A Gaussian Process Emulator Based Approach for Bayesian Calibration of a Functional Input

Zhaohui Li

Georgia Institute of Technology

March 23, 2022

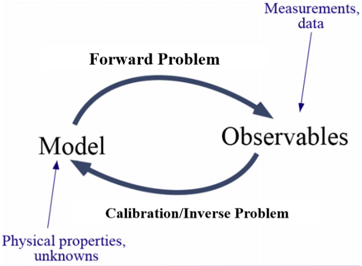
Joint work with Matthias Tan (City University of Hong Kong).

Introduction

- In this talk, we will introduce a new method for calibration of functional input with a time consuming simulator.
- Problem in consideration: Estimating the functional input of a Partial Differential Equation (PDE) from the observation data of physical experiment.

Model Calibration/Inverse Problem

- Model calibration/inverse problem: Estimating from a set of observations the parameters of model that produced them;
- Starts with the effects and then calculates the causes;
- It is the inverse of a forward problem, which starts with the causes and then calculates the effects.



Motivation Example

- Consider the groundwater equation (a PDE with initial and boundary conditions):

$$\begin{aligned} \nabla \cdot T(\boldsymbol{x})\nabla u(\boldsymbol{x},t) &= Su_t(\boldsymbol{x},t) + h(\boldsymbol{x},t) && \boldsymbol{x} \in \Omega, 0 \le t \le t_e \\ T(\boldsymbol{x})\nabla u(\boldsymbol{x},t)\cdot \boldsymbol{n} &= g_1(\boldsymbol{x},t) && \boldsymbol{x} \in \Gamma_1, 0 \le t \le t_e \\ u(\boldsymbol{x},t) &= g_2(\boldsymbol{x},t) && \boldsymbol{x} \in \Gamma_2, 0 \le t \le t_e \\ u(\boldsymbol{x},0) &= g_0(\boldsymbol{x}) && \boldsymbol{x} \in \Gamma_1 \end{aligned}$$

- The main issue is to estimate the **transmissivity** $T(\boldsymbol{x})$ as a function of spatial coordinates $\boldsymbol{x}$.
- The observation is water heads on selected observation wells at specific time:

$$y^p(\boldsymbol{s}_i) = u(\boldsymbol{s}_i) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_e^2), \boldsymbol{s}_i = (\boldsymbol{x}_i, t_i). i = 1, \cdots, n,$$

  Let $\boldsymbol{y}^p = (y^p(\boldsymbol{s}_1, \cdots, y^p(\boldsymbol{s}_n)))^T$.

Literature Review

- Tuo, R., and Wu, C. F. J. (2015) Efficient calibration for imperfect computer models. AOS
- Proposes L_2 calibration, and show its semiparametric efficiency.
- Plumlee, M., et.al (2016). Calibrating Functional Parameters in the Ion Channel Models of Cardiac Cells. JASA.
- Develops a new modeling strategy to functional parameter using GP prior and proposes new sampling scheme.
- Sraj, I., et.al (2016). "Coordinate transformation and Polynomial Chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function," CMAM.
- Tagade, P. M. and Choi, H. L. (2014). "A generalized polynomial chaos-based method for efficient Bayesian calibration of uncertain computational models," IPSE.
- Propose to use the GP prior with random correlation parameters in functional input calibration. However, their dimensional reduction methods are far more cumbersome than our method.
- Sinsbeck, M., and Nowak, W. (2017). "Sequential design of computer experiments for the solution of Bayesian inverse problems," SIAM-UQ.
- Damblin, G., et.al(2018). "Adaptive numerical designs for the calibration of computer codes," SIAM-UQ.
- Propose new sequential deign criteria for model calibration with time consuming simulator. The calibration parameter is bounded scalar/vector parameters.

Bayesian Calibration Framework

- Prior for the spatial field $T(\cdot)$.
  - Log transformation to positive parameter $f(\cdot) = \ln(T(\cdot))$ is common.
  - Prior $\pi(f(\cdot))$, usually assumed a Gaussian process (GP).
- Likelihood $p(\boldsymbol{y}^p|\boldsymbol{y}^s(f(\cdot)))$ depends on the distribution of ϵ_i, i.e.,

$$p(\boldsymbol{y}^p|\boldsymbol{y}^s(f(\cdot))) \propto \exp(-\frac{1}{2\sigma_e^2}||\boldsymbol{y}^p - \boldsymbol{y}^s(f(\cdot))||^2),$$

where $\boldsymbol{y}^s(f(\cdot)) = (y^s(\boldsymbol{s}_1, f(\cdot)), \cdots, y^s(\boldsymbol{s}_n, f(\cdot)))^T$, $y^s(\boldsymbol{s}_i, f(\cdot))$ is simulator output (PDE solution) at space-time coordinate $\boldsymbol{s}_i$ under functional input $f(\cdot)$.

- When the simulator $y^s(\boldsymbol{s}, f(\cdot))$ is time-consuming to run, it is common to employ an emulator to replace $\boldsymbol{y}^s(f(\cdot))$ (Introduce later).
- The posterior:

$$\pi(f(\cdot)|\boldsymbol{y}^p) \propto p(\boldsymbol{y}^p|\boldsymbol{y}^s(f(\cdot))) \cdot \pi(f(\cdot)).$$

Our Main Contributions

- For functional input:
  - We employ a **GP prior with a Gaussian correlation function and a uniform prior for its correlation parameter** to model the standardized functional input. We propose the use of a truncated KL expansion of the prior process for dimension reduction.
- For time-consuming simulator:
  - When employing GP emulator to approximate time-consuming simulator, we propose a **sequential design criterion** that select follow-up design points from high posterior regions with large prediction variance. The criterion can be **efficiently computed when the input and output are high dimensional**. Moreover, we prove the **convergence of the emulator-based likelihood when design is selected using WPV criterion**.

Prior for Functional Input

- GP prior for functional input:

$$f(\cdot)|\sigma_f^2, \boldsymbol{\eta} \sim \mathrm{GP}(0, \sigma_f^2 T(\cdot,\cdot|\boldsymbol{\eta})),$$

  - Use Gaussian correlation function

$$R(\mathbf{x}, \mathbf{x}'|\boldsymbol{\eta}) = \prod_{i=1}^{d} \eta_i^{(x_i - x_i')^2},$$

    $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_d) \in H \subset (0,1)^d$ is defined in a compact set.
  - We assign $\boldsymbol{\eta}$ a nondegenerate prior: $\boldsymbol{\eta} \sim \mathrm{Uniform}(H)$.
- The unconditional standardized process

$$f(\cdot)/\sigma_f \sim \mathrm{SP}(0, C(\mathbf{x}, \mathbf{x}'))$$

  is a **non-Gaussian Process**, where $C(\mathbf{x}, \mathbf{x}') = E_{\boldsymbol{\eta}} R(\mathbf{x}, \mathbf{x}'|\boldsymbol{\eta})$.
- Karhunen-Loève (KL) expansion of $f(\cdot)/\sigma_f$:

$$f(\mathbf{x})/\sigma_f = \sum_{i=1}^{\infty} z_i v_i(\mathbf{x}).$$

  $v_i(\mathbf{x})$ eigenfunctions scaled by square root of eigenvalues of C.

Karhunen-Loève (KL) Expansion

- Distribution of coefficient z_i is **unknown**.
- We derive the conditional distribution of $(z_1, \cdots, z_M)^T | \boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma}(\boldsymbol{\eta}))$, where

$$\boldsymbol{\Sigma}(\boldsymbol{\eta}) = \text{cov}(z_i, z_j | \boldsymbol{\eta}) = (\lambda_i, \lambda_j)^{-1} \int v_i(\boldsymbol{x}) R(\boldsymbol{x}, \boldsymbol{x}' | \boldsymbol{\eta}) v_j(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}'.$$

- $f(\boldsymbol{x}) = \sum_{i=1}^{M} \zeta_i v_i(\boldsymbol{x})$, finite dimensional approximation, where $\zeta_i = \sigma_f z_i, i = 1, \ldots, M$. $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_M)^T$.
- Determine the truncation number M:
  - M is set as the smallest integer M^* such that $\{v_i(\boldsymbol{x}), i = 1, \ldots, M\}$ capture at least a fraction $\tau(95\%)$ of the total variance of $f(\boldsymbol{x})/\sigma_f | \boldsymbol{\eta}$ for all $\boldsymbol{\eta} \in H$.
- Functional parameter $f(\cdot)$ is parametrized by $\boldsymbol{\xi} = (\boldsymbol{\zeta}, \sigma_f^2, \boldsymbol{\eta})$.
- Prior: $\pi(f(\cdot)) = \pi(\boldsymbol{\xi}) = \pi(\boldsymbol{\zeta} | \sigma_f^2, \boldsymbol{\eta}) \pi(\sigma_f^2, \boldsymbol{\eta})$. This fact allows us to perform Bayesian computations based on the KL expansion.

The Emulator-based Posterior

- Second Problem: Simulator $\boldsymbol{y}^s(\zeta)$ (PDE solution with $f(\boldsymbol{x}) = \sum_{i=1}^{M} \zeta_i v_i(\boldsymbol{x})$) is time consuming! Needs surrogate model for the simulator.
- We employ the GP emulator for simulator $\boldsymbol{y}^s(\zeta_0)$:

$$\boldsymbol{Y}^s(\zeta_0) \sim GP(\boldsymbol{\mu}, \sigma_0^2 \boldsymbol{R}_s R_f(\zeta_0, \zeta_0')).$$

  i.e., using separable mean function and correlation function.

- Train the emulator with experimental design $\mathcal{D} = \{\zeta_{01}, \ldots, \zeta_{0q}\}$
- Prediction of GP emulator: $\boldsymbol{Y}^s(\zeta_0|\mathcal{D}) \sim GP(\boldsymbol{m}^s(\zeta_0), v(\zeta_0)\boldsymbol{R}_s)$.
- Replacing the simulator $\boldsymbol{y}^s(\zeta)$ with $\boldsymbol{Y}^s(\zeta|\mathcal{D})$ we have

$$\boldsymbol{y}^p|\zeta, \boldsymbol{y}^s(\mathcal{D}) \sim N(\boldsymbol{m}^s(\zeta_0), v(\zeta_0)\boldsymbol{R}_s + \sigma_e^2 \boldsymbol{I}_n).$$

The Emulator-based Posterior

- The emulator-based posterior for the parameters of the functional input:

$$
\begin{aligned}
&\pi(\boldsymbol{\xi}|\boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D})) \\
\propto&|\zeta)\boldsymbol{R}_s + \sigma_e^2 \boldsymbol{I}_n|^{-1/2} \exp\{-\frac{1}{2}[\boldsymbol{y}^p - \boldsymbol{m}^s(\zeta)]^T(\zeta_0)\boldsymbol{R}_s + \sigma_e^2\boldsymbol{I}_n)^{-1}[\boldsymbol{y}^p - \boldsymbol{m}^s(\zeta)]\} \\
\cdot&|\sigma_f^2\boldsymbol{\Sigma}(\boldsymbol{\eta})|^{-1/2} \exp\{-\frac{1}{2\sigma_f^2}\zeta^T\boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1}\zeta\}\pi(\boldsymbol{\eta})\pi(\sigma_f^2).
\end{aligned}
$$

- It is important to improve GP emulator quality within high posterior region.
  - Experimental design plays a crucial role to improve the GP emulator quality.
  - It is vital to place more design points at regions of values ζ_0 of ζ with high simulator based posterior density, which is unknown priori.

Sequential Experimental Design

- In order to produce good prediction around the areas which has high posterior densities. We propose to use sequential design strategy.
  - Initial design
    - Generate a large candidate set from prior distribution. Choose a subset from the candidate set to achieve better space filling properties (Tan 2013). This gives the initial design $\mathcal{D}_0 = \{\zeta_{01}, \ldots, \zeta_{0q_0}\}$.
  - Follow-up design is achieved by selecting design points through optimizing specific design criterion.
    - The follow-up design criterion aim at produce the accurate emulator-based posterior in the high simulator based posterior region.
- **Main difficulties of follow-up design construction**
  - **Functional calibration parameter lies in high dimensional space. Existing criteria are difficult to compute for problems with high dimensional input or high dimensional output, which makes optimization of the criteria to determine follow-up design points difficult.**
- We propose a new criterion:
  - It can be computed efficiently when the input and output dimensions are large.
  - We prove the convergence of the emulator-based likelihood when design is selected using proposed criterion.

Tan M H Y. Minimax designs for finite design regions[J]. Technometrics, 2013, 55(3): 346-358.

Weighted Prediction Variance (WPV)

- Define expected prediction variance (EPV):

$$\int_{\mathbb{R}^M} v(\zeta)\pi(\zeta|\boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D}))d\zeta$$

- It is reasonable to construct design that minimizes the EPV.
- A greedy approach to minimize EPV is to select the next design point $\zeta_{0,q+1}$ the maximizer of the integrand (WPV) in expression of EPV, i.e.,

$$\zeta_{0,q+1} = \text{argmax}_{\zeta_0 \in \mathcal{Z}}[v(\zeta)\pi(\zeta_0|\boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D}))]$$

where design region $\mathcal{Z} = [L, U]^M$ is a hypercube, $L = -U$ is a large number.

Weighted Prediction Variance (WPV)

- This is called WPV criterion.
  - It tends to select high posterior density points with large prediction variance as follow-up design points.
  - Is an intuitively appealing criterion for reducing the difference between the emulator based posterior density and the simulator based posterior density.
  - Is efficient to compute. It has computation cost similar to one evaluation of the unnormalized version of $\pi(\zeta|\boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D}))$.

Theorem

Under mild conditions for design region and simulator. If follow-up design points are chosen sequentially (one-at-a-time) using the WPV criterion, then the emulator-based likelihood $(\boldsymbol{y}^p|\zeta, \boldsymbol{y}^s(\mathcal{D}))$ *converges to the simulator-based likelihood.*

MCMC Algorithm

- Metropolis within Gibbs algorithm: updates $\sigma_f^2, \eta_1, \ldots, \eta_d$ and ζ cyclically:
  1. $\pi(\sigma_f^2 | \boldsymbol{\eta}, \zeta, \boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D}))$: inverse Gamma distribution;
  2. $\pi(\eta_i | \sigma_f^2, \boldsymbol{\eta}_{-i}, \zeta, \boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D})), i = 1, \ldots, d$: slice sampling;
  3. $\pi(\zeta | \sigma_f^2, \boldsymbol{\eta}, \boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D}))$: Metropolis step;
- Proposal for Metropolis step is crucial for good mixing of the chain and full exploration of the posterior.
- Constructing proposal using the approximation of posterior.
  - The linear approximation of emulator near current state ζ':

    $$\boldsymbol{m}^s(\zeta) \approx \boldsymbol{m}^s(\zeta) + \nabla \boldsymbol{m}^s(\zeta')(\zeta' - \zeta).$$

    Replace $\boldsymbol{m}^s(\zeta)$ with its linear approximation and $v(\zeta)$ with $v(\zeta')$ in the posterior. This gives an approximation for full conditional distribution for ζ that is a Gaussian distribution $\pi(\zeta | \sigma_f^2, \boldsymbol{\eta}, \boldsymbol{y}^p, \boldsymbol{y}^s(\mathcal{D})) \approx N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\zeta'})$
  - Choose the random walk proposal $g(\zeta | \zeta', \sigma_f^2, \boldsymbol{\eta}) = N(\zeta | \zeta', 2.4^2 M^{-1} \hat{\boldsymbol{\Sigma}}_{\zeta'})$.
- New proposal gives good acceptance probability.

Calibration of Functional Parameter

1. Determine the prior model for functional input and apply dimension reduction for functional calibration parameter.
2. Construct Initial experiment design and fit GP emulator.
3. Select Follow-up experiment design points using WPV criterion. Update the GP emulator.
4. Posterior inference using MCMC algorithm.

Numerical Study

- Recall the motivation example

$$
\begin{aligned}
\nabla \cdot T(\boldsymbol{x}) \nabla u(\boldsymbol{x}, t) &= S u_t(\boldsymbol{x}, t) + h(\boldsymbol{x}, t) & \boldsymbol{x} \in \Omega, 0 \le t \le t_e \\
T(\boldsymbol{x}) \nabla u(\boldsymbol{x}, t) \cdot \boldsymbol{n} &= g_1(\boldsymbol{x}, t) & \boldsymbol{x} \in \Gamma_1, 0 \le t \le t_e \\
u(\boldsymbol{x}, t) &= g_2(\boldsymbol{x}, t) & \boldsymbol{x} \in \Gamma_2, 0 \le t \le t_e \\
u(\boldsymbol{x}, 0) &= g_0(\boldsymbol{x}) & \boldsymbol{x} \in \Gamma_1
\end{aligned}
$$

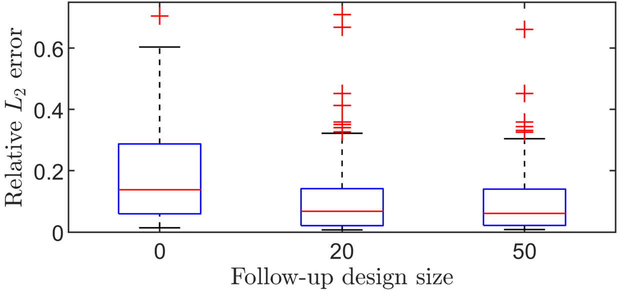
where $\Omega = [0, 1]^2, \Gamma_1 = [0, 1] \times \{0, 1\}, \Gamma_2 = \{0, 1\} \times [0, 1]$, $S = 1.234 \times 10^{-4}$, $h(\boldsymbol{x}, t)$: the source term is determined by pumping test.

- Prior assumption: $\log(T(\cdot)) = -6 + \log(10^{-6}) + f(\cdot), f(\cdot)|\boldsymbol{\eta}, \sigma_f^2 \sim GP(0, \sigma_f^2 R(\cdot, \cdot|\boldsymbol{\eta}))$, $\boldsymbol{\eta} \in H = [0.01, 0.99]^2, \pi(\boldsymbol{\eta}) = \text{Uniform}(H), \pi(\sigma_f^2) = \text{InvGamma}(3, 3)$.

Numerical Study

- Apply MCMC algorithm for posterior inference:
  - True value of functional input is randomly generated from $GP(0, R(\cdot, \cdot|\boldsymbol{\eta}_0)), \boldsymbol{\eta}_0 = (0.99, 0.01)$;
  - 90 initial design points and 20 follow-up design points generated using WPV criterion.
- Comparison of prior models for functional input:
  - Unif-GP (Proposed prior): uniform prior for the correlation parameter of the GP;
  - L-GP: fixed correlation parameter at lower bound GP;
  - M-GP: fixed correlation parameter at midpoint GP;
  - U-GP: fixed correlation parameter at upper bound GP.

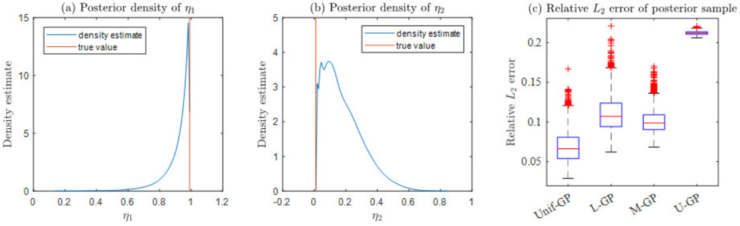
Numerical Study



(a), (b). When Unif-GP prior is used, posterior density of η_1 (True value 0.99) and η_2 (True value 0.01) concentrate around their true values.

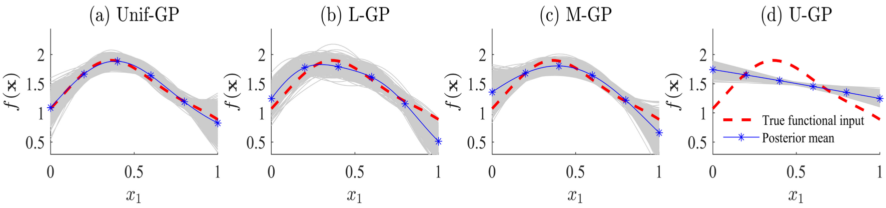
(c). Boxplots of the relative L_2 errors of posterior samples for the functional input when different priors are used. Proposed Unif-GP prior outperforms other fixed correlation parameter GP priors.

Numerical Study



- Plot of posterior mean of functional input (line marked with asterisk) and true functional input (dotted line) versus x_1 when $x_2 = x_1$.
- Light lines, which together produce shaded areas when many are close together, plot the posterior samples of the functional input.
- Unif-GP prior gives most accurate point estimate of the functional input.

Numerical Study



Boxplots of relative L_2 errors of posterior modes obtained with follow-up design of sizes 0, 20, 50 generated by WPV criterion.

The large relative L2 errors obtained when there is no follow-up run ($q_1 = 0$) indicate the importance of follow-up runs.

The lack of difference between errors given by $q_1 = 20$ and $q_1 = 50$ follow-up runs suggests that the convergence result stated in Theorem is nearly achieved with $q_1 = 20$ follow-up runs.

Summary

- We generate a new method for Bayesian calibration of a functional input to a time-consuming simulator based on a PDE model.
- Main contributions:
  - Employing a GP prior with a Gaussian correlation function and a uniform prior for its correlation parameter to model the standardized functional input. We propose the use of a truncated KL expansion of the prior process for dimension reduction.
  - Proposing WPV sequential design criterion that select follow-up design points from high posterior regions with large prediction. It can be efficiently computed when the input and output are high dimensional. Moreover, we prove the convergence of the emulator-based likelihood when design is selected using WPV criterion.

Q & A

Thank You!

zhaohui.li@gatech.edu