mathcal{X}$ where $\mathbf{h}(\mathbf{x})$ is a q-dimensional row vector of basis functions, $\boldsymbol{\theta}$ is an unknown regression parameter and $c(\mathbf{x}, \mathbf{x}')$ is a correlation function.

Let $\mathbf{x}^D = \{\mathbf{x}_1^D, \ldots, \mathbf{x}_n^D\}$, which is chosen design points, and $\mathbf{x}^*$ denotes a new input. Recall that conditional of a joint Gaussian is also Gaussian, i.e. if $(\mathbf{x}_1, \mathbf{x}_2)^T \sim N\left((\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$, then $p(\mathbf{x}_1|\mathbf{x}_2) = N(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Using this we obtain

$$y(\mathbf{x}^*)|\mathbf{y}^D, \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2 \sim N(m^*(\mathbf{x}^*), \sigma^2 c^*), \tag{3}$$

where

$$m^*(\mathbf{x}^*) = \mathbf{h}(\mathbf{x}^*)\boldsymbol{\theta} + \mathbf{r}^T(\mathbf{x}^*)\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}^D)\boldsymbol{\theta}), \tag{4}$$

$$c^* = c(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{r}^T(\mathbf{x}^*)\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}^*) \tag{5}$$

with $\mathbf{r}(\mathbf{x}^*) = (c(\mathbf{x}^*, \mathbf{x}_1^D), \ldots, c(\mathbf{x}^*, \mathbf{x}_n^D))^T$, $\mathbf{R}$ is a correlation matrix of inputs $\mathbf{x}^D$, and $\mathbf{h}(\mathbf{x}^D)$ is a $n \times q$ data matrix.

Furthermore, for any $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T \in \mathcal{X}$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})^T \in \mathcal{X}$, $i = 1, \ldots, n$, we are going to consider the Gaussian process where the correlation function is a special case of power exponential correlation function

$$c(\mathbf{x}_i, \mathbf{x}_j) = \prod_{l=1}^{p} \exp\left(\left(\frac{|x_{il} - x_{jl}|}{\gamma_l}\right)^{1.9}\right). \tag{6}$$

The power 1.9 is a default setting of **RobustGasp**, and this parametrization was from (Gu et al., 2018).

2.2 Predictive Inference with Unknown $\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}$

2.2.1 Known $\boldsymbol{\gamma}$

Our aim is to find a distribution $y(\mathbf{x}^*)|\mathbf{y}$ by integrating $\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}$ out. First, assume $\boldsymbol{\gamma}$ to be known. We know that

$$p(\mathbf{y}^D|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}) = \frac{|\mathbf{R}|^{-1/2}}{(2\pi)^{n/2}(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}^D - \mathbf{h}(\mathbf{x}^D)\boldsymbol{\theta})^T\mathbf{R}^{-1}(\mathbf{y}^D - \mathbf{h}(\mathbf{x}^D)\boldsymbol{\theta})\right), \tag{7}$$

where $\mathbf{h}(\mathbf{x}^D)$ is the $n \times q$ basis matrix with the (i, j) entry $h_j(\mathbf{x}_i)$.

We consider the reference prior of the form:

$$\pi(\boldsymbol{\theta}, \sigma^2) \propto \frac{1}{\sigma^2}. \tag{8}$$

In order to derive the distribution of $y(\mathbf{x}^*)|\mathbf{y}^D, \boldsymbol{\gamma}$, we rewrite the equation (7) as

$$p(\mathbf{y}^D|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}) = \frac{|\mathbf{R}|^{-1/2}}{(2\pi)^{n/2}(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + S^2\right), \tag{9}$$

where

$$\hat{\boldsymbol{\theta}} = (\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D))^{-1}\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{y}^D, \tag{10}$$

and

$$S^2 = (\mathbf{y}^D)^T\mathbf{Q}\mathbf{y}^D \tag{11}$$

with $\mathbf{Q} = \mathbf{R}^{-1}\mathbf{P}_\mathbf{R}$ where

$$\mathbf{P}_\mathbf{R} = \mathbf{I}_n - \mathbf{h}(\mathbf{x}^D)\{\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)\}^{-1}\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}. \tag{12}$$

Then,

$$\begin{aligned} p(\boldsymbol{\theta}, \sigma^2|\mathbf{y}^D, \boldsymbol{\gamma}) &\propto p(\mathbf{y}^D|\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\theta}, \sigma^2) \\ &\propto (\sigma^{-2})^{\frac{n+2}{2}} \exp\left(-\frac{1}{2\sigma^2}((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + S^2)\right) \\ &= (\sigma^{-2})^{\frac{n-q}{2}+1+\frac{q}{2}} \exp\left(-\frac{1}{2\sigma^2}((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + S^2)\right) \end{aligned} \tag{13}$$

Note that $p(\boldsymbol{\theta}, \sigma^2|\mathbf{y}, \boldsymbol{\gamma})$ is proportional to density of a normal-inverse gamma distribution,

which implies

$$\boldsymbol{\theta}|\sigma^2, \mathbf{y}, \boldsymbol{\gamma} \sim N(\hat{\boldsymbol{\theta}}, \sigma^2(\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D))^{-1}), \tag{14}$$

and

$$\sigma^2|\mathbf{y}, \boldsymbol{\gamma} \sim IG(\frac{n-q}{2}, \frac{S^2}{2}) \tag{15}$$

(Banerjee, 2008). Then from the result of Oakely(1999, p.14), by combining (3) and (13) and integrating out $\boldsymbol{\theta}$ we obtain

$$y(\mathbf{x}^*)|\mathbf{y}^D, \sigma^2, \boldsymbol{\gamma} \sim N(m^{**}, \sigma^2 c^{**}), \tag{16}$$

where

$$m^{**} = \mathbf{h}^T(\mathbf{x}^*)\hat{\boldsymbol{\theta}} + \mathbf{r}^T(\mathbf{x}^*)\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}^D)\hat{\boldsymbol{\theta}}), \tag{17}$$

and

$$c^{**} = c^* + \left(\mathbf{h}^T(\mathbf{x}^*) - \mathbf{r}^T(\mathbf{x}^*)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)\right)\left(\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)\right)^{-1}\left(\mathbf{h}^T(\mathbf{x}^*) - \mathbf{r}^T(\mathbf{x}^*)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)\right)^T \tag{18}$$

Furthermore, since

$$\mathbf{y}^*|\mathbf{y} \sim t_{2a}(\mu, \frac{b}{a-1}\Sigma) \tag{19}$$

if

$$\sigma^2|\mathbf{y} \sim IG(a, b) \quad \text{and} \quad \mathbf{y}^*|\mathbf{y}, \sigma^2 \sim N(\mu, \sigma^2\Sigma) \tag{20}$$

(Englezou, 2018), combining (15) and (16) we find that

$$y(\mathbf{x}^*)|\mathbf{y}^D, \boldsymbol{\gamma} \sim t_{n-q}(m^{**}, \frac{S^2}{n-q-2}c^{**}). \tag{21}$$

2.2.2 Estimating the Range Parameter $\boldsymbol{\gamma}$

Now assume $\boldsymbol{\gamma}$ to be unknown. For the prior for $\boldsymbol{\gamma}$, we can use

$$\pi^{JR}(\boldsymbol{\gamma}) = \left(\sum_{l=1}^{p} C_l\beta_l\right)^{0.2} \exp\left(-b\sum_{j=1}^{p} C_l\beta_l\right), \tag{22}$$

with $\beta_l = \frac{1}{\gamma_l}$, C_l equal to the mean of $|x_{il}^D - x_{jl}^D|$, for $1 \leq i, j \leq n$, $i \neq j$, and $b = n^{-1/p}(0.2+p)$, for $l = 1, \ldots, p$.

This prior was proposed by Gu (2018) which efficiently approximates a reference prior for

the Gaussian process model with the product correlation suggested in Paulo (2005). Then, the reference prior for $\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}$ becomes, by independence,

$$\pi(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}) = \pi(\boldsymbol{\theta}, \sigma^2) \times \pi^{JR}(\boldsymbol{\gamma}) \propto \frac{\pi^{JR}(\boldsymbol{\gamma})}{\sigma^2}. \tag{23}$$

Then the posterior of $\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}$ is given by

$$p(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}|\mathbf{y}^D) \propto \frac{|\mathbf{R}|^{-1/2}}{(\sigma^2)^{\frac{n+2}{2}}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}^D - \mathbf{h}(\mathbf{x}^D)\boldsymbol{\theta})^T\mathbf{R}^{-1}(\mathbf{y}^D - \mathbf{h}(\mathbf{x}^D)\boldsymbol{\theta})\right) \times \pi^{JR}(\boldsymbol{\gamma}) \tag{24}$$

Integrating out $\boldsymbol{\theta}$ and σ^2 gives (Lopes, 2011)

$$\begin{aligned} p(\boldsymbol{\gamma}|\mathbf{y}^D) &= \int_{\mathbb{R}^q \times \mathbb{R}} p(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\gamma}|\mathbf{y}^D) d\boldsymbol{\theta} d\sigma^2 \times \pi^{JR}(\boldsymbol{\gamma}) \\ &\propto |\mathbf{R}|^{-1/2}|\mathbf{h}^T(\mathbf{x}^D)\mathbf{R}^{-1}\mathbf{h}(\mathbf{x}^D)|^{-1/2}(S^2)^{-(\frac{n-q}{2})} \times \pi^{JR}(\boldsymbol{\gamma}), \end{aligned} \tag{25}$$

which is the posterior distribution for $\boldsymbol{\gamma}$ given $\mathbf{y}^D$.

Then, $\boldsymbol{\gamma}$ can be estimated by either using the marginal posterior mode

$$\hat{\boldsymbol{\gamma}} = (\hat{\gamma_1}, \ldots, \hat{\gamma_p})^T = \text{argmax}_{\gamma_1,\ldots,\gamma_p} \left\{p(\boldsymbol{\gamma}|\mathbf{y}^D\right\}, \tag{26}$$

or sampling from the posterior using the the Metropolis-Hastings algorithm. Gu (2018) indicates that the $\boldsymbol{\gamma}$ the marginal posterior mode under the parametrization given in (6) satisfies a number of robustness criteria.

2.3 Different Parameterization for the Correlation Function

However, it turns out that the parametrization in (6) leads us to have difficulty describing a full posterior density by implementing the Metropolis-Hasting algorithm especially for large p, since $p(\boldsymbol{\gamma}|\mathbf{y}^D)$ slowly gets close to zero and shows periodic behavior at the tails, which makes the chain diverge. This is because the samples in the chain get too large to come back to near the true mode, causing the numerical error, e.g. $\det(\mathbf{R}) \approx 0$.

This issue can be found even in one dimensional example. Figure 1 plots the diverging

Markov chain sampled from $p(\gamma|\mathbf{y}^D)$, where Latin Hypercube Design ranging between -10 to 10 is used for x^D and $y^D = sin(x^D)$. Therefore due to the posterior's behavior at the tails, large p is more likely to cause the divergence, making the MCMC implementation to be difficult.

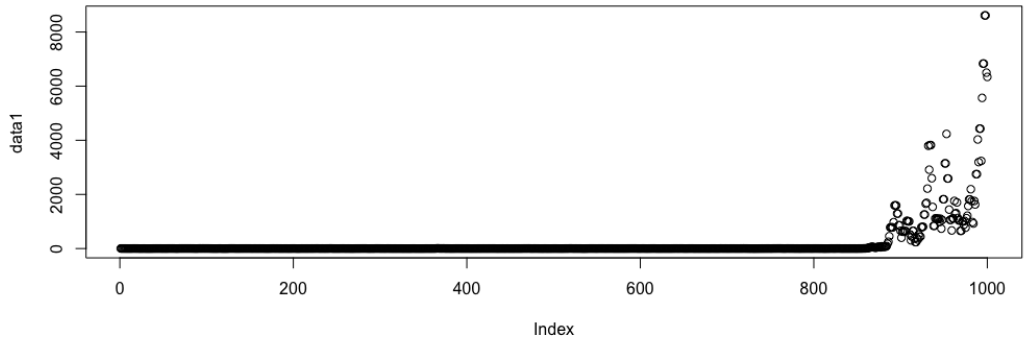

Figure 1: Samples sampled from $p(\gamma|\mathbf{y}^D)$ which diverges.

Furthermore, when the mode of $p(\gamma_l|\mathbf{y}^D)$ is too large for some l (i.e. when there is an inert input), a very long chain is required to describe a posterior density properly. Unfortunately, the computation for longer chain is difficult as a Gaussian process to a data set of size n requires a matrix inversion which has complexity $O(n^3)$. Plus, as dimensions increases, a chance that a chain diverges at least one of l^{th} coordinates increases as p increases. Therefore, we conclude that the $\boldsymbol{\gamma}$ parametrization with the JR prior may not appropriate for implementing MCMC although it has certain advantages for the marginal posterior mode.

To overcome this issue, we re-parametrize $\boldsymbol{\gamma}$ as following:

$$\rho_l = \frac{1}{\exp(\beta_l)}, \quad l = 1, \ldots, p, \tag{27}$$

with $\beta_l = \frac{1}{\gamma_l}$ (Higdon et al., 2008). An advantage of this parametrization is that ρ is now bounded between 0 and 1, and thus prevents divergence and is free from large γ_l. Note that our correlation function now becomes

$$c(\mathbf{x}_i, \mathbf{x}_j) = \prod_{l=1}^{p} \exp\left(\left(\log(\frac{1}{\rho_l})|x_{il} - x_{jl}|\right)^{1.9}\right). \tag{28}$$

Lastly, we will assume $\rho_l \sim \text{Beta}(\alpha_l, \beta_l)$.

3 Computer Model Emulation

From the above Bayesian approaches, using the Gaussian process as a prior for unknown functions provides flexible realizations interpolating given design points as well as Bayesian uncertainty. In this sense, Gaussian process can be used as a prior for a computationally expensive deterministic computer model (Oakley, 1999).

In this example, the Gaussian process regression is applied to an emulation of computational fluid dynamics model of a bubbling fluidized bed. Specifically, the Multiphase Flow with Interphase eXchanges (MFIX) is used as the simulator.

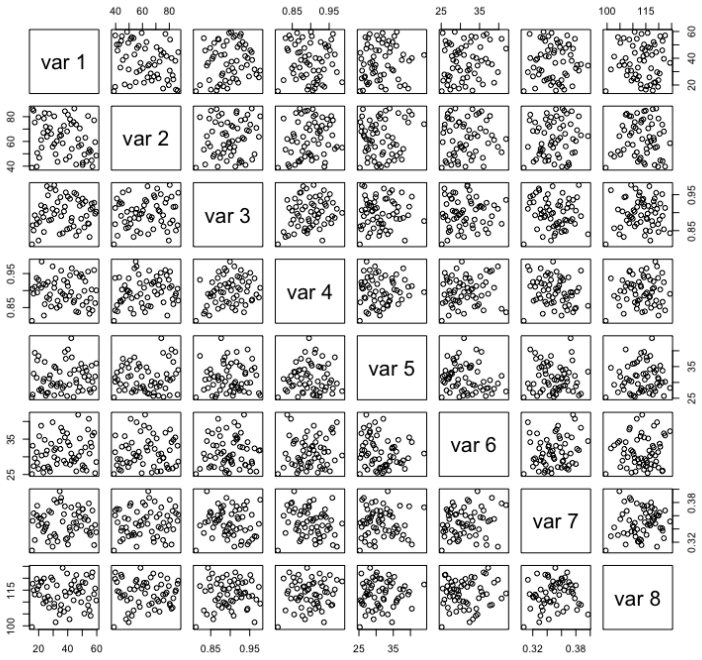

Figure 2: Pairwise scatterplot of 60 design points of the MFIX simulated data.

The 60 MFIX simulated datas from Wong et al. (2017) are used where the input variables, flow rate (x_1, FRate) and bed temperature (x_2, Temp), are involved, and model parameters Res-PP ($\theta 1$): the particle-particle coefficient of restitution, ResPW ($\theta 2$): the particle-wall coefficient of restitution, FricAng-PP ($\theta 3$): the particle-particle friction angle, FricAng-PW ($\theta 4$): the particle-wall friction angle, PBVF ($\theta 5$): Packed bed void fraction, and Part-Size ($\theta 6$). Note that $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)^T$ is a calibration parameter that should be estimated, yet as mentioned earlier, we will treat it as a fixed known parameter. The output variable y is the time averaged value of the pressure drop at location P3820 once it was oscillating in

steady state. Lastly, a constant mean is used as the Gaussian process trend.

Its corresponding pairwise scatterplot is plotted in Figure 2, and it can be noticed that each of these variables have significantly different ranges. For example, flow rate x_1 is between 15 and 60 whereas FricAng-PP θ_3 is between 0.8 and 1. Therefore, $\mathbf{x}^D$ is re-scaled to be bounded between 0 and 1 so that the effect of the term $|x_{il} - x_{jl}|$ in $c(\mathbf{x}_i, \mathbf{x}_j)$ is the same for all variables. Lastly, we are going to let $\alpha_l = \beta_l = 1$ for all l, i.e. uniform distribution, as a prior for $\boldsymbol{\rho}$.

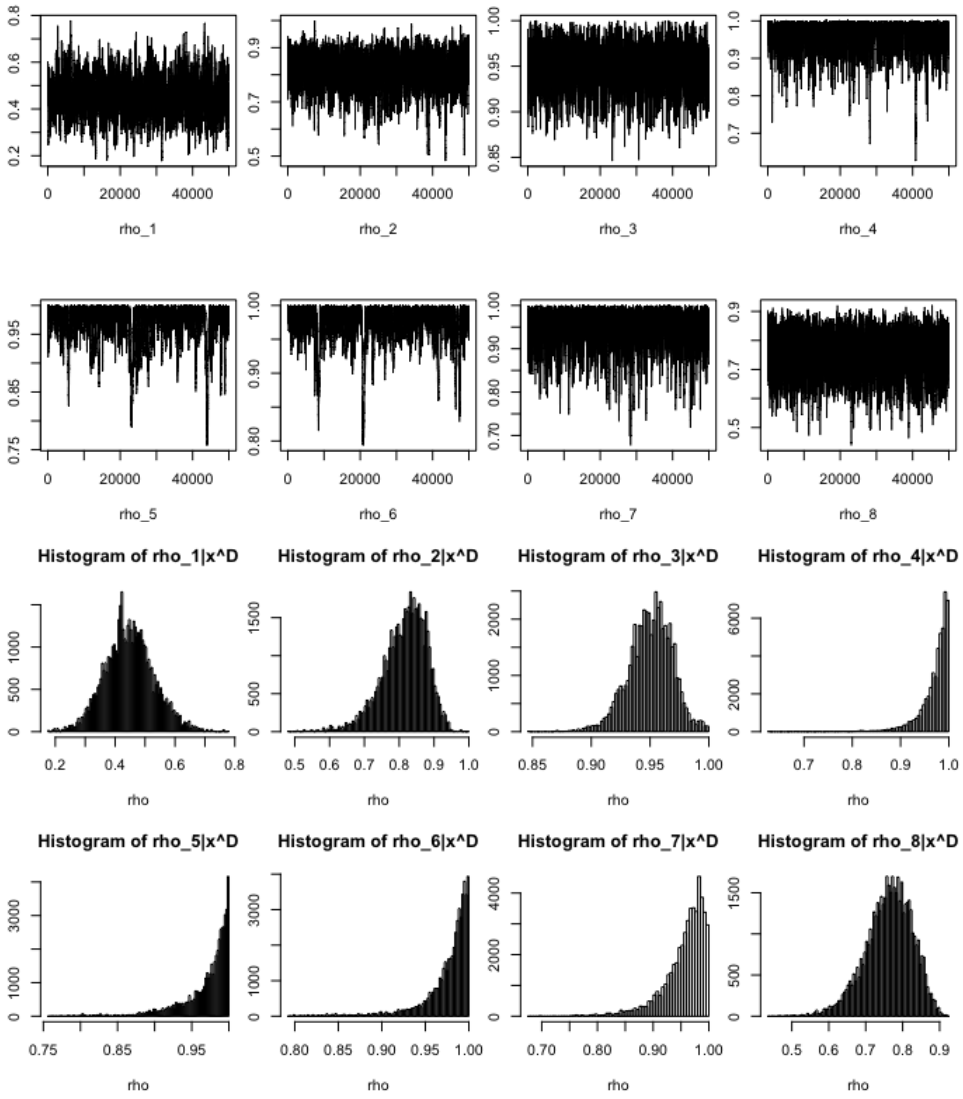

Figure 3: (a) Trace plots with respect to $\boldsymbol{\rho}$ (b) Histograms with respect to $\boldsymbol{\rho}$

Now we are going to implement the Metropolis-Hasting algorithm with the uniform proposal

distribution

$$\rho_1^*|\rho_1 \sim U\left[\frac{1}{\lambda_1}\rho_1, \lambda_1\rho_1\right], \rho_2^*|\rho_2 \sim U\left[\frac{1}{\lambda_2}\rho_2, \lambda_2\rho_2\right], \rho_3^*|\rho_3 \sim U\left[\frac{1}{\lambda_3}\rho_3, \lambda_3\rho_3\right] \tag{29}$$

$$\rho_4^*|\rho_4 \sim U\left[\frac{1}{\lambda_4}\rho_4, \lambda_4\rho_4\right], \rho_5^*|\rho_5 \sim U\left[\frac{1}{\lambda_5}\rho_5, \lambda_5\rho_5\right], \rho_6^*|\rho_6 \sim U\left[\frac{1}{\lambda_6}\rho_6, \lambda_6\rho_6\right] \tag{30}$$

$$\rho_7^*|\rho_7 \sim U\left[\frac{1}{\lambda_7}\rho_7, \lambda_7\rho_7\right], \rho_8^*|\rho_8 \sim U\left[\frac{1}{\lambda_8}\rho_8, \lambda_8\rho_8\right] \tag{31}$$

with $\boldsymbol{\lambda} = (1.2, 1.1, 1.1, 1.1, 1.02, 1.02, 1.1, 1.2)^T$. Then with the initial point $(0.6, 0.8, 0.99, 0.997, 0.9999, 0.9999, 0.991, 0.9)^T$ we get the trace plots and histograms accordingly, given in Figure 3. The Metropolis-Hasting tuning parameter and the initial point were chosen empirically after several implementations, so that we have appropriate rejection rate and have no burn-in period. The relevant interpretation linking the plots and the result (21) will be given in the later paragraph.

Then, using the result (21) and the achieved description of the density of $\gamma|\mathbf{y}^D$ from the Metropolis hasting algorithm, we can predict the time averaged values of the pressure drop at location P3820 given the input variables. We first randomly choose 250 samples from the posterior $\boldsymbol{\rho}|\mathbf{y}^D$ to calculate $E[\mathbf{y}^*|\mathbf{y}^D]$ using the law of total expectation, i.e.

$$E[\mathbf{y}^*|\mathbf{y}^D] = E[E[\mathbf{y}^*|\boldsymbol{\rho}, \mathbf{y}^D]|\mathbf{y}^D]. \tag{32}$$

Furthermore, a credible band can also be achieved by using the density of $\gamma|\mathbf{y}^D$, which is given by

$$\left[E[\mathbf{y}^*|\mathbf{y}^D] - t_{n-q,\alpha/2}\sqrt{var(\mathbf{y}^*|\mathbf{y}^D)}, E[\mathbf{y}^*|\mathbf{y}^D] + t_{n-q,1-\alpha/2}\sqrt{var(\mathbf{y}^*|\mathbf{y}^D)}\right], \tag{33}$$

where $var(\mathbf{y}^*|\mathbf{y}^D)$ is achieved by using the law of total variance:

$$var(\mathbf{y}^*|\mathbf{y}^D) = E[var(\mathbf{y}^*|\boldsymbol{\rho}, \mathbf{y}^D)] + var(E[\mathbf{y}^*|\boldsymbol{\rho}, \mathbf{y}^D]). \tag{34}$$

$t_{n-q,\alpha/2} = 2.262$ and $\alpha = 0.05$ is used, yet this is not a rigorous 95% credible interval since $t_{n-q,\alpha/2}$ is not an exact 95% quantile in this case.

An easier approach would be simply plugging in a single estimate of $\boldsymbol{\rho}$ (e.g. the mode) so

that

$$y(\mathbf{x}^*)|\mathbf{y}^D, \hat{\boldsymbol{\rho}} \sim t_{n-q}(m^{**}(\hat{\boldsymbol{\rho}}), \frac{S^2}{n-q-2}c^{**}(\hat{\boldsymbol{\rho}})). \tag{35}$$

Then we can similarly construct a credible interval as in equation (33).

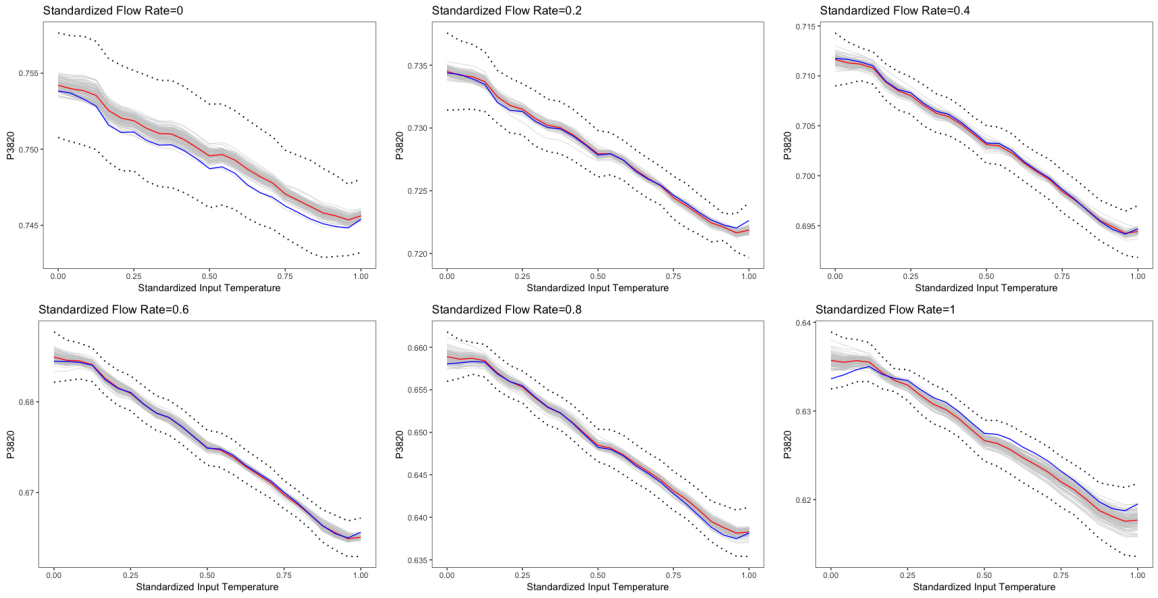

Figure 4: Values of the pressure drop P3820 against standardized input temperature on six distinct standardized flow rates and mid-points of standardized fixed model parameters (0.5). The grey line represents 200 realizations, i.e. $E[\mathbf{y}^*|\mathbf{y}^D, \tilde{\boldsymbol{\rho}}_i]$, where $\tilde{\boldsymbol{\rho}}_i$ is a sample from the chain for $i = 1, \ldots, 200$. The red line represents predictive mean $E[\mathbf{y}^*|\mathbf{y}^D]$, and the blue line represents $E[\mathbf{y}^*|\mathbf{y}^D, \hat{\boldsymbol{\rho}}]$, where $\hat{\boldsymbol{\rho}}$ is the marginal posterior mode. Lastly, the black dotted line represents the credible bands; equation (33).

The fitted simulator is summarized in Figure 4 by plotting the pressure drop against the input temperatures on six distinct values of Flow rates ($x_2 = 20, 30, 40, 50, 60, 70$) and the fixed model parameters ($\theta_1 = 0.9, \theta_2 = 0.9, \theta_3 = 30, \theta_4 = 30, \theta_5 = 0.33, \theta_6 = 110$). The red line and the black dotted line represent the predicted time averaged values of the pressure drop P3820 and its corresponding 95% credible band, respectively. Lastly, the blue line represents $E[\mathbf{y}^*|\mathbf{y}^D, \hat{\boldsymbol{\rho}}]$, where $\hat{\boldsymbol{\rho}}$ is the marginal posterior mode.

From the figures, it can be understood how the changes of input temperatures (x_2) and Flow rate (x_1) affects the pressure drop at location P3820 given the model parameters. For example, as the input temperatures increases from 40(C) to 90(C), the pressure drop decreases regardless of Flow rates. Furthermore, increase in Flow rate causes the decreases of the pressure drop. Lastly, the fitted emulator exhibits more uncertainty at the $x_1 = 20, 70$

since they are at the end points of the design points.

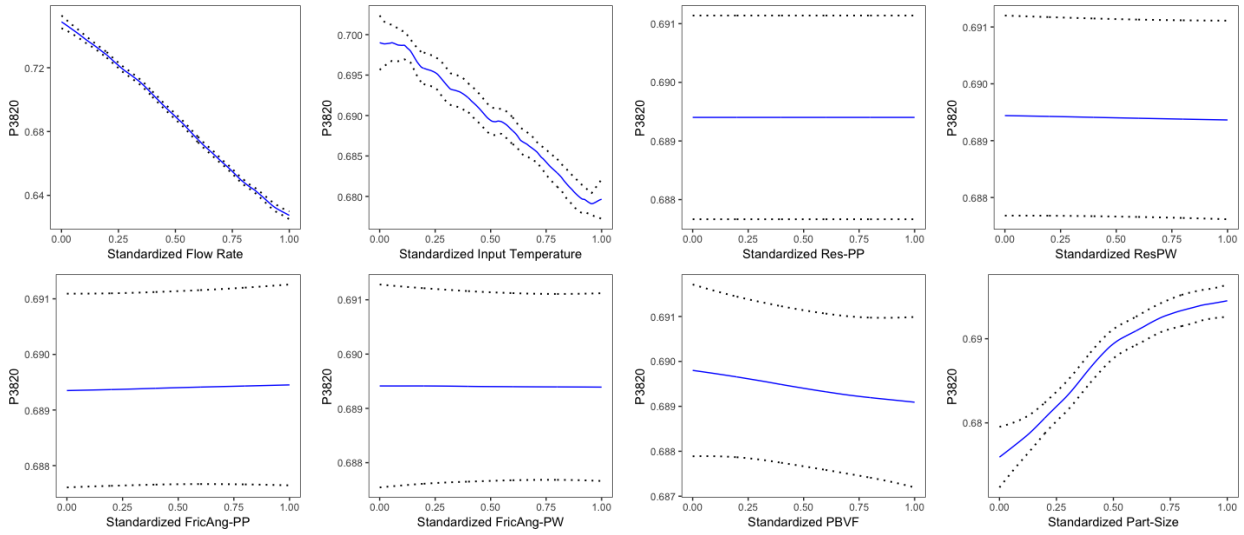

Figure 5: Values of the pressure drop P3820 by varying one input at a time. Each standardized inputs were fixed as following: $x_1 = 0.5, x_2 = 0.5, \theta_1 = 0.5, , \theta_2 = 0.5, \theta_3 = 0.5, \theta_4 = 30, \theta_5 = 0.5, \theta_6 = 0.5$.

Moreover, we analyze a magnitude of effects of $x_1, x_2, \theta_1, \theta_2, \theta_3, \theta_,\theta_5, \theta_6$ to the values of the pressure drop P3820 by varying one input at a time. This time, for the computational efficiency, $E[\mathbf{y}^*|\mathbf{y}^D, \hat{\boldsymbol{\rho}}]$ and $var(\mathbf{y}^*|\mathbf{y}^D, \hat{\boldsymbol{\rho}})$ is used for the visualization, where $\hat{\boldsymbol{\rho}}$ is the marginal posterior mode. From the figure 5, $\theta_1, \theta_2, \theta_4$ and θ_5 seems to have little effect on the output of the model.

This result agrees with the values of $p(\boldsymbol{\rho}|y^D)$ in Figure 3. We can notice that the modes of $\rho_3, \rho_4, \rho_5, \rho_6, \rho_7$ are significantly skewed toward to 1. Then $c(\mathbf{x}_i, \mathbf{x}_j)$ is not significantly affected by the change in distance $|x_{il} - x_{jl}|$ for $l = 3, 4, 5, 6, 7$. That is, the change does not affect the predictive mean function (17) and thus it can be concluded that $\theta_1, \theta_2, \theta_4$ and θ_5 might be inert inputs. Note that this analogy was possible because every variable is in the same scale.

Lastly, we visualize this by plotting the actual versus predicted values and see if it is close to a regressed diagonal line, yet the Gaussian process regression always gives $y = x$ since this always interpolates given design points and outputs by its construction. To avoid this issue, we fit the model without the i^{th} sample, giving $\hat{y}_i$. Repeat this for $i = 1, \ldots, 60$ and plot $\hat{y}_i$ against on $\mathbf{y}$ provides Leave-one-out plot in Figure 6. The plot clearly shows a linear

pattern, and thus we conclude that the model fits well.

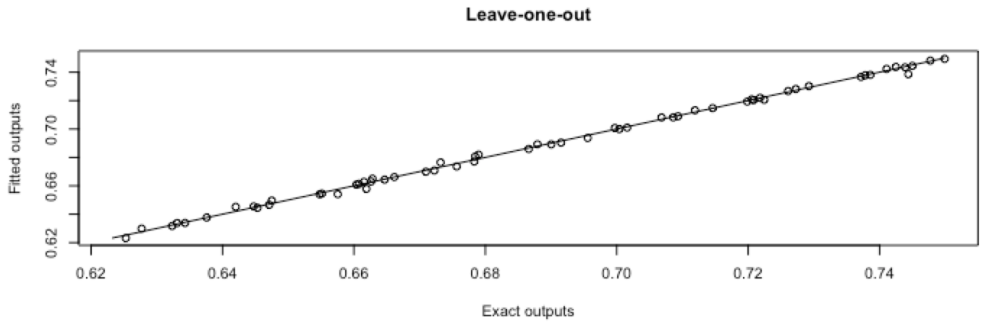

Figure 6: Leave-one-out plot.

4 Discussion

This paper has presented a full Bayesian approach for computer model emulation, using the Metropolis-Hasting algorithm. Section 2 provides detailed explanations for deriving the predictive mean and variance function in (21). Using these functions, Section 3 describes how the MFIX model behave by plotting the values of the pressure drop P3820 against standardized input temperature on six distinct standardized flow rates and mid-points of standardized fixed model parameters. By marginalizing all the parameters in the model out using the Metropolis-Hasting algorithm, we could take the uncertainty involved in the parameters, specifically $\boldsymbol{\rho}$, fully into account to the predictive distribution.

We have also presented an empirical Bayes approach, by simply plugging the single estimate (e.g. the mode of $\boldsymbol{\rho}$) into the equation (21). This approach has considerably less computing time; while sampling $50,000$ samples using the Metropolis-Hasting algorithm took more than 30 minutes, the marginal posterior mode $\hat{\boldsymbol{\rho}}$ can be achieved through some algorithms (e.g. Newton-Raphson algorithm) within a second.

Even though the uncertainty involved in $\boldsymbol{\rho}$ is simplified in exchange for the computational efficiency, the cost does not seem to be significant; we can notice that there is no big difference between blue and red line in Figure 4. This is because the mean and the mode of the density of $\boldsymbol{\rho}|\mathbf{y}^D$ are close each other; $E[\boldsymbol{\rho}|\mathbf{y}^D] = (0.4569938, 0.8132497, 0.9509005, 0.9728466, 0.9775522, 0.9742655, 0.95911740.7669581)^T$, $\hat{\boldsymbol{\rho}} = (0.4473867, 0.8525918, 0.9543105, 0.9961552, 0.9934416, 0.9882758, 0.97638130.7437625)^T$. Also, the effect of $\boldsymbol{\rho}$ is decreased as $\boldsymbol{\rho}$ is close to 1 so that

skewed density for $\rho_4, \rho_5, \rho_6, \rho_7$ does not significant effect on the predictive mean function in the equation (21). Therefore, this paper concludes that when emulating computer models without considering estimation of calibration parameters, the plug-in method is sufficient and more desirable.

There are further topics that were not discussed in this paper. The choice of the correlation function in Gaussian process model should be carefully addressed. Since the correlation function determines features of the Gaussian process such as its smoothness, if it is mis-specified, the predictive mean function might be significanlty different from the desirable function (e.g. the emulation model). Moreover, we shall treat the calibration parameters as random variables in order to predict the behaviour of the processes. The expectation is that the calibration requires the full Bayesian approach so that the estimation contains all sources of uncertainties. The according computational cost is another key issue.

References

[1] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

[2] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583, 2008.

[3] J. Oakley. *Bayesian uncertainty analysis for complex computer codes*. PhD thesis, University of Sheffield, 1999.

[4] M. Gu et al. Jointly robust prior for gaussian stochastic process in emulation, calibration and variable selection. *Bayesian Analysis*, 2018.

[5] M. Gu, J. Palomo, and J. O. Berger. Robustgasp: Robust gaussian stochastic process emulation in r. *arXiv preprint arXiv:1801.01874*, 2018.

[6] M. Gu, X. Wang, J. O. Berger, et al. Robust gaussian stochastic process emulation. *The Annals of Statistics*, 46(6A):3038–3066, 2018.

[7] R. K. Wong, C. B. Storlie, and T. C. Lee. A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):635–648, 2017.

[8] S. Banerjee. Bayesian linear model: Gory details. *Dowloaded from http://www. biostat. umn. edu/˜ ph7440*, 2008.

[9] Y. Englezou. *Bayesian design for calibration of physical models*. PhD thesis, University of Southampton, 2018.