California State University, Monterey Bay

Digital Commons @ CSUMB

6-2023

Development of the Ames Global Hyperspectral Synthetic Data Set: Surface Bidirectional Reflectance Distribution Function

Weile Wang
NASA Ames Research Center

Jennifer Dungan
NASA Ames Research Center

Vanessa Genovese
California State University, Monterey Bay

Yohei Shinozuka
Bay Area Environmental Research Institute

Qiguang Yang
Science Systems and Applications, Inc. (SSAI)

See next page for additional authors

Follow this and additional works at: https://digitalcommons.csumb.edu/aes_fac

Authors

Weile Wang, Jennifer Dungan, Vanessa Genovese, Yohei Shinozuka, Qiguang Yang, Xu Liu, Benjamin Poulter, and Ian Brosnan

Development of the Ames Global Hyperspectral Synthetic Data Set: Surface Bidirectional Reflectance Distribution Function

Weile Wang[1] , Jennifer Dungan[1], Vanessa Genovese[2], Yohei Shinozuka[3], Qiguang Yang[4], Xu Liu[5] , Benjamin Poulter[6] , and Ian Brosnan[1]

[1]NASA Ames Research Center, Moffett Field, CA, USA, [2]California State University Monterey Bay, Seaside, CA, USA, [3]Bay Area Environmental Research Institute, Moffett Field, CA, USA, [4]Science Systems and Applications, Inc. (SSAI), Hampton, VA, USA, [5]NASA Langley Research Center, Hampton, VA, USA, [6]NASA Goddard Space Flight Center, Greenbelt, MD, USA

Abstract This study introduces the Ames Global Hyperspectral Synthetic Data set (AGHSD), in particular the surface bidirectional reflectance distribution function (BRDF) product, to support the NASA Surface Biology and Geology (SBG) mission development. The data set is generated based on the corresponding multispectral BRDF products from NASA's MODIS satellite sensor. Based on theories of radiative transfer in vegetation canopies, we derive a simple but robust relationship that indicates that the hyperspectral surface BRDF can be accurately approximated as a weighted sum of the soil surface reflectance, the leaf single albedo, and the canopy scattering coefficient, where the weights or coefficients are spectrally invariant and thus readily estimated from the multispectral MODIS products. We validate the algorithm with simulations by a Monte Carlo Ray Tracing model and find the results highly consistent with the theoretic derivation. Using reflectance spectra of soil and vegetation derived from existing spectral libraries, we apply the algorithm to generate the AGHSD BRDF product at 1 km and 8-day resolutions for the year of 2019. The data set is biogeochemically and biogeophysically coherent and consistent, and serves the goal to support the SBG community in developing sciences and applications for the future global imaging spectroscopy mission.

Plain Language Summary This paper develops a emulated (or synthetic) global hyperspectral data set, based on existing multispectral satellite products, to support the NASA Surface Biology and Geology (SBG) mission development. The challenge of the task is to find a simple yet robust algorithm to interpolate multispectral observations to (unobserved) hyperspectral bands. Based on theoretic analyses and numerical model simulations, we develop an accurate and efficient algorithm that accomplishes this task by combining information from the optical properties of the surface elements and the structural properties of the landscape and vegetation canopies. The former are obtained from existing spectral libraries and the latter can be estimated from the multispectral satellite products. We apply the algorithm to generate a scientifically coherent and consistent global hyperspectral data set, which are publicly available to serve the needs of the science community of SBG and other global hyperspectral remote sensing satellite missions.

1. Introduction

Surface Biology and Geology (SBG) is a forthcoming NASA mission recommended by the 2017–2027 National Academies' Decadal Survey (NASEM, 2018), as one of the high priority Designated Observables, to acquire hyperspectral measurements in the visible to shortwave infrared (VSWIR, 380–2,500 nm) range and multispectral measurements in the thermal infrared (TIR, 4–12 μm) range with a sub-monthly revisit cadence over the entire globe (Cawse-Nicholson et al., 2021; Schimel & Poulter, 2022; Thompson et al., 2021). The development of SBG continues the legacy from EO-1/Hyperion (Folkman et al., 2001), the NASA HyspIRI pathfinder study (Lee et al., 2015), the airborne AVIRIS/AVIRIS-NG instrument (Hamlin et al., 2011), and the latest EMIT mission (R. O. Green et al., 2020), joined by steadily growing international hyperspectral missions like the HISUI (Matsunaga et al., 2017), DESIS (Krutz et al., 2019), PRISMA (Loizzo et al., 2018), EnMAP (Guanter et al., 2015), and the future CHIME (Rast et al., 2021). Observations of SBG are expected to help significantly improve our understanding of global ecology, hydrology, weather, climate, as well as solid earth (NASEM, 2018).

One of the scientific goals—and grand challenges—of SBG is to retrieve vegetation functional traits at the global scale. Such plant functional traits, including pigment and nutrient concentrations, leaf/canopy structure and

Methodology: Weile Wang, Qiguang Yang, Xu Liu
Project Administration: Jennifer Dungan, Benjamin Poulter
Resources: Jennifer Dungan, Benjamin Poulter
Software: Yohei Shinozuka, Qiguang Yang, Xu Liu
Supervision: Ian Brosnan
Validation: Vanessa Genovese
Visualization: Vanessa Genovese
Writing – original draft: Weile Wang
Writing – review & editing: Ian Brosnan

morphology, and metabolic characteristics (Serbin & Townsend, 2020), are drivers and indicators of vegetation physiological functions, community assembly, and species diversities (Cavender-Bares et al., 2020; Kunstler et al., 2016). However, spectral reflectance observed by remote sensors is not only shaped by the biogeochemical properties of leaves but also influenced by the biogeophysical properties (e.g., leaf distribution and plant structure) of the canopy as well as the background environment (e.g., soil and surface topography). The transportation of light in vegetation canopies can modulate the directional distribution of photons, introduce surface reflectance anisotropy, and obscure the leaf spectral characteristics. Vegetation retrieval algorithms, especially those intended for global scales, thus must account for the effects of radiative transfer in the imaging spectroscopy data, which requires us to gain a deeper understanding of light transmissions at the land surface.

The anisotropy of surface reflectance is described by the bidirectional reflectance distribution function (BRDF) or BRDF. BRDF is a fundamental optical property of the surface, which intrinsically modulates the spatial and angular distributions of the radiation signals reflected by the surface. The distrubutions of the signals are further transformed and complicated by the radiative transfer processes through the atmosphere. Therefore, measurements of the reflected solar radiation by satellite or airborne instruments are influenced by the surface BRDF effects to various degrees. Depending on the requirement of the applications, the BRDF effects on the measurements may be undesirable and need to be "corrected." More challengingly, because accurate knowledge of surface BRDF is difficult to obtain, atmospheric correction algorithms such as ATREM (Gao et al., 1993) and ISOFIT (Thompson et al., 2018) often make simplifications that the surface scatters incident radiation equally in all directions. Such an idealized surface is called "Lambertian." The Lambertian assumption inevitably introduce uncertainties in the retrieved surface reflectance and other downstream products. To meet the requirement of the SBG scientific objectives, we must quantitatively evaluate the BRDF effects and the associated uncertainties on the hyperspectral measurements and derived science products.

The Modeling Working Group (ModWG) of SBG Research and Applications team was established to inform the SBG mission development and support the science and applications traceability with wall-to-wall model simulations (Schimel & Poulter, 2022). A key task of ModWG is to generate a global synthetic hyperspectral data set spanning from surface reflectance to top-of-atmosphere (TOA) radiances for the development and testing of community science algorithms and applications. The simulation of synthetic data is essentially a forward modeling process in which we examine how plant functional traits translate into spectral signals transferring from the surface through the atmosphere to the remote sensor. Therefore, the synthetic data are not necessarily "true" (i.e., reflecting the ground truth), but should be "reasonable" (i.e., reflecting the model truth).

Although the task of synthetic data generation can be accomplished with process-based vegetation and radiative transfer models (Chernetskiy et al., 2018; Poulter et al., 2023), the computational costs associated with such global model simulations at landscape resolutions are very expensive. In this study, therefore, we seek a physics-based but computationally efficient approach to address the task. The result of this study is the Ames Global Hyperspectral Synthetic Data set (AGHSD) set or AGHSD. We first develop a biogeophyscially and biogeochemically coherent algorithm to generate global hyperspectral surface BRDF product based on the corresponding multispectral MODIS products. We then propagate the signals to TOA reflectance with the help of atmospheric radiative transfer models. This paper introduces the land surface BRDF product of AGHSD, focusing on its algorithm development and validation. We will introduce the corresponding TOA product in an accompanying paper. The development of synthetic data sets over ocean and inland water surfaces is led by another team at NASA Ames Research Center, who will also report their progress in separate papers.

To give a brief background, BRDF is typically defined as the ratio between the light flux (L_λ) reflected from the surface in the view direction (described by the solid angle Ω) to the parallel light beam ($E_{\lambda,0}$) incident on the surface in the illumination direction (Ω_0) Martonchik et al., 2000 that is,

$$\rho_\lambda(\Omega_0, \Omega) = \frac{L_\lambda(\Omega)}{E_{\lambda,0}(\Omega_0)} \tag{1}$$

where $E_{\lambda,0}$ is irradiance ($w \cdot m^{-2}$), L_λ is radiance ($w \cdot m^{-2} \cdot sr^{-1}$), and the subscript "$\lambda$" indicates that these variables are wavelength dependent. The BRDF ρ_λ is generally anisotropic and varies with the illumination-view geometry. Theoretical BRDF models can be very complex and thus are rarely used in practical studies. Instead, remote sensing applications commonly use kernel-based semiempirical BRDF models. For instance, the MODIS BRDF products adopts the Ross-thick Li-sparse (RTLS) model (Lucht et al., 2000; Schaaf et al., 2002; Wanner

et al., 1997), which represents surface BDRF (ρ_λ) as the sum of Lambertian (isotropic), volume scattering, and geometric optical components

$$\rho_\lambda(\Omega_0, \Omega) = k_\lambda^L + k_\lambda^V f_V(\Omega_0, \Omega) + k_\lambda^G f_G(\Omega_0, \Omega). \tag{2}$$

where f_V and f_G are the kernel functions for volume scattering and geometric optical components (the Lambertian kernel is normalized to be 1), and $\boldsymbol{K}_\lambda = \left[k_\lambda^L, k_\lambda^V, k_\lambda^G\right]^T$ are the coefficients or weights for the corresponding components. Note that the kernel functions are purely determined by the illumination-view geometry and thus spectrally invariant. Fitting the surface BRDF with the RTLS model is thus to determine the coefficient vector $\boldsymbol{K}_\lambda$.

Existing multispectral BRDF products from satellite sensors like MODIS and VIIRS provide global estimates of $\boldsymbol{K}_\lambda$ at several spectral bands in the VSWIR range. To derive synthetic hyperspectral surface BRDF is essentially to identify an interpolation scheme that estimate $\boldsymbol{K}_\lambda$ at the hyperspectral bands from the corresponding multi-band products. There is rich literature about the interpolation/reconstruction of hyperspectral reflectance from the corresponding multispectral data with statistical or machine learning methods (Guanter et al., 2009; Liu et al., 2009; Rivera et al., 2015; Sun et al., 2014; Verrelst et al., 2019). These methods generally compare the multispectral $\boldsymbol{K}_\lambda^M$ (the superscript "M" denotes the multispectral cases) to a library of hyperspectral $\boldsymbol{K}_\lambda^H$ (the superscript "H" denotes the hyperspectral cases) at the corresponding spectral bands and choose the best matches ($\boldsymbol{K}_\lambda^H$) as the solution to the problem.

The hyperspectral library of $\boldsymbol{K}_\lambda^H$ can be constructed from existing libraries of reflectance (ρ_λ^H) such as the USGS Spectral Library (Kokaly et al., 2017) and the Aster/ECOSTRESS Spectral Library (Baldridge et al., 2009). For instance, the original derivation of the RTLS kernels can be used to relate the $\boldsymbol{K}_\lambda$ parameters to the reflectance spectra of multiple landscape elements (horizontal soil surface, the scattering facets, and the sunlit crown), the vegetation leaf area index, the area proportion of vegetation coverage, and a couple of other variables (Lucht et al., 2000). A previous study by Q. Yang et al. (2020) selected a subset of ~900 reflectance spectra from the Aster Spectral Library (Baldridge et al., 2009) and created a library of ~40,000 BRDF spectra. These spectra can be conveniently and accurately represented by the leading ~20 principal components (PCs), which allows Q. Yang et al. (2020) derive an efficient formula to calculate hyperspectral spectra from the MODIS BRDF parameters $\boldsymbol{K}_\lambda^M$ at seven bands.

However, the algorithm of Q. Yang et al. (2020) is purely empirical and there are important scientific questions that are left unanswered. The spectra samples from the Aster (or USGS) library reflect optical properties of small elements (e.g., leaves) measured under laboratory settings, while satellite BRDF data represent the "convoluted" signals contributed by all elements over landscape scales. There are strong interests in gaining deeper understanding of how the two quantities are physically coupled in the scaling processes. On one hand, we would like to understand how the landscape patterns and structures regulate the leaf-level spectra into the satellite measured BRDF characteristics. On the other hand, this knowledge will also be very useful for us to evaluate (and improve) inverse algorithms that de-convolute satellite BRDF signals to retrieve optical properties of vegetation leaves, soil surfaces, and other elements of the landscape.

Therefore, this study introduces a physics-based algorithm that generates synthetic satellite BRDF from landscape element optical properties and structural characteristics. The algorithm is rigorously derived with the theories of radiative transfer in vegetation canopy. It has a simple linear form as reported in Q. Yang et al. (2020), but different from the previous study, the coefficients of the formula have physical explanations. We validate the algorithm with simulation results from Monte Carlo Ray Tracing model experiments. The algorithm is then applied to generate global hyperspectral synthetic BRDF data set for 2019 based on MODIS products. We will also analyze the results and discuss possible ways of improvement.

2. Data and Methods

2.1. MODIS BRDF Data Sets

The global multi-band BRDF data set used in this study is the MODIS MCD19A3 product generated by the NASA Multi-Angle Implementation Atmospheric Correction (MAIAC) algorithm (Lyapustin et al., 2018). The product provides estimates of the RTLS BRDF parameters ($\boldsymbol{K}_\lambda$) at an 8-day frequency and a 1 km spatial resolution for

Figure 1. An example true-color image of global land surface (60°N–60°S) generated with MODIS MCD19A3 bidirectional reflectance distribution function parameters collected on 4 July 2019. For each pixel, the solar zenith, the view zenith, and the relative azimuth angles are 20°, 45°, and 90°, respectively.

the first eight MODIS bands, including the seven land bands (Band 1 to Band 7) and the first ocean color band (Band 8). For this study we only use the K_λ parameters of the seven land bands. Figure 1 shows an example image of global land surface between 60°N and 60°S calculated with the MCD19A3 BRDF parameters collected on 4 July (Julian Day 185), 2019. The reflectance at each pixel is normalized to the same illumination-view geometry, where the solar zenith, the view zenith, and the relative azimuth angles are 20°, 45°, and 90°, respectively.

2.2. Monte Carlo Radiative Transfer Model

1. We use the radiative transfer model Forest Light Environmental Simulator (FLiES) (Kobayashi & Iwabuchi, 2008) to simulate the transportation of photons in 3-D vegetation canopies over heterogeneous landscapes. FLiES uses a Monte Carlo Ray Tracing approach to simulate the scatterings and absorption of photons between leaves, trunks, and soil surfaces. The current FLiES code considers a rectangular forest stand with trees, shrubs, and grasses distributed in both horizontal and vertical dimensions. Trees are represented by geometric objects in four different shapes (cones, cylinders, spheroids, and semi-spheroids). The shape and the size of the crowns as well as the horizontal location and the height of the trees are specified for each individual trees. The optical properties of plants are specified by "groups," where trees in the same group have the same spectral reflectance and/or transmittance for leaves and barks. FLiES supports up to four different tree groups in a forest stand and uses an addition group to describe the properties of the forest floor vegetation (e.g., grasses). The specification of a tree group also includes canopy structural information such as the leaf area density and branch area index. The performance and robustness of FLiES for simulating bidirectional reflectance factors have been demonstrated in previous studies (Widlowski Jean-Luc et al., 2015).

We have conducted model sensitivity experiments with various combinations of spectral optical properties of leaves, barks, and soil surfaces, landscape/forest structures, as well as model parameters (see Appendix A for details). Figure 2a shows an example of the forest stand we generated for FLiES. The horizontal locations of the trees are generated by the Poisson's distribution. For simplicity, in this example we specify only one group of trees with spheroidal crowns, where the ratios between the radius (in the horizontal dimensions) and the height (in the vertical dimension) of the tree crown are about ∼0.5. The crowns largely fall into three size classes (3, 10, and 20 m) and the tree heights are approximately proportional to the crown sizes. The leaf area density of the tree canopies is set to 1 and the overall area ratio of tree crowns to the total spatial domain is ∼0.7.

Figure 2b shows the optical properties, including leaf reflectance and transmittance, bark reflectance, and soil reflectance, of the forest stand. The spectra of the optical properties range from 400 to 2,400 nm at a 10-nm resolution (e.g., 201 spectral bands). The leaf reflectance and the leaf transmittance are usually very similar to each other, and their sum is called as leaf single albedo (ω_l). The leaf optical properties are simulated with the PROSPECT model (Jacquemoud & Baret, 1990) with typical leaf chemistry parameters. The soil reflectance is a representative example from the Brightness-Shape-Moisture model (P. Yang et al., 2020). The bark reflectance is from the Aster/ECOSTRESS Spectral Library (Baldridge et al., 2009; Meerdink et al., 2019). Figure 2b also shows the Spectral Response Function of the first eight MODIS bands. We use SRFs to sample and integrate the simulated hyperspectral data into the corresponding MODIS bands so that they can be used in the regression analysis.
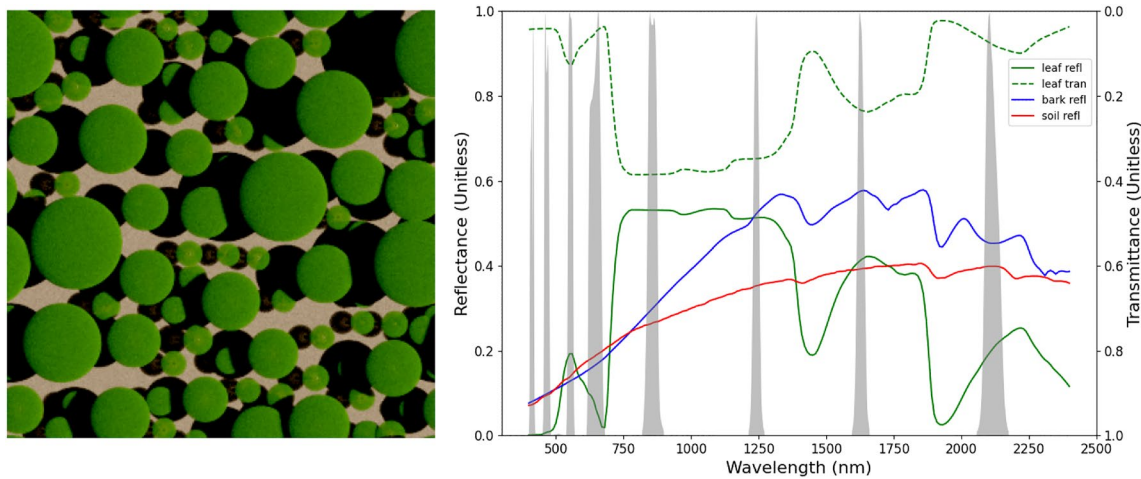
Figure 2. (a) A forest stand example generated using the ellipsoidal tree models. The boundary of the site is set to be periodic so that the spatial pattern of the trees is supposed to repeat in horizontal dimensions. (b) The optical properties of the forest elements shown in panel (a). Note that all the trees in this example are in the same group and we have neglected the grass on the forest floor. The gray shades in panel (b) shows the Spectral Response Function of the first eight MODIS bands, which are used to sample and integrate the hyperspectral data into the corresponding MODIS bands.

2.3. AGHSD Algorithm Development

Our hypothesis in developing the AGHSD algorithm is that the RTLS parameterization of hyperspectral surface BRDF can be accurately approximated as a weighted sum of the soil surface reflectance ($\omega_{\lambda,s}$), the leaf single albedo ($\omega_{\lambda,l}$), and the canopy scattering coefficient ($\omega_{\lambda,c}$).

The spectra of soil surface reflectance and leaf single albedo are well understood and illustrated in Figure 2b. The concept of canopy scattering coefficient is relatively new and needs some explanation. Simply speaking, the canopy scattering coefficient is an up-scaled version of leaf single albedo: it measures the probability that a photon intercepted by a vegetation canopy will be scattered by the canopy (Knyazikhin et al., 2013; Smolander & Stenberg, 2005; W. Wang et al., 2018). This variable is related to both the leaf single albedo and the structural characteristics of the canopy. When a photon enters the canopy and collides with a leaf element, it will be scattered with a probability of ω_l (leaf single albedo). The scattered photon will either recollides with another leaf element or escapes the canopy through a gap. The process iterates until the photon finally escapes or is absorbed the canopy. We define the probability that the scattered photon recollides with another leaf element as the recollision probability. If the recollision probability has a stable value p (see below for explanation), the canopy scattering coefficient can then be expressed as a geometric series

$$\omega_{\lambda,c} = \omega_{\lambda,l}(1-p)\left[1 + \omega_{\lambda,l}p + (\omega_{\lambda,l}p)^2 + (\omega_{\lambda,l}p)^3 + \cdots\right] = \frac{\omega_{\lambda,l}(1-p)}{1 - \omega_{\lambda,l}p} \tag{3}$$

where $1 - p$ represents the escape probability.

It is easy to see that the scattered photons in the vegetation canopy must be spatially distributed in a certain way so that the statistical mean of their recollision probabilities takes the stable value p. Fortunately, the theory of radiative transfer indicates that the spatial distribution of the scattered photons will converge to this characteristic form (which is intrinsic to the given canopy) as the order of scattering increases (Huang et al., 2007). Simulations with 3D Monte Carlo Radiative Transfer models indicate the scattering process converges rapidly and the recollision probability becomes stable at the second (and higher order) scattering (Lewis & Disney, 2007), which is also verified by the FLiES experiments in this study (see the Appendix A). Therefore, we can decompose the bidirectional reflectance of vegetation canopy into two components. The first scattering component includes photons scattered by leaves only once, which is thus associated with the leaf single albedo ($\omega_{\lambda,l}$). The multiple scattering component includes photons scattered twice or more times by leaf elements before they escape the canopy, which is described by the canopy scattering coefficient (or more rigorously, the product of canopy scattering coefficient and the single albedo, $\omega_{\lambda,l} \cdot \omega_{\lambda,c}$).

Not all scattered photons are reflected by vegetation canopies. Some of them are directly reflected by the soil surface and therefore modulated by the soil reflectance $\omega_{\lambda,s}$. The total surface BRDF thus must include the soil surface reflectance as the third component, in addition to the two vegetation components. Therefore, we represent the surface BRDF as

$$\rho_\lambda(\Omega_0, \Omega) = \omega_{\lambda,s} \cdot F_s(\Omega_0, \Omega) + \omega_{\lambda,l} \cdot F_l(\Omega_0, \Omega) + \frac{\omega_{\lambda,l}^2 (1-p)}{1 - \omega_{\lambda,l} p} \cdot F_c(\Omega_0, \Omega) \tag{4}$$

Note that here we neglect the multiple scatterings between the atmosphere and the soil surface as well as those between the canopy and the soil surface.

The $F_s(\Omega_0, \Omega)$, $F_l(\Omega_0, \Omega)$, and $F_c(\Omega_0, \Omega)$ terms in Equation 4 denote the bidirectional scattering factors for the corresponding components. These factors are determined by the structural characteristics of the land surface elements. Like the kernel functions of the RTLS model, they are functions of the illumination-view geometry but not the wavelength. By expanding these factors with the RTLS kernel functions, we transform Equation 4 into

$$\rho_\lambda(\Omega_0, \Omega) = \begin{bmatrix} 1 & f_V & f_G \end{bmatrix} \cdot \left(\omega_{\lambda,s} \boldsymbol{K}_s + \omega_{\lambda,l} \boldsymbol{K}_l + \frac{\omega_{\lambda,l}^2 (1-p)}{1 - \omega_{\lambda,l} p} \boldsymbol{K}_c \right) \tag{5}$$

where $\boldsymbol{K}_s$, $\boldsymbol{K}_l$, $\boldsymbol{K}_c$ are the vectors of expansion coefficients. Comparing Equation 5 with Equation 2 we immediately see that

$$\boldsymbol{K}_\lambda = \omega_{\lambda,s} \boldsymbol{K}_s + \omega_{\lambda,l} \boldsymbol{K}_l + \frac{\omega_{\lambda,l}^2 (1-p)}{1 - \omega_{\lambda,l} p} \boldsymbol{K}_c \tag{6}$$

or explicitly

$$k_\lambda^i = \omega_{\lambda,s} k_s^i + \omega_{\lambda,l} k_l^i + \frac{\omega_{\lambda,l}^2 (1-p)}{1 - \omega_{\lambda,l} p} k_c^i \tag{7}$$

where the superscript "i" represents one of the Lambertian (L), the volumetric (V), and the geometric (G) components.

Equation 6 or Equation 7 is a mathematical expression of our hypothesis stated at the beginning of this section. Note that the k_s^i, k_l^i, k_c^i coefficients and the photon recollision probability p are all spectrally independent. They can be estimated from existing multi-band satellite BRDF products ($\boldsymbol{K}_\lambda^M$) if we have prior knowledge of soil reflectance ($\omega_{\lambda,s}$) and leaf single albedo ($\omega_{\lambda,l}$), which may be obtained from ground-based measurements. To do this we first use the Spectral Response Functions of the satellite instruments to resample the spectra of the soil reflectance ($\omega_{\lambda,s}$) and the leaf single albedo ($\omega_{\lambda,l}$) into the same broad bands as the satellite sensors (Figure 2). We then conduct regression analysis, using the satellite products of $\boldsymbol{K}_\lambda^M$ as the independent variables and the resampled $\omega_{\lambda,s}^M$ and $\omega_{\lambda,l}^M$ as the dependent variables, to estimate the coefficients $\boldsymbol{K}_s$, $\boldsymbol{K}_L$, and $\boldsymbol{K}_c$ following Equation 6. Last, once the coefficients are determined, we use Equation 6 to predict $\boldsymbol{K}_\lambda^H$ at any other (hyperspectral) bands with the optical properties $\omega_{\lambda,s}$ and $\omega_{\lambda,l}$.

In applications when site-specific measurements of $\omega_{\lambda,s}$ and $\omega_{\lambda,l}$ are not available, they must be approximated with spectra libraries or model simulations. The task is to search for a pair of $\omega_{\lambda,s}$ and $\omega_{\lambda,l}$ that minimize the differences between the multi-spectral satellite measurements ($\boldsymbol{K}_\lambda^M$) and the predictions ($\boldsymbol{K}_\lambda^H$) of Equation 6 under certain constraints. To speed up the search, we follow the method described in Q. Yang et al. (2020) and P. Yang et al. (2020) to map a multi-spectral vector $\boldsymbol{K}_{M \times 1}^i$ into the corresponding space of hyperspectral vectors $\boldsymbol{K}_{H \times 1}^i$ with the linear transform

$$\boldsymbol{K}_{H \times 1}^i = \boldsymbol{A}_{H \times M}^i \cdot \boldsymbol{K}_{M \times 1}^i \tag{8}$$

where "i" represents either the Lambertian (L), the volumetric (V), or the geometric (G) components. $\boldsymbol{A}_{H \times M}^i$ is a $H \times M$ matrix estimated from the training data set generated with sample spectra of $\omega_{\lambda,s}$ and $\omega_{\lambda,l}$ from the Aster/ECOSTRESS Library as well as varying combinations of $\boldsymbol{K}_s$, $\boldsymbol{K}_L$, and $\boldsymbol{K}_c$. The details of the estimation of $\boldsymbol{A}_{H \times M}^i$ are described in Q. Yang et al. (2020) and P. Yang et al. (2020). Note that the hyperspectral vector $\boldsymbol{K}_{H \times 1}^i$ estimated
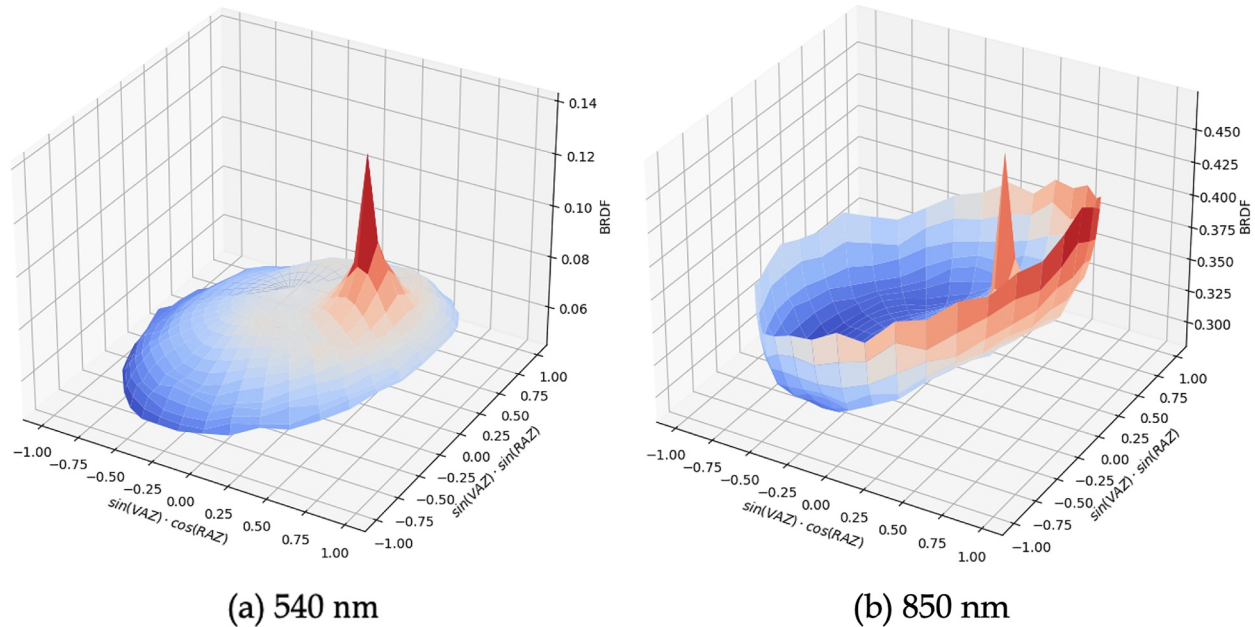
Figure 3. Surface bidirectional reflectance distribution function (BRDF) by simulated with Forest Light Environmental Simulator as a function of view zenith and (relative) azimuth angles. The solar zenith and azimuth angles used in the simulations are 30° and 0°, respectively. Panels (a, b) show the 3D BRDF surface for the Green (540 nm) and the NIR (850 nm) bands. The hot-spot spectra at the 30° VZA is apparent.

with Equation 8 are not necessarily "accurate" because the mapping is not bijective and multiple hyperspectral vectors $\mathbf{K}^i_{H\times 1}$ may correspond to the same multispectral vector $\mathbf{K}^i_{M\times 1}$. The accuracy of the linear transform ($\mathbf{A}^i_{H\times M}$) also depends on the representativeness of the training pairs. These intrinsic uncertainties associated with the estimated hyperspectral BRDF parameters, which cannot be quantified at this stage, emphasize that the synthetic data do not necessarily reflect the reality and are not intended to be used in place of observational datasets.

3. Results

We use FLiES to simulate surface or Top-of-Canopy (TOC) BRDF at 17 view zenith angles (0°–80° at a 5° step) and 19 azimuth angles (0°–180° at a 10° step) at various sun zenith angles (SZA). Figure 3 show an example of the simulated results for SZA = 30° with the forest stand shown in Figure 2. For simplicity, we assume that the soil surface is Lambertian in the simulations. The simulated TOC BRDF is clearly anisotropic. The 3D plots of the BRDF generally show a "bell" shape at absorptive bands (Figure 3a, 540 nm), where the single albedo (ω_s) is relatively low and thus the contribution from multiple scattering is small. In contrast, they show a "bowl" shape at more reflective bands (Figure 3b, 850 nm), where the single albedo is relatively high and the contribution from multiple scattering is more significant. The bi-directional reflectance of the 850 nm band (0.3–0.4 sr^{-1}) is apparently higher than the 540 nm band (0.05–0.1 sr^{-1}), reflecting the difference in their single albedos. Figure 3 also illustrates the "hot-spot" phenomenon, where BRDF sharply rises when the view angle coincides with the illuminate direction.

We can decompose the simulations shown in Figure 3 into a sequence of components by their orders of scattering. Figure 4 shows the results of the first scattering ($k = 1$) and the multiple scattering components ($k \geq 2$). Further decomposition of the multiple scattering component and the evaluation of their convergence can be found in the Appendix A. The first scattering components of both the 540 and the 850 nm bands show very similar "bell" shapes with the signature "hot-spot." The difference in the magnitudes between the two bands largely reflect their different single albedos (Figure 4). In comparison, the multiple scattering components of the two bands show similar "bowl" shapes. Note that the contrast between the magnitude of the 540 nm band (~0.01) and the 850 nm band (~0.2) is much bigger than the differences of their single albedos. The increased magnitude differences between the two bands reflect the effects of multiple scattering in the vegetation canopy: according to Equation 4,
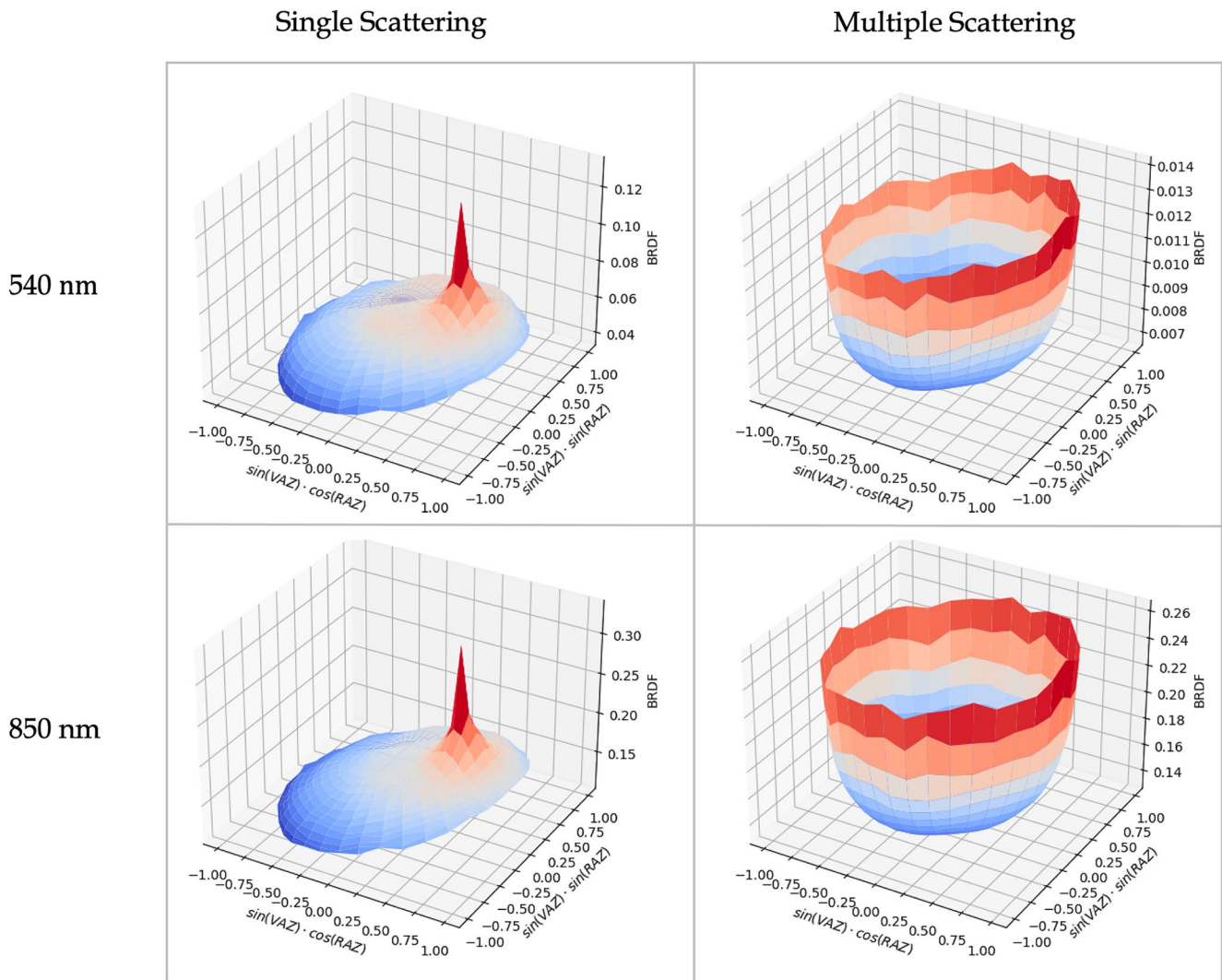
Figure 4. Decomposition of the bidirectional reflectance distribution function of Figure 3 into the first scattering ($k = 1$) and the multiple scattering ($k \geq 2$) components.

the ratio of the multiple scattering reflectance components between two bands are direct proportional to the square of the corresponding ratio between the single albedos. For dense vegetation, where the photon recollision probability (p) approaches 1, the difference is further enlarged by the inverse ratio between the single absorption coefficient ($1 - \omega_{\lambda,l}$; Equation 4).

The results shown in Figure 4 provide visual confirmation to our algorithm derivation outlined in the Methodology section. The nearly identical *shapes* of the first and the multiple scattering BRDF components between the 540 and the 850 nm bands are determined by the corresponding bidirectional scattering factors ($F_l(\Omega_0, \Omega)$ and $F_c(\Omega_0, \Omega)$ in Equation 4), which only depend on the illumination-view geometry. Yet the *magnitudes* of the first and the multiple scattering BRDF components are modulated by the leaf single albedo and the canopy scattering coefficients, which depend on the wavelength of the bands. The agreement between the theoretical derivation (Equation 4) and the MCRT model simulation results (Figure 4) provides strong support to the AGHSD algorithm.

Next, we examine how well the RTLS model can represent the FLiES simulated BRDFs by regressing the latter on the corresponding values (evaluated under the same geometry) of the kernel functions to determine the model parameters K_λ. The results are shown in (Figure 5). In general, the RTLS model well capture ($r^2 > 0.8$) the smooth variations of the original BRDF, as the directional reflectance generated by the RTLS model
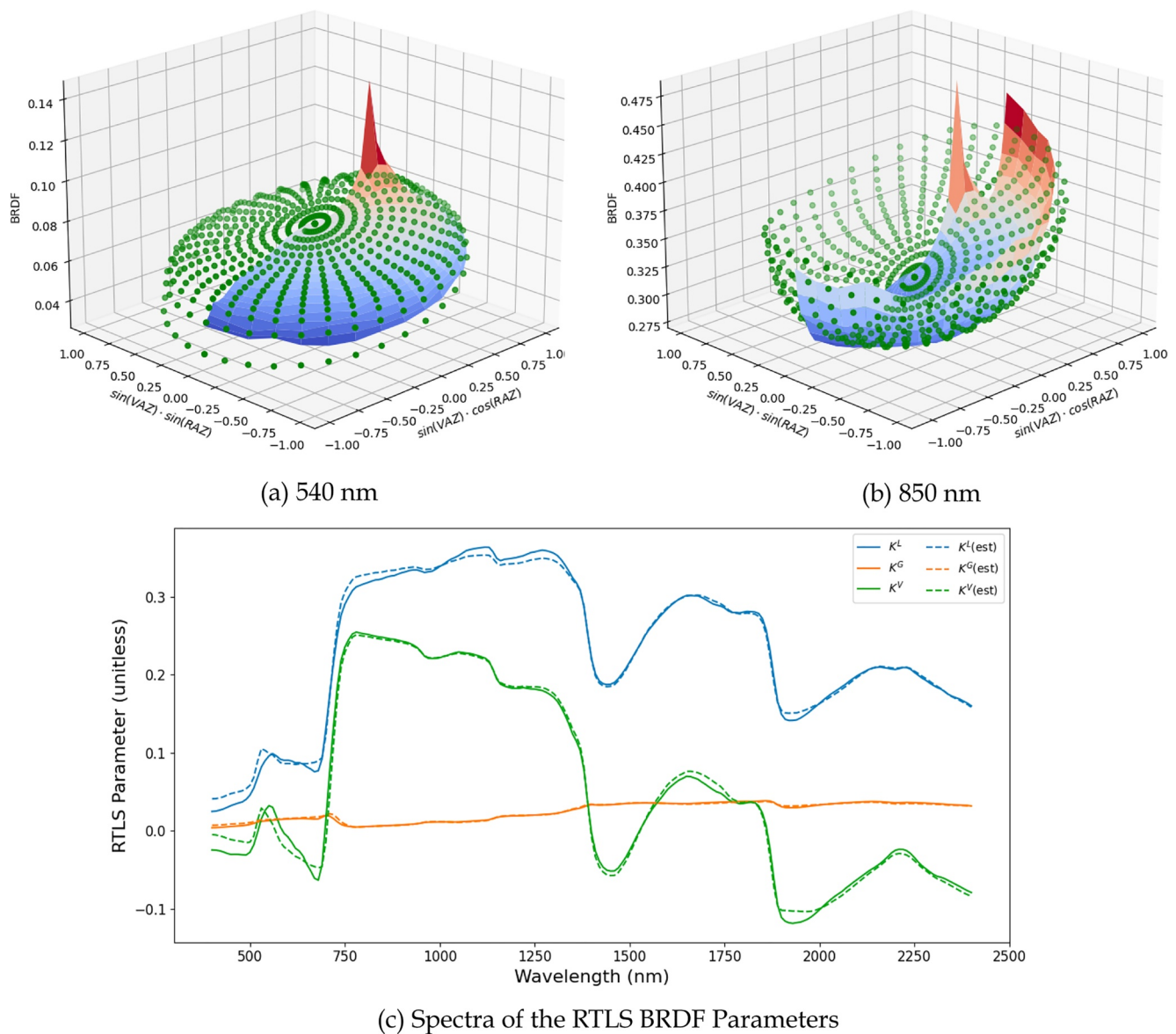
(a) 540 nm

(b) 850 nm

(c) Spectra of the RTLS BRDF Parameters

Figure 5. Ross-thick Li-sparse (RTLS) representation of the simulated surface bidirectional reflectance distribution function (BRDF). The top two panels show comparison of the original BRDF (color surface) with the BRDF generated with the RTLS parameters (green dots) for the (a) 540 and the (b) 850 nm bands. In Panel (c), the solid lines show the RTLS parameters for all the 201 spectral bands ($\boldsymbol{K}_\lambda$), where the dashed lines indicate the estimation of $\boldsymbol{K}_\lambda$ by Equation 6 with the spectra of soil reflectance and leaf single albedo shown in Figure 2. The r^2 statistics between $\boldsymbol{K}_\lambda$ and its estimates are ~0.99 for all three components.

closely resemble those simulated by FLiES (Figure 5). The comparison also indicates that the RTLS model has difficulties in approximating the sharp variations of the BRDFs induced by the hot-spot effects, which is a known challenge in BRDF modeling (Jiao et al., 2016). In this study, we neglect the BRDF discrepancies induced by the hot-spot effects and assume that the representation of surface BRDF by the RTLS model is accurate.

Shown in Figure 5, the spectra of the regressed RTLS parameters ($k_\lambda^L, k_\lambda^V, k_\lambda^G$) reflect features from the optical properties of the forest stand (Figure 2). Both k_λ^L (the Lambertian or isotropic component) and k_λ^V (the volumetric scattering component) show clear signatures of the vegetation single albedo, including the absorption by chlorophyll in the blue and the red bands, the sharp rise in the 700–800 nm range ("red edge"), and the water absorption around the 1,400 and 1,900 nm bands. The spectra of k_λ^G (the geometric scattering component) appears to be flat with little variations. This feature seems reasonable because the geometric scattering component mainly reflects
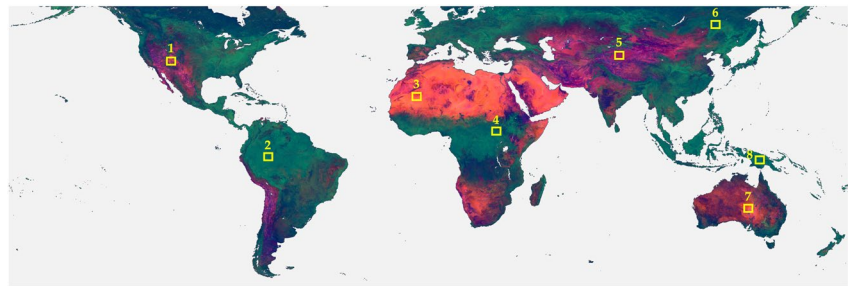
Figure 6. A false-color image of global land surface generated with the synthetic hyperspectral bidirectional reflectance distribution function data using the same illumination-view geometry as in Figure 1. A minimum noise fraction transform is applied to the data and the first three leading principal components are used as the Red, the Green, and the Blue channels, respectively. The yellow boxes indicate the location of the sites examined in Figure 7.

the mutual shadowing effects of the landscape elements (e.g., leaves, stems, etc.), which are supposed to be independent on wavelength.

We should point that the RTLS parameters shown in Figure 5, especially the volumetric and the geometric scattering components (k_λ^V, k_λ^G), are allowed to be negative as far as the overall bidirectional reflectance is positive. If we force the parameters to be nonnegative, the explanation power (by the r^2 statistics) of the simulated BRDF by the RTLS model drops from 80% to 50% at some wavelengths. Allowing the RTLS parameters to be negative is thus a compromise between the higher explanation power and the lower certainty in the physical interpretation of the corresponding BRDF components. A fundamental reason for this compromise is that the RTLS kernel functions are far from a complete basis of the functional space spanned by all possible BRDFs. The kernel functions are not orthogonal, either, which implies multicollinearity between the predicting variables (the RTLS kernel functions) in the regression analysis to fit the predicted variable (the model simulated BRDF). Therefore, the values of individual RTLS parameters (k_λ^V, k_λ^G, in particular) estimated from the regression analysis may be associated with large uncertainties and not suitable for literal physical interpretation. Given both reasons, we tolerate negative values of k_λ^V or k_λ^G in certain ranges. More discussions on the topic of interpreting the physical meaning of the RTLS components can be found in the Appendix A.

We use the spectra of the soil reflectance and the leaf single albedo shown in Figure 2 to approximate the RTLS parameters shown in Figure 5, using Equation 6 as the regression equation. The task can be accomplished with non-linear regression. As shown in Figure 5c, the estimated RTLS parameters (indicated by dash lines) are almost identical to the original ones. The root-mean-square-errors (RMSE) between the true and the estimated RTLS parameters are ~0.008 for the isotropic (k^L) and the volumetric (k^V) components and ~0.001 for the geometric component (k^G). The corresponding r^2 statistics are ~0.99 for all the components. Therefore, if we know the spectra of soil surface reflectance and leaf single albedo of a vegetation stand, we can very accurately estimate the hyperspectral RTLS parameters and therefore the surface BRDF of the landscape. This is exactly what we hypothesized and theoretically proved in developing the AGHSD algorithm. The results in Figure 5 provide the experimental support to our algorithm.

We proceed to apply the AGHSD algorithm to generate the global hyperspectral synthetic BRDF data set from the corresponding MODIS products. For real satellite data, we generally do not know the optical properties of the landscape represented by each of the pixels. Therefore, we use Equation 8 to map the multispectral MODIS BRDF parameters ($\boldsymbol{K}_\lambda^M$) into the spectral library generated in Q. Yang et al. (2020) and P. Yang et al. (2020) to estimate the corresponding hyperspectral BRDF parameters ($\boldsymbol{K}_\lambda^H$). The data set is hosted at our data portal (https://data.nas.nasa.gov/aghsd/index.html) and publicly accessible.

Figure 6 shows a false-color image of global land surface (between 60°N and 60°S) generated with the synthetic hyperspectral BRDF data using the same illumination-view geometry as Figure 1. We applied a Minimum-Noise Fraction (MNF) transformation (Chang & Du, 1999; A. A. Green et al., 1988) to the data and used the first three MNF PCs respectively as the red, green, and blue bands in creating the image. MNF transformation is a special PC Analysis (PCA) technique commonly used in hyperspectral image processing to derive information from the spectral content (Cawse-Nicholson et al., 2021). The leading PCs can be thought as key spectral signatures that are found in the spectra samples in the scene. In our example, at the global scale, the first two MNF components
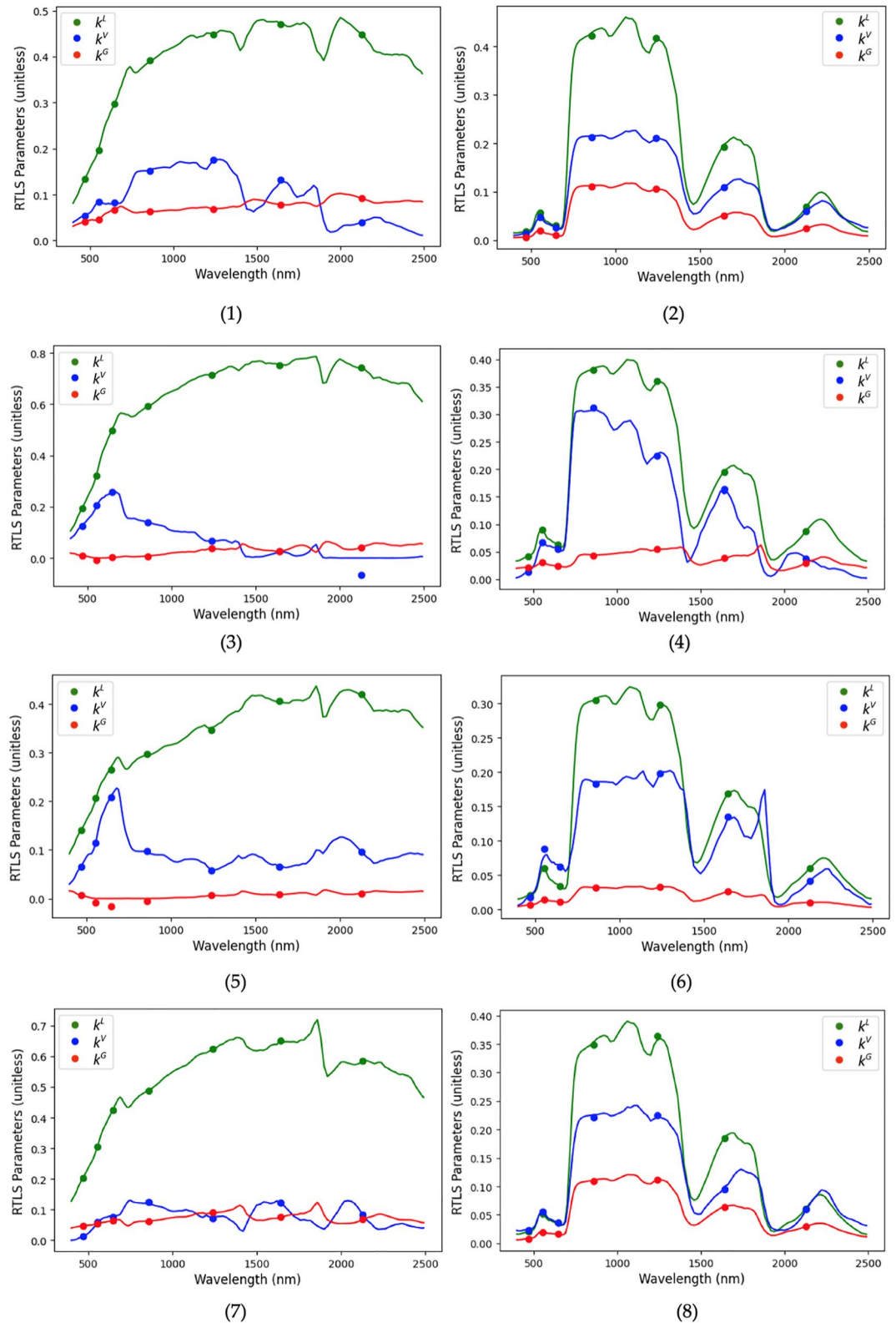
Figure 7. Comparison between the synthetic hyperspectral bidirectional reflectance distribution function Ross-thick Li-sparse parameters (lines) with the corresponding MODIS data (dots) at the eight chosen sites.

(i.e., the red and the green band) clearly represent spectral characteristics of bare soil and vegetation, respectively (Figure 6). Visual examination of the false-color image (Figure 6) indicates that the synthetic data well reveal the same spatial details as the MODIS image (Figure 1), where the landscape variations and landcover changes across the globe are accurately captured. Note that the MNF PCs depend on the spatial domain of the analysis and can be a powerful tool to illustrate subtle variations in spectral features, and thus vegetation species and mineral contents, at regional or local scales (Cawse-Nicholson et al., 2021). This is a distinct advantage of the MNF false-color image over the true-color representation (i.e., using only three spectral bands) of hyperspectral data.

We randomly pick a few sites of different land cover types and over different continent to further examine the performance of the AGHSD algorithm. Figure 7 shows the spectra of the estimated hyperspectral RTLS parameters compared with the corresponding MODIS data over the chosen sites. In general, the spectra of the synthetic RTLS parameters are highly consistent with MODIS data at the selected bands. These spectra tend to closely follow the trajectories of the leaf single albedo ($\omega_{\lambda,l}$) or the soil reflectance ($\omega_{\lambda,s}$). This is particularly true for the vegetated sites (Sites 2, 4, 6, and 8), where the regulation of the leaf optical properties on the RTLS parameters are similar as those identified in the Monte Carlo model simulations (Figure 5). Similar influences of soil surface reflectance are observed on the Lambertian component parameter (k_λ^L) for the non-vegetated sites (Sites 1, 3, 5, and 7).

The spectra of the k_λ^G and the k_λ^V parameters occasionally show signs of overfitting, which are typically induced by the inflation of coefficients and translate into increased variability of the predicted variables. At Site 6, for instance, the curve of the k_λ^V parameter exhibits sharp variations around 1,900 nm, which are most likely artifacts resulted from the regression analysis. Weaker overfitting signs are also recognizable at Site 7 around several wavelengths. We are testing additional regression techniques, including the Tikhonov regularization or ridge regression method (Li et al., 2015) and the bounded-error estimation of parameters (Jaulin & Walter, 1993), to mitigate the problem. Overfitting may also be a reminder that the functional space of the hyperspectral BRDF vectors ($\boldsymbol{K}_\lambda^H$) is under-represented by the spectral library, so that some of the target vectors may not be close to any of the 40,000 example spectra in the library. The only way to address this issue is to populate the spectral library with more diverse spectra examples around the globe. We shall report the progress of these efforts, including the improvement of the AGHSD data, in the future.

4. Discussions

We have emphasized that the synthetic data generated by our algorithm do not necessarily reflect the real-world truth. Yet it is still interesting—and natural—to ask: how realistic are they? Here we discuss this question in a few different contexts.

First, if we ask the question without consideration of the geolocation of the synthetic data, we claim that a well-represented BRDF spectra, such as those of Sites 1 and 2 in Figure 7, can be as realistic as any physics-based models (e.g., FLiES). That is, if there is a forest stand (or bare ground) that has the same (or similar) leaf single albedo ($\omega_{\lambda,l}$) and soil reflectance ($\omega_{\lambda,s}$) as contained in our library, the BRDF of the forest (or the ground) could be accurately represented by our synthetic data. Such a forest stand potentially exists somewhere on Earth.

Second, if we bring in the geolocation context and ask whether a synthetic BRDF spectra realistically represents the "ground truth" at the specific pixel, the answer becomes less certain: it may or may not. The uncertainty is ultimately induced by the fact that the mapping from the multispectral data (MODIS) to the hyperspectral space is not fully constrained. The same multispectral BRDF vector could correspond to several candidate hyperspectral solutions, from which we must pick one. Without further information, our algorithm simply does not know which solution is more realistic than the others.

Uncertainties also exist in the multispectral BRDF products. If such uncertainties are totally random, they should have little impact on the regression results with Equation 6. However, if they are systematic biases and, in particular, if such biases are correlated with surface optical properties ($\omega_{\lambda,s}$ and $\omega_{\lambda,l}$), the retrieved parameters from the regression analysis will be biased as well. Such uncertainties can propagate to the estimated RTLS coefficients for the other wavelengths with scaling factors determined by the spectra of surface optical properties. If we further consider uncertainties in the prior knowledge of surface optical properties, the propagation of errors will be more complicated.

Hyperspectral observations can be brought in to help us determine the optimal solution of surface BRDF and to detect and suppress uncertainties in the multispectral data and the prior surface optical properties. Existing hyperspectral observations include those from AVIRIS, DESIS, PRISMA, EMIT, as well as the EO-1/Hyperion. Strictly speaking, using satellite or airborne observations to retrieve or improve the prior knowledge of surface/atmosphere variables is the task of inversion modeling. For the development of AGHSD, we are interested in identifying the sources of uncertainties in light transport from the surface to the top of the atmosphere so that we can forwardly simulate their impacts on the at-sensor signals. We are developing a forward modeling algorithm based on the well-known concept of Green's function (Lyapustin & Knyazikhin, 2001) to accurately and efficiently model TOA radiance from surface BRDF. By adjusting the level of uncertainties in the surface BRDF parameters, we expect to derive a set of more realistic "model truths" to test the inverse scientific algorithms. We plan to report the development of the synthetic TOA hyperspectral data in a separate paper.

Finally, suppose the hyperspectral remote sensing observations allow us to choose the realistic BRDF spectra ($\boldsymbol{K}_\lambda^H$) for a given pixel, are we able to separate the optical properties ($\omega_{\lambda,l}$ and $\omega_{\lambda,s}$) from the influence of the landscape structure characteristics ($\boldsymbol{K}_s$, $\boldsymbol{K}_l$, and $\boldsymbol{K}_c$—see Equation 6 for details)? This is another important yet challenging problem that requires intensive research (Knyazikhin et al., 2013; Z. Wang et al., 2017). Theoretically speaking, an inverse problem of this kind can only be solved with additional information (Baret & Buis, 2008; Combal et al., 2002; Rodgers, 2000), which may be derived from hyperspectral data or other type of observations (e.g., LIDAR measurements of vegetation structure). Further discussion on this topic is out of the scope of this paper and we refer the interested audience to recent papers on surface imaging algorithms with hyperspectral remote sensing data (Cawse-Nicholson et al., 2021; Raiho et al., 2023). Nevertheless, we hope the development of AGHSD and the associated forward modeling framework provide an experimental environment that may help test scientific algorithms with this regards.

5. Conclusions

This paper introduces the AGHSD, in particular the surface BRDF product. We derive the synthetic surface BRDF, represented by the parameters of the RTLS model, from the corresponding MODIS products. Based on the theories of radiative transfer in vegetation canopies, we derive a simple but robust relationship that indicates the RTLS parameters of hyperspectral surface BRDF can be accurately expressed as a weighted sum of the soil surface reflectance, the leaf single albedo, and the canopy scattering coefficient. The weighting coefficients of these components are determined by the structural characteristics of the landscape and thus do not depend on the wavelength of the spectra. Therefore, they can be estimated from the multispectral MODIS data.

We validate our algorithm with the Monte Carlo Ray Tracing radiative transfer model, FLiES. The simulation results are highly consistent with the theoretic derivations. The simulated BRDF resulted from the first and the multiple scattering components show distinct spatial distribution signatures associated with the geometric-optical and the volumetric-scattering kernels, which are independent on wavelength, of the RTLS model. In comparison, their magnitudes are regulated by the leaf single albedo and the canopy scattering coefficient, which are dependent on wavelengths. With prior knowledge of these optical properties, our algorithm can very accurately estimate the hyperspectral RTLS parameters for the simulated surface BRDF.

We apply the algorithm to generate a global hyperspectral BRDF data at 1 km and 8-day resolutions for 2019 based on the MODIS (MCD19A3) products. The spectra of the synthetic RTLS parameters are highly accurate as the MODIS data at the selected spectral bands. The spectra are also consistent and coherent at other wavelengths in general. These results indicate that, in most cases, we can use weighted spectra of surface soil and vegetation optical properties to satisfactorily interpolate/extrapolate the RTLS parameters from the few MODIS bands to any hyperspectral SBG bands.

We also detect weakness of the current algorithm. It is found that at unvegetated pixels, the algorithm may overfit the model to the multispectral data and introduce spurious variations in the spectra of the RTLS parameters. This problem was essentially caused by the under-representativeness of the spectral library used currently by the algorithm. Future work thus is required to populate the spectral library with more diverse spectra samples in order to mitigate the problem. In addition, we are testing alternative regression techniques, including Tikhonov regularization and Bounded-Error Estimation, to detect and address the occurrence of model overfitting in our

algorithm. We also plan to compare the AGHSD data with available hyperspectral observations from satellite and airborne sensors (e.g., AVIRIS-NG and DESIS) to further examine the quality of the data set.

Finally, we emphasize that AGHSD is specifically generated to support the development and testing of SBG community science algorithms. The data set is publicly accessible from our data portal (https://data.nas.nasa.gov/aghsd/index.html). All SBG communities, including terrestrial ecology, aquatic ecosystems, surface minerology, and the cryosphere community, can use AGHSD to test if their algorithms or processing pipelines can correctly handle the format and the volume of weekly global hyperspectral data. On the other hand, we emphasize that the community should not expect to discover new scientific results from the synthetic data set. As described in this paper, no hyperspectral "ground truth" is used in generating AGHSD. Indeed, such a wall-to-wall global hyperspectral data set still awaits to be derived from future hyperspectral satellite sensors, which highlights the importance and excitement of the SBG mission.

Appendix A: Additional Results of Monte Carlo Ray Tracing Model Experiments

A1. Convergence of Multiple Scattering

Figure A1 shows the decomposition of the multiple scattering ($k \geq 2$) bidirectional reflectance distribution function (BRDF) component of Figure 4 in the main text into the second ($k = 2$), the third ($k = 3$), and the fourth ($k = 4$) scattering components. For both the 540 nm and the 850 bands, all the components have similar "bowl-shaped" distributions. We calculate the spatial correlation coefficients between the BRDF components as a means to evaluate the convergence of the shape of the angular distributions. For both bands, the correlation coefficients (r) between the second and the third scattering components are ∼0.985, which increase to ∼0.996 between the third and the fourth scattering components. Therefore, these results support the argument that the spatial/angular distribution of the radiation fields resulted from multiple ($k \geq 2$) scatterings converge rapidly. In comparison, the correlation coefficients between the first and the second scattering components are negative (−0.6 or −0.7) for both bands (Figure 4 in the main text). The contrast between the first and the multiple scattering BRDF components also support our approach to treat them separately, as expressed in Equations 4–7.
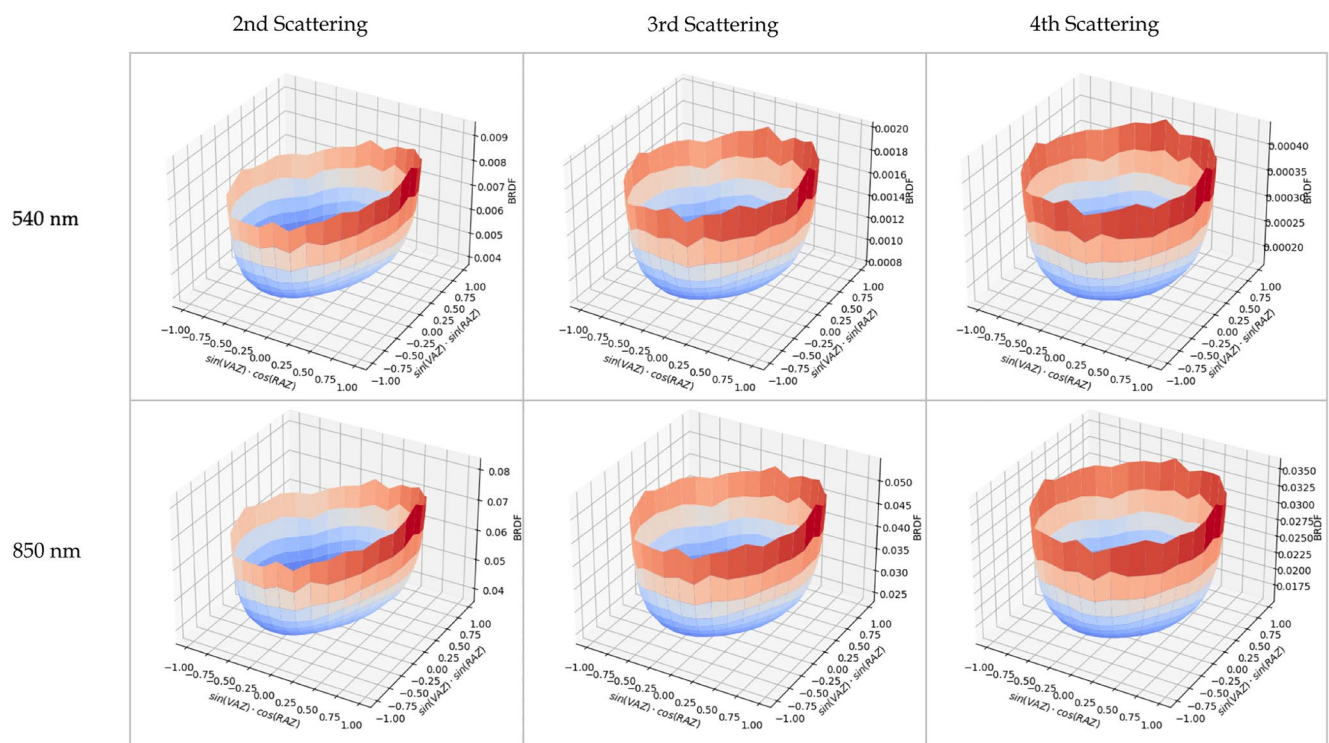

Figure A1. Decomposition of the multiple scattering ($k \geq 2$) component of the bidirectional reflectance distribution function shown in Figure 4 into the second ($k = 2$), the third ($k = 3$), and the fourth ($k = 4$) scattering components, respectively. For both spectral bands, the correlation coefficients (r) between the second and the third scattering components are ∼0.985, which increase to ∼0.996 between the third and the fourth scattering components.

Table A1
Values of the K_s, K_l, K_c Vectors Corresponding to the Ross-Thick Li-Sparse Coefficients Shown in Figure 5

	Isotropic	Volumetric	Geometric
K_s	0.387	0.084	−0.300
K_l	0.312	0.024	0.379
K_c	−0.067	−0.046	−0.027

A2. BRDF Components From Soil Surface, Vegetation Boundary, and Vegetation Interior

The key idea of the Ames Global Hyperspectral Synthetic Data set algorithm is that the surface BRDF can be largely decomposed to contributions from the soil surface, the first scattering component of vegetation canopies (vegetation boundary), and the multiple scattering component of vegetation canopies (vegetation interiors), and each of these components can be represented as the product of the spectrally dependent optical properties (soil reflectance $\omega_{\lambda,s}$, leaf single albedo $\omega_{\lambda,l}$, and canopy scattering coefficient $\omega_{\lambda,c}$) and a spectrally independent weight function that depends on the illumination-view geometries (Equations 4–6 in the main text). Each of the weight functions can be looked at as a normalized BRDF and thus represented by its own Ross-Thick-Li-Sparse (RTLS) parameters (K_s, K_l, K_c). These parameters can be determined by regression analysis with Equation 6. Table A1 shows the values of K_s, K_l, K_c corresponding to the BRDF RTLS parameters in Figure 5 in the main text. Examples of 3D plots of these functions are shown in Figure A2. As shown, the BRDF of the soil surface has a "dome" shape and the one representing the multiple scattering component of vegetation canopies ("vegetation interior") has a "bowl" shape. The BRDF representing the first scattering component of vegetation canopies ("vegetation boundary") falls somewhere in between (Figure A2).

However, we should avoid strict interpretations of the physical meaning of the K_s, K_l, K_c parameters. Although the RTLS kernels are semi-physical, they are neither a complete nor an orthogonal) basis for the functional space of BRDF. As discussed in the main text, although negative component of BRDF appears non-physical, we have to allow some of the parameters to be negative to increase the RTLS models explanation power of the simulated BRDF. The surface optical properties ($\omega_{\lambda,s}$, $\omega_{\lambda,l}$, and $\omega_{\lambda,c}$) are not totally mutually independent either. The multicollinearity among these variables also can induce ambiguity in the retrieved k_λ^i parameters (where the superscript "i" represents one of the Lambertian (L), the volumetric (V), and the geometric (G) components). Based on these considerations, we didn't discuss much about the values of the K_s, K_l, K_c parameters in the main text.

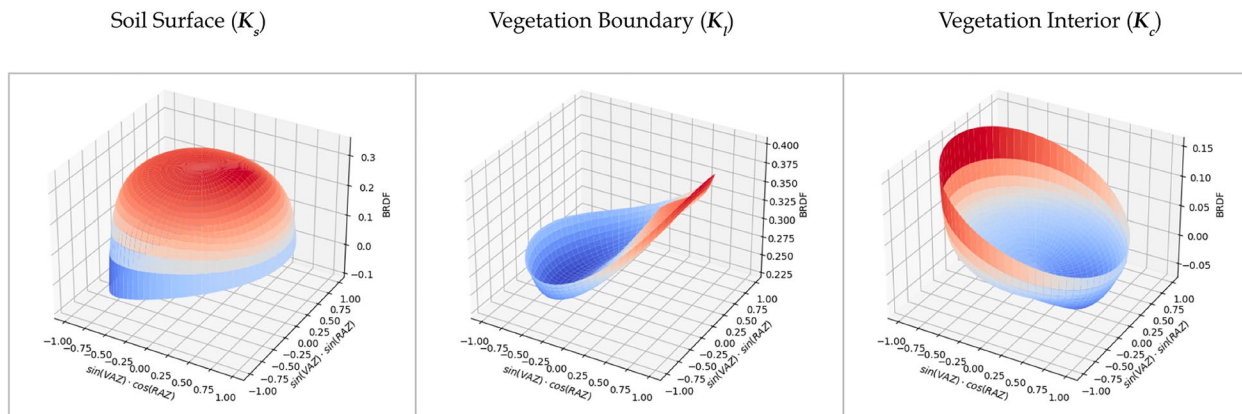**Figure A2.** Surface bidirectional reflectance distribution function corresponding to the Ross-thick Li-sparse vectors K_s, K_l, K_c shown in Table A1. The solar zenith angle is 30° in these plots.

A3. Model Sensitivity Experiments

Although the Forest Light Environmental Simulator (FLiES) experiment results reported in the main text are based on one model configuration, we have conducted model sensitivity experiments with various combinations of spectral optical properties of leaves, barks, and soil surfaces (Figure A3), landscape/forest structures (Figure A4), as well as model parameters (solar illumination geometry, leaf surface angle distribution function, etc.). We have repeated the model experiments and the analysis for over 100 model configurations, the results are all very similar to those presented in the paper. They all support our argument that the hyperspectral surface BRDF can be accurately approximated as a weighted sum of the soil surface reflectance, the leaf single albedo, and the canopy scattering coefficient.

A particular model sensitivity experiment we conducted is regarding the differences between the radiation fields in broadleaf and needleleaf forests. Because FLiES does not explicitly distinguish between broadleaf and needleleaf forests, we simulated the contrasts between the two types of forests by varying the shapes of tree crowns (e.g., cones vs. ellipsoids) and the ratio of the spherical average shoot silhouette area to the total needle area (referred to as "*sbar*" in the FLiES document). According to the document and the literature, *sbar* is fixed to be 0.25 for broadleaf forests and is generally lower (e.g., 0.15) for needleleaf forests (e.g., Oker-Blom & Smolander, 1988), which also suggests stronger clumping effects. We specify the crown shapes in our study scene to be cones and set the *sbar* parameters to be 0.15 and 0.1 in the model simulations. The results are shown in Figure A5 below. Under the otherwise the same model configurations, changing the crown shapes from ellipsoids to cones and changing the *sbar* parameter from 0.25 to 0.15 (or 0.10) tend to reduce the magnitude of the surface BRDF, especially the multiple scattering components. This is likely due to the increased clumping effects (Kobayashi et al., 2010). The spatial/angular distributions of the BRDFs slightly change with the changes in crown shapes and the values of *sbar*. But we can always accurately retrieve the RTLS parameters from the simulations with Equation 6. The rates of convergence by the multiple-scattering components of the simulated BRDF are also similar to those described above. All these results, though not necessarily exhaustive, suggest that our modeling approach (Equations 4–7) is robust to the extent as allowed by FLiES.
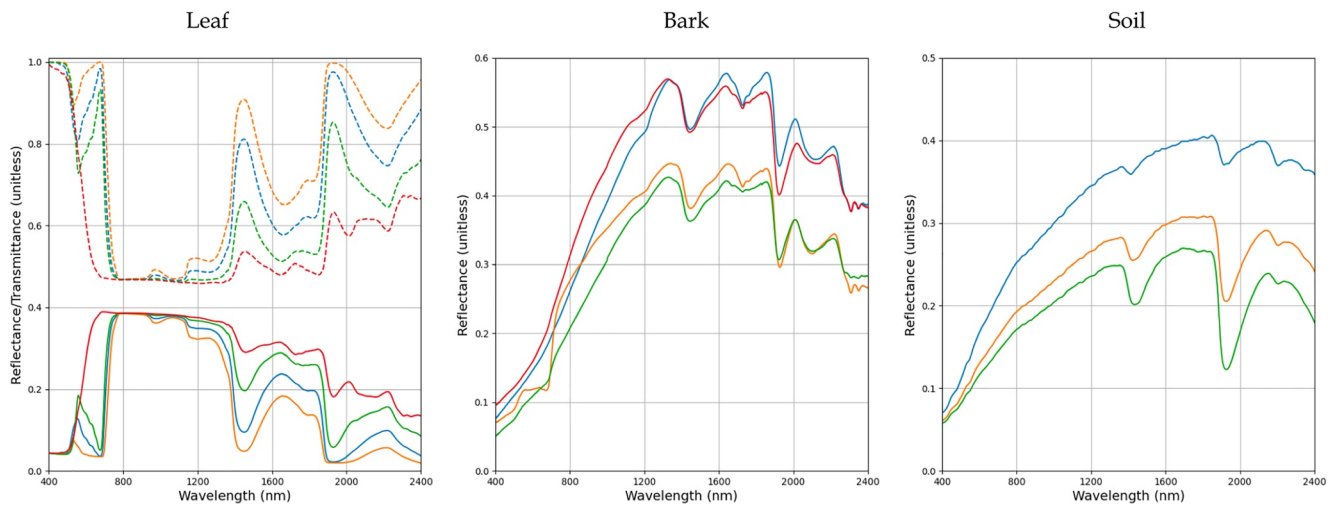

Figure A3. Examples of forest spectral optical properties, including reflectance/transmittance of leaves, barks, and soil surfaces used in the model sensitivity experiments of Forest Light Environmental Simulator.
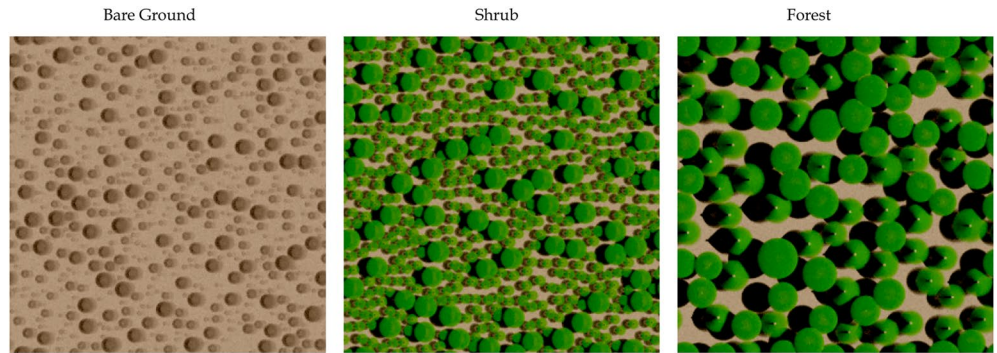
Figure A4. Examples of landscape/vegetation canopies, including bare ground, shrubs, and forests, used in the model sensitivity experiments of Forest Light Environmental Simulator.
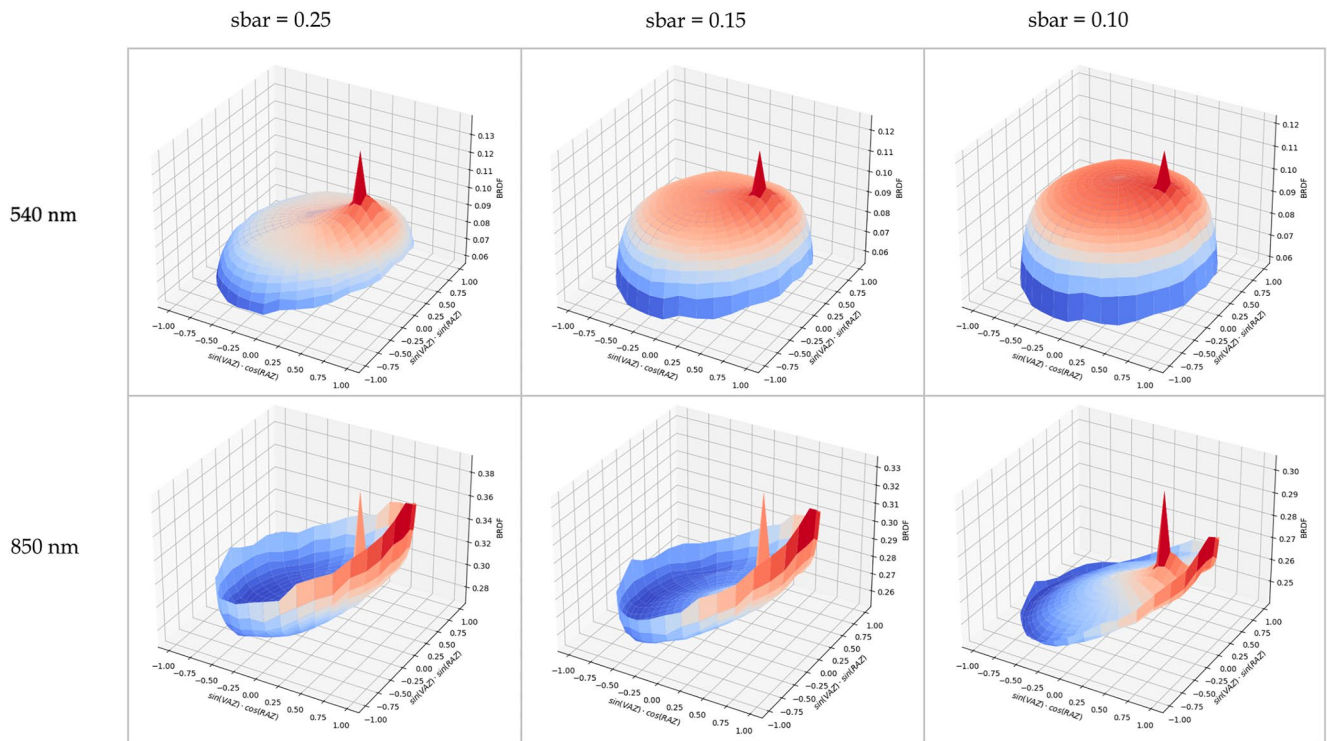

Figure A5. Surface bidirectional reflectance distribution function simulated by Forest Light Environmental Simulator similar to Figure 3 but with cone-shaped tree crowns and the ratios of the spherical average shoot silhouette area to the total needle area ("sbar") decreasing from 0.25 to 0.10.

Acronyms

AGHSD	NASA Ames Global Hyperspectral Synthetic Data set
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
BRDF	Bidirectional Reflectance Distribution Function
CHIME	Copernicus Hyperspectral Imaging Mission for the Environment
DESIS	DLR Earth Sensing Imaging Spectrometer
ECOSTRESS	Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station
EMIT	Earth Surface Mineral Dust Source Investigation
EnMAP	Environmental Mapping and Analysis Program
FLiES	Forest Light Environment Simulator

HISUI	Hyperspectral Imager Suite
HyspIRI	Hyperspectral Infrared Imager
MAIAC	Multi-Angle Implementation Atmospheric Correction
MNF	Minimum Noise Fraction
MODIS	Moderate Resolution Imaging Spectroradiometer
PCA	Principal Component Analysis
PRISMA	Hyperspectral Precursor of the Application Mission
RTLS	Ross-Thick-Li-Sparse
SBG	Surface Biology and Geology
TIR	Thermal Infrared
TOA	Top-of-Atmosphere
TOC	Top-of-Canopy
VSWIR	Visible to Short-Wave Infrared

Data Availability Statement

The AGHSD surface BRDF data set, with a total volume of ∼48 TB, is publicly available for download through NASA Advanced Supercomputing (NAS) Data Portal (https://data.nas.nasa.gov/aghsd/aghsddata/version2.0/). The code example to illustrate the algorithm is available via Zenodo at https://zenodo.org/record/7877752#.ZExFg-zMKWA.

References

Baldridge, A. M., Hook, S. J., Grove, C. I., & Rivera, G. (2009). The ASTER spectral library version 2.0. *Remote Sensing of Environment*, *113*(4), 711–715. https://doi.org/10.1016/j.rse.2008.11.007

Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. In S. Liang (Ed.), *Advances in land remote sensing*. Springer.

Cavender-Bares, J., Gamon, J. A., & Townsend, P. A. (2020). *Remote sensing of plant biodiversity*. Springer Nature.

Cawse-Nicholson, K., Townsend, P. A., Schimel, D., Assiri, A. M., Blake, P. L., Buongiorno, M. F., et al. (2021). NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. *Remote Sensing of Environment*, *257*, 112349. https://doi.org/10.1016/j.rse.2021.112349

Chang, C.-I., & Du, Q. (1999). Interference and noise-adjusted principal components analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(5), 2387–2396. https://doi.org/10.1109/36.789637

Chernetskiy, M., Gobron, N., Gómez-Dans, J., Morgan, O., Disney, M., Lewis, P., & Schmullius, C. (2018). Simulating arbitrary hyperspectral bandsets from multispectral observations via a generic Earth Observation-Land Data Assimilation System (EO-LDAS). *Advances in Space Research*, *62*(7), 1654–1674. https://doi.org/10.1016/j.asr.2018.07.015

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., & Pragnére, A. (2002). Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, *84*, 1–15. https://doi.org/10.1016/S0034-4257(02)00035-4

Folkman, M. A., Pearlman, J., Liao, L. B., & Jarecke, P. J. (2001). EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration. *Hyperspectral Remote Sensing of the Land and Atmosphere*, *4151*, 40–51. https://doi.org/10.1117/12.417022

Gao, B. C., Heidebrecht, K. B., & Goetz, A. F. (1993). Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of Environment*, *44*(2–3), 165–178. https://doi.org/10.1016/0034-4257(93)90014-o

Green, A. A., Berman, M., Paul, S., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, *26*(1), 65–74. https://doi.org/10.1109/36.3001

Green, R. O., Mahowald, N., Ung, C., & Thompson, D. R. (2020). Lori Bator, Matthew Bennet, Michael Bernas et al. The Earth surface mineral dust source investigation: An Earth science imaging spectroscopy mission. In *2020 IEEE aerospace conference* (pp. 1–15). IEEE.

Guanter, L., Kaufmann, H., Karl, S., Foerster, S., Rogass, C., Chabrillat, S., et al. (2015). The EnMAP spaceborne imaging spectroscopy mission for Earth observation. *Remote Sensing*, *7*, 8830–8857. https://doi.org/10.3390/rs70708830

Guanter, L., Segl, K., & Kaufmann, H. (2009). Simulation of optical remote-sensing scenes with application to the EnMAP hyperspectral mission. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2340–2351. https://doi.org/10.1109/tgrs.2008.2011616

Hamlin, L., Green, R. O., Mouroulis, P., Eastwood, M., Wilson, D., Dudik, M., & Paine, C. (2011). Imaging spectrometer science measurements for terrestrial ecology: AVIRIS and new developments. In *2011 Aerospace conference* (pp. 1–7). IEEE.

Huang, D., Knyazikhin, Y., Dickinson, R. E., Rautiainen, M., Stenberg, P., Disney, M., et al. (2007). Canopy spectral invariants for remote sensing and model applications. *Remote Sensing of Environment*, *106*(1), 106–122. https://doi.org/10.1016/j.rse.2006.08.001

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, *34*(2), 75–91. https://doi.org/10.1016/0034-4257(90)90100-z

Jaulin, L., & Walter, E. (1993). Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, *29*(4), 1053–1064. https://doi.org/10.1016/0005-1098(93)90106-4

Jiao, Z., Schaaf, C. B., Dong, Y., Román, M., Hill, M. J., Chen, J. M., et al. (2016). A method for improving hotspot directional signatures in BRDF models used for MODIS. *Remote Sensing of Environment*, *186*, 135–151. https://doi.org/10.1016/j.rse.2016.08.007

Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., et al. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences*, *110*(3), E185–E192. https://doi.org/10.1073/pnas.1210196109

Kobayashi, H., Delbart, N., Suzuki, R., & Kushida, K. (2010). A satellite-based method for monitoring seasonality in the overstory leaf area index of Siberian larch forest. *Journal of Geophysical Research*, *115*(G1), G01002. https://doi.org/10.1029/2009jg000939

Kobayashi, H., & Iwabuchi, H. (2008). A coupled 1-D atmosphere and 3-D canopy radiative transfer model for canopy reflectance, light environment, and photosynthesis simulation in a heterogeneous landscape. *Remote Sensing of Environment*, *112*(1), 173–185. https://doi.org/10.1016/j.rse.2007.04.010

Kokaly, R. F., Clark, R. N., Swayze, G. A., Livo, K. E., Hoefen, T. M., Pearson, N. C., et al. (2017). *USGS spectral library version 7 data: US geological survey data release*. United States Geological Survey (USGS).

Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., et al. (2019). The instrument design of the DLR Earth sensing imaging spectrometer (DESIS). *Sensors*, *19*(7), 1622. https://doi.org/10.3390/s19071622

Kunstler, G., Falster, D., Coomes, D. A., Hui, F., Kooyman, R. M., Laughlin, D. C., et al. (2016). Lourens Poorter et al. "Plant functional traits have globally consistent effects on competition". *Nature*, *529*(7585), 204–207. https://doi.org/10.1038/nature16476

Lee, C. M., Morgan, L., Cable, S. J., Hook, R. O., Green, S. L., Ustin, D. J. M., & Middleton, E. M. (2015). An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment*, *167*, 6–19. https://doi.org/10.1016/j.rse.2015.06.012

Lewis, P., & Disney, M. (2007). Spectral invariants and scattering across multiple scales from within-leaf to canopy. *Remote Sensing of Environment*, *109*(2), 196–206. https://doi.org/10.1016/j.rse.2006.12.015

Li, W., Du, Q., & Xiong, M. (2015). Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification. In *IEEE geoscience and remote sensing letters* (Vol. 12, pp. 48–52).

Liu, B., Zhang, L., Zhang, X., Zhang, B., & Tong, Q. (2009). Simulation of EO-1 hyperion data from ALI multispectral data based on the spectral reconstruction approach. *Sensors*, *9*(4), 3090–3108. https://doi.org/10.3390/s90403090

Loizzo, R., Guarini, R., Longo, F., Scopa, T., Formaro, R., Facchinetti, C., & Varacalli, G. (2018). PRISMA: The Italian hyperspectral mission. In *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium* (pp. 175–178). IEEE.

Lucht, W., Schaaf, C. B., & Strahler, A. H. (2000). Algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 977–998. https://doi.org/10.1109/36.841980

Lyapustin, A., & Knyazikhin, Y. (2001). Green's function method for the radiative transfer problem. I. Homogeneous non-Lambertian surface. *Applied Optics*, *40*(21), 3495–3501. https://doi.org/10.1364/ao.40.003495

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, *11*(10), 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Martonchik, J. V., Bruegge, C. J., & Strahler, A. H. (2000). A review of reflectance nomenclature used in remote sensing. *Remote Sensing Reviews*, *19*(1–4), 9–20. https://doi.org/10.1080/02757250009532407

Matsunaga, T., Iwasaki, A., Tsuchida, S., Iwao, K., Tanii, J., Kashimura, O., et al. (2017). Current status of hyperspectral imager suite (HISUI) onboard International Space Station (ISS). In *2017 IEEE international geoscience and remote sensing symposium (IGARSS)* (pp. 443–446). IEEE.

Meerdink, S. K., Hook, S. J., Roberts, D. A., & Abbott, E. A. (2019). The ECOSTRESS spectral library version 1.0. *Remote Sensing of Environment*, *230*, 111196. https://doi.org/10.1016/j.rse.2019.05.015

National Academies of Sciences, Engineering, Aand Medicine. (2018). *Thriving on our changing planet: A decadal strategy for Earth observation from space*. The National Academies Press.

Oker-Blom, P., & Smolander, H. (1988). The ratio of shoot silhouette area to total needle area in Scots pine. *Forest Science*, *34*(4), 894–906.

Poulter, B., Currey, B., Calle, L., Brookshire, J., Campbell, P., Adam, C., et al. (2023). Simulating global dynamic surface reflectances for imaging spectroscopy spaceborne missions-LPJ-PROSAIL. *Journal of Geophysical Research: Biogeosciences*, *128*(1), e2022JG006935. https://doi.org/10.1029/2022jg006935

Raiho, A., Cawse-Nicholson, K., Adam, C., Dozier, J., Gierach, M. M., Miner, K., et al. (2023). Exploring mission design for imaging spectroscopy retrievals for land and aquatic ecosystems. *Journal of Geophysical Research: Biogeosciences*, *128*(4), e2022JG006833. https://doi.org/10.1029/2022jg006833

Rast, M., Jens Nieke, J., Adams, C. I., & Ferran, G. (2021). Copernicus hyperspectral imaging mission for the environment (Chime). In *2021 IEEE international geoscience and remote sensing symposium IGARSS* (pp. 108–111). IEEE.

Rivera, J. P., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., & Camps-Valls, G. (2015). An emulator toolbox to approximate radiative transfer models with statistical learning. *Remote Sensing*, *7*, 9347–9370. https://doi.org/10.3390/rs70709347

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and practice* (p. 238). World Scientific Publishing Co. Pte. Ltd.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, *83*(1–2), 135–148. https://doi.org/10.1016/s0034-4257(02)00091-3

Schimel, D. S., & Poulter, B. (2022). The Earth in living color-NASA's surface biology and geology designated observable. In *2022 IEEE aerospace conference (AERO)* (pp. 1–6). IEEE.

Serbin, S. P., & Townsend, P. A. (2020). Scaling functional traits from leaves to canopies. In J. Cavender-Bares, J. A. Gamon, & P. A. Townsend (Eds.), *Remote sensing of plant biodiversity* (pp. 43–82). Springer International Publishing.

Smolander, S., & Stenberg, P. (2005). Simple parameterizations of the radiation budget of uniform broadleaved and coniferous canopies. *Remote Sensing of Environment*, *94*(3), 355–363. https://doi.org/10.1016/j.rse.2004.10.010

Sun, X., Zhang, L., Yang, H., Wu, T., Cen, Y., & Guo, Y. (2014). Enhancement of spectral resolution for remotely sensed multispectral image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(5), 2198–2211. https://doi.org/10.1109/jstars.2014.2356512

Thompson, D. R., Brodrick, P. G., Cawse-Nicholson, K., Dana Chadwick, K., Green, R. O., Poulter, B., et al. (2021). Spectral fidelity of Earth's terrestrial and aquatic ecosystems. *Journal of Geophysical Research: Biogeosciences*, *126*(8), e2021JG006273. https://doi.org/10.1029/2021jg006273

Thompson, D. R., Natraj, V., Green, R. O., Helmlinger, M. C., Gao, B. C., & Eastwood, M. L. (2018). Optimal estimation for imaging spectrometer atmospheric correction. *Remote Sensing of Environment*, *216*, 355–373. https://doi.org/10.1016/j.rse.2018.07.003

Verrelst, J., Rivera Caicedo, J. P., Vicent, J., Morcillo Pallarés, P., & Moreno, J. (2019). Approximating empirical surface reflectance data through emulation: Opportunities for synthetic scene generation. *Remote Sensing*, *11*(2), 157. https://doi.org/10.3390/rs11020157

Wang, W., Nemani, R., Hashimoto, H., Ganguly, S., Huang, D., Knyazikhin, Y., et al. (2018). An interplay between photons, canopy structure, and recollision probability: A review of the spectral invariants theory of 3D canopy radiative transfer processes. *Remote Sensing*, *10*(11), 1805. https://doi.org/10.3390/rs10111805

Wang, Z., Skidmore, A. K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., et al. (2017). Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International Journal of Applied Earth Observation and Geoinformation*, *54*, 84–94. https://doi.org/10.1016/j.jag.2016.09.008

Wanner, W., Strahler, A., Hu, B., Lewis, P., Muller, J. P., Li, X., et al. (1997). Global retrieval of bidirectional reflectance and albedo over land from EOS MODIS and MISR data: Theory and algorithm. *Journal of Geophysical Research*, *102*(D14), 17143–17161. https://doi.org/10.1029/96jd03295

Widlowski, J.-L., Mio, C., Disney, M., Adams, J., Andredakis, I., Atzberger, C., et al. (2015). The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing. *Remote Sensing of Environment*, *169*, 418–437. https://doi.org/10.1016/j.rse.2015.08.016

Yang, P., van der Tol, C., Yin, T., & Verhoef, W. (2020). The SPART model: A soil-plant-atmosphere radiative transfer model for satellite measurements in the solar spectrum. *Remote Sensing of Environment*, *247*, 111870. https://doi.org/10.1016/j.rse.2020.111870

Yang, Q., Liu, X., & Wu, W. (2020). A hyperspectral bidirectional reflectance model for land surface. *Sensors*, *20*(16), 4456. https://doi.org/10.3390/s20164456