Surrogates 7020

Chapter 8: Calibration Model

Dr. Alex Bledar Konomi

Department of Mathematical Sciences
University of Cincinnati

K-O Calibration Framework

The calibration data are composed of field (experimental) measurements and computer model data. We denote the response of the n field measurement by $\boldsymbol{z} = (z_1, \ldots, z_n)$, where each component z_i is being subject to:

$$z_i = \zeta(\boldsymbol{x}_i) + e_i,$$

where $\zeta(\boldsymbol{x}_i)$ denotes the response of the actual physical system, $\boldsymbol{x}_i$ are the observable inputs, and e_i denotes the nugget error the i^{th} observation.
The computer model aims to simulate the real system. The input of the computer model consists of $q-$dimensional observable input $\boldsymbol{x}$ and $p-$dimensional calibration input values t.
The computer model output is an unknown function $\eta(\boldsymbol{x}, t)$ of the observable and calibration input, which simulates a physical system.

Calibration Model

The observations z_i, the true process $\zeta(\cdot)$, and the computer model function $\eta(\cdot, \cdot)$ are linked through:

$$z_i = \zeta(\boldsymbol{x}_i) + e_i = \eta(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i) + e_i, \tag{1}$$

- $\delta(\cdot)$ is a model disagreement between the real system from the computer model.
- $\boldsymbol{\theta}$ denotes the best fixed but unknown setting for the calibration input t.
- Each of the n field measurements $\eta(\cdot, \cdot)$ consists of the known observable input value $\boldsymbol{x}_i$, and the unknown $p-$dimensional vector calibration parameter $\boldsymbol{\theta}$, which is considered fixed (for each of the n field measurements).
- Here the $\boldsymbol{x}_i$ is considered fixed and we are interested to infer the calibration parameters $\boldsymbol{\theta}$

Computer model output: dealing with the nugget

- The output of the computer model is $y = \eta(\boldsymbol{x}, t) + v$, where v represents the nugget error of the computer model.
- For m computer experiment runs at input points $((\boldsymbol{x}_1^*, t_1), \ldots, (\boldsymbol{x}_m^*, t_m))$ (both observable and calibration input) we denote the output as $\boldsymbol{y} = (y_1, \ldots, y_m)$, where $y_j = \eta(\boldsymbol{x}_j^*, t_j) + v_j$.
- Although some computer models are generated by deterministic solvers with no random error, to avoid an infinite differentiability covariance function it is better to add a nugget effect in the statistical model (Stein, 1999).
- Moreover, Gramacy and Lee (2012) argue that the use of a nugget helps protect against poor fits when assumptions are violated. Statistical emulators for computer models are not exact in practice.

Gaussian Process Model

Typically, the unknown functions $\eta(\cdot,\cdot)$ and $\delta(\cdot)$ are modeled as two independent Gaussian processes (GP) (Kennedy and O'Hagan, 2001; Higdon et al., 2004; Williams et al., 2006; Higdon et al., 2008), that is:

$$\eta(\cdot,\cdot) \sim N(\mu_\eta(\cdot,\cdot), c_\eta((\cdot,\cdot),(\cdot,\cdot)))$$

and

$$\delta(\cdot) \sim N(\mu_\delta(\cdot), c_\delta(\cdot,\cdot)).$$

We will refer to this method as standard Bayesian Gaussian process calibration (SBGPC). For $\eta(\cdot,\cdot)$ and $\delta(\cdot)$ the mean is usually assumed to be a linear model as: $\mu_\eta(\boldsymbol{x},t) = \boldsymbol{h}_\eta(\boldsymbol{x},t)^T\boldsymbol{\beta}_\eta$ and $\mu_\delta(\boldsymbol{x}) = \boldsymbol{h}_\delta(\boldsymbol{x})^T\boldsymbol{\beta}_\delta$.

Choise for the Covaraince Function

The covariance function of $\eta(\cdot, \cdot)$ is modeled in a separable form, as in Kennedy and O'Hagan (2001):

$$c_\eta((\boldsymbol{x}, t), (\boldsymbol{x}', t')) = \sigma_\eta^2 \rho(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\phi}_{\eta,x}) \rho(t, t'; \boldsymbol{\phi}_{\eta,t}),$$

and the covariance function of $\delta(\cdot)$ is:

$$c_\delta(\boldsymbol{x}, \boldsymbol{x}') = \sigma_\delta^2 \rho(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\phi}_{\delta,x}),$$

where σ_η^2 and σ_δ^2 are the variance of $\eta(\cdot, \cdot)$ and $\delta(\cdot)$, correspondingly, and ρ denotes the correlation function.

Choice for the Covariance Function

- The correlation function, ρ, is of particular importance as it defines the smoothness of the random field.
- Different choices, such as Matérn and power exponential covariance family, can be made.
- The separable power exponential covariance family is considered as a standard choice in the computer experiments (Santner et al., 2003) where the dimensionality of the input can be usually high.
- In specific, for the squared exponential family $\rho(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\phi}_{\eta,x}) = \exp\left(-\frac{1}{2} \sum_{l=1:q} \frac{||\boldsymbol{x}_l - \boldsymbol{x}'_l||^2}{\phi^2_{l,x}} \right)$, where $\phi_{l,x}$ is the correlation strength in the l direction.
- Different inputs usually have different meaning. Therefore, it is preferable to have different correlation parameters.
- The same formulations can be applied for $\rho(t, t'; \boldsymbol{\phi}_{\eta,t})$ and $\rho(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\phi}_{\delta,x})$.

Define Observation

- For simplicity in the formulation, we denote the $(q+p)$ dimensional input space, both observational and calibration, by $\mathcal{D}$.
- We also denote the observable input points of the n field measurements by $\boldsymbol{D}^1 = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, and the set of observable input points augmented with the calibration parameter by $\boldsymbol{D}^1(\boldsymbol{\theta}) = \{(\boldsymbol{x}_1, \boldsymbol{\theta}), \ldots, (\boldsymbol{x}_n, \boldsymbol{\theta})\}$.
- In addition, we denote the set of input points (both observable and calibration) of the computer model by $\boldsymbol{D}^2 = \{(\boldsymbol{x}_1^*, t_1), \ldots, (\boldsymbol{x}_m^*, t_m)\}$. We also represent all of the input dataset augmented by θ as $\boldsymbol{D} = (\boldsymbol{D}^1(\boldsymbol{\theta}), \boldsymbol{D}^2)$ and their output as $\boldsymbol{d} = (\boldsymbol{z}, \boldsymbol{y})$.
- The dataset consist of both field and computer model output. We denote by $\boldsymbol{D}^1(\boldsymbol{\theta})$ the input from the field and $\boldsymbol{D}^2$ the input from the computer model. $\boldsymbol{D} = \{\boldsymbol{D}^1(\boldsymbol{\theta}), \boldsymbol{D}^2\}$ input points and the corresponding output $\boldsymbol{d} = (\boldsymbol{z}, \boldsymbol{y})$.

Likelihood:

We model the output $\{\boldsymbol{d}\}$ with the SBGPC (explained above). For a given calibration parameter $\boldsymbol{\theta}$, and GP parameters $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}\} = (\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2)$, the likelihood is proportional to,

$$f(\boldsymbol{d}|\boldsymbol{\Theta}, \boldsymbol{\theta}) \propto |\boldsymbol{V_d}|^{-1/2} \exp[-\frac{1}{2}(\boldsymbol{d} - E(\boldsymbol{d}))^T \boldsymbol{V_d}^{-1}(\boldsymbol{d} - E(\boldsymbol{d}))],$$

where $E(\boldsymbol{d})$ and $\boldsymbol{V_d}$ are the mean and variance of the output $\boldsymbol{d}$.

Mean and Variance of the parameters

In order to represent more explicitly the mean and variance of each component in the preceding formulation, we introduce some new symbols:

- Let $\boldsymbol{H}_\eta(\boldsymbol{D}^2)$ denotes the matrix with rows $\boldsymbol{h}_\eta(\boldsymbol{x}, t)$ for each $(\boldsymbol{x}, t) \in \boldsymbol{D}^2$, and $\boldsymbol{H}_\delta(\boldsymbol{D}^1)$ denotes the matrix with rows $\boldsymbol{h}_\delta(\boldsymbol{x})$ for each $\boldsymbol{x} \in \boldsymbol{D}^1$.
- Let $\boldsymbol{V}_\eta(\boldsymbol{D}) = \boldsymbol{C}_\eta(\boldsymbol{D}, \boldsymbol{D})$ be the covariance matrix with (i, i') elements, $c_\eta((\boldsymbol{x}_i, t_i), (\boldsymbol{x}_{i'}, t_{i'})$ for every pair $(\boldsymbol{x}_i, t_i) \in \boldsymbol{D}$, and $(\boldsymbol{x}_{i'}, t_{i'}) \in \boldsymbol{D}$.
- Similarly, we define $\boldsymbol{V}_\delta(\boldsymbol{D}^1) = \boldsymbol{C}_\delta(\boldsymbol{D}^1, \boldsymbol{D}^1)$.

Mean and Variance of the parameters

The mean of the output is:

$$E(\boldsymbol{d}) = \boldsymbol{H}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{H}_\eta(\boldsymbol{D}^1(\boldsymbol{\theta})) & \boldsymbol{H}_\delta(\boldsymbol{D}^1) \\ \boldsymbol{H}_\eta(\boldsymbol{D}^2) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_\eta \\ \boldsymbol{\beta}_\delta \end{bmatrix},$$

and its covariance matrix is:

$$\boldsymbol{V_d} = cov(\boldsymbol{d}, \boldsymbol{d}) = \boldsymbol{V}_\eta(\boldsymbol{D}) + \begin{pmatrix} \boldsymbol{V}_\delta(\boldsymbol{D}^1) + \tau_e^2 \boldsymbol{I}_n & 0 \\ 0 & \tau_v^2 \boldsymbol{I}_m \end{pmatrix},$$

where τ_e^2 and τ_v^2 are the variances of the nuggets, and $\boldsymbol{I}_n$ and $\boldsymbol{I}_m$ are identity matrices of dimension $n \times n$ and $m \times m$ respectively.

Prior:

We assign a prior distribution on the parameter $(\mathcal{T}, \boldsymbol{\theta}, \boldsymbol{\Theta})$, such as:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\Theta}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{\Theta}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{\beta}_\eta, \sigma_\eta^2)\pi(\boldsymbol{\phi}_\eta)\pi(\boldsymbol{\beta}_\delta, \sigma_\delta^2)\pi(\boldsymbol{\phi}_\delta)\pi(\tau_e)\pi(\tau_v).$$

the prior distributions of the GP hyper-parameters are:

$$\pi(\boldsymbol{\beta}_\eta, \sigma_\eta^2)\pi(\boldsymbol{\phi}_\eta) \propto \frac{1}{\sigma_{\eta_k}^2} \prod_{l=1:(q+p)} [G(\phi_{\eta,l}|\alpha_{G,1}, \beta_{G,1})]$$

$$\pi(\boldsymbol{\beta}_\delta, \sigma_\delta^2)\pi(\boldsymbol{\phi}_\delta) \propto \frac{1}{\sigma_\delta^2} \prod_{l=1:p} [G(\phi_{\delta,l}|\alpha_{G,1}, \beta_{G,1})],$$

where $(q+p)$ is the dimension of the experimental input and calibration space, q is the dimension of the experimental input space, and $\alpha_{G,1}, \beta_{G,1}$ express a prior knowledge.

Calibration and nugget parameters prior

- We define the prior for $\boldsymbol{\theta}$ with a modified *Beta* distribution with parameters defined to represent previous studies or the domain scientist opinion.
- We also assign priors for the nugget hyper-parameter $\pi(\tau_e)$ and $\pi(\tau_v)$ as exponential distribution to ensure positive values.

Posterior:

The posterior distribution is known up to a normalizing constant as:

$$\begin{aligned} p(\boldsymbol{\Theta}, \boldsymbol{\theta}|\boldsymbol{d}) \propto & \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi}_\eta, \boldsymbol{\beta}_\eta, \sigma_\eta^2)|\boldsymbol{V_d}|^{-1/2} \\ & \times \exp[-\frac{1}{2}(\boldsymbol{d} - E(\boldsymbol{d}))^T \boldsymbol{V_d}^{-1}(\boldsymbol{d} - E(\boldsymbol{d}))]. \end{aligned} \tag{2}$$

Posterior inference for the proposed model is facilitated by an MCMC sampler.

Bayesian Inference

- Conditional on the calibration parameters $\boldsymbol{\theta}$, the full joint posterior distribution of the GP hyperparameters $(\boldsymbol{\Theta}|\boldsymbol{d}, \boldsymbol{\theta}, \mathcal{T})$ is analytically intractable.
- Exact posterior inference is performed by a customized MCMC algorithm. Analytically, we firstly sample from the closed posterior distribution of $\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\tau}^2, \boldsymbol{d}$.
- Then, we sample from the posterior distribution of $\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2|\boldsymbol{d}, \boldsymbol{\theta}$, which we find by integrating out $\boldsymbol{\beta}$ with Metropolis-Hastings (M-H).

Cont: Bayesian Inference

Given the prior specification for β and $\boldsymbol{\sigma}^2$, the close form of the posterior distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{d}$ is a multivaraite Normal distribution with mean $\hat{\boldsymbol{\beta}} = \boldsymbol{W}\boldsymbol{H}^T\boldsymbol{V_d}^{-1}\boldsymbol{d}$ and variance $\boldsymbol{W} = \boldsymbol{H}^T\boldsymbol{V_d}^{-1}\boldsymbol{H}$ is:

$$\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\tau}, \boldsymbol{d} \sim \mathcal{N}(\hat{\boldsymbol{\beta}}, \boldsymbol{W}). \tag{3}$$

Both $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{W}$ depend on $(\boldsymbol{\phi}, \boldsymbol{\tau}^2)$ and $\boldsymbol{\theta}$. Using properties of the normal density function, we integrate out $\boldsymbol{\beta}$ and compute the joint posterior distribution of $\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2|\boldsymbol{d}, \boldsymbol{\theta}$ as:

$$\begin{aligned} p(\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2|\boldsymbol{d}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{\phi})\pi(\boldsymbol{\sigma}^2)\pi(\boldsymbol{\tau}^2)|\boldsymbol{V_d}|^{-1/2}|\boldsymbol{W}|^{1/2} \\ \exp[-\frac{1}{2}(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}})^T\boldsymbol{V_d}^{-1}(\boldsymbol{d} - \boldsymbol{H}\hat{\boldsymbol{\beta}})]. \end{aligned} \tag{4}$$

Cont: Bayesian Inference

The conditional posteriors of $\boldsymbol{\phi}|\boldsymbol{\sigma}^2, \boldsymbol{\tau}$, $\boldsymbol{\sigma}^2|\boldsymbol{\phi}, \boldsymbol{\tau}$ and $\boldsymbol{\tau}|\boldsymbol{\sigma}^2, \boldsymbol{\phi}$ cannot be sampled directly. Therefore, we use Metropolis-Hastings updates within a Gibbs sampler, (Mueller, 1993; Gelfand and Smith, 1990; Hastings, 1970).
Let
$\boldsymbol{\chi} = (\boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}) = (\phi_{\eta,1}, \ldots, \phi_{\eta,(q+p)}, \phi_{\delta,1}, \ldots, \phi_{\eta,q}, \sigma_\eta^2, \sigma_\delta, \tau_e, \tau_\nu)$.
For each component of $\boldsymbol{\chi}$, χ_j for $j = 1, \ldots, (2q + p + 4)$, we perform Metropolis within Gibbs as in (Mueller, 1993). For any step of the Gibbs sampler that does not have a close form conditional posterior distribution
$p(\chi_j|\boldsymbol{d}_j, \chi_1, \ldots, \chi_{j-1}, \chi_{j-1}, \ldots, \chi_{j,(2q+p+4)})$, substitute a MH sampler.

Metropolis-Hastings updates within a Gibbs

For $j = 1, \ldots, (2q + p + 4)$, given
$\boldsymbol{\chi}_{(-j)}^{(t)} = (\chi_{k,1}^{t+1}, \ldots, \chi_{j-1}^{t+1}, \chi_{j-1}^{t}, \ldots, \chi_{(2q+p+4)}^{t})$:

1. Generate $\chi_j^* \sim q_j(\chi_j^* | \chi_j^{(t)}) \equiv \log N(\chi_j^* | \chi_j^{(t)})$ from a log Normal distribution.
2. Calculate:

$$r = \frac{p(\chi_j^* | \boldsymbol{d}, \boldsymbol{\chi}_{(-j)}^{(t)}) q_j(\chi_j^{(t)} | \chi_j^*)}{p(\chi_j^{(t)} | \boldsymbol{d}, \boldsymbol{\chi}_{(-j)}^{(t)}) q_j(\chi_j^* | \chi_j^{(t)})}. \tag{5}$$

3. Set $\chi_j^{(t+1)} = \chi_j^*$ with probability $\min(1, r)$ and $\chi_j^{(t+1)} = \chi_j^{(t)}$ with the remaining probability.

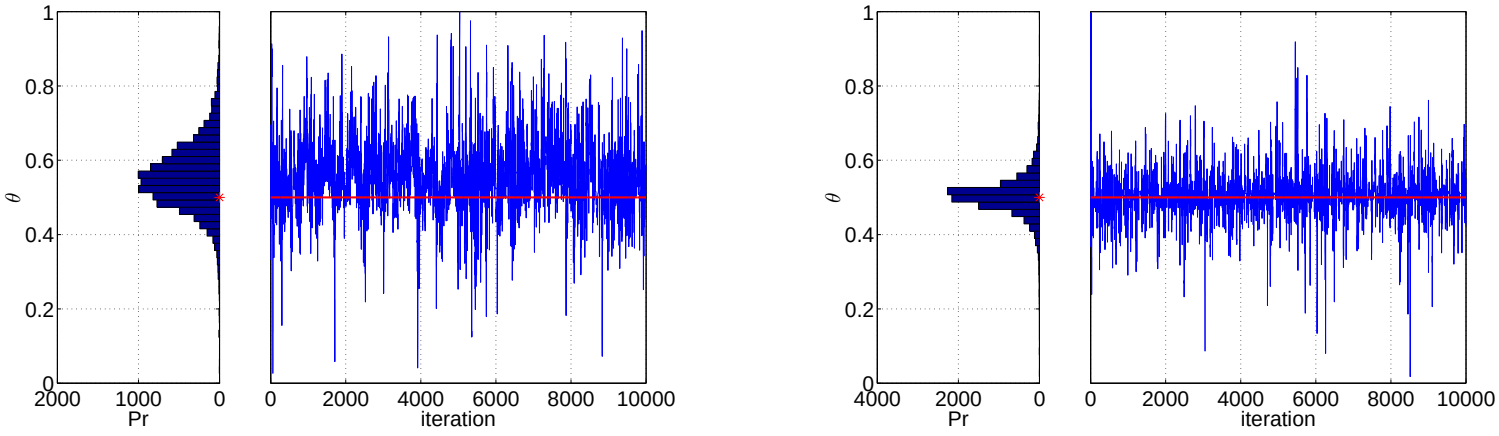
First case study:

In our first example, we assume a function of $\eta(\cdot)$ in a three-dimensional input space (one observable input x_1 and two calibration inputs t_1, t_2):

$$\eta(x_1, t_1, t_2) = \begin{cases} (1 - x_1)\cos(\pi t_1) + 0t_2, & 0 \leq x_1 < 0.4 \\ (1 - x_1)\cos(\pi t_1) + 0t_2 + 0.5, & 0.4 \leq x_1 < 1, \end{cases}$$

where the calibration variable t_2 does not affect the output of the $\eta(\cdot, \cdot, \cdot)$. From this formulation, it is clear that the computer model has a discontinuity in the observable input x_1. To make the case more realistic, we assume the computer model output has a normally distributed nugget effect with zero mean and variance $\tau_v^2 = 0.3$. We also assume the discrepancy function is $\delta(x_1) = 0.7$, and the real experimental output is $\zeta(x_1) = \eta(x_1, 0.5, 0.5) + \delta(x_1) + e$ for calibration parameters $\boldsymbol{\theta} = (0.5, 0.5)$.

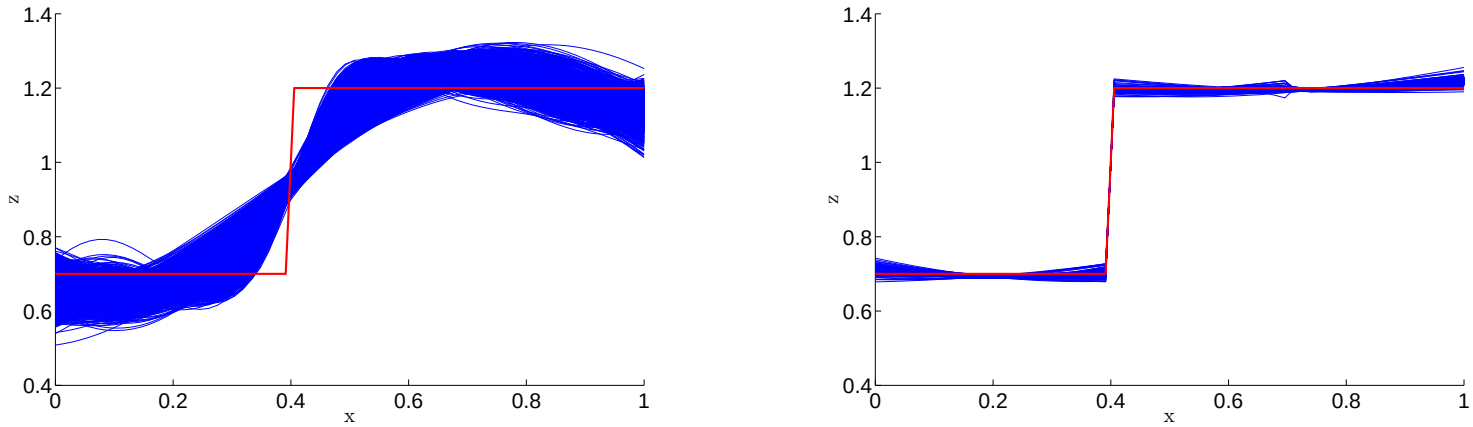
A



(a) Calibration of θ_1 using GP calibration

(b) Calibration of θ_1 using TGP calibration

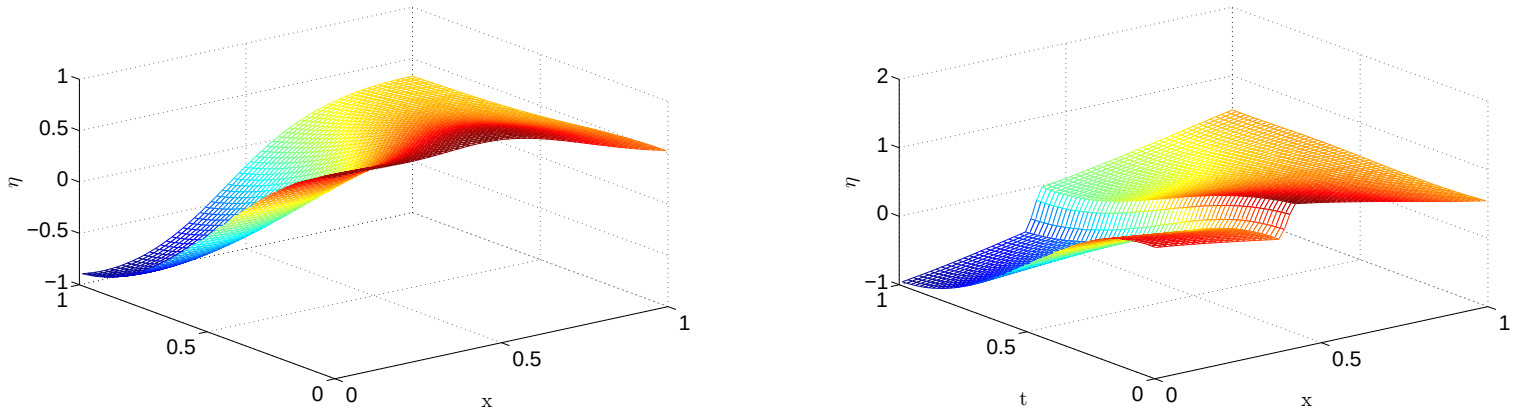
Figure: MCMC for the calibration parameter θ_1 for two different calibration methods (a) GP calibration and (b) proposed TGP calibration.



(a) Prediction of the output, ζ, using GP calibration

(b) Prediction of the output using TGP calibration

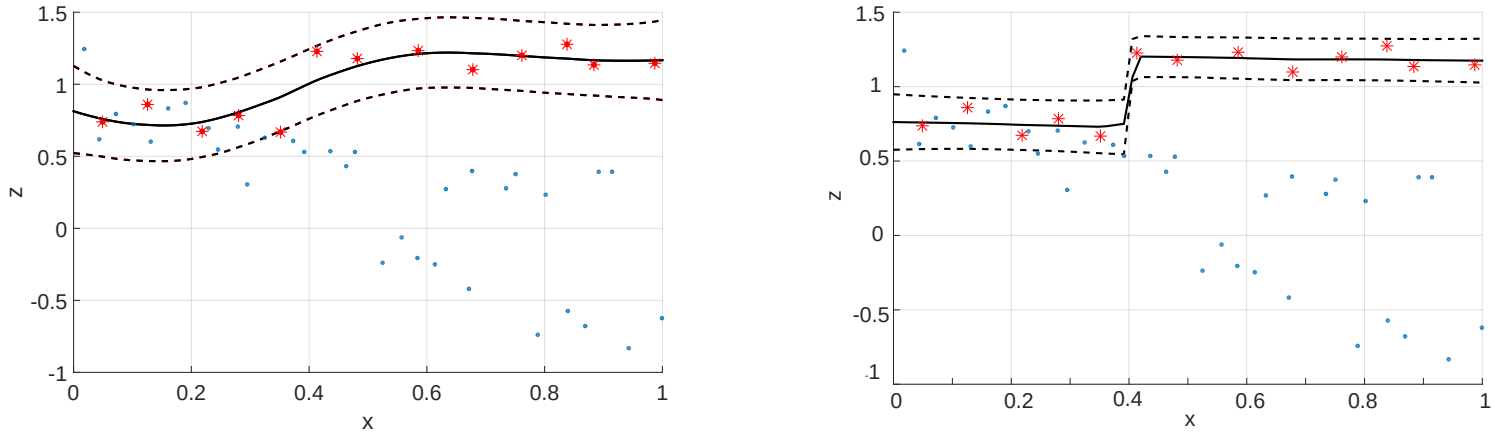
Figure: Prediction performance of the output (a) GP calibration and (b) proposed TGP calibration.



(a) Prediction of the η using GP calibration

(b) Prediction of η using TGP calibration

Figure: MCMC for the calibration parameter θ_1 for two different calibration methods (a) GP calibration and (b) proposed TGP calibration.



(a) Prediction of the real system using GP calibration

(b) Prediction of the real system using TGP calibration

Figure: Prediction mean and 95% prediction intervals using two different calibration methods (a) GP calibration and (b) proposed TGP calibration.

Field Observation

Let $Y^F(\boldsymbol{x})$ denote a field observation under m_x-dimensional conditions $\boldsymbol{x}$, and $y^R(\boldsymbol{x})$ denote the real output under condition $\boldsymbol{x}$. Assume R and F are related as follows.

$$Y^F(\boldsymbol{x}) = y^R(\boldsymbol{x}) + \epsilon,$$

where $\epsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$.

- This isn't much different from typical modeling apparatuses where observations are corrupted by independent and identically distributed idiosyncratic Gaussian noise.
- Considering the expense of setting up a physical experiment in the field, we presume that only a small number n_F of field observations Y_{n_F} are available at $\boldsymbol{x}$ locations X_{n_F}.
- Replicates can be helpful for separating signal from noise, especially when σ_ϵ^2 is large.

Computer Model

- Let $y^M(\boldsymbol{x}, \boldsymbol{u})$ denote output from a computer model run under conditions $\boldsymbol{x}$ and tuning or calibration parameters $\boldsymbol{u}$.
- We shall presume that $y^M(\cdot, \cdot)$ is deterministic to simplify the following discussion.
- There's no reason why stochastic simulation must be precluded by the framework, however such setups are far less well investigated in the literature.
- Inputs $\boldsymbol{x}$ to computer model $y^M(\boldsymbol{x}, \boldsymbol{u})$ coincide with $\boldsymbol{x}$'s from the field experiment(s). Inputs $\boldsymbol{u}$, in dimension m_u, represent any aspect of M which can't be controlled in F and/or are unknown in R.
- It's quite typical for a mathematical model, or its computer implementation, to have more knobs than can be controlled in the field.

Calibration Parameter

Example $\boldsymbol{u}$ coordinates may arise from an artificial aspect of computer implementation, like mesh size. Or they might have real physical meaning, like acceleration due to gravity, which is not known (precisely enough) to be recorded in the field. Some practitioners make a distinction between the two, calling the former a tuning parameter (omitting from probabilistic modeling enterprises), and treating only the latter as a calibration parameter $\boldsymbol{u}$. I'll be lazy by using those two terms interchangeably and modeling in a unified fashion.

K-O Calibration Framework

Kennedy and O'Hagan (2001) proposed a Bayesian framework for coupling M and F. KOH, hereafter, represent a real process R as the computer model output at the best setting of calibration parameters, $\boldsymbol{u}^*$, plus a discrepancy term acknowledging that there can be systematic disagreement between model and truth.

$$y^R(x) = y^M(x, \boldsymbol{u}^*) + b(x)$$

so that

$$Y^F(\boldsymbol{x}) = y^M(\boldsymbol{x}, \boldsymbol{u}^*) + b(\boldsymbol{x}) + \epsilon$$

The quantity $b(\cdot)$ is a functional discrepancy, or bias correction.

$$-b(\boldsymbol{x}) = y^M(\boldsymbol{x}, \boldsymbol{u}^*) - y^R(\boldsymbol{x}).$$

The point here is that a computer model has systematic imperfections, even under its best tuning $\boldsymbol{u}^*$, but KOH specify an a priori belief that reasonable correction can be learned

K-O Calibration Framework

- The point here is that a computer model has systematic imperfections, even under its best tuning $\boldsymbol{u}^*$
- KOH specify an a priori belief that reasonable correction can be learned through $b(\cdot)$.
- Errors ϵ are independent zero-mean Gaussian with variance σ_ϵ^2.

KOH framework

- Altogether, unknowns are $\boldsymbol{u}^*$, σ^2_ϵ, and discrepancy $b(\cdot)$. KOH emphasized Bayesian inference, particularly averaging over trade-offs between calibration values $\boldsymbol{u}$ and discrepancies $b(\cdot)$ under a GP prior.
- Known information or restrictions on u-values can be specified through prior $p(\boldsymbol{u})$. Otherwise a uniform prior (over a finite domain) can be used.
- Often, and especially when little prior information is available on $\boldsymbol{u}$, a regularizing prior with mass somewhat more concentrated on a default or midway value can prevent over-concentration of posterior density on boundary settings.
- Reference priors for σ^2_ϵ are typical (Berger, De Oliveira, and Sanso 2001).
- KOH utilized a GP specification with linear mean for $b(\cdot)$, but the presentation here considers a zero-mean.

Additive Discrepancy

If evaluating the computer model is fast, then inference (Bayesian or otherwise) is made rather straightforward via residuals between computer model outputs and field observations at n_F field locations X_{n_F}

$$Y_{n_F}^{b|\boldsymbol{u}} = y^b(X_{n_F}, \boldsymbol{u}) = Y_{n_F} - Y_{n_F}^{M|\boldsymbol{u}} = Y_{n_F} - y^M(X_{n_F}, \boldsymbol{u})$$

which can be computed at will for any $\boldsymbol{u}$ (Higdon et al. 2004). An "r" superscript may have been more appropriate for residuals. Besides avoiding clash with "R" for "real", superscript "b" was chosen instead to emphasize the role of residuals in training $b(\cdot)$. Eq. (is characterizing a new n_F-dimensional response vector $Y^b|\boldsymbol{u}_{n_F}$ at inputs X_{n_F}.

Likelihood with Additive Discrepancy

With a GP prior for $b(\cdot)$, $Y_{F_n}^{b|\boldsymbol{u}}$ is n_F-variate MVN with covariance derived through inverse exponentiated squared Euclidean distances between rows of X_{n_F}. This implies a likelihood on parameters $(\boldsymbol{u}, \theta_b)$, where θ_b may collect scale,m_x lengthscales and nugget hyperparameters. Let $\Sigma_{n_F}^b$ denote the $n_F \times n_F$ covariance matrix built from X_{n_F} and θ_b. Note that by including both scale and nugget in θ_b, $\Sigma_{n_F}^b$ captures field data variance σ_ϵ^2 implicitly through their product. The likelihood is thus proportional to

$$|\Sigma_{n_F}^b|^{-1/2} exp\{-\frac{1}{2}(Y_{F_n}^{b|\boldsymbol{u}})^T(\Sigma_{n_F}^b)^{-1}Y_{F_n}^{b|\boldsymbol{u}}\}$$

That likelihood can be maximized over all unknown coordinates, or fully Bayesian inference may be used to sample from the joint posterior.

Making Inference

- If evaluating the computer model is expensive or otherwise indirectly available, a surrogate $\hat{y}^M(\cdot,\cdot)$ can be fit to n_M simulations of M run over a design $[XnM; UnM]$ in $(\boldsymbol{x}, \boldsymbol{u})$-space.
- KOH recommend a GP prior for y^M, i.e., a coupled pair of GPs including $b(\cdot)$.
- Rather than performing inference for y^M separately, using just n_M runs as typical of computer experiments in isolation, KOH recommend joint posterior inference for all unknowns $\Theta = (y^M, b(\cdot), \boldsymbol{u}^*, \sigma^2_\epsilon)$ using the full corpus of data from computer model and field experiment $[Y_{n_M}, Y_{n_F}]$.
- From a Bayesian perspective, this is the coherent thing to do: infer all unknowns jointly given all data.

Likelihood evaluation

The likelihood involves evaluating a mean-zero MVN density, but this time it's $n_M + n_F$ variate for stacked computer model and field data.

$$\begin{pmatrix} Y_{n_M} \\ Y_{n_F} \end{pmatrix} \sim \mathcal{N}_{n_M+n_F}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{n_M} & \Sigma_{n_M}(X_{n_F}, \boldsymbol{u}) \\ \Sigma_{n_M}(X_{n_F}, \boldsymbol{u})^T & \Sigma_{n_F}(\boldsymbol{u}) + \Sigma^b_{n_F} \end{pmatrix}\right)$$

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

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

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

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

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J., and Ryne, J. (2004), "Combining field data and computer simulations for calibration and prediction," *SIAM Journal on Scientific Computing*, 26, 448–466.

Kennedy, M. and O'Hagan, A. (2001), "Bayesian calibration of computer models (with discussion)," *Journal of the Royal Statistical Society (Series B)*, 68, 425–464.

Mueller, P. (1993), "Alternatives to the Gibbs sampling scheme," Tech. rep., Institute Statistics and Decision Sciences, Duke University.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York, NY : Springer-Verlag, 1st ed.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging. 2nd edition*, New York: Springer.

Williams, B., Higdon, D., Gattiker, J., Moore, L., McKay, M., and Keller-McNult, S. (2006), "Combining Experimental Data and Computer Simulations, With an Application to Flyer Plate Experiments," *Bayesian Analysis*, 1, 765–792.