An Analysis of the Uncertainties in Radiative Transfer Models Used in Remote Sensed Data Product Generation

Robin D. Morris
USRA-RIACS
444 Castro St, Suite 320
Mountain View, CA 94041

Athanasios Kottas and Matt Taddy
Applied Mathematics and Statistics
Baskin School of Engineering
University of California, Santa Cruz
1156 High Street, Santa Cruz, CA 95064

Roberto Furfaro and Barry D. Ganapol
Aerospace and Mechanical Engineering
University of Arizona
1130 N. Mountain, PO Box 210119
Tucson, AZ 85721

***Abstract*— The process of predicting a satellite observation of a vegetated region (e.g. a MODIS scene) involves running a Radiative Transfer Model (RTM). The RTM takes as input various biospheric and illumination parameters and computes the upwelling radiation at the top of the canopy that is ultimately observed by the satellite mounted sensor. The question we address is the following: which of the inputs to the RTM has the greatest impact on the computed observation?**

We use the Leaf Canopy Model (LCM) RTM as a surrogate for the RTM used as the basis of the MODIS production algorithm. The LCM was designed to study the feasibility of observing leaf chemistry remotely. It takes as input leaf chemistry variables (chlorophyll, water, lignin, cellulose) and canopy structural parameters (leaf area index, leaf angle distribution, soil reflectance, sun angle). The influence of each input variable, or small subsets of the inputs, is captured through the determination of the "main effects". Computing these quantities numerically requires extensive runs of the RTM, which is computationally expensive. Using a Gaussian Process approximation to the LCM RTM, we compute efficiently the main effects and determine those input variables that are vital for accurate prediction.

I. Introduction

The accurate estimation of properties of the biosphere is critical for our understanding of the Earth's coupled system. The atmosphere, oceans and land comprise a complex, coupled dynamical system, and valid statistical prediction of the properties of this system, and its changes, require inputs that are both accurate and have their uncertainties accurately quantified.

Global models require global observations, and the only effective method for making routine global measurements is via sensors mounted on orbiting satellites. Typically, however, satellite mounted sensors do not measure directly the quantity of interest. Passive visible/near infra-red sensors measure upwelling radiation, and it is from these measurements that the biospherical parameters of interest must be inferred.

This inference process is complex. It is the inversion of the process of sunlight passing through the atmosphere, being reflected off vegetation on the ground, and then passing again through the atmosphere before being detected by the satellite mounted sensor. Clearly in this scenario the uncertainty introduced by traditional "noise" in the actual sensor will be swamped by the number of places at which uncertain process models enter the estimation. In the brief outline above we have two, one for the propagation of light through the atmosphere, and the second for the reflection of light by the vegetation on the ground. It is the uncertainty characteristics of the second of these process models that we will analyse in this paper.

Analysing, quantifying and reporting the uncertainty in remote sensed data products is of great importance. It is the only way in which the uncertainty of further analyses using these data products as inputs can be quantified. Analysing the source of the data product uncertainties can identify where the models must be improved, or where better input information must be obtained. Both of these aspects are known; the editorial for the Special Issue on Global Land Product Validation [10] wrote

> users need access to quantitative information on product uncertainties

and that

> [m]aking quantified accuracy information available to the user can ultimately provide developers the necessary feedback for improving the products.

In terms of actually implementing these ideas, there is still work to be done. For example, the current MODIS LAI/fPAR algorithm has been improved continuously since the satellite's launch. The main improvements have been in the use of a better biome map (reducing the uncertainty in that input); improvements in atmospheric correction; and improved models of surface reflectance from different biomes [16].

These improvements have reduced the uncertainty in the resulting data product, but have not necessarily improved the quantification of the uncertainties, and have not specifically addressed the statistical identification of the sources of the uncertainties. Here we will address one aspect of this overall process. Models of surface reflectance are typically Radiative Transfer Models (RTMs). We analyse in detail the effects of the inputs to an RTM in terms of the sensitivity of the RTM's output to each of the inputs. Specifically we analyse the Leaf Canopy Model (LCM) RTM [4], used as a surrogate for the

RTM used as the basis for the MODIS production algorithm [7]. See section II for a discussion of the LCM. In section III, we use and develop methods from the statistical literature on sensitivity analysis [14] to compute the main effects, which graphically show the relative importance of each input on the RTM output.

Computing the main effects requires the evaluation of multidimensional integrals over the input space of the model. Evaluating RTMs is typically computationally expensive, and so standard numerical integration methods (e.g. multidimensional quadrature or Monte Carlo integration) would be computationally prohibitive in terms of the number of times the RTM would have to be run. Instead, we adopt the approach of approximating the RTM by a Gaussian Process (GP) model [8], [13], [11], which can be constructed using a comparatively small number of carefully chosen RTM evaluations. See section IV. Using the GP approximation instead of the actual RTM will introduce uncertainty into the evaluation of the main effects, but this can also be quantified [12]. See section V.

Finally, in section VI we present the main effects for the LCM RTM, and show how they enable the identification of the relative importance of each input to the model output. This also gives information as to how well these inputs can be predicted from observations of the model output at different wavelengths.

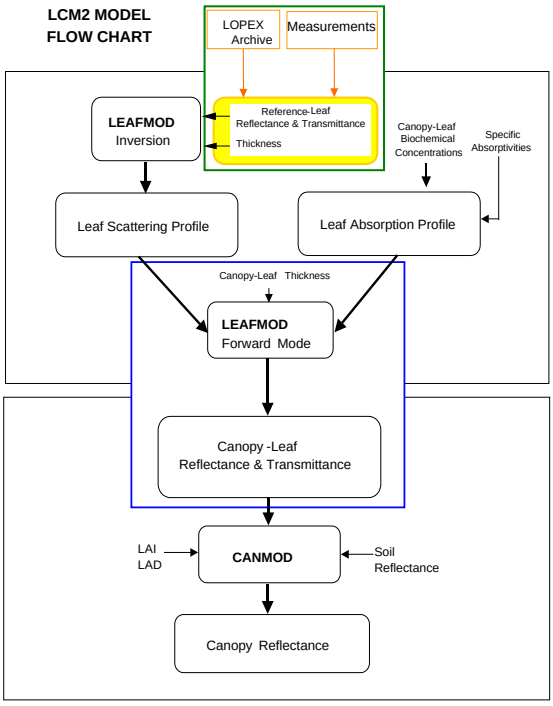

Fig. 1. LCM2 Flow Chart

II. A Coupled Leaf-Canopy Radiative Transfer Model

Over the past decade, in collaboration with the Ecosystem science and technology branch at NASA Ames, the Vegetation Modeling Transport Group (University of Arizona) has developed a coupled Leaf-Canopy Model (LCM) in order to capture the essential biophysical processes associated with the interaction between light and vegetation [4]. LCM was developed to provide a tool to aid in remote sensing as applied to ecosystem dynamics in support of the TERRA platform and it is specifically used to investigate the feasibility of observing chemistry remotely. The model combines two different radiative transfer models, one at leaf level (LEAFMOD) and one at canopy level (CANMOD) to predict the radiative regime inside the vegetation canopy under consideration.

LEAFMOD [3] is the model that simulates the radiative regime inside the single leaf. From a morphological point of view, the leaf element is an extremely complex and rich object. Any model that attempts to describe each single interaction process for the light moving in such a medium will face this enormous complexity. The strength of the LEAFMOD algorithm is its simplicity through natural averaging. The model relies on the fact that while light is moving in a complicated medium, natural averaging occurs in such way that the simpler assumption of isotropic scattering and uniform absorption seems to capture the transport effects. Moreover, the model has the ability to include chemistry as a key element dominating the absorption process. Different concentrations of chlorophyll, water, lignin and cellulose can be specified to model the optical properties of the single leaf species. The model is calibrated over the LOPEX leaf species archive [6], where experimental leaf property data are stored. The calibration occurs in the sense that the optical properties required by the canopy model are retrieved through a procedure that uses the LOPEX archive as input data.

The CANMOD (CANopy Model) algorithm [5], [4] takes the information coming from LEAFMOD regarding the single leaf characteristic (transmittance and reflectance) and together with canopy structural parameters, (LAI and Leaf Angle Distribution), soil reflectance and sun angle inclination, computes, at any given wavelength, the radiative regime within and at the top of the canopy by solving a radiative transfer equation. The strengths of the model are simplicity and the ability to take into account leaf chemistry, which is important to properly describe the light absorption environment.

Figure 1 shows a flowchart that demonstrates the operation of the coupled algorithm. The algorithm can be explained as follows. The first module uses LEAFMOD in the forward and inverse mode to compute the leaf optical properties (i.e. leaf reflectance and transmittance). The second module uses the CANMOD forward mode to compute the spectral canopy hemispherical reflectance factor. The code requires the specification of the input parameters. In addition to the parameters listed in table I, the model also takes as input wavelength (between 400nm and 2100nm), canopy architecture (LAD - leaf angle distribution, which takes one of 5 discrete values) and the sun angle.

Note that the soil reflectance depends on the wavelength. Indeed, usually the spectral soil reflectance is specified depending on the type of soil of interest but once the wavelength

input	min	max
LAI	0	8
chlorophyll $(\mu g/cm^2)$	0	100
water fraction	0.1	0.8
protein (g/cm^2)	0.0001	0.001
lignin/cellulose (g/cm^2)	0.0001	0.006
thickness (cm)	0.001	0.01
soil	0.3	1.3

TABLE I

RANGES OF VALUES OF THE INPUTS TO THE LCM. LAI AND WATER FRACTION ARE DIMENSIONLESS. THE SOIL PARAMETER IS MULTIPLIED BY A STANDARD SOIL SPECTRUM. SEE TEXT.

is set, the algorithm will work with the value of soil reflectance relative to the specified wavelength. We assumed a typical visible/near infra-red spectrum for a dry soil, and we considered a multiplicative brightness parameter varying between 0.3-1.3 (see table I) to account for the possible variations of the background (soil) reflectance level [1]. This multiplicative parameter is assumed to be wavelength independent.

Once the leaf type is specified, the LOPEX database contains the measured leaf optical properties for the leaf of interest. Nevertheless, we can tune the canopy by considering leaves that are of the same type but with different biochemistry and thickness. This gives the code great flexibility in modeling the effect of biochemistry on the overall canopy reflectance. The algorithm begins by analyzing the leaf under consideration. Assume, for example that the canopy of interest is a maple canopy. The LOPEX database is accessed to retrieve the measured spectral reflectance and transmittance for a nominal maple leaf. Note that, as before, since the wavelength is set, reflectance and transmittance for the nominal leaf are selected for the specific wavelength of interest. The LEAFMOD inverse mode accepts the reflectance and transmittance and retrieves scattering and absorption coefficients. It is assumed that, to first order, the scattering depends on the anatomical structure of the leaf, while the absorption depends only on the biochemical components [3]. Thus, the scattering coefficient for maple leaves is assumed to be the same and it is retained. A new maple leaf having the biochemical components and thickness specified by the inputs is constructed retaining the same scattering coefficient and constructing the new absorption coefficient for the wavelength of interest. Both absorption and scattering coefficient are fed to the LEAFMOD forward mode to compute the reflectance and transmittance of the desired leaf, i.e. the leaf with thickness, water, chlorophyll, lignin and protein specified by the inputs. Reflectance and transmittance are fed to the second module together with LAI, LAD, soil reflectance and sun angle to compute the hemispherical reflectance.

III. SENSITIVITY ANALYSIS

Sensitivity analysis aims to determine how the variation in the output of a model can be apportioned amongst the inputs [15, ch 7]. That is, it attempts to determine how much of the variation seen in the output is due to variation in each of the inputs. The type of sensitivity analysis we are interested in here is *global* sensitivity analysis, looking at how the output changes as all the inputs vary continuously, rather than the more common *local* sensitivity analyses, which look at how the output changes as the inputs are each varied about a fixed point. Clearly this latter type of analysis will give limited information about how the output varies for substantial changes in the inputs.

How the inputs vary is determined by a probability distribution that defines the expected distributions of the inputs. Using $\boldsymbol{v}$ to denote the vector of model inputs, this distribution is $H(\boldsymbol{v})$. The actual form of this distribution is problem dependent, and dependent on the amount of knowledge available about each input variable. It may be that for some inputs all that can be given is a physically plausible range (e.g. water fraction is limited to the range 0-1), whereas for others a more precise distributions may be known (e.g. the distribution of leaf thickness for a particular tree type may be known from field measurements). The distribution $H(\boldsymbol{v})$ also encodes correlations between variables that are known to vary together. The authors in [1] give truncated Gaussian distributions for the variables in table I. In this work we use the simpler formulation of independent uniform distributions over the ranges given in table I for each input variable.

A. Main Effects

Denote the response of the model to input $\boldsymbol{v}$ as $y = f(\boldsymbol{v})$. The function $f(\boldsymbol{v})$ can be decomposed as

$$\begin{aligned} y = f(\boldsymbol{v}) &= \mathrm{E}(Y) + \sum_{i=1}^{d} z_i(v_i) + \sum_{i<j} z_{i,j}(v_i, v_j) + \dots \\ &\quad + z_{1,2,\dots,d}(v_1, v_2, \dots, v_d) \end{aligned} \tag{1}$$

where $\boldsymbol{v} = (v_1, ..., v_d)$ is $d-$dimensional (with $d = 7$ in our sensitivity analysis of the LCM). The first term is the expected value of $f(\boldsymbol{v})$, i.e.,

$$\mathrm{E}(Y) = \int_{v_j, j=1\dots d} f(\boldsymbol{v}) \mathrm{d}H(\boldsymbol{v})$$

and the next d terms are the *main effects*, given by

$$\begin{aligned} z_i(v_i) &= E(Y|v_i) - \mathrm{E}(Y) \\ &= \int_{\boldsymbol{v}_{-i}} f(\boldsymbol{v}) \mathrm{d}H(\boldsymbol{v}_{-i}|v_i) - \mathrm{E}(Y) \end{aligned} \tag{2}$$

where $\boldsymbol{v}_{-i}$ denotes all the elements of $\boldsymbol{v}$ except v_i. The later terms of the decomposition are the interactions. They give information about the combined influence of two or more inputs taken together. We will not consider them further here.

Plotting the main effects, $z_i(v_i)$ for each i gives a visual impression of the relative importance of each input to the variation in the output. This visual impression is heightened if the inputs are normalized (to the range 0-1, for example, for uniformly distributed inputs), allowing all the main effects to be plotted together on the same plot. See section VI where we present main effects plots for the LCM RTM.

To compute the main effects requires the evaluation of a $(d-1)-$dimensional integral. For even moderately complex functions $f(\boldsymbol{v})$ it will be impossible to evaluate this integral analytically; indeed, for most cases of interest an analytic form for $f(\boldsymbol{v})$ does not exist, rather, $f(\boldsymbol{v})$ only exists as a computer program. In these cases the $z_i(v_i)$ must be computed numerically. If evaluating $f(\boldsymbol{v})$ for a given $\boldsymbol{v}$ requires appreciable computation then the standard methods of numerical integration, multidimensional quadrature and Monte Carlo integration, will be too computationally intensive to be practical. In these cases we can approximate $f(\boldsymbol{v})$ and compute the main effects of the approximation, and also compute the uncertainty introduced by the approximation to $f(\boldsymbol{v})$. This is given in sections IV and V. Details of the Gaussian Process approximation we use for $f(\boldsymbol{v})$ is given in section IV, and its application to computing the main effects in section V.

IV. Approximating the LCM Using a Gaussian Process

Gaussian Processes (GPs) are probability distributions over *functions*. Rather than placing a distribution over a (small) set of parameters, a GP places a distribution directly over the function of interest. Under a GP probability model for function $f(\cdot)$, the joint distribution of $(f(\boldsymbol{v}_1), ..., f(\boldsymbol{v}_k))$ is multivariate Gaussian, for any finite set of input points $\boldsymbol{v}_1, ..., \boldsymbol{v}_k$. It is this property that allows for tractable computation – whilst the GP is defined over an infinite dimensional quantity (the continuous function, $f(\boldsymbol{v})$), any computation is necessarily done over only a finite set of locations.

A GP is specified by its mean function, $\mathrm{E}(f(\boldsymbol{v}))$, and its covariance function $\mathrm{Cov}(f(\boldsymbol{v}), f(\boldsymbol{v}'))$. The flexibility of choosing and adapting the mean and covariance functions allows a GP model to be successfully used to approximate a wide spectrum of functions $f(\boldsymbol{v})$, based on a set of training examples, $\boldsymbol{d} = \{\boldsymbol{y}, \boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, where $\boldsymbol{y} = (y_1, ..., y_n)$ and y_i is the response $f(\boldsymbol{x}_i)$ at observed input point $\boldsymbol{x}_i$, $i = 1, ..., n$. The set of training examples is chosen carefully to optimally sample the input space. Here we used a Latin Hypercube design [9] to choose the set of inputs to the LCM. The other choices made were to use a constant mean function, $\mathrm{E}(f(\boldsymbol{v})) = \mu$, a constant variance $\mathrm{Var}(f(\boldsymbol{v})) = \sigma^2$, and the product Gaussian correlation function

$$\mathrm{Corr}(f(\boldsymbol{v}), f(\boldsymbol{v}'); \boldsymbol{\theta}) = \exp\left(-\sum_{\ell=1}^{d} \frac{(v_\ell - v'_\ell)^2}{\gamma_\ell}\right)$$

where $\boldsymbol{\theta} = (\gamma_1, ..., \gamma_d)$, and d is the number of dimensions in the inputs, $\boldsymbol{v}$. The γ parameters give a measure of the scale over which the function $f(\boldsymbol{v})$ varies in each input dimension, and σ^2, the variance of the GP, determines the overall scale of $f(\boldsymbol{v})$. Using this mean and correlation function, the GP defines the joint distribution

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}|C|^{1/2}} \times \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \mu\mathbf{1}_n)^T C^{-1}(\boldsymbol{y} - \mu\mathbf{1}_n)\right) \quad (3)$$

where C is the correlation matrix with (i,j)-th element $\mathrm{Corr}(f(\boldsymbol{x}_i), f(\boldsymbol{x}_j))$, and $\mathbf{1}_n$ denotes an n-dimensional vector with all elements equal to 1.

We use the set of training examples, $\boldsymbol{d}$, to estimate the parameters $\{\boldsymbol{\theta}, \mu, \sigma^2\}$ of the GP model using maximum likelihood estimation. From equation 3 the log-likelihood is

$$\mathcal{L} = -\frac{1}{2\sigma^2}(\boldsymbol{y} - \mu\mathbf{1}_n)^T C(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \mu\mathbf{1}_n) - \frac{1}{2}\log|C(\boldsymbol{\theta})| - \frac{n}{2}\log(2\pi\sigma^2) \quad (4)$$

where we have made explicit the dependence of C on the parameters $\boldsymbol{\theta}$. The derivatives of $\mathcal{L}$ with respect to each of the parameters can be straightforwardly derived. Maximizing $\mathcal{L}$ results in a point estimate for the parameters, denoted by $\{\hat{\boldsymbol{\theta}}, \hat{\mu}, \hat{\sigma^2}\}$, that we use when evaluating the main effects. Note that using point estimates for these parameters will cause the uncertainty of the main effects to be underestimated. In future work we will consider a fully inferential Bayesian approach where expectations are also taken with respect to these parameters.

Once the GP model parameters are estimated, the first quantities of interest are the conditional predictive distributions for sets of new inputs. From the definition of the GP, these distributions will be Gaussian. For a single new input $\boldsymbol{v}$ the predictive distribution for $f(\boldsymbol{v})$ has mean

$$m \equiv m(\boldsymbol{v}; \hat{\mu}, \hat{\boldsymbol{\theta}}, \boldsymbol{d}) = \hat{\mu} + \boldsymbol{r}^T(\hat{\boldsymbol{\theta}})C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n)$$

and variance

$$S \equiv S(\boldsymbol{v}; \hat{\mu}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}}, \boldsymbol{d}) = \hat{\sigma}^2\left(1 - \boldsymbol{r}^T(\hat{\boldsymbol{\theta}})C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{r}(\hat{\boldsymbol{\theta}})\right).$$

Here $\boldsymbol{r}(\hat{\boldsymbol{\theta}})$ is the $n \times 1$ vector with i-th element given by $\mathrm{Corr}(f(\boldsymbol{v}), f(\boldsymbol{x}_i)) = \exp(-\sum_{\ell=1}^{d}(v_\ell - x_{i\ell})^2/\hat{\gamma}_\ell)$, and $C(\hat{\boldsymbol{\theta}})$ is the observed $n \times n$ correlation matrix with (i,j)-th element given by $\exp(-\sum_{\ell=1}^{d}(x_{i\ell} - x_{j\ell})^2/\hat{\gamma}_\ell)$. Recall that the $\boldsymbol{x}_i$ are the input values of the training examples.

The joint predictive distribution for $(f(\boldsymbol{v}), f(\boldsymbol{v}'))$ corresponding to generic inputs $\boldsymbol{v} = (v_1, ..., v_d)$ and $\boldsymbol{v}' = (v'_1, ..., v'_d)$ is bivariate normal with (2×1) mean vector

$$\boldsymbol{w} = \hat{\mu}\mathbf{1}_2 + R^T(\hat{\boldsymbol{\theta}})C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n) \quad (5)$$

and (2×2) covariance matrix

$$W = \hat{\sigma}^2\left(B(\hat{\boldsymbol{\theta}}) - R^T(\hat{\boldsymbol{\theta}})C^{-1}(\hat{\boldsymbol{\theta}})R(\hat{\boldsymbol{\theta}})\right), \quad (6)$$

where $B(\hat{\boldsymbol{\theta}})$ is the (2×2) correlation matrix for $(f(\boldsymbol{v}), f(\boldsymbol{v}'))$ with off-diagonal element given by $\exp(-\sum_{\ell=1}^{d}(v_\ell - v'_\ell)^2/\hat{\gamma}_\ell)$, and $R(\hat{\boldsymbol{\theta}})$ is the $(n \times 2)$ matrix with first row elements $\exp(-\sum_{\ell=1}^{d}(v_\ell - x_{i\ell})^2/\hat{\gamma}_\ell)$, $i = 1, ..., n$, and analogously for the second row elements replacing v_ℓ with v'_ℓ.

V. Approximating the Main Effects using the Gaussian Process Approximation to the LCM

To compute the main effects requires evaluating $\mathrm{E}(Y \mid v_j)$ and $\mathrm{E}(Y)$, as indicated in equation 2. However, we recall that we are approximating the function $y = f(\boldsymbol{v})$ by a GP model,

and we must account for this approximation by computing $\mathrm{E}^*\{\mathrm{E}(Y \mid v_j)\}$ and $\mathrm{E}^*\{\mathrm{E}(Y)\}$, where, following [12], we use $\mathrm{E}^*\{\,\}$, $\mathrm{Var}^*\{\,\}$ and $\mathrm{Cov}^*\{\,\}$ to indicate expectation, variance and covariance, respectively, with respect to the GP predictive distributions. We give details of these quantities here.

For the global mean, we have

$$\mathrm{E}(Y) = \int_{\boldsymbol{v}} f(\boldsymbol{v}) \prod_{\ell=1}^{d} \mathrm{d}H_\ell(v_\ell)$$

where $H(\boldsymbol{v}) = \prod_{\ell=1}^{d} H_\ell(v_\ell)$ is the input distribution, comprising independent components $H_\ell(v_\ell)$, which are uniform distributions over ranges (a_ℓ, b_ℓ), $\ell = 1, ..., d$. Therefore,

$$\begin{aligned}
\mathrm{E}^*\{\mathrm{E}(Y)\} &= \int \mathrm{E}(Y)\mathrm{d}N(f(\boldsymbol{v}); m, S) \\
&= \int_{\boldsymbol{v}} m(\boldsymbol{v}) \prod_{\ell=1}^{d} \mathrm{d}H_\ell(v_\ell) \\
&= \int_{\boldsymbol{v}} \{\hat{\mu} + \boldsymbol{r}^T(\hat{\boldsymbol{\theta}}) C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n)\} \\
&\quad \times \prod_{\ell=1}^{d} \mathrm{d}H_\ell(v_\ell) \\
&= \hat{\mu} + \boldsymbol{T}^T C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n),
\end{aligned}$$

where $\boldsymbol{T}$ is the $n \times 1$ vector with i-th element given by $\prod_{\ell=1}^{d} \left\{ \int_{a_\ell}^{b_\ell} \exp(-(v_\ell - x_{i\ell})^2/\hat{\gamma}_\ell)(b_\ell - a_\ell)^{-1} \mathrm{d}v_\ell \right\}$.

For $\mathrm{E}(Y \mid v_j)$, for each value u_j of the $j-$th input, we have

$$\mathrm{E}(Y \mid u_j) = \int_{\{v_\ell : \ell \neq j\}} f(v_1, ..., u_j, ..., v_d) \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v_\ell)$$

and thus

$$\begin{aligned}
&\mathrm{E}^*\{\mathrm{E}(Y \mid u_j)\} \\
&= \int \mathrm{E}(Y \mid u_j)\mathrm{d}N(f(v_1, ..., u_j, ..., v_d); m, S) \\
&= \int_{\{v_\ell : \ell \neq j\}} m(v_1, ..., u_j, ..., v_d) \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v_\ell) \\
&= \hat{\mu} + \boldsymbol{T}_j^T(u_j) C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n), \qquad (7)
\end{aligned}$$

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

$$\exp\left(-\frac{(u_j - x_{ij})^2}{\hat{\gamma}_j}\right) \times \prod_{\{\ell : \ell \neq j\}} \left\{ \int_{a_\ell}^{b_\ell} \exp\left(-\frac{(v_\ell - x_{i\ell})^2}{\hat{\gamma}_\ell}\right) \frac{1}{b_\ell - a_\ell} \mathrm{d}v_\ell \right\}. \qquad (8)$$

The previous expressions provide point estimates for all main effects associated with the d inputs. In particular, for each input $j = 1, ..., d$, $\mathrm{E}^*\{\mathrm{E}(Y \mid u_j)\}$ can be computed over a grid of u_j values to obtain point estimates for the functions $\mathrm{E}(Y \mid u_j)$ (or for $\mathrm{E}(Y \mid u_j) - \mathrm{E}(Y)$ using also $\mathrm{E}^*\{\mathrm{E}(Y)\}$). These estimates can be compared graphically (linear transformations can be applied so that all inputs are on the same scale).

For a measure of the uncertainty associated with these estimates, we use

$$\mathrm{Var}^*\{\mathrm{E}(Y \mid u_j)\} = \mathrm{E}^*\left\{(\mathrm{E}(Y \mid u_j))^2\right\} - (\mathrm{E}^*\{\mathrm{E}(Y \mid u_j)\})^2.$$

Because we already have the expression for $\mathrm{E}^*\{\mathrm{E}(Y \mid u_j)\}$ through (7) and (8), what is needed is an expression for $\mathrm{E}^*\left\{(\mathrm{E}(Y \mid u_j))^2\right\}$. Note that,

$$\begin{aligned}
&(\mathrm{E}(Y \mid u_j))^2 = \\
&\qquad \left(\int_{\{v_\ell : \ell \neq j\}} f(v_1, ..., u_j, ..., v_d) \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v_\ell) \right)^2 \\
&\qquad = \iint_{\substack{\{v_\ell : \ell \neq j\} \\ \{v'_\ell : \ell \neq j\}}} f(v_1, ..., u_j, ..., v_d) f(v'_1, ..., u_j, ..., v'_d) \\
&\qquad\quad \times \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v_\ell) \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v'_\ell)
\end{aligned}$$

and thus we need to take $\mathrm{E}^*\{\cdot\}$ with respect to the bivariate predictive distribution for $(f(v_1, ..., u_j, ..., v_d), f(v'_1, ..., u_j, ..., v'_d))$. Specifically,

$$\begin{aligned}
&\mathrm{E}^*\left\{(\mathrm{E}(Y \mid u_j))^2\right\} = \\
&\quad \iint_{\substack{\{v_\ell : \ell \neq j\} \\ \{v'_\ell : \ell \neq j\}}} \mathrm{E}^*\{f(v_1, ..., u_j, ..., v_d) f(v'_1, ..., u_j, ..., v'_d)\} \\
&\qquad\qquad \times \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v_\ell) \prod_{\{\ell : \ell \neq j\}} \mathrm{d}H_\ell(v'_\ell), \qquad (9)
\end{aligned}$$

where

$$\begin{aligned}
&\mathrm{E}^*\{f(v_1, ..., u_j, ..., v_d) f(v'_1, ..., u_j, ..., v'_d)\} = \\
&\qquad \mathrm{Cov}^*\{f(v_1, ..., u_j, ..., v_d), f(v'_1, ..., u_j, ..., v'_d)\} \\
&+ (\mathrm{E}^*\{f(v_1, ..., u_j, ..., v_d)\}\, \mathrm{E}^*\{f(v'_1, ..., u_j, ..., v'_d)\}). \qquad (10)
\end{aligned}$$

Denote by $\boldsymbol{R}_1^T(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{R}_2^T(\hat{\boldsymbol{\theta}})$ the first and second rows, respectively, of the $(n \times 2)$ matrix $R(\hat{\boldsymbol{\theta}})$ defined in section IV. Note that here the input vectors we are working with, $(v_1, ..., u_j, ..., v_d)$ and $(v'_1, ..., u_j, ..., v'_d)$, have common element u_j. Therefore, $\boldsymbol{R}_1(\hat{\boldsymbol{\theta}})$ is the $n \times 1$ vector with elements

$$\exp\left(-\frac{(u_j - x_{ij})^2}{\hat{\gamma}_j} - \sum_{\{\ell : \ell \neq j\}} \frac{(v_\ell - x_{i\ell})^2}{\hat{\gamma}_\ell}\right), \quad i = 1, ..., n,$$

and analogously for $\boldsymbol{R}_2(\hat{\boldsymbol{\theta}})$, replacing v_ℓ with v'_ℓ. Then, using (5) and (6), we obtain

$$\begin{aligned}
\mathrm{E}^*\{f(v_1, ..., u_j, ..., v_d)\} &= \hat{\mu} + \boldsymbol{R}_1^T(\hat{\boldsymbol{\theta}}) C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n) \\
\mathrm{E}^*\{f(v'_1, ..., u_j, ..., v'_d)\} &= \hat{\mu} + \boldsymbol{R}_2^T(\hat{\boldsymbol{\theta}}) C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n)
\end{aligned}$$

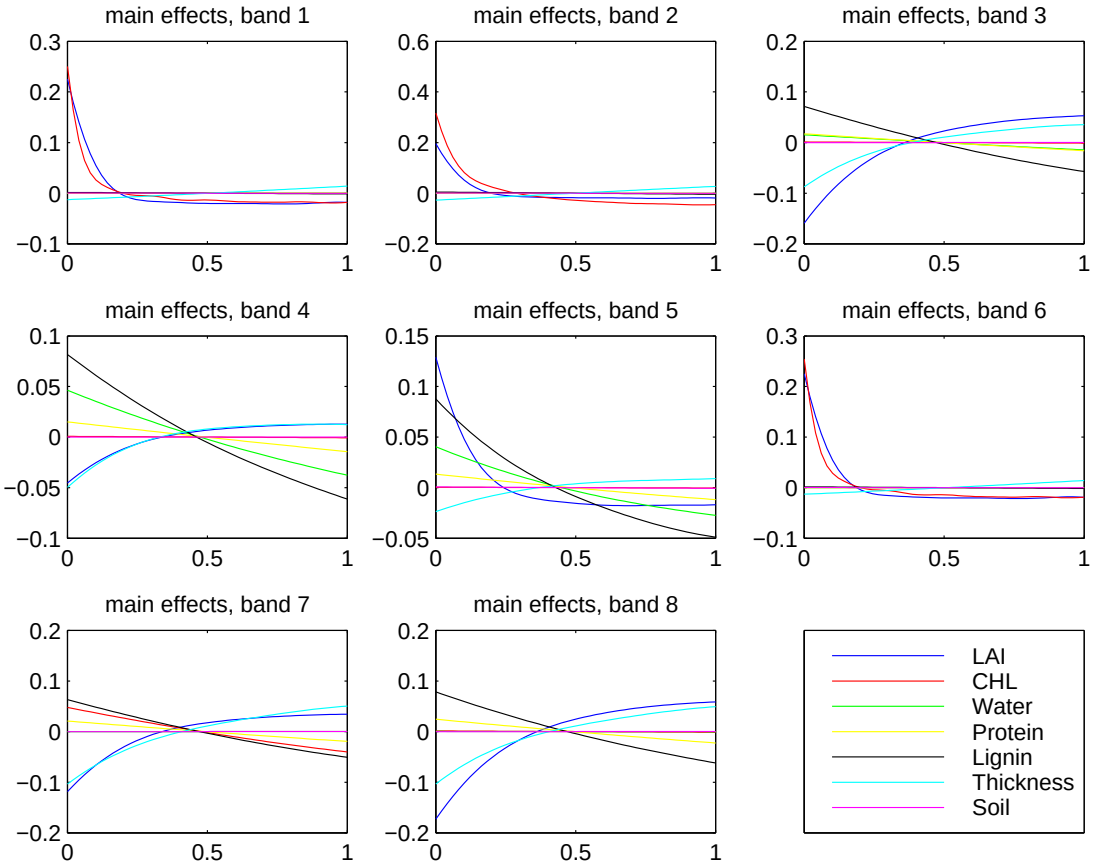

Fig. 2. The main effects for the LCM RTM

$$\text{Cov}^*\left\{f(v_1,...,u_j,...,v_d), f(v_1',...,u_j,...,v_d')\right\} = \hat{\sigma}^2\left\{\exp\left(-\sum_{\{\ell:\ell\neq j\}}\frac{(v_\ell - v_\ell')^2}{\hat{\gamma}_\ell}\right) - \boldsymbol{R}_1^T(\hat{\boldsymbol{\theta}})C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{R}_2(\hat{\boldsymbol{\theta}})\right\}. \quad (11)$$

Finally, substituting (10) and (11) in (9), we obtain for each $j = 1, ..., d$,

$$\text{E}^*\left\{(\text{E}(Y \mid u_j))^2\right\} = \hat{\sigma}^2\left(e(\hat{\boldsymbol{\theta}}) - \boldsymbol{T}_j^T(u_j)C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{T}_j(u_j)\right) + \left(\hat{\mu} + \boldsymbol{T}_j^T(u_j)C^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{y} - \hat{\mu}\mathbf{1}_n)\right)^2, \quad (12)$$

where $\boldsymbol{T}_j(u_j)$ is the $n \times 1$ vector with elements given in (8), and

$$e(\hat{\boldsymbol{\theta}}) = \prod_{\{\ell:\ell\neq j\}}\left\{\int_{a_\ell}^{b_\ell}\int_{a_\ell}^{b_\ell}\exp\left(-\frac{(v_\ell - v_\ell')^2}{\hat{\gamma}_\ell}\right)\frac{\text{d}v_\ell \text{d}v_\ell'}{(b_\ell - a_\ell)^2}\right\}.$$

Note that the second term in equation 12 is $(\text{E}^*\{E(Y \mid u_j)\})^2$, and so the required variance has the simpler expression

$$\text{Var}^*\left\{\text{E}(Y \mid u_j)\right\} = \hat{\sigma}^2\left(e(\hat{\boldsymbol{\theta}}) - \boldsymbol{T}_j^T(u_j)C^{-1}(\hat{\boldsymbol{\theta}})\boldsymbol{T}_j(u_j)\right)$$

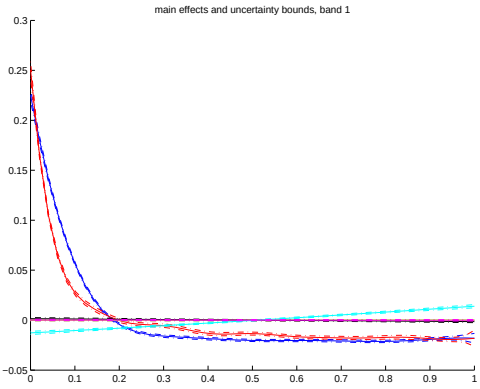

Fig. 3. The uncertainty in the main effects due to using the GP approximation to the LCM RTM. Band 1. Line colours as in figure 2.

VI. Results

To generate the training data for the GP model we generated a 250 point Latin Hypercube design over the 7-dimensional space of inputs given in table I. The Leaf Angle Distribution (LAD) variable was set to planophile (leaves mostly horizontal) and the sun angle was set to zenith. While the sun angle will vary, for any given satellite scene it will be known, and so

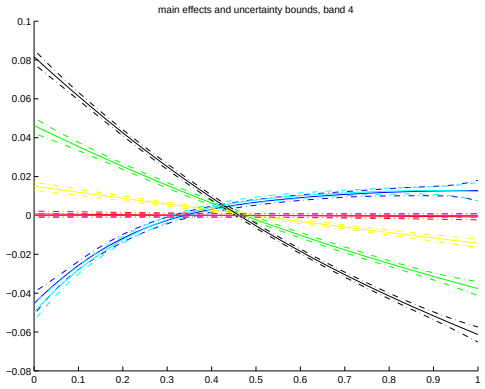

Fig. 4. The uncertainty in the main effects due to using the GP approximation to the LCM RTM. Band 4. Line colours as in figure 2.

band number	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 II
THE WAVELENGTH FOR EACH BAND USED, AND THE CORRESPONDING MODIS BAND NUMBER.

we do not consider it as one of the inputs for the uncertainty analysis. The LCM was run at 8 wavelengths, given in table II, corresponding to eight of the MODIS bands that are sensitive to vegetation. The corresponding MODIS band is also given in table II. Note that the bands are in MODIS band order, not in wavelength order.

Figure 2 shows plots of the main effects for the 7 input variables for each of the 8 bands. The larger the variation of the main effect plot, the greater the influence of that input on the LCM response.

In bands 1, 2 and 6 the response is dominated by LAI and chlorophyll. This is consistent with the results of a much more restricted sensitivity analysis in [2]. The analysis here, however, reveals that for other spectral bands other input variables become important. For example, in bands 4 and 5 there is a strong response to the lignin input, and the leaf thickness is important in bands 7 and 8.

Figures 3 and 4 show the same main effects plots for band 1 and 4, but include the uncertainty bounds due to approximating the LCM by the GP. The uncertainties are sufficiently small that they do not affect any conclusions drawn from the main effects plots.

These results show that analysing the uncertainty characteristics of RTMs used in remote sensed data product generation is practical and important. It gives information on the level of accuracy needed in the model's inputs, can guide data collection efforts to most effectively reduce the uncertainties, and can guide further development effort for the RTMs themselves.

VII. Acknowledgments

This work was supported by the NASA AISR program, through grant number NNG06G177G.

References

[1] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau. Neural network estimation of LAI, $fAPAR$, $fCover$ and $LAI \times C_{ab}$ from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of the Environment*, 105:313–325, 2006.

[2] J. Dungan and B. Ganapol. Sources of uncertainty in the prediction of LAI/fPAR from MODIS. In *AGU Fall Meeting*, San Francisco, 2002.

[3] B.D. Ganapol, L.F. Johnson, C.A. Hlavka, and D.L. Peterson. LEAFMOD: A new within-leaf radiative transfer model. *Remote Sensing of the Environment*, 63:182–193, 1998.

[4] B.D. Ganapol, L.F. Johnson, C.A. Hlavka, D.L. Peterson, and B. Bond. LCM2: A coupled leaf/canopy radiative transfer model. *Remote Sensing of the Environment*, 70:153–166, 1999.

[5] B.D. Ganapol and R.B Myneni. The FN method for the one-angle radiative transfer equation applied to plant canopies. *Remote Sensing of the Environment*, 39:213–231, 1992.

[6] B. Hosgood, S. Jacquemound, G. Andreoli, J. Verdebout, G. Peredrini, and G. Schmuck. Leaf OPtical Properties EXperiment (LOPEX93: Report EUR16095EN. Technical report, Joint Research Center-European Commission, Institute for Remote Sensing Applications, 1995.

[7] Y. Knyazikhin, J.V. Martonchik, R.B. Myneni, D.J. Diner, and S.W. Running. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, 103:32257–32276, 1998.

[8] D.J.C. Mackay. Introduction to Gaussian process. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166, 1998.

[9] M.D. McKay, R.J. Beckman, and W.J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

[10] J.T. Morisette, F. Baret, and S. Liang. Special issue on global land product validation. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1695–1697, 2006.

[11] R.M. Neal. Regression and classification using Gaussian process priors. In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 6*, pages 475–501. Oxford University Press, 1998.

[12] J.E. Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society, Series B*, 66:751–769, 2004.

[13] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[14] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis*. John Wiley and Sons, Chichester, 2000.

[15] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, 2003.

[16] W. Yang, B. Tan, D. Huang, M. Rautiainen, N.V. Shabanov, Y. Wang, J.L. Privette, K.F. Huemmrich, R. Fenshold, I. Sandholt, M. Weiss, D.E. Ahl, S.T. Glower, R.R. Nemani, Y. Knyazikhin, and R.B. Myneni. MODIS leaf area index products: From validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1885–1989, 2006.