Statistica Sinica **25** (2015), 243-259
doi:http://dx.doi.org/10.5705/ss.2013.219w

A COMPOSITE LIKELIHOOD APPROACH TO COMPUTER MODEL CALIBRATION WITH HIGH-DIMENSIONAL SPATIAL DATA

Won Chang[1], Murali Haran[2], Roman Olson[3] and Klaus Keller[2]

[1] *University of Chicago,* [2] *The Pennsylvania State University and* [3] *University of New South Wales*

Abstract: In this paper, we introduce a composite likelihood-based approach to perform computer model calibration with high-dimensional spatial data. While composite likelihood has been studied extensively in the context of spatial statistics, computer model calibration using composite likelihood poses several new challenges. We propose a computationally efficient approach for Bayesian computer model calibration using composite likelihood. We also develop a methodology based on asymptotic theory for adjusting the composite likelihood posterior distribution so that it accurately represents posterior uncertainties. We study the application of our approach in the context of calibration for a climate model.

Key words and phrases: Block composite likelihood, computer model calibration, Gaussian process emulator, high-dimensional spatial data.

1. Introduction

Complex computer models are often used to approximate real-world processes and study complex physical phenomena. A central source of uncertainty regarding computer models, and hence the behavior of the process they are approximating, stems from uncertainty about the value of model input parameters. Computer model calibration can pose nontrivial inferential challenges. In many applications computer model runs are computationally expensive. In this case, model runs are often available at only a limited number of parameter settings. A popular method to overcome this hurdle is the Gaussian process approach (cf. Sacks et al. (1989), Kennedy and O'Hagan (2001)). This method enables calibration with a limited number of model runs using probabilistic interpolation between the model runs. However, this approach faces computational challenges when applied to computer model output that are in the form of high-dimensional spatial data, which are increasingly common in modern science and engineering applications (see e.g., Higdon et al. (2009), Bhat, Haran, and Goes (2010); Bhat et al. (2012), Chang et al. (2014)).

The current approaches to overcome such limitation for high-dimensional data rely on dimension reduction or basis expansion (Bayarri et al. (2007), Bhat et al. (2012), Chang et al. (2014), Higdon et al. (2008)). Here we construct a somewhat different method based on block composite likelihood (Caragea and Smith (2006), Eidsvik et al. (2013)). To our knowledge this is the first time composite likelihood methodology has been used in this context. In particular, we adopt the idea of hybrid composite likelihood proposed by Caragea and Smith (2006) that relies on two components: dependence between block means, and dependence within each block conditioning on its block mean. This composite likelihood approach allows for a substantial reduction in the computational burden for maximum likelihood inference with high-dimensional spatial data. Also, this opens up possibilities for flexible spatial covariance structure that vary depending on each block. Moreover, since the composite likelihood from the block composite likelihood framework is a valid probability model, no further justification is necessary for its use in Bayesian inference.

The remainder of this paper is organized as follows. In Section 2 we outline the basic model calibration framework using Gaussian random fields. In Section 3 we formulate the Bayesian calibration model using block composite likelihood, discuss relevant asymptotic theory, and explain how Godambe information may be used to adjust posterior uncertainty when using composite likelihood. In Section 4 we describe an application of our method to a climate model calibration problem using two-dimensional spatial patterns of ocean temperature change and a relevant simulated example. Finally, in Section 5, we conclude with a discussion and future directions for research.

2. Calibration Using Gaussian Processes

Here we introduce our computer model calibration framework which consists of two stages: model emulation and parameter calibration (Bayarri et al. (2007), Bhat et al. (2012), Chang et al. (2014)). We first construct an 'emulator', which is a statistical model interpolating the computer model outputs as well as providing interpolation uncertainties (Sacks et al. (1989)). Using the emulator, we find the posterior density of computer model parameters while taking into account important sources of uncertainty including interpolation uncertainty, model-observation discrepancy, and observational error (Kennedy and O'Hagan (2001)).

We use the following notation henceforth. $Y(\mathbf{s}, \boldsymbol{\theta})$ is the computer model output at the spatial location $\mathbf{s} \in \mathcal{S}$ and the parameter setting $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. $\mathcal{S}$ is the spatial field that we are interested in, usually a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. $\boldsymbol{\Theta} \subset \mathbb{R}^q$ is the open set of all possible computer model parameter settings with an integer $q \geq 1$. Let $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p\} \subset \boldsymbol{\Theta}$ be a collection of p design points in the

parameter space and $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{S}$ be the set of n model grid locations. $\mathbf{Y}_i = (Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \ldots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$ is computer model output at the model grid locations at the parameter setting $\boldsymbol{\theta}_i$. The concatenated $np \times 1$ vector of all computer model outputs is $\mathbf{Y} = (\mathbf{Y}_1^T, \ldots, \mathbf{Y}_p^T)^T$. Typically $p \ll n$ since computer model runs with high-resolution are computationally expensive. Finally, we let $Z(\mathbf{s})$ be an observation at spatial location $\mathbf{s}$ and $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$ be the observational data, a spatial process observed at n locations. The notation introduced in this section is summarized in Table S1 in Supplementary Material.

Model Emulation Using Gaussian Proccesses. Following Bhat et al. (2012) and Chang et al. (2014), we construct a Gaussian process that interpolates computer model outputs as $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \Sigma(\boldsymbol{\xi}_y))$, where $\mathbf{X}$ is an $np \times b$ covariate matrix containing all the spatial locations and climate parameters used to define the covariance matrix $\Sigma(\boldsymbol{\xi}_y)$. $\boldsymbol{\beta}$ and $\boldsymbol{\xi}_y$ are the vectors of regression coefficients and covariance parameters respectively. We construct an interpolation process by finding the maximum likelihood estimate (MLE) of these parameters. This interpolation model provides the predictive distribution of a computer model run at any given location $\mathbf{s} \in \mathcal{S}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (Sacks et al. (1989)). We call this predictive process an emulator and denote it by $\eta(\mathbf{s}, \boldsymbol{\theta})$. Throughout this paper, $\boldsymbol{\beta}$ is set to $\mathbf{0}$ since the Gaussian process provides enough flexibility in modeling the output process.

Model Calibration Using Gaussian Random Processes. We model the observational data $\mathbf{Z}$ as

$$\mathbf{Z} = \boldsymbol{\eta}(\boldsymbol{\theta}^*) + \boldsymbol{\delta}, \tag{2.1}$$

where $\boldsymbol{\theta}^*$ is the true or fitted value of computer model parameter for the observational data (Bayarri et al. (2007)), $\boldsymbol{\eta}(\boldsymbol{\theta}^*) = (\eta(\mathbf{s}_1, \boldsymbol{\theta}^*), \ldots, \eta(\mathbf{s}_n, \boldsymbol{\theta}^*))^T$ is the emulator output at $\boldsymbol{\theta}^*$ on the model grid, and $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), \ldots, \delta(\mathbf{s}_n))^T$ is a term that includes both data-model discrepancy as well as observational error. The discrepancy process $\delta(\mathbf{s})$ is also modeled as a Gaussian process with spatial covariance between the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Model calibration with high-dimensional spatial data leads to computational challenges as described in the following section.

3. Calibration with High-Dimensional Spatial Data

In this section we briefly examine the challenges in model calibration using high-dimensional spatial data and the existing approaches to the problem. We then proceed to the formulation of our composite likelihood approach. We summarize the notation used in this section in Tables S2 and S3 in the Supplementary Material.

3.1. Challenges with high-dimensional spatial data

The basic challenge with the approach in Section 2 stems from the fact that the computational cost for a single likelihood evaluation is $\mathcal{O}(n^3p^3)$. For large n, evaluating the likelihood function repeatedly when using algorithms like Markov chain Monte Carlo (MCMC) can become computationally prohibitive. One can reduce the computational cost by assuming a separable covariance structure between the spatial dependence and the dependence due to computer model parameters, but the computational cost is still $\mathcal{O}(n^3)$, and hence does not scale well with n. Here we introduce an approach that relies on the block composite likelihood for spatial data (Caragea and Smith (2006), Eidsvik et al. (2013)).

3.2. Composite likelihood for model calibration

In this framework, we partition the spatial field $\mathcal{S}$ into small blocks to avoid the computational issues related to high-dimensional data. In Section 4 we describe an example of how such a partition can be constructed in practice. The block composite likelihood method substitutes the original likelihood by a composite likelihood that utilizes the spatial blocks, thereby resulting in a likelihood function that requires much less computational effort. In particular, we adopt the block composite likelihood formulation of Caragea and Smith (2006). This framework assumes conditional independence between outcomes in different blocks given the block means, and the dependence between blocks is modeled through the covariance between block means. This framework gives a valid probability model, and therefore the posterior distribution defined using the composite likelihood function based on this approach is also a valid probability model. Obtaining a valid probability model is important because we are embedding the likelihood in a Bayesian approach; having a valid probability model automatically assures us that the resulting posterior distribution is proper when all the prior distributions used are proper.

Note that we do not use spatial blocking to fully specify the probability model here. Instead, the blocking provides a way to find a computationally feasible pseudo-likelihood estimator that has good asymptotic properties under the original probability model described in Section 2. Since the composite likelihood estimator based on blocking is not the correct likelihood estimator under the original probability model, an additional adjustment step for its asymptotic variance is required and this is discussed below.

We divide the spatial area for the computer model output into M different blocks and denote the output for each block by $\mathbf{Y}_{(1)}, \ldots, \mathbf{Y}_{(M)}$. The blocks are made according to the spatial field, not the parameter space, because the number of computer model runs is usually quite limited due to the

high computational costs of running the model. However, in principle our approach can be extended to blocking in parameter space as well if the number of model runs is also large. Let n_i denote the number of computer model outcomes in the ith block. We denote the spatial locations in the ith block by $\mathbf{s}_{i1}, \ldots, \mathbf{s}_{in_i}$. Each $\mathbf{Y}_{(i)}$ is a stack of (n_i-1)-dimensional spatial output for p different parameter settings, $\mathbf{Y}_{(i)} = (Y(\mathbf{s}_{i1}, \cdot)^T, Y(\mathbf{s}_{i2}, \cdot)^T, \ldots, Y(\mathbf{s}_{in_i-1}, \cdot)^T)^T$, where $Y(\mathbf{s}_{ij}, \cdot) = (Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \ldots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T$ is the $p \times 1$ vector of computer model outcomes for all the parameter settings $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$. We omit one spatial location for each block in defining the output vectors to avoid degeneracy. The composite likelihood function changes slightly depending on which spatial location in each block is omitted. However, this difference is negligible unless the block sizes $n_1, \ldots, n_M$ are very small. We let $\bar{\mathbf{Y}}_{(i)} = \sum_{j=1}^{n_i} (Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_1), \ldots, Y(\mathbf{s}_{ij}, \boldsymbol{\theta}_p))^T / n_i$ be the p-dimensional mean vector of model outcomes for the ith block, the means for the spatial block consisting of same set of locations across all model parameter settings. We define the vector of all block means by $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_{(1)}^T, \ldots, \bar{\mathbf{Y}}_{(M)}^T)^T$. Similarly, we divide the observational data into M blocks in the same way and omit one observation for each block to have $\mathbf{Z}_{(1)}, \ldots, \mathbf{Z}_{(M)}$, the vectors of observational data in different blocks. We let $\bar{\mathbf{Z}}_{(i)} = \sum_{j=1}^{n_i} Z(\mathbf{s}_{ij})/n_i$ be the ith block mean of observational data and $\bar{\mathbf{Z}} = (\bar{\mathbf{Z}}_{(1)}, \ldots, \bar{\mathbf{Z}}_{(M)})^T$ be the collection of them.

Assuming separability, we model the covariance between the process at two different spatial locations and parameter settings $Y(\mathbf{s}, \boldsymbol{\theta})$ and $Y(\mathbf{s}', \boldsymbol{\theta}')$ by $\mathrm{Cov}\,(Y(\mathbf{s}, \boldsymbol{\theta}), Y(\mathbf{s}', \boldsymbol{\theta}')) = K_s(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_s) K_\theta(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\xi}_\theta)$, where K_s and K_θ are valid covariance functions respectively in $\mathcal{S}$ and $\boldsymbol{\Theta}$ with parameters $\boldsymbol{\xi}_s$ and $\boldsymbol{\xi}_\theta$. The covariance between discrepancy process at $\mathbf{s}$ and $\mathbf{s}'$ is given by $\mathrm{Cov}\,(\delta(\mathbf{s}), \delta(\mathbf{s}')) = K_d(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_d)$ with a valid covariance function K_d in $\mathcal{S}$ and a vector of parameters $\boldsymbol{\xi}_d$. More specific definition of the covariance functions is discussed below.

Computer Model Emulation. The first component of our composite likelihood is the model for block means that captures the large scale trend. The covariance between the block means is $\Sigma^{\bar{\mathbf{Y}}} = \mathrm{H} \otimes \Sigma_\theta$ where Σ_θ is the covariance matrix for the random variable across p parameter settings and H is the $M \times M$ covariance matrix between the blocks. It is straightforward to see that the block covariance is

$$\{\mathrm{H}\}_{ij} = \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} K_s(\mathbf{s}_{ik}, \mathbf{s}_{jl}; \boldsymbol{\xi}_s), \tag{3.1}$$

the mean of all possible cross covariances between two blocks.

The second component is the sum of the conditional likelihoods for each block, which models the small scale dependence and variation. For the ith block, the conditional distribution of output $\mathbf{Y}_{(i)}$ given the block mean $\bar{\mathbf{Y}}_{(i)}$ is a normal distribution with the mean and covariance given by $\mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}} = E(\mathbf{Y}_{(i)}|\bar{\mathbf{Y}}_{(i)}) =$

$(\gamma^{(i)}/\{\mathrm{H}\}_{ii} \otimes \mathbf{I}_p)\bar{\mathbf{Y}}_{(i)}$ and $\Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}} = Var(\mathbf{Y}_{(i)}|\bar{\mathbf{Y}}_{(i)}) = (\Gamma_i - \gamma^{(i)}\left(\gamma^{(i)}\right)^T/\{\mathrm{H}\}_{ii}) \otimes \Sigma_\theta$ where Γ_i is the $(n_i-1)\times(n_i-1)$ spatial covariance matrix for the ith block and $\gamma^{(i)}$ is the $(n_i-1)\times 1$ covariance vector between the ith block mean and the ith block locations (see Table S2 in the Supplementary Material for their definitions). The log composite likelihood function for the model output is then $c\ell(\boldsymbol{\xi}_s, \boldsymbol{\xi}_\theta) \propto -(\log|\Sigma^{\bar{\mathbf{Y}}}| + \bar{\mathbf{Y}}^T(\Sigma^{\bar{\mathbf{Y}}})^{-1}\bar{\mathbf{Y}})/2 - \sum_{i=1}^M(\log|\Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}}| + (\mathbf{Y}_{(i)} - \mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}})^T(\Sigma_i^{\mathbf{Y}|\bar{\mathbf{Y}}})^{-1}(\mathbf{Y}_{(i)} - \mu_i^{\mathbf{Y}|\bar{\mathbf{Y}}}))/2$. We construct the emulator by finding the MLE of $\boldsymbol{\xi}_\theta$ and $\boldsymbol{\xi}_s$, denoted by $\hat{\boldsymbol{\xi}}_\theta$ and $\hat{\boldsymbol{\xi}}_s$. The computational cost for a single likelihood evaluation is reduced from $n^3/3$ flops to $\sum_{i=1}^M\sum_{j=i}^M n_i n_j + M^3/3 + \sum_{i=1}^M(n_i-1)^3/3$ flops, where the first term is the computational cost for finding H. This is a reduction from 6.86×10^{10} flops to 5.92×10^7 flops in the climate model calibration example in Section 4.

Computer Model Calibration. We formulate the composite likelihood for observational data in the same manner as above. Let $\boldsymbol{\Omega}$ be the $M\times M$ covariance between the M block means of the discrepancy $\boldsymbol{\delta}$, defined in the same way as H, but with a different set of parameters $\boldsymbol{\xi}_d$ (see Table S3 in the Supplementary Material for its definition). The conditional mean and covariance for the block means of observational data $\bar{\mathbf{Z}}$ are $\mu^{\bar{\mathbf{Z}}} = \left(\mathbf{I}_M \otimes \Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\right)\bar{\mathbf{Y}}$, an $M\times 1$ vector, and $\Sigma^{\bar{\mathbf{Z}}} = \mathrm{H} \otimes \left(\Sigma_{\theta^*} - \Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T\right) + \Omega$, an $M\times M$ matrix where $\Sigma_{\theta^*\theta}$ is the $1\times p$ covariance vector between the fitted computer model parameter value $\boldsymbol{\theta}^*$ and the design points $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$. Likewise, we define Λ_i and $\lambda^{(i)}$ as the discrepancy counterparts of Γ_i and $\gamma^{(i)}$ with the covariance parameter $\boldsymbol{\xi}_d$. Hence, Λ_i and $\lambda^{(i)}$ are the $(n_i-1)\times(n_i-1)$ ith block discrepancy covariance matrix and the $(n_i-1)\times 1$ covariance vector between the block outputs and the block mean respectively (see Table S3 in the Supplementary Material for their definitions). The conditional mean and covariance for observational data in the ith block are therefore $\mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}} = (\mathbf{I}_{n_i-1} \otimes \Sigma_{\theta^*\theta}\Sigma_\theta^{-1})\mathbf{Y}_{(i)} + (\tau^{(i)} + \lambda^{(i)})\{\Sigma^{\bar{\mathbf{Z}}}\}_{ii}^{-1}(\bar{\mathbf{Z}}_i - \{\mu^{\bar{\mathbf{Z}}}\}_i)$ and $\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} = (\Gamma_i \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T) + \Lambda_i) - (\tau^{(i)} + \lambda^{(i)})(\tau^{(i)} + \lambda^{(i)})^T/\{\Sigma_{\bar{\mathbf{Z}}}\}_{ii}$, where $\tau^{(i)} = \gamma^{(i)} \otimes (\Sigma_{\theta^*} - \Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T)$. The log composite likelihood for the observational data is then

$$
\begin{aligned}
c\ell_n(\boldsymbol{\psi}) \propto & -\frac{1}{2}\left(\log\left|\Sigma^{\bar{\mathbf{Z}}}\right| + \left(\bar{\mathbf{Z}} - \mu^{\bar{\mathbf{Z}}}\right)^T\left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1}\left(\bar{\mathbf{Z}} - \mu^{\bar{\mathbf{Z}}}\right)\right) \\
& -\frac{1}{2}\sum_{i=1}^M\left(\log\left|\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right| + \left(\mathbf{Z}_{(i)} - \mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^T\left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1}\left(\mathbf{Z}_{(i)} - \mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)\right), \quad (3.2)
\end{aligned}
$$

where the first line in (3.2) is the log likelihood corresponding to the block means and the second line corresponding to the observations within each block. $\boldsymbol{\psi}$ denotes all the parameters being estimated in the calibration stage including $\boldsymbol{\theta}^*$ and $\boldsymbol{\xi}_d$.

By choosing a proper prior for $\boldsymbol{\psi}$, $f(\boldsymbol{\psi})$, we define the approximate log posterior density, $\log(\pi_n(\boldsymbol{\psi})) \propto \log f(\boldsymbol{\psi}) + c\ell_n(\boldsymbol{\psi})$ and infer $\boldsymbol{\psi}$ using the standard Metropolis-Hastings algorithm. We allow the scale parameters for the emulator to be re-estimated along with the other parameters but fix the other emulator parameters in $\boldsymbol{\xi}_s$ and $\boldsymbol{\xi}_\theta$ at their estimated values from the emulation stage (Bayarri et al. (2007), Bhat et al. (2012), Chang et al. (2014)). The formulation results in the same computational gain as in the emulation stage.

In both the emulation and calibration stages, calculation of the covariance matrix for the block means is a computational bottleneck, requiring $\sum_{i=1}^{M}\sum_{j=i}^{M} n_i n_j$ flops of computation. While computationally very demanding, its contribution to the likelihood function is usually not significant (Caragea and Smith (2006)). Therefore, instead of using all cross covariances between spatial locations, we randomly sample a subset of cross covariances to approximate the covariance between block means H. The computation of H in (3.1) is substituted by

$$\{\mathrm{H}\}_{ij} = \frac{1}{m_i m_j}\sum_{k=1}^{m_i}\sum_{l=1}^{m_j} K_s(\mathbf{u}_{ik}, \mathbf{u}_{jl}; \boldsymbol{\xi}_s), \tag{3.3}$$

with $m_i \le n_i$ and $m_j \le n_j$, where $\mathbf{u}_{i1}, \ldots, \mathbf{u}_{im_i}$ and $\mathbf{u}_{j1}, \ldots, \mathbf{u}_{jm_j}$ are randomly chosen respectively from $\mathbf{s}_{i1}, \ldots, \mathbf{s}_{in_i}$ and $\mathbf{s}_{j1}, \ldots, \mathbf{s}_{jn_j}$. This reduces the computational cost from $\sum_{i=1}^{M}\sum_{j=i}^{M} n_i n_j$ to $\sum_{i=1}^{M}\sum_{j=i}^{M} m_i m_j$, that is, 1.32×10^7 flops to 2.86×10^5 flops for the calibration problem in Section 4.2. The same approximation can be applied to Ω with $\boldsymbol{\xi}_d$.

Covariance Function and Prior Specification. We use the exponential covariance function with a nugget term to define the covariance between parameter settings (K_θ), spatial covariance for the emulator (K_s), and the spatial covariance for the discrepancy (K_d). Thus the covariance between the process at two parameter settings $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$ and $\boldsymbol{\theta'} = (\theta'_1, \ldots, \theta'_q)^T$ is $K_\theta(\boldsymbol{\theta}, \boldsymbol{\theta'}; \boldsymbol{\xi}_\theta) = \zeta_\theta 1(\boldsymbol{\theta} = \boldsymbol{\theta'}) + \kappa_\theta \exp(-\sum_{i=1}^{q}\phi_{\theta,i}|\theta_i - \theta'_i|)$, where $\boldsymbol{\xi}_\theta = (\zeta_\theta, \kappa_\theta, \phi_{\theta,1}, \ldots, \phi_{\theta,q})$, and $\zeta_\theta, \kappa_\theta, \phi_{\theta,1}, \ldots, \phi_{\theta,q} > 0$. Likewise, the covariance between the process at two spatial locations $\mathbf{s}$ and $\mathbf{s}'$ for the emulator and the discrepancy term are

$$\begin{aligned} K_s(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_s) &= \kappa_s\left(\zeta_s 1(\mathbf{s} = \mathbf{s}') + \exp\left(-\phi_s g(\mathbf{s}, \mathbf{s}')\right)\right), \\ K_d(\mathbf{s}, \mathbf{s}'; \boldsymbol{\xi}_d) &= \kappa_d\left(\zeta_d 1(\mathbf{s} = \mathbf{s}') + \exp\left(-\phi_d g(\mathbf{s}, \mathbf{s}')\right)\right), \end{aligned} \tag{3.4}$$

respectively, with $\boldsymbol{\xi}_s = (\zeta_s, \kappa_s, \phi_s)$, $\boldsymbol{\xi}_d = (\zeta_d, \kappa_d, \phi_d)$, and $\zeta_s, \kappa_s, \phi_s, \zeta_d, \kappa_d, \phi_d > 0$. $g(\mathbf{s}, \mathbf{s}')$ denotes the distance between two points. In the climate model calibration problem in Section 4, for example, g is the geodesic distance between two points on the earth's surface.

The parameters inferred by the Bayesian approach in the calibration stage are κ_s, ζ_d, κ_d, ϕ_d, and $\boldsymbol{\theta}^*$. Following Bayarri et al. (2007), the sill parameter for

the emulator κ_s is initially inferred via maximum likelihood estimate in the emulation stage and re-estimated by Bayesian inference in the calibration stage. We impose informative priors on the above parameters to avoid potentially obtaining improper posterior distributions (cf., Berger, De Oliveira, and Sansó (2001)) and identifiability issues. The latter is explained further in Section 4. The sill parameters, κ_s and κ_d receive inverse-Gamma priors $IG(a_{\kappa_s}, b_{\kappa_s})$ and $IG(a_{\kappa_d}, b_{\kappa_d})$. We also impose an Inverse-Gamma prior $IG(a_{\zeta_d}, b_{\zeta_d})$ for the nugget parameter ζ_d. The prior density for the range parameter ϕ_d is assumed to be uniform with a wide support. The fitted computer model parameter $\boldsymbol{\theta}^*$ also receives a uniform prior over a wide range. Note that one can also assume a more informative prior for $\boldsymbol{\theta}^*$ such as a unimodal distribution based on some physical knowledge. However, in the calibration problem in Section 4 we do not impose such a prior for $\boldsymbol{\theta}^*$; this allows us to study the characteristics of the posterior density of $\boldsymbol{\theta}^*$ more transparently.

Asymptotics and Adjustment using Godambe Information. The composite likelihood in (3.2) is not based on the true probability model in (2.1), and therefore the 'composite' posterior density based on (3.2) is quite different from the true posterior based on (2.1). In this section, we discuss how the Godambe information matrix (Godambe (1960)) for estimating equations may be used to adjust for using the composite likelihood when making inferences.

We first provide the asymptotic justification for the adjustment using the Godambe information matrix. We show that, for large n and p, the mode of the approximate posterior $\hat{\boldsymbol{\psi}}_n^B = \arg\max_{\boldsymbol{\psi}} \pi_n(\boldsymbol{\psi})$ is consistent and asymptotically normally distributed, with a covariance matrix given by the inverse of the Godambe information matrix. If we let $p \to \infty$, then the emulator converges to the measurement-error model such that $\boldsymbol{\eta}(\boldsymbol{\theta}) \sim N(\mathbf{Y}(\boldsymbol{\theta}), \zeta_\theta \Sigma^s)$, where $\mathbf{Y}(\boldsymbol{\theta})$ is the $n \times 1$ vector of model output at the parameter setting $\boldsymbol{\theta}$ and the spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. This result holds as long as the computer model output varies reasonably smoothly in the parameter space (Yakowitz and Szidarovszky (1985)). The model for observational data is $\mathbf{Z} \sim N(\mathbf{Y}^*, \zeta_\theta \Sigma^s + \Sigma^d)$, where $\mathbf{Y}^* = \mathbf{Y}(\boldsymbol{\theta}^*)$, $\{\Sigma^s\}_{ij} = K_s(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\xi}_s)$, and $\{\Sigma^d\}_{ij} = K_d(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\xi}_d)$. The composite likelihood in (3.2) then has means and covariances $\mu^{\bar{\mathbf{Z}}} = \bar{\mathbf{Y}}^*$, an $M \times 1$ vector, $\Sigma^{\bar{\mathbf{Z}}} = \zeta_\theta \mathrm{H} + \Omega$, an $M \times M$ matrix, $\mu_i^{\mathbf{Z}|\bar{\mathbf{Z}}} = \mathbf{Y}^*_{(i)} + (\zeta_\theta \gamma^{(i)} + \lambda^{(i)}) \{\Sigma^{\bar{\mathbf{Z}}}\}_{ii}^{-1} (\bar{\mathbf{Z}}_i - \{\mu^{\bar{\mathbf{Z}}}\}_i)$, and $\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}} = (\zeta_\theta \Gamma_i + \Lambda_i) - (\zeta_\theta \gamma^{(i)} + \lambda^{(i)})(\zeta_\theta \gamma^{(i)} + \lambda^{(i)})^T / \{\Sigma_{\bar{\mathbf{Z}}}\}_{ii}$, where $\bar{\mathbf{Y}}^*_{(i)} = \sum_{j=1}^{n_i} Y(\mathbf{s}_{ij}, \boldsymbol{\theta}^*)/n_i$ is the ith block mean of the computer model output at $\boldsymbol{\theta}^*$ and $\bar{\mathbf{Y}}^* = (\bar{\mathbf{Y}}^*_{(1)}, \ldots, \bar{\mathbf{Y}}^*_{(M)})^T$ is the collection of all their block means.

We now show the consistency and the asymptotic normality of the posterior mode $\hat{\boldsymbol{\psi}}_n^B$ as $n \to \infty$. We utilize expanding domain asymptotic results (see e.g., Mardia and Marshall (1984), Cressie (1993), Cox and Reid (2004),

Zhang and Zimmerman (2005), Varin (2008)). The first step is establishing consistency and asymptotic normality of the maximum composite likelihood estimator.

Proposition 1. *The following holds for the maximum composite likelihood estimator* $\hat{\boldsymbol{\psi}}_n^{CL} = \arg\max_{\boldsymbol{\psi}} c\ell_n(\boldsymbol{\psi})$*:* (i) (*Consistency*) $\hat{\boldsymbol{\psi}}_n^{CL} \xrightarrow{\mathcal{P}} \boldsymbol{\psi}^0$ *as* $n \to \infty$*, where* $\boldsymbol{\psi}^0$ *is the vector of true values of parameters in* $\boldsymbol{\psi}$*;* (ii) (*Asymptotic Normality*) $\mathbf{G}_n^{1/2}(\hat{\boldsymbol{\psi}}_n^{CL} - \boldsymbol{\psi}^0) \xrightarrow{\mathcal{D}} N(0, \mathbf{I})$ *where* $\mathbf{G}_n = \mathbf{Q}_n\mathbf{P}_n^{-1}\mathbf{Q}_n$ *is the Godambe information matrix (Godambe (1960)).* $\mathbf{P}_n$ *is the covariance matrix of the gradient* $\nabla c\ell_n$ *and* $\mathbf{Q}_n$ *is the negative expected value of the Hessian matrix of* $c\ell_n$*, where both are evaluated at* $\boldsymbol{\psi} = \boldsymbol{\psi}^0$*.*

The main result of this section establishes the consistency and asymptotic normality of the posterior mode, $\hat{\boldsymbol{\psi}}_n^B$.

Proposition 2. (i) (*Posterior consistency*) $|\pi_n(\boldsymbol{\psi}) - \pi_n^0(\boldsymbol{\psi})|_{TV} \xrightarrow{\mathcal{P}} 0$ *as* $n \to \infty$ *where* $|\cdot|_{TV}$ *is the total variation norm and* $\pi_n^0(\boldsymbol{\psi})$ *is a normal density with the mean* $\boldsymbol{\psi}^0 + \mathbf{Q}_n^{-1}\nabla c\ell_n(\boldsymbol{\psi}^0)$ *and the covariance* $\mathbf{Q}_n^{-1}$ *with* $\mathbf{Q}_n^{-1} \to \mathbf{0}$ *as* $n \to \infty$*;* (ii) (*Asymptotic normality*) *as* $n \to \infty$*,* $\mathbf{G}_n^{1/2}(\hat{\boldsymbol{\psi}}_n^B - \boldsymbol{\psi}^0) \xrightarrow{\mathcal{D}} N(0, \mathbf{I})$

The proofs of the propositions are given in Section S1 in the Supplementary Material.

Application of Gobambe Adjustment. We have several options for adjusting our composite likelihood-based inference. These include (a) direct use of the asymptotic distribution in Proposition 2. (ii); (b) The 'open-faced sandwich' post-hoc adjustment (Shaby (2013)) of MCMC sample from the composite posterior distribution $\pi_n(\boldsymbol{\psi})$; (c) 'curvature' adjustment (Ribatet, Cooley, and Davison (2012)) for our MCMC procedure. We utilize (b) and (c) because these MCMC-based methods can capture the higher-order moments of the posterior distribution, which may be important in finite sample inference.

For any of these methods, it is necessary to evaluate $\mathbf{P}_n$ and $\mathbf{Q}_n$. See Section S2 in the Supplementary Material for an example of their analytic computation. Here $\mathbf{Q}_n$ can also be obtained using MCMC runs from the posterior distribution $\pi_n(\boldsymbol{\psi})$ by the asymptotic result in Proposition 2. (i). The matrix operation for computing $\mathbf{Q}_n$ and $\mathbf{P}_n$ is not computationally demanding as it can be done in less than a few minutes using a high-performance single core and needs to be done only once.

We caution that the adjustment procedures here rely on the identifiability of parameters in $\boldsymbol{\psi}$. In order to evaluate $\mathbf{P}_n$ and $\mathbf{Q}_n$ under the correct probability model $\mathbf{Z} \sim N(\mathbf{Y}^*, \zeta_\theta\Sigma^s + \Sigma^d)$, we need to be able to estimate the true value $\boldsymbol{\psi}^0$

accurately by the posterior mode $\hat{\boldsymbol{\psi}}_n^B$. This may not always hold as there is a trade-off between the discrepancy parameters in $\boldsymbol{\xi}_d$ for finite sample sizes.

The open-faced sandwich adjustment is one approach for adjusting the covariance based on Proposition 2 (Shaby (2013)). For any MCMC sample of $\boldsymbol{\psi}$ from $\pi_n(\boldsymbol{\psi})$, the open-faced sandwich adjustment is defined by $\tilde{\boldsymbol{\psi}}^{open} = \hat{\boldsymbol{\psi}}_n^B + \mathbf{C}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n^B)$ with $\mathbf{C} = \mathbf{Q}_n^{-1}\mathbf{P}_n^{1/2}\mathbf{Q}_n^{1/2}$. Similar to the curvature adjustment, this approach guarantees that the distribution of the adjusted posterior sample has the same posterior mode and the desired asymptotic covariance $\mathbf{G}_n^{-1}$. This method can be either embedded in each step of MCMC run, or applied after an entire MCMC run is finished. Another possible approach is curvature adjustment (Ribatet, Cooley, and Davison (2012)), which likely leads to a similar result as shown in Shaby (2013). We provide the details of this approach in Section S3 in the Supplementary Material. It is also possible to infer the parameters using the Godambe information matrix. However, Bayesian inference based on the adjustment allows us to incorporate prior information and learn about the shape of the posterior density which might be important for a finite sample. In computer model calibration it is important to use prior information for identifying the discrepancy function; a Bayesian approach is therefore a convenient framework for doing this.

4. Application to UVic ESCM Calibration

We demonstrate the application of our approach to a climate model calibration problem. The computer model used here is the University of Victoria Earth system climate model (UVic ESCM) of intermediate complexity (Weaver et al. (2001)). The input parameter of interest is climate sensitivity (CS), defined as the equilibrium global mean surface air temperature change due to a doubling of carbon dioxide concentrations in the atmosphere (see e.g. Knutti and Hegerl (2008)). We use ten UVic ESCM runs with different CS values ($p = 10$). Each model run is a spatial pattern of ocean temperature anomaly on a regular 1.8° latitude by 3.6° longitude grid, defined as change between 1955-1964 mean and 2000-2009 mean in degree Celsius times meter (°C m). At each location, the ocean temperature anomaly is vertically integrated from 0 to 2,000 m in depth. See Sriver et al. (2012) for a more detailed description of the model runs.

The model output has regions of missing data since it covers only the ocean, and partition of the spatial area needs careful consideration. We partitioned the spatial area using a random tessellation; this is also the approach followed by Eidsvik et al. (2013). We first randomly chose M different centroids out of total n locations and then assigned the spatial locations to different subregions according to the nearest centroid in terms of geodesic distance. When finding the nearest centroid for each point, we only considered the centroids in the same

ocean to avoid assigning locations separated by land to the same block. This random tessellation ensured, on average, that we have more subregions where data points are densely distributed.

Choosing the number of blocks requires balancing two conflicting objectives, using enough number of blocks to make the computation feasible, and ensuring enough sample size for each block for asymptotic convergence. We provide a guideline in Section S4 in the Supplementary Material.

4.1. Simulated examples

We conducted some perfect model experiments to answer the questions of whether the the posterior density based on the composite likelihood (composite posterior) is similar to the posterior density based on the original likelihood (original posterior), of whether the posterior density with approximated block mean covariance computation (approximated composite posterior) described in (3.3) is close to the true composite posterior, and of how the number of spatial blocks and the magnitude of the discrepancy affect the composite posterior density.

Each experiment followed four key steps: choose one of the parameter settings for model runs as the synthetic truth; leave the corresponding model run out and superimpose a randomly generated error on it to construct a synthetic observation; emulate the computer model using the remaining model runs; calibrate the computer model using the emulator and compare the resulting density with the synthetic truth. To be able to compute the original posterior density with a reasonable computational effort, we restricted ourselves to a subset of spatial locations consisting of 1,000 randomly selected points ($n = 1,000$) and assumed separable covariance structure for the spatial field and the computer model parameter space. The synthetic truth for the climate sensitivity used here was 2.153, but choosing other parameter settings gave similar results shown here.

A comparison between the composite posterior densities with 10 blocks and the original posterior densities are shown in Figure 3(a) and 3(b). We used two different realizations of the model-observation discrepancy. These were generated from a Gaussian process model with exponential covariance (3.4) with $\zeta_d^* = 0.01$, κ_d^* =160,000, and $\phi_d^* = 690^{-1}$ km, where $(\zeta_d^*, \kappa_d^*, \phi_d^*)$ are assumed true values of $(\zeta_d, \kappa_d, \phi_d)$. We also conducted the same comparison for the approximated composite posterior densities (Figure 3(c) and 3(d)). The posterior densities and the resulting credible intervals from all three approaches are reasonably similar. The composite posterior densities after adjustment are slightly more dispersed than the original posterior due to the information loss caused by blocking, but the modes are quite close to the original ones confirming the consistency result in Proposition 2 (i).

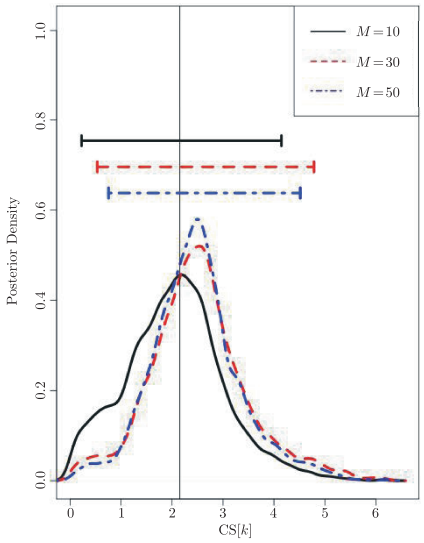


Figure 1. Comparison of posterior densities between three simulated examples with different block numbers: $M = 10$ (solid curve), $M = 30$ (dashed curve), and $M = 50$ (dotted-dashed curve). The vertical line is the assumed true value for our simulated example and the horizontal bars above are 95% credible intervals. Posterior modes based on 30 and 50 blocks show slight biases, but the width of intervals does not show notable differences.

.

We also compared the adjusted composite posterior densities with different numbers of blocks to examine the effect of the number of blocks on calibration results (Figure 1). The results show that using more than 30 blocks introduces a slight bias for the posterior mode that might be due to the reduced number of data points in each block or the smaller area covered by each block. When using more than 30 blocks, there are several blocks for which the maximum distance is less than the effective range. However, the credible intervals are again reasonably similar to each other and our approach still gives a reasonable answer in this setting. Similarly, we compare the adjusted composite posterior densities based on datasets generated using different assumed sill values, κ_d^* =40,000, 90,000, and 160,000 to investigate the effect of magnitude of discrepancy on calibration results (Figure S1 in the Supplementary Material). As one would expect, the posterior density becomes more dispersed as we increase the value of the sill.

We used informative priors for the statistical parameters, important in reducing the identifiability issues occurring in the calibration based on observational data in Section 4.2. We imposed a vague prior for the nugget parameter $\zeta_d \sim IG(2, 0.01(2+1))$ and a highly informative prior for the sill parameter $\kappa_d \sim IG(10{,}000, \kappa_d^*(10{,}000+1))$. The sill parameter for the emulator κ_s was

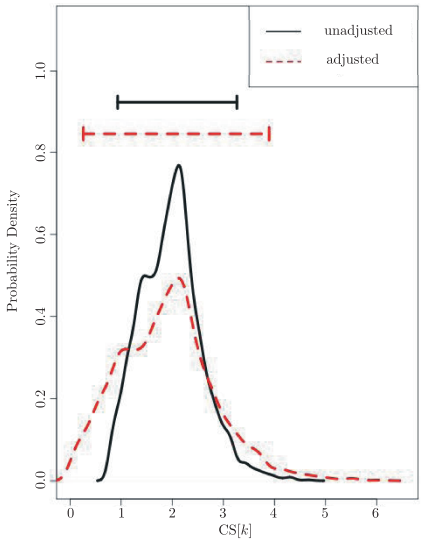


Figure 2. Posterior densities of the climate sensitivity calibrated based on the observational data from Levitus et al. (2012) using our composite likelihood approach. The adjusted posterior density (dashed curve) is notably more dispersed than the unadjusted one (solid curve), and the corresponding 95% credible intervals (horizontal bars above) for the adjusted posterior density is also much wider than the one for the unadjusted density.

given a mildly informative prior with $IG(20, \hat{\kappa}_s(20+1))$, $\hat{\kappa}_s$ the MLE of κ_s computed in the emulation stage. The shape parameters for the inverse-Gamma distributions are specified in the way that the prior modes are aligned with certain target values. Note that inference for simulated examples does not suffer from identifiability issues without the informative priors; we use these priors only to be consistent with the calibration based on observational data. For example, changing the shape parameter b_{κ_d} for the inverse-Gamma prior from 10,000 to 2 does not change the results in Figures 1−3.

4.2. Calibration using observed ocean temperature anomalies

As an illustrative example, we calibrated the climate sensitivity using the observed spatial pattern of ocean temperature anomaly from the data product constructed by Levitus et al. (2012). We interpolated the observational data onto the UVic model grid using a simple bilinear interpolator. By placing the model output and the observational data on the same grid, we take full advantage of the computational gains from the separability assumption. We divided the 5,903 locations ($n = 5,903$) into 50 blocks ($M = 50$) using the random tessellation method described above. The covariance matrices for block means were approximated using (3.3) with $m_i = \min(10, n_i)$ for $i = 1, \ldots, 50$. The prior specification

was the same as in the simulated example with assumed sill (κ_d^*) of 160,000, except that the inverse of the discrepancy range parameter ϕ_d^{-1} was restricted to be greater than 800 km to reduce identifiability issues. Figure 2 shows the posterior density of climate sensitivity. The length of the MCMC chain was 15,000, and the computing time was about 15 hours (wall time) via parallel computing using 32 high-performance cores for a system with Intel Xeon E5450 Quad-Core 434 at 3.0 GHz. We verified that our MCMC algorithm and chain length were adequate by ensuring that the MCMC standard errors for our parameter estimates (Jones et al. (2006), Flegal, Haran, and Jones (2008)) were small enough and by comparing posterior density estimates after various run lengths to see that the results, namely posterior pdfs, had stabilized.

Unlike simulated examples described above, the calibration results with and without the adjustment described in Section 3.2 were clearly different due to the reduced block sizes. This indicates that the adjustment using the Godambe information is an important step to account for the information loss caused by blocking, especially when using many smaller blocks is unavoidable due to the large amount of data. In our particular application blocking causes underestimation of the variance of the parameter $\boldsymbol{\theta}$. Underestimation of the variances of important input parameters can lead to an incorrect analysis in follow-up studies based on the calibration results. For example, underestimating the uncertainty in the important climate parameters such as the climate sensitivity can result in incorrect risk assessment in climate change impact studies.

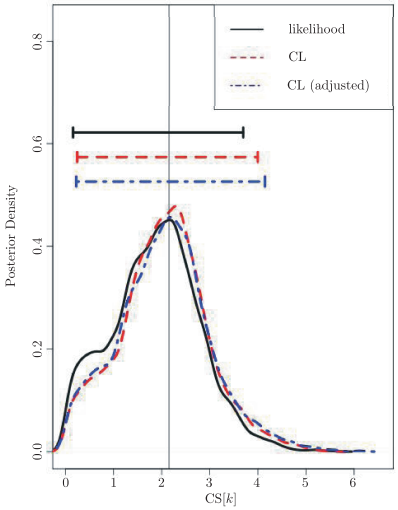
5. Discussion

This work is, to our knowledge, the first application of composite likelihood to the computer model calibration problem. Our composite likelihood approach enables computationally efficient inference in computer model calibration using high-dimensional spatial data. We proved consistency and asymptotic normality of our posterior estimates and established covariance adjustment for posterior density based on them. The adjustment can be easily integrated into common MCMC algorithms such as the Metropolis-Hastings algorithm. The block composite likelihood used here yields a valid probability model, and therefore no additional verification for the propriety of the posterior distribution is necessary.

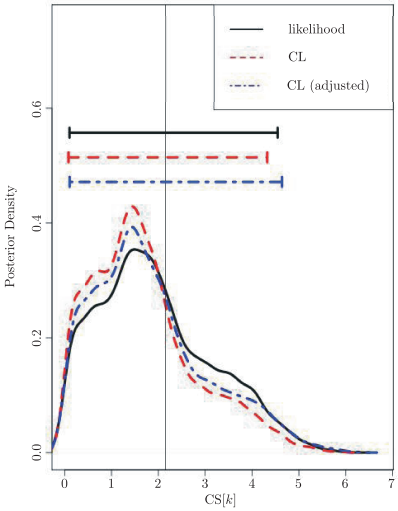
An attractive benefit of this general framework is that it is relatively easy, in principle, to extend the approach to a more complicated and easy-to-interpret covariance model. For example, by allowing covariance parameters to vary across the different spatial blocks, our approach can introduce non-stationarity in the spatial processes of model output and observational data.

The method and the scientific result are subject to a few caveats that commonly apply to most calibration problems and composite likelihood approaches. We refer to Section S5 in the Supplementary Material for a detailed discussion.

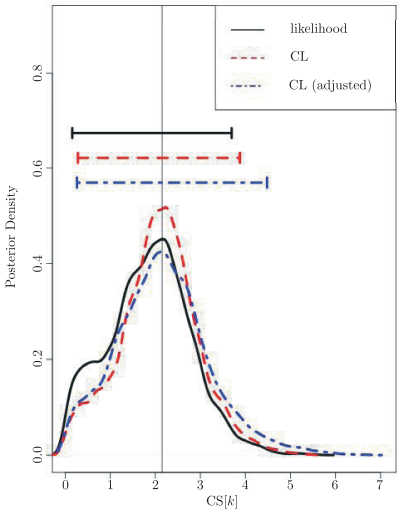
(a) Realization 1 without approximation



(b) Realization 2 without approximation



(c) Realization 1 with approximation



(d) Realization 2 with approximation

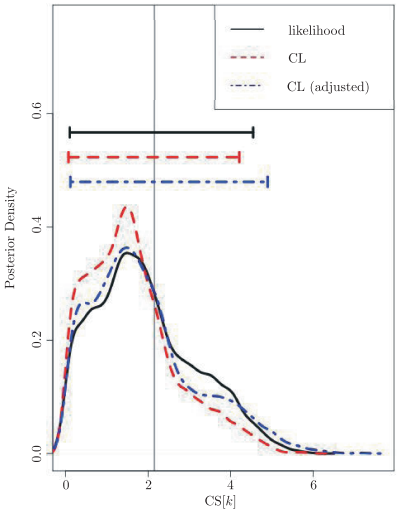


Figure 3. Comparison between calibration results using i) the original likelihood without blocking (solid curves), ii) the block composite likelihood without the variance adjustment (dashed line), and iii) the block composite likelihood with the variance adjustment (dashed-dotted line). The vertical lines represent the assumed true value for our simulation, and the horizontal bars above show the 95% credible intervals. The results shown here are based on two different realizations (two left panels for Realization 1 and two right panels for Realization 2) from the same GP model. The posterior densities with the approximation for the block means (two lower panels) are reasonably close to the densities without the approximation (two upper panels) when the variance adjustment is applied.

Acknowledgement

This work was partly supported by NSF through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management (CLIMA). All views, errors, and opinions are solely those of the authors. We thank Ming-Hui Chen, a Co-Editor and two anonymous referees for helpful comments and suggestions.

References

Bayarri, M., Berger, J., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R., Paulo, R., Sacks, J. and Walsh, D. (2007). Computer model validation with functional output. *Ann. Statist.* **35**, 1874-1906.

Berger, J., De Oliveira, V. and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* **96**, 1361-1374.

Bhat, K., Haran, M. and Goes, M. (2010). Computer model calibration with multivariate spatial output: A case study. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (Edited by M. Chen, P. Müller, D. Sun, K. Ye, and D. Dey), 168-184. Springer, New York.

Bhat, K., Haran, M., Olson, R. and Keller, K. (2012). Inferring likelihoods and climate system characteristics from climate models and multiple tracers. *Environmetrics* **23**, 345-362.

Caragea, P. and Smith, R. (2006). Approximate likelihoods for spatial processes. *Preprint.*

Chang, W., Haran, M., Olson, R. and Keller, K. (2014). Fast dimension-reduced climate model calibration and the effect of data aggregation. arXiv preprint arXiv:1303.1382, *Ann. Appl. Statist.*, in press.

Cox, D. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729-737.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

Eidsvik, J., Shaby, B., Reich, B., Wheeler, M. and Niemi, J. (2013). Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.*, doi:10.1080/10618600.2012.760460.

Flegal, J. M., Haran, M. and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23**, 250-260.

Godambe, V. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **31**, 1208-1211.

Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008). Computer model calibration using high-dimensional output. *J. Amer. Statist. Assoc.* **103**, 570–583.

Higdon, D., Reese, C., Moulton, J., Vrugt, J. and Fox, C. (2009). Posterior exploration for computationally intensive forward models. In *Handbook of Markov Chain Monte Carlo* (Edited by S. Brooks, A. Gelman, G. Jones, X. Meng), 401-418. CRC Press, New York.

Jones, G., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101**, 1537-1547.

Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models. *J. Roy. Statist. Soc. Ser. B.* **63**, 425-464.

Knutti, R. and Hegerl, G. C. (2008). The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nat. Geosci.* **1**, 735-743.

Levitus, S., Antonov, J., Boyer, T., Baranova, O., Garcia, H., Locarnini, R., Mishonov, A., Reagan, J., Seidov, D., Yarosh, E., et al. (2012). World ocean heat content and thermosteric sea level change (0-2000 m), 1955-2010. *Geophys. Res. Lett.* **39**.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.

Ribatet, M., Cooley, D. and Davison, A. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist. Sinica* **22**, 813-846.

Sacks, J., Welch, W., Mitchell, T. and Wynn, H. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4**, 409-423.

Shaby, B. (2013). The open-faced sandwich adjustment for MCMC using estimating functions. *J. Comput. Graph. Statist.* in press.

Sriver, L., Nathan, M., Olson, R. and Keller, K. (2012). Toward a physically plausible upper bound of sea-level rise projections. *Nat. Clim. Chang.* **115**, 893-902.

Varin, C. (2008). On composite marginal likelihoods. *Adv. Statist. Anal.* **92**, 1-28.

Weaver, A., Eby, M., Wiebe, E., Bitz, C., Duffy, P., Ewen, T., Fanning, A., Holland, M., MacFadyen, A. and Matthews, H. (2001). The UVic earth system climate model: Model description, climatology, and applications to past, present and future climates. *Atmos.-Ocean.* **39**, 361-428.

Yakowitz, S. and Szidarovszky, F. (1985). A comparison of kriging with nonparametric regression methods. *J. Multivariate Anal.* **16**, 21-53.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92**, 921-936.

Department of Statistics, University of Chicago, Chicago, IL 60637, USA.

E-mail: wonchang@uchicago.edu

Department of Statistics, The Pennsylvania State University, University park, PA 16802, USA.

E-mail: mharan@stat.psu.edu

Climate Change Research Center, University of New South Wales, Sydney, NSW 2052, Australia.

E-mail: roman.olson@unsw.edu.au

Department of Geosciences, The Pennsylvania State University, University park, PA 16802, USA.

E-mail: klaus@psu.edu

(Received July 2013; accepted March 2014)