Simulating Imaging Spectroscopy in Tropical Forest with 3D Radiative Transfer Modeling
Dav Ebengo, Florian de Boissieu, Grégoire Vincent, Christiane Weber, Jean-Baptiste Feret

To cite this version:

HAL Id: hal-03255512
https://hal.inrae.fr/hal-03255512
Submitted on 9 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Simulating Imaging Spectroscopy in Tropical Forest with 3D Radiative Transfer Modeling

Dav M. Ebengo 1,*, Florian de Boissieu 1,2,3*, Grégoire Vincent 4*, Christiane Weber 1* and Jean-Baptiste Féret 1*

Abstract: Optical remote sensing can contribute to biodiversity monitoring and species composition mapping in tropical forests. Inferring ecological information from canopy reflectance is complex and data availability suitable to such a task is limiting, which makes simulation tools particularly important in this context. We explored the capability of the 3D radiative transfer model DART (Discrete Anisotropic Radiative Transfer) to simulate top of canopy reflectance acquired with airborne imaging spectroscopy in a complex tropical forest, and to reproduce spectral dissimilarity within and among species, as well as species discrimination based on spectral information. We focused on two factors contributing to these canopy reflectance properties: the horizontal variability in leaf optical properties (LOP) and the fraction of non-photosynthetic vegetation (NPVf). The variability in LOP was induced by changes in leaf pigment content, and defined for each pixel based on a hybrid approach combining radiative transfer modeling and spectral indices. The influence of LOP variability on simulated reflectance was tested by considering variability at species, individual tree crown and pixel level. We incorporated NPVf into simulations following two approaches, either considering NPVf as a part of wood area density in each voxel or using leaf brown pigments. We validated the different scenarios by comparing simulated scenes with experimental airborne imaging spectroscopy using statistical metrics, spectral dissimilarity (within crowns, within species, and among species dissimilarity) and supervised classification for species discrimination. The simulation of NPVf based on leaf brown pigments resulted in the closest match between measured and simulated canopy reflectance. The definition of LOP at pixel level resulted in conservation of the spectral dissimilarity and expected performances for species discrimination. Therefore, we recommend future research on forest biodiversity using physical modeling of remote-sensing data to account for LOP variability within crowns and species. Our simulation framework could contribute to better understanding of performances of species discrimination and the relationship between spectral variations and taxonomic and functional dimensions of biodiversity. This work contributes to the improved integration of physical modeling tools for applications, focusing on remotely sensed monitoring of biodiversity in complex ecosystems, for current sensors, and for the preparation of future multispectral and hyperspectral satellite missions.

Keywords: diversity mapping; imaging spectroscopy; leaf traits; radiative transfer; DART; PROSPECT

1. Introduction

Tropical forest ecosystems host at least two-thirds of the terrestrial biodiversity and contain about 25% of the carbon in the terrestrial biosphere [1,2]. The loss of biodiversity currently observed worldwide strongly impacts tropical forests [3]. This decline of biodiversity is induced by anthropogenic factors including land-use change, exploitation of natural
resources, and climate change, leading to forest disturbance and disappearance [4–7], with potential feedback due to climate-biosphere interactions [8].

Many studies have reported the relationship between ecosystem functions and biodiversity [9–12], others highlight ecosystem services provided by the biodiversity of such forest ecosystems and their importance to sustaining life on Earth [13,14]. Thus in order to monitor biodiversity and ensure their conservation, a set of so-called Essential Biodiversity Variables (EBVs) has been defined by the Group on Earth Observations—Biodiversity Observation Network (GEOBON) [15]. EBVs enable the study, reporting and integration of change in biodiversity into local- to global-scale management plans [16]. The difficulty to collect ecological information from ground observation over large spatial scales and in a consistent and repeatable manner led to the use of remote sensing [17], and the identification of remote-sensing enabled EBVs (RS-EBVs) and satellite remote sensing (SRS) EBVs [18].

Remote sensing has been shown to be a powerful tool to track biodiversity [19–21]. As such, studies focusing on remote sensing of terrestrial plant diversity have been classified into four categories according to Wang and Gamon [22]: (i) indirect estimation of biodiversity derived from habitat mapping performed at moderate spatial resolution; (ii) fine scale mapping of species distribution [23–25] or functional types [26] (iii) estimation of functional diversity through leaf traits estimated from remote-sensing information [27–29]; (iv) estimation of diversity based on the spectral variation hypothesis (SVH). The SVH initially formulated by Palmer et al. [30] relates the spatial variability of spectral data (spectral heterogeneity or variation) to environmental heterogeneity, a major factor explaining species distribution and functional diversity. Reviews on this hypothesis stated that this assumption, however, overlooks a number of factors such as sensor specifications, target vegetation types or metrics derived from remotely sensed data [31–33].

Imaging spectroscopy has shown strong potential for diversity mapping using the aforementioned approaches [25,28,34,35], as these methods can take advantage of fine spatial scale, high spectral sampling and high signal-to-noise ratio (SNR) provided by airborne acquisitions. Despite the potential of airborne imaging spectroscopy for the study of biodiversity, its use for regional- to global-scale monitoring of ecosystems is currently impossible for logistical and financial reasons. Current and forthcoming spaceborne imaging spectroscopy sensors open perspectives for improved capacity to transfer methodologies developed to monitor biodiversity based on airborne imaging spectroscopy at broader scales. In this perspective, we need to better understand the influence of key drivers linking vegetation diversity to spectral diversity, and understand the contribution of multiple factors which can be intrinsic and extrinsic to vegetation. Such understanding will contribute to preparing, testing, and identifying limitations of methodologies dedicated to forest diversity mapping and applicable to spaceborne imaging spectroscopy. Optical remote-sensing information acquired from forest ecosystems is driven by interactions between incoming sunlight in the visible and infrared domains, and biophysical, biochemical and structural properties of vegetation [36]. These include leaf traits (chemical constituents, morphological properties), foliage density and orientation, and the structural properties of woody elements (trunks, branches, twigs). In addition to these factors intrinsic to vegetation, other factors such as background (ground reflectance), and the conditions of acquisition, including atmospheric properties, geometry of acquisition, as well as sensor configuration also influence the signal acquired with an airborne or spaceborne sensor [36,37].

In order to better understand the influence of the different factors influencing canopy reflectance measured with an optical sensor, simulation tools based have been developed during the past 50 years. These models usually rely on simulations of bi-directional reflectance factor (BRF) [38]. Three major types of BRF model are used to simulate canopy reflectance. Empirical models give a mathematical description of BRF without attempting to explain the biophysical and chemical processes that govern BRF [39]. Geometric optical (GO) reflectance models consider physical dimension and structure of forest to understand the BRF [40,41]. Radiative transfer models (RTMs), also known as physical modeling,
describe interactions between incident solar radiation and physical environments such as tropical forests [42]. They allow explicit description of interactions between incoming sun radiation and vegetation through structural and biochemical properties based on the fundamental equations of radiative transfer. RTMs are more robust than empirical methods and GO reflectance models.

Numerous approaches have been implemented to model radiative transfer in terrestrial environments. One-dimensional models such as PROSAIL [43] have been widely used to simulate radiative transfer in spatially homogeneous canopies like crops, grasslands, and closed forests [44,45]. In complex and spatially heterogeneous scenes such as tropical forests, 3D models are more suitable since they account for vertical and horizontal heterogeneity, and include shading and multiple scattering [46]. Three-dimensional (3D) RTMs reconstruct the architecture of a medium through geometric primitives such as triangles, cubes, cones, spheres or cylinders [47]. The combination of these elements with the other inputs of the RTMs allows generating 3D scenes of different levels of complexity.

Simple geometric structures derived from dendrometric measurements were initially used to approximate the forest structure in the first studies using 3D RTM [47]. The recent development of light detection and ranging (LiDAR) provided an unprecedented level of detail to characterize forest 3D architecture [48]. Schneider et al. [49] evidenced the interest of using detailed 3D structure of vegetation derived from LiDAR compared to simpler 3D representations of individual trees defined as geometric shapes and located in the scene based on field inventories: they obtained increased accuracy when comparing experimental imaging spectroscopy with simulations using LiDAR-derived 3D structure.

In the current era of deep learning, RTMs are also strong candidates in the perspectives of data augmentation [50] to alleviate challenges related to data sparsity in order to understand interactions between vegetation and solar radiation, to test potential and limits of methods, and to prepare future missions. DART (Discrete Anisotropic Radiative Transfer) [51,52] is currently one of the most comprehensive 3D RTMs for simulating remotely sensed data, in particular in the context of forest studies [42,53]. However, the proper parameterization of DART in order to account for forest structural and radiometric properties requires particular care, and needs to be tuned for the application considered. Malenovský et al. [54] showed the influence of including woody elements in 3D simulations in order to retrieve Leaf Area Index (LAI) and simulate canopy reflectance for a Norway spruce forest. Strong assumptions on the homogeneity of leaf optical properties (LOP) within species or individuals are also usually made when using 3D models to study heterogeneous vegetation. Schneider et al. [49] simulated imaging spectroscopy with DART on a temperate mixed forest in order to study the potential of upscaling leaf level observations, and put efforts on characterization of the vertical heterogeneity of LOP for each tree species, but used the same LOP for a given layer and all individual trees from the same species. Ferreira et al. [55] studied the contribution of DART simulations to the retrieval of structural and chemical properties of individual tree crowns (ITCs) in dry tropical forests, and assumed homogeneity of LOP in ITCs. The assumptions made by the authors were appropriate to meet their objective. However, accounting for intra-species and intra-crown variability of LOP is potentially crucial for the production of realistic simulations aiming at exploring the potential of methods dedicated to biodiversity monitoring since there is spectral variability within crowns, within species and among species in tropical forest [56]. Thus, the hypothesis of uniform LOP at either crown or species scale may lead to dramatic underestimation of spectral heterogeneity in simulation of forest stand scale remotely sensed data, making these simulations unsuitable for applications built upon the fine analysis of spatial heterogeneity of spectral information, such as biodiversity mapping.

The lack of information on the non-photosynthetic vegetation fraction (NPVF) and an inadequate characterization of the diversity of LOP have been shown to contribute to the relatively low radiometric agreement between simulation and experimental data [49,54,57]. Simulating the variability in canopy reflectance is challenging in many ways, as it requires definition of fine scale 3D structure and heterogeneity of LOP. While airborne laser
scanning (ALS) potentially fulfills the requirements for the definition of the 3D structure, the definition of LOP also needs to be addressed using indirect estimation, as sampling within and among species variability in optical properties in hyperdiverse ecosystems is not feasible. To address those challenges, we designed a methodology aiming at simulating spatial heterogeneity in spectral data over a complex tropical forest using the DART model. We used ALS data in order to reconstruct 3D forest architecture. Then, we defined scenarios to account for NPVf using different proxies based on 3D structure or canopy reflectance. Once an optimal scenario to characterize NPVf was defined, we performed simulations accounting for LOP variability at different scales (pixel, crown, species). Finally, we compared simulations to airborne imaging spectroscopy using statistical metrics, dissimilarity metric and supervised classification methods. The main objective of this study was to define a framework for producing realistic simulations in the perspective of supporting methodological development for biodiversity assessment using 3D modeling, considering: (i) the effects of incorporation of NPVf into simulation and (ii) the influence of LOP variability on the spectral response simulated by the DART model.

2. Materials

2.1. Study Area

The study area is located in the coastal part of French Guiana at Paracou research station (5°18’ N, 52°53’ W). It is characterized by a lowland terra firme rainforest. The forest is dominated by Leguminosae-Caesalpinioideae, Lecythidaceae, Chrysobalanaceae, and Sapotaceae families. Field plot network inventories carried out almost every year for more than 35 years highlight a large specific diversity with more than 750 tree species identified over 118.75 ha. A complete description of the Paracou experimental site can be found on https://paracou.cirad.fr/website (accessed on 2 March 2021).

2.2. Remotely Sensed Data

In September 2016, an airborne campaign funded by the Centre National d’Etudes Spatiales (CNES) and the Center for the study of biodiversity in Amazonia (labex CEBA) led to the acquisition of imaging spectroscopy, LiDAR data, and very high spatial resolution red, green and blue (RGB) imagery. The acquisitions were performed in clear sky conditions at an average flying altitude of 890 m.

2.2.1. Imaging Spectroscopy

Imaging spectroscopy data were acquired with a Hyspex VNIR-1600 sensor (Hyspex NEO, Skedsmokorset, Norway) over 160 spectral bands in the visible and near infrared (VNIR) regions, ranging from 415 nm to 993 nm. Spectral information corresponding to spectral bands > 902 nm were discarded because of the low signal-to-noise ratio. Images were orthorectified and geo-referenced using the PARGE software (http://www.rese.ch/products/parge/ accessed on 4 March 2021) with the canopy DSM (Digital Surface Model) produced from the LiDAR point cloud. Atmospheric corrections were performed using ATCOR-4 software [58] to produce top of canopy (TOC) reflectance with 1-m spatial resolution. Airborne acquisition over the full site and its surroundings resulted in 23 flight lines covering 14 Km².

2.2.2. Airborne Laser Scanning (ALS)

ALS data were acquired with a RIEGL LMS Q780 simultaneously with imaging spectroscopy data. The sensor operated with a frequency of 400 kHz and a scanning angle of ±30° which ensured an average pulse density of 33 pulses/m². A point cloud filtering was applied to remove points classified as Low point (noise) according to the ASPRS Standard LIDAR Point classification [59].
2.2.3. Red, Green and Blue (RGB) Imagery

Very high spatial resolution (VHSR) imagery was acquired during the same airborne campaign using an iXU 180 Phase One camera. This imagery includes red, green and blue channels and the final spatial resolution of the orthorectified mosaic was 10 cm.

2.3. Ground Information

2.3.1. Field Plot Network and Selected Sites for Simulation

The Paracou site hosts an experimental setup including multiple permanent plots contributing to long-term monitoring of the influence of human induced disturbances since 1986 [60]. Species inventories combined with accurate tree trunk location have been performed every 1–2 years for each plot for more than 35 years for the oldest plots.

We selected two subplots in the control undisturbed forest plots, dimensioned to fulfill the different aspects of our analysis:

- A plot of 80 \times 80 m, hereafter referred to as Site A was dedicated to the adjustment of statistical models for the estimation of vegetation biochemical properties;
- A plot of 300 \times 300 m, hereafter referred to as Site B, was dedicated to the comparison of the simulations with experimental data.

The extent of the airborne imaging spectroscopy acquisition and the location of the plot network as well as Sites A and B are presented in Figure 1.

![Figure 1. Red, green and blue (RGB) composite of airborne imaging spectroscopy acquired (R: 665 nm, G: 560 nm, B: 490 nm): (a) over the study site; (b) Site A; (c) Site B. Delineated individual tree crowns (ITCs) are represented by yellow polygons.](image-url)
2.3.2. Individual Tree Crown Delineation

Manual delineation of emerging crowns identified from the inventories was conducted, based on visual interpretation of the VHSR image, combined with a canopy height model derived from ALS. This delineation was validated during field surveys [61–63]. More than 2500 ITCs were delineated in the area covered by the imaging spectroscopy acquisition, but our analysis focused on ITCs used as support to the modeling activities performed on site B. We selected the nine most abundant species from these delineated crowns in site B, corresponding to 162 ITCs and 10,262 pixels. Table 1 provides basic statistics on size and distribution of ITCs per species in sites A and B.

Table 1. Statistical description of ITCs corresponding to the nine most abundant species delineated in site A and B.

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of ITCs</th>
<th>Number of Pixels</th>
<th>Mean Number of Pixels Per ITC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bocoa prouacensis</td>
<td>1</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>Dicorynia guianensis</td>
<td>1</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Eschweilera sagotiana</td>
<td>5</td>
<td>286</td>
<td>57</td>
</tr>
<tr>
<td>Licania alba</td>
<td>1</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Licania heteromorpha</td>
<td>2</td>
<td>44</td>
<td>22</td>
</tr>
<tr>
<td>Pradosia cochlearia</td>
<td>2</td>
<td>433</td>
<td>216</td>
</tr>
<tr>
<td>Sextonia rubra</td>
<td>2</td>
<td>223</td>
<td>111</td>
</tr>
<tr>
<td>Vouacapoua americana</td>
<td>6</td>
<td>387</td>
<td>64</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>20</strong></td>
<td><strong>1581</strong></td>
<td><strong>85</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of ITCs</th>
<th>Number of Pixels</th>
<th>Mean Number of Pixels Per ITC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dicorynia guianensis (DG)</td>
<td>11</td>
<td>1147</td>
<td>104</td>
</tr>
<tr>
<td>Eperua falcata (EF)</td>
<td>45</td>
<td>3102</td>
<td>68</td>
</tr>
<tr>
<td>Eperua grandiflora (EG)</td>
<td>8</td>
<td>694</td>
<td>86</td>
</tr>
<tr>
<td>Eschweilera sagotiana (ES)</td>
<td>30</td>
<td>1270</td>
<td>42</td>
</tr>
<tr>
<td>Licania heteromorpha (LH)</td>
<td>8</td>
<td>251</td>
<td>31</td>
</tr>
<tr>
<td>Moronobea coccinea (MC)</td>
<td>12</td>
<td>609</td>
<td>50</td>
</tr>
<tr>
<td>Pradosia cochlearia (PC)</td>
<td>14</td>
<td>1538</td>
<td>109</td>
</tr>
<tr>
<td>Recordoxylon speciosum (RS)</td>
<td>14</td>
<td>944</td>
<td>67</td>
</tr>
<tr>
<td>Tapura capitulifera (TC)</td>
<td>20</td>
<td>707</td>
<td>35</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>162</strong></td>
<td><strong>10262</strong></td>
<td><strong>66</strong></td>
</tr>
</tbody>
</table>

3. Methods
3.1. Overview of the Methodology

Our main objective was to define a simulation framework aiming at simulating realistic remote sensing data measured with optical sensors over tropical forests. To reach this goal, we focused on identifying the optimal method for the introduction of two major contributors to the heterogeneity measured with remote sensing, NPVF and LOP. Figure 2 provides the general workflow of our study. During the first step, the global parameterization of DART simulations was defined for all the factors remaining unchanged for all simulations, based on experimental data. This global parameterization included sensor characteristics, geometry of acquisition, topography definition as well as 3D forest reconstruction from ALS data. The second step aimed at building different scenarios for the integration of two factors into the simulations, NPVF and spatial variability in leaf optical properties. NPVF was based either on the introduction of woody elements (wood area density, WAD) into the 3D reconstruction, or on the addition of brown pigments ($C_{brown}$) into leaf constituents. The spatial variability in leaf optical properties was defined at different levels of organization (pixel, crown, species) based on inventoried and delineated tree crowns. The third step consisted in running the DART model in order to produce simulated imaging spectroscopy data for each scenario. Finally, the fourth step aimed at comparing simulations with airborne acquisitions at different levels: we compared
experimental airborne acquisitions with simulations based on pixel reflectance, spectral dissimilarity within and among species, and species discrimination performed with a supervised approach.

Figure 2. Flowchart of the approach proposed to test the influence of different scenarios for the integration of non-photosynthetic vegetation fraction (NPVf) and heterogeneity in leaf optical properties (LOP) when simulating imaging spectroscopy of complex canopies.

3.2. Presentation of Leaf and Canopy Models
3.2.1. Leaf Optical Properties Modeling with PROSPECT

We used the leaf model PROSPECT [64,65] to simulate leaf optical properties. PROSPECT is a leaf model simulating directional-hemispherical reflectance using biophysical and biochemical parameters. It is based on Allen’s generalized plate model [66]: a leaf is modeled as a pile of identical elementary layers characterized by an absorption coefficient and a refractive index defined over the optical domain from 400 nm to 2500 nm. Here, we used PROSPECT-D [64] including three main pigments that control the absorptive properties of fresh leaves in the visible (VIS) domain, i.e., chlorophylls a + b (C_{ab}), carotenoids (xanthophylls and carotenes) (C_{xc}), and anthocyanins (C_{anth}). Additional constituents accounted for by PROSPECT-D include brown pigments (C_{brown}), which is used to simulate senescent leaves, and contributes to leaf absorption in the VNIR, leaf mass per area (LMA) and equivalent water thickness (EWT), which are mainly contributing to leaf absorption in the short wave infrared domain. The scattering properties of the leaf are controlled by the leaf
The DART model [51,52] was used to simulate airborne imaging spectroscopy. DART simulates the radiative transfer in the 3D “Earth-Atmosphere” system from the visible to thermal infrared domain of the electromagnetic spectrum. It is used to study both natural and urban landscapes, and allows for the integration of atmospheric and topographic effects. The DART scenes were based on 3D representations made of individual volume elements (voxels), each of them characterized by a set of properties such as vegetation density, leaf angle distribution, optical properties corresponding to leaves, trunk and branches, soil and any relevant elements included in the scene. Simulations were performed with the package pytools4dart (https://pytools4dart.gitlab.io/pytools4dart/ accessed on 12 March 2021) a python API (Application Programming Interface) aiming at preparing, running and analyzing DART simulations, specially dedicated to the design of large and complex simulations.

DART was set in flux tracking mode to simulate TOC reflectance, with the sun and the atmosphere as the only radiation sources. Topography was defined with a digital surface model produced from ALS point clouds. A generic litter spectrum was used for the optical properties of the soil/background. Simulations were performed for 135 bands ranging from 415 nm to 902 nm, following instrumental characteristics of the airborne acquisition for sites A and B. Sensor height, solar zenith and azimuth angles as well as spatial resolution were set according to the geometry of acquisition during the airborne campaign. Simulations were performed using nadir viewing geometry for all pixels. Hence, no effects induced by the field of view of the sensor and the angles of rotation of the aircraft were accounted for in the simulated reflectance.

3.3. Defining Scenarios for the Integration of Leaf Optical Properties (LOP) Heterogeneity in 3D Modeling

Spectral variability in canopy reflectance among species, within species, and within individual trees has been documented [56] and multiple factors such as position in canopy or topography contribute to this variability. In our study, we specifically focused on the influence of spatial variations in LOP on this spectral variability: we defined three scenarios for the integration of LOP in radiative transfer simulations at pixel level (LOP_{pixel}), crown level (LOP_{ITC}), or species level (LOP_{species}). Then we compared simulated canopy reflectance with corresponding experimental acquisitions in order to describe the influence of LOP variability on spectral heterogeneity of canopy reflectance and assess the degree of realism of the simulation.

The definition of LOP at different scales could not be performed using field sampling for logistic reasons. Instead, we implemented a method to estimate a set of relevant chemical constituents from canopy reflectance: here we focused on leaf pigments, and estimated the combination of pigment content for each pixel (see Section 3.6 for details). Then we used the pigment content estimation at pixel level, or its average at crown or species level using the species inventory and the crown delineation, to simulate and assign LOP values in the simulation scene depending on the scenario.
3.4. Definition of 3D Structure of the Canopy

The 3D forest structure was derived from ALS acquisition: the ALS point cloud and associated trajectory file were voxelized using the AMAPvox v.1.7.3-BETA software (http://amap-dev.cirad.fr/projects/amapvox accessed on 3 April 2021). AMAPvox tracks every laser pulse through a 3D grid (voxelized space) from the top of the canopy to the last recorded hit. The effective sampling area of each laser pulse (or fraction of pulse in case of multiple hits) is computed from the theoretical beam section (a function of distance from laser and divergence of laser beam) and the remaining beam fraction entering a voxel. In case more than one hit is recorded for a given pulse then it is assumed that the surface intercepted by each hit (the “hit area”) is equal to the inverse of the total number of hits times the beam section. This information is combined with the optical path length of each pulse entering a voxel to compute the local transmittance per voxel. Different estimation procedures are provided in the AMAPVox software [67]. In the present study we used a numerical resolution to compute for each voxel the transmittance, or gap probability $P_{\text{gap}}$, from the equation below:

$$\sum_i n_i B_{\text{ent}} \times S_i \times P_{\text{gap}}^l = \sum_i n_i B_{\text{out}} \times S_i$$

with $l_i$ the optical potential path length of pulse $i$ (from entry point to the expected exit point).

$S_i$ the cross section of pulse $i$, computed at voxel center.

$B_{\text{ent}}$ and $B_{\text{out}}$ being the beam surface fraction entering and outgoing the voxel.

The equation above expresses that the estimated $P_{\text{gap}}$ per unit length of a voxel is the value that yields the light transmission rate (given the actual path length and effective surface of all contributing pulses) equal to the observed transmission rate.

The gap fraction probability per unit path length per view direction $\theta$ depends only on the foliage projection ratio and the plant area density ($P_{\text{AD}}$, $m^2 \cdot m^{-3}$).

$$P_{\text{gap}} = e^{-G(\theta) \times P_{\text{AD}}}$$

$G(\theta)$ projection function will depend on the leaf inclination distribution function (LIDF) which is also an important factor among canopy structure properties influencing canopy reflectance [48,49,68]. However available data did not allow us to include estimated LIDF in our simulations, and we used a unique LIDF corresponding to spherical distribution, the most widely used since it offers straightforward computation of the projection function, equal to 0.5 in any direction [68].

Here, we used a uniform $1 \times 1 \times 1$ m voxel size matching with the spatial resolution and pixel grid of the imaging spectroscopy acquisition. Some post-processing steps are required to adjust the raw 3D distribution of $P_{\text{AD}}$ obtained from AMAPvox [69–72] and ensure accurate description of horizontal and vertical structure. These adjustments were performed as follows:

• A hierarchical re-estimation of the transmittance based on a mixed linear model was performed in order to minimize the uncertainty associated with poor sampling of lower canopy voxels resulting from the gradual extinction of the laser beam (occlusion) [72].

• A reduction factor of 0.8 was applied to the re-estimated $P_{\text{AD}}$ in order to compensate for the limited penetration achieved by the LMSQ780 LiDAR when flying at 900 m. This factor was set to fit the extinction profile obtained over the same area while operating at lower altitude (450 m) (G. Vincent, in prep).

3.5. Integration of Non-Photosynthetic Vegetation Fraction (NPVf) into Simulations

Each voxel in the 3D scene is characterized by a specific $P_{\text{AD}}$. $P_{\text{AD}}$ is an important factor driving the interactions of electromagnetic radiation with vegetation [36,73], which combines the contribution of various elements including leaves of all development stages (from juvenile to senescent and dead leaves), branches and stems. The vertical integration
of the \(PAD\) corresponds to the Plant Area Index (\(PAI\)). \(PAI\) is defined as the sum of Leaf Area Index (\(LAI\)) and Wood Area Index (\(WAI\)) [74]. \(LAI\) is a key parameter to consider for physiological, ecological and climatological studies [75]. Old-growth forests are characterized by a non-negligible fraction of \(NPV_f\) corresponding to dead leaves and woody components like standing deadwood, twigs and branches distributed within tree crowns [76]. Zhu et al. [77] showed an overestimation of \(LAI\) estimated by terrestrial LiDAR of 3.0 to 31.9\%, caused by woody elements, with differences observed among forest types. Calders et al. [78] analyzed the relationship between effective \(PAI\) (\(ePAI\)), effective \(LAI\) (\(eLAI\)), and effective \(WAI\) (\(eWAI\)), which corrects for clumping effect, using the librat Monte Carlo ray tracing model. They obtained a linear relationship between \(eLAI\) and \(ePAI - eWAI\). However, no strict complementarity was obtained between these three area indices, as woody structure is preferentially obscured by the leaves. Accounting for the relative effect of these woody elements is particularly challenging when using radiative transfer modeling [79] since they are important photon-absorbing and -scattering components [36], hence both their spectral response and their 3D distribution in a forest scene should be considered in radiative transfer approaches.

In our study, we compared two approaches to take into account \(NPV_f\) in our simulations, based on different hypotheses: the first approach is based on the disaggregation of \(PAD\) into \(LAD\) and \(WAD\), while the second approach assumes equality between \(PAD\) and \(LAD\) and uses brown pigment content in leaves to account for the influence of \(NPV_f\) on canopy reflectance. The former will be referred to as \(NPV_f_{WAD}\), while the latter will be referred to as \(NPV_f_{Cbrown}\). We also compared these two scenarios accounting for \(NPV_f\) with the case where \(PAD\) is exclusively defined as leaf materials without brown pigments, hereafter referred to as \(NPV_f_0\).

### 3.5.1. Estimation of \(NPV_f_{WAD}\)

The definition of \(WAI\) from the voxelized mock-up is based on two hypotheses. First, \(PAD\) is assumed to be the sum of \(WAD\) and \(LAD\): although literature showed that this hypothesis is incorrect when accounting for clumping [74,80], we used 3D forest representation based on individual voxels characterized by turbid medium with no clumping. Second, the contribution of \(WAD\) was assumed to be a fixed proportion of \(PAD\). This assumption is not validated by experimental observations, but this simplification is imposed by the impossibility to accurately differentiate wood and leaves from ALS data.

In order to adjust the proportion of \(WAD\) to be applied to all voxels, we first identified a set of \(ITC\) corresponding to defoliated and foliated individuals based on visual analysis of VHSR imagery. This visual identification resulted in 10 \(ITCs\) unambiguously identified as defoliated, and 20 \(ITCs\) unambiguously identified as fully foliated and non-senescent. The visual identification was not assisted by inventory data, so our selection included a random selection of species. Then, we extracted the \(PAD\) corresponding to the three uppermost voxels from the 3D mockup, and we computed the mean \(PAD\) from foliated and defoliated voxels. We defined the reference \(NPV_f_{WAD}\) as the ratio between the mean \(PAD\) from defoliated voxels and the mean \(PAD\) from foliated voxels, assuming that on average the \(WAD\) of the defoliated crowns is similar to the \(WAD\) of foliated crowns. In order to account for woody structure being preferentially obscured by the leaves, we also tested an intermediate scenario in which \(NPV_f\) is half of \(NPV_f_{WAD}\). This intermediate scenario will be referred to as \(NPV_f_{WAD/2}\).
3.5.2. Estimation of Leaf Brown Pigment Content

In the second approach, we assumed equality between \( P_{\text{AD}} \) and \( L_{\text{AD}} \), and used a spectral index in order to estimate the brown pigment content \( (C_{\text{brown}}) \) corresponding to each pixel. This approach is based on the hypothesis that the radiometric contribution of non-photosynthetic elements in a voxel can be compared to the radiometric contribution of \( C_{\text{brown}} \), which absorbs in the VNIR domain. Brown pigments are characterized by a broad absorption domain, with maximum absorption in lower wavelengths of the VIS domain, gently decreasing until 1100 nm [81]. The introduction of brown pigments results in smoothing of the green reflectance peak, and more remarkably a less abrupt increase in reflectance in the red-edge. The influence of \( C_{\text{brown}} \) in the red-edge domain can be used to define a spectral index linking canopy reflectance to leaf brown pigment content. The definition of this spectral index and the procedure used to adjust the regression model linking canopy reflectance to leaf pigment content is described in the next section.

3.6. Simulation of Leaf Optical Properties through the Estimation of Leaf Pigment Content

Leaf pigment content controls \( \text{LOP} \) and canopy reflectance in the VIS region of the electromagnetic spectrum as well as in the red-edge domain and first part of the NIR domain [36]. Here, our first goal was to define a strategy to assign pigment content to each pixel, crown or species in our simulations, based on canopy reflectance. A variety of methods can be used to assess foliar pigment content from remotely sensed canopy reflectance acquired with imaging spectroscopy. Verrelst et al. [82] categorized these methods into four categories: (i) parametric regression, including vegetation indices, shape indices and spectral transformations, and assuming a simple relationship between spectral data and vegetation biophysical variable, such as a linear monovariate relationship; (ii) non-parametric regression, based on data-driven techniques that directly link spectral data and biophysical variables (e.g., linear and non-linear machine learning regression algorithms); (iii) inversion of RTM using numerical optimization and look-up table approaches; and (iv) hybrid methods, which combine RTM for the simulation of a training dataset with parametric or non-parametric regression methods.

Parametric regressions based on vegetation indices and linear or polynomial models are the most straightforward means of estimating pigment content [83–85]. However, the adjustment of such empirical models requires a large amount of field data. Hybrid approaches provide a means to overcome this dependency on field data: they are based on the simulation of a training dataset (corresponding to \( \text{LOP} \) or canopy reflectance) using appropriate physical models and sampling strategy, in order to adjust a regression model applicable to experimental data. Therefore, we chose a hybrid approach in this study, in order to train regression models linking spectral indices with leaf pigment contents, including \( C_{\text{ab}}, C_{\text{xc}}, C_{\text{anth}} \) and \( C_{\text{brown}} \). This approach is based on two steps: First, the regression models were adjusted based on a simulation including a broad range of pigment content. Here, we adjusted a linear model specific to each pigment and each \( NPVf \) scenario, based on simulations performed on Site A. Then the linear models were applied on experimental data corresponding to Site B, in order to define specific pigment content for each \( NPVf \) scenario. Figure 3 illustrates the procedure used to adjust regression models for the estimation of leaf pigment content from spectral indices, and to estimate pigment content from airborne imaging spectroscopy on site B based on these regression models.
3.6.1. Adjustment of Models for the Estimation of Leaf Pigments (Site A)

We used three spectral indices to estimate \( C_{ab} \), \( C_{xc} \) and \( C_{anth} \) from remotely sensed data: the Chlorophyll Index (CI) (Equation (3)), the Carotenoids Reflectance Index (CRI) (Equation (4)) and the Anthocyanin Reflectance Index (ARI) (Equation (5)) [86]. These spectral indices were applied on imaging spectroscopy for various ecosystems including forest [28,29,87]. Due to the lack of documented spectral index for \( C_{brown} \) in the literature, we defined a Brown Pigment Reflectance Index (BPI) (Equation (6)) based on reflectance measured in the red-edge and NIR domains. The spectral band selected in the red-edge was centered on 750 nm, which is characterized by negligible absorption of \( C_{ab} \) compared to \( C_{brown} \).

\[
CI = \frac{R_{760-800}}{R_{690-725}} - 1,
\]

\[
CRI = \left( \frac{1}{R_{510-520}} - \frac{1}{R_{690-710}} \right) \times R_{760-800},
\]

\[
ARI = \left( \frac{1}{R_{530-570}} - \frac{1}{R_{690-710}} \right) \times R_{760-800},
\]

\[
BPI = \frac{R_{850} - R_{750}}{R_{850} + R_{750}},
\]
where \( R_{\lambda_1-\lambda_2} \) corresponds to the mean reflectance in the spectral domain defined by \( \lambda_1 \) and \( \lambda_2 \).

DART was used to simulate the hyperspectral acquisitions on site A coupled with PROSPECT-D to simulate LOP. We followed the global parameterization setting for instrumental configuration, the structure of the 3D mockup was parameterized as described in Section 3.4. A sampling strategy based on experimental distribution and co-distribution of leaf chemistry \cite{88} was designed in order to define a combination of leaf pigments to be associated with each pixel, i.e., vertical column of voxels, of the 3D mockup. We then generated 6400 combinations of \( C_{ab}, C_{xc}, C_{anth} \) and \( C_{brown} \) following a normal multivariate distribution for \( C_{ab} \) and \( C_{xc} \), a log normal distribution for \( C_{anth} \), and a normal distribution for \( C_{brown} \). Each of these 6400 combinations was assigned to a pixel from the simulation corresponding to site A.

The other input parameters of PROSPECT-D, \( N \), \( EWT \), and \( LMA \), were set to a constant value. The leaf structure parameter \( N \) influences leaf scattering properties and the ratio between reflectance and transmittance, but does not influence leaf absorption. Hence, the \( N \) parameter was set to 1.8, assuming that the vegetation scattering induced by the detailed canopy structure would provide sufficient variability. The influence of \( EWT \) and \( LMA \) on canopy reflectance is limited in the VNIR spectral domain, thus they were set to 12.9 mg cm\(^{-2}\) and 12 mg cm\(^{-2}\) respectively, which correspond to the average values reported by Féret et al. \cite{88} for a leaf dataset collected at the Paracou research station.

The linear regression models between leaf pigment content and the spectral indices (computed from simulated canopy reflectances) were trained using sunlit pixels only, i.e., pixels with more than 20% reflectance in the NIR spectral band centered on 800 nm, in order to reduce confusion induced by low illumination in further comparisons \cite{23}.

### 3.6.2. Application of Models for the Estimation of Leaf Pigments (Site B)

The regression models adjusted on site A were applied to the pixel of the airborne hyperspectral observations of Site B in order to estimate pigment content site B 3D mockup. Same as for site A, all voxels in the same vertical columns shared the same pigment content. The pixel-scale pigment content was then aggregated at crown scale following delineated ITCs and at species scale using the taxonomic identification of these ITCs, using the average value of sunlit pixels. A generic spectrum simulated using a combination of canopy pigment was used to characterize pixels of non-delineated ITCs in species and ITC scenarios. Finally, for each level of LOP heterogeneity (pixel, crown, species), the four \( NPVf \) scenarios (\( NPVf_0 \), \( NPVf_WAD \), \( NPVf_WAD/2 \), \( NPVf_Cbrown \)) were produced, resulting in 12 simulations of imaging spectroscopy data corresponding to Site B.

### 3.7. Comparing Simulations with Airborne Acquisitions

We compared airborne acquisitions with simulated imaging spectroscopy data based on different levels of information. They included radiometric comparison between measured and estimated canopy reflectance, as well as comparison of crown/species scale spectral variability between measured and estimated canopy reflectance, species discrimination using supervised classification algorithms and spectral dissimilarity within ITCs, between ITCs of the same species and between ITCs of different species using spectral angle metrics.

#### 3.7.1. Radiometric Comparison

Different statistical criteria were used to compare the observed (\( R_{\text{SENSOR},\lambda} \)) and simulated TOC reflectance (\( R_{\text{DART},\lambda} \)) in the VNIR spectral domain. These included the root mean square error (RMSE) (Equation (7)) and the coefficient of determination \( (R^2) \) (Equation (8)), computed as follows for each spectral band of the sensor:

\[
RMSE(\lambda) = \sqrt{\frac{\sum_{i=1}^{n}(R_{\text{DART},\lambda}(i) - R_{\text{SENSOR},\lambda}(i))^2}{n}},
\]
Remote Sens. 2021, 13, 2120

3.7.2. Analysis of Spectral Dissimilarity

Tropical tree species discrimination using supervised methods remains challenging, and assuming a direct link with within- and among-species variability of canopy reflectance is a simplification [89]. A number of statistical metrics of variability such as the coefficient of variation, the spectral angle or the minimum distance from the spectral centroid can be used to relate spectral diversity to plant diversity [22]. In this study, the spectral angle [90], defined in (Equation (9)), was used to compare spectral dissimilarity between experimental data and simulated data. The spectral angle was computed on sunlit pixels at three level using the 135 spectral bands: (i) within crown for all pairwise combinations of pixels of a given crown, (ii) within species for all pairwise combinations of crowns of a given species, and (iii) among species, for all pairwise combinations of species. The comparison between the dissimilarity matrices of observations and simulations was done at each level with Pearson’s correlation coefficient and RMSE.

\[
R^2(\lambda) = 1 - \frac{\sum_{i=1}^{n}(R_{\text{DART}}(i) - R_{\text{SENSOR}}(i))^2}{\sum_{i=1}^{n}(R_{\text{SENSOR}}(i) - R_{\text{SENSOR}})^2}, \quad (8)
\]

where \(S_i\) and \(S_j\) are the spectra of elements \(i\) and \(j\), and \(\lambda_a\) and \(\lambda_b\) are the wavelength boundaries i.e., 415 to 902 nm in this study.

3.7.3. Species Discrimination

A large number of publications evidenced the interest of imaging spectroscopy data for the discrimination among species, even for tropical forests showing strong diversity [24,62,89,91–94]. On a higher ecological level, this capacity to discriminate among species based on spectral information is also a fundamental property for methods aiming at estimating diversity indices based on the heterogeneity of spectral information [28,29,34]. In this study, we used the performances of species discrimination based on supervised classification as an indicator of the realism of our simulated imaging spectroscopy data. Species discrimination applied in real situations inherently leads to successful classification for species characterized by high discriminability, and confusion between species showing lower discriminability. This capacity to accurately identify a species or misidentify it as another species is explained by the expression of vegetation traits. In our study, we applied supervised classification for species identification on both experimental and simulated imaging spectroscopy data, hypothesizing that realistic simulations would lead to similarity in the confusion matrix derived from classification when compared to the confusion matrix derived from experimental data. We used linear discriminant analysis (LDA) as a supervised classification algorithm in order to perform species classification [62,89,92].

LDA is a parametric classification algorithm based on the hypothesis of multivariate normality of input variables, and aims to maximize the between-class variance and minimize the within-class variance, through a linear discriminant function. In our study, we performed classification on ITCs corresponding to the nine most abundant species detailed in Table 1. For each species the ITCs were randomly distributed between two folds of equal size: one for training and the other for testing. The split is made at the crown level instead of the pixel level in order to avoid using pixels from the same crown in both training and test datasets. Sampling pixels from the same crown in both training and test dataset may lead to overestimated scores due to spatial correlation among neighboring pixels and possibly lower variability within crown than among crowns. This resulted in 4 to 23 ITCs/species used to train the LDA model, and 4 to 22 ITCs/species used to test the model. The procedure was applied on imaging spectroscopy data, with training and test datasets built from the same pixel selection for experimental and simulated data, and repeated
100 times in order to account for the influence of the random selection. The Pearson’s correlation coefficient was used to compare the average confusion matrices of tests on observations against simulations. Two correlation analyses were performed, one to compare the success of classification using the diagonal elements of the confusion matrix, and another one comparing confusion rate between species and using off-diagonal elements.

4. Results

4.1. Definition of the 3D Structure of the Forest

The PAI was computed for each pixel in site B, first from the initial 3D mock-up produced with AMAPVox, after application of the hierarchical re-estimation of transmittance, and finally after application of the attenuation factor. The mean PAI following was 9.71 m²/m² for Site A and 9.4 m²/m² for Site B.

4.2. Estimation of NPVf

4.2.1. NPVf Derived from Wood Area Density (WAD)

The PAD of three uppermost voxels of the selection of foliated and defoliated ITCs was averaged in order to analyze their distribution. The mean PAD corresponding to defoliated and foliated tree crowns was 0.33 m²·m⁻³ and 1.54 m²·m⁻³, respectively (Figure 4). NPVf/WAD, defined as the ratio of these two mean densities, was then set to 21% in simulations. Other studies reported NPVf ranging between 10% and 50% depending on forest type and stage [55,95,96]. In order to account for potential shadowing of woody elements by leaves [62], we also tested a scenario corresponding to 50% of NPVf/WAD hereafter referred to as NPVf/WAD2. For each of these scenarios, the optical properties corresponding to generic wood optical properties were assigned to a constant proportion of PAD defined as NPVf/WAD or NPVf/WAD2. The complementary fraction of PAD corresponding to LAD was characterized by LOP with pigment content defined as described in Section 4.3.

![Figure 4. Distribution of mean plant area density (PAD) of the three uppermost voxels corresponding to foliated and defoliated tree crowns delineated from VHSR data interpretation.](image)

4.2.2. NPVf Derived from Leaf Brown Pigment Content

A linear regression model was adjusted in order to describe the relationship between BPI and CBrown based on DART simulation over Site A. The Pearson’s correlation coefficient (r) corresponding to this linear relationship was 0.67 (Figure 5a). The linear model was then applied to sunlit pixels from Site B in order to account for non-photosynthetic vegetation in the scenario considering the PAD as equivalent to LAD. The value of CBrown ranged between 0 and 0.6 arbitrary units (Figure 5b). The scenario accounting for NPVf through CBrown is referred to as NPVf/Cbrown hereafter.
4.3. Influence of NPVf on the Pigment Content Estimate

4.3.1. Impact on the Adjustment of the Relationship between Spectral Indices and Pigment Content

Once DART simulations were performed on Site A for the different scenarios accounting for NPVf, the relationships between pigments content and their corresponding spectral indices were studied (Figure 6). Overall, a linear relationship with a positive correlation was observed between pigment content and corresponding spectral index, for all pigments and all scenarios accounting for NPVf. The relationship between \( C_{\text{ab}} \) and CI showed the strongest correlation, with Pearson correlation coefficient ranging between 0.89 and 0.93. However, the correspondence between \( C_{\text{ab}} \) and CI strongly varied depending on the scenario used to account for NPVf: \( \text{NPVf}_{\text{WAD}} \) and \( \text{NPVf}_{\text{WAD/2}} \) resulted in a dramatic decrease in the range of CI (Figure 6, X-axis), which is explained by the decreased effect of chlorophyll absorption on the signal when mixing photosynthetic elements with non-photosynthetic elements.

The relationship between \( C_{\text{xy}} \) and CRI showed lower correlation, with Pearson’s correlation coefficient ranging between 0.55 and 0.65, with maximum correlation obtained for \( \text{NPVf}_0 \), and comparable correlation for the three other scenarios. As for CI, the range of variation observed for CRI decreased with \( \text{NPVf}_{\text{WAD}} \) and \( \text{NPVf}_{\text{WAD/2}} \) due to reduced contribution of photosynthetic vegetation.

Finally, the relationship between \( C_{\text{anth}} \) and ARI showed relatively different behavior: The Pearson’s correlation coefficient was similar for \( \text{NPVf}_0 \) and \( \text{NPVf}_{\text{Cbrown}} \) (Pearson’s \( r = 0.83-0.82 \)), while it increased to 0.91–0.90 when including wood optical properties in the voxels, along with a reduced range of variation of ARI, as for the two other indices.

While all spectral indices appeared to be sensitive to the pigment content they were initially designed to be used with, the linear models adjusted based on DART simulations are expected to result in very different pigment content when applied to experimental data, depending on the method used to account for NPVf.
4.3.2. Implication of the Application of the Models on Experimental Data

The regression models adjusted based on DART simulations performed on Site A (Figure 7) were applied to sunlit pixels from Site B in order to estimate pigment content for each NPVf scenario. The distribution obtained for $C_{ab}$ and $C_{xc}$ was very similar for NPVf$_0$ and NPVf$_{brown}$. $C_{ab}$ ranged from 2.9 to 82.7 µg·cm$^{-2}$ with a mean of 35.2 ± 10.1 µg·cm$^{-2}$ for NPVf$_0$ and from 0.9 to 93.7 µg·cm$^{-2}$ with a mean value of 38.5 ± 11.7 µg·cm$^{-2}$ for NPVf$_{brown}$. The two scenarios NPVf$_{WAD}$ and NPVf$_{WAD/2}$ resulted in increase in $C_{ab}$ ranging from 1.6 to 105.4 µg·cm$^{-2}$ with a mean of 43.7 ± 13.1 µg·cm$^{-2}$, and from 1.1 to 133.6 µg·cm$^{-2}$ with a mean value of 54.8 ± 16.7 µg·cm$^{-2}$, respectively.

The influence of the different NPVf scenarios on the estimation of $C_{xc}$ was comparable to the results obtained with $C_{ab}$. $C_{xc}$ ranged from 1.5 to 17.9 µg·cm$^{-2}$ with a mean value of 7 ± 2.4 µg·cm$^{-2}$ for NPVf$_0$ and from 1.7 to 23.7 µg·cm$^{-2}$ with a mean value of 9.1 ± 3.2 µg·cm$^{-2}$ for NPVf$_{brown}$. The distribution of $C_{xc}$ strongly increased for NPVf$_{WAD}$ and NPVf$_{WAD/2}$ scenarios.

The influence of the different NPVf scenarios on the estimation of $C_{anth}$ differed from the influence observed for photosynthetic pigments. $C_{anth}$ distributions including negative values were obtained for all scenarios with various occurrences, suggesting discrepancies between simulated and measured reflectance for the spectral domain used to compute $ARI$. These negative values were set to zero when introduced into DART simulations.
Figure 7. Pigment distribution computed for Site B, based on linear models relating spectral indices to pigment content, adjusted from the four NPVf scenarios applied on Site A. (a) C_{ab}; (b) C_{xc} (c) C_{anth}.

4.4. Influence of NPVf on Simulated Reflectance

Imaging spectroscopy and DART simulations were compared using spectral $R^2$ and the RMSE for each NPVf scenario (Figure 8). We assumed that higher $R^2$ and lower RMSE would indicate higher radiometric realism. Overall, NPVf_{Cbrown} and NPVf_0 scenarios resulted in similar $R^2$ over all the spectral domain, with stronger correlations than those obtained for NPVf_{WAD} and NPVf_{WAD/2} over the full spectral domain. The scenario corresponding to NPVf_{WAD} resulted in lowest $R^2$ overall. The spectral profile of $R^2$ was similar among NPVf scenarios: lower correlations were obtained in the far-blue (<450 nm), green (520 nm), red-edge and NIR domains, while higher correlations were obtained between blue and green domains (500 nm), and in the yellow-red domains (570–680 nm).

Figure 8. Spectral variation of (a) $R^2$ and (b) root mean square error (RMSE) for the different scenarios accounting for NPVf, and accounting for pixel-specific leaf optical properties.

The choice of NPVf scenario resulted in strong differences in RMSE between measured and simulated reflectance in the red-edge/NIR domains. The NPVf_{brown} scenario resulted
in lower RMSE compared to other scenarios in the full spectral domain, except in the red domain.

Spectral bands located in the blue, green, red, red edge and NIR spectral domains were selected in order to analyze the influence of NPVf on the simulated reflectance (Figure 9). The influence of NPVf on reflectance was particularly strong at 491 nm (blue) and 666 nm (red), corresponding to domains of absorption of photosynthetic vegetation. In these regions, an underestimation of reflectance was observed for NPVf0. The integration of NPVf using WAD led to an increase of simulated reflectance up to an overestimation for NPVfWAD. For NPVfCbrown, reflectance was underestimated at 491 nm and 666 nm. In the green domain (560 nm), the effect of NPVf was moderate, except for NPVfCbrown which removes the positive bias observed with the other scenarios. An overestimation of reflectance was observed for NPVf0 at 720 nm (red edge) compared to scenarios where NPVf were considered. Simulated reflectance was overestimated at 800 nm overall but no significant effect of NPVf was detected. Given these results, we decided to only use the NPVfCbrown in the following analyses.

Figure 9. Relationship between Imaging spectroscopy and DART simulations top of canopy (TOC) reflectance. (a) 491 nm; (b) 560 nm; (c) 666 nm; (d) 702 nm and (e) 800 nm. From top to bottom: NPVf0; NPVfWAD/2; NPVfWAD and NPVfCbrown.

4.5. Influence of LOP Variability on Simulated Reflectance

Spectral \( R^2 \) and RMSE were used to compare reflectance from airborne acquisitions and DART simulations for each spectral band and each LOP level, when selecting NPVfCbrown. LOPpixel outperformed LOPITC and LOPspecies in the visible region based on \( R^2 \) and in the NIR based on the RMSE (Figure 10). Relatively similar results were obtained for other NPVf scenarios (results not shown).
4.6. Analysis of Spectral Dissimilarity

The spectral dissimilarity was computed within crowns, among crowns corresponding to the same species and among crowns corresponding to different species using the spectral angle. DART simulations were compared to imaging spectroscopy to test if the spectral dissimilarity observed in imaging spectroscopy data was correctly simulated for the different LOP scenarios (Figure 11 and Table 2). When defining LOP per species in DART (LOPspecies), low agreement was obtained between spectral dissimilarity computed from imaging spectroscopy and from simulations for all levels (within crowns, within species and among species), with systematic underestimation. Intermediate performances were obtained when defining LOP per ITC in DART simulations. LOPITC scenario showed a systematic underestimation of spectral dissimilarity within crowns and a good agreement between spectral dissimilarities within species and among species. The definition of LOP per pixel in DART simulations outperformed other scenarios, with strong agreement of spectral dissimilarity when comparing experimental data with simulations for all levels.

Figure 10. Spectral variation of (a) $R^2$ and (b) RMSE derived from NPVfCbrown for the different scenarios accounting for LOP variability.

Figure 11. Relationship between spectral dissimilarity computed for experimental imaging spectroscopy and DART simulations based on spectral angle. (a) LOPspecies level; (b) LOPITC level; (c) LOPpixel level.
Table 2. Pearson’s correlation coefficient and RMSE computed when comparing spectral dissimilarity from imaging spectroscopy and from DART simulations for the different LOP variability scenarios. Spectral dissimilarity is compared within crowns, within species and among species.

<table>
<thead>
<tr>
<th>Spectral Dissimilarity</th>
<th>LOP_{species}</th>
<th>LOP_{ITC}</th>
<th>LOP_{pixel}</th>
<th>LOP_{species}</th>
<th>LOP_{ITC}</th>
<th>LOP_{pixel}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Crowns</td>
<td>0.10</td>
<td>0.34</td>
<td>0.88</td>
<td>0.84</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>Within Species</td>
<td>0.28</td>
<td>0.91</td>
<td>0.91</td>
<td>1.84</td>
<td>0.69</td>
<td>0.54</td>
</tr>
<tr>
<td>Among Species</td>
<td>0.48</td>
<td>0.93</td>
<td>0.92</td>
<td>1.75</td>
<td>0.81</td>
<td>0.67</td>
</tr>
</tbody>
</table>

4.7. Assessing the Potential of Species Discrimination

Performances for species classification based on LDA were compared when applied on airborne imaging spectroscopy and DART simulations (Figure 12 and Table 3). The simulation corresponding to LOP_{species} resulted in an overestimation of the potential of species discrimination compared to airborne imaging spectroscopy, with close to 100% overall classification accuracy. The confusion matrix derived from classification applied on LOP_{ITC} and LOP_{pixel} Scenarios showed stronger consistency with the confusion matrix derived from classification applied on experimental. Overall, LOP_{pixel} outperformed LOP_{ITC}, with particular ability to reproduce confusion among species.

Figure 12. Confusion matrix obtained from linear discriminant analysis (LDA) performed for the classification of the nine most abundant species reported in in situ data for Site B. (a) imaging spectroscopy; (b) LOP_{species}; (c) LOP_{ITC}; (d) LOP_{pixel}.
Table 3. Pearson’s correlation coefficient comparing LDA classification of DART simulations to imaging spectroscopy for LOP level scenarios. The consistency in correct classification and misclassification is compared between classification results obtained from airborne acquisition and classification results obtained from simulations.

<table>
<thead>
<tr>
<th>Confusion Matrix Elements</th>
<th>LOP\textsubscript{species}</th>
<th>LOP\textsubscript{ITC}</th>
<th>LOP\textsubscript{pixel}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct classification</td>
<td>0.05</td>
<td>0.86</td>
<td>0.92</td>
</tr>
<tr>
<td>Misclassification</td>
<td>−0.04</td>
<td>0.71</td>
<td>0.79</td>
</tr>
</tbody>
</table>

4.8. Visual Comparison

Figure 13 allows for visual comparison between imaging spectroscopy and the DART simulation of site B with the LOP\textsubscript{pixel}/NPV\textsubscript{fcbrown} scenario. Similar patterns and object dimensions were observed after stretching the nadir true color composites of DART simulation (Figure 13b) and imaging spectroscopy (Figure 13e). The nadir true color composites (Figure 13a) and imaging spectroscopy (Figure 13d) highlight the need for spectral characterization of defoliated crowns. This aspect was not considered in this study. In addition, the comparison in Figure 13f highlights the need for accurate characterization of biophysical and structural parameters which particularly affect the NIR region. This, in addition to the improvement of pigment contents estimates, will potentially increase radiometric agreement between simulations and imaging spectroscopy.

Figure 13. RGB composite Color. Top: imaging spectroscopy: (a) True color (R: 640 nm, G: 549 nm, B: 458 nm); (b) True color stretched (R: 640 nm, G: 549 nm, B: 458 nm); (c) False color stretched (R: 800 nm, G: 640 nm, B: 549 nm). Bottom: NPV\textsubscript{fcbrown}: (d) True color (R: 640 nm, G: 549 nm, B: 458 nm); (e) True color stretched (R: 640 nm, G: 549 nm, B: 458 nm); (f) False color stretched (R: 800 nm, G: 640 nm, B: 549 nm).
5. Discussion

5.1. Three-Dimensional (3D) Forest Reconstruction

The 3D forest structure was reconstructed by discretizing ALS data in 3D space using AMAPVox software. By applying Beer-Lambert’s turbid medium approximation, local transmittance enables estimation of \( \text{P}_{\text{AD}} \) and \( \text{P}_{\text{AI}} \). Infravoxel clumping may have caused an underestimation of LAI and \( \text{P}_{\text{AI}} \) \([77,97]\). Qu et al. \([98]\) estimated the \( \text{P}_{\text{AI}} \) of a moist tropical forest at Fazenda Cauaxi in the Paragominas Municipality of Pará State, Brazil. They reported a mean \( \text{P}_{\text{AI}} \) of 6.94 m\(^2\)/m\(^2\). Another study was conducted by Pimmasarn et al. \([69]\) in Khao Yai National Park in central Thailand. They obtained a mean \( \text{P}_{\text{AI}} \) of 8.02 m\(^2\)/m\(^2\). These results are slightly lower than the \( \text{P}_{\text{AI}} \) reported here (mean \( \text{P}_{\text{AI}} \) of 9.4 m\(^2\)/m\(^2\)). By contrast, Vincent et al. \([72]\) mapped \( \text{P}_{\text{AI}} \) at the same site (Paracou experimental station) by ALS (using a different laser system, i.e., a Riegl LMSQ560) and reported a mean \( \text{P}_{\text{AI}} \) of 13.2 m\(^2\)/m\(^2\). In that study the authors also compared simulated and measured light levels below canopy which revealed a slight underestimation of the ALS estimated transmittance \([72]\). Moreover, it is worth noting that the formula used to compute voxel transmittance from ray tracing in Vincent et al. \([72]\) (Equation (7)) can be shown to be biased by the fact that it considered an average optical path length rather than the actual path lengths as was done here. The bias might be responsible for the underestimation in transmittance (and overestimation of \( \text{P}_{\text{AI}} \)) observed at that time. This may indicate that the present estimate is more accurate than the one previously published for the site.

An independent measurement of \( \text{P}_{\text{AI}} \) on site A (80 \( \times \) 80 m) was conducted using high density TLS in 2016 (G. Vincent and F. Heuschmidt (2018) unpublished). The plot was voxelized at 50 cm resolution using AMAPvox. The estimated plot level \( \text{P}_{\text{AI}} \) was found to be equal to 11.4 m\(^2\)/m\(^2\).

Finally, it should be noted that the uncertainty in \( \text{P}_{\text{AI}} \) which to a large extent is related to the limited sampling density in the lower canopy due to occlusion. The upper canopy is much better sampled and the \( \text{P}_{\text{AD}} \) of the top canopy layers which indeed are of prime interest here are probably more accurately estimated than \( \text{P}_{\text{AI}} \) (\( \text{P}_{\text{AD}} \) integrated over the entire profile).

5.2. Influence of NPVf on Leaf Chemistry Estimate

A set of linear models was adjusted to estimate pigment contents from spectral indices, for each of the scenarios accounting for NPVf. The relationship between leaf pigment content and spectral indices varied depending on the scenario, due to the contribution of woody elements or brown pigments to the canopy reflectance. Therefore, these statistical models aiming at estimating leaf pigment content integrated different mixtures of materials assumed to be found in experimental data (wood, senescent leaves).

Despite differences in these relationships, a strong correlation was obtained between \( C_{\text{ab}} \) and the Chlorophyll Index, and between \( C_{\text{anth}} \) and the Anthocyanin Reflectance Index, for all scenarios accounting for NPVf in Site A. We found a moderate correlation between carotenoid content and the Carotenoid Reflectance Index. The main influence of NPVf was on the range of spectral indices, thus revealing different distributions of pigment content expected when applying Site A regressions to Site B. The poor suitability of some of these models (such as the models leading to negative anthocyanin content) highlights weaknesses in the integration of some of the factors influencing canopy reflectance into simulations.

Further work would be required in order to perform direct validation of these models for the estimation of leaf pigments based on imaging spectroscopy. However, it was expected that such a straightforward approach based on spectral indices would lead to inaccurate estimation when applied to such complex ecosystems. The main purpose of our approach was to provide a basis for the integration of relative variations in pigment content on our study site, and more refined multivariate approaches may be needed to minimize absolute bias in simulated values, that we could not control or correct here.
We compared the estimated pigment contents following NPVf scenarios with those reported in the literature in the context of tropical forests in order to determine their realism for our purposes. Arellano et al. [99] reported $C_{ab}$ ranging from 19.2 to 105.4 µg cm$^{-2}$ and $C_{xc}$ ranging from 1.0 to 44.2 µg cm$^{-2}$ in an Amazon rainforest of Ecuador. The ranges obtained for the different simulations then remained in a realistic range for $C_{ab}$ and $C_{xc}$. $C_{anth}$ showed discrepancies in some cases, with negative values. Ranges for $C_{anth}$ reported in the literature can reach higher values than those derived from our hybrid models [64,86], but such high values correspond to leaves which are unlikely to be found as main contributors to canopy reflectance in tropical forests. In tropical forests, anthocyanins mainly accumulate in juvenile leaves during the expansion phase and in vegetation found in the shaded parts of the forest. Further investigations on the definition of optimal NPVf could be relevant for improving the matching between estimated and expected pigment contents in tropical forests.

Despite the ubiquity of spectral indices in remote sensing, there are some issues when upscaling from leaf to canopy level since other factors such as canopy architecture, solar geometry, background properties and LAI also affect the relationship between pigment content and vegetation index [100]. Intensive sampling in situ combined with the partial least square (PLS) method, which has proved to be a successful empirical approach to retrieve leaf traits from canopy spectral data [101–103], could be relevant for local calibration and improvement of pigment content estimation.

5.3. Influence of NPVf on Simulation of Canopy Reflectance

The consideration of woody elements within forest architecture is one of the main issues in forest RTM [49]. In this respect, all DART NPVf scenarios were compared to imaging spectroscopy in order to investigate the influence of NPVf on simulation of canopy reflectance. Particularly important influence of NPVf was observed in the blue and red parts of the VIS domain, corresponding to the main domains of absorption of photosynthetic pigments with a major contribution of $C_{ab}$. Pigments are major contributors of pixel’s reflectance in the visible region [37,104]. Our results showed that the introduction of NPVf from WAD (NPVf$_{WAD}$ and NPVf$_{WAD/2}$) resulted in an increase of reflectance in the blue and the red domains (Figure 8). This observation is due to the mixture effect of green vegetation (strong absorption in the blue and red domains) with non-photosynthetic vegetation (lower absorption in the visible) which leads to an increase of reflectance. These findings were in agreement with Malenovský et al. [54] who evidenced an increase of 2% of canopy reflectance in the red domain when woody elements were integrated into simulations. The introduction of NPVf based on brown pigments only resulted in the addition of absorbing constituents to the leaves in the simulation, and no mixture with elements of higher reflectance. Consistent with previous observations [54,105], the introduction of NPVf led to a decrease of simulated reflectance in the red edge and NIR, either using WAD or $C_{brown}$. Overall, we found lower agreement between imaging spectroscopy and DART simulations in the NIR.

The reflectance in the red region was underestimated for NPVf$0$ and our optimal scenario (NPVf$_{Cbrown}$) compared to their counterpart in the imaging spectroscopy. Defining NPVf as a fraction of PAD (NPVf$_{WAD}$) needs further attention. Combining the integration of woody elements through WAD and senescent materials through brown pigments in simulations may further improve the agreement between imaging spectroscopy and DART simulations.

Two assumptions made when considering NPVf as a uniform fraction of PAD should be discussed. First, this leaf-to-wood area ratio is known to be spatially heterogeneous (across species, with canopy depth, etc.). Second, the turbid medium assumption is expected to lead to discrepancies when introducing woody elements which are spatially hierarchically organized and interact with light as volumes instead of surfaces. The additional implicit assumption of homogenous mixing between leaf and wood is also not met. Improved representation of tree branching structure based on connected cylinders may
contribute to improve the integration of woody elements [106,107]. A mixed representation using voxels for foliage and meshed volumes for woody parts is entirely compatible with the DART model, however the ALS point cloud alone does not provide sufficient resolution to achieve such level of detail of canopy geometry description. Leaf and wood are virtually inseparable in ALS point clouds as many returns are mixed points (with both leaves and wood contributing to a particular return) due to the laser footprint size (around 25 cm diameter here). Tree geometry reconstruction would further be hampered by the limited pulse density, the geometric accuracy (limited by the laser footprint size), and the high occlusion rate occurring in dense forests. Finally, small woody elements such as the twigs have been shown to significantly affect the reflectance more than large woody parts such as branches or trunks [54], which suggests that the improved integration of the main wood structure in a mixed representation, which could be considered with TLS measurements, should also be refined with a procedure allocating small elements.

5.4. Influence of LOP Variability on Species Separability and Discrimination

The inadequate representation of LOP spatial variability within crowns, among crowns or species was reported as one of the main sources of error in RTM studies [49,108,109]. To solve these issues, we tested three scenarios accounting for LOP spatial variability: LOP\textsubscript{species}, LOP\textsubscript{ITC} and LOP\textsubscript{pixel}.

Agreement between imaging spectroscopy and DART simulations was assessed by comparing the whole spectrum using statistical metrics, species separability using spectral dissimilarity metrics and potential of species discrimination using discriminant analysis. The lowest agreement was obtained by LOP\textsubscript{species} representation. This scenario led to overestimation of species separability. The LOP\textsubscript{ITC} representation showed intermediate results while the most faithful scenario was LOP\textsubscript{pixel}.

The efforts made in the definition of LOPs turned out to be an important discriminating axis of species. However, we were constrained by the lack of detailed description of other sources of variability of the LOPs such as the foliar clumping, N, LIDF or LMA homogenized on the whole 3D mockup. This potentially explain the lower discrimination of species in simulated images (LOP\textsubscript{pixel} and LOP\textsubscript{ITC}).

The fine scale definition of leaf optical properties partly allows conservation of the spatial heterogeneity of reflectance, and appears to be a promising framework to explore the spectral variation hypothesis by means of remote-sensing simulations.

6. Conclusions

DART RTM was used to simulate imaging spectroscopy in tropical forests. We took advantage of ALS to reconstruct a detailed forest 3D architecture. With the aim of defining a framework for producing realistic simulation in tropical forest, the influence of NPVf and LOP variability on simulated reflectance was explored. Analyses were performed using sunlit pixels of a set of ITCs. In order to estimate pigment content used as input for LOP characterization of each pixel, we used a hybrid approach combining 3D RTM and spectral indices.

NPVf were incorporated into simulations either considering WAD or C\textsubscript{brown} as proxies. Three scenarios including LOP\textsubscript{species}, LOP\textsubscript{ITC} and LOP\textsubscript{pixel} enabled the assessment of the effect of LOP variability on simulated reflectance. Comparison between imaging spectroscopy and DART simulated images was performed using statistical metrics, spectral dissimilarity metric and supervised classification.

Results showed that simulations which did not account for NPVf resulted in moderate correlation ($R^2$ 0.10–0.62) between measured and simulated canopy reflectance in the VIS domain, and consistently low correlation ($R^2$ 0.08–0.18) in the red edge and NIR domain, associated with relatively high RMSE. Using C\textsubscript{brown} as a proxy for NPVf resulted in similar performances in the VIS domain, lower RMSE between measured and simulated canopy reflectance in the red edge and NIR domains, but still relatively low correlation. Since good agreement between imaging spectroscopy and DART simulation were obtained
for $NPVf_{\text{brown}}$, the underestimation of the reflectance observed in the blue and the red spectral domains when including $C_{\text{brown}}$ suggested the possibility to combine the two approaches, with an appropriate value of $NPVf_{WAD}$ in order to optimize the agreement between measured and simulated reflectance over the full VNIR range.

The definition of specific LOP for each pixel in a DART simulation resulted in more realistic simulations than when using the same LOP for all pixels in a tree crown, or when using the same LOP for all pixels in all tree crowns of a species. This was observed for radiometry, spectral dissimilarity and performances when applying supervised classification. We conclude that the strategy usually adopted in the literature, consisting in assigning one set of LOP per species results in strong overestimation of the capacity of supervised approaches to perform species discrimination. Simulations of remote sensing data based on physical modeling should put more efforts into accounting for variations of leaf optical properties within species, and within individual crowns when relevant regarding the spatial resolution of the simulations. Spectral variation hypothesis needs to consider the large variability within species when inferring species diversity patterns from spectral diversity.

Overall, results showed that the incorporation of $NPVf$ combined with LOP variability are relevant for simulating spatial heterogeneity in remotely sensed data in tropical forests. However, the improvement of leaf trait definition and characterization of structural parameters should be addressed in order to increase radiometric agreement between simulation and imaging spectroscopy data. Our simulation framework (LOP per pixel approach) showed an interesting potential for benchmarking remote-sensing applications dedicated to biodiversity monitoring when applied to satellite missions currently operational as well as for future missions currently in preparation [110–112].

Author Contributions: Conceptualization, D.M.E., C.W. and J.-B.F.; formal analysis, D.M.E., F.d.B., G.V. and J.-B.F.; methodology, D.M.E. and J.-B.F.; supervision, C.W. and J.-B.F.; visualization, D.M.E.; writing—original draft preparation, D.M.E.; writing—review and editing, F.d.B., G.V., C.W. and J.-B.F.; project administration, J.-B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: D. Ebengo, F. de Boissieu and J.-B. Féret acknowledged financial support from Agence Nationale de la Recherche (BioCop project—ANR-17-CE32-0001-01) and from TOSCA program grant of the French Space Agency (CNES) (HyperTropik/HyperBIO projects). We thank the Region Occitanie for funding of D. Ebengo doctoral researches. Vincent has benefited from Investissement d’Avenir grants of the ANR (CEBA: ANR-10-LABX-0025). We also acknowledge Jean-Philippe Castellu-Etchegorrory, Nicolas Laurent and the DART Team for their advice and support in using the DART model. Thanks to A. Laybros, M. Aubry-Kientz, R. Dutrieux and C. Baltzer for field support and data cleaning and preparation Hyperspectral imagery acquisition was funded by the CNES (Centre National des Etudes Spatiales) and acquired by Hi-Tech Imaging.

Conflicts of Interest: The authors declare no conflict of interest.

References


