A composite likelihood approach to computer model calibration with high-dimensional spatial data

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

1. Department of Statistics, The Pennsylvania State University.
2. Department of Geosciences, The Pennsylvania State University.

Supplementary Material

S1 Proof of Propositions 1 and 2

Proof of Proposition 1: For a composite likelihood, it is sufficient to verify the same regularity conditions as for the usual maximum likelihood estimators (Lindsay (1988)). In the context of expanding domain asymptotics in spatial statistics, the spatial covariance function and its first and second derivatives need to be absolutely summable. From Theorem 3 in Mardia and Marshall (1984), this condition holds for the exponential covariance function that we are using here. (i) and (ii) follow immediately.

Proof of Proposition 2: When the maximum composite likelihood estimator $\boldsymbol{\psi}_n^{CL}$ is consistent and asymptotically normal, (i) and (ii) follow (Theorems 1 and 2, respectively, in Chernozhukov and Hong (2003)). Hence the result follows directly from Proposition 1.

S2 Computation of $\mathbf{P}_n$ and $\mathbf{Q}_n$

In this supplementary material, we describe the matrix computation for $\mathbf{P}_n = \mathrm{Cov}\left(c\dot{\ell}_n(\boldsymbol{\psi})\right)$ and $\mathbf{Q}_n = \mathrm{E}\left(c\ddot{\ell}_n(\boldsymbol{\psi})\right)$. For ease of computation, it is useful to rewrite the composite likelihood function when $p = \infty$ as

$$
\begin{aligned}
c\ell_n(\boldsymbol{\psi}) \propto &-\frac{1}{2}\left(\log|\Sigma^{\bar{\mathbf{Z}}}| + \left(\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*\right)^T \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1} \left(\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*\right)\right) \\
&-\frac{1}{2}\left(\sum_{i=1}^{M} \log|\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}| + \sum_{i=1}^{M} \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]}\right)^T \mathbf{A}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{A}_i \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]}\right)\right),
\end{aligned}
$$

where $\mathbf{A}_i$ is a $(n_i - 1) \times n_i$ matrix such that

$$
\mathbf{A}_i = \left(\mathbf{I}_{(n_i-1)\times(n_i-1)}\ \mathbf{0}_{(n_i-1)\times 1}\right) - \mathbf{a}_i\left(\frac{1}{n_i}, \ldots, \frac{1}{n_i}\right)_{1\times n_i},
$$

and $\mathbf{a}_i$ is a $(n_i - 1) \times 1$ vector such that

$$
\mathbf{a}_i = \left(\zeta_\theta \gamma^{(i)} + \lambda^{(i)}\right) \left\{\Sigma^{\bar{\mathbf{Z}}}\right\}_{ii}^{-1}.
$$

$\mathbf{Z}_{[i]}$ is a $n_i \times 1$ vector containing all the n_i observational data in the ith spatial block without omission, and $\mathbf{Y}_{[i]}$ is a $n_i \times 1$ vector of model output at $\boldsymbol{\theta}^*$ defined in the same way. Omitting the part irrelevant to the data, the partial derivative of $c\ell_n(\boldsymbol{\psi})$ with respect to the jth computer model parameter, θ_j^*, is given by

$$
\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \theta_j^*} \propto \bar{\mathbf{B}}_j^* (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) + \sum_{i=1}^{M} \mathbf{B}_{i,j}^* \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]}\right),
$$

where

$$
\begin{aligned}
\bar{\mathbf{B}}_j^* &= \frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta_j^*} \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1}, \\
\mathbf{B}_{i,j}^* &= \left(\frac{\partial \mathbf{Y}^*_{[i]}}{\partial \theta_j^*}\right)^T \mathbf{A}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{A}_i.
\end{aligned}
$$

We let $\boldsymbol{\xi}$ be the vector containing all the parameters in $\boldsymbol{\xi}_d$ as well as the emulator parameter being re-estimated. The partial derivative with respect to the kth parameter in $\boldsymbol{\xi}$, ξ_k, can be written as

$$\begin{aligned}\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \xi_k} \propto & \quad \frac{1}{2}(\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*)^T \bar{\mathbf{B}}_k^d (\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*) \\ & + \frac{1}{2}\sum_{i=1}^{M} \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]} \right)^T \mathbf{B}_{i,k}^d \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]} \right) \\ & + \sum_{i=1}^{M} \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]} \right)^T \tilde{\mathbf{B}}_{i,k}^d \left(\mathbf{Z}_{[i]} - \mathbf{Y}^*_{[i]} \right)\end{aligned}$$

where

$$\begin{aligned}\bar{\mathbf{B}}_k^d &= \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1} \frac{\partial \Sigma^{\bar{\mathbf{Z}}}}{\partial \xi_k} \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1}, \\ \mathbf{B}_{i,k}^d &= \mathbf{A}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \frac{\partial \Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}}{\partial \xi_k} \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{A}_i, \\ \tilde{\mathbf{B}}_{i,k}^d &= -\left(\frac{\partial \mathbf{A}_i}{\partial \xi_k}\right)^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{A}_i.\end{aligned}$$

Inference on $\boldsymbol{\theta}^*$, our main goal, requires only calculating the asymptotic covariance of $\hat{\boldsymbol{\theta}}_n^B$ due to the asymptotic independence between $\hat{\boldsymbol{\theta}}_n^B$ and $\hat{\boldsymbol{\xi}}_n^B$, the posterior modes of $\boldsymbol{\theta}^*$ and $\boldsymbol{\xi}$ respectively. More specifically, for any j and k,

$$\text{Cov}\left(\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \theta_j^*}, \frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial \xi_k}\right) = 0,$$

because a linear combinations of zero-mean normal random variables and a quadratic form of the same variables are uncorrelated to one another. As a result, $\hat{\boldsymbol{\theta}}_n^B$ and $\hat{\boldsymbol{\xi}}_n^B$ have zero cross-covariance in $\mathbf{G}_n$ and are asymptotically independent due to normality. Let $\mathbf{P}_n^*$ be the part of $\mathbf{P}_n$, the covariance matrix between partial derivatives with respect to the parameters in $\boldsymbol{\theta}^*$ only. Likewise, let $\mathbf{Q}_n^*$ be the part of $\mathbf{Q}_n$ that contains only the negative expected Hessian of the parameters in $\boldsymbol{\theta}^*$. For inference on $\boldsymbol{\theta}^*$, it is sufficient to compute $\mathbf{P}_n^*$ and $\mathbf{Q}_n^*$ instead of $\mathbf{P}_n$ and $\mathbf{Q}_n$.

We compute the (k,l)th element of $\mathbf{P}_n^*$ by plugging in $\hat{\boldsymbol{\psi}}_n^B$ in place of $\boldsymbol{\psi}$ in

$$\begin{aligned}
\mathrm{Cov}\left(\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial\theta_k^*}, \frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial\theta_l^*}\right) = \quad & \bar{\mathbf{B}}_k^* \Sigma^{\bar{\mathbf{Z}}} \left(\bar{\mathbf{B}}_l^*\right)^T \\
+ \quad & \sum_{i=1}^M \sum_{j=1}^M \mathbf{B}_{i,k}^* \; \Sigma_{i,j}^{\mathbf{Z}} \left(\mathbf{B}_{j,l}^*\right)^T \\
+ \quad & \sum_{i=1}^M \bar{\mathbf{B}}_k^* \Sigma_i^{\bar{\mathbf{Z}},\mathbf{Z}} \left(\mathbf{B}_{i,l}^*\right)^T \\
+ \quad & \sum_{i=1}^M \bar{\mathbf{B}}_l^* \Sigma_i^{\bar{\mathbf{Z}},\mathbf{Z}} \left(\mathbf{B}_{i,k}^*\right)^T,
\end{aligned}$$

where $\Sigma_{i,j}^{\mathbf{Z}}$ is the $n_i \times n_j$ covariance matrix between $\mathbf{Z}_{[i]}$ and $\mathbf{Z}_{[j]}$, and $\Sigma_i^{\bar{\mathbf{Z}},\mathbf{Z}}$ is the $1 \times n_i$ covariance matrix between $\bar{\mathbf{Z}}$ and $\mathbf{Z}_{[i]}$ under the probability model in (3). Similarly, the second order partial derivative of $c\ell_n(\boldsymbol{\psi})$ with respect to θ_j^* and θ_k^* is given by

$$\begin{aligned}
\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial\theta_j^*\partial\theta_k^*} \propto \quad & \left(\frac{\partial^2\bar{\mathbf{Y}}^*}{\partial\theta_j^*\partial\theta_k^*}\right)^T \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1} \left(\bar{\mathbf{Z}} - \bar{\mathbf{Y}}^*\right) \\
- \quad & \left(\frac{\partial\bar{\mathbf{Y}}^*}{\partial\theta_j^*}\right)^T \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1} \frac{\partial\bar{\mathbf{Y}}^*}{\partial\theta_k^*} \\
+ \quad & \sum_{i=1}^M \left(\frac{\partial^2\mathbf{Y}_{[i]}^*}{\partial\theta_j^*\partial\theta_k^*}\right)^T \mathbf{B}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{B}_i \left(\mathbf{Z}_{[i]} - \mathbf{Y}_{[i]}^*\right) \\
- \quad & \sum_{i=1}^M \left(\frac{\partial\mathbf{Y}_{[i]}^*}{\partial\theta_j^*}\right)^T \mathbf{B}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{B}_i \frac{\partial\mathbf{Y}_{[i]}^*}{\partial\theta_k^*}.
\end{aligned}$$

The (j,k)th element of $\mathbf{Q}_n^*$ is computed by substituting $\boldsymbol{\psi}$ with $\hat{\boldsymbol{\psi}}_n^B$ in

$$\begin{aligned}
-E\left(\frac{\partial c\ell_n(\boldsymbol{\psi})}{\partial\theta_j^*\partial\theta_k^*}\right) \propto \quad & \left(\frac{\partial\bar{\mathbf{Y}}^*}{\partial\theta_j^*}\right)^T \left(\Sigma^{\bar{\mathbf{Z}}}\right)^{-1} \frac{\partial\bar{\mathbf{Y}}^*}{\partial\theta_k^*} \\
+ \quad & \sum_{i=1}^M \left(\frac{\partial\mathbf{Y}_{[i]}^*}{\partial\theta_j^*}\right)^T \mathbf{B}_i^T \left(\Sigma_i^{\mathbf{Z}|\bar{\mathbf{Z}}}\right)^{-1} \mathbf{B}_i \frac{\partial\mathbf{Y}_{[i]}^*}{\partial\theta_k^*}.
\end{aligned}$$

Computing $\mathbf{P}_n^*$ and $\mathbf{Q}_n^*$ requires finding the first-order derivatives of $\mathbf{Y}_{[1]}^*, \ldots, \mathbf{Y}_{[M]}^*$, and $\bar{\mathbf{Y}}^*$. Since they are unknown functions of $\boldsymbol{\theta}^*$, we approximate them using the corresponding derivatives of the emulator output. The approximated derivatives of $\bar{\mathbf{Y}}^*$ and $\mathbf{Y}^*_{[i]}$ with respect to θ^*_j are given by

$$\begin{aligned}\frac{\partial \bar{\mathbf{Y}}^*}{\partial \theta^*_j} &= \left(\mathbf{I}_M \otimes \left(\frac{\partial \Sigma_{\theta^*\theta}}{\partial \theta^*_j}\Sigma_\theta^{-1}\right)\right)\bar{\mathbf{Y}},\\ \frac{\partial \mathbf{Y}^*_{[i]}}{\partial \theta^*_j} &= \left(\mathbf{I}_{n_i} \otimes \left(\frac{\partial \Sigma_{\theta^*\theta}}{\partial \theta^*_j}\Sigma_\theta^{-1}\right)\right)\mathbf{Y}_{[i]}.\end{aligned}$$

The derivative term $\frac{\partial \Sigma_{\theta^*\theta}}{\partial \theta^*_j}$ is determined by the covariance function for the parameter space. For the exponential covariance function used in our example, the derivative is

$$\left\{\frac{\partial \Sigma_{\theta^*\theta}}{\partial \theta^*_i}\right\}_j = \phi_{\theta,i}(-1)^{1(\theta^*_i > \theta_{ij})} \exp\left(-\sum_{k=1}^{q} \phi_{\theta,k}\,|\theta^*_k - \theta_{kj}|\right),\ i = 1,\dots,q,\ j = 1,\dots,p,$$

where $1(\cdot)$ is the indicator function, and θ_{ij} is the ith parameter value of the jth design point $\boldsymbol{\theta}_j$.

S3 Details of curvature adjustment

The approach due to Ribatet et al. (2012) substitutes $\boldsymbol{\psi}$ in (3.2) with $\tilde{\boldsymbol{\psi}}^{curv} = \hat{\boldsymbol{\psi}}_n^B + \mathbf{D}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_n^B)$, where $\boldsymbol{\psi}$ is the posterior mode from (3.2). $\mathbf{D}$ is the matrix that satisfies $\mathbf{D}^T\mathbf{Q}_n\mathbf{D} = \mathbf{Q}_n\mathbf{P}_n^{-1}\mathbf{Q}_n$. This approach ensures that the resulting posterior distribution has the same mode as the original composite likelihood $c\ell_n(\boldsymbol{\psi})$ and the asymptotic covariance $\mathbf{G}_n^{-1}$, as described in Proposition 2. (ii). The choice for $\mathbf{D}$ is not unique, and Ribatet et al. (2012) suggested using $\mathbf{D} = \mathbf{Q}_n^{\frac{1}{2}}\left(\mathbf{Q}_n\mathbf{P}_n^{-1}\mathbf{Q}_n\right)^{\frac{1}{2}}$, where the square roots of the matrices are computed using singular value decomposition. Here we use the open-faced adjustment; the curvature adjustment approach can also be used but, as shown in (Shaby (2013)), the difference between the two approaches is likely to be minimal.

S4 Guideline for choosing number of blocks

One heuristic approach is to ensure that the maximum distance within each block exceeds the effective range (Eidsvik et al. (2013), Zhang and Zimmerman (2005)), which is the spatial distance at which correlation reduces to 0.05. For the exponential covariance function we use for the discrepancy term in this work, this corresponds to $3\frac{1}{\phi_d}$. In practice, however, it might be difficult to strictly follow this guideline due to the long range dependence for the discrepancy process $\boldsymbol{\delta}$. In such case, one needs to choose the smallest number of blocks that still enables feasible likelihood computation. In a simulated example below, we examine how this affects the inference results.

S5 Caveats

While our approach is helpful in mitigating computational issues for various calibration problems, There is still more work to be done to make the computation more efficient. As n continues to get large the number of spatial locations in each block may become excessively large and evaluation of composite likelihood may not be computationally tractable. One can consider increasing the number of blocks until the computation becomes feasible, but then the convergence of the posterior modes may be very slow due to too small block sizes (Cox and Reid (2004), Varin (2008)). Another perhaps simpler approach is to use a composite likelihood framework that does not involve blocks though this may involve the need for analytical work to establish posterior propriety.

Another possible issue is related to the use of a Gaussian emulator in place of the true computer model in computing $\mathbf{P}_n$ and $\mathbf{Q}_n$. Using a Gaussian process emulator, we approximate not only the true computer model itself, but also its first and second derivatives. In our particular example, this did not cause any problem due to very regular behavior of the computer model output with respect to the input parameters. However, this may not be true in general and therefore $\mathbf{P}_n$ and $\mathbf{Q}_n$ calculations may be inaccurate.

The asymptotic independence between input parameters and discrepancy parameters does not usually hold in a finite sample. It is well known that calibration models usually suffer from identifiability issues (Wynn (2001)). One way to avoid the issues is to impose discrepancy prior information on the discrepancy term (Arendt et al. (2012)) as we did in Section 4.

The scientific result shown in Figure 3 requires some caution in its interpretation. First, besides climate sensitivity, climate system response to changes in radiatively active gases in the atmosphere also depends on the magnitude of the radiative effects of these gases ("radiative forcing"), and on the vertical mixing of heat into the deep ocean (Hansen et al. (1985), Knutti et al. (2002), Schmittner et al. (2009), Urban and Keller (2010)). The parameters controlling the forcing and the vertical mixing were kept fixed, respectively, at 0.1 and 1.5 in the model runs we use. Including these additional uncertainties is expected to make the posterior density of CS more dispersed. The example serves as a demonstration of computational feasibility of our approach when applied to high-dimensional spatial datasets rather than providing an improved estimate of CS. Second, the variability of the posterior density is sensitive to the prior information for the discrepancy term; this is a common problem for many calibration problems, as discussed earlier.

References

Arendt, P. D., Apley, D. W., and Chen, W. (2012). Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability. *J. Mech. Design* **134**, 100908.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *J. Econometrics* **115**, 293–346.

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

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

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., and Stone, P. (1985). Climate response times - dependence on climate sensitivity and ocean mixing. *Science* **229**, 857–859.

Knutti, R., Stocker, T. F., Joos, F., and Plattner, G. K. (2002). Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221–39.

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. *Stat. Sinica* **22**, 813–846.

Schmittner, A., Urban, N., Keller, K., and Matthews, D. (2009). Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate–carbon cycle projections. *Global Biogeochem. Cy.* **23**, 737–750.

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

Urban, N. and Keller, K. (2010). Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: a Bayesian fusion of century-scale observations with a simple model. *Tellus A* **62**, 737–750.

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

Wynn, H. (2001). Contribution to the discussion on the paper by Kennedy and O'Hagan. *J. R. Stat. Soc. B.* **63**, 425–464.

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

Table S1: Summary of notation used in Section 2

Symbol	Definition
p	number of design points in parameter space
n	number of spatial locations for computer model grid
$\boldsymbol{\Theta}$	open set of all possible computer model parameter settings
$\boldsymbol{\theta}_i$	parameter setting for ith model run
$\mathcal{S}$	spatial field of interest
$\mathbf{s}_j$	jth location on computer model grid
$Y(\mathbf{s}_j, \boldsymbol{\theta}_i)$	model output at location $\mathbf{s}_j$ for parameter setting $\boldsymbol{\theta}_i$
$\mathbf{Y}_i$	model output for ith parameter setting, $(Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \ldots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$
$Z(\mathbf{s}_j)$	observation at spatial location $\mathbf{s}_j$
$\mathbf{Z}$	observational data, $(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$
$\mathbf{X}$	covariate matrix
$\boldsymbol{\beta}$	vector of regression coefficients
$\boldsymbol{\xi}_y$	vector of covariance parameters for emulator
$\eta(\mathbf{s}, \boldsymbol{\theta})$	emulator process at location $\mathbf{s}$ and $\boldsymbol{\theta}$
$\boldsymbol{\eta}(\boldsymbol{\theta})$	emulator output at parameter setting $\boldsymbol{\theta}$, $(\eta(\mathbf{s}_1, \boldsymbol{\theta}), \ldots, \eta(\mathbf{s}_n, \boldsymbol{\theta}))^T$
$\boldsymbol{\theta}^*$	true or fitted value of computer model parameter for observational data
$\delta(\mathbf{s})$	model-observation discrepancy at location $\mathbf{s}$
$\boldsymbol{\delta}$	discrepancy process observed at grid locations, $(\delta(\mathbf{s}_1), \ldots, \delta(\mathbf{s}_n))^T$

Table S2: Summary of notation used for emulation model in Section 3

Symbol	Definition
M	number of blocks
n_i	number of spatial locations in ith block
$\mathbf{s}_{ij}$	jth spatial location in ith block
$\mathbf{Y}_{(i)}$	model output for ith block, $\left(Y(\mathbf{s}_{i1},\cdot)^T, Y(\mathbf{s}_{i2},\cdot)^T, \ldots, Y(\mathbf{s}_{in_i-1},\cdot)^T\right)^T$
$Y(\mathbf{s}_{ij},\cdot)$	model output at location $\mathbf{s}_{ij}$, $(Y(\mathbf{s}_{ij},\boldsymbol{\theta}_1),\ldots,Y(\mathbf{s}_{ij},\boldsymbol{\theta}_p))^T$
$\bar{\mathbf{Y}}_{(i)}$	mean vector for ith block, $\frac{1}{n_i}\sum_{j=1}^{n_i}(Y(\mathbf{s}_{ij},\boldsymbol{\theta}_1),\ldots,Y(\mathbf{s}_{ij},\boldsymbol{\theta}_p)))^T$
$\bar{\mathbf{Y}}$	vector of all block means, $\left(\bar{\mathbf{Y}}_{(1)}^T,\ldots,\bar{\mathbf{Y}}_{(M)}^T\right)$
$\boldsymbol{\xi}_s$	parameters for spatial covariance function K_s
$K_s(\mathbf{s},\mathbf{s}';\boldsymbol{\xi}_s)$	spatial covariance function between $\mathbf{s}$ and $\mathbf{s}'$ for emulator
$\boldsymbol{\xi}_\theta$	parameters for covariance function K_θ
$K_\theta(\boldsymbol{\theta},\boldsymbol{\theta}';\boldsymbol{\xi}_\theta)$	covariance function between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$
$\Sigma^{\bar{\mathbf{Y}}}$	covariance between the block means, $\mathrm{H}\otimes\Sigma_\theta$
H	spatial covariance for block means, $\{\mathrm{H}\}_{ij}=\frac{1}{n_in_j}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}K_s(\mathbf{s}_{ik},\mathbf{s}_{jl};\boldsymbol{\xi}_s)$
Σ_θ	covariance between p parameter settings, $\{\Sigma_\theta\}_{ij}=K_\theta(\boldsymbol{\theta}_i,\boldsymbol{\theta}_j;\boldsymbol{\xi}_\theta)$
$\mu_i^{\mathbf{Y}\mid\bar{\mathbf{Y}}}$	conditional mean of $\mathbf{Y}_{(i)}$ given $\bar{\mathbf{Y}}_{(i)}$, $(\gamma^{(i)}/\{\mathrm{H}\}_{ii}\otimes\mathbf{I}_p)\bar{\mathbf{Y}}_{(i)}$
$\Sigma_i^{\mathbf{Y}\mid\bar{\mathbf{Y}}}$	conditional variance of $\mathbf{Y}_{(i)}$ given $\bar{\mathbf{Y}}_{(i)}$, $\left(\Gamma_i-\gamma^{(i)}\left(\gamma^{(i)}\right)^T/\{\mathrm{H}\}_{ii}\right)\otimes\Sigma_\theta$
$\gamma^{(i)}$	spatial covariance between ith block mean and block locations, $\{\gamma^{(i)}\}_j=\frac{1}{n_i}\sum_{k=1}^{n_i}K_s(\mathbf{s}_{ij},\mathbf{s}_{ik};\boldsymbol{\xi}_s)$
Γ_i	spatial covariance matrix for ith block, $\{\Gamma_i\}_{jk}=K_s(\mathbf{s}_{ij},\mathbf{s}_{ik};\boldsymbol{\xi}_s)$

Table S3: Summary of notation used for calibration model in Section 3

Symbol	Definition
$\mathbf{Z}_{(i)}$	observations for ith block, $(Z(\mathbf{s}_{i1}),\ldots,Z(\mathbf{s}_{in_i}))^T$
$\bar{\mathbf{Z}}_{(i)}$	ith block mean for observations
$\bar{\mathbf{Z}}$	collection of block means for observations, $\left(\bar{\mathbf{Z}}_{(1)},\ldots,\bar{\mathbf{Z}}_{(M)}\right)^T$
$\boldsymbol{\xi}_d$	parameters for spatial covariance function K_d
$K_d(\mathbf{s},\mathbf{s}';\boldsymbol{\xi}_d)$	spatial covariance function between $\mathbf{s}$ and $\mathbf{s}'$ for discrepancy
$\boldsymbol{\Omega}$	spatial covariance matrix for block means of discrepancy, $\{\boldsymbol{\Omega}\}_{ij}=\frac{1}{n_in_j}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}K_d(\mathbf{s}_{ik},\mathbf{s}_{jl};\boldsymbol{\xi}_d)$
$\Sigma_{\theta^*\theta}$	covariance between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}$
$\mu^{\bar{\mathbf{Z}}}$	mean vector for the block means $\bar{\mathbf{Z}}$, $\left(\mathbf{I}_M\otimes\Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\right)\bar{\mathbf{Y}}$
$\Sigma^{\bar{\mathbf{Z}}}$	covariance matrix for $\bar{\mathbf{Z}}$, $\mathrm{H}\otimes\left(\Sigma_{\theta^*}-\Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T\right)+\Omega$
$\lambda^{(i)}$	spatial covariance between ith block mean and block locations for discrepancy, $\{\lambda^{(i)}\}_j=\frac{1}{n_i}\sum_{k=1}^{n_i}K_d(\mathbf{s}_{ij},\mathbf{s}_{ik};\boldsymbol{\xi}_d)$
Λ_i	spatial covariance for ith block locations for discrepancy, $\{\Lambda_i\}_{jk}=K_d(\mathbf{s}_{ij},\mathbf{s}_{ik};\boldsymbol{\xi}_d)$
$\mu_i^{\mathbf{Z}\mid\bar{\mathbf{Z}}}$	mean of $\mathbf{Z}_{(i)}$ given $\bar{\mathbf{Z}}_{(i)}$, $\left(\mathbf{I}_{n_i-1}\otimes\Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\right)\mathbf{Y}_{(i)}+(\tau^{(i)}+\lambda^{(i)})\left\{\Sigma^{\bar{\mathbf{Z}}}\right\}_{ii}^{-1}(\bar{\mathbf{Z}}_i-\left\{\mu^{\bar{\mathbf{Z}}}\right\}_i)$
$\Sigma_i^{\mathbf{Z}\mid\bar{\mathbf{Z}}}$	covariance of $\mathbf{Z}_{(i)}$ given $\bar{\mathbf{Z}}_{(i)}$, $\left(\Gamma_i\otimes\left(\Sigma_{\theta^*}-\Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T\right)+\Lambda_i\right)-(\tau^{(i)}+\lambda^{(i)})(\tau^{(i)}+\lambda^{(i)})^T/\{\Sigma_{\bar{\mathbf{Z}}}\}_{ii}$
$\tau^{(i)}$	$\gamma^{(i)}\otimes\left(\Sigma_{\theta^*}-\Sigma_{\theta^*\theta}\Sigma_\theta^{-1}\Sigma_{\theta^*\theta}^T\right)$

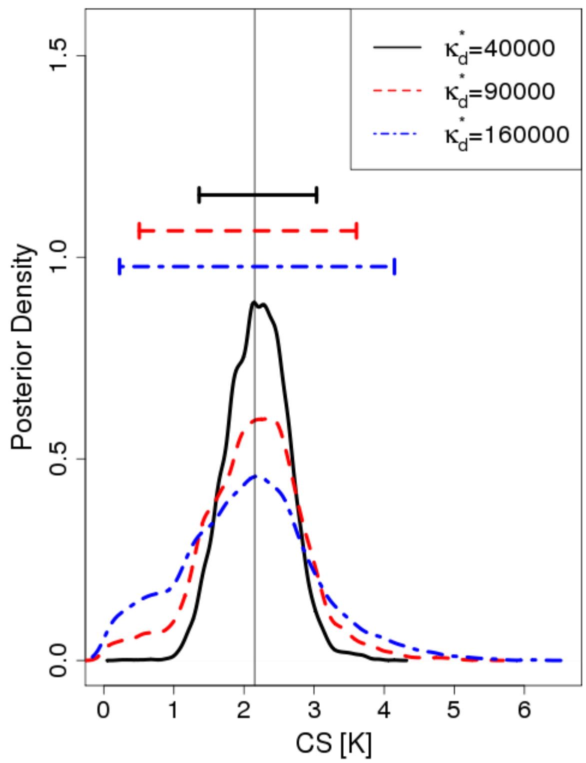

Figure S1: Comparison of posterior densities between three simulated examples with different assumed magnitudes of the discrepancies: $\kappa_d^* = 40000$ (solid black curve), $\kappa_d^* = 90000$ (dashed red curve), and $\kappa_d^* = 160000$ (dotted-dashed blue curve). The vertical line indicates the assumed true value, and the horizontal bars above show the 95% credible intervals. As the discrepancy grows, the densities become more dispersed but the posterior modes stay similar.