Calibration of mathematical computer models

Ksenia N. Kyzyurova
The University of Sheffield, UK
kseniak.ucoz.net, ksenia.kyzyurova@gmail.com

November 26, 2019

Abstract

The problem of calibration of mathematical computer models with respect to collected data occasionally occurs in contemporary research. Calibration task is analogous to identifying the *preimage* of a set of experimental or observational data under a certain computer model — function over an input space of parameters to that model. Identified preimage is formed as a subset of the input space to the computer model. In its turn, collected data is typically described by means of probability distributions; thus, leading to performing probabilistic calibration within the Bayesian framework. Bayesian inversion is considered advantageous over other non-Bayesian or pseudo-Bayesian approaches because of interpretability of its results.

In practice, calibration quickly runs into computational obstacles: the shape of the posterior distribution resulting from the Bayesian inversion may be "ugly", such that standard approaches to its estimate (including Markov chain Monte-Carlo (MCMC) approximation) are prohibitive. Instructional examples are provided for illustration. We conclude that calibration must be performed together with investigation of its scientific computing issues and software/hardware numerics.

1 Introduction

A computer model is a simulator based on mathematical representation of a natural phenomenon. Given a set of initial values of input parameters in the input space X to a computer model f, a value in the output space Y is obtained. If observations corresponding to the output of the model are collected, a theoretical model may be assessed on the agreement to the collected data. Moreover, given a computer model and collected data, one might inquire which values of inputs to the model could have generated the collected data; thus, performing *calibration* of a model.

Calibration is, therefore, analogous to finding an inverse image of function $f : X \to Y$, given a set of n values $\mathbf{y} = (y_1, \ldots, y_n) \in Y$, that is $f^{-1}(\mathbf{y}) = \{\mathbf{x} = (x_1, \ldots, x_n) \in X : f(x_i) = y_i \ \ \forall\, i = 1, \ldots, n\}$. However, this is indeed only an analogy, since in practice calibration involves (a) a computationally intensive computer model or a system of such models, and (b) *noisy* collected data which does not arise from the computer model itself but is obtained in either an experiment or observed in nature. Calibration framework which accounts for these two levels of complexity is presented in this study.

Examples of computer models with the properties described above include scientific models of volcano pyroclastic flows (Pitman et al. 2003, Pitman & Le 2005), volcano ash transport and dispersal model (Bursik et al. 2012) and engineering system for carbon sequestration (United States Department of Energy carbon-capture initiative). The input space of parameters X to such models varies from low to high dimensionality.

We assume throughout that the number of parameters is at most only moderately large, consisting of no more than several dozens of parameters, as have been observed in these and other models (Ghanem 2018). In practice, high-dimensional space has been already reduced with suitable sensitivity analysis for an input space, or simply main inputs have been identified by experts (Madankan et al. 2014); thus, the input space X is irreducible further.

Analytic solution f to such a computer model (or simply a computer model itself) of a kind given in the examples, for a particular configuration of input parameters, is typically not available and is demanding for numerical approximation to the solution. In its turn numerical procedure takes considerable time to run, so that execution of a computer model at all desired input configurations is either not feasible and/or involves a trade-off between computational time and numerical error associated with that procedure (Gramacy & Lee 2012, Chkrebtii et al. 2016).

Another issue is logistical: storage and sharing of large volumes of input and output data from a computer model is problematic and not desirable. Thus, we assume that a computer model is required to be approximated by a statistical surrogate. Such approximations are constructed relying only on a handful of inputs-outputs available from the execution of a model. One example is a Gaussian Process emulator of a computer model (Santner et al. 2003). This type of surrogates has been proved to be useful and is employed in this work.[1]

Once a computer model and/or its approximation is made available, its calibration must incorporate collected data. Experimental or observational, measured or otherwise collected data, due to noise in the data, is convenient to characterize by distributional assumptions on the data. This leads to the natural choice of performing calibration within the framework of Bayesian inversion. This framework proposes a probabilistic version of identifying an inverse image of a deterministic function, a computer model f.

We start the calibration study with the exposition of a forward problem in Section 2: construction of an emulator of a computer model and that of a system of computer models. Section 3 introduces Bayesian inversion, that is calibration of a computer model involving its surrogate with respect to collected data. Section 4 concludes with further research directions.

2 Forward problem

Gaussian process emulator of a computer model. Let function $g(\cdot)$ denote a computer model. Suppose a computer model has been evaluated at m input sets of parameters $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$. Each input $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^{\mathrm{T}}$ is d-dimensional. Once a computer model has been evaluated at $\mathbf{x}$, output values $\mathbf{g}(\mathbf{x})$ are available. In this exposition output space $Y \in \mathcal{R}$ is one-dimensional.

Fast statistical approximation to a computer model g is constructed assuming that model g represents a sample from a Gaussian process $g^M(\cdot)$ and given pairs of inputs-outputs $\{\mathbf{x}, \mathbf{g}(\mathbf{x})\}$. While these inputs-outputs are from the model g, they also denote pairs of inputs-outputs $\{\mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})\}$ from the approximation, such that $\{\mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})\} = \{\mathbf{x}, \mathbf{g}(\mathbf{x})\}$.

For brief mathematical introduction we only state the general notation of a Gaussian process emulator (GASP), that is, the predictive Gaussian process, conditional on pairs $\{\mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})\}$,

$$g^M(\cdot) \mid \cdot, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}) \sim \mathcal{GASP}(\mu_g^*(\cdot), \sigma_g^{*2}(\cdot, \cdot)) \,, \tag{1}$$

where $\mu_g^*(\cdot)$ is the predictive mean and $\sigma_g^{*2}(\cdot, \cdot)$ is the predictive variance-covariance matrix. These functions are specified parametrically.[2]

[1] The consistency of Gaussian process approximations with the original solution to the model has been shown in Stuart & Teckentrup (2018).

[2] This introduction to the GASP methodology suffices for the purpose of this work. Detailed exposition may be found in, e.g. (Kyzyurova et al. 2018). In there a GASP emulator builds on developments in (Bayarri et al. 2007) and objective Bayesian implementation of the emulator following (Gu et al. 2018, Berger et al. 2001).

The notation (1) means that probabilistic approximation to a computer model output at any set of inputs $\mathbf{x}^*$ is given by a corresponding joint predictive normal distribution, conditional on m pairs of inputs-outputs $\{\mathbf{x}_i, g(\mathbf{x}_i)\}_{i=1}^m$,

$$g^M(\mathbf{x}^*) \mid \mathbf{x}^*, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}) \sim \mathcal{N}(\mu_g^*(\mathbf{x}^*), \sigma_g^{*2}(\mathbf{x}^*, \mathbf{x}^*)) \,.$$

where $\mu_g^*(\mathbf{x}^*)$ denotes the mean and $\sigma_g^{*2}(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance matrix evaluated at inputs $\mathbf{x}^*$.

Emulator of a system of computer models. Having introduced the framework for an emulator of a single computer model we turn to a question on how to approximate a coupled system of computer models mathematically expressed as a composition $g \circ f(\cdot)$ of two parts $g(\cdot)$ and $f(\cdot)$, if only the two submodels $g(\cdot)$ and $f(\cdot)$ are independently observed, while the composition $g \circ f(\cdot)$ is not.

Let two conditional stochastic processes $g^M(\cdot)$ and $f^M(\cdot)$ be emulators of two computer models $g : X \to Y$ and $f : \Upsilon \to X$

$$g^M(\cdot) \mid \cdot, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}) \sim \mathcal{GASP}(\mu_g^*(\cdot), \sigma_g^{*2}(\cdot, \cdot)) \,, \tag{2}$$

$$f^M(\cdot) \mid \cdot, \mathbf{u}, \mathbf{f}^{\mathbf{M}}(\mathbf{u}) \sim \mathcal{GASP}(\mu_f^*(\cdot), \sigma_f^{*2}(\cdot, \cdot)) \,. \tag{3}$$

Then the coupled system of models $g \circ f(\cdot)$ which has not been observed may be approximated using the *linked emulator* given by

$$(g \circ f)^M(\cdot) \mid \cdot, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}), \mathbf{u}, \mathbf{f}^{\mathbf{M}}(\mathbf{u})$$

or its Gaussian approximation *linked GASP*. What has been missing in that work is the *correlation structure* emerging in these two emulators.

Lemma. Following the notation of the linked emulator, let $\xi_1 = (g \circ f)^M(\mathbf{u_1})$ and $\xi_2 = (g \circ f)^M(\mathbf{u_2})$ be two random variables of a linked emulator $(g \circ f)^M$ at new respective inputs $\mathbf{u_1} \in \Upsilon$ and $\mathbf{u_2} \in \Upsilon$, then covariance between ξ_1 and ξ_2 by total law of covariance is

$$\begin{aligned}\mathrm{Cov}(\xi_1, \xi_2) = \mathrm{E}_{f^M(\mathbf{u_1}), f^M(\mathbf{u_2})} \mathrm{Cov}\left(g^M \mid f^M(\mathbf{u_1}), g^M \mid f^M(\mathbf{u_2})\right) + \\ \mathrm{Cov}_{f^M(\mathbf{u_1}), f^M(\mathbf{u_2})}\left(\mathrm{E}_{g^M \mid f^M(\mathbf{u_1})} g^M, \mathrm{E}_{g^M \mid f^M(\mathbf{u_2})} g^M\right) \,.\end{aligned} \tag{4}$$

Using this lemma, closed-form expression of covariance between variables of a linked emulator stochastic process in some special cases may be found. We state the corresponding theorem in the appendix 5.1 for completeness of this work.

Correlation structure of *any* emulator we consider — GASP emulator, linked emulator or linked GASP — determines the smoothness of sample trajectories of that emulator. This feature is illustrated in the next paragraph.

Influence of an emulator's correlation structure on its sample trajectories. Illustrative example. A damping sine function, namely, $f(t, \omega) = \exp(-t)\cos(\omega\, t)$ where $\omega \in [1, 10]$ and $t \in [0, 1]$, is chosen to be a simulator. Assuming that this function is not known to us, the emulator is to be constructed. We choose 24 training points $\mathbf{x}^{tr} = (\mathbf{t}^{tr}, \boldsymbol{\omega}^{tr})$ over a grid 6×4 of the two-dimensional space $\{t, \omega\}$. The grid is shown in Figure 1, left panel. The simulator f has been evaluated at the training points resulting in outputs $\mathbf{f}(\mathbf{x}^{tr}) = \{f(\mathbf{t}^{tr}{}_i, \boldsymbol{\omega}^{tr}{}_i)\}_{i=1}^{24}$.

Emulator f^M of the simulator, conditional on simulator input-output pairs of data $\{\mathbf{x}_i^{tr}, f(\mathbf{t}^{tr}{}_i, \boldsymbol{\omega}^{tr}{}_i)\}_{i=1}^{24}$, has been constructed. For parametrization of the Gaussian stochastic process we have chosen squared-exponential correlation function and linear mean function within the objective Bayesian implementation of its construction. Resulting emulator is shown in Figure 1, right panel. This emulator has been tested for adequacy of performance on five testing points, also shown in the left panel of that Figure.

Testing points have been captured in the corresponding 95% credible area, resulting in 100% empirical frequency coverage. The average length of credible intervals is 0.53 (ranging from 0.48 to 0.55), overall demonstrating excellent performance.

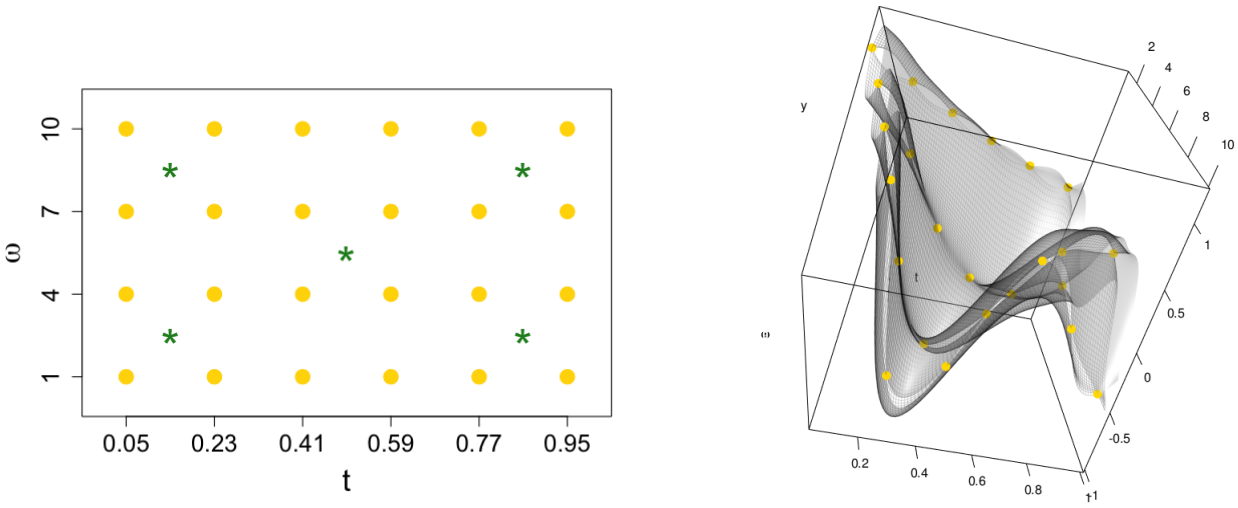

Figure 1: Illustration. Left panel shows chosen uniform design for training points shown with golden circles. Five testing points, shown with green asterisks, comprise about 20% of training design points. Right panel shows the constructed emulator. The emulator's 95% central credible area is enclosed within the two grey surfaces. Golden circles are the design points.

At a set of m new inputs $\mathbf{x}' = (\mathbf{t}', \boldsymbol{\omega}')$, the emulator f^M is given by the normal predictive distribution with the mean $\mu_f^*(\mathbf{x}')$ and covariance matrix $\sigma_f^{*2}(\mathbf{x}', \mathbf{x}')$, that is

$$f^M(\mathbf{x}') \mid \mathbf{x}', \mathbf{x}^{tr}, \mathbf{f}(\mathbf{x}^{tr}) \sim \mathcal{N}(\mu_f^*(\mathbf{x}'), \sigma_f^{*2}(\mathbf{x}', \mathbf{x}'))\,. \tag{5}$$

Samples from this distribution may be obtained. One sample is shown in Figure 2, left panel. The sample is smooth, because variance-covariance matrix is accounted for.

If correlation structure is omitted, this induces independence between all of its pairs of variables. One sample from

$$\mathcal{N}(\mu_f^*(\mathbf{x}'), \text{diag}\left(\sigma_f^{*2}(\mathbf{x}', \mathbf{x}')\right)) \tag{6}$$

is shown in the right panel of Figure 2. As can be seen, the sample trajectory has lost all information about smoothness of the Gaussian process samples, resulting in rough samples. Independence is imposed through the identity correlation matrix $I_{m\times m}$.

3 Inverse problem

We now assume that an emulator (1) of the computer model g has been constructed. We turn to the question of calibrating this model using its approximation (1) with respect to, first, noiseless observations, and, second, with respect to noisy observations.

3.1 Calibration with respect to noiseless collected data.

Suppose that n experimental or observational data points $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ have been collected; each ith point $y_i, i = 1, \ldots, n$ has been obtained under some input conditions $\mathbf{x}_i^* \in X$. These input conditions are assumed to belong to the same input space $\mathbf{x}_i^* \in X$ as the computer model inputs.

Each ith input $\mathbf{x}_i^*$ is comprised of a set of control variables $\mathbf{x}_i^{*,c} \in \mathcal{R}^{d_1}$ (say, first d_1 inputs) and a set of calibration variables $\mathbf{x}_i^{*,p} \in \mathcal{R}^{d_2}$, the set of the rest $d_2 = d - d_1$ variables; i.e. $\mathbf{x}^* = \{\mathbf{x}^{*,c}, \mathbf{x}^{*,p}\}$. A subset of control variables $\mathbf{x}^{*,c}$ is set and known. A subset of $\mathbf{x}^{*,p}$ is unknown. However, we assume that the unknown conditions are the same for all data points, i.e. $\mathbf{x}_1^{*,p} = \mathbf{x}_2^{*,p} = \ldots = \mathbf{x}_n^{*,p} = \boldsymbol{\omega}$.

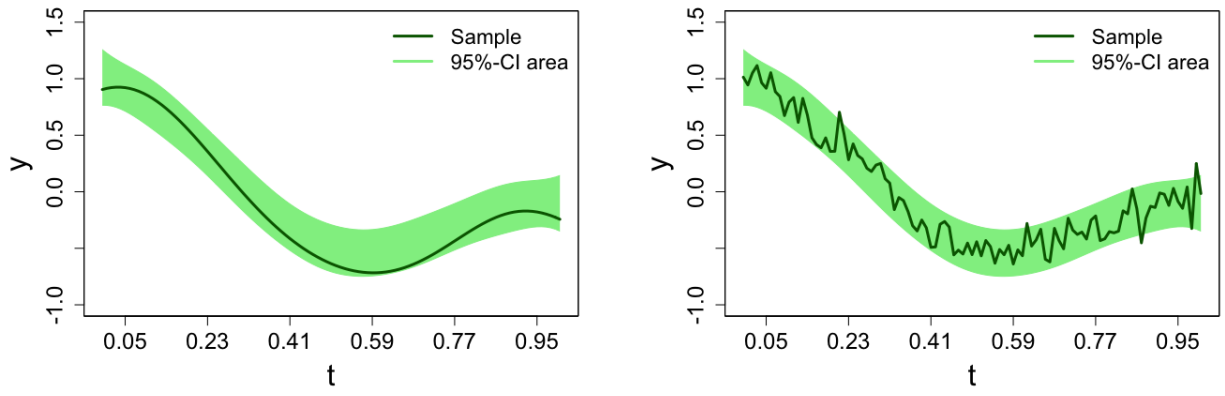

Figure 2: Illustration. Left panel: one sample path from the emulator (5) which accounts for the correlation structure, resulting in a smooth trajectory. Right panel: one sample path from the distribution (6) for which the correlation structure is absent, thus using only information from marginals of the predictive GASP.

Calibration aims at identifying such values of parameters $\boldsymbol{\omega}$, that computer model output (that is of its approximation) $\mathbf{g}^{\mathbf{M}}(\mathbf{x}^*) = g^M(\mathbf{x}_1^{*,c}, \boldsymbol{\omega}), \ldots, g^M(\mathbf{x}_n^{*,c}, \boldsymbol{\omega})$ obtained at inputs $\mathbf{x}^*$ with parameters $\boldsymbol{\omega}$ plugged in matches closely experimental data $\mathbf{y}$.

Conditional on the computer model data $\mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})$, within the emulation framework, computer model output $\mathbf{g}^{\mathbf{M}}(\mathbf{x}^*) = \mathbf{g}^{\mathbf{M}}(\mathbf{x}^{*,c}, \boldsymbol{\omega})$ is given by the predictive distribution

$$\mathbf{g}^{\mathbf{M}}(\mathbf{x}^{*,c}, \boldsymbol{\omega}) \mid \mathbf{x}^{*,c}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}) = (g^M(\mathbf{x}_1^{*,c}, \boldsymbol{\omega}), \ldots, g^M(\mathbf{x}_n^{*,c}, \boldsymbol{\omega}))^{\mathrm{T}} \mid \mathbf{x}^{*,c}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})\,. \tag{7}$$

This yields the likelihood $L(\boldsymbol{\omega}; \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x}^{*,c}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}))$ for to-be calibrated parameters $\boldsymbol{\omega}$, which is given by the distribution of experimental data $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ observed at $\mathbf{x}^*$

$$L(\boldsymbol{\omega}; \mathbf{y}) = p(g^M(\mathbf{x}_1^{*,c}, \boldsymbol{\omega}) = y_1, \ldots, g^M(\mathbf{x}_n^{*,c}, \boldsymbol{\omega}) = y_n \mid \mathbf{x}^{*,c}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}))\,. \tag{8}$$

To complete Bayesian specification of the model, one must specify the prior distribution for unknown parameters $\boldsymbol{\omega}$. In the calibration setting prior distribution $p(\boldsymbol{\omega})$ on parameters $\boldsymbol{\omega}$ is, though, possibly vague, but an expert-elicited prior.

The solution to the calibration problem is the posterior distribution $p(\boldsymbol{\omega} \mid \mathbf{y}, \mathbf{x}^{*,c}, \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}))$. For convenience the short notation is $p(\boldsymbol{\omega} \mid \mathbf{y})$ which implicitly assumes conditioning on known controlled settings $\mathbf{x}^{*,c}$ and computer model data, that is, pairs of observations $\{\mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x})\}$. Posterior distribution is then obtained via the Bayes' theorem

$$p(\boldsymbol{\omega} \mid \mathbf{y}) = \frac{L(\boldsymbol{\omega}; \mathbf{y}) p(\boldsymbol{\omega})}{\int L(\boldsymbol{\omega}; \mathbf{y}) p(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega}}\,. \tag{9}$$

This distribution is well-defined since the prior $p(\boldsymbol{\omega})$ is defined on closed space of parameters, and, therefore, the posterior is proper. Thus, the numerical solution (9) to the calibration problem always exists.

Qualitative assessment of calibration results. Ideally, posterior $p(\boldsymbol{\omega} \mid y)$ identifies a small set of most probable parameters $\boldsymbol{\omega}$ such that computer model data evaluated at these values matches closely collected data $\mathbf{y}$. The assessment of agreement between the theoretical computer model and experimental or observational data may be done subjectively or objectively. Mathematical methodology of doing so is discussed in (Kyzyurova 2019*a*) and shown in the study (Kyzyurova 2019*b*).

This step is important in practice because, in the unfortunate scenario, one may find that even being evaluated at the most probable values of $\boldsymbol{\omega}$, computer model output may not match the collected data, thus,

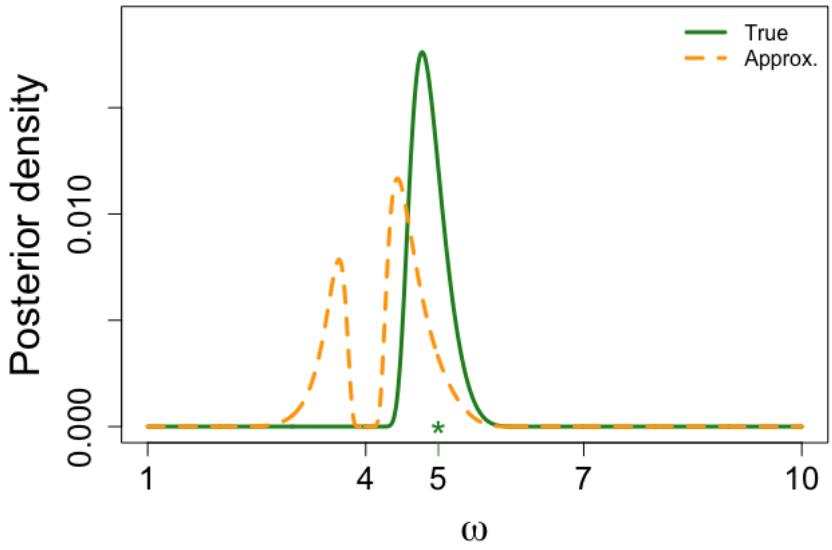

Figure 3: Illustrative example. The true posterior density is shown with a green solid line and obtained under the model (11). The posterior obtained with absent correlation structure (12) is shown with a dashed orange line. A green star indicates the true value of the parameter ω.

resulting in *no agreement* between the theory and the data. Such a case would indicate a crucial problem with either a theoretical computer model or a prior on calibration parameters $p(\boldsymbol{\omega})$, or both; demanding for more inquiry into domain knowledge and work with field experts for problem analysis.

Illustrative example. Continuing the previous example which started in Section 2: let the true value of parameter ω be equal to 5. Let also experimental data $\mathbf{y}$ be collected at controlled input values $\mathbf{t} = (0.1, 0.7, 0.75)^{\mathrm{T}}$, that is $\mathbf{y} = (f(0.1, 5), f(0.7, 5), f(0.75, 5))^{\mathrm{T}}$.

The GASP approximation to a computer model f provides the likelihood for ω via the predictive distribution $L(\omega; \mathbf{y}) = p(\mathbf{f}^{\mathbf{M}}(\mathbf{t}, \omega) = \mathbf{y} \mid \mathbf{t}, \omega, \mathbf{x}^{tr}, f(\mathbf{x}^{tr}))$ given by its mathematical representation (5)

$$L(\omega; \mathbf{y}) = \mathcal{N}(\mathbf{y} \mid \mu_f^*(\mathbf{t}, \omega), \sigma_f^{*2}((\mathbf{t}, \omega), (\mathbf{t}, \omega)))\,, \tag{10}$$

where this normal distribution of experimental data $\mathbf{y}$ is determined by the mean vector $\mu_f^*(\mathbf{t}, \omega)$ and the variance-covariance matrix $\sigma_f^{*2}((\mathbf{t}, \omega), (\mathbf{t}, \omega))$; both of which depend on an unknown parameter ω which we would like to calibrate.

Assuming that experts may only provide broad prior information on the parameter ω, a typical prior information is set as uniform on all possible values identified by the experts. In this particular simulation example, that would be a uniform distribution $\omega \sim \mathcal{U}(1, 10)$.[3] The following posterior corresponding to the calibration with the uniform (flat) prior on parameter ω is shown in Figure 3 with a green solid line

$$p(\omega \mid \mathbf{y}) \propto \mathcal{N}\left(\mathbf{y} \mid \mu_f^*(\mathbf{t}, \omega), \sigma_f^{*2}((\mathbf{t}, \omega), (\mathbf{t}, \omega))\right) p(\omega)\,. \tag{11}$$

The mass of the true posterior is concentrated around the true value of the parameter $\omega = 5$ showing good parameter recovery.

[3]This distribution has a restricted area of support, which is advantageous from computational perspective. However, if the range of support is desired to be unrestricted, $\mathcal{N}(\text{mean} = 5.5, \text{sd} = 2.25)$ may be an alternative distribution which serves as the prior on ω. Almost 95% of the mass of this distribution falls within the range $[1, 10]$.

Bayesian inversion (11) involves numerical computation of the inverse of the covariance matrix $\sigma_f^{*2}((\mathbf{t},\omega),(\mathbf{t},\omega))$ from the log-likelihood for different values of ω. This matrix may provide strong correlation between points, such that its correct numerical inversion becomes a challenging problem even if the number of experimental data points is small. Numerical errors coming from the inversion of the matrix in log-densities may preclude obtaining the correct solution. In order to overcome this computational problem the information about the correlation structure in the likelihood may be omitted, effectively resulting in the following distribution (12) as an approximation to the true posterior (11)

$$p(\omega \mid \mathbf{y}) \propto \mathcal{N}\left(\mu_f^*(\mathbf{t},\omega), \operatorname{diag}\left(\sigma_f^{*2}((\mathbf{t},\omega),(\mathbf{t},\omega))\right)\right) p(\omega)\,. \tag{12}$$

This distribution does not take information about GASP sample paths into account but only uses the information from the *marginals* of the corresponding likelihood. Computing this approximated solution is trivial.

Proposition. Let the GASP correlation function be dependent of and decaying with distance along its inputs, then the further apart from each other experimental data inputs are, the more their corresponding cross-correlations are approaching zero. Therefore, the closer approximation (12) is to the true posterior (11).

Proof. Following exposition of the GASP, correlation between two inputs $\mathbf{x}_1$ and $\mathbf{x}_2$ is $c(\mathbf{x}_1,\mathbf{x}_2) = \prod_{j=1}^{d} c(x_{1j},x_{2j})$. For the jth coordinate correlations depend only on the distance $c(x_{1j},x_{2j}) = c(|x_{1j}-x_{2j}|)$, whose properties are $c(x_{1j},x_{2j}) \longrightarrow 1$ and $c(x_{1j},x_{2j}) \longrightarrow 0$ as $|x_{1j}-x_{2j}| \longrightarrow \infty$. The last property proves the proposition.

The resulting numerical approximation (12) from the illustrative example is shown with a dashed orange line in Figure 3. In this example accounting for correlation structure in the GASP emulator is advantageous because it provides valuable information on the smoothness of the emulator's trajectories; resulting in that the true posterior captures the true mode well (and only this mode) while the approximated solution has an additional spurious mode.

In case of multimodality typical numerical solutions to the posterior based on variants of Markov chain Monte-Carlo (MCMC) technique is going to be problematic, since switching between the modes is difficult. One has to resort to the numerical integration involving analysis of the software/hardware error. Additionally, multimodality would most likely indicate that more collected data is necessary to discern between the correct and spurious modes.

3.2 Calibration with respect to noisy collected data

In practice collected data is not perfect and is observed with noise $\boldsymbol{\epsilon}$. That is, true values of $\mathbf{y}$ are not known but may be estimated from the collected data. Collected data is also called *field data* and is denoted as $\mathbf{y}^F$. In the previous section we assumed the ideal scenario of the absence of any noise in the collected data, resulting in that $\mathbf{y}^F = \mathbf{y}$. In practice, true data $\mathbf{y}$ is not known, but posterior estimate on the true values of $\mathbf{y}$ given the collected data, i.e. the distribution $p(\mathbf{y} \mid \mathbf{y}^F)$ may be obtained.

The more information about $\boldsymbol{\epsilon}$ is available, the better the estimate of posterior $p(\mathbf{y} \mid \mathbf{y}^F)$ is going to be. The common knowledge is that the larger the noise in the data, the more replicates is required to be taken in order to estimate the noise and reduce the noise in the derived estimates of the truth. Ultimately, the quality of collected data matters most.

Having obtained posterior belief on the true values $\mathbf{y} \mid \mathbf{y}^F$, Bayesian calibration becomes

$$p(\boldsymbol{\omega} \mid \mathbf{y}^F) = \int_{\mathcal{Y}} p(\boldsymbol{\omega} \mid \mathbf{y}) p(\mathbf{y} \mid \mathbf{y}^F) d\mathbf{y}\,, \tag{13}$$

where $p(\boldsymbol{\omega} \mid \mathbf{y})$ is given by equation (9).[4]

[4]Computationally approximation to $p(\boldsymbol{\omega} \mid \mathbf{y}^F)$ may be obtained through sampling from the distribution $p(\mathbf{y} \mid \mathbf{y}^F)$. For each of the r independent samples $\mathbf{y}_{smpl_i} \sim p(\mathbf{y} \mid \mathbf{y}^F)$, $i = 1,\ldots,r$, the distribution $\boldsymbol{\omega} \mid \mathbf{y}_{smpl_i}$ may be obtained. All combined (to integrate out $\mathbf{y} \mid \mathbf{y}^F$), marginal $p(\boldsymbol{\omega} \mid \mathbf{y}^F)$ is the posterior estimate of the unknown parameter ω. This approach is easily parallelizable, since $p(\boldsymbol{\omega} \mid \mathbf{y}_{smpl_i})$ is independent from $p(\boldsymbol{\omega} \mid \mathbf{y}_{smpl_j})$ for all independent samples $\mathbf{y}_{smpl_i}$, $i = 1,\ldots,r$.

This form of the posterior reflects that we may be unsure about the truth, true values $\mathbf{y}$. The more sure we are about the truth, the closer calibration framework (13) is to (9). Calibration (13) also states that we are able to calibrate only to the extent that the observed data allow us to do it. This acknowledges the possibility that calibration task with corrupt data (e.g., data with large noise or biased data or data with limited information on how the data has been obtained) does not allow to perform calibration.

3.3 Other approaches to calibration

Other approaches to calibration are not Bayesian, therefore we omit a detailed discussion of them. These include methods based on minimization of various metrics in L^p spaces between the collected data and computer model data.[5] These approaches necessarily provide a numerical answer but lack the interpretability of its results. This is also true for the approach of Kennedy and O'Hagan.[6] We demonstrate below that this approach is not equivalent to the Bayesian solution. For brevity of exposition, assuming no bias in the computer model, KOH framework is given as follows

$$y^F(\cdot) = g^M(\cdot) + \epsilon(\cdot),$$
$$g^M(\cdot) \mid \mathbf{x}, \mathbf{g}^{\mathbf{M}}(\mathbf{x}) \sim \mathcal{GASP}(\mu^*(\cdot), \sigma^{*2}(\cdot,\cdot)).$$

In this framework ϵ is assumed to be white noise, so that for the ith observation

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \text{ iid } \forall\ i = 1, \ldots, n,$$

where *iid* stands for independently and identically distributed.

KOH framework targets the following posterior distribution of $\boldsymbol{\omega}$ within its framework

$$p_{KOH}(\boldsymbol{\omega} \mid \mathbf{y}^F) = \frac{\int_{\boldsymbol{\Xi}} p(\mathbf{y}^F \mid \boldsymbol{\omega}, \boldsymbol{\xi}) p(\boldsymbol{\omega}, \boldsymbol{\xi}) d\boldsymbol{\xi}}{\int_{\boldsymbol{\Omega}, \boldsymbol{\Xi}} p(\mathbf{y}^F \mid \boldsymbol{\omega}, \boldsymbol{\xi}) p(\boldsymbol{\omega}, \boldsymbol{\xi}) d\boldsymbol{\omega} d\boldsymbol{\xi}}. \tag{14}$$

The targeting posterior from Bayesian inverse calibration framework (13) is given by the distribution $p(\boldsymbol{\omega} \mid \mathbf{y}^F)$.

$$p(\boldsymbol{\omega} \mid \mathbf{y}^F) = \int_{\mathcal{Y}} \frac{p(\mathbf{y} \mid \boldsymbol{\omega}) p(\boldsymbol{\omega})}{\int_{\boldsymbol{\Omega}} p(\mathbf{y} \mid \boldsymbol{\omega}) p(\boldsymbol{\omega}) d\boldsymbol{\omega}} \frac{\int_{\boldsymbol{\Xi}} p(\mathbf{y}^F \mid \mathbf{y}, \boldsymbol{\xi}) p(\mathbf{y} \mid \boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi}}{p(\mathbf{y}^F)} d\mathbf{y}. \tag{15}$$

Easy to see that the Bayesian answer (15) and KOH answer (14) are different in principle, and two frameworks target different distributions, which are not equivalent; with KOH answer leading to biased estimates of parameters. [7]

Bayesian inverse framework (15) acknowledges that there are two approximations to the reality, i.e. to the truth. One approximation comes from the computer model (and, thus, an emulator of that model), another approximation comes from experimental observations. Bayesian calibration tries to reconcile these two approximations. The true Bayesian answer also allows for incorporation of *any* noise structure in the collected data, i.e. beyond the standard assumption of Gaussian additive noise.

A case study is provided in the Appendix for illustrative purposes of implementation of the framework in practice.

[5] If $p = 2$, the minimization corresponds to the least squares method, that of the L^2-norm.

[6] "Bayesian calibration of computer models", Journal of Royal Statistical Society, Series B, 2001.

[7] This is especially so for the estimate of the noise parameter in the data. See illustrative example in the PhD thesis of Ksenia N. Kyzyurova "On Uncertainty Quantification for Systems of Computer Models", Duke University, 2017 (Kyzyurova 2017)

4 Conclusion

Fully Bayesian approach to calibration of a complex computer model with respect to collected data has been outlined and illustrated with simulation examples. Next research directions are to formulate a number of relevant practical examples together with investigation of its implementation challenges.

One potential example is to perform calibration of the model of the volcano eruption plume (Bursik et al. 2012) from a system of volcano ash transport and dispersal computer model with respect to collected data reported in Gudmundsson et al. (2012) on sizes of observed volcano ash particles during Eyjafjallajökull, Iceland volcano eruption in April-May 2010. This may lead to useful inference for construction of systems which may predict volcano eruptions.

Another interesting research direction is investigation of simultaneous construction of designs for computer model data together with designs for experimental data. Since both designs influence the quality of calibration result, this research should lead to formulation of design strategies for robust inference in calibration tasks.

5 Appendix

5.1 Appendix A: Correlation structure of the linked GASP in a special case

Theorem. Let g^M with given parameters $\theta_{\mathbf{g}} = (\beta, \sigma^2, \{\delta_j\}_{j=1,\ldots,d}, \eta)$ be a GASP emulator of a simulator g evaluated at training input points $\mathbf{x}$, i.e. $g^M(\cdot)$. Suppose the mean function is linear in the bth cooordinate of an input $\mathbf{x}'$, so that $\mu(\mathbf{x}') = \beta_0 + \beta_1 x'_b$. Let $g^M(\cdot)$ have a product power exponential correlation function $c(\mathbf{x_k}, \mathbf{x_l}) = \prod_{j=1}^{d} c(x_{kj}, x_{lj})$, where

$$c(x_{kj}, x_{lj}) = \exp\left\{-\left(\frac{|x_{kj} - x_{lj}|}{\delta_j}\right)^{\alpha_j}\right\}, \tag{16}$$

with a range parameter $\delta_j \in (0, \infty)$ and a smoothness parameter $\alpha_j \in (0, 2]$ along the jth coordinate.

Let smoothness parameters α_j of coordinates $j \in b, \ldots, d$ be equal to 2. For each $j \in b, \ldots, d$ let f_j^M be an independent emulator of a simulator f_j, corresponding to the coordinate j of the input to the simulator g. Then covariance $\mathrm{Cov}(\xi_1, \xi_2)$ between two variables of the linked emulator $\xi_1 = (g \circ (f_b, \ldots, f_d))(\mathbf{u_1})$ and $\xi_2 = (g \circ (f_b, \ldots, f_d))(\mathbf{u_2})$ evaluated at new inputs $\mathbf{u_1}$ and $\mathbf{u_2}$ to the coupled simulator $(g \circ (f_b, \ldots, f_d))(\cdot)$ is

$$\begin{aligned}
\mathrm{Cov}(\xi_1, \xi_2) = \beta_0^2 + \beta_0\beta_1(\mu_{b1}^* + \mu_{b2}^*) + \beta_1^2(\mu_{b1}^*\mu_{b2}^* + \rho_b\sigma_{b1}^*\sigma_{b2}^*) - (\mathrm{E}\xi_1)(\mathrm{E}\xi_2) + \\
\sigma^2 \prod_{j=1}^{b-1} c(u_{1j}, u_{2j}) \prod_{j=b}^{d} J_j + \sum_{k,l=1}^{m} (a_k a_l - \sigma^2\{C_x^{-1}\}_{k,l}) \prod_{j=1}^{b-1} c(u_{1j}, x_{kj}) c(x_{lj}, u_{2j}) \prod_{j=b}^{d} J_j^{k,l} + \\
\beta_0 \sum_{i=1}^{m} a_i \left(\prod_{j=1}^{b-1} c(u_{2j}, x_{ij}) \prod_{j=b}^{d} J_{j,2}^i + \prod_{j=1}^{b-1} c(u_{1j}, x_{ij}) \prod_{j=b}^{d} J_{j,1}^i \right) + \\
\beta_1 \sum_{i=1}^{m} a_i \left(\prod_{j=1}^{b-1} c(u_{2j}, x_{ij}) J_{2\,b}^{+i} \prod_{j=b+1}^{d} J_{j,2}^i + \prod_{j=1}^{b-1} c(u_{1j}, x_{ij}) J_{1\,b}^{+i} \prod_{j=b+1}^{d} J_{j,1}^i \right),
\end{aligned} \tag{17}$$

where $a = (a_1, \ldots, a_m)^{\mathrm{T}} = C_x^{-1}(\mathbf{g^M}(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{x}))$ and

$$\begin{pmatrix} f_{j_1}^M(\mathbf{u_1}) \\ f_{j_2}^M(\mathbf{u_2}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{j_1}^* \\ \mu_{j_2}^* \end{pmatrix}, \begin{pmatrix} \sigma_{j_1}^{*2} & \rho_j\sigma_{j_1}^*\sigma_{j_2}^* \\ \rho_j\sigma_{j_1}^*\sigma_{j_2}^* & \sigma_{j_2}^{*2} \end{pmatrix} \right) \tag{18}$$

and

$$J_j = \frac{\delta_j e^{-\frac{(\mu_{j1}^* - \mu_{j2}^*)^2}{\delta_j^2 + 2(\sigma_{j1}^{*2} + \sigma_{j2}^{*2} - 2\rho_j \sigma_{j1}^* \sigma_{j2}^*)}}}{\sqrt{\delta_j^2 + 2(\sigma_{j1}^{*2} + \sigma_{j2}^{*2} - 2\rho_j \sigma_{j1}^* \sigma_{j2}^*)}}, \tag{19a}$$

$$J_j^{k,l} = \frac{\delta_j^2 e^{-2\frac{(\frac{\delta_j^2}{2} + \sigma_{j2}^{*2})(x_{kj} - \mu_{j1}^*)^2 + (\frac{\delta_j^2}{2} + \sigma_{j1}^{*2})(x_{lj} - \mu_{j2}^*)^2 - 2\rho_j \sigma_{j1}^* \sigma_{j2}^* (x_{kj} - \mu_{j1}^*)(x_{lj} - \mu_{j2}^*)}{\delta_j^4 + 2\delta_j^2(\sigma_{j1}^{*2} + \sigma_{j2}^{*2}) + 4(1-\rho_j^2)\sigma_{j1}^{*2}\sigma_{j2}^{*2}}}}{\sqrt{\delta_j^4 + 2\delta_j^2(\sigma_{j1}^{*2} + \sigma_{j2}^{*2}) + 4(1-\rho_j^2)\sigma_{j1}^{*2}\sigma_{j2}^{*2}}}, \tag{19b}$$

$$J_{j,k}^i = \frac{1}{\sqrt{1 + 2\frac{\sigma_{jk}^{*2}}{\delta_j^2}}} \exp\left(-\frac{(x_{ij} - \mu_{jk}^*)^2}{\delta_j^2 + 2\sigma_{jk}^{*2}}\right), \tag{19c}$$

$$J_{1\ b}^{+i} = \frac{\delta_b e^{-\frac{(\mu_{b1}^* - x_{ib})^2}{\delta_b^2 + 2\sigma_{b1}^{*2}}} ((\delta_b^2 + 2\sigma_{b1}^{*2})\mu_{b2}^* + 2\rho_b \sigma_{b2}^* \sigma_{b1}^* (x_{ib} - \mu_{b1}^*)}{\sqrt{(\delta_b^2 + 2\sigma_{b1}^{*2})^3}}, \tag{19d}$$

$$J_{2\ b}^{+i} = \frac{\delta_b e^{-\frac{(\mu_{b2}^* - x_{ib})^2}{\delta_b^2 + 2\sigma_{b2}^{*2}}} ((\delta_b^2 + 2\sigma_{b2}^{*2})\mu_{b1}^* + 2\rho_b \sigma_{b1}^* \sigma_{b2}^* (x_{ib} - \mu_{b2}^*)}{\sqrt{(\delta_b^2 + 2\sigma_{b2}^{*2})^3}}. \tag{19e}$$

5.2 Appendix B: Gibbs sampling from the model for experimental data with iid Gaussian noise

Typically, however, experimental data $\mathbf{y}$ are observed with noise ϵ. An additive structure on the truth $\mathbf{y}$ and noise ϵ is usually assumed

$$y_i^F = y_i + \epsilon_i, i = 1, \ldots, n. \tag{20}$$

This means that true value of experimental data $\mathbf{y}$ are not known, but may be estimated from the observed field data $\mathbf{y}^F$ if distribution of ϵ is known. Within the Bayesian framework posterior distribution $p(\mathbf{y} \mid \mathbf{y}^F)$ on the true values of $\mathbf{y}$ may be obtained.

Often ϵ is assumed to be white noise, so that for the ith observation

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2) \text{ iid } \forall\ i = 1, \ldots, n. \tag{21}$$

$$y_{ij} \sim \mathcal{N}(\mu_j, \sigma^2) \tag{22}$$

for ith observation within group $j \in 1, \ldots, J$. The model is complete with priors $p(\mu_1, \ldots, \mu_J, \sigma^2) \propto \frac{1}{\sigma^2}$. Estimates of parameters are easy to get via Gibbs sampling.

Full conditionals

$$\mu_j \mid \sigma^2, \mathbf{y} \sim \mathcal{N}\left(\frac{\sum_{i=1}^{N_j} y_{ij}}{N_j}, \frac{\sigma^2}{N_j}\right), \tag{23}$$

$$\sigma^2 \mid \boldsymbol{\mu}, \mathbf{y} \sim \mathcal{IG}\left(\text{shape} = \frac{\sum_{j=1}^{J} N_j}{2}, \text{rate} = \frac{\sum_{j=1}^{J} \sum_{i=1}^{N_j} (y_{ij} - \mu_j)^2}{2}\right), \tag{24}$$

where N_j is the number of observations in group j.

Table 1: Performance of emulator of the second simulator.

$\text{RMSPE}_{\text{base}}/\text{RMSPE}$	EFC	L_{CI}
1.8352	98%	5.43

5.3 Appendix C: Correlated noisy observations

$$\begin{pmatrix} y_{i,1} \\ \vdots \\ y_{i,N_j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{N_j} \end{pmatrix}, \sigma^2 \Sigma \right) . \tag{25}$$

for the ith set of observations in groups $j \in 1, \ldots, J$. We assume that the sets of observations are independent of each other. Matrix Σ is a matrix of correlations between observations in different groups. We will assume that this matrix is known.

The inference in this model, completing model specification as in the previous section 5.2, via Gibbs sampling requires a change in the full conditional for variance.

$$\sigma^2 \mid \mathbf{y}, \boldsymbol{\mu} \sim \mathcal{IG}\left(\text{shape} = \frac{\sum_{j=1}^{J} N_j}{2}, \text{rate} = \frac{\sum_{i=1}^{N_i} \mathbf{v}\Sigma^{-1}\mathbf{v}^{\mathrm{T}}}{2} \right) \tag{26}$$

where $\mathbf{v} = \begin{pmatrix} y_{i,1} - \mu_1, & \ldots, & y_{i,J} - \mu_J \end{pmatrix}^{\mathrm{T}}$ and N_i is the number of experiments.

5.4 Appendix D: Case study

Motivated originally by the problem of calibration of large-scale engineering systems which could not or have not yet been observed, a case study is provided for instructional purpose. As before, the coupled system of models $z = g \circ f$ consists of individual models g and f. The model g is mathematical, while the model f is described by an empirical physical law.

Damped sine function of the form

$$g(\mathbf{x} = (x_1, x_2), \boldsymbol{\omega} = (\omega_1, \omega_2), y_1) = x_2 \exp\{-x_1/y_1\} \cos\{\omega_1 x_1 + \omega_2\}$$

is a simulator g which has 5 inputs $X_1, X_2, \omega_1, \omega_2$ and Y_1 and the output is Z. The construction and the performance of the corresponding emulator of this model is described below.

A GASP emulator of the model 27, based on 50 runs of this simulator, is constructed. The emulator is tested on another 50 test points. Parameters $\boldsymbol{\omega} = (\omega_1, \omega_2)$ are parameters to calibrate.

$$Z = \mathcal{GASP}(X_1, X_2, \omega_1, \omega_2, Y_1) \tag{27}$$

The emulator performs very well, as can be seen from the summary provided in Table 1. Root-mean square predictive error (RMSPE) is 1.83 times smaller than the base estimate $\text{RMSPE}_{\text{base}}$, while frequency coverage is 98%.

As for the model f: quantity Y_1 is assumed to be proportional to X_1 due to some empirical linear physical law, which requires two main parameters, intercept θ_1 and slope θ_2 for its statistical description

$$Y_1 = \mathcal{N}\left(\theta_1 + \theta_2 X_1, \sigma^2\right) . \tag{28}$$

Additional parameter is variance σ^2 (or standard deviation σ) which accounts for variability in the data.

In this illustrative example the data are generated from this model with true values of parameters $\boldsymbol{\theta} = (19, 6)$ and true noise variance $\sigma^2 = 0.04$. We imagine that 11 data points corresponding to this model has been

Table 2: Case study. Posterior summaries for parameters $\boldsymbol{\theta}$. Namely, 2.5% and 97.5% of the posteriors distributions are shown.

	2.5%	97.5%
$\theta_1 \mid \mathbf{X}_1, \mathbf{Y}_1$	18.79	19.30
$\theta_2 \mid \mathbf{X}_1, \mathbf{Y}_1$	5.33	6.25

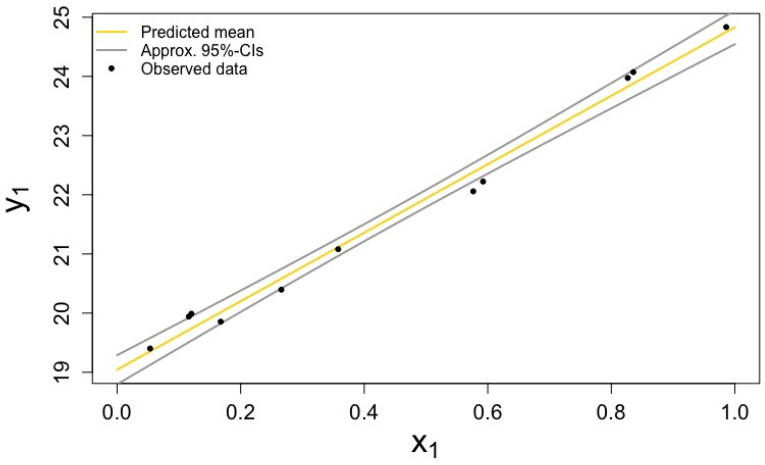

Figure 4: The first model. Black dots correspond to 11 collected data points.

collected. Now the goal is to perform calibration for parameters $\boldsymbol{\omega}$, by first calibrating the model f, and then the coupled model $z = g \circ f$.

With 11 experimental pairs of data points $\mathbf{X}_1, \mathbf{Y}_1$ corresponding to model f, we perform Bayesian inversion and obtain inferential relationship between X_1 and Y_1. Posterior summaries of parameters θ are available in Table 2. The resulting relationship with a 95% credible region is shown in Figure 4. The implementation of inference from this statistical model is based on the Gibbs sampling algorithm. The sampling involves full conditionals which are derived and stated in the Appendix 5.5.

We construct the linked emulator of two models and then calibrate parameters $\boldsymbol{\omega}$ using the linked GASP of the complex model and 20 experimental data points corresponding to the output of the model g. Experimental data are sampled from the true bivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu = (5, 9\pi/10)^{\mathrm{T}}$ and Σ is variance-covariance matrix with corresponding standard deviations along the two dimensions being 0.1 and 0.05 respectively, with correlation coefficient between variables in two dimensions equal to $\rho = -0.6$.

We assume that experimental observations are noise-free, while true parameters $\boldsymbol{\omega}$ follow some distribution (rather than being point-wise truth with some assumed noise structure). The result of the calibration is shown with bivariate density plot for parameters $\boldsymbol{\omega}$ in Figure 5.

We check that the match between the output of an emulator and the experimental data is acceptable. Figure 6 shows the predictive distribution evaluated at the highest posterior density point of calibrated parameters $\boldsymbol{\omega}$ together with the experimental data. From the plot we see that, indeed, the developed emulator matches the experimental data well. All 20 points of experimental data points stay within 95% credible area of the linked emulator.

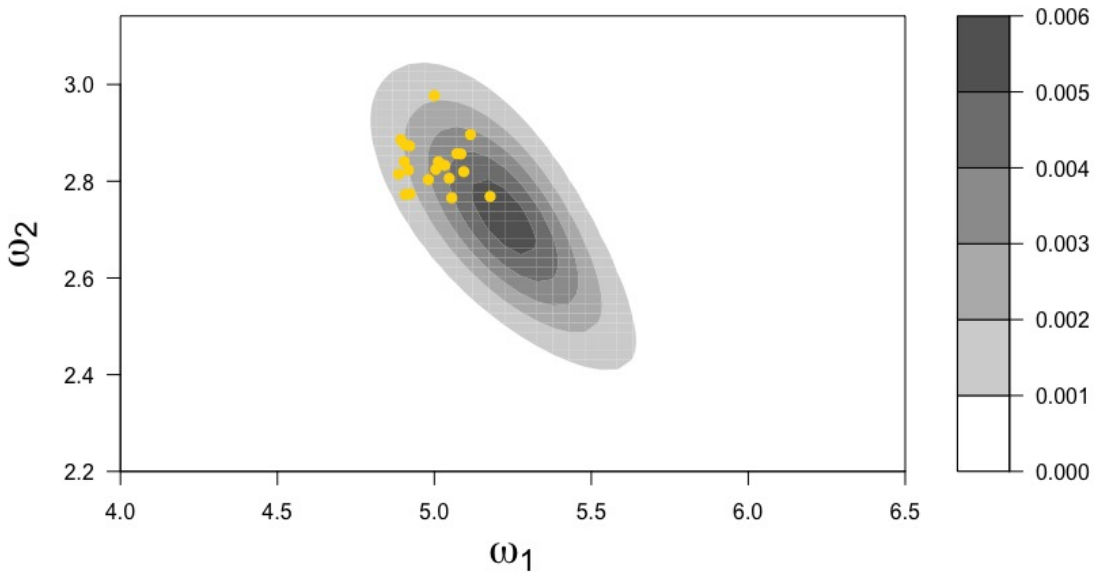

Figure 5: Posterior probability density for $\boldsymbol{\omega}$.

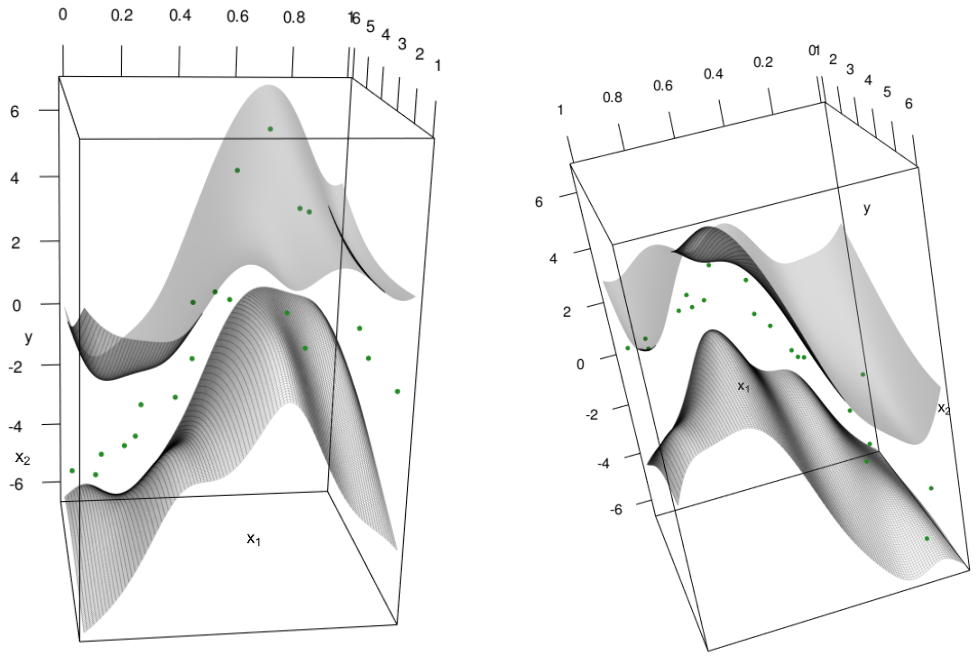

Figure 6: Predictive distribution from the linked emulator evaluated at the highest posterior density point of calibrated parameters $\boldsymbol{\omega}$, together with experimental data. Two grey surfaces show central 95% credible region of the emulator over the region of inputs x_1 and x_2 and output Z. Gridded surface shows the mean of the emulator. Green points correspond to experimental data. Two perspectives on the emulators surfaces summaries are given for clear visualization.

5.5 Appendix E: Gibbs sampling from the statistical model from the case study

$$\theta_1 \mid \theta_2, \sigma^2, \mathbf{y}, \mathbf{x^f} \sim \mathcal{N}\left((\overline{\mathbf{y}} - \theta_2\overline{\mathbf{x^f}}), \text{var} = \frac{\sigma^2}{n}\right), \tag{29}$$

$$\theta_2 \mid \theta_1, \sigma^2, \mathbf{y}, \mathbf{x^f} \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n}(y_i - \theta_1)x_i^f}{\sum_{i=1}^{n}\{x_i^f\}^2}, \text{var} = \frac{\sigma^2}{\sum_{i=1}^{n}\{x_i^f\}^2}\right), \tag{30}$$

$$\sigma^2 \mid \mathbf{y}, \mathbf{x^f}, \boldsymbol{\theta} \sim \mathcal{IG}\left(\text{shape} = \frac{n}{2}, \text{rate} = \frac{\sum_{i=1}^{n}(y_i - \boldsymbol{\theta}^{\mathrm{T}}(1, x_i)^{\mathrm{T}})^2}{2}\right). \tag{31}$$

where n is the number of observational data points.

References

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H. & Tu, J. (2007), 'A framework for validation of computer models', *Technometrics* **49**(2), 138–154.

Berger, J. O., De Oliveira, V. & Sansó, B. (2001), 'Objective Bayesian analysis of spatially correlated data', *Journal of the American Statistical Association* **96**(456), 1361–1374.

Bursik, M., Jones, M., Carn, S., Dean, K., Patra, A., Pavolonis, M., Pitman, E. B., Singh, T., Singla, P. & Webley, P. (2012), 'Estimation and propagation of volcanic source parameter uncertainty in an ash transport and dispersal model: application to the Eyjafjallajokull plume of 14–16 April 2010', *Bulletin of volcanology* **74**(10), 2321–2338.

Chkrebtii, O. A., Campbell, D. A., Calderhead, B., Girolami, M. A. et al. (2016), 'Bayesian solution uncertainty quantification for differential equations', *Bayesian Analysis* **11**(4), 1239–1267.

Ghanem, R. (2018), 'Modeling and Algorithmic Aspects of UQ for Material with Multiscale Behavior', *SAMSI workshop "Modeling Uncertainty: Mathematical and Statistical (MUMS)"* .

Gramacy, R. B. & Lee, H. K. (2012), 'Cases for the nugget in modeling computer experiments', *Statistics and Computing* **22**(3), 713–722.

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

Gudmundsson, M. T., Thordarson, T., Höskuldsson, Á., Larsen, G., Björnsson, H., Prata, F. J., Oddsson, B., Magnússon, E., Högnadóttir, T., Petersen, G. N. et al. (2012), 'Ash generation and distribution from the April-May 2010 eruption of eyjafjallajökull, iceland', *Scientific reports* **2**, 572.

Kyzyurova, K. N. (2017), On Uncertainty Quantification for Systems of Computer Models, PhD thesis, Duke University.

Kyzyurova, K. N. (2019*a*), *Analysis of scientific computer models. Methodology in numerical simulator data analysis*.

Kyzyurova, K. N. (2019*b*), 'On scoring rules and frequency predictive measures'.

Kyzyurova, K. N., Berger, J. O. & Wolpert, R. L. (2018), 'Coupling Computer Models through Linking Their Statistical Emulators', *SIAM/ASA Journal on Uncertainty Quantification* **6**(3), 1151–1171.

Madankan, R., Pouget, S., Singla, P., Bursik, M., Dehn, J., Jones, M., Patra, A., Pavolonis, M., Pitman, E., Singh, T. et al. (2014), 'Computation of probabilistic hazard maps and source parameter estimation for volcanic ash transport and dispersion', *Journal of Computational Physics* **271**, 39–59.

Pitman, E. B. & Le, L. (2005), 'A two-fluid model for avalanche and debris flows', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **363**(1832), 1573–1601.

Pitman, E. B., Nichita, C. C., Patra, A., Bauer, A., Sheridan, M. & Bursik, M. (2003), 'Computing granular avalanches and landslides', *Physics of fluids* **15**(12), 3638–3646.

Santner, T. J., Williams, B. J., Notz, W. I. & Williams, B. J. (2003), *The design and analysis of computer experiments*, Vol. 1, Springer.

Stuart, A. & Teckentrup, A. (2018), 'Posterior consistency for Gaussian process approximations of Bayesian posterior distributions', *Mathematics of Computation* **87**(310), 721–753.