

ARTMO'S GLOBAL SENSITIVITY ANALYSIS (GSA) TOOLBOX TO QUANTIFY DRIVING VARIABLES OF LEAF AND CANOPY RADIATIVE TRANSFER MODELS

Jochem Verrelst, Juan Pablo Rivera, and Jose Moreno

University of Valencia, Image Processing Laboratory (IPL),
Laboratory of Earth Observation (LEO), Valencia, Spain;
[jochem.verrelst](mailto:jochem.verrelst@uv.es) / [rijuanpa](mailto:rijuanpa@uv.es) / [jose.moreno](mailto:jose.moreno@uv.es){at}uv.es

ABSTRACT

This contribution introduces ARTMO's new global sensitivity analysis (GSA) toolbox. With the GSA toolbox the majority of ARTMO's available radiative transfer models (RTMs) can be used for global sensitivity analysis with little user interaction. For a selected RTM output, the global sensitivity analysis identifies most influential and non-influential input variables according to Sobol' first-order and total-order indices. The toolbox can process RTM spectral outputs for any kind of optical sensor setting within the spectral range of 400-2400 nm. Multiple model outputs can be automatically analysed within the same execution, which is of interest for more advanced, multi-output RTMs such as SCOPE. To illustrate its functioning, total-order sensitivity indices results for spectral outputs of PROSPECT-4/5 (leaf) and PROSPECT+SAIL (canopy) RTMs were calculated. Wave-length-dependent key driving and non-influential input variables were identified. The toolbox can be beneficial to the broader remote sensing community to gain insight into vegetation-light interactions and RTM input-output functioning.

KEYWORDS

ARTMO, global sensitivity analysis, radiative transfer models, sensitivity indices, leaf and canopy

INTRODUCTION

Since the advent of optical remote sensing, physically based radiative transfer models (RTMs) have helped in the understanding of light interception by plant canopies and the interpretation of vegetation reflectance in terms of biophysical characteristics (1). RTMs attempt to describe absorption and scattering, and are useful in a wide range of applications, including designing vegetation indices, conducting sensitivity analyses, developing inversion procedures to accurately retrieve vegetation properties from remotely sensed data (see (2) for a review), and generation of artificial scenes as observed by an optical sensor. Plant and atmospheric RTMs are currently used in an End-to-End simulator that functions as a virtual laboratory in the development of new optical sensors, for instance in preparation of ESA's candidate Earth Explorer mission FLEX (Fluorescence Explorer) (3).

In all of these studies, an important requirement is to know the key input variables driving the spectral output in a specific spectral region. Such knowledge can lead to a simplified model that is driven only by the key variables, which makes exploration of a broad range of target and observation conditions easier and more effective (4). To achieve this, a sensitivity analysis (SA) is required.

A sensitivity analysis evaluates the relative importance of each input variable in a model and can be used to identify the most influential variables affecting model outputs to which the model output is most sensitive (5). Hence, sensitivity analysis can be applied with RTMs to identify the key determinants of outputs such as reflectance, fluorescence, etc. Less influential variables can also be identified and be safely set to default values under relatively wide ranges of conditions. In general, sensitivity analysis methods can be categorized as either local or global. Local sensitivity analysis

methods are often referred to as “one-factor-at-a-time” (OAT), because they involve changing one input variable at a time whilst holding all others at their default values, then measuring variation in the outputs. A drawback of OAT methods is their being informative only at the default point where the calculation is executed and do not encompass the entire input variable space. Thus, local sensitivity analysis methods are inadequate for analysis of complex models with many variables that may be highly dimensional and/or non-linear (6,7,8).

Unlike local sensitivity analysis, global sensitivity analysis (GSA) explores the full input variable space (6). The contribution of each input variable to the variation in outputs is averaged over the variation of all input variables, i.e., all input variables are changed together. In general, variance-based sensitivity analysis methods aim to quantify the amount of variance that each input variable contributes to the unconditional variance (variance across all simulations) of the model output (8). GSA is thus required to identify the driving variables of a RTM. However, GSA is computationally intensive and the diversity of GSA methods may be perceived as overly complicated. Currently, no user-friendly toolbox exists that enables GSA methods to be applied to RTMs. Accordingly, the development of a GSA toolbox dedicated to the analysis of RTMs may greatly benefit the use of RTMs in all kinds of remote sensing applications.

Over the past few years, various leaf and canopy RTMs have been implemented and standardized into a single scientific graphical user interface (GUI) toolbox called ARTMO (Automated Radiative Transfer Models Operator) (9). In ARTMO, RTMs can be operated in a semi-automatic fashion for any kind of optical sensor in the visible, near-infrared and shortwave infrared range (400-2500 nm). Therefore, having a diverse range of RTMs with varying complexity at hand, this platform can perfectly serve as a benchmark in the development and evaluation of a GSA toolbox.

This study aims at addressing the following objectives: (i) to develop a global sensitivity analysis toolbox within the ARTMO framework that enables a GSA to be applied to the available RTMs; and (ii) to demonstrate the usefulness of the GSA toolbox by identifying the driving variables of leaf and canopy RTMs across the optical spectral range.

VARIANCE-BASED GLOBAL SENSITIVITY ANALYSIS

The most popular variance-based methods include the Fourier Amplitude Sensitivity Test (FAST) (10), the Sobol’ method (11), and a modified version of the Sobol’ method proposed by (12). This modification has been demonstrated to be effective in identifying the so-called Sobol’ sensitivity indices. These indices quantify both the main sensitivity effects (first-order effects) (the contribution to the variance of the model output by each input variables) and total sensitivity effects (the first-order effect plus interactions with other input variables) of input variables (13). This method has been implemented into the GSA toolbox. A description according to (13) is outlined below.

Formally, given a model $Y = f(X)$, where Y is the model output, $X = (X_1, X_2, \dots, X_k)$ is the input variable vector. A variance decomposition of f suggested by (11) is:

$$V(Y) = \sum_{i=1}^k V_i + \sum_{i=1}^k \sum_{j=i+1}^k V_{ij} + \dots + V_{1,\dots,k} \quad (1)$$

where X is rescaled to a k -dimensional unit hypercube Ω^k , where $\Omega^k = |X|$, $0 \leq X_i \leq 1$, and $i = 1, \dots, k$; $V(Y)$ is the total unconditional variance; V_i is the partial variance or ‘main effect’ of X_i on Y and given by $V_i = V[E(Y|X_i)]$; V_{ij} is the joint impact of X_i and X_j on the total variance minus their first-order effects.

Here, the first-order sensitivity index S_i (Eq. 2) and total effect sensitivity index S_{T_i} (Eq. 3) are given as (6):

$$S_i = \frac{V_i}{V(Y)} = \frac{V[E(Y|X_i)]}{V(Y)} \quad (2)$$

$$S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \dots = \frac{E[V(Y|X_{-i})]}{V(Y)} \quad (3)$$

where X_{-i} denotes variation in all input variables and X_i , S_{ij} is the contribution to the total variance by the interactions between variables.

Following (12), to compute S_i and S_{T_i} , two independent input variable sampling matrices P and Q with dimensions (N,k) were created, where N is the sample size and k is the number of input variables. Each row in matrix P and Q represents a possible value of X . The variable ranges in the matrices are scaled between 0 and 1. The Monte Carlo approximations for $V(Y)$, S_i (Eq. 6) and S_{T_i} (Eq. 7) are defined as follows (8,12):

$$\hat{f}_o = \frac{1}{N} \sum_{j=1}^N f(P)_j \quad (4)$$

$$\hat{V}(Y) = \frac{1}{N} \sum_{j=1}^N (f(P)_j)^2 - \hat{f}_o^2 \quad (5)$$

$$\hat{S}_i = \frac{1}{N} \sum_{j=1}^N \frac{f(Q)_j (f(P_Q^{(i)}) - f(P)_j)}{\hat{V}(Y)} \quad (6)$$

$$\hat{S}_{T_i} = \frac{1}{2N} \sum_{j=1}^N \frac{(f(P)_j - f(P_Q^{(i)})_j)^2}{\hat{V}(Y)} \quad (7)$$

Where $\hat{\cdot}$ is the estimate; \hat{f}_o is the estimated value of the model output; $P_Q^{(i)}$ represents all columns from P except the i^{th} column, which is from Q , using a radial sampling scheme (6). To compute S_i and S_{T_i} simultaneously, a scheme suggested by (14) was used which reduced the model runs to $N(k+2)$.

To sample the P and Q matrices, the Sobol' quasi-random sampling sequence, also known as LP-s, or LPTAU (15), was used. This sequence helps to distribute the sampling points as uniformly as possible in the variable space to avoid clustering and increases the convergence rate (6). Therefore, the use of these sequences enhances the convergence of the Monte Carlo integrals. While the Monte Carlo integration, and thus the Sobol' sensitivity analysis, normally converges at a rate of $1/\sqrt{n}$, the Sobol' quasi-random sampling converges this to almost $1/n$ (8).

ARTMO

ARTMO brings multiple leaf and canopy radiative transfer models (RTMs) together in one modular toolbox, along with essential tools required for semi-automatic retrieval of biophysical variables. ARTMO is freely downloadable at <http://ipl.uv.es/artmo/>. In short, the toolbox permits the user:

- i. to run various invertible leaf and canopy RTMs of a low to high complexity (e.g., PROSPECT-4/5 (16), LIBERTY (17), SAIL (18), FLIGHT (19), INFORM (20), SCOPE (21))
- ii. to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems

- iii. to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400 to 2500 nm)
- iv. to generate look-up tables (LUTs), which are directly stored in a relational SQL database management system (MySQL, version 5.5 or higher; local instalment required), and finally
- v. to configure and run various retrieval scenarios using EO reflectance datasets for biophysical variable mapping applications.

ARTMO is developed in Matlab (2011 version or higher). Figure 1 presents ARTMO's v. 3.16 main window and a systematic overview of the drop-down menu below. To start with, in the main window a new project can be initiated, a sensor chosen and a comment added, whereas all processing modules are accessible by means of drop-down menus at the top bar.

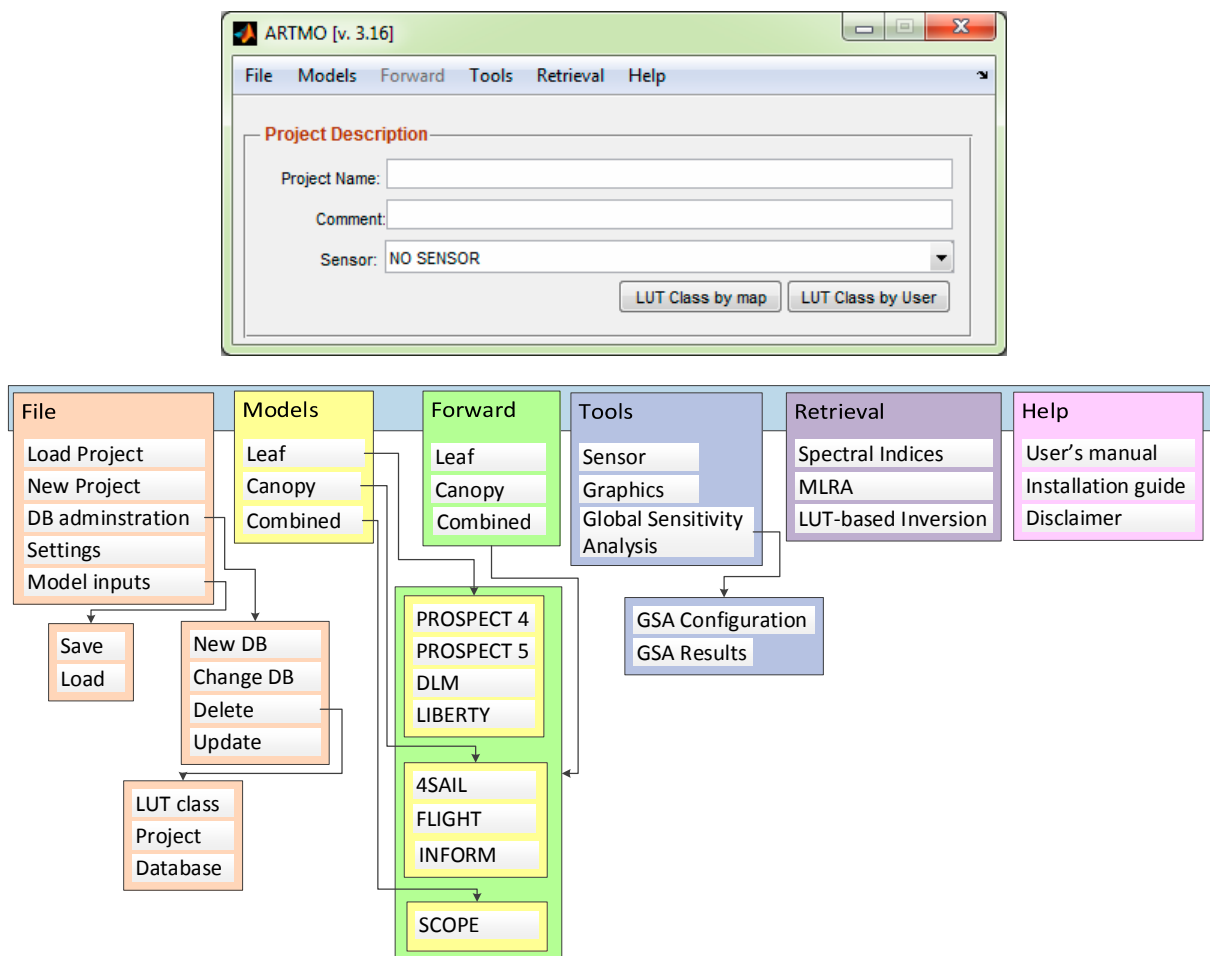


Figure 1: Screenshot of ARTMO's v3.16 main window (top) and schematic overview of its drop-down menu (bottom).

A first rudimentary version of ARTMO was used in LUT-based inversion applications (9). The toolbox has been improved and expanded since then, with most importantly the implementation of retrieval toolboxes. These toolboxes are based on parametric and non-parametric regression as well as physically-based inversion using a LUT. They resulted in a:

- i. 'Spectral Indices assessment toolbox' (22)
- ii. 'Machine Learning Regression Algorithm (MLRA) toolbox' (23); and
- iii. 'LUT-based inversion toolbox' (24).

ARTMO v3.16 with the Global Sensitivity Analysis (GSA) Toolbox' (v. 1.01) as most important novelty is formally presented in this paper.

GSA toolbox

The GSA toolbox essentially consists of two modules:

- i. the ‘**GSA configuration**’ module
- ii. the ‘**GSA results**’ visualization module.

Both modules are outlined below.

a. GSA configuration module

The core component that enables us to select an RTM and configure the GSA is displayed in Figure 2. The toolbox can be divided into three parts: (i) Main settings, (ii) RTM input settings, and (iii) RTM output settings.

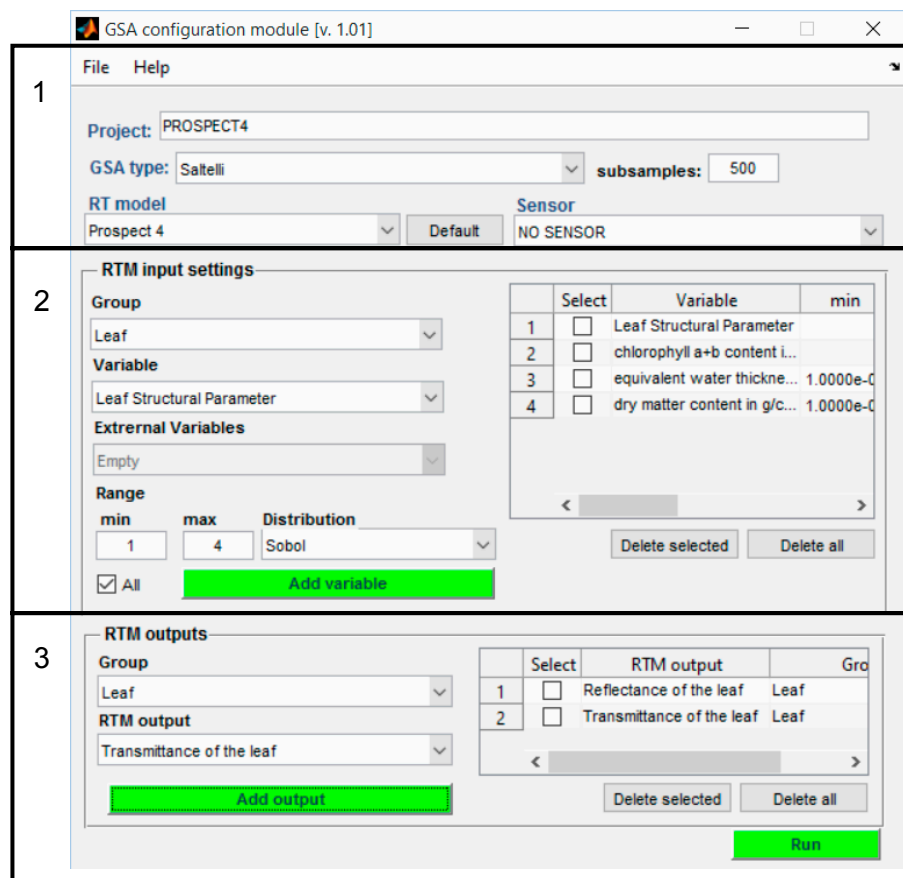


Figure 2: Screenshot of GSA configuration module.

In the first panel, the main GSA configuration settings need to be given. These include: (i) to give a project name (e.g. ‘PROSPECT4’), (ii) to choose a GSA method, (iii) to choose the number of samples per variable, (iv) to choose the RT model to be analyzed, and optionally: (v) to choose for resampling of the spectra to a sensor setting. These general settings are briefly outlined below:

- i. **Name:** The GSA toolbox stores its results in a MySQL database. By giving a name the main information is stored in a MySQL overview table. GSA results are stored in an associated table.
- ii. **GSA method:** Only the GSA method described in (12) has been successfully implemented. This method was proved to be effective in several studies (e.g., 4,7,8,13). Other tested methods (eFAST, Sobol) did not lead to consistent results and were therefore discarded.
- iii. **Number of samples:** Here, the number of samples (N) per variable (k) is given. Following (14), in total $N(k + 2)$ simulations are generated.

- iv. **RTM:** The following RTMs can be analysed with the GSA toolbox: PROSPECT4, PROSPECT5, DLM, LIBERTY, PROSPECT4-SAIL, PROSPECT5-SAIL, LIBERTY-SAIL, PROSPECT4-INFORM, PROSPECT5-INFORM, LIBERTY-INFORM, SCOPE. Note hereby that a canopy model is always coupled to a leaf model, because the leaf model provides necessary input data to a canopy model.
- v. **Sensor:** By selecting a sensor the GSA results will be given for the spectral bands according to the chosen optical sensor. In ARTMO's Sensor module, any kind of sensor within the 400-2400 nm (range defined by the leaf models) can be inserted. This option allows us to identify driving RTM variables for spectral bands as configured by a sensor. By default the following sensors are provided: LANDSAT-7, SPOT-4, CHRIS-M3, MODIS, MERIS, Sentinel-2, Sentinel-3 OLCI & SLSTR, Landsat-8, Pleiades-1A and QuickBird. Alternatively, users can also define their own sensor settings within ARTMO's Sensor module.

In the second panel, the RTM input setting can be configured. Depending on the complexity of the RTM, inputs are organized per 'Group'. This is especially the case with the SVAT model SCOPE, which consists of several sub-models. Otherwise, 'Inputs' is given. The RTM input variables are listed in the scroll-down menu below. The input variables have to be selected one-by-one, and for each variable the minimum and maximum boundaries and a data distribution have to be given. By default the Sobol' quasi-random sequence is provided, but also a uniform, normal and exponential distribution can be selected. By clicking on 'Add variable' the variable with its boundaries is entered into the right panel. The boundaries can be modified and inserted variables can be removed.

The bottom panel enables selection of the RTM outputs. Often, RTMs provide more than one output variable, usually spectral output, but it can also be fluxes (i.e., point outputs). The toolbox allows the relative contribution of the input variables to multiple outputs to be analysed at once. Hence, the user can select multiple outputs. They will appear in the right-bottom panel. Finally, by clicking on 'Run' the toolbox starts running the GSA. Depending on the RTM and given number of samples this can take a while (e.g., a few days in case of SCOPE).

b. GSA results

Once the analysis is completed, the 'GSA results' window will appear (Figure 3). This window shows the Project name and the other general settings (GSA type, RT model, # subsamples and Sensor). The panel below lists the inserted variables and their boundaries and sampling distribution. Finally, the analysed RTM output variable can be chosen in the bottom panel. First order (S_i) and total order (S_{Ti}) Sobol' sensitivity indices can be plotted for inspection or exported. To facilitate the interpretation of obtained results across the spectral domain, it is recommended to normalize the GSA results, i.e. the sum of all relative contributions is 100%.

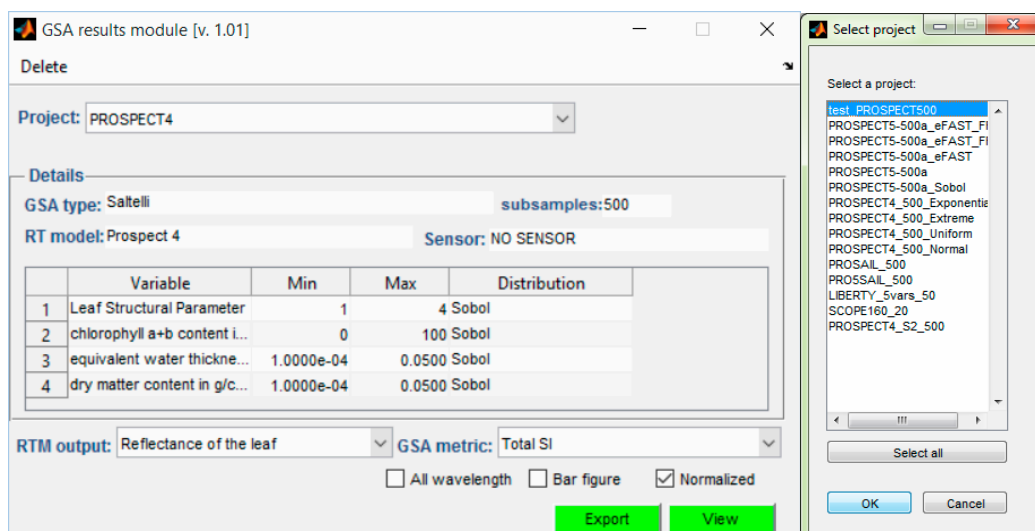


Figure 3: Screenshot of GSA Results module (left) and overview of Projects to be deleted (right).

In the GUI, it is also possible to view earlier projects by scrolling down the Project and list, and to delete earlier GSA exercises. Also, when clicking on ‘Delete’, a GUI appears that provides an overview of all GSA Projects as stored in MySQL. Selected projects are then deleted.

CASE STUDIES

Having the toolbox outlined, some illustrative case studies will be given here. Total sensitivity indices (S_{Ti}) will be calculated for leaf RTMs (PROSPECT-4, PROSPECT-5) and leaf and canopy RTMs (PROSPECT-4/5 + SAIL). The impact of different sampling distributions on the GSA outputs was firstly tested. These distributions hardly have an impact on GSA results, therefore, only the default Sobol semi-quasi sampling sequence was further used. The GSA analysis was set to sample each input variable 500 times. RTM input ranges for PROSPECT-4 and -5 (two additional variables) and SAIL are provided in Table 1.

Table 1: PROSPECT-4, -5 and SAIL input variables boundaries.

Model Variable	Min.	Max.
<i>Leaf: PROSPECT</i>		
4	Leaf Structural variable (N)	1 4
	chlorophyll a+b content in $\mu\text{g}/\text{cm}^2$ (Chl)	0 100
	equivalent water thickness in g/cm^2 or cm (Cw)	0.0001 0.05
	dry matter content in g/cm^2 (Cd)	0.0001 0.05
5	Carotenoids in $\mu\text{g}/\text{cm}^2$ (Car)	0 10
	brown pigments in g/cm^2 (Bp)	0 10
<i>Canopy: SAIL</i>		
	Total Leaf Area Index (LAI)	0 10
	Leaf angle distribution (LAD)	0 90
	Diffuse/direct light	0 100
	Soil Coefficient	0 1
	Solar Zenith Angle (SZA)	0 90
	Relative Azimuth Angle (RAA)	0 180

RESULTS

Since the total-order effect (S_{Ti}) of a variable includes the first-order plus all the interaction effects, it is the preferred sensitivity indicator to identify the dominant and negligible input variables. S_{Ti} results on leaf reflectance and transmittance of PROSPECT-4/5 along the 400-2400 nm region are given in Figure 4. The relative contributions of each input variable to the selected model output can be inspected. For instance, the figure reveals the role of chlorophyll content in the visible part and the importance of the structural variables N and dry matter content across the whole spectral range. Leaf water content only governs reflectance and transmittance from around 1200 nm onwards. For PROSPECT-5, the visible part is also strongly governed by brown pigments. Carotenoids play a smaller role, only in the 400-550 nm range.

When coupling PROSPECT with the canopy model SAIL, both leaf and canopy variables shape the variability of the spectral output. As can be observed in Figure 5, LAI is the prime driving variable. LAI alone can explain up to 40% of the total variability (i.e. with interactions). Other important variables are leaf angle distribution (LAD), soil coefficient and solar zenith angle. GSA also identifies the difference between directional reflectance and hemispherical reflectance. For hemispherical reflectance, the solar zenith angle is more important, as well the variable that controls diffuse/direct light.

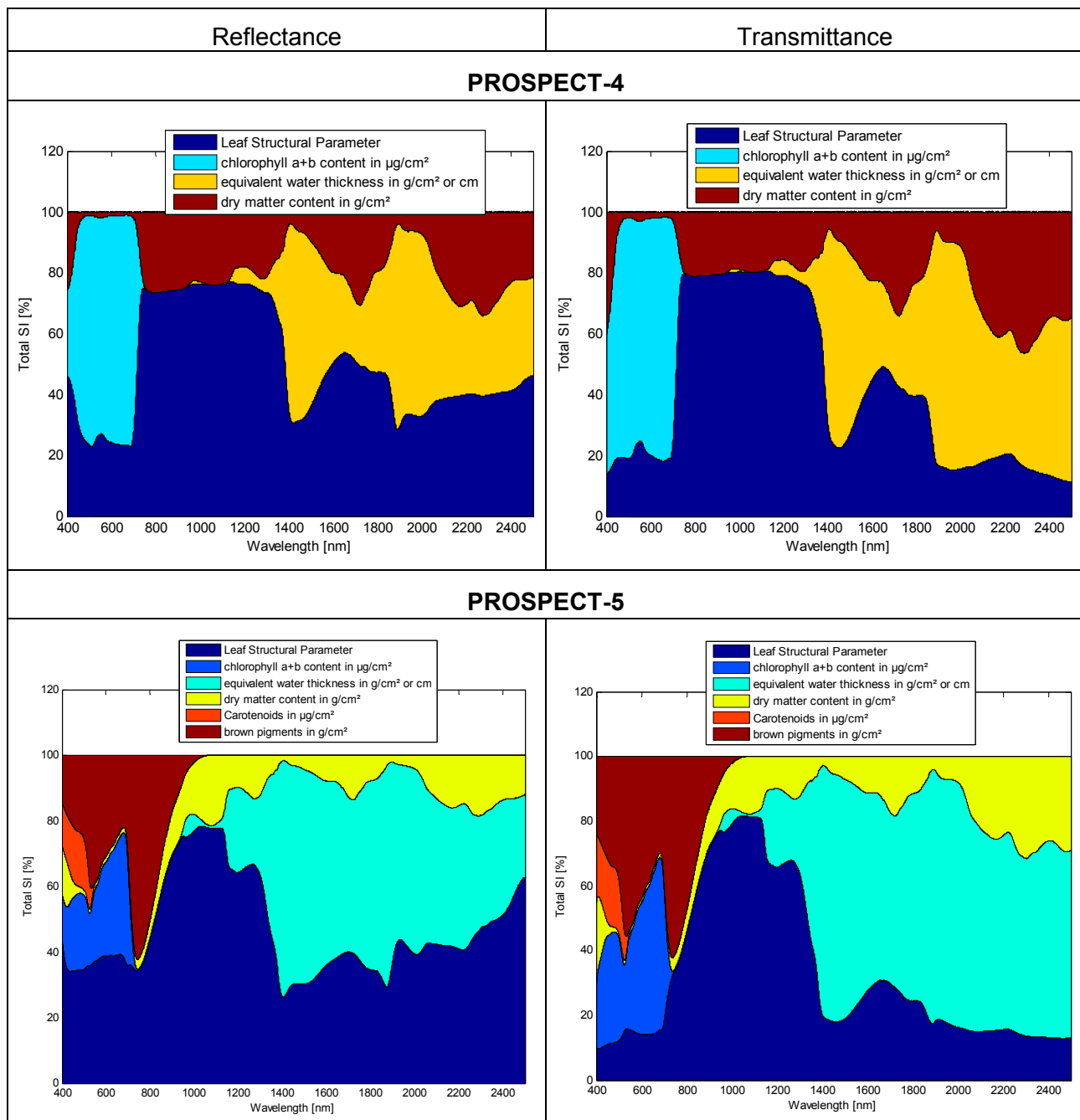


Figure 4: S_{T_i} results for PROSPECT-4, -5 leaf reflectance and transmittance outputs.

DISCUSSION

Global sensitivity analysis (GSA) evaluates the relative importance of each input variable in a model and can be used to identify the most influential variables affecting model outputs. GSA can be applied with radiative transfer models (RTMs) to identify the key determinants of spectral outputs such as fluorescence, reflectance, radiance. Non-influential variables can also be identified and be safely set to arbitrary values under relatively wide ranges of conditions. While this methodology has earlier identified the spectral dependency of driving variables of popular RTMs (e.g., 4,25), until now, no GSA software packages readily applicable to RTMs are available to the broader remote sensing community.

