Remote Sensing of Environment 315 (2024) 114467

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Illumination correction for close-range hyperspectral images using spectral invariants and random forest regression

Olli Ihalainen [a,*], Theresa Sandmann [b,c], Uwe Rascher [c], Matti Mõttus [a]

[a] VTT Technical Research Centre of Finland Ltd, Otaniemi, Finland
[b] University of Bonn, Bonn, Germany
[c] Forschungszentrum Jülich GmbH, Jülich, Germany

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Close-range
Hyperspectral
Imaging spectroscopy
Radiative transfer
p-theory
Spectral invariants
Monte Carlo ray tracing
Random forest
Inversion

ABSTRACT

Identifying materials and retrieving their properties from spectral imagery is based on their spectral reflectance calculated from the ratio of reflected radiance to the incident irradiance. However, obtaining the true reflectances of materials within a vegetation canopy is challenging given the varying illumination conditions across the canopy – i.e., the irradiance incident on a surface inside the canopy – caused by its complex 3D structure. Instead, in remote sensing, reflectances are calculated from the ratio of the spectral radiance measured by the sensor to the top-of-canopy (TOC) spectral irradiance, resulting in apparent reflectances that can significantly differ from the true reflectance spectra. To address this issue, we present a physically based illumination correction method for retrieving the true reflectances from close-range hyperspectral TOC reflectance images. The method uses five spectral invariant parameters to predict the illumination conditions from TOC reflectance and compute the corrected spectrum using a physically based model. For computational efficiency, the spectrally invariant parameters were retrieved using random forest regression trained with Monte Carlo ray tracing simulations. The method was tested on close-range imaging spectroscopy data from dense and sparse vegetation canopies for which reference in situ spectral measurements were available. This work is a step toward resolving the 3D radiation regime in vegetation canopies from TOC hyperspectral imagery. The retrieved spectral invariants provide a physical connection to the structure of the observed vegetation canopy. The true spectra of artificial and natural materials in a vegetation canopy, determined under various illumination conditions, allow their more robust (bio)chemical characterization, opening new applications in vegetation monitoring and material detection, and machine learning makes it possible to apply the method rapidly to large hyperspectral image sets.

1. Introduction

Close-range imaging spectroscopy measurements of vegetation canopies enable non-invasive, high throughput monitoring applications of the health and overall status of the vegetation very high resolution (Mishra et al., 2020; Peng et al., 2022). A key quantity for these applications is the spectral hemispherical-directional reflectance factor (reflectance or HDRF from hereafter), defined as the ratio between the radiance scattered by a target surface to the radiance scattered by a Lambertian reference surface under identical illumination conditions and observation geometry (Schaepman Strub et al., 2006). However, obtaining these conditions and geometries for each surface visible in the data is practically impossible given the large number of leaves. Furthermore, placing a reference surface within a canopy affects the illumination conditions via multiple scattering. A solution to this issue is typically sought by calibrating the spectra with respect to the top-of-canopy (TOC) surface. However, this TOC reflectance can be significantly different from the true reflectance of an observed surface due to the varying illumination conditions inside the canopy (Takala and Mõttus, 2016). These differences may appear as artifacts in the measured spectra and they can mask important spectral features leading to a misestimation of any material properties. For instance, the multiply scattered photons from within the canopy on an observed surface can cause the TOC reflectance to show spectral features similar to leaf reflectance on materials with no such features Leblanc and Chen (2001). Thus, illumination correction is crucial for accurate vegetation monitoring applications relying on imaging spectroscopy data with a sub-meter spatial resolution.

Generally, illumination correction methods can be categorized into data driven methods and physically based methods. Most data driven

* Corresponding author.
E-mail address: olli.ihalainen@vtt.fi (O. Ihalainen).

https://doi.org/10.1016/j.rse.2024.114467
Received 20 July 2024; Received in revised form 20 September 2024; Accepted 9 October 2024
Available online 30 October 2024

methods are based on normalizing the measured spectra which can reduce the variation between sunlit and shaded pixels of the same material (Ge et al., 2016; Asaari et al., 2018; Roger et al., 2020). However, the methods do not account for wavelength-dependent illumination effects such as multiply scattered radiation from neighboring surfaces.

Physically based methods attempt to directly model the illumination conditions via parametric models. For instance, the COSINE model (Jay et al., 2016), coupled with the PROSPECT leaf radiative transfer model (Feret et al., 2008), can accurately reproduce the true reflectance of a single leaf in a laboratory setup with minimal multiple scattering. More recently, Ihalainen et al. (2023b,a) demonstrated the use of the spectral invariant theory, also called the p-theory, for retrieving the true reflectances of sunlit leaves inside a vegetation canopy. They also showed that the p-theory is not valid for shaded leaves without modifications to the theory. Overall, these methods are only applicable to leaves, given their reliance on forward modeling the leaf reflectance, and they are not applicable to shaded surfaces. Recently, a novel approach to slopes of surfaces from hyperspectral radiance was presented by Carmon et al. (2023), illustrating the potential of the spectral information in a pixel.

In this paper, we present an extended formulation of the spectral invariant theory for vegetation canopies. Using the theory, we can convert the TOC reflectance values in each pixel of a hyperspectral image to the true reflectances of the materials filling these pixels. Next, using random forest regression, we implement the illumination correction method in a computationally fast algorithm and demonstrate its performance on simulated and measured hyperspectral images.

2. Theory

Consider the spectral radiance scattered into the instantaneous field-of-view of a detector above a vegetation canopy by a locally flat area of a surface within the canopy, such that each point on the observed area receives approximately the same irradiance. The hemispherical-directional reflectance factor of the surface, measured with respect to the TOC, is given by

$$\begin{aligned} R_{\mathrm{TOC}}(\Omega_i, \Omega_i', \Omega_s'; \lambda) &:= \pi \frac{I_s(\Omega_i', \Omega_s'; \lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\Omega_i; \lambda)} \\ &= \frac{\phi(\Omega_i'; \lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\Omega_i; \lambda)} S(\Omega_i', \Omega_s'; \lambda) \\ &= k(\Omega_i, \Omega_i'; \lambda) S(\Omega_i', \Omega_s'; \lambda), \end{aligned} \tag{1}$$

where λ is the wavelength, I_s is the radiance scattered by the surface in the direction Ω_s', $\boldsymbol{\Phi}_{\mathrm{TOC}}$ is the total incident irradiance on the TOC from above the canopy, Ω_i is the direction of the Sun in the TOC coordinate system, ϕ is the total incident irradiance on both sides of the surface (4π), including from the direction opposite to the observer for (partially) transmitting surfaces, Ω_i' is the direction of the Sun in the surface coordinate system, $S := \pi \frac{I_s}{\phi}$ is the scattering factor of the surface (Ihalainen and Mõttus, 2022), and $k := \frac{\phi}{\boldsymbol{\Phi}_{\mathrm{TOC}}}$ is the irradiance ratio between the observed surface and the TOC (also called the reflectance conversion factor; Markiet and Mõttus, 2020). In the following text, we refer to the scattering factor as the true reflectance given that the radiance scattered by a surface within a canopy is often mostly due to reflectance rather than transmittance (Ihalainen et al., 2023a). Additionally, we omit the use of the directional terms Ω_i, Ω_i', and Ω_s' in further analysis for the sake of brevity. The irradiance ϕ can be written in terms of the downwelling irradiances from the sky, ϕ_S, and the multiply scattered irradiance from within the canopy, ϕ_M. Thus, we can write the irradiance ratio in terms of these components as

$$k(\lambda) = \frac{\phi_S(\lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\lambda)} + \frac{\phi_M(\lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\lambda)} \tag{2}$$

$$= k_S(\lambda) + k_M(\lambda). \tag{3}$$

Let us next parameterize the irradiance ratios k_S and k_M. To that end, the downwelling irradiances from the sky can generally be decomposed as $\phi_S = \phi_\odot + \phi_D$ where $\phi_\odot$ is the direct solar irradiance and ϕ_D is the diffuse sky irradiance. Consider first the ratio $\frac{\phi_\odot}{\boldsymbol{\Phi}_{\mathrm{TOC}}}$. We note that, $\phi_\odot = \boldsymbol{\Phi}_0 \alpha_\odot |\cos\theta_0'|$ where $\boldsymbol{\Phi}_0$ is the direct normal component of the solar irradiance on the TOC, $\theta_0' \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is the angle between the surface normal and solar direction, i.e., the solar zenith angle (SZA) in the surface coordinate system and $\alpha_\odot$ is a solar-view factor defined such that $\alpha_\odot = 0$ if the Sun is completely obstructed to the surface, $\alpha_\odot \in (0, 1)$ if the surface is in the geometric penumbra of the Sun, and $\alpha_\odot = 1$ otherwise. The absolute value of $\cos\theta_0'$ accounts for the fact that the side of the surface seen by the sensor may not be the sun-facing side of the surface. Correspondingly, for the direct solar irradiance at the TOC we have $\boldsymbol{\Phi}_\odot = \boldsymbol{\Phi}_0 \cos\theta_0$ where $\theta_0 \in [0, \frac{\pi}{2}]$ is the SZA with respect to the TOC surface normal. Given that $\boldsymbol{\Phi}_{\mathrm{TOC}}$ and $\boldsymbol{\Phi}_\odot$ can be related via the direct-to-global irradiance ratio, $d = \frac{\boldsymbol{\Phi}_\odot}{\boldsymbol{\Phi}_{\mathrm{TOC}}}$, we can write

$$\frac{\phi_\odot(\lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\lambda)} = \beta_\odot d(\lambda). \tag{4}$$

where we have defined the solar geometry factor $\beta_\odot := \alpha_\odot \frac{|\cos\theta_0'|}{\cos\theta_0}$. A similar expression can be sough for $\frac{\phi_D}{\boldsymbol{\Phi}_{\mathrm{TOC}}}$. First, assuming isotropic diffuse sky irradiance, we can write $\phi_D = \beta_D \boldsymbol{\Phi}_D$, where β_D is the fraction of unobstructed sky visible on both sides of the surface, i.e., the canopy gap fraction on the surface, also called the sky-view factor by Carmon et al. (2023), and $\boldsymbol{\Phi}_D$ is the diffuse sky irradiance at the TOC. Then, since $\boldsymbol{\Phi}_D$ and $\boldsymbol{\Phi}_{\mathrm{TOC}}$ are related via the diffuse-to-global irradiance ratio, $1 - d = \frac{\boldsymbol{\Phi}_D}{\boldsymbol{\Phi}_{\mathrm{TOC}}}$, we get

$$\frac{\phi_D(\lambda)}{\boldsymbol{\Phi}_{\mathrm{TOC}}(\lambda)} = \beta_D (1 - d(\lambda)). \tag{5}$$

The downwelling irradiance ratio thus depends on only two geometric parameters, $\beta_\odot$ and β_D, as

$$k_S(\lambda) = \beta_D + (\beta_\odot - \beta_D) d(\lambda). \tag{6}$$

Note, that a nearly equivalent expression for k_S can be derived for convex surfaces or shoots, although with slightly different definitions for $\beta_\odot$ and β_D (Mõttus et al., 2015). The multiple scattering irradiance ratio k_M can be derived using the spectral invariant theory (see Appendix) which results in

$$k_M(\lambda) = \frac{\tilde{\rho}\, \tilde{\omega}_0(\lambda) + \tilde{s}_L}{1 - \tilde{p}\tilde{\omega}_0(\lambda)}, \tag{7}$$

where $\tilde{\rho}$, $\tilde{p}$, and $\tilde{s}_L$ are scaled spectral invariant parameters related to directional gap density, photon recollision probability, and canopy interceptance via Eqs. (A.11)–(A.13), respectively, and $\tilde{\omega}_0$ is an empirically determined reference leaf albedo (Knyazikhin et al., 2013).

Given the parameterization of k described above, we arrive at a forward model for the TOC reflectance:

$$R_{\mathrm{TOC}}(\boldsymbol{\vartheta}; \lambda) = \left(\beta_D + (\beta_\odot - \beta_D) d(\lambda) + \frac{\tilde{\rho}\, \tilde{\omega}_0(\lambda) + \tilde{s}_L}{1 - \tilde{p}\, \tilde{\omega}_0(\lambda)} \right) S(\lambda) + \varepsilon(\lambda), \tag{8}$$

where the state vector $\boldsymbol{\vartheta}$ contains the spectral invariants $\boldsymbol{\vartheta}_k = [\beta_\odot, \beta_D, \tilde{\rho}, \tilde{p}, \tilde{s}_L]^T$ as well as the spectra of d, $\tilde{\omega}_0$ and S, and ε contains the modeling and measurement errors. Importantly, if $\boldsymbol{\vartheta}_k$, d and $\tilde{\omega}_0$ are known, we can directly compute the true reflectance S using the measured R_{TOC} and the modeled k as

$$S(\lambda) = \frac{R_{\mathrm{TOC}}(\lambda)}{k(\boldsymbol{\vartheta}_k, d, \tilde{\omega}_0; \lambda)}. \tag{9}$$

3. Materials and methods

3.1. Study sites

We collected hyperspectral images and reference spectrometer data for validating the presented method from two sites: Campus Klein-Altendorf (CKA, 50° 37′ N, 6° 59′ E) in Rheinbach, Germany, between

June 12 and July 6, 2023, and the Otaniemi campus (60° 11′ N, 24° 49′ E) in Espoo, Finland, on August 16th, 2022. The CKA site contained five different types of Brassica in 3 × 3 m^2 field plots. Namely, one *B. juncea*, two *B. napus* and two *B. carinata* genotypes grown in a randomized field design. At the acquisition dates, the plants were in stem elongation phase, shortly before flowering. In Otaniemi, the hyperspectral images were taken of a *Diervilla lonicera* Mill. shrubbery (Ihalainen et al., 2023a). The canopy was dense, homogeneous, and completely closed. At both sites, we placed artificial materials within the canopy as markers and for testing the proposed method on non-vegetation surfaces. At CKA, we placed a brown paper coffee filter in the canopy for testing the method and at Otaniemi we used a plastic leaf as well as a green, a yellow, and a red piece of paper.

3.2. Hyperspectral measurements

The hyperspectral measurements at CKA were carried out using a Specim IQ portable pushbroom imaging spectrometer (SN: 190-1100171, Specim, Finland) on June 17th, 2023. Specim IQ has a spectral range of 400 to 1000 nm, a spectral resolution (full-width-at-half-maximum) of 7 nm, and a peak signal-to-noise ratio of 400:1. The images were captured under nearly cloudless conditions between 14:55 and 15:45, corresponding to SZA values between 27.5° and 30.7°.

The instrument was placed on a tripod, at around 1 m above the TOC in a nadir view geometry, resulting in a spatial resolution of approx. 1.1 mm. To compute the TOC reflectance, a calibrated five-step reference panel (Sphere Optics Zenite lite) with nominal reflectances of approx. 10%, 20%, 30%, 44% and 50% was positioned at the TOC on a tripod, in the field of view of the camera. To avoid saturating the measurements the panel was placed so that only steps with 10%, 20% reflectances were visible in the image. To further prevent saturating the spectra and to obtain good a signal-to-noise ratio, the integration times were selected manually.

The hyperspectral measurements at Otaniemi were performed using another Specim IQ device (SN: 190-1100152) in a setup similar to the CKA measurements, as detailed by Ihalainen et al. (2023a). The TOC reflectance images from both study sites were computed by first performing dark-frame subtraction on the raw image and then dividing the result with the mean spectra of the dark-frame subtracted reference pixels. Finally, we used linear interpolation to resample the bands the closest integer wavelength value. The SZA was approx. 50° during the measurements. Since the reflectance spectra measured outdoors under clear sky conditions with Specim IQ devices can be unstable from around 400 to 415 nm and 925 to 100 nm (Behmann et al., 2018), we clipped the hyperspectral data to contain only wavelengths between 420 and 915 nm.

3.3. Reference and training spectral data

We measured the HDRF of natural and artificial reference materials appearing in the hyperspectral images with spectrometry. At CKA, the measurements were performed using a NaturaSpec field spectrometer (Spectral Evolution, USA). The spectral range of NaturaSpec is from 350 to 2500 nm with spectral resolutions (full-width-at-half-maximum) of 2.7 nm at 700 nm, 5.5 nm at 1500 nm, and 5.8 nm at 2100 nm. The device measures and subtracts the dark current signal automatically from the target spectrum and interpolates the data into 1 nm intervals. To obtain the HDRF of a target within a canopy at CKA, the radiance scattered by the target was measured with a hand-held radiance-calibrated fiber optic cable with a 30° field of view, placed approx. 5 cm above the target at a nadir view angle. A black cardboard sheet with 2.5% nominal reflectance was placed under the sample during measurement to avoid background irradiances. Then, a calibrated 13 cm×13 cm white reference panel was measured in the same position and orientation as the target. The HDRF was computed by dividing the target spectrum with the white reference panel spectrum. This setup was used for measuring the reflectance of 4 leaves in different plots as well as a piece of coffee filter paper placed in these plots. We refer to these leaves as L1–L4 and the coffee filter papers as C1–C4 in the following sections. The genotype of L1 was *Brassica carinata* (Ethiopian mustard) while L2–L4 were of the *Brassica napus* (rapeseed) genotype.

In Otaniemi, the HDRF measurements were performed using an AVASPEC-ULS 2048x64TEC-EVO spectrometer (Avantes, the Netherlands) with a bare fiber optic cable (for details, see Ihalainen et al. (2023a)). The reflectances of 6 reference materials were measured: two *Diervilla lonicera* Mill. leaves (L5 and L6), a plastic leaf (PL), a red paper leaf (RP), a green paper leaf (GP), and a yellow paper leaf (YP).

To provide spectral data for training the proposed illumination correction method (Section 3.7) we performed spectral measurements of non photosynthetic materials and selected leaf spectra from a spectral library by Hovi et al. (2017) and Hovi and Rautiainen (2022). For the non photosynthetic materials, we measured the HDRF of the soil at CKA and conducted laboratory measurements of directional-hemispherical reflectance (DHR) for 5 artificial materials visually similar to the coffee filter paper: four different sheets of brown cardboard and a yellowed newspaper. The soil HDRF was measured from three different positions and then averaged together, using the same measurement setup as with the reference measurements at CKA. The DHR laboratory measurements were conducted with the NaturaSpec spectrometer and an AvaSphere-50-REFL (Avantes, the Netherlands) integrating sphere. The measured target was placed on the sample port of the integrating sphere and covered with the black cardboard sheet to reduce background radiation. For the leaf spectra, we selected the broadleaved species data from the spectral library and computed the mean DHR of the adaxial and abaxial leaf sides of each broadleaved species resulting in 28 leaf spectra. In addition, we selected the calibration spectrum of a gray reference panel with 10% nominal reflectance. The total number of spectra from different materials used in training the proposed method was 35.

Table 1
The 6S radiative transfer model input parameters for computing the direct-to-global irradiance ratio.

Parameter	Unit	Value (Otaniemi)	Value (CKA)
H_2O	g cm^{-2}	1.8	1.7
Ozone	Dobson unit	340	320
AOT at 550 nm		0.18	0.19
Fractions of aerosol types			
Dust-like	%	0.29	0.30
Water-soluble	%	0.66	0.65
Oceanic	%	0.02	0.01
Soot	%	0.03	0.04
SZA	Degree	50	30

3.4. Top-of-canopy spectral irradiance modeling

For both measurement sites, and the ray tracing simulations, we modeled the spectral direct-to-global irradiance ratio at the TOC using the Py6S interface (Wilson, 2013) of the 6S radiative transfer model (Vermote et al., 1997). For the input, 6S takes the atmospheric parameters and SZA. To obtain the atmospheric input parameters, we retrieved the GOCART (Goddard Chemistry Aerosol Radiation and Transport, Chin et al., 2002) atmospheric composition for the CKA and Otaniemi sites for the measurement dates using the Giovanni Web-based tool (http://disc.sci.gsfc.nasa.gov/giovanni). The SZA was computed based on the site coordinates and the time of the hyperspectral measurement.

3.5. Monte Carlo ray traced data

We simulated hyperspectral data using a Monte Carlo ray tracing (MCRT) software developed by the authors (see Mõttus and Stenberg,

Table 2
The minimum and maximum values of the spectral invariants for the sunlit and shaded leaf scenarios used in the global sensitivity analysis.

	$\beta_\odot$		β_D		$\tilde{\rho}$		$\tilde{p}$		$\tilde{s}_L$	
	min	max	min	max	min	max	min	max	min	max
Sunlit	0.32	1.21	0.06	1.0	0.02	0.46	0.32	0.84	−0.02	0.06
Shaded	0.00	0.00	0.00	0.73	0.01	0.45	0.37	0.83	−0.01	0.05

2008, for more details). We refer to this data as MCRT data in the following text to distinguish it from the synthetic data used for training the random forest model. The software was selected given its ability to produce images of the incident irradiances on the visible surfaces (direct solar, diffuse sky, and multiply scattered) and the corresponding spectral invariants, $\beta_\odot$ and β_D (as demonstrated by Ihalainen and Mõttus (2022)) in addition to an orthographic TOC reflectance image. The remaining spectral invariants, ρ and p, and s_L, were computed via least squares regression from Eq. (7). The input of the program includes scene geometry (positions and orientations of the canopy elements), optical properties of each element (spectral and directional scattering properties), camera view direction, solar zenith and azimuth angles, and a spectrum for direct-to-global irradiance ratio at the TOC.

In more detail, we simulated a scene consisting of 138 randomly oriented and positioned identical, disk-like, bi-Lambertian leaves within a 1 m × 1 m area. The leaf radius was 13 cm. For the leaf reflectance and transmittance, we used the spectrum of the leaf #202 (Acer pseudoplatanus L.) in the ANGERS leaf optical properties database (Feret et al., 2008). The forest floor was modeled as a flat Lambertian surface and we simulated two scenarios: a forest floor reflectance factor taken from the understory reflectance data measured by Rautiainen et al. (2012) or set to zero. The simulated spectral bands corresponded to the resampled bands of the Specim IQ, resulting in a total of 204 bands per hypercube between 400 to 1000 nm. The simulated photons were traced from the light sources (the Sun and the sky) to a camera placed at nadir with a spatial resolution 1 cm. The Sun was modeled as a distant light source, i.e., each photon had the initial same direction, and the solar zenith angle (SZA) was 30°. The diffuse sky irradiance was modeled as isotropic radiation. The direct-to-global irradiance ratio spectrum was generated with the 6S model (Section 3.4) using SZA = 30° and atmospheric input parameters corresponding to the Otaniemi site parameters (Table 1).

3.6. Sensitivity analysis of the spectral invariant model

We performed a variance-based global sensitivity analysis (GSA) to quantify the sensitivity of the TOC reflectance model (Eq. (8)) to the spectral invariant parameters for wavelengths between 400 and 1000 nm. Specifically, we used the GSA method by Saltelli et al. (2010) to compute total sensitivity index S_{Ti} for each spectra invariant parameter and the interactions between the parameters.

The sensitivity analysis was performed for two scenarios: a sunlit leaf and a shaded. The GSA used a Saltelli sampler with 512 samples per parameter which yielded 6144 combinations for the five spectral invariants. The minimum and maximum values of $\boldsymbol{\vartheta}_k$ for both scenarios were determined from the MCRT data as the mean value of each spectral invariant plus or minus two times its standard deviation (Table 2). For d, we used the 6S-generated spectra for the CKA site. For $\tilde{\omega}_0$, we used the PROSPECT-generated spectrum with the input parameter values given by Knyazikhin et al. (2013). For S, we used the reflectance of the coffee filter paper, although should be noted that the model sensitivity does not depend on S since none of the model parameters change the spectra of S.

The GSA was implemented in Python using the SALib library (Iwanaga et al., 2022; Herman and Usher, 2017). To compare the relative importance of the total sensitivity index, we normalized it as percentage. We refer to this value as the contribution of a given parameter.

Table 3
The optimized hyperparameters for the random forest regression model. The scikit-learn variable names are in parentheses.

Name	Value
Number of trees (`n_estimators`)	110
Maximum depth of trees (`max_depth`)	26
Minimum samples split (`min_samples_split`)	20
Minimum samples leaf (`min_samples_leaf`)	21
Maximum features (`max_features`)	sqrt
Bootstrap samples (`bootstrap`)	True

3.7. Random forest regression and illumination correction

We trained a random forest (RF) regression algorithm (Breiman, 2001) on synthetic data to predict the spectral invariants from hyperspectral TOC reflectance images. We created the synthetic R_{TOC} spectra for all bands of the Specim IQ camera using the forward model (Eq. (8)) which takes in the direct-to-global irradiance ratio d, reference leaf albedo $\tilde{\omega}_0$, true reflectance S, and the five spectral invariant parameters $\boldsymbol{\vartheta}_k$. For d, we used the 6S-generated spectra for the CKA site. For $\tilde{\omega}_0$, we used the PROSPECT-generated spectrum with the input parameter values given by Knyazikhin et al. (2013). For S, we used the 35 spectra of leaves and non photosynthetic materials (Section 3.3). The values of $\boldsymbol{\vartheta}_k$ were taken from the MCRT data for the scene with a non-zero forest reflectance. More specifically, we identified the sunlit and shaded pixels of the surfaces visible in the image, resulting in a total of 226 surfaces. Then, we computed the corresponding mean values of the spectral invariants, $\langle \boldsymbol{\vartheta}_k \rangle$, for each surface yielding 226 mean values. Note that a surface having one part of its surface sunlit, and another part shaded is counted as two surfaces.

To generate synthetic R_{TOC} spectra for each of the 35 spectra from the spectral measurements and libraries (Section 3.3), we drew 30 random samples of $\boldsymbol{\vartheta}_k$ from a Gaussian distribution for each $\langle \boldsymbol{\vartheta}_k \rangle$ from the MCRT data. We used empirically determined standard deviations for the spectral invariants to produce values not appearing in the simulations: $\sigma_{\boldsymbol{\vartheta}_k} = 10^{-2} \times [5.0, 5.0, 2.0, 2.0, 0.1]^T$. Any sampled spectral invariant with an unrealistic value was adjusted back within range via reflective clipping. To simulate spectral measurement noise, we added zero-mean Gaussian noise with a standard deviation of 0.001 to each R_{TOC} spectrum. The total number of R_{TOC} spectra and $\boldsymbol{\vartheta}_k$ used in the synthetic data was $30 \times 226 \times 35 = 237\,300$.

RF regression is a supervised machine learning method based on an ensemble of decision trees. The predicted values are generated by aggregating the outcomes from all individual decision trees within the ensemble, taking advantage of the overall predictive capability of the ensemble. Additionally, RF can measure the importance of each input feature to the predicted values. We computed this feature importance to discard redundant spectral bands for each Specim IQ band between 420 and 915 nm using a permutation based approach (Breiman, 2001) and we selected the 10 most influential bands based on the importance values. The wavelengths corresponding to these bands were 430, 450, 480, 550, 650, 680, 700, 718, 770, and 790 nm. The RF algorithm was implemented in Python using the scikit-learn library (Pedregosa et al., 2011) and the hyperparameters of the model (Table 3) were tuned using the Optuna hyperparameter optimization framework library (Akiba et al., 2019).

After predicting the spectral invariants with RF, we computed the illumination-corrected spectrum, i.e., the true reflectance, from Eq. (9), using the predicted parameters and along with the site-specific spectra for d and the PROSPECT-generated spectrum for $\tilde{\omega}_0$ as the input. The code for computing the illumination-corrected spectrum was optimized and parallelized for the CPU using the Numba high performance Python compiler (Lam et al., 2015). The code for performing illumination correction with RF is made available on our GitHub repository (https://github.com/mottus/spectralinvariant).

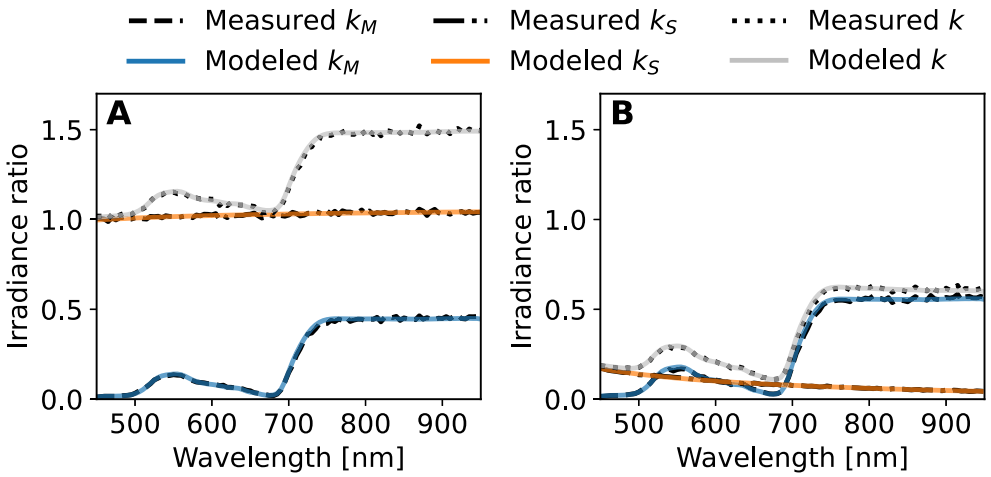

Fig. 1. An example of the differences between the modeled and simulated irradiance ratios, k_S, k_M, and their sum k, for a sunlit pixel (A) and a shaded leaf pixel (B).

3.8. Accuracy estimation

We evaluated the agreement between the modeled and measured or simulated data using the root-mean-squared deviation (RMSD). The result of the random forest inversion was evaluated by first selecting the pixels corresponding to the in-situ measured surfaces in the image and computing their average reflectance. Then, we computed RMSD between average illumination corrected spectra and the corresponding in-situ measured reference reflectance spectra. Similarly, we computed the RMSD between the average measured TOC reflectances and the reference spectra.

4. Results

4.1. Accuracy of the spectral invariant model

The spectra modeled with Eq. (8) showed close agreement with the simulated spectra, and the largest source of error was the simulation noise (Fig. 1). The RMSD values for R_{TOC} was below 0.008 for all pixels, and the averaged RMSD was 0.003. For the irradiance ratios, the RMSD of k_S was always below 0.009, and the averaged RMSD was 0.004. The RMSD for k_M was below 0.017 and the averaged RMSD was 0.008. For the ratio of the total incident irradiances, i.e., the reflectance correction factor, k, the RMSD was below 0.018 for all pixels and the averaged RMSD was 0.009.

4.2. Global sensitivity analysis

The results of the GSA (Fig. 2) showed that the spectral invariants $\tilde{p}$ and $\tilde{\rho}$, corresponding to the multiply scattered irradiance from within the canopy, produce most of the variability in the TOC reflectance spectrum above 710 nm. For the sunlit leaf, the contributions of $\tilde{p}$ and $\tilde{\rho}$ above 710 nm were approx. 50% and 20%, respectively. For the shaded leaf, the contribution of $\tilde{p}$ ranged between approx. 70% and 80% while the contribution of $\tilde{\rho}$ was approx. 30%.

For wavelengths below 710 nm, the solar geometry factor $\beta_{\odot}$ was the largest contributor on the TOC reflectance of a sunlit leaf (Fig. 2a), with values ranging between around 70% and 95%. The contribution of the sky-view factor β_D decreased monotonically, starting from approx. 30% at 400 nm. The contributions of $\tilde{\rho}$ and $\bar{s}_L$ were nearly negligible while $\tilde{p}$ had a slight contribution (<6%), peaking at around 550 nm. The GSA for the shaded leaf showed a much larger contribution for $\tilde{p}$ close to 550 nm, peaking at approx. 50%. The contribution of $\tilde{rho}$ was below 1% whereas the contribution of $\bar{s}_L$ ranged between approx. 5% to 25%. By far, the largest contribution below 710 nm was from β_D, ranging between 40% and 95%.

Table 4
RMSD values between the in-situ measured reflectance spectra and the averaged spectra of the hyperspectral image pixels corresponding to the reference materials. The RMSD was computed for the illumination-corrected image and the TOC reflectance image. Entries without data are marked with ND (no data).

Site	ID	Sunlit		Shaded	
		Corrected RMSD (%)	TOC RMSD (%)	Corrected RMSD (%)	TOC RMSD (%)
CKA	L1	5.1	10.9	4.6	20.0
	L2	7.1	13.4	6.4	15.4
	L3	5.1	10.8	4.3	11.4
	L4	4.7	5.9	5.1	13.9
	C1	5.5	9.1	8.1	45.3
	C2	1.2	9.3	15.5	45.2
	C3	3.6	12.0	8.6	40.5
	C4	2.8	7.1	6.5	35.4
Otaniemi	L5	0.8	16.8	ND	ND
	L6	6.9	22.9	ND	ND
	PL	3.1	11.8	5.3	18.0
	GP	7.0	21.7	38.3	46.6
	RP	7.2	18.3	23.1	49.9
	YP	30.5	36.9	29.4	48.9

4.3. Random forest performance and retrieval algorithm performance

Comparing the RGB images of the TOC reflectance data and the illumination-corrected data from both study sites shows that the proposed method generally reduced the visual differences between the sunlit and shaded pixels of the same material (Fig. 3). Indeed, surfaces hidden completely in shadow such as soil and leaves became more visible in the illumination-corrected images. The algorithm produced reasonable results even for materials that did not have spectrally similar counterparts in the training data such as the red plastic stick at CKA (Fig. 3b), or the red paper leaf at Otaniemi (Fig. 3d).

Overall, the illumination correction produced similar RMSD values for data from both sites, it performed better for sunlit than shaded surfaces, and the RMSD was lower for the corrected spectra than for the TOC reflectances (Table 4). For the CKA data, the highest RMSD value for the corrected spectra was for the shaded pixels of the paper C2 (RMSD = 15.5%) and the lowest value was for the sunlit pixels of the same material (RMSD = 1.2%). For the data from Otaniemi, the highest RMSD for the corrected spectra was for the sunlit pixels of the yellow paper, YP, (RMSD = 30.5%) and the lowest value was for leaf 5, L5, (RMSD = 0.8%) which was entirely sunlit.

A comparison between the in-situ measured spectra, the TOC reflectance spectra, and the illumination-corrected spectra of the reference materials clearly demonstrated how the multiple scattering within

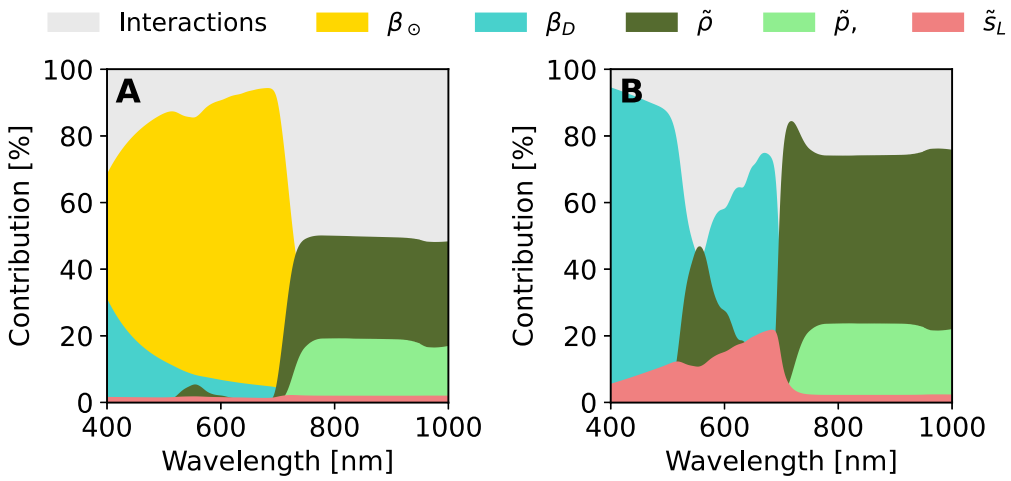

Fig. 2. Contributions of the different input variables to the TOC reflectance modeled using the spectral invariants (Eq. (8)) in the case of a sunlit leaf (A) and shaded leaf (B).

Fig. 3. TOC reflectance images (A, C), and the illumination-corrected images in RGB (B, D; wavelengths 600, 550, 480 nm) from the two sites. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the canopy affected the TOC reflectance spectra while it was diminished in the illumination corrected spectra (Fig. 4). Indeed, typical features of vegetation reflectance spectra such as the small increase in reflectance close to 550 nm and the sharp increase at the red edge (710 to 790 nm) are clearly visible in both sunlit and shaded pixels of the reference sample C1 (Fig. 4e, f).

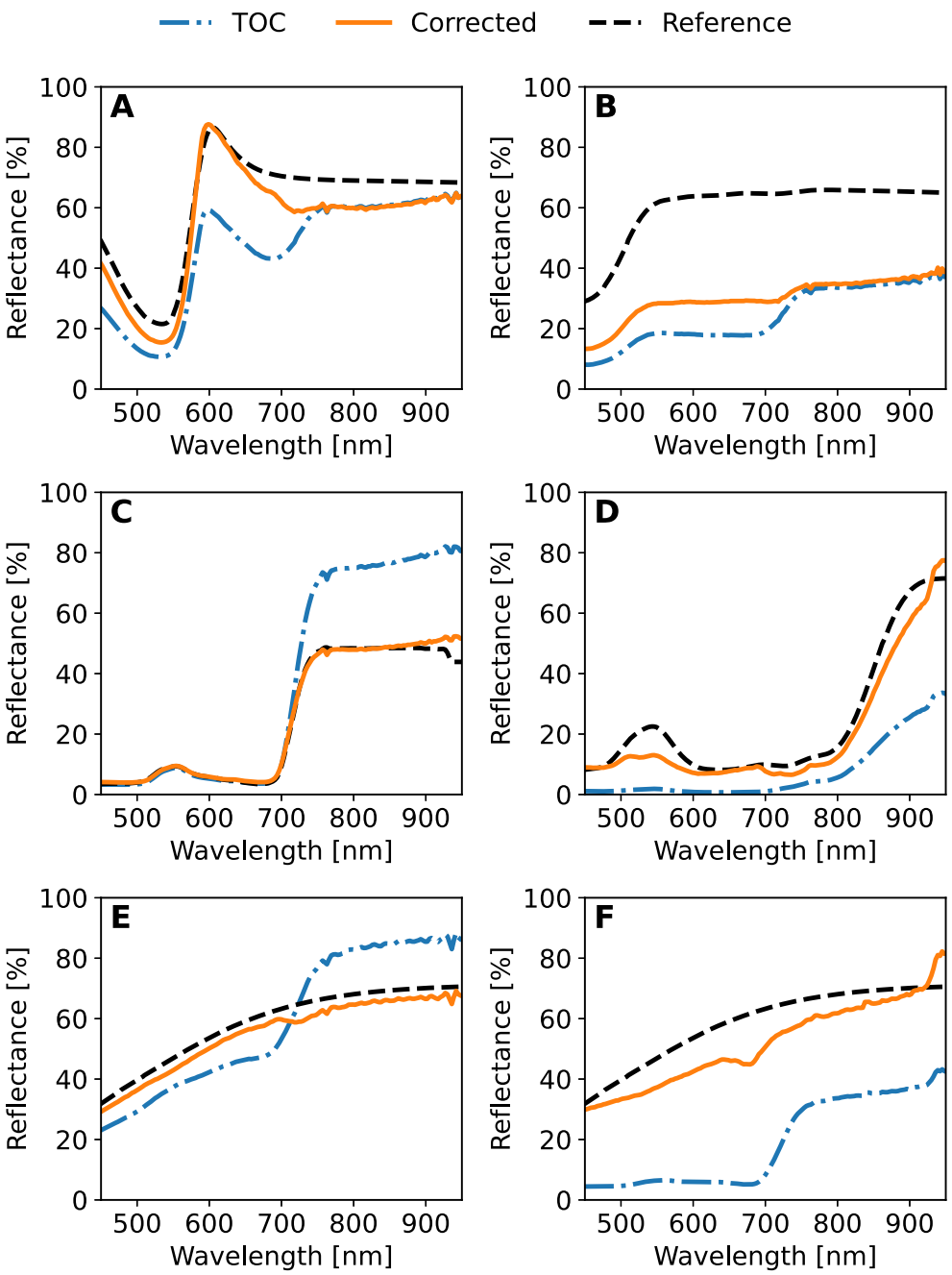

Fig. 4. The in-situ measured reference reflectance spectra and the corresponding averaged TOC reflectance spectra and illumination-corrected spectra from the hyperspectral images for the sunlit pixels of the red paper RP (A), yellow paper YP (B), sunlit leaf L5 (C), shaded plastic leaf PL (D), sunlit paper C3 (E), and shaded paper C3 (F). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For the best result by the algorithm, the sunlit leaf, L5, the illumination-corrected spectrum closely followed the-situ measured spectrum while the TOC reflectance was much higher than the reference spectrum after 700 nm (Fig. 4c). The algorithm produced the worst result for the yellow paper where the corrected spectrum was much lower than the reference spectrum, although the effects of multiple scattered irradiance were still mostly eliminated (Fig. 4B).

The illumination correction algorithm was computationally fast. Predicting the spectral invariants from the TOC reflectance via RF regression and then computing the illumination-corrected hyperspectral images took on the average 0.7 s when benchmarked on a laptop with a 12th Gen Intel® Core i5-1245U processor that had a base clock speed of 1.6 GHz and 10 cores.

5. Discussion

5.1. Modeling TOC reflectance with spectral invariants

Overall, the presented spectral invariant model (Eq. (8)) for TOC reflectance can accurately account for the varying illumination conditions on leaves and other surfaces within a vegetation canopy with five spectral invariant parameters and two spectra. Importantly, the presented model applies to both sunlit and shaded surfaces, whereas previous approaches based on the spectral invariant theory (Ihalainen et al., 2023b,a) were only applicable to sunlit leaves.

When comparing the spectra of the modeled irradiance ratios with their simulated counterparts, the irradiance ratio for the single-scattered radiation, k_S, showed a very low RMSD (Fig. 1). This was

to be expected as the ratio derives from well-established radiometric principles. A more surprising finding was the similarly low RMSD for the multiply scattered irradiance ratio, k_M, given that its model (Eq. (7)) was derived with several simplifications and approximations. Therefore, the results for k_M only confirm that a parametric model taking the form of Eq. (7) (Stenberg et al., 2016; Wang et al., 2018) can closely approximate the spectra of k_M. Hence, the model offers a simple and accurate tool for estimating the magnitude of the multiply scattered irradiance from vegetation on R_{TOC}. While the present study focused on investigating the performance of the model in the VNIR region, we expect the theory also to be applicable to the shortwave infrared region based on our preliminary studies (results not show here). However, confirming this is left for future studies.

The global sensitivity analysis results show which wavelength regions are best suited for the retrieval of a given parameter, whether the used spectral region can be used to reliably estimate the different parameters, and how the uncertainty in a parameter may affect the forward model output. In the case of a sunlit surface, the solar geometry factor $\beta_\odot$ was by far the largest contributor to the variation in R_{TOC} between 400 to 1000 nm (Fig. 2a). The contribution was the largest at around 600 to 700 nm due to the low magnitude of the diffuse sky irradiance and the strong absorption of photons by vegetation in this range. Given the nature of Rayleigh scattering, the magnitude of the diffuse sky irradiance, and thereby the contribution of the sky-view factor β_D to R_{TOC} decreased from 400 to 700 nm. Above 700 nm, the influence of β_D became nearly negligible whereas the contributions of the parameters accounting for the multiply scattered radiation from within the canopy started showing an effect. The contributions of both $\tilde{\rho}$ and $\tilde{p}$ showed relatively similar behavior, indicating that independent retrievals of both parameters can be difficult. In the case of a shaded surface, $\beta_\odot = 0$ and the sky-view factor had the largest influence on R_{TOC} in the 400–700 nm region, particularly at both ends of the region. Notably, the parameter $\tilde{\rho}$ had a prominent contribution at around 550 nm, corresponding to the relatively low absorption of photons by the leaf pigments. Hence, the influence of multiple scattering from vegetation at around 550 nm to R_{TOC} is largely controlled by $\tilde{\rho}$. The parameter $\tilde{s}_L$ had a notable contribution for the shaded leaf below 700 nm whereas the contribution was overall nearly negligible for the sunlit leaf.

The presented forward model shares similarities with the model used by Carmon et al. (2023) for retrieving topographic variables from the radiance measured by an airborne sensor. Carmon et al. model the overall radiance measured by the instrument, including the path radiance from to atmospheric scattering which requires the use of atmospheric radiative transfer modeling software as part of the forward model whereas the presented model is formulated for the TOC reflectance. In contrast, our approach reduces the number of free parameters, which is advantageous for model inversion, but requires knowledge of the downwelling radiation via the direct-to-global irradiance ratio, d. Other differences between the models are related to the definitions of the parameters related to the direct and diffuse fluxes on the observed surfaces which may be of little consequence for surfaces in open areas but can play a larger role in vegetation canopies. Moreover, of their parameters, Carmon et al. "caution against interpreting them directly as geometric properties". The reason for caution is most evident in the definition for their parameter related to direct solar irradiance, which they set equal to the cosine of the SZA with respect to the observed surface, $\cos\theta_0'$, i.e., they do not consider the effect of the geometric penumbra. In the present study, the penumbra is accounted for via the solar-view factor $\alpha_\odot$. A more general difference is that Carmon et al. use the traditional approach of vegetation reflectance factors, making use of an abstract canopy surface, whereas our approach aims to attribute the observed reflected radiance to the physical properties of a scattering element which can be irradiated from any direction inside a canopy treated as a 3D scattering medium.

5.2. Illumination correction

The illumination correction results showed that the proposed method can retrieve the true reflectance of leaves and artificial materials in direct sunlight and shade at very different vegetation canopies with good accuracy (Table 4). Using random forest (RF) regression to retrieve the parameters describing the illumination conditions provided a computationally efficient method for illumination correction that is applicable to any material within a canopy. In comparison, previous illumination correction methods by Ihalainen et al. (2023b,a) were based on minimizing a (regularized) objective function and could describe the optical properties of only one type of material, namely leaves. Also, minimizing an objective function is often computationally costly as it often requires several rounds of iteration to reach the minimum. Depending on the data volume, level of software optimization, and computational resources, this can take from minutes to hours, whereas the method proposed here can produce results in under a second for the whole image, making it well-suited for applications requiring a high-throughput or near real-time application. As a downside, being based on machine learning, our method is limited by the availability data. Hopefully, in the future, spectral data can be retrieved from the images themselves by identifying sunlit pixels where multiple scattering by vegetation has only a small effect.

The RF algorithm performed surprisingly well given that the spectral invariant parameters were taken from relatively simplistic simulated hyperspectral images. This corroborates the usefulness of 3D radiative transfer ray tracing simulations as a source of training data for universal approximators such as RF or neural networks (Verger et al., 2011). However, a lack of spectrally similar training data led to an underperformance for pixels with artificial materials whereas the performance was much better for the pixels containing leaves which did not suffer from this issue (Fig. 3). For instance, the training data did not contain many samples from partially shaded surfaces nor the spectra of artificial materials, resulting in a corrected spectrum with much lower values than the reference spectrum for the yellow paper in the Otaniemi data (Fig. 4b). Furthermore, the training data only contained TOC reflectance spectra from pure pixels, i.e., pixels containing one material. Hence, the trained RF model mispredicts the parameters for spectrally mixed signals causing regions such as the border between leaf edge and ground to appear as yellow. Nevertheless, the proposed method was able to reduce the effect of multiple scattering in pure pixels with natural and artificial materials. The good performance of the RF algorithm on the two sites with very different canopy structures and leaf types indicates its transferability to other sites as well. To improve the overall performance of the RF algorithm, the training spectra should contain a larger coverage of different material types from spectral libraries as well as spectrally mixed signals.

Regardless of how well the presented forward model describes the TOC reflectance or how accurately the RF algorithm predicts the spectral invariants, the illumination-corrected spectra can still contain modeling errors depending on the spectra used for describing the atmosphere and leaves, d and $\bar{\omega}_0$, respectively. Since d was generated by a well-established atmospheric radiative transfer software, the error due to d should generally be relatively low, provided accurate input parameters for the software. The error due to d is the lowest for sunlit pixels where k_s is relatively constant whereas it is much more prominent for shaded pixels where $k_S \approx \beta_D(1-d)$. Additionally, d can be expected to be the same for each pixel, especially for close-range data. In comparison, determining the spectrum of $\bar{\omega}_0$ is more heuristic and an ill-chosen spectrum can cause anomalous features in the corrected spectra. This can be seen, e.g., in the spectra of the coffee filter paper C2 (Fig. 4f) where the feature at around 700 nm can be attributed to the modeled k_M having higher values than the true k_M, meaning that the albedo fits badly to the true k_M regardless of the fitting parameters. Moreover, using a leaf albedo for $\bar{\omega}_0$ for an open and heterogeneous canopy ignores the influence of non photosynthetic

materials on the multiple scattered irradiance. On the other hand, for relatively dense and homogeneous vegetation canopies, using the a leaf albedo for $\tilde{\omega}_0$ is a reasonable approach, as evident from the results for the leaves L5 (Fig. 4c) and L6. A possible solution to obtaining an optimal reference albedo for a given canopy would be to apply the inversion method by Ihalainen et al. (2023a) to Eq. (A.10). For instance, one can take the mean HDRF spectrum of the canopy to estimate the spectral invariants $\tilde{\rho}$, $\tilde{p}$, and $\tilde{s}_L$ that best fit Eq. (A.10) as well as $\tilde{\omega}_0(\tilde{\boldsymbol{\beta}})$, parameterized by leaf traits, $\tilde{\boldsymbol{\beta}}$. The resulting spectrum $\tilde{\omega}_0(\tilde{\boldsymbol{\beta}})$ then describes the reference leaf albedo of the canopy. However, the validity and performance of this method needs to be investigated by future studies.

In addition to the modeling errors, the hyperspectral data and the spectrometer data may also contain errors due to instrument noise, calibration, and measurement protocol, as discussed by Ihalainen et al. (2023a). The errors were mitigated via careful consideration of the sensor properties, meteorological and radiation conditions, measurement setup, and measurement protocol (Section 3). Despite these considerations, the signal-to-noise ratio of shaded pixels in hyperspectral images can still be suboptimal. This results in the overall better performance of the presented method for sunlit than for shaded surfaces.

We used RF to predict the spectral invariants, which is one of the many possible universal approximators usable for the task, such as XGBoost (Chen and Guestrin, 2016), lightGBM (Ke et al., 2017), Gaussian processes (Rasmussen, 2003), neural networks (Heaton, 2018). Alternatively, convolutional neural networks could also leverage information from the neighboring pixels to better predict the spectral invariants as the multiply scattered irradiance does usually not vary rapidly inside a canopy. Using the pixel neighborhood could also be used in methods based on finding the minimum of an objective function to produce informative constraints to the solution.

Our method is also applicable to large mixed pixels containing multiple materials such as data collected via airborne measurements, as indicated by the results of Mõttus et al. (2017), Markiet and Mõttus (2020), Carmon et al. (2023), and the derivations by Mõttus et al. (2015). For vegetation canopies, the resulting illumination-corrected spectrum would be that of an average canopy element in the pixel (Takala and Mõttus, 2016). However, this still needs to be confirmed by future studies. While the illumination-corrected images are useful for revealing objects hidden in shadow, the corrected spectra should also provide more accurate retrievals for traits such as leaf chlorophyll content than the TOC reflectance. Hence, future research should focus on combining the presented method with a leaf model such as PROSPECT (Féret et al., 2017) or a spectral invariant based parameterization (Wu et al., 2021) for leaf trait retrieval. Finally, since the presented method effectively estimates the amount of radiation arriving on and scattered by a given surface, the method could also be extendable to account for emitted radiation, such as solar induced chlorophyll fluorescence.

6. Conclusion

In this work, we developed a physically based illumination correction method for hyperspectral TOC reflectance images of vegetation in the VNIR wavelength region. The illumination conditions were accounted for by a mathematical model for the TOC reflectance that considers the irradiance incident on a physical surface (leaf or another material) in the vegetation canopy using a set of spectral invariant parameters related to the different irradiance components: direct solar beam, diffuse sky illumination, and multiple scattering inside the canopy.

To obtain a computationally efficient illumination correction method that works for various types of materials, we trained a random forest regression algorithm to predict the spectral invariants from the measured TOC reflectance spectra. Applying the method to measured hyperspectral data demonstrated that it can accurately retrieve the true reflectance spectrum of the observed surface when the random forest has been trained on sufficiently similar spectral data.

CRediT authorship contribution statement

Olli Ihalainen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Theresa Sandmann:** Writing – review & editing, Writing – original draft, Resources, Methodology, Data curation. **Uwe Rascher:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition. **Matti Mõttus:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the Research Council of Finland (grants 322256, and 348035), as well as the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Germany under Germany's Excellence Strategy – EXC 2070 – 390732324, and by the 'Strukturwandel-Projekt BioökonomieREVIER', which is funded by the German Federal Ministry of Education and Research, Germany (project identification number 031B0918A). We greatly thank the staff at Campus Klein-Altendorf for providing the field experiments, which were used to record the data for this study. We also thank Yunseon Lee for her assistance in testing the applicability of decision tree-based regression methods of hyperspectral data during her internship at VTT.

Appendix. Irradiance ratio for multiply scattered photons within a canopy

Let us first consider the traditional formulation of p-theory, where we assume that the impact of canopy background on canopy reflectance is negligible. More specifically, we follow the supporting Information by Knyazikhin et al. (2013). They approximate canopy spectral reflectance as

$$\mathrm{HDRF}(\Omega,\lambda)=\frac{i_0\rho(\Omega)}{1-p\omega_L(\lambda)}\omega_L(\lambda), \tag{A.1}$$

where $\tilde{\omega}_0$ is the leaf albedo, i_0 is the canopy interceptance, $\rho(\Omega)$ is the directional gap density and p is the photon recollision probability. The leaf albedo can be written in terms of leaf interceptance, i_L, fraction of surface-reflected radiation, s_L, and a transformed leaf albedo, $\tilde{\omega}$ (S3.1)

$$\omega_L(\lambda)=i_L\tilde{\omega}(\lambda)+s_L. \tag{A.2}$$

The fraction $s_L = 1 - i_L$ is assumed to be a wavelength-independent function of leaf surface properties. Knyazikhin et al. (2013) approximate $\omega_L \approx i_L\tilde{\omega}$ in the 710–790 nm spectral interval, since the diffuse leaf albedo dominates in this region. However, we will take a different route and keep the term s_L in the equations. Substituting Eq. (A.2) into Eq. (A.1) and solving for $\mathrm{HDRF}/\tilde{\omega}$ yields

$$\frac{\mathrm{HDRF}(\Omega,\lambda)}{\tilde{\omega}(\lambda)}=\frac{1}{1-ps_L}\left(i_Li_0\rho(\Omega)+i_Lp\mathrm{HDRF}(\Omega,\lambda)+\frac{i_0\rho(\Omega)s_L}{\tilde{\omega}(\lambda)}\right). \tag{A.3}$$

Let us now define a new set of spectral invariant parameters:

$$\rho'(\Omega)=\frac{i_Li_0\rho(\Omega)}{1-ps_L} \tag{A.4}$$

$$p'=\frac{i_Lp}{1-ps_L} \tag{A.5}$$

$$s'_L(\Omega)=\frac{i_0\rho(\Omega)s_L}{1-ps_L}. \tag{A.6}$$

Therefore, Eq. (A.3) simplifies to

$$\frac{\mathrm{HDRF}(\Omega,\lambda)}{\tilde{\omega}(\lambda)} = \rho'(\Omega) + p'\mathrm{HDRF}(\Omega,\lambda) + \frac{s'_L}{\tilde{\omega}(\lambda)}, \tag{A.7}$$

from which we can get a new formulation for the HDRF as

$$\mathrm{HDRF}(\Omega,\lambda) = \frac{\rho'(\Omega)\tilde{\omega}(\lambda) + s'_L(\Omega)}{1 - p'\tilde{\omega}(\lambda)}. \tag{A.8}$$

This formulation extends the p-theory to wavelengths outside of the red edge (710–790 nm) spectral interval.

Similar to the spectral invariant relationship between the canopy HDRF and the albedo of an average leaf in the canopy (Eq. (A.8)), any transformed leaf albedo can be related to a fixed reference spectrum, $\tilde{\omega}_0$ as

$$\tilde{\omega}(\lambda) = \frac{\rho_L\,\tilde{\omega}_0(\lambda) + c_L}{1 - p_L\,\tilde{\omega}_0(\lambda)}, \tag{A.9}$$

where p_L is the within-leaf recollision probability, $\rho_L = 1 - p_L$ is the within-leaf escape probability, and c_L is the fraction of reflected incident radiation. Substituting Eq. (A.9) into Eq. (A.1) via Eq. (A.2) thus yields

$$\mathrm{HDRF}(\Omega,\lambda) = \frac{\tilde{p}\,\tilde{\omega}_0(\lambda) + \tilde{s}_L}{1 - \tilde{p}\,\tilde{\omega}_0(\lambda)}, \tag{A.10}$$

where the new spectral invariant parameters are

$$\tilde{\rho} = \frac{1}{1 - p(s_L - i_L c_L)}\left(p_L(1 - p s_L) + i_L \rho_L p\right) \tag{A.11}$$

$$\tilde{p} = \frac{1}{1 - p(s_L - i_L c_L)}\left(i_L i_0 \rho_L \rho - s_L p_L\right) \tag{A.12}$$

$$\tilde{s}_L = \frac{1}{1 - p(s_L - i_L c_L)}\left(s_L + i_L \rho_L c_L\right). \tag{A.13}$$

An expression for the irradiance ratio, k_M, can be sought from basic radiometric definitions and Eq. (A.1). Given that the canopy reflectance is defined as $\mathrm{HDRF} = \pi I_C/\Phi_{\mathrm{TOC}}$, we can write the radiance, I_C, scattered from within the canopy toward a leaf as

$$\begin{aligned} I_C(\Omega) &= \pi^{-1}\Phi_{\mathrm{TOC}}(\lambda)\mathrm{HDRF}(\lambda,\Omega) \\ &= \pi^{-1}\Phi_{\mathrm{TOC}}(\lambda)\frac{i_0\rho(\Omega)}{1 - p\omega_L(\lambda)}\omega_L(\lambda). \end{aligned} \tag{A.14}$$

Thus, the incident irradiance on the leaf from all directions within the canopy is given by

$$\begin{aligned} \phi_C(\lambda) &= \int_{4\pi} I_C(\Omega)|\mu|d\Omega \\ &= \pi^{-1}\Phi_{\mathrm{TOC}}(\lambda)\frac{i_0}{1 - p\omega_L(\lambda)}\omega_L(\lambda)\int_{4\pi}\rho(\Omega)|\mu|d\Omega \\ &= \Phi_{\mathrm{TOC}}(\lambda)\frac{i_0\rho}{1 - p\omega_L(\lambda)}\omega_L(\lambda), \end{aligned}$$

where $\rho = \pi^{-1}\int_{4\pi}\rho(\Omega)|\mu|d\Omega$. Hence, we finally get an expression for k_M:

$$k_M(\lambda) = \frac{\phi_C(\lambda)}{\Phi_{\mathrm{TOC}}(\lambda)}$$

$$= \frac{i_0\rho\omega_L(\lambda)}{1 - p\omega_L(\lambda)} \tag{A.15}$$

$$= \frac{\rho'\tilde{\omega}(\lambda) + s'_L}{1 - p'\tilde{\omega}(\lambda)} \tag{A.16}$$

$$= \frac{\tilde{\rho}\tilde{\omega}_0(\lambda) + \tilde{s}_L}{1 - \tilde{p}\tilde{\omega}_0(\lambda)}, \tag{A.17}$$

where the spectral invariant parameters have been integrated over 4π.

Data availability

I have shared a link to the code. The data will be made available on request.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2623–2631.

Asaari, M.S.M., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., Scheunders, P., 2018. Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. ISPRS J. Photogramm. Remote Sens. 138, 121–138.

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M.T., Jussila, J., Salo, H., et al., 2018. Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. Sensors 18 (2), 441.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Carmon, N., Berk, A., Bohn, N., Brodrick, P.G., Dozier, J., Johnson, M., Miller, C.E., Thompson, D.R., Turmon, M., Bachmann, C.M., et al., 2023. Shape from spectra. Remote Sens. Environ. 288, 113497.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B.N., Duncan, B.N., Martin, R.V., Logan, J.A., Higurashi, A., Nakajima, T., 2002. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements. J. Atmos. Sci. 59 (3), 461–483.

Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., Ustin, S.L., Le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112 (6), 3030–3043.

Féret, J.-B., Gitelson, A., Noble, S., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215.

Ge, Y., Bai, G., Stoerger, V., Schnable, J.C., 2016. Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. Comput. Electron. Agric. 127, 625–632.

Heaton, J., 2018. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618. Genet. Program. Evol. Mach. 19 (1), 305–307.

Herman, J., Usher, W., 2017. SALib: An open-source Python library for sensitivity analysis. J. Open Source Softw. 2 (9), http://dx.doi.org/10.21105/joss.00097.

Hovi, A., Raitio, P., Rautiainen, M., 2017. A spectral analysis of 25 boreal tree species. Silva Fennica 51 (4), http://dx.doi.org/10.14214/sf.7753.

Hovi, A., Rautiainen, M., 2022. Leaf and needle spectra for 25 boreal tree species. http://dx.doi.org/10.17632/nvgjcn5nsx.1.

Ihalainen, O., Juola, J., Mõttus, M., 2023a. Physically based illumination correction for sub-centimeter spatial resolution hyperspectral data. Remote Sens. Environ. 298, 113810.

Ihalainen, O., Mõttus, M., 2022. Spectral invariants in ultra-high spatial resolution hyperspectral images. J. Quant. Spectrosc. Radiat. Transfer 108265.

Ihalainen, O., Mõttus, M., Juola, J., 2023b. Inversion of true leaf reflectance from very high spatial resolution hyperspectral images. In: IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 7579–7582.

Iwanaga, T., Usher, W., Herman, J., 2022. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. Soc.-Environ. Syst. Model. 4, 18155. http://dx.doi.org/10.18174/sesmo.18155, URL: https://sesmo.org/article/view/18155.

Jay, S., Bendoula, R., Hadoux, X., Féret, J.-B., Gorretta, N., 2016. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. Remote Sens. Environ. 177, 220–236.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.

Knyazikhin, Y., Schull, M.A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., Marshak, A., Carmona, P.L., Kaufmann, R.K., Lewis, P., Disney, M.I., Vanderbilt, V., Davis, A.B., Baret, F., Jacquemoud, S., Lyapustin, A., Myneni, R.B., 2013. Hyperspectral remote sensing of foliar nitrogen content. Proc. Natl. Acad. Sci. USA 110 (3), http://dx.doi.org/10.1073/pnas.1210196109.

Lam, S.K., Pitrou, A., Seibert, S., 2015. Numba: A llvm-based python jit compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. pp. 1–6.

Leblanc, S.G., Chen, J.M., 2001. A practical scheme for correcting multiple scattering effects on optical LAI measurements. Agricult. Forest. Meterol. 110 (2), 125–139.

Markiet, V., Mõttus, M., 2020. Estimation of boreal forest floor reflectance from airborne hyperspectral data of coniferous forests. Remote Sens. Environ. 249, 112018.

Mishra, P., Lohumi, S., Khan, H.A., Nordon, A., 2020. Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. Comput. Electron. Agric. 178, 105780.

Mõttus, M., Hernández-Clemente, R., Perheentupa, V., Markiet, V., 2017. In situ measurement of Scots pine needle PRI. Plant Methods 13 (1), http://dx.doi.org/10.1186/s13007-017-0184-4.

Mõttus, M., Stenberg, P., 2008. A simple parameterization of canopy reflectance using photon recollision probability. Remote Sens. Environ. 112 (4), 1545–1551. http://dx.doi.org/10.1016/j.rse.2007.08.002.

Mõttus, M., Takala, T.L., Stenberg, P., Knyazikhin, Y., Yang, B., Nilson, T., 2015. Diffuse sky radiation influences the relationship between canopy PRI and shadow fraction. ISPRS J. Photogramm. Remote Sens. 105, 54–60. http://dx.doi.org/10.1016/j.isprsjprs.2015.03.012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Peng, H., Cendrero-Mateo, M.P., Bendig, J., Siegmann, B., Acebron, K., Kneer, C., Kataja, K., Muller, O., Rascher, U., 2022. HyScreen: A ground-based imaging system for high-resolution red and far-red solar-induced chlorophyll fluorescence. Sensors 22 (23), 9443.

Rasmussen, C.E., 2003. Gaussian processes in machine learning. In: Summer School on Machine Learning. Springer, pp. 63–71.

Rautiainen, M., Mõttus, M., Yáñez-Rausell, L., Homolová, L., Malenovskỳ, Z., Schaepman, M.E., 2012. A note on upscaling coniferous needle spectra to shoot spectral albedo. Remote Sens. Environ. 117, 469–474.

Roger, J.-M., Biancolillo, A., Marini, F., 2020. Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy. Chemometr. Intell. Lab. Syst. 199, 103975.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181 (2), 259–270.

Stenberg, P., Mõttus, M., Rautiainen, M., 2016. Photon recollision probability in modelling the radiation regime of canopies — A review. Remote Sens. Environ. 183, 98–108. http://dx.doi.org/10.1016/j.rse.2016.05.013, URL: https://www.sciencedirect.com/science/article/pii/S0034425716302140.

Schaepman Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensing-definitions and case studies. Remote Sens. Environ. 103 (1), 27–42. http://dx.doi.org/10.1016/j.rse.2006.03.002.

Takala, T.L., Mõttus, M., 2016. Spatial variation of canopy PRI with shadow fraction caused by leaf-level irradiation conditions. Remote Sens. Environ. 182, 99–112. http://dx.doi.org/10.1016/j.rse.2016.04.028.

Verger, A., Baret, F., Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens. Environ. 115 (2), 415–426.

Vermote, E.F., Tanré, D., Deuze, J.L., Herman, M., Morcette, J.-J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. IEEE Trans. Geosci. Remote Sens. 35 (3), 675–686.

Wang, W., Nemani, R., Hashimoto, H., Ganguly, S., Huang, D., Knyazikhin, Y., Myneni, R., Bala, G., 2018. An interplay between photons, canopy structure, and recollision probability: A review of the spectral invariants theory of 3d canopy radiative transfer processes. Remote Sens. 10 (11), 1805.

Wilson, R.T., 2013. Py6S: A python interface to the 6S radiative transfer model. Comput. Geosci. 51, 166–171.

Wu, S., Zeng, Y., Hao, D., Liu, Q., Li, J., Chen, X., Asrar, G.R., Yin, G., Wen, J., Yang, B., et al., 2021. Quantifying leaf optical properties with spectral invariants theory. Remote Sens. Environ. 253, 112131.