Bayesian Inference for Sensitivity Analysis of Computer Simulators, with an Application to Radiative Transfer Models

Marian Farah
MRC Biostatistics Unit
Institute of Public Health
Cambridge, CB20SR, U.K.
(marian.farah@mrc-bsu.cam.ac.uk)

Athanasios Kottas
Department of Applied Mathematics and Statistics
University of California Santa Cruz
Santa Cruz, California 95064
(thanos@ams.ucsc.edu)

REVISED April 12, 2012

Computer simulators are used in science and technology to model physical processes or the behavior of real-world systems. Sensitivity analysis provides a useful tool for quantifying the impact of uncertainty in the computer simulator inputs on the computed output. We focus on global sensitivity analysis, which quantifies output uncertainty as all the inputs vary continuously over the input space. The influence of each input and how uncertainty in the output is apportioned amongst the inputs are determined by calculating the main effects and sensitivity indices of the computer simulator inputs. Typically, these quantities are computed using Monte Carlo methods, which require a large number of computer simulator runs, making the calculations infeasible if the simulator is computationally expensive. Bayesian methods have been used to tackle sensitivity analysis of computationally expensive simulators through building a statistical emulator for the computer simulator output, typically, based on a Gaussian process prior for the simulator output function. In this work, we develop an approach for integrating global sensitivity analysis tools and extending semi-Bayesian approaches to a fully Bayesian methodology. The approach is utilized to carry out sensitivity analysis of the Leaf-Canopy Model, a radiative transfer model that simulates the interaction of sunlight with vegetation.

KEY WORDS: Gaussian process; Leaf-Canopy Model; Main effects; Sensitivity indices.

1 Introduction

Complex process models are widely used in science and engineering to study various real-world systems in order to understand underlying processes and make predictions about their future behavior. These mathematical models are implemented in computer code, which can take from a fraction of a second to several days in order to obtain a single run. We refer to the mathematical model and its computer code implementation as a simulator. Here, we consider deterministic simulators, i.e., different runs of the simulator using the same inputs result in the same output. In many cases, the process model inputs are not easily observable, and thus, there is uncertainty about the values of the simulator inputs. Describing and quantifying the induced uncertainty in the simulator output due to uncertainty in its inputs is known as sensitivity analysis.

Sensitivity analysis is a valuable tool in model development, calibration, and validation, since it can be used to identify where a model can be improved by obtaining better input information. In general, there are two approaches for sensitivity analysis, local and global. Local sensitivity analysis determines how the output changes as the inputs are each varied about a fixed point, while global sensitivity analysis studies how the output changes as all the inputs vary continuously over the entire input space; see Saltelli et al. (2000) for a review of both approaches. For global sensitivity analysis, which is the subject of this paper, the sensitivity of the simulator output is measured via the calculation of the "main effects", which provide a summary of the influence of each input on the model output, and by the "sensitivity indices", which are variance-based measures that give the expected amount by which the uncertainty in the output would be reduced if the true value of the input was known.

Calculating the main effects and sensitivity indices requires the evaluation of multidimensional integrals over the input space of the simulator. Thus, standard numerical integration methods (e.g., Monte Carlo integration or multidimensional quadrature) are infeasible when the simulator is computationally expensive. This problem has been tackled through building a statistical emulator, which is a computationally efficient statistical approximation of the simulator output. The Gaussian process (GP) provides a popular approach to developing the emulator because it is a convenient and, in general, flexible statistical model. In Bayesian modeling, it is commonly used as a prior model for an unknown function, such as the output of a computer simulator. Using GP emulators for deterministic simulators dates back to the work of Sacks et al. (1989); see, e.g., the book by Santner et al. (2003). Moreover, in more recent years there has been an upsurge in research activity on Bayesian methods for analysis of computer simulators; see, e.g., Kennedy and O'Hagan (2001), Craig et al. (2001), Oakley and O'Hagan (2002),

Higdon et al. (2004), Goldstein and Rougier (2006), Bayarri et al. (2007), Bayarri et al. (2009), Han et al. (2009), and Gramacy and Lian (2012).

Once posterior inference is obtained using the GP model, the main effects and sensitivity indices can be determined using runs from the emulator's posterior predictive distribution, which are substantially faster than simulator runs. Oakley and O'Hagan (2004) develop a semi-Bayesian estimation approach for the main effects based on fixed GP "range of dependence" parameters. They also propose approximate point estimates for the sensitivity indices based on ratios of posterior expectations to estimate the posterior expectation of ratios of variances.

A similar approach, albeit based on likelihood estimation for the GP parameters, is used by Morris et al. (2008). Also under a likelihood estimation setting, Marrel et al. (2009) compare sensitivity indices built from either the entire stochastic process for the GP emulator or only its mean. An example of an approach that does not utilize GP emulation can be found in Ziehn and Tomlin (2009), where orthonormal polynomial expansions are used to approximate the variance components needed for the calculation of sensitivity indices. Finally, seeking to extend the scope of variance-based sensitivity analysis methods, Oakley (2009) presents a decision-theoretic framework which allows the model user (or decision maker) to relate the importance of each uncertain input to the model user's optimal decision.

The impetus for our work is full inference for sensitivity analysis, including appropriate uncertainty quantification for main effects and sensitivity indices. The starting point of our approach involves approximating a computationally expensive simulator by a fully Bayesian GP model. Based on runs of the GP posterior predictive distribution, we develop an approach to full inference for global sensitivity analysis. First, we calculate Bayesian point estimates of the main effects and their associated uncertainties. This approach is appealing because it utilizes analytic expressions to estimate the main effects based on the GP model, which results in efficient computation. Next, we design a method to obtain full posterior distributions of different types of sensitivity indices over the input space of the model. This latter method expands the inference scope of the earlier work in Oakley and O'Hagan (2004), Morris et al. (2008), and Marrel et al. (2009). A similar approach to full inference for sensitivity indices was discussed in Taddy et al. (2009), based on a treed GP prior (Gramacy and Lee, 2008) for the simulator output; this approach is implemented in the tgp package for `R` as detailed in Gramacy and Taddy (2010).

The motivating application for this work is provided by the Leaf-Canopy Model (LCM), a radiative transfer model for the interaction of sunlight with vegetation. The methodology is applied to the LCM to estimate the main effects and sensitivity indices of each of its inputs at 8

different MODIS spectral bands that are sensitive to vegetation; MODIS (Moderate Resolution Imaging Spectroradiometer) is a key instrument aboard the Terra and Aqua satellites.

The outline of the paper is as follows. Section 2 develops the methodology for sensitivity analysis, with some of the technical details provided in the appendices. Section 3 includes a description of the LCM model and reports inference results on sensitivity analysis for its inputs. Finally, Section 4 concludes with a summary and discussion.

2 Methods

To prepare the ground for the proposed methodology, Section 2.1 reviews GP-based emulation for computer simulators. The approach to inference for sensitivity analysis is presented in Section 2.2. In particular, Section 2.2.1 develops a computationally efficient approach to point and interval estimation for the main effects, whereas in Section 2.2.2, we describe a method to sample the entire posterior distribution of the sensitivity indices.

2.1 Gaussian process emulation

An emulator is a computationally efficient statistical model that is used to approximate a computationally expensive simulator. Denote by $f(\boldsymbol{v})$ the simulator output as a function of input vector $\boldsymbol{v} = (v_1, \ldots, v_k)$. Given a set of training model runs $D = \{(\boldsymbol{x}_i, y_i) : i = 1, \ldots, n\}$, where $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{ki})$ is the i–th realized design input vector, and $y_i = f(\boldsymbol{x}_i)$ is the corresponding output, the emulator treats the computer simulator as a black box and uses D to estimate $f(\cdot)$. Following the work of Sacks et al. (1989) and Kennedy and O'Hagan (2001), GPs are widely used to model the computer code output function $f(\cdot)$. The advantage of a GP emulator is that it is a fully specified statistical model that requires one carefully chosen set of model runs. While the GP approximation introduces uncertainty into the computation of the main effects and sensitivity indices, this uncertainty is quantifiable.

Under a GP prior for function $f(\cdot)$, for any finite set of input points $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N)$, where $\boldsymbol{v}_i = (v_{1i}, \ldots, v_{ki})$ for $i = 1, \ldots, N$, the joint distribution of the outputs $(f(\boldsymbol{v}_1), \ldots, f(\boldsymbol{v}_N))$ is multivariate normal. Furthermore, GP models typically assume that the output is a smooth function of its inputs, that is, nearby locations in the input space produce outputs that are stochastically close in value. A GP is fully specified by its mean function, $\mathrm{E}\left(f(\boldsymbol{v})\right)$, and positive definite covariance function, $\mathrm{Cov}\left(f(\boldsymbol{v}_i), f(\boldsymbol{v}_j)\right)$. We assume constant mean function, $\mathrm{E}\left(f(\boldsymbol{v})\right) = \mu$, and covariance function, $\mathrm{Cov}\left(f(\boldsymbol{v}_i), f(\boldsymbol{v}_j)\right) = \tau^2 \mathrm{Corr}\left(f(\boldsymbol{v}_i), f(\boldsymbol{v}_j)\right)$, which is taken to be

isotropic with constant variance, τ^2, and a product power exponential correlation of the form,

$$R_{\boldsymbol{\phi}} = \text{Corr}\left(f(\boldsymbol{v}_i), f(\boldsymbol{v}_j)\right) = \exp\left\{-\sum_{\ell=1}^{k} \phi_\ell |v_{\ell i} - v_{\ell j}|^{a_\ell}\right\}, \tag{1}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)$, with $\phi_\ell > 0$, is the vector of "range of dependence" parameters, which control the dependence strength in each of the component directions of $\boldsymbol{v}$. Here, $a_\ell \in [1, 2]$ are the "smoothness" parameters, which are typically fixed based on a combination of prior knowledge about $f(\cdot)$ and computational considerations (e.g., Higdon et al., 2004). For example, a value of $a_\ell = 2$ implies that $f(\cdot)$ is a smooth infinitely differentiable function, whereas smaller values of a_ℓ result in rougher (continuous) realizations. The form of the correlation function in (1) corresponds to a choice commonly used for GP emulation of computer simulators. In our context, it facilitates computing of some of the integrals needed for estimation of the main effects (see Section 2.2.1). However, the approach to sensitivity analysis presented in Section 2.2 is sufficiently generic to allow application under more general GP covariance functions.

To obtain the set of training data, D, we use a Latin Hypercube design (McKay et al., 1979) to generate the design matrix of the model inputs and calculate the corresponding outputs using the simulator. We treat the functional form of the simulator output, $f(\cdot)$, as unknown and specify a prior for it in the form of the isotropic GP discussed above. We note that it is often useful (or necessary) to add a small (fixed) *jitter* term to the covariance function for numerical stability. This is a standard computational strategy in Bayesian nonparametric regression as well as analysis of simulators with GP priors (e.g., Neal, 1998; Higdon et al., 2004).

Given D, there are n induced variables from the GP representation for $f(\cdot)$, that is, $y_i = f(\boldsymbol{x}_i)$, for $i = 1, \ldots, n$, with induced prior $(f(x_1), \ldots, f(x_n)) \sim N_n\left(\mu \mathbf{1}_n, \tau^2 R_{\boldsymbol{\phi}}\right)$. Here, $\mathbf{1}_n$ is the n–dimensional vector with all elements equal to 1, and $R_{\boldsymbol{\phi}}$ is the $n \times n$ observed correlation matrix with (i, j)–th element given by $\exp\left\{-\sum_{\ell=1}^{k} \phi_\ell |x_{\ell i} - x_{\ell j}|^{a_\ell}\right\}$. To complete the Bayesian model for the GP emulator, we fix a_ℓ, $\ell = 1, \ldots, k$, and place (independent) priors on the hyperparameters of the GP, μ, τ^2, and ϕ_ℓ, $\ell = 1, \ldots, k$. Thus, the joint posterior distribution of all parameters, $\boldsymbol{\psi} = (\mu, \tau^2, \boldsymbol{\phi})$, is given by $p(\boldsymbol{\psi} \mid D) \propto N_n\left(\boldsymbol{y} | \mu \mathbf{1}_n, \tau^2 R_{\boldsymbol{\phi}}\right) p(\mu) p(\tau^2) p(\phi_1) \ldots p(\phi_k)$, where $\boldsymbol{y} = (y_1, ..., y_n)$. Samples from $p(\boldsymbol{\psi} \mid D)$ are obtained using Markov chain Monte Carlo (MCMC) posterior simulation as discussed in Appendix A.

Analysis of simulator output performed using runs of the emulator have an additional level of uncertainty, since those runs are an approximation of the computer code output. We account for this uncertainty by performing any further analysis over the posterior predictive distribution

of the GP. For any generic input, $\boldsymbol{v} = (v_1, \ldots, v_k)$, which is not part of the design, we can obtain the posterior predictive distribution for $\tilde{y} = f(\boldsymbol{v})$. Specifically,

$$p\left(\tilde{y} \mid D\right) = \int N\left(\tilde{y} \mid m(\boldsymbol{v}), s^2(\boldsymbol{v})\right) p\left(\boldsymbol{\psi}|D\right)\, d\boldsymbol{\psi} \tag{2}$$

with

$$m(\boldsymbol{v}) \equiv \mathrm{E}\left(\tilde{Y}|\boldsymbol{\psi}, D\right) = \mu + \boldsymbol{r}^T(\boldsymbol{v}) R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right), \tag{3}$$

$$s^2(\boldsymbol{v}) \equiv \mathrm{Var}\left(\tilde{Y}|\boldsymbol{\psi}, D\right) = \tau^2\left(1 - \boldsymbol{r}^T\left(\boldsymbol{v}\right) R_{\boldsymbol{\phi}}^{-1} \boldsymbol{r}(\boldsymbol{v})\right), \tag{4}$$

where $\boldsymbol{r}(\boldsymbol{v})$ is the $n \times 1$ vector with i–th element given by $\mathrm{Corr}\left(f(\boldsymbol{v}), f(\boldsymbol{x}_i)\right)$.

The joint predictive distribution for $(\tilde{y}, \tilde{y}') = (f(\boldsymbol{v}), f(\boldsymbol{v'}))$ corresponding to generic inputs $\boldsymbol{v} = (v_1, \ldots, v_k)$ and $\boldsymbol{v'} = (v_1', \ldots, v_k')$ is given by $p(\tilde{y}, \tilde{y}' \mid D) = \int p(\tilde{y}, \tilde{y}' \mid \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi}$, where $p(\tilde{y}, \tilde{y}' \mid \boldsymbol{\psi})$ is bivariate normal with (2×1) mean vector

$$\boldsymbol{\omega}(\boldsymbol{v}, \boldsymbol{v'}) = \mu \mathbf{1}_2 + R^T\left(\boldsymbol{v}, \boldsymbol{v'}\right) R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right), \tag{5}$$

and (2×2) covariance matrix

$$C(\boldsymbol{v}, \boldsymbol{v'}) = \tau^2\left(B\left(\boldsymbol{v}, \boldsymbol{v'}\right) - R^T\left(\boldsymbol{v}, \boldsymbol{v'}\right) R_{\boldsymbol{\phi}}^{-1} R\left(\boldsymbol{v}, \boldsymbol{v'}\right)\right), \tag{6}$$

where $B\left(\boldsymbol{v}, \boldsymbol{v'}\right)$ is the (2×2) correlation matrix for $(f(\boldsymbol{v}), f(\boldsymbol{v'}))$, and $R\left(\boldsymbol{v}, \boldsymbol{v'}\right)$ is the $(n \times 2)$ matrix, where the elements of the first column are given by $\mathrm{Corr}\left(f(\boldsymbol{v}), f(\boldsymbol{x}_i)\right)$, $i = 1, \ldots, n$, and the elements of the second column by $\mathrm{Corr}\left(f(\boldsymbol{v'}), f(\boldsymbol{x}_i)\right)$, $i = 1, \ldots, n$.

2.2 Fully Bayesian inference for global sensitivity analysis

Variance-based sensitivity analysis builds from a decomposition of the simulator output function into summands of increasing dimensionality. Specifically, for a k–dimensional input space,

$$y = f\left(\boldsymbol{v}\right) = f_0 + \sum_{\ell=1}^{k} f_\ell(v_\ell) + \sum_{1 \le \ell < m \le k} f_{\ell,m}\left(v_\ell, v_m\right) + \cdots + f_{1,2,\ldots,k}\left(v_1, \ldots, v_k\right).$$

Here, f_0 is the global mean given by $f_0 = \mathrm{E}\left(Y\right) = \int_{\boldsymbol{v}} f\left(\boldsymbol{v}\right) dH\left(\boldsymbol{v}\right)$, where $H\left(\boldsymbol{v}\right) = \prod_{\ell=1}^{k} H_\ell(v_\ell)$ is the uncertainty distribution of the inputs comprising independent components $H_\ell(v_\ell)$. The next k terms are the main effects, where $f_\ell(v_\ell)$ is the main effect of input v_ℓ, providing a measure

of the influence of input v_ℓ on the computed output. For $\ell = 1, \ldots, k$,

$$f_\ell(v_\ell) = \mathrm{E}\left(Y|v_\ell\right) - \mathrm{E}\left(Y\right) = \int_{\boldsymbol{v}_{-\ell}} f\left(\boldsymbol{v}\right) dH\left(\boldsymbol{v}_{-\ell}|v_\ell\right) - \mathrm{E}\left(Y\right),$$

where $\boldsymbol{v}_{-\ell}$ denotes input vector $\boldsymbol{v}$ excluding element v_ℓ. Because of the independent components of the uncertainty distribution, the conditional distribution $H\left(\boldsymbol{v}_{-\ell}|v_\ell\right)$ simplifies to $H\left(\boldsymbol{v}_{-\ell}\right)$. The remaining terms of the decomposition are the interactions, which quantify the combined influence on the output of two or more inputs taken together. For instance, the first-order interactions, $f_{\ell,m}(v_\ell, v_m) = \mathrm{E}(Y|v_\ell, v_m) - f_\ell(v_\ell) - f_m(v_m) - \mathrm{E}(Y)$.

Sobol (1993) shows that based on this output decomposition, and assuming independence between the input variables in the uncertainty distribution, the total variance, $\mathrm{Var}(Y) = W$, can also be decomposed as the sum of partial variances,

$$W = \sum_{\ell=1}^{k} W_\ell + \sum_{1 \le \ell < m \le k} W_{\ell,m} + \cdots + W_{1,2,\ldots,k}, \tag{7}$$

where $W_\ell = \mathrm{Var}\left(f_\ell(v_\ell)\right) = \mathrm{Var}(\mathrm{E}(Y|v_\ell))$, $W_{\ell,m} = \mathrm{Var}\left(f_{\ell,m}(v_\ell, v_m)\right)$, and analogously for the higher order terms. Hence, the sensitivity indices are given by

$$S_\ell = \frac{W_\ell}{W}, \;\; S_{\ell,m} = \frac{W_{\ell,m}}{W}, \;\; \ldots \;\; , \;\; S_{1,2,\ldots,k} = \frac{W_{1,2,\ldots,k}}{W},$$

where S_ℓ is the *first-order sensitivity index* for input v_ℓ, which measures the fractional contribution of that input to the variance of $f(\boldsymbol{v})$, $S_{\ell,m}$, for $\ell \neq m$, is the *second-order sensitivity index*, which measures the contribution of interaction due to inputs v_ℓ and v_m on the variance of $f(\boldsymbol{v})$, and analogously for the higher order terms. The decomposition in (7) standardizes the sensitivity indices, that is, $\sum_{\ell=1}^{k} S_\ell + \sum_{1 \le \ell < m \le k} S_{\ell,m} + \ldots + S_{1,2,\ldots,k} = 1$.

Introduced by Homma and Satelli (1996), the *total sensitivity index*, S_ℓ^T, is a further related measure, defined by the sum of all the sensitivity indices involving input v_ℓ. Specifically,

$$S_\ell^T = 1 - \frac{W_{-\ell}}{W}, \quad \ell = 1, \ldots, k,$$

where $W_{-\ell} = \mathrm{Var}(\mathrm{E}(Y|\boldsymbol{v}_{-\ell}))$ is the total contribution to $\mathrm{Var}\left(f(\boldsymbol{v})\right)$ due to all inputs except v_ℓ. A large difference between S_ℓ and S_ℓ^T for the ℓ–th input indicates an important role of interaction terms involving that input on the variation in the output

The definition of the main effects and sensitivity indices involves expectations with respect

to the simulator output function $y = f(\boldsymbol{v})$. Therefore, if we approximate the output function by a GP model, we must account for this approximation by computing $\mathrm{E}^*\{\mathrm{E}(Y) \mid D\}$, $\mathrm{E}^*\{\mathrm{E}(Y|v_\ell) \mid D\}$, $\mathrm{E}^*\{S_\ell \mid D\}$, and $\mathrm{E}^*\{S_\ell^T \mid D\}$, where $\mathrm{E}^*\{\cdot \mid D\}$ indicates expectations with respect to the GP posterior predictive distribution, $p(\tilde{y} \mid D)$, developed in Section 2.1. As pointed out by Oakley and O'Hagan (2004), $\mathrm{E}^*\{S_\ell \mid D\}$, and $\mathrm{E}^*\{S_\ell^T \mid D\}$, which are posterior expectations of ratios or random variables, cannot be derived analytically. Instead, Oakley and O'Hagan (2004) obtain approximate point estimates for S_ℓ and S_ℓ^T by computing the ratio of expectations over $p(\tilde{y} \mid D)$, where $\mathrm{E}^*\{S_\ell \mid D\}$ is approximated by the ratio of $\mathrm{E}^*\{\mathrm{Var}(\mathrm{E}(Y|v_\ell)) \mid D\}$ and $\mathrm{E}^*\{\mathrm{Var}(Y) \mid D\}$, and analogously for $\mathrm{E}^*\{S_\ell^T \mid D\}$. The approximation of the sensitivity indices through ratios of expectations is also used in the likelihood approach of Morris et al. (2008).

2.2.1 Point estimates and uncertainty bands for the main effects

Here, we develop fully Bayesian point estimates for the main effects accompanied by a measure of posterior predictive uncertainty. The corresponding expressions result in relatively straightforward computing owing to the conditional normality structure of the GP emulator. The approach is similar to the one in Oakley and O'Hagan (2004), but extends their empirical Bayesian method based on fixed range parameters for the GP correlation function.

Given the generic input $\boldsymbol{v} = (v_1, \ldots, v_k)$, the distribution of the predicted emulator output, $\tilde{y} = f(\boldsymbol{v})$, is given by (2). In order to determine the main effect of input v_ℓ, we need to calculate $\mathrm{E}^*\{\mathrm{E}(Y) \mid D\}$ and $\mathrm{E}^*\{\mathrm{E}(Y|v_\ell) \mid D\}$. We assume independent components in the uncertainty distribution for the inputs, which, for simpler notation, are taken to be uniform over a normalized range of values in $(0, 1)$ for each input. Then, using (2)–(4), in Appendix B we obtain

$$\mathrm{E}^*\{\mathrm{E}(Y) \mid D\} = \int_{f(\boldsymbol{v})} \mathrm{E}(Y)\, p(f(\boldsymbol{v}) \mid D)\, df(\boldsymbol{v}) = \int_{\boldsymbol{\psi}} \left\{\mu + \boldsymbol{T}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)\right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi}.$$

Here, $\boldsymbol{T}$ is the $n \times 1$ vector with i–th element given by $\prod_{\ell=1}^k \left\{\int_0^1 \exp(-\phi_\ell |v_\ell - x_{\ell i}|^{a_\ell})\, dv_\ell\right\}$. Note that the elements of $\boldsymbol{T}$ can be computed analytically if $a_\ell = 1$, for $\ell = 1, \ldots, k$, which is the specification for the exponential correlation function. Under this specification, the i–th element of $\boldsymbol{T}$ is written as $\prod_{\ell=1}^k \left\{{\phi_\ell}^{-1}\left(2 - e^{-\phi_\ell x_{\ell i}} - e^{-\phi_\ell (1 - x_{\ell i})}\right)\right\}$.

For each specified value u_j of the j–th input,

$$\mathrm{E}(Y|u_j) = \int_{\{v_\ell : \ell \neq j\}} f(v_1, \ldots, u_j, \ldots, v_k) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell). \tag{8}$$

Again, using (2)–(4), in Appendix B we derive

$$\begin{aligned}\mathrm{E}^*\left\{\mathrm{E}\left(Y|u_j\right) \mid D\right\} &= \int_{f(v_1,\ldots,u_j,\ldots,v_k)} \mathrm{E}\left(Y|u_j\right) p\left(f(v_1,\ldots,u_j,\ldots,v_k) \mid D\right) df\left(v_1,\ldots,u_j,\ldots,v_k\right) \\ &= \int_{\boldsymbol{\psi}} \left\{\mu + \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right)\right\} p\left(\boldsymbol{\psi} \mid D\right) d\boldsymbol{\psi},\end{aligned}$$

where $\boldsymbol{T}_j(u_j)$ is the $n \times 1$ vector with i–th element given by

$$\exp\left(-\phi_j |u_j - x_{ji}|^{a_j}\right) \times \prod_{\{\ell:\ell\neq j\}} \left\{\int_0^1 \exp\left(-\phi_\ell |v_\ell - x_{\ell i}|^{a_\ell}\right) dv_\ell\right\}.$$

For a measure of (posterior predictive) uncertainty associated with the estimate of the main effects, we use $\mathrm{Var}^*\left\{\mathrm{E}\left(Y|u_j\right) - \mathrm{E}(Y) \mid D\right\}$, which is given by

$$\begin{aligned}&\mathrm{Var}^*\left\{\mathrm{E}(Y \mid u_j) \mid D\right\} + \mathrm{Var}^*\left\{\mathrm{E}(Y) \mid D\right\} - 2\mathrm{Cov}^*\left\{\mathrm{E}(Y \mid u_j), \mathrm{E}(Y) \mid D\right\} = \\ &\quad \mathrm{E}^*\left\{\left(\mathrm{E}(Y \mid u_j)\right)^2 \mid D\right\} - \left(\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j) \mid D\right\}\right)^2 + \mathrm{E}^*\left\{\left(\mathrm{E}(Y)\right)^2 \mid D\right\} - \left(\mathrm{E}^*\left\{\mathrm{E}(Y) \mid D\right\}\right)^2 \\ &\quad - 2\left(\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\} - \mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j) \mid D\right\}\mathrm{E}^*\left\{\mathrm{E}(Y) \mid D\right\}\right) \qquad (9)\end{aligned}$$

Because we already have the expressions for $\mathrm{E}^*\left\{\mathrm{E}\left(Y|u_j\right) \mid D\right\}$ and $\mathrm{E}^*\left\{\mathrm{E}\left(Y\right) \mid D\right\}$, what is needed is expressions for $\mathrm{E}^*\left\{\left(\mathrm{E}\left(Y|u_j\right)\right)^2 \mid D\right\}$, $\mathrm{E}^*\left\{\left(\mathrm{E}\left(Y\right)\right)^2 \mid D\right\}$, and $\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\}$. Extending the arguments in the derivation of $\mathrm{E}^*\left\{\mathrm{E}\left(Y\right) \mid D\right\}$ and $\mathrm{E}^*\left\{\mathrm{E}\left(Y|u_j\right) \mid D\right\}$, it can be shown (see Appendix B) that

$$\mathrm{E}^*\left\{\left(\mathrm{E}\left(Y|u_j\right)\right)^2 \mid D\right\} = \int_{\boldsymbol{\psi}} \left\{\tau^2\left(e - \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1} \boldsymbol{T}_j(u_j)\right) + \left(\mu + \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right)\right)^2\right\} p\left(\boldsymbol{\psi} \mid D\right) d\boldsymbol{\psi}$$

$$\mathrm{E}^*\left\{\left(\mathrm{E}(Y)\right)^2 \mid D\right\} = \int_{\boldsymbol{\psi}} \left\{\tau^2\left(g - \boldsymbol{T}^T R_{\boldsymbol{\phi}}^{-1} \boldsymbol{T}\right) + \left(\mu + \boldsymbol{T}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)\right)^2\right\} p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi} \qquad (10)$$

and

$$\begin{aligned}\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\} \;=\; &\int_{\boldsymbol{\psi}} \left\{\tau^2\left[e\phi_j^{-1}\left(2 - e^{-\phi_j u_j} - e^{-\phi_j(1-u_j)}\right) - \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1} \boldsymbol{T}\right] + \right. \\ &\left. \left(\mu + \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)\right)\left(\mu + \boldsymbol{T}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)\right)\right\} p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi} \qquad (11)\end{aligned}$$

where $e = \prod_{\{\ell:\ell\neq j\}} \left\{\int_0^1 \int_0^1 \exp\left(-\phi_\ell |v_\ell - v'_\ell|^{a_\ell}\right) dv_\ell dv'_\ell\right\}$, and $g = \prod_{\ell=1}^k \left\{\int_0^1 \int_0^1 \exp\left(-\phi_\ell \mid v_\ell - v'_\ell \mid^{a_\ell}\right) dv_\ell dv'_\ell\right\}$. Note that, again, e and g are available analytically under the exponential correlation function.

In particular, letting $a_\ell = 1$, for $\ell = 1, \ldots, k$, we obtain $e = \prod_{\{\ell:\ell\neq j\}} \left\{2\phi_\ell^{-2}\left(e^{-\phi_\ell} + \phi_\ell - 1\right)\right\}$, and $g = \prod_{\ell=1}^{k} \left\{2\phi_\ell^{-2}\left(e^{-\phi_\ell} + \phi_\ell - 1\right)\right\}$.

2.2.2 Full inference for the sensitivity indices

The approach of Section 2.2.1 cannot be extended to estimate the sensitivity indices. Instead of relying on approximate point estimates for S_ℓ and S_ℓ^T, $\ell = 1, ..., k$, we propose to sample from the posterior distributions for the sensitivity indices by computing at every MCMC sample of the GP emulator all the expectations needed for the definition of the S_ℓ and S_ℓ^T, that is, $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2$, $\mathrm{E}\left((\mathrm{E}(Y|u_j))^2\right)$, and $\mathrm{E}\left((\mathrm{E}(Y|\boldsymbol{u}_{-j}))^2\right)$. Letting $\{\boldsymbol{v} = (v_1, \ldots, v_k), y = f(\boldsymbol{v})\}$ be a generic run of the simulator, the expectation and variance of y are given by

$$\mathrm{E}(Y) = \int_{\boldsymbol{v}} f(\boldsymbol{v}) \prod_{\ell=1}^{k} dH_\ell(v_\ell) \quad \text{and} \quad \mathrm{Var}(Y) = \int_{\boldsymbol{v}} f^2(\boldsymbol{v}) \prod_{\ell=1}^{k} dH_\ell(v_\ell) \; - \; (\mathrm{E}(Y))^2 .$$

For a generic value u_j of the j–th input, squaring the expression for $\mathrm{E}(Y|u_j)$ in (8) and then taking its expectation, we obtain

$$\begin{aligned}
\mathrm{E}\left((\mathrm{E}(Y|u_j))^2\right) &= \int \left\{\int_{\{v_\ell:\ell\neq j\}} f(v_1, \ldots, u_j, \ldots, v_k) \prod_{\{\ell:\ell\neq j\}} dH_\ell(v_\ell)\right\}^2 dH_j(u_j) \\
&= \int \{\int_{\{v_\ell:\ell\neq j\}} \int_{\left\{v'_\ell:\ell\neq j\right\}} f(v_1, \ldots, u_j, \ldots, v_k)\, f(v'_1, \ldots, u_j, \ldots, v'_k) \\
&\qquad \prod_{\{\ell:\ell\neq j\}} dH_\ell(v_\ell) \prod_{\{\ell:\ell\neq j\}} dH_\ell(v'_\ell)\}\, dH_j(u_j) \\
&= \int_{\boldsymbol{v}} \int_{\left\{v'_\ell:\ell\neq j\right\}} f(\boldsymbol{v}) f(v'_1, \ldots, v_j, \ldots, v'_k) \prod_{\ell=1}^{k} dH_\ell(v_\ell) \prod_{\{\ell:\ell\neq j\}} dH_\ell(v'_\ell).
\end{aligned}$$

Regarding the expectations needed for the total sensitivity indices, let $\boldsymbol{u}_{-j} = (u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_k)$. Then, $\mathrm{E}(Y|\boldsymbol{u}_{-j}) = \int f(v_j, \boldsymbol{u}_{-j})\, dH_j(v_j)$, and, analogously to the derivation above,

$$\mathrm{E}\left((\mathrm{E}(Y|\boldsymbol{u}_{-j}))^2\right) = \int \int f(u_j, \boldsymbol{u}_{-j})\, f\left(v'_j, \boldsymbol{u}_{-j}\right) dH_j(v'_j) \prod_{\ell=1}^{k} dH_\ell(u_\ell).$$

At each MCMC posterior sample of the GP emulator, the posterior distributions for the first-order sensitivity index, S_j, and the total sensitivity index, S_j^T, are sampled by evaluating all the expectations that enter their definition,

$$S_j = \frac{\mathrm{Var}\left(\mathrm{E}(Y|u_j)\right)}{\mathrm{Var}(Y)} = \frac{\mathrm{E}\left((\mathrm{E}(Y|u_j))^2\right) - (\mathrm{E}(Y))^2}{\mathrm{Var}(Y)} \tag{12}$$

$$S_j^T = \frac{\mathrm{Var}(Y) - \mathrm{Var}\left(\mathrm{E}(Y|\boldsymbol{u}_{-j})\right)}{\mathrm{Var}(Y)} = 1 - \frac{\mathrm{E}\left((\mathrm{E}(Y|\boldsymbol{u}_{-j}))^2\right) - (\mathrm{E}(Y))^2}{\mathrm{Var}(Y)}. \tag{13}$$

The computing involves Monte Carlo integration based on samples from the uncertainty distribution, in the spirit of techniques from Saltelli (2002), but extending the approach to account for the GP approximation to the computer simulator output.

Specifically, we begin by generating input sample matrix $\mathbf{M}$ of size $B \times k$,

$$\mathbf{M} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,k} \\ \vdots & \vdots & \dots & \vdots \\ v_{B,1} & v_{B,2} & \dots & v_{B,k} \end{pmatrix}$$

where each row of $\mathbf{M}$ is drawn independently from the uncertainty distribution over the simulator inputs, $H(\boldsymbol{v}) = \prod_{\ell=1}^{k} H_\ell(v_\ell)$.

Next, we generate k input sample matrices, $\mathbf{N}_j$, for $j = 1, \dots, k$, of size $B \times k$ each,

$$\mathbf{N}_j = \begin{pmatrix} v'_{1,1} & v'_{1,2} & \dots & v_{1,j} & \dots & v'_{1,k} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ v'_{B,1} & v'_{B,2} & \dots & v_{B,j} & \dots & v'_{B,k} \end{pmatrix}$$

where the j–th column of matrix $\mathbf{N}_j$ equals the j–th column of matrix $\mathbf{M}$, but the remaining elements of each row of $\mathbf{N}_j$ form independent random samples from the corresponding marginal of the uncertainty distribution, $\prod_{\{\ell:\ell\neq j\}} H_\ell(v_\ell)$.

Finally, we generate k input sample matrices, $\mathbf{N}_{-j}$, for $j = 1, \dots, k$, of size $B \times k$ each,

$$\mathbf{N}_{-j} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v'_{1,j} & \dots & v_{1,k} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ v_{B,1} & v_{B,2} & \dots & v'_{B,j} & \dots & v_{B,k} \end{pmatrix}$$

where matrices $\mathbf{N}_{-j}$ and $\mathbf{M}$ have all columns in common except the j–th one; the $v'_{b,j}$, $b = 1, ..., B$, are randomly sampled from $H_j(v_j)$.

Now, the Monte Carlo simulation, based on the posterior samples from the GP emulator, proceeds as follows:

- For each MCMC posterior sample for $\boldsymbol{\psi} = (\mu, \tau^2, \boldsymbol{\phi})$, obtain the following posterior predictive samples according to (2): for each row b of M, sample $\tilde{y}_b$, then compute $\tilde{y}_b^2$; for each row b of N_j, sample $\tilde{y}'_{b,j}$; and for each row b of N_{-j}, sample $\tilde{y}'_{b,-j}$.

- Obtain the posterior sample for $\mathrm{E}(Y)$ and $\mathrm{E}(Y^2)$ by computing $B^{-1}\sum_{b=1}^{B} \tilde{y}_b$ and $B^{-1}\sum_{b=1}^{B} \tilde{y}_b^2$, respectively.

- For $j = 1, \ldots, k$, obtain the posterior sample for $\mathrm{E}\left((\mathrm{E}(Y|u_j))^2\right)$ and $\mathrm{E}\left((\mathrm{E}(Y|u_{-j}))^2\right)$ through $B^{-1}\sum_{b=1}^{B}\tilde{y}_b\tilde{y}'_{b,j}$ and $B^{-1}\sum_{b=1}^{B}\tilde{y}_b\tilde{y}'_{b,-j}$, respectively.

- Compute the posterior realizations for the first-order sensitivity indices, S_j, and the total sensitivity indices, S_j^T, by evaluating expressions (12) and (13), respectively, using the posterior samples above for the required expectations.

Because the posterior samples for the variances in expressions (12) and (13) are evaluated through differences of expectations, negative values can arise for the S_j and/or S_j^T posterior realizations. This issue can be overcome by appropriate choice of the Monte Carlo sample size B. For instance, for the results reported in Section 3.2, we used $B = 250,000$; with the exception of the occurrence of negative values for sensitivity indices supported by values close to 0, the estimated posterior distributions in Figure 3 were similar under $B = 25,000$.

This approach enables estimation of the entire distribution for each sensitivity index allowing for the uncertainty of the sensitivity indices to be determined. While repeated samples from the GP predictive distribution are required for computing (12) and (13), these are computationally inexpensive emulator runs, which are substantially faster than those obtained using the computer simulator. The computational burden imposed by the large Monte Carlo sample size, required as discussed above, can be relaxed through a simple parallel implementation of the method, using a number of nodes over blocks of the MCMC posterior samples for the GP emulator. Nevertheless, the methodology remains computationally feasible for computer simulators with small to moderate number of inputs. An additional challenge for simulators with large number of inputs involves the GP emulation approach, which is restricted by the inversion of high-dimensional covariance matrices associated with the larger number of training simulator runs that are needed for high-dimensional input spaces.

Finally, note that the method discussed in this section can, in principle, be also applied to obtain the posterior distribution for the main effects. However, the approach requires for each input a (relatively large) number of sample matrices (of size $B \times k$ each) over a (relatively fine) grid in the input space; each of these matrices can be used to obtain the posterior sample for $\mathrm{E}(Y|v_j = w_q)$, where w_q is the q–th grid point for the j–th input. Hence, in practice, the approach becomes prohibitively computationally expensive even for moderate dimensions for the input space. The approach of Section 2.2.1 offers a practically feasible alternative to point and interval estimation for the main effects. At the same time, we note that the method of this section is entirely generic with regard to the statistical model emulator utilized for the computer simulator, whereas the approach of Section 2.2.1 is specific to GP emulators.

3 Application

Measurements from Earth observing satellites are used to produce a wide array of data products, e.g., sea surface temperature, polar ice coverage, and plant type. Such data products are used in a wide variety of further scientific studies, and also as inputs to important policy decisions, especially those concerning the impact of human activity on the biosphere. Many of the data products are produced by inverting a Radiative Transfer Model (RTM). RTMs are implemented in complex computer programs to simulate light reflected off the surface of the Earth and its propagation through the atmosphere as a function of biospheric parameters, e.g., land cover type, available water, leaf chemistry. The upwelling radiation at the top of the atmosphere simulated by an RTM is also observed by the satellite, and thus RTMs can be used in conjunction with satellite measurements to invert for key data products (inputs to the RTM) in the study of biosphere and ecosystem dynamics. Hence, RTMs are widely used in geoscience and remote sensing for the prediction of the properties of Earth's coupled dynamical system.

While RTMs are deterministic computer simulators, there is uncertainty about the values of their inputs. Here, we study the impact of this uncertainty on the computed output, using the methods presented in Section 2. Specifically, we work with the Leaf-Canopy Model (LCM) (Ganapol et al., 1999), a particular RTM which simulates light reflected by vegetation. In Section 3.1, we provide a description of the LCM simulator, and in Section 3.2, we study the sensitivity of the LCM output to uncertainty in its inputs.

3.1 Leaf-Canopy Model

The LCM was developed by the Vegetation Modeling Transport Group (University of Arizona), in collaboration with the Ecosystem Science and Technology Branch at NASA Ames in support of MODIS, a key instrument aboard Terra and Aqua satellites. In particular, the LCM was developed to capture the essential biophysical processes associated with the interaction between light and vegetation. The LCM simulator combines two different radiative transfer algorithms: LEAFMOD, which simulates the radiative regime inside the single leaf, and CANMOD, which combines the information coming from LEAFMOD with canopy structural parameters to compute the radiative regime within and at the top of the canopy.

LEAFMOD is run in the forward and inverse modes to compute the leaf optical properties. In the forward mode, it uses the leaf's thickness, scattering profile, and absorption profile to calculate hemispherical reflectance and transmittance and the directional distribution of the radiance exiting the leaf surface. The leaf absorption profile is constructed from biochemical

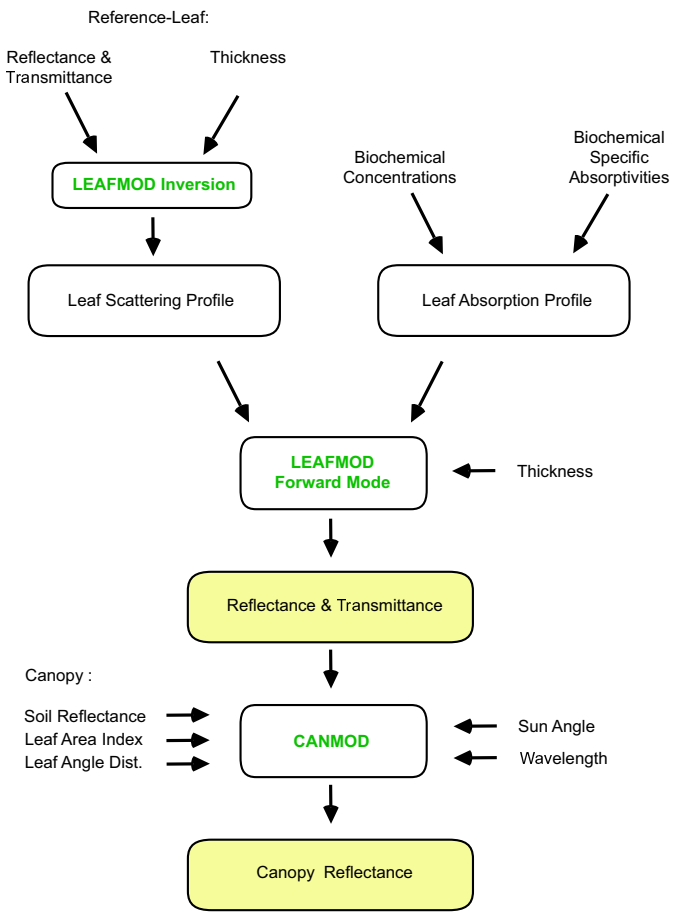

Figure 1: LCM simulator flowchart. First, a leaf type is specified, and its absorption profile is constructed based on its biochemical components. Next, the leaf scattering profile is determined by executing LEAFMOD in the inverse mode. Given the constructed scattering and absorption profiles of the leaf and its thickness, LEAFMOD is run in the forward mode to compute the leaf reflectance and transmittance properties. Then, the output of LEAFMOD is fed to CANMOD together with LAI, LAD, soil reflectance, the sun angle, and the wavelength to compute the canopy hemispherical reflectance coefficient.

concentrations, and absorptivity properties of chlorophyll and carotenoids, protein, lignin and cellulose, and water (Ganapol et al., 1998). In the inverse mode, LEAFMOD uses leaf thickness and spectral measurements from the LOPEX leaf database to determine the scattering profile of the leaf. The LOPEX leaf species archive stores experimentally obtained spectral properties for many common species (Hosgood et al., 1995)

The CANMOD algorithm combines the leaf spectral information coming from LEAFMOD with Leaf Area Index (LAI), leaf angle distribution (LAD), soil reflectance, and sun angle, and computes the radiative regime, at any given wavelength (between 400 and 2100 nm), within and at the top of the canopy. LAI is the area of the leaves on a canopy divided by the area of the ground covered by the canopy, and is thus a dimensionless quantity. LAD describes the

Input	Min	Max
LAI	0	8
Chlorophyll ($\mu g/cm^2$)	0	100
Water fraction	0.1	0.8
Protein (mg/cm^2)	0.1	1
Lignin/Cellulose (mg/cm^2)	0.1	6
Thickness (cm)	0.01	0.1
Soil reflectance	0.3	1.3

Table 1: Ranges of inputs to the LCM. LAI, water fraction, and soil reflectance parameters are dimensionless.

band #	wavelength (nm)	MODIS band
1	469	ref3
2	555	ref4
3	1240	ref5
4	1640	ref6
5	2130	ref7
6	667	ref13
7	748	ref15
8	870	ref16

Table 2: Wavelength for each band used and the corresponding MODIS band number. Bands are in the MODIS band order, not in the wavelength order.

orientation of the leaves and it takes 5 discrete values: planophile (leaves mainly horizontal), erectophile (leaves mainly vertical), plagiophile (leaves mainly at 45 degrees), extremophile (leaves mainly both horizontal and vertical), unophile (leaves mainly spherical).

Figure 1 shows a flowchart for the operation of the LCM simulator, including details on its implementation steps. From the two coupled algorithms, the LCM inputs include leaf chemistry variables (chlorophyll, water fraction, lignin/cellulose, and protein), leaf thickness, soil reflectance, canopy architecture (LAI and LAD), wavelength, and sun angle. In our analysis, the LAD variable is set to planophile, and the sun angle is set to zenith. Table 1 lists the LCM inputs and their ranges, and Table 2 includes the 8 bands (or groups of wavelengths) used by the LCM along with their corresponding MODIS band numbers.

In the next section, we perform sensitivity analysis of the LCM in order to identify inputs that are the main contributors to variability in the computed canopy reflectance. Such analysis is an important model development tool, since it provides guidance as to where better input information should be obtained in order to reduce variability in the output. Additionally, identification of the more influential inputs gives information as to how well these inputs can be estimated by combining LCM output at different wavelengths and remote sensed measurements. Of particular interest is estimation of LAI, which is a key parameter used in climate and ecological models that quantify the exchange of fluxes of energy, mass, and momentum between the land surface and the atmosphere (e.g., Houborg et al., 2007).

3.2 Sensitivity Analysis Results for the LCM Simulator

We apply the Bayesian approach to the GP emulator using a training set of 250 LCM runs based on a Latin Hypercube design at each of the 8 MODIS bands (see Table 2). We use the

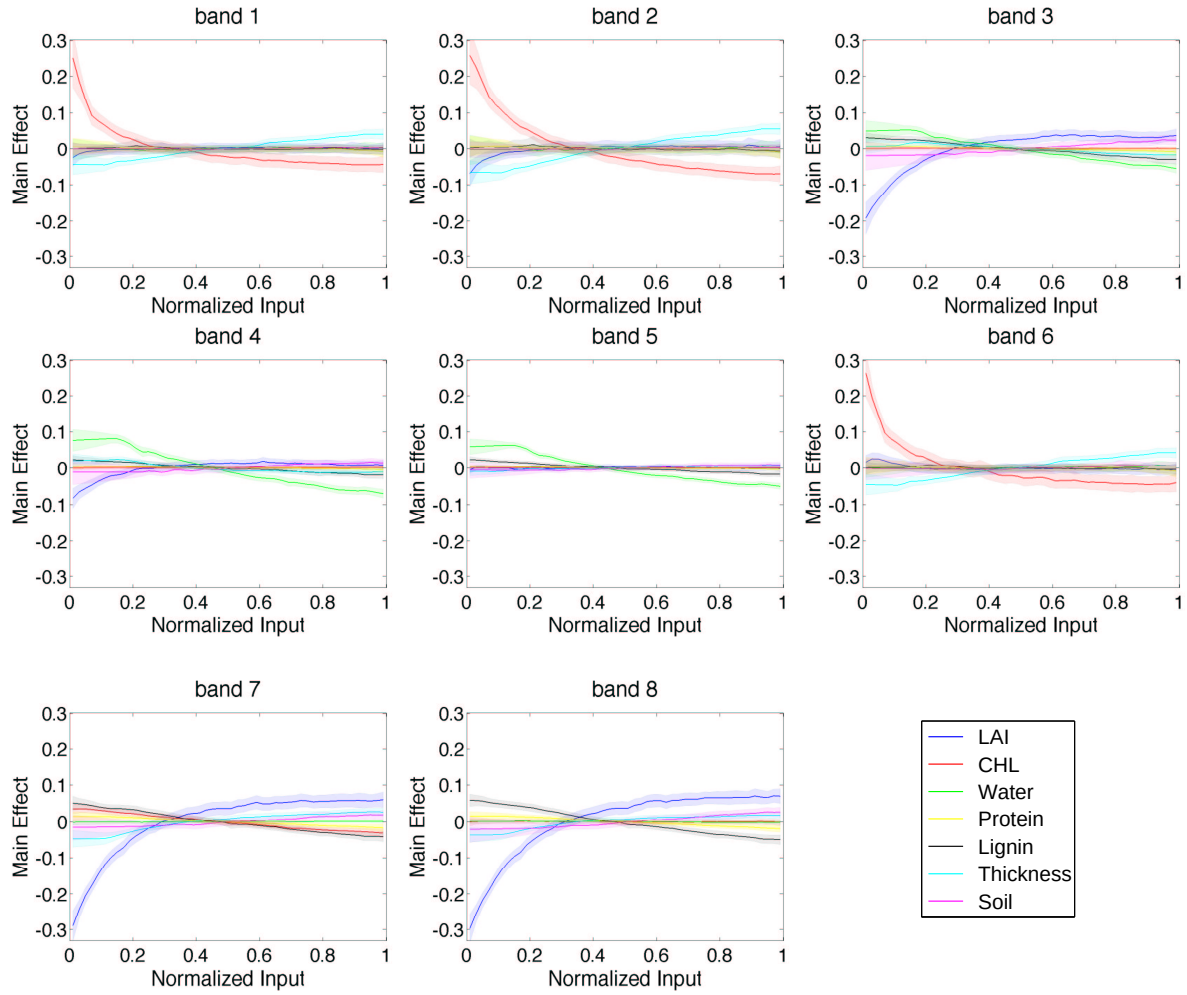

Figure 2: Posterior point estimates ± 2 standard deviations of the main effects for the LCM simulator at 8 MODIS bands.

exponential correlation function, setting in (1) $a_\ell = 1$, for $\ell = 1, \ldots, k = 7$ (a jitter term was not needed). We place a normal prior on μ, an inverse-gamma prior on τ^2, and a Unif$(0, b_{\phi_\ell})$ prior on each ϕ_ℓ, $\ell = 1, \ldots, 7$, assuming prior independence for all hyperparameters. Details on prior specification as well as MCMC posterior simulation for the GP model parameters are provided in Appendix A. We have also experimented with gamma priors of varying dispersion for each ϕ_ℓ, which resulted in nearly identical posteriors. For the uncertainty distribution, we assume independent uniform components over the ranges given in Table 1 for each input variable.

Figure 2 shows plots of the main effects for the 7 normalized input variables and their uncertainty intervals for each of the 8 MODIS bands, based on the approach of Section 2.2.1. The interval estimates are computed by adding and subtracting two standard deviations, calculated from the variance point estimate given in (9), to the posterior mean estimate $\mathrm{E}^* \{\mathrm{E}(Y|u_j) - \mathrm{E}(Y) \mid D\}$.

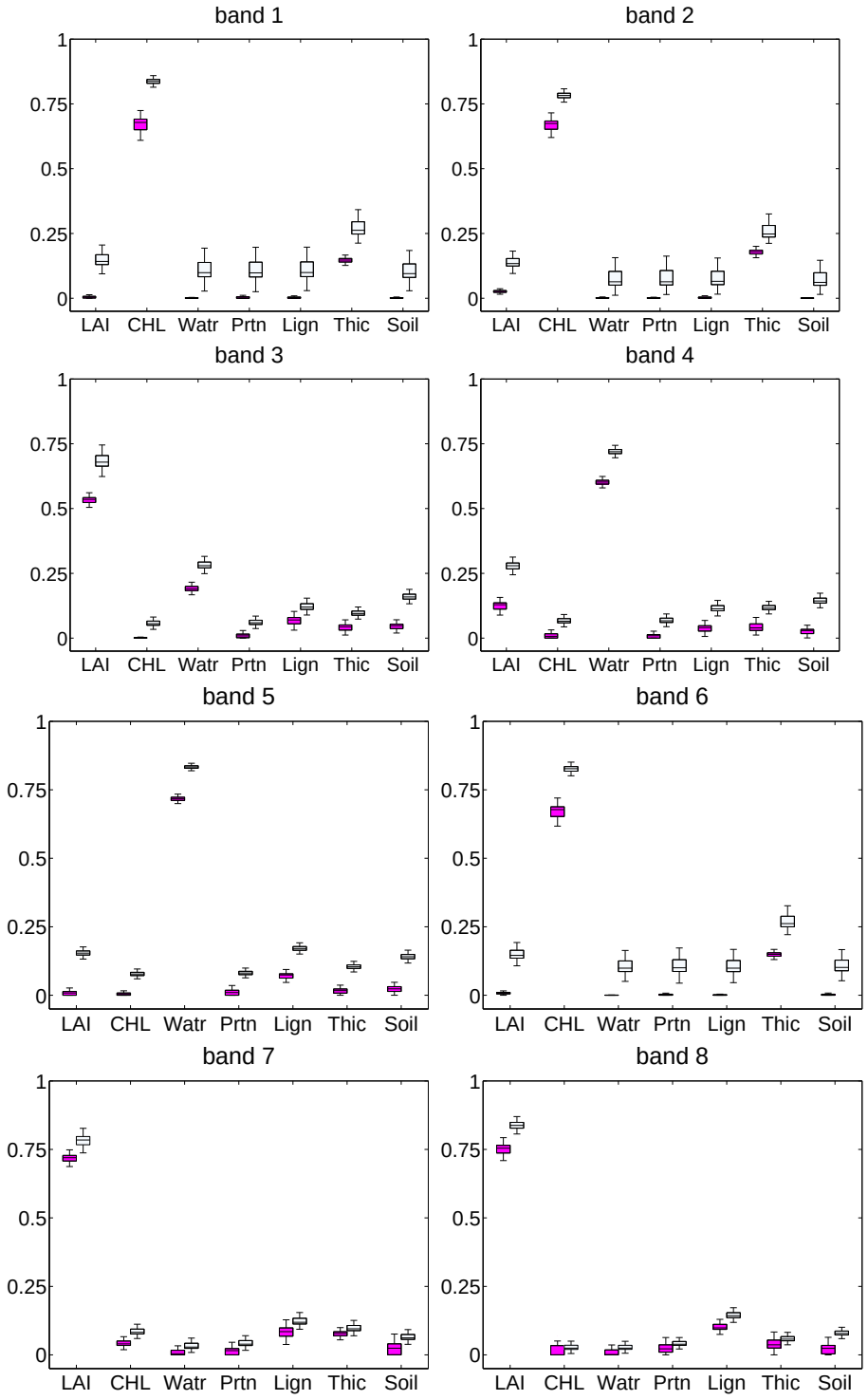

Figure 3: Box plots of the posterior distributions of the first-order and total sensitivity indices for the LCM inputs at 8 MODIS bands (first-order indices are in magenta). The horizontal line inside each box indicates the median; the edges of the box correspond to the first and third quartile; and the whiskers stretch to horizontal lines outside the box that cover the central 95% of the probability mass for each distribution (outliers have been removed to avoid cluttering the plots).

Normalizing the range of values of the inputs to the unit interval allows all the main effects to be plotted together on the same plot. In general, the larger the variation of the main effect plot, the greater the influence of that input on the LCM output. For visible spectrum (bands 1, 2 and 6), the LCM is most sensitive to chlorophyll, where an increase in chlorophyll produces a decrease in the LCM output. For near infrared (bands 3, 7 and 8), the LCM is most sensitive to LAI, where an increase in LAI produces an increase in the LCM output. Finally, for short infrared bands (bands 4 and 5), the chlorophyll effect is diminished, while LAI and water are dominant for band 4, and water becomes more influential for band 5. In general, we observe that all dominant inputs have non-linear main effects. Figure 3 shows boxplots summarizing posterior inference for the first-order and total sensitivity indices, obtained using the method of Section 2.2.2. The results indicate that inputs with influential main effects are also major contributors to the variation in the LCM, i.e., they have large sensitivity indices.

We also study the extent to which the proposed methodology improves on methods that do not fully incorporate uncertainty by fixing some of the GP emulator parameters and/or using ad-hoc point estimates for the sensitivity indices. First, we fix the GP range of dependence parameters, ϕ_ℓ, $\ell = 1, \ldots, 7$, to their posterior modes (obtained from the fully Bayesian fit of the GP emulator) and apply the approach of Section 2.2.1 to obtain point and interval estimates for the LCM main effects (not shown). The point estimates were similar to the ones in Figure 2. However, there was a decrease in the width of the uncertainty intervals, which was particularly noticeable at the regions of steeper slope for the dominant inputs main effects. Next, we compute approximate point estimates for first-order and total sensitivity indices. Specifically, for the S_j, we work again with fixed parameters ϕ_ℓ, and employ the Monte Carlo simulation method of Section 2.2.2 to obtain samples for the numerator and denominator in expression (12). From these samples, we compute point estimates for $\mathrm{Var}\left(\mathrm{E}\left(Y|u_j\right)\right)$ and $\mathrm{Var}(Y)$, which are then divided to produce the approximate point estimates for the first-order sensitivity indices. The approach is analogous for the total sensitivity indices. Although implemented in a different fashion, this is the essence of the method in Oakley and O'Hagan (2004). Table 3 reports representative results for three MODIS bands, one each from the visible spectrum, the near infrared, and the short infrared (bands 1, 3, and 4, respectively). Contrasting these results with Figure 3, we note that the approximate point estimates for band 1 successfully identify the inputs with the largest first-order sensitivity indices; this was also the case for bands 2, 5, and 6, and, to a smaller extent, bands 7 and 8. However, for band 4 and, more detrimentally, for band 3 qualitatively different conclusions would be drawn from Table 3. The shortcomings of this approach are exacerbated

	band 1		band 3		band 4	
Input	first-order	total	first-order	total	first-order	total
LAI	0.008	0.012	0.169	0.184	0.024	0.052
Chlorophyll	0.680	0.699	0.022	0.029	0.037	0.041
Water fraction	0.016	0.026	0.338	0.346	0.740	0.799
Protein	0.035	0.041	0.059	0.073	0.021	0.033
Lignin/Cellulose	0.052	0.058	0.225	0.323	0.084	0.097
Thickness	0.210	0.233	0.076	0.103	0.088	0.099
Soil reflectance	0.033	0.046	0.000	0.021	0.108	0.111

Table 3: Approximate point estimates of the first-order and total sensitivity indices for the LCM inputs at three of the MODIS bands (see Section 3.2 for details).

by the total sensitivity indices estimates, where essentially for all bands, the difference from the first-order indices for the dominant inputs is significantly underestimated.

The results in Figure 3 suggest that many inputs with negligible first-order sensitivity indices have non-negligible total sensitivity indices. A substantial difference between S_j and S_j^T of the j–th input indicates an important role of interaction terms involving that input on the variation in the output. The approach of Section 2.2.2 to full inference for S_j and S_j^T enables formal probabilistic statements for the differences $S_j^T - S_j$, and thus a more informative approach to sensitivity analysis than comparison of point estimates for the sensitivity indices.

4 Discussion

We have presented a framework for Bayesian global sensitivity analysis of deterministic simulators. The basis of the methodology is statistical model approximation (emulation) of the simulator output, which is built from Gaussian process (GP) priors. We have discussed an approach to full inference for sensitivity indices of the simulator inputs. The approach combines draws from the posterior predictive distribution of the GP emulator and Monte Carlo samples from the input uncertainty distribution to obtain samples from the posterior distributions of the sensitivity indices. We have also derived expressions which enable ready computing of point estimates and standard errors for the main effects of the simulator inputs. The methodology has been applied to the Leaf-Canopy Model (LCM), a radiative transfer model for the interaction of sunlight with vegetation, to identify the most influential inputs at different spectral bands.

We used the standard specification in the GP emulation literature based on the product power exponential correlation function, and thus we are bound by the assumptions of isotropic covariance and variance homogeneity. Although for the LCM simulator this specification resulted in flexible interpolation, the use of a non-stationary GP prior model might be more appropriate

in other applications. We found that in applying the inference method for sensitivity indices, one must monitor the resulting posterior distributions under different Monte Carlo sample sizes until results stabilize, which took longer for some LCM bands than others. The methodology can be extended to estimate main effects and sensitivity indices associated with first and higher order interactions. However, this extension comes at an increasing computational cost, and thus a practical strategy may be to investigate terms that include inputs with large total sensitivity indices. Finally, an interesting direction for future work involves the emulation of a *dynamic* implementation of the LCM which takes into account seasonal effects.

ACKNOWLEDGMENTS

This research is part of the Ph.D. dissertation of Marian Farah, completed at University of California, Santa Cruz, and was supported in part by NASA AISR program grant NNX07AV69G, and by an award from the NASA UARC Aligned Research Program. The authors thank Robin Morris, Matt Taddy, and Roberto Furfaro for helpful discussions, and Mark Riehl for the LCM flowchart in Figure 1. They also wish to thank the Editor, Hugh Chipman, an Associate Editor, and two referees for several useful comments and suggestions.

APPENDIX A

Posterior inference for the GP emulator: Following the GP model formulation in Section 2.1, the posterior distribution is $p(\mu, \tau^2, \boldsymbol{\phi} \mid D) \propto N_n\left(\boldsymbol{y}|\mu\mathbf{1}_n, \tau^2 R_{\boldsymbol{\phi}}\right) p(\mu)p(\tau^2)p(\phi_1)...p(\phi_k)$. We place independent priors on the GP parameters, specifically, we use a $N(a_\mu, b_\mu)$ prior for μ, an $\Gamma^{-1}(a_\tau, b_\tau)$ prior for τ^2, and Unif$(0, b_{\phi_\ell})$ priors for ϕ_ℓ, $\ell = 1, \ldots, k$. Here, $\Gamma^{-1}(a, b)$ denotes the inverse-gamma distribution with mean $b/(a-1)$, provided $a > 1$.

Posterior simulation from $p(\mu, \tau^2, \boldsymbol{\phi} \mid D)$ proceeds via Gibbs sampling. The full conditional posterior distribution for μ is normal with variance $S = \left(\tau^{-2}\mathbf{1}_n^T R_{\boldsymbol{\phi}}^{-1}\mathbf{1}_n + b_\mu^{-1}\right)^{-1}$, and mean $M = S\left(\tau^{-2}\mathbf{1}_n^T R_{\boldsymbol{\phi}}^{-1}\boldsymbol{y} + a_\mu b_\mu^{-1}\right)$. For τ^2, the posterior full conditional is $\Gamma^{-1}(A, B)$, with $A = a_\tau + 0.5n$ and $B = b_\tau + 0.5\left(\boldsymbol{y} - \mu\mathbf{1}_n\right)^T R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu\mathbf{1}_n\right)$. The posterior full conditional for each ϕ_ℓ, $\ell = 1, \ldots, k$, is proportional to $|R_{\boldsymbol{\phi}}|^{-\frac{1}{2}} \exp(-0.5\tau^{-2}\left(\boldsymbol{y} - \mu\mathbf{1}_n\right)^T R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu\mathbf{1}_n\right)) \times 1_{(0, b_{\phi_\ell})}(\phi_\ell)$, which cannot be sampled directly. We use Metropolis-Hastings steps for each ϕ_ℓ based on a right-truncated exponential proposal distribution with density $d_\ell \exp(-d_\ell\phi_\ell)/\{1 - \exp(-d_\ell b_{\phi_\ell})\}$. To choose the rate parameter d_ℓ (which is the only tuning parameter), we obtain an estimate $\tilde{\phi}_\ell$ of ϕ_ℓ (e.g., the MLE), set $\tilde{\phi}_\ell$ equal to the median of the proposal distribution, and solve for d_ℓ.

Prior specification: We set $a_\tau = 2$, a value that yields infinite variance for the corresponding inverse-gamma prior. To specify the hyperparameters a_μ, b_μ and b_τ, note that for each i, $\mathrm{E}(Y_i) = \mathrm{E}\left(\mathrm{E}\left(Y_i|\mu\right)\right) = \mathrm{E}(\mu) = a_\mu$, and $\mathrm{Var}(Y_i) = \mathrm{E}\left(\mathrm{Var}\left(Y_i|\tau^2\right)\right) + \mathrm{Var}\left(\mathrm{E}\left(Y_i|\mu\right)\right) = b_\tau + b_\mu$. Now, assume we have a prior guess for the center, c_y, and range, r_y, of the simulator output values. (For instance, such information is readily available for our application to the LCM simulator.) Then, we set $a_\mu = c_y$, and $b_\tau = b_\mu \approx (r_y/4)^2$, using $2(r_y/4)^2 \approx b_\tau + b_\mu$, with the extra inflation factor 2, and splitting the variance estimate equally between b_τ and b_μ.

Specifying prior information for the ϕ_ℓ is more difficult. One way to specify b_{ϕ_ℓ} is based on the interpretation of ϕ_ℓ under the correlation function in (1): for any fixed α_ℓ, it controls how fast the correlation decays with distance in the direction of the ℓ–th input x_ℓ. In particular, for $\alpha_\ell = 1$, $3/\phi_\ell$ is the "range of dependence", i.e., the value of the distance $d = |x_\ell - x'_\ell|$ that yields correlation approximately 0.05. Hence, we could use, say, $0.1d_{max}$, where $d_{max} = \max|x_\ell - x'_\ell|$, as a rough guess at $3/\phi_\ell$ and specify b_{ϕ_ℓ} from $0.1d_{max} = 3/b_{\phi_\ell}$. For the application to the LCM simulator, we used the available range for each input variable to specify d_{max} and thus b_{ϕ_ℓ}. The resulting uniform priors for the ϕ_ℓ led to a significant amount of prior to posterior learning.

APPENDIX B

Here, we provide details for the derivation of the expressions for $\mathrm{E}^*\left\{\mathrm{E}\left(Y\right) \mid D\right\}$, $\mathrm{E}^*\left\{\mathrm{E}\left(Y|u_j\right) \mid D\right\}$, $\mathrm{E}^*\left\{\left(\mathrm{E}(Y|u_j)\right)^2 \mid D\right\}$, $\mathrm{E}^*\left\{\left(\mathrm{E}\left(Y\right)\right)^2 \mid D\right\}$, and $\mathrm{E}^*\left\{\mathrm{E}\left(Y|u_j\right)\mathrm{E}\left(Y\right) \mid D\right\}$, which form the basis of the approach to estimation of the main effects discussed in Section 2.2.1.

Regarding the posterior point estimate for the global mean, we obtain

$$\begin{aligned}
\mathrm{E}^*\left\{\mathrm{E}\left(Y\right) \mid D\right\} &= \int_{f(\boldsymbol{v})} \mathrm{E}\left(Y\right) p\left(f(\boldsymbol{v}) \mid D\right) df(\boldsymbol{v}) \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\boldsymbol{v}} \left\{ \int_{f(\boldsymbol{v})} f\left(\boldsymbol{v}\right) p\left(f(\boldsymbol{v})|\boldsymbol{\psi}\right) df(\boldsymbol{v}) \right\} dH\left(\boldsymbol{v}\right) \right\} p\left(\boldsymbol{\psi} \mid D\right)\, d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\boldsymbol{v}} m(\boldsymbol{v}) \prod_{\ell=1}^{k} dH_\ell\left(v_\ell\right) \right\} p\left(\boldsymbol{\psi} \mid D\right) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\boldsymbol{v}} \left\{ \mu + \boldsymbol{r}^T(\boldsymbol{v}) R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right) \right\} \prod_{\ell=1}^{k} dH_\ell\left(v_\ell\right) \right\} p\left(\boldsymbol{\psi} \mid D\right) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \mu + \boldsymbol{T}^T R_{\boldsymbol{\phi}}^{-1}\left(\boldsymbol{y} - \mu \mathbf{1}_n\right) \right\} p\left(\boldsymbol{\psi} \mid D\right) d\boldsymbol{\psi},
\end{aligned}$$

where $\boldsymbol{T}$ is the $(n \times 1)$ vector with i–th element $\prod_{\ell=1}^{k} \left\{ \int_0^1 \exp\left(-\phi_\ell |v_\ell - x_{\ell i}|^{a_\ell}\right) dv_\ell \right\}$.

Turning to the posterior point estimate for $\mathrm{E}(Y|u_j)$, for any specified value u_j of the j–th input, let $\boldsymbol{v}_j = (v_1, \ldots, u_j, \ldots, v_k)$ and $f(\boldsymbol{v}_j) = f(v_1, \ldots, u_j, \ldots, v_k)$. Then, we can derive

$$\begin{aligned}
\mathrm{E}^*\left\{\mathrm{E}(Y|u_j) \mid D\right\} &= \int_{f(\boldsymbol{v}_j)} \mathrm{E}(Y|u_j)\, p\left(f(\boldsymbol{v}_j) \mid D\right) df(\boldsymbol{v}_j) \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\{v_\ell : \ell \neq j\}} \left\{ \int_{f(\boldsymbol{v}_j)} f(\boldsymbol{v}_j)\, p\left(f(\boldsymbol{v}_j)|\boldsymbol{\psi}\right) df(\boldsymbol{v}_j) \right\} \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) \right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \int_{\{v_\ell : \ell \neq j\}} m(\boldsymbol{v}_j) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) \right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \mu + \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1} (\boldsymbol{y} - \mu \mathbf{1}_n) \right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi},
\end{aligned}$$

with $\boldsymbol{T}_j(u_j)$ an $(n \times 1)$ vector with i–th element $\exp\left(-\phi_j |u_j - x_{ji}|^{a_j}\right) \prod_{\{\ell : \ell \neq j\}} \left\{ \int_0^1 \exp\left(-\phi_\ell |v_\ell - x_{\ell i}|^{a_\ell}\right) dv_\ell \right\}$.

Next, we derive the expression for $\mathrm{E}^*\left\{(\mathrm{E}(Y|u_j))^2 \mid D\right\}$. Denote $\boldsymbol{v}_j$ as before and $\boldsymbol{v'}_j = (v'_1, \ldots, u_j, \ldots, v'_k)$. Then,

$$\left(\mathrm{E}(Y|u_j)\right)^2 = \left(\int_{\{v_\ell : \ell \neq j\}} f(\boldsymbol{v}_j) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) \right)^2 = \iint_{\substack{\{v_\ell : \ell \neq j\} \\ \{v'_\ell : \ell \neq j\}}} f(\boldsymbol{v}_j) f(\boldsymbol{v'}_j) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) dH_\ell(v'_\ell).$$

Therefore, taking expectation with respect to the bivariate posterior predictive distribution for $(f(\boldsymbol{v}_j), f(\boldsymbol{v'}_j))$, developed in Section 2.1, we obtain

$$\begin{aligned}
\mathrm{E}^*\left\{ \left(\mathrm{E}(Y|u_j)\right)^2 \mid D \right\} &= \int \left(\mathrm{E}(Y|u_j)\right)^2 p(f(\boldsymbol{v}_j), f(\boldsymbol{v'}_j) \mid D) df(\boldsymbol{v}_j) df(\boldsymbol{v'}_j) \\
&= \int_{\boldsymbol{\psi}} \left\{ \iint_{\substack{\{v_\ell : \ell \neq j\} \\ \{v'_\ell : \ell \neq j\}}} \left\{ \int f(\boldsymbol{v}_j)\, f\left(\boldsymbol{v'}_j\right) p\left(f(\boldsymbol{v}_j), f(\boldsymbol{v'}_j) \mid \boldsymbol{\psi}\right) df(\boldsymbol{v}_j) df(\boldsymbol{v'}_j) \right\} \right. \\
&\qquad \left. \times \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) dH_\ell(v'_\ell) \right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \left\{ \iint_{\substack{\{v_\ell : \ell \neq j\} \\ \{v'_\ell : \ell \neq j\}}} \mathrm{E}\left(f(\boldsymbol{v}_j) f(\boldsymbol{v'}_j) \mid \boldsymbol{\psi}\right) \prod_{\{\ell : \ell \neq j\}} dH_\ell(v_\ell) dH_\ell(v'_\ell) \right\} p(\boldsymbol{\psi} \mid D)\, d\boldsymbol{\psi}.
\end{aligned}$$

Using the standard covariance identity, we obtain

$$\mathrm{E}\left(f(\boldsymbol{v}_j) f(\boldsymbol{v'}_j) \mid \boldsymbol{\psi}\right) = \mathrm{Cov}\left(f(\boldsymbol{v}_j), f(\boldsymbol{v'}_j) \mid \boldsymbol{\psi}\right) + \mathrm{E}\left(f(\boldsymbol{v}_j) \mid \boldsymbol{\psi}\right) \mathrm{E}\left(f(\boldsymbol{v'}_j) \mid \boldsymbol{\psi}\right) \tag{14}$$

where the expectation and covariance terms are taken over the conditional bivariate normal distribution for $(f(\boldsymbol{v}_j), f(\boldsymbol{v'}_j)) \mid \boldsymbol{\psi}, D$ with mean vector and covariance matrix given by (5) and (6), respectively. Denote by $\boldsymbol{R_1} \equiv \boldsymbol{R}_1(v_1, \ldots, u_j, \ldots, v_k)$ and $\boldsymbol{R_2} \equiv \boldsymbol{R_2}(v'_1, \ldots, u_j, \ldots, v'_k)$ the first and second columns, respectively, of the $(n \times 2)$ matrix $R(\boldsymbol{v}, \boldsymbol{v'})$ in (6). Note that here the input vectors, $(v_1, \ldots, u_j, \ldots, v_k)$ and $(v'_1, \ldots, u_j, \ldots, v'_k)$, have common element u_j. Therefore, $\boldsymbol{R}_1$ is the $(n \times 1)$ vector with elements $\exp\left(-\phi_j|u_j - x_{ji}|^{a_j} - \sum_{\{\ell:\ell\neq j\}} \phi_\ell|v_\ell - x_{\ell i}|^{a_\ell}\right)$, for $i = 1, ..., n$, and analogously for $\boldsymbol{R}_2$, replacing v_ℓ with v'_ℓ. Then using (3) and (4), we obtain

$$\mathrm{E}(\tilde{Y} \mid \boldsymbol{\psi}) = \mu + \boldsymbol{R}_1^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \boldsymbol{1}_n) \quad \text{and} \quad \mathrm{E}(\tilde{Y}' \mid \boldsymbol{\psi}) = \mu + \boldsymbol{R}_2^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \boldsymbol{1}_n) \tag{15}$$

$$\mathrm{Cov}(\tilde{Y}, \tilde{Y}' \mid \boldsymbol{\psi}) = \tau^2 \left\{ \exp\left(- \sum\nolimits_{\{\ell:\ell\neq j\}} \phi_\ell |v_\ell - v'_\ell|^{a_\ell} \right) - \boldsymbol{R}_1^T R_{\boldsymbol{\phi}}^{-1} \boldsymbol{R}_2 \right\} \tag{16}$$

Substituting (15) and (16) in (14), we obtain for each $j = 1, ..., k$,

$$\mathrm{E}^*\left\{ (\mathrm{E}(Y|u_j))^2 \mid D \right\} = \int_{\boldsymbol{\psi}} \left\{ \tau^2 \left(e - \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1} \boldsymbol{T}_j(u_j) \right) + \left(\mu + \boldsymbol{T}_j^T(u_j) R_{\boldsymbol{\phi}}^{-1} (\boldsymbol{y} - \mu \boldsymbol{1}_n) \right)^2 \right\} p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi},$$

where $e = \prod_{\{\ell:\ell\neq j\}} \left\{ \int_0^1 \int_0^1 \exp\left(-\phi_\ell |v_\ell - v'_\ell|^{a_\ell}\right) dv_\ell dv'_\ell \right\}$.

Next, we derive the expression for $\mathrm{E}^*\left\{ (\mathrm{E}(Y))^2 \mid D \right\}$. We have

$$\begin{aligned}
\mathrm{E}^*\left\{ (\mathrm{E}(Y))^2 \mid D \right\} &= \int \int \int_{\boldsymbol{\psi}} (\mathrm{E}(Y))^2 \, p(f(\boldsymbol{v}), f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi} df(\boldsymbol{v}) \, df\left(\boldsymbol{v'}\right) \\
&= \int_{\boldsymbol{\psi}} \int \int \int_{\boldsymbol{v'}} \int_{\boldsymbol{v}} f(\boldsymbol{v}) \, f\left(\boldsymbol{v'}\right) p(f(\boldsymbol{v}), f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D) \\
&\qquad \prod_{\ell=1}^{k} dH_\ell(v_\ell) \prod_{\ell=1}^{k} dH_\ell\left(v'_\ell\right) df(\boldsymbol{v}) \, df\left(\boldsymbol{v'}\right) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\boldsymbol{v'}} \int_{\boldsymbol{v}} \mathrm{E}\left(f(\boldsymbol{v}) \, f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi} \right) \prod_{\ell=1}^{k} dH_\ell(v_\ell) \prod_{\ell=1}^{k} dH_\ell\left(v'_\ell\right) p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi}
\end{aligned}$$

As before, we obtain $\mathrm{E}\left(f(\boldsymbol{v}) \, f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right)$ using the covariance formula $\mathrm{E}\left(f(\boldsymbol{v}) f(\boldsymbol{v'}) \mid \boldsymbol{\psi}\right) = \mathrm{Cov}\left(f(\boldsymbol{v}), f(\boldsymbol{v'}) \mid \boldsymbol{\psi}\right) + \mathrm{E}\left(f(\boldsymbol{v}) \mid \boldsymbol{\psi}\right) \mathrm{E}\left(f(\boldsymbol{v'}) \mid \boldsymbol{\psi}\right)$, where:

- $\mathrm{E}\left(f(\boldsymbol{v}) \mid \boldsymbol{\psi}\right) = \mu + \boldsymbol{r}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \boldsymbol{1}_n)$, with $\boldsymbol{r}$ the $(n \times 1)$ vector with i–th element given by $\mathrm{Corr}\left(f(\boldsymbol{v}), f(\boldsymbol{x}_i) \mid \boldsymbol{\psi}\right) = \exp\left[- \sum_{\ell=1}^{k} \phi_\ell \mid v_\ell - x_{\ell i} \mid^{a_\ell} \right]$
- $\mathrm{E}\left(f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right) = \mu + \boldsymbol{r'}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \boldsymbol{1}_n)$, with $\boldsymbol{r'}$ the $(n \times 1)$ vector with i–th element given by $\mathrm{Corr}\left(f(\boldsymbol{v'}), f(\boldsymbol{x}_i) \mid \boldsymbol{\psi}\right) = \exp\left[- \sum_{\ell=1}^{k} \phi_\ell \mid v'_\ell - x_{\ell i} \mid^{a_\ell} \right]$

- $\mathrm{Cov}\left(f\left(\boldsymbol{v}\right), f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right) = \tau^2 \left\{\exp\left[-\sum_{\ell=1}^{k} \phi_\ell \mid v'_\ell - v_\ell \mid^{a_\ell}\right] - \boldsymbol{r}^T R_{\boldsymbol{\phi}}^{-1} \boldsymbol{r'}\right\}$.

Combining the expressions above, we obtain $\mathrm{E}^*\left\{\left(\mathrm{E}(Y)\right)^2 \mid D\right\}$ as given in equation (10).

Finally, we derive an expression for $\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\}$. We can write

$$\begin{aligned}
\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\} &= \int\int\int_{\boldsymbol{\psi}} \mathrm{E}(Y \mid u_j)\mathrm{E}(Y) p(f\left(\boldsymbol{v}_j\right), f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi} df\left(\boldsymbol{v}_j\right) df\left(\boldsymbol{v'}\right) \\
&= \int_{\boldsymbol{\psi}} \int\int\int_{\boldsymbol{v'}} \int_{\{v_\ell : \ell \neq j\}} f\left(\boldsymbol{v}_j\right) f\left(\boldsymbol{v'}\right) p(f\left(\boldsymbol{v}_j\right), f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}) p(\boldsymbol{\psi} \mid D) \\
&\qquad \prod_{\{\ell : \ell \neq j\}} dH_\ell\left(v_\ell\right) \prod_{\ell=1}^{k} dH_\ell\left(v'_\ell\right) df\left(\boldsymbol{v}_j\right) df\left(\boldsymbol{v'}\right) d\boldsymbol{\psi} \\
&= \int_{\boldsymbol{\psi}} \int_{\boldsymbol{v'}} \int_{\{v_\ell : \ell \neq j\}} \mathrm{E}\left(f\left(\boldsymbol{v}_j\right) f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right) \prod_{\{\ell : \ell \neq j\}} dH_\ell\left(v_\ell\right) \prod_{\ell=1}^{k} dH_\ell\left(v'_\ell\right) p(\boldsymbol{\psi} \mid D) d\boldsymbol{\psi}.
\end{aligned}$$

We obtain $\mathrm{E}\left(f\left(\boldsymbol{v}_j\right) f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right)$ using the covariance formula, which requires:

- $\mathrm{E}\left(f\left(\boldsymbol{v}_j\right) \mid \boldsymbol{\psi}\right) = \mu + \boldsymbol{r}_j^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)$, with $\boldsymbol{r}_j$ the $(n \times 1)$ vector with i–th element given by $\mathrm{Corr}\left(f(\boldsymbol{v}_j), f(\boldsymbol{x}_i) \mid \boldsymbol{\psi}\right) = \exp\left[-\sum_{\{\ell : \ell \neq j\}} \phi_\ell \mid v_\ell - x_{\ell i} \mid^{a_\ell}\right] \times \exp\left[-\phi_j \mid u_j - x_{ji} \mid^{a_j}\right]$
- $\mathrm{E}\left(f\left(\boldsymbol{v'}\right) \mid \boldsymbol{\psi}\right) = \mu + \boldsymbol{r'}^T R_{\boldsymbol{\phi}}^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)$, with $\boldsymbol{r'}$ the $(n \times 1)$ vector with i–th element given by $\mathrm{Corr}\left(f(\boldsymbol{v'}), f(\boldsymbol{x}_i) \mid \boldsymbol{\psi}\right) = \exp\left[-\sum_{\ell=1}^{k} \phi_\ell \mid v'_\ell - x_{\ell i} \mid^{a_\ell}\right]$
- $\mathrm{Cov}\left(f(\boldsymbol{v}_j), f(\boldsymbol{v'}) \mid \boldsymbol{\psi}\right) = \tau^2 \left\{\exp\left[-\sum_{\{\ell : \ell \neq j\}} \phi_\ell \mid v'_\ell - v_\ell \mid^{a_\ell}\right] \exp\left[-\phi_j \mid v'_j - u_j \mid^{a_j}\right] - \boldsymbol{r}_j^T R_{\boldsymbol{\phi}}^{-1} \boldsymbol{r'}\right\}$.

The final result for $\mathrm{E}^*\left\{\mathrm{E}(Y \mid u_j)\mathrm{E}(Y) \mid D\right\}$ given in equation (11) arises by combining the expressions above.

References

Bayarri, M. J., Berger, J. O., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H., and Tu, J. (2009), "Predicting Vehicle Crashworthiness: Validation of Computer Models for Functional and Hierarchical Data," *Journal of the American Statistical Association*, 104, 929–943.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154.

Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001), "Bayesian forecasting for complex systems using computer simulators," *Journal of the American Statistical Association*, 96, 717–729.

Ganapol, B. D., Johnson, L. F., Hammer, P. D., Hlavka, C. A., and Peterson, D. L. (1998), "LEAFMOD: A New Within-Leaf Radiative Transfer Model," *Remote Sensing of Environment*, 63, 182–193.

Ganapol, B. D., Johnson, L. F., Hlavka, C. A., Peterson, D. L., and Bond, B. (1999), "LCM2: A Coupled Leaf/Canopy Radiative Transfer Model," *Remote Sensing of Environment*, 70, 153–166.

Goldstein, M. and Rougier, J. C. (2006), "Bayes linear calibrated prediction for complex systems," *Journal of the American Statistical Association*, 101, 1132–1143.

Gramacy, R. B. and Lee, H. K. H. (2008), "Bayesian Treed Gaussian Process Models With an Application to Computer Modeling," *Journal of the American Statistical Association*, 103, 1119–1130.

Gramacy, R. B. and Lian, H. (2012), "Gaussian Process Single-Index Models as Emulators for Computer Experiments," *Technometrics*, 54, 30–41.

Gramacy, R. B. and Taddy, M. (2010), "Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models," *Journal of Statistical Software*, 33, 1–48.

Han, G., Santner, T. J., Notz, W. I., and Bartel, D. L. (2009), "Prediction for Computer Experiments Having Quantitative and Qualitative Input Variables," *Technometrics*, 51, 278–288.

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

Homma, T. and Satelli, A. (1996), "Importance Measures in Global Sensitivity Analysis of Nonlinear Models," *Reliability Engineering and System Safety*, 52, 1–17.

Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, G., and Schmuck, G. (1995), "Leaf OPtical Properties EXperiment (LOPEX93)," Report EUR16095EN, Joint Research Center-European Commission, Institute for Remote Sensing Applications.

Houborg, R., Soegaard, H., and Boegh, E. (2007), "Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data," *Remote Sensing of Environment*, 106, 39–58.

Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian calibration of computer models (with discussion)," *Journal of the Royal Statistical Society B*, 63, 425–464.

Marrel, A., Iooss, B., Laurent, B., and Roustant, O. (2009), "Calculations of Sobol indices for the Gaussian process metamodel," *Reliability Engineering and System Safety*, 94, 742–751.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21, 239–245.

Morris, R. D., Kottas, A., Taddy, M., Furfaro, R., and Ganapol, B. D. (2008), "A Statistical Framework for the Sensitivity Analysis of Radiative Transfer Models," *IEEE on Geoscience and Remote Sensing*, 46, 4062–4074.

Neal, R. M. (1998), "Regression and Classification Using Gaussian Process Priors," *Bayesian Statistics*, 475–501, bernando, J., Berger, J., Dawid, A., and Smith, A., eds.

Oakley, J. and O'Hagan, A. (2002), "Bayesian inference for the uncertainty distribution of computer model outputs," *Biometrika*, 89, 769–784.

Oakley, J. E. (2009), "Decision-Theoretic Sensitivity Analysis for Complex Computer Models," *Technometrics*, 121–129.

Oakley, J. E. and O'Hagan, A. (2004), "Probabilistic Sensitivity Analysis of Complex Models: a Bayesian Approach," *Journal of the Royal Statistical Society, Series B*, 66, 751–769.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and analysis of computer experiments (C/R: p423-435)," *Statistical Science*, 4, 409–423.

Saltelli, A. (2002), "Making Best Use of Model Evaluations to Compute Sensitivity Indices," *Computer Physics Communications*, 45, 280–297.

Saltelli, A., Chan, K., and Scott, E. M. (2000), *Sensitivity Analysis*, John Wiley and Sons.

Santner, T., Williams, B., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer-Verlag.

Sobol, I. M. (1993), "Sensitivity Estimates For Non Linear Mathematical Models," *Mathematical Modelling and Computational Experiments*, 1, 407–414.

Taddy, M., Lee, H. K. H., Gray, G. A., and Griffin, J. D. (2009), "Bayesian Guided Pattern Search for Robust Local Optimization," *Technometrics*, 51, 389–401.

Ziehn, T. and Tomlin, A. S. (2009), "GUI-HDMR - A software tool for global sensitivity analysis of complex models," *Environmental Modelling & Software*, 24, 775–785.