Estimating vegetation traits of mediterranean open canopies using imaging spectroscopy

Thomas Miraglio

To cite this version:

Thomas Miraglio. Estimating vegetation traits of mediterranean open canopies using imaging spectroscopy. Engineering Sciences [physics]. ISAE - Institut Supérieur de l'Aéronautique et de l'Espace, 2021. English. NNT : . tel-03927504

HAL Id: tel-03927504
https://hal.science/tel-03927504

Submitted on 6 Jan 2023

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut Supérieur de l'Aéronautique et de l'Espace

Présentée et soutenue par

Thomas MIRAGLIO

Le 7 octobre 2021

Estimation de traits de végétation de canopées ouvertes méditerranéennes par télédétection hyperspectrale

Ecole doctorale : **GEETS - Génie Electrique Electronique,Télécommunications et Santé : du système au nanosystème**

Spécialité : **Photonique et Systèmes Optoélectronique**

Unité de recherche :
ISAE-ONERA OLIMPES Optronique, Laser, Imagerie Physique et Environnement Spatial

Thèse dirigée par
Xavier BRIOTTET et Susan USTIN

Jury
M. Stéphane JACQUEMOUD, Rapporteur
M. Kamel SOUDANI, Rapporteur
Mme Agnès BÉGUÉ, Examinatrice
M. Clement ATZBERGER, Examinateur
M. Jochem VERRELST, Examinateur
Mme Karine ADELINE, Examinatrice
M. Xavier BRIOTTET, Directeur de thèse
Mme Susan USTIN, Co-directrice de thèse

Acknowledgements

Compte-tenu du fait que les personnes ayant gravité autour de ma thèse n'étaient pas forcément francophones, j'ai préféré écrire ces remerciements en franglais.

Je tiens tout d'abord à remercier les rapporteurs de cette thèse pour le temps qu'ils ont consacré à sa relecture, et plus généralement les membres du jury pour avoir accepté d'évaluer ces travaux. J'ai une pensée particulière pour Stéphane Jacquemoud, que j'ai rencontré il y a maintenant quatre ans alors que je souhaitais me reconvertir, et qui m'a donné envie de suivre le master dont il était responsable et m'a fait découvrir le monde de la télédétection appliquée à la végétation.

Pour ces trois ans passés à suivre mes travaux, je me dois également de remercier Susan Ustin, Karine Adeline, Margarita Huesca, et surtout Xavier Briottet. Les échanges que j'ai pu avoir avec Xavier, malgré son emploi du temps plus que chargé, m'ont permis de garder un cap, et plus encore, de garder le moral malgré les déconvenues. Unfortunately, I was not able to visit Susan Ustin due to the pandemic, however I am very thankful to both her and Xavier for the trust they had in me and in my work. Hopefully one day I will finally be able to come to Davis, and visit these sites that I studied for three years.

I also thank Crystal Schaaf and Peter Boucher, for their warm welcome during the two weeks I spent in their lab at the University of Massachussets, Boston, and who gave me a taste of what life was like on the other side of the Atlantic.

To the thesis assessment committee members, Guerric Le Maire and Zbyněk Malenovský, who accepted to review my work on two occasions and gave suggestions, insight, and advices in order to improve the quality of my work, I express my gratitude.

Malheureusement, les différents confinements et le télétravail ont grandement réduit le temps que j'ai pu passer avec eux, mais je tiens à remercier l'équipe du DOTA, et en particulier Thomas Rivière, Stéphanie Doz et Philippe Déliot pour la coinche du midi ; Ahmed Moussous, pour les cafés et les parties de Age of Empires 2 ; Pierre-Yves, qui m'a poussé à m'inscrire à ce trail de 12 heures à Bruniquel que jamais je n'aurais cru finir ; Véronique Achard, Sophie Fabre, Nicolas Rivière, Laurent Poutier, Laure Rupioz.

Bien entendu, je n'ai pas encore évoqué les doctorants du DOTA, avec qui j'ai pu partager cette aventure. Je ne connaissais personne en arrivant à Toulouse, mais cela a bien vite changé. Pour leur accueil lors de mon arrivée, je remercie tout d'abord Guillaume, Florian, Simon et Sandra. Merci également à Nicolas, Matthieu, Arnaud, Rollin, mais aussi à Romain, Thierry, William, Gabriel, Erwan, Killian et Yohann pour toutes ces pauses café, ces belotes et (encore) ces coinches, ces moments passés à discuter, et ces sorties. Je n'aurai probablement plus de collègue de bureau mettant Barbara, Brel ou Céline Dion comme fond sonore pour travailler (ou pour chanter, selon l'humeur). Enfin, j'ai une pensée pour Lucas, qui a commencé sa thèse peu après moi et devrait la finir bientôt, pour ces moments à se soutenir mutuellement et à échanger sur nos déboires respectifs.

Merci à Étienne, qui m'a initié à la plongée sous-marine, ainsi qu'à Stéphane, Éric, Wilfrid, et aux autres membres du club plongée de l'AS, pour tous ces moments passés sous l'eau ou bien à désaturer autour d'un verre.

To Adrien, Ameline, Lorena, to whom I wish all the best for the last year of her PhD, Merve, and my international correspondents of /Culture Pals/, for their support during at least a part of those three years: thank you.

Enfin, merci à mes parents, Roger et Sophie, pour leur soutien depuis toujours, mais également à Margaretha, Joseph, Françoise et Claude, mes grand-parents, et à Manon, ma sœur. J'espère les rendre fiers.

Contents

List of Figures

List of Tables

List of Acronyms

1D	*One Dimensional*
3D	*Three Dimensional*
ALA	*Average Leaf Angle*
ANN	*Artificial Neural Network*
ANOVA	*Analysis Of VAriance*
AVIRIS-C	*AVIRIS Classic*
AVIRIS-NG	*AVIRIS Next-Generation*
BRF	*Bidirectional Reflectance Factor*
$\mathbf{C}_{ab}$	*Chlorophyll Content*
Car	*Carotenoid Content*
$\mathbf{C}_{ant}$	*Leaf Anthocyanins Content*
CBL	*Compact Biomass Lidar*
CC	*Canopy Cover*
DHP	*Digital Hemispherical Photography*
DOY	*Day Of Year*
EBV	*Essential Biodiversity Variables*
EWT	*Equivalent Water Thickness*
FMC	*Fuel Moisture Content*
FWHM	*Full Width at Half Maximum*
GEO BON	*Group on Earth Observations Biodiversity Observation Network*
GPR	*Gaussian Process Regression*
GSD	*Ground Sampling Distance*
LAD	*Leaf Angle Distribution*
LAI	*Leaf Area Index*
LMA	*Leaf Mass per Area*

LUT	*Look-Up Table*
NEdL	*Noise-Equivalent Delta Radiance*
NEdR	*Noise-Equivalent Delta Reflectance*
NIR	*Near InfraRed*
NPV	*Non Photosynthetic Vegetation*
PAI	*Plant Area Index*
PDT	*Pacific Daylight Time*
PST	*Pacific Standard Time*
PLSR	*Partial Least-Square Regression*
RAMI	*RAdiation transfer Model Intercomparison*
RFR	*Random Forest Regression*
RMSE	*Root-Mean-Square Error*
RTM	*Radiative Transfer Model*
SAC	*Spectral Absorption Coefficient*
SAM	*Spectral Angle Mapper*
SFR	*Simple Forest Representation*
SMLR	*Stepwise MultiLinear Regression*
SNR	*Signal-to-Noise Ratio*
STDB	*standard deviation of the difference between estimated points and the regression line*
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*
SWIR	*Short-Wave InfraRed*
TLS	*Terrestrial Laser Scanner*
VI	*Vegetation Index*
VIP	*Variable Importance in Projection*

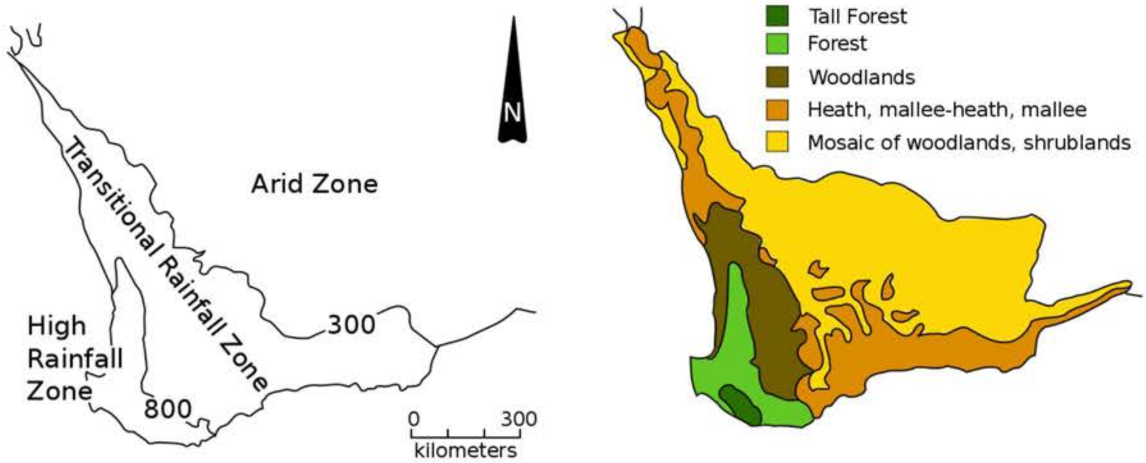
Introduction

The Mediterranean Ecoregions and their Characteristics

Mediterranean-climate ecoregions are found in California, Chile, South Africa, Australia, and of course around the Mediterranean Basin, within 30° and 40° of the equator and always on the western sides of the continents (Figure 1). They are characterized by prolonged hot, dry summer periods, and mild, wet winters (Fath, 2018). Extreme events, either in the form of storms and massive precipitations or recurrent forest fires, also define the Mediterranean climate. These ecosystems saw the early development of many human settlements and civilizations that, in turn, have shaped them through burning, logging and livestock grazing, extending back to 10,000 years around the Mediterranean Basin (Caldwell et al., 1995; Jiménez-Ruano, Mimbrero, and de la Riva Fernández, 2009; Scarascia-Mugnozza et al., 2000).

Figure 1: Location of the Mediterranean ecoregions (in brown). Cowling et al., 1996.

Although these ecoregions span only 2% of land surfaces, they are home to an exceptional biodiversity. Mediterranean ecosystems are considered to be major biodiversity hot spots (Medail and Quezel, 1997; Myers et al., 2000), and it is estimated that almost 20% of the Earth's plant species reside there, with a significant portion being endemic (Cowling et al., 1996). These plant species are distributed over numerous habitats and vegetation types, encompassing grasslands, forests and shrublands such as the emblematic maquis and chapparal. Overall, true Mediterranean forests are scarce due to historically strong climate and anthropogenic pressures, and preforests and presteppic forests, presenting a lower tree density, are a more common occurence. For instance, in Southwestern Australia, open canopies compose the majority of plant communities due to the limited extent of high-precipitation areas (Figure 2).



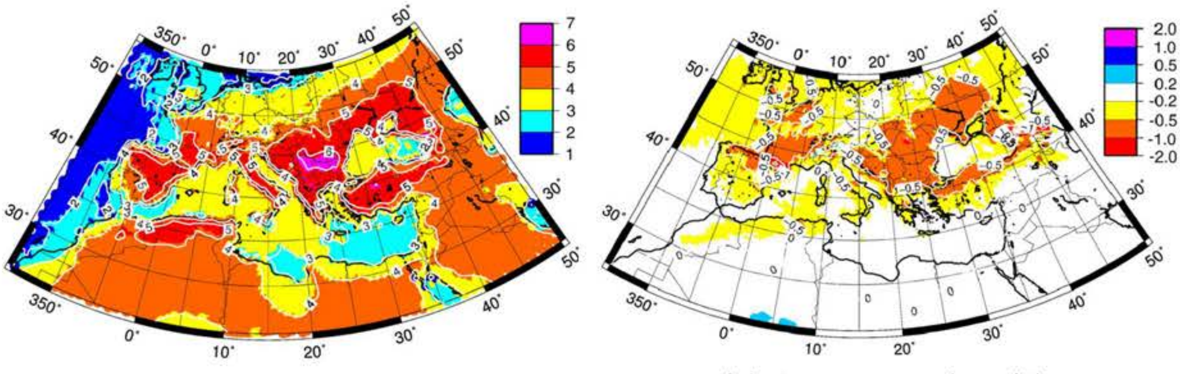
(a) Rainfall isoyets (mm/yr).

(b) Major vegetation zones as defined by Beard, 1975.

Figure 2: Precipitation and vegetation zones over the South West Botanical Province and South West Interzone in Australia. As per Beard, 1975, forests correspond to closed canopies, woodlands to open canopies, and mallee to eucalypts shrublands. Adapted from Hopper, 1979.

Environmental and Anthropogenic Pressures on Vegetation

Current projections for climate change at the global scale indicate a general warming. Mediterranean ecoregions in general are expected to face increased temperatures and decreased precipitations. Focusing on the Mediterranean Basin, climate models agree that temperatures will have considerably risen by 2100 and that precipitations and most importantly soil moisture will significantly decrease, leading to increased frequency of heatwaves and drought periods (Füssel et al., 2012; Giorgi and Lionello, 2008; Lionello et al., 2014; Somot et al., 2008; Vautard et al., 2007). Figure 3 illustrates the projected climatic changes over Western Europe and North Africa.



(a) 2 m surface temperature (°C).

(b) Precipitation (mm/d).

Figure 3: Projected average changes for surface temperatures and precipitation between the 1961–1990 and 2070–2099 periods (adapted from Somot et al., 2008).

Vegetation has developed numerous strategies to face the harsh conditions of Mediterranean ecosystems. The recurrent droughts, that are the main drivers for restricting plant productivity and growth, have led to deep root systems and small leaves to limit evapotranspiration (Mooney and Dunn, 1970). Under drought stress, plant species also show decreases in growth, hydraulic conductivity or general photosynthetic activity (Sardans and Peñuelas, 2013). However, new climatic patterns are threatening the biodiversity of the Mediterranean biome. Even longer and frequent droughts and general aridity will have considerable detrimental effect on vegetation health status, leading to massive forest die-offs (Choat et al., 2012; Goulden and Bales, 2019).

Forest fires are another strong selective factor in Mediterranean ecosystems, and can participate in vegetation rejuvenation through increased light and water availability in the burned areas due to decreased competition (Fath, 2018; Jiménez-Ruano, Mimbrero, and de la Riva Fernández, 2009). Alternatively, the general drying that is expected to occur should significantly increase the frequency of widlfires, which could lead to land degradation, erosion and nutrient loss through runoffs. While endemic Mediterranean plant species are adapted to forest fires, complete vegetation regeneration would no longer be possible if the time span between burnings was reduced too much (Médail, 2017; Moriondo et al., 2006; Syphard et al., 2009; Tessler et al., 2016).

Finally, anthropogenic land use through habitation, cultures and herbivore grazing has historically created environments with open canopies and dominant low species. Because of the mild winters and extended warm periods, the region's favorable weather conditions have encouraged population growth in the five Mediterranean areas faster than the global average. Between 1990 and 2000, population density increased on average by 14% and the extent of urban areas by 10%, so much so that urban areas covered 9.2% of Mediterranean ecoregions (Underwood et al., 2009). In conjunction, agricultural areas occupy a significant portion of the land, averaging 24% over the five regions. The fragmenting of Mediterranean ecosystems that results from human land use has already led to significant alteration of ecosystem processes and vegetation dynamics, and will be detrimental to the adaptation of native plant species to climate change. Indeed, Mediterranean climate extent will shift over the present century, and the potential new habitats may be isolated from historical ones due to the absence of colonization corridors, and would therefore be inaccessible to endemic Mediterranean species (Caldwell et al., 1995; Klausmeyer and Shaw, 2009).

Overall, it is projected that Mediterranean ecosystems will have experienced a large, if not the largest, loss of biodiversity by 2100 (Sala et al., 2000). A thorough monitoring of vegetation health status is therefore critical in order to guide present and future conservation policies.

Monitoring Vegetation Health Status

Measuring forest health first requires to define health indicators. For instance, forestry usually only considers forest productivity, overlooking vegetation structure and biomass. A healthy forest may also be defined as a functional community whose future is not threatened, that is resilient to change and able to recover from stressors (Edmonds, Agee, and Gara, 2000). However, various parts of a forest may be dying at a given time, before being regenerated as part of a natural cycle. A healthy forest could therefore be defined as a forest that continuously presents various patches at different health levels (Trumbore, Brando, and Hartmann, 2015). A significant increase of unhealthy trees from the norm would therefore indicate a degradation of the ecosystem: Figure 4 shows the effects of a bark-beetle outbreak in North American forests, with a unusually high tree die-off (in red). Monitoring forests' health status thus requires the ability to make measurements over large regions on a regular time basis.

Figure 4: A high number of dead trees (in red) in a North American forest, after massive bark-beetle outbreaks facilitated by the warming temperatures. From Ramsfield et al., 2016.

In order to organize biodiversity monitoring efforts, *Essential Biodiversity Variables* (EBV) have been developed by the *Group on Earth Observations Biodiversity Observation Network* (GEO BON) and organized into six major classes: genetic composition (*e.g.* genotypes of species); species populations (population surveys, distribution); species traits (*e.g.* phenology); community composition (taxa surveys at different locations); ecosystem structure (canopy closure, height); ecosystem function (nutrient, carbon and water cycles) (Pereira et al., 2013). Much of the variables selected as EBV require extensive and intensive *in situ* surveys and measurements, which may hinder their estimation. While a clear framework for it has not yet been defined, satellite remote sensing presents a great opportunity to dramatically expand both spatially and temporally the EBV data sets by providing a global, frequent coverage (Pettorelli et al., 2016) for some classes (species traits, community composition, ecosystem structure, ecosystem function) due to existing or emerging remote sensing technics (Pereira et al., 2013).

Remote Sensing and its Limitations for Ecosystem Monitoring

Airborne and satellite remote sensing, either active or passive, provides information about land surfaces by measuring their reflected or emitted radiations. Active remote sensing of surfaces can involve radar, which emits radio waves and measures the intensity and polarization of the reflected signal to infer the surfaces characteristics, or lidar, which illuminates its environment with laser light and measures light return times and intensities to make *Three Dimensional* (3D) representations of its surroundings. Active techniques are used in the context of vegetation monitoring: lidar technology has been sucessfully used for forest inventory purposes (White et al., 2016; Wulder et al., 2012), and was also coupled with other remote sensing methods through data fusion: for instance, Asner et al., 2015 took advantage of lidar measurements to better isolate sunlit crown pixels and improve estimation accuracy.

However, most methods rely on passive remote sensing over the 0.4–2.5 μm spectral range (Gamon et al., 2019) using multispectral or hyperspectral sensors (Figure 5). Indeed, a lot of plant traits influence plant optical properties, both at the leaf and canopy scale (Cavender-Bares et al., 2017; Ustin and Gamon, 2010), which makes it possible to estimate them from the spectral reflectance. Asner et al., 2011a further demonstrated the potential of hyperspectral remote sensing for vegetation estimation at the global scale.

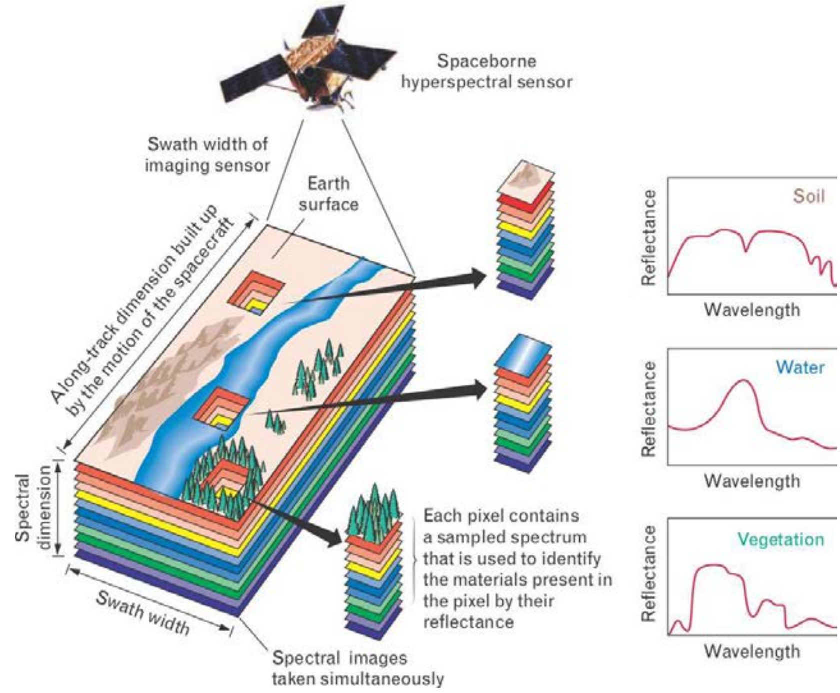


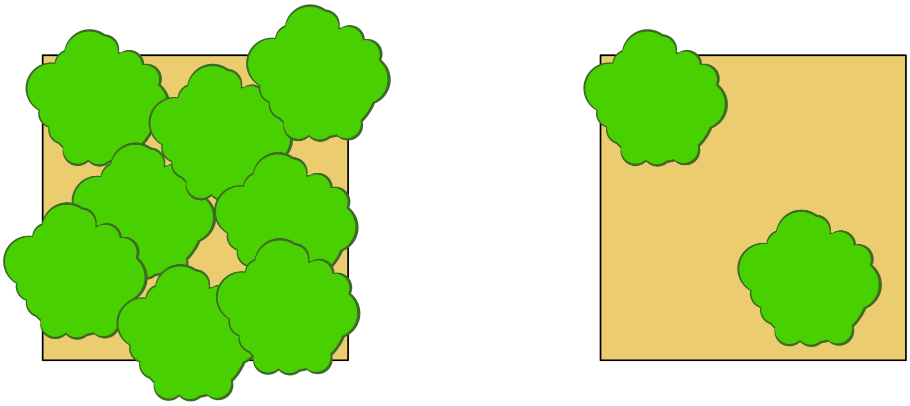
Figure 5: The concept of passive remote sensing of surfaces in the reflective domain: a sensor measures spectral reflectances at various wavelengths over large swathes. The resulting signal in each pixel is then processed to retrieve information about the scene. From Shaw and Burke, 2003.

Most hyperspectral imaging spectrometers currently in use are airborne. They usually present high spectral resolutions from 10 to below 1 nm (Schaepman et al., 2015; Vane et al., 1993; Wieneke et al., 2016). While the *Ground Sampling Distance* (GSD) of the acquired images can range from a few dozen of centimeters to 20 meters, a usual pixel surface for airborne acquisitions is about 1 m^2. Images acquired with these systems, with fine spectral and spatial resolutions, can resolve individual tree crowns and allow for the retrieval of several plant traits, as demonstrated over various tropical forests in the works of Dana Chadwick and Asner, 2016 and Asner et al., 2011b. However, airborne acquisitions are temporally and spatially limited, with at best a few overflights over some zones of interest, and a thorough knowledge of the study sites thanks to field acquisitions.

Satellite remote sensing addresses both of these problems. Historically, most satellite-based sensors were multispectral, *i.e.* only measured a few large spectral bands, which considerably limited the number of vegetation traits that could be retrieved (Barnes, Xiong, and Salomonson, 2002; Frampton et al., 2013; Markham et al., 2004). Hyperspectral satellite sensors, which would potentially allow for the retrieval of considerably more traits, are more recent (EO-1/Hyperion was launched in 2000 (Pearlman et al., 2003)) and have only been a handful. The last couple of years, new hyperspectral missions such as PRISMA (Stefano et al., 2013) have begun, while others, such as EnMAP (Guanter et al., 2015), CHIME (Rast et al., 2019), SBG (inheritor mission of HyspIRI (Lee et al., 2015)), Biodiversity (inheritor mission of HYPXIM (Carrere et al., 2013)) or PRISMA Second generation, are still in preparatory phase.

Unfortunately, the spatial resolutions of satellite sensors range from 8 to 30 m: at this spatial resolution, individual tree crowns are often not resolved, and vegetation reflectance may only represent a fraction of the total signal, greatly increasing retrieval difficulties (Figure 6). Zarco-Tejada et al., 2019 showed that leaf chlorophyll content could be retrieved from 20 m resolution Sentinel-2 images over sparse conifer forests by using a *Vegetation Index* (VI). However, this required sufficient field knowledge to calibrate the VI. In the context of global satellite remote sensing, knowledge about the imaged ecosystems is by nature limited, which adds to the difficulties of designing appropriate estimation methods. Still, Darvishzadeh et al., 2019, again using Sentinel-2 imagery, demonstrated that needle chlorophyll content could be retrieved through *Look-Up Table* (LUT)-based inversion without field calibration data over a temperate, closed conifer canopy.

Overall, hyperspectral sensors have been shown to have a great potential for vegetation studies, with possible estimation of several vegetation traits over large areas, in particular when working with airborne high-resolution data. The most commonly estimated biophysical and biochemical traits in remote sensing studies are *Leaf Area Index* (LAI) and the leaf chlorophyll content (Berger et al., 2018): the variety of traits retrievable from satellite medium spatial resolution hyperspectral instruments, and the methods to retrieve them, must still be identified (Hill, Buddenbaum, and Townsend, 2019). This is especially true for open canopies, that are dominant in the Mediterranean biome, due to the reduced vegetation signal at such spatial resolutions. In particular, *Equivalent Water Thickness* (EWT) and *Leaf Mass per Area* (LMA) are two other vegetation traits that could be of prime interest in the context of increased forest fires as they are used to compute the *Fuel Moisture Content* (FMC), an important predictor of wildfire occurence.



(a) Closed canopy.

(b) Open canopy.

Figure 6: Illustration of the pixel content of medium spatial resolution images acquired over two kinds of canopies.

Objectives of the Thesis

In the context of the recent and future satellite hyperspectral mission (namely SBG (NASA) and Biodiversity (CNES)), the present thesis aims at developing methods for the estimation of vegetation traits over open canopies from medium resolution hyperspectral imagery, when field knowledge is insufficient to calibrate regression models.

Due to previous field campaigns, inventories of two botanicals groups (broad-leaved trees and conifers) were either already available or easy to retrieve using classification methods. As such, the research work could focus specifically on vegetation traits retrieval, under the hypothesis that mixed conifer-broad-leaved pixels were equivalent to pure broad-leaved pixels for estimation purposes.

Chapter 1 of this manuscript gives a state-of-the-art of the current knowledge concerning the remote sensing of vegetation traits and the various retrieval methods in use.

In Chapter 2, descriptions of the study sites, of the various *in situ* datasets and of the hyperspectral images available are given.

Due to the lack of hyperspectral satellite data acquired at the same time as the *in situ* measurements, the works presented in the subsequent three chapters of this document focused on airborne imagery.

In Chapter 3, the objective was to study the appropriateness of a basic forest representation in a 3D *Radiative Transfer Model* (RTM) for the estimation of LAI and foliar pigments from medium-resolution imagery. First, a database aggregating field measurements done over the study sites was built to later serve for validation purposes. Then, the idea was to model the canopy with "lollipop trees" in a 3D RTM. LAI and foliar pigments contents were then estimated through LUT-based inversion. Estimates accuracy was then compared to the literature values in order to validate the modeling.

In Chapter 4, the objective was to reduce the LUT building times resulting from the use of a 3D RTM by developing a new LUT building method, in order to estimate a multitude a vegetation traits (gap fraction, leaf pigments, EWT, LMA). Two LUT were generated using a *One Dimensional* (1D) and a 3D RTM in order to calibrate transformation factors from one to the other. Then, an assessment criterion was established and the transformation error was estimated in order to verify that it was acceptable with regards to the spectral variations of the original LUT. Then, estimation of the various vegetation traits was done with the new LUT building method in order to assess its performances over an operational case.

In Chapter 5, the work focused again on the modeling of the canopy, to identify structural parameters with a significant influence over canopy reflectance in order to better model canopies and obtain acceptable EWT and LMA estimations. A sensitivity study of the 3D RTM outputs to the canopy structural parameters was done in order to identify those with a statistically significant influence. Then, EWT and LMA were estimated from synthetic hyperspectral images using machine learning methods trained with database obtained with various canopy modelings, in order to compare their performances. EWT and LMA were subsequently estimated from real measured hyperspectral images in order to confirm the previous results.

Finally, the objective of Chapter 6 was to assess the retrieval capabilities of the various vegetation traits previously studied from synthetic hyperspectral Biodiversity and SBG images, to better understand the limitations of each imager concerning vegetation monitoring. Synthetic images were generated from airborne imagery to serve as surrogates. Then, regressors were trained over databases generated from a 3D RTM with one of the modelings identified in the previous chapter in order to estimate various vegetation traits. Lastly, estimates from SBG and Biodiversity images were both intercompared and compared to field measurements to better understand the limits of each sensor.

A general discussion, a conclusion, and perspectives are then given in chapter 6.

The present document is written in English and followed by a synthesis in French summarizing the main results of this thesis.

Chapter 1

State-of-the-Art of Hyperspectral Remote Sensing of Vegetation and Associated Retrieval Methods

---

Contents

1.1 Introduction

The present chapter presents the state of the art of current methods dedicated to the remote sensing of vegetation traits. It first describes the physical basis of interactions between radiation and vegetation, from the leaf to the canopy. The second section gives an overview of the radiative transfer codes presently in use to model these interactions on a physical basis. Finally, the last section presents the main families of retrieval methods and in which context they have been employed.

1.2 Leaf and Canopy Traits and Their Optical Properties

The optical properties of canopies depend on numerous factors, as the radiation reflected from a canopy is influenced by the understory, non-photosynthetic elements of the vegetation, and the leaves themselves. In Europe, it would seem that the study of leaf colors can be dated as far back as Aristotle and his *De Coloribus* in the 4th century BC. However, effective measurements of leaves' optical properties have only been done for a little more than a century, with the pioneering work of Coblentz, 1913. All leaf constituents influence the spectral signature of vegetation in a specific manner in the visible domain (0.4–0.76 µm) as well as the near- and short-wave-infrared (0.76–1.4 µm and 1.4–2.5 µm, respectively), as illustrated in Figure 1.1. The effect of leaves on canopy reflectance is further dependent on the characteristics of the vegetation type and cover, that will determine how much they will interact with the radiation. The present section presents the influence of various leaf and canopy traits on vegetation reflectance.

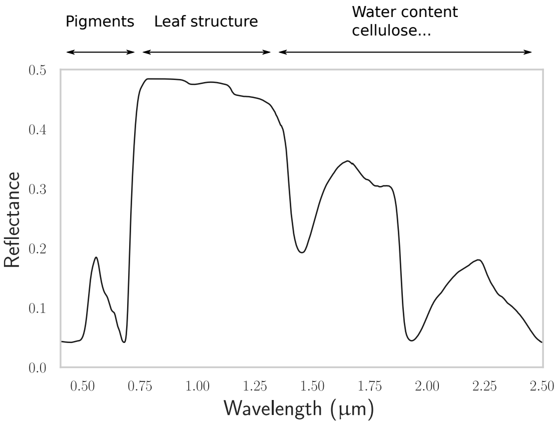
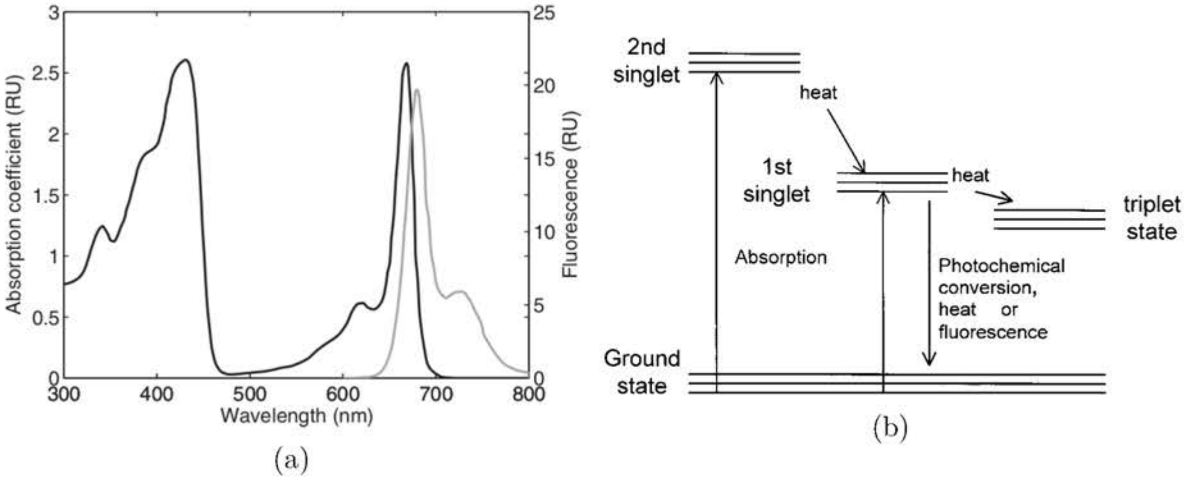


Figure 1.1: Spectral ranges of influence of some leaf parameters over a typical leaf reflectance spectrum.

1.2.1 Leaf Pigments

Plants are photoautotrophs, organisms capturing photons and carbon to create complex energy-storing organic compounds. Photon capture occurs in the chloroplast, through absorption by the two main leaf pigments: chlorophylls *a* and *b*. These pigments absorb light over the ultaviolet and visible spectral domain, with absorption peaks located in the violet, blue and red regions (Figure 1.2a). The captured photons raise chlorophylls energy level depending on their wavelength. However, chlorophylls excited by blue light (2nd singlet, see Figure 1.2b) are unstable, and reach the stable 1st singlet through heat loss. Chlorophylls can therefore attain an appropriate excitation state no matter the absorbed photon's wavelength (Lambers, Chapin, and Pons, 2008). From this stable excited state, in order to return to their ground state, chlorophylls will either: transfer their energy so that it is used in photochemistry, convert it to heat, or emit a photon as part of the fluorescence process (Krause and Weis, 1991).



(a)

(b)

Figure 1.2: (a) Absorption (black) and emission (gray) spectra of chlorophyll *a* in a methanol solution. (adapted from Jacquemoud and Ustin, 2019) and (b) qualitative energy levels of a chlorophyll molecule. Blue and red light absorption lead to the 2nd and 1st states, respectively. Transition from the 2nd singlet to the 1st one is obtained through heat loss. From the 1st state, photosynthetic processes can occur (from Chow, 2003).

Chloroplasts contain the majority of plant nutrients (Makino and Osmond, 1991), and, provided favorable conditions, chlorophylls are constantly renewed during plants' period of activity (Croft and Chen, 2017). In the Fall, chlorophylls are no longer replaced when they are broken down, ensuring relocation of the nutrients in the storage organs of the plants before the shedding of the leaves (Migita, Chiba, and Tange, 2007). Due to their importance in the photosynthetic process, their strong spectral signature, and because of their high relative abundance, chlorophylls are the most commonly estimated leaf pigments.

Carotenoids represent a family of accessory leaf pigments encompassing two subgroups, carotenes and xanthophylls. These pigments have a double function, and serve both in the photosynthetic and photoprotective processes of the leaves (Demmig-Adams and Adams, 1996; Ritz et al., 2000). They complement the sunlight harvest of chlorophylls by absorbing photons at different wavelengths, and prevent damage of leaf tissues by dissipating the excess energy as heat through the xanthophyll cycle (Thayer and Björkman, 1990). An illustration of the various energy transfers among leaf pigments, in which carotenoids take part, is given in Figure 1.3. While carotenoids are present during all the leaf life-cycle, they become visible during the senescence period, as chlorophylls are broken before carotenoids (Sanger, 1971).

Anthocyanins represent an ensemble of pigments mostly responsible for the red color of leaves. Their specific role is not well understood, however it is believed that they are created in periods of environmental stress to protect the photosystems of excessive sunlight (Flint, Jordan, and Caldwell, 1985). Excluding the few red-leaved species, anthocyanins are only present during specific stages of a leaf's lifetime: the juvenile and senescent stages, and as a response to droughts or cold temperatures (Chalker-Scott, 1999; Merzlyak et al., 2008).

An illustration of the evolution of these pigments' concentrations in the leaves during

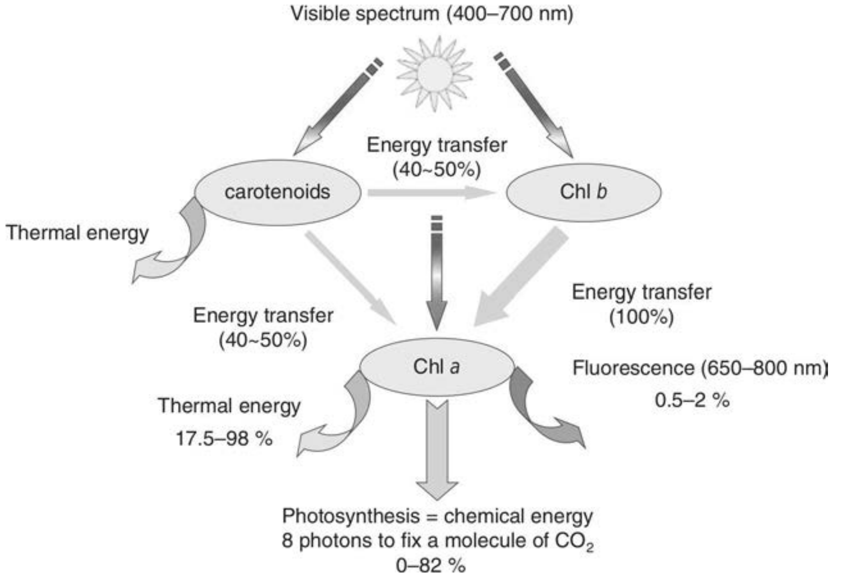


Figure 1.3: Energy transfers from the various photosynthetic pigments of leaves. From Jacquemoud and Ustin, 2019.

senescence is given Figure 1.4.

Pigments almost exclusively affect the visible region of the light spectrum. The spectral behaviour of leaves in the infrared is driven by different components, more closely related to the plants' ecological strategies.

1.2.2 Water and Dry Matter Contents

Water represents from 70 to 95% of the leaf biomass, and allows for the transport of metabolites in the cells and of various materials generally within the plant. It also ensures vegetation support, as thanks to water, plant cells exert a positive pressure on their walls (turgor). Vegetation requires a large amount of water, as much of it is used for thermoregulation purposes: 99% of the water absorbed by the plants is lost to transpiration from the leaves' stomates, the "air ducts" through which the CO_2 required for the photosynthesis is captured (Lambers, Chapin, and Pons, 2008), therefore cooling the plant. Water being present in abundant quantities in the leaves, its spectral signature is clearly visible (Figure 1.5a) at the leaf level. However, difficulties arise when imaging ecosystems from airborne or satellite sensors: indeed, several bands are also strongly affected by atmospheric water vapor absorption and have a low *Signal-to-Noise Ratio* (SNR), which limits the spectral bands of interest (San and Süzen, 2011; Thenkabail, Lyon, and Huete, 2018).

In the event of water stress, plants can partially or completely close their stomates and limit the transport of water from the roots to the leaves. While this helps prevent excessive water loss, it comes at the cost of limiting the photosynthetic process and risking permanent damage to the leaves through overheating. Leaves grown under water-stress conditions will

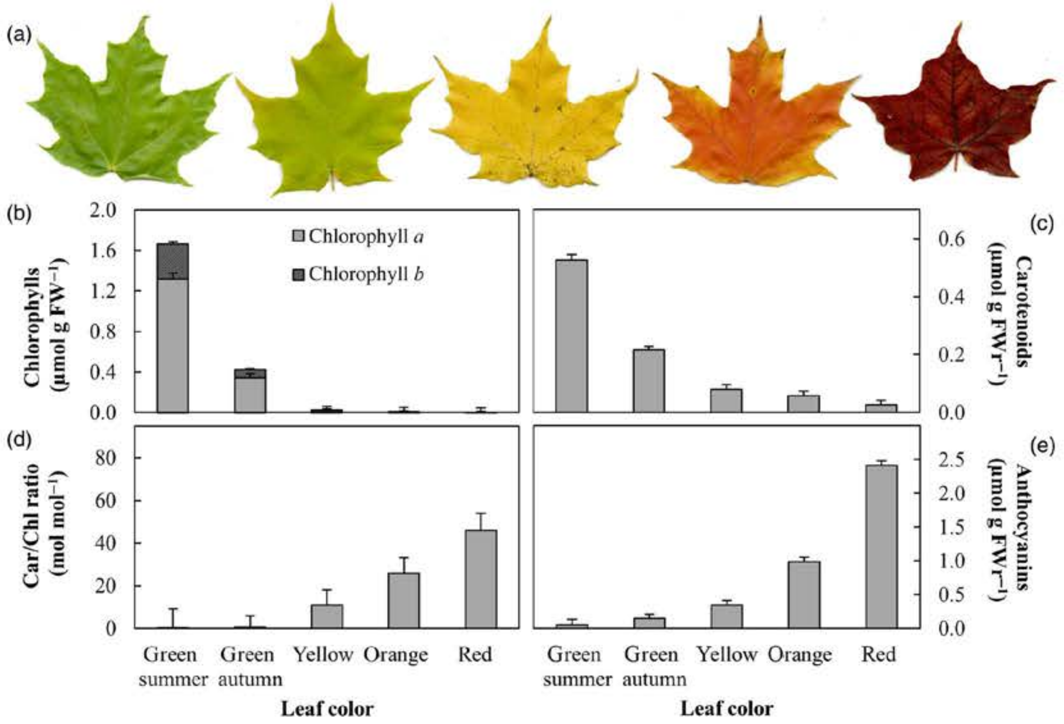
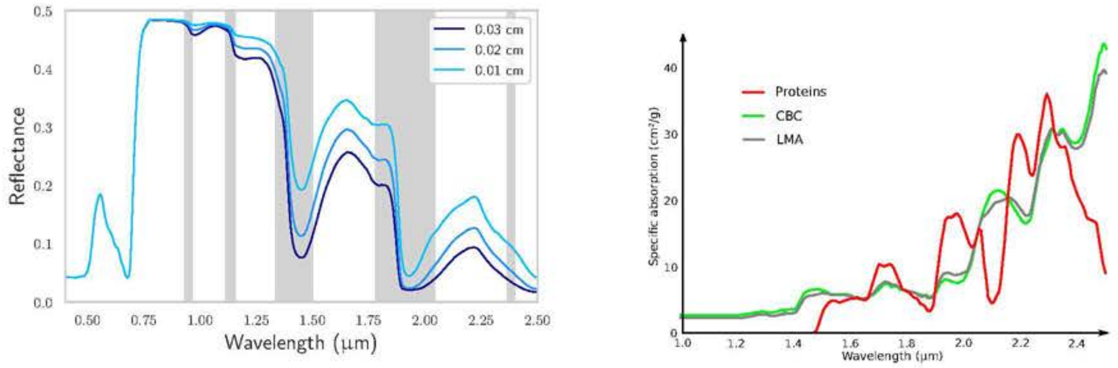


Figure 1.4: (a) picture of maple leaves sampled in summer and autumn, as well as variations of (b) chlorophylls and (c) carotenoids contents, (d) carotenoids-chlorophylls ratio and (e) anthocyanins content. Adapted from Junker and Ensminger, 2016.



(a) Leaf reflectance spectra at various equivalent water thicknesses as simulated by PROSPECT. The grey areas correspond to atmospheric water vapor absorption bands.

(b) Specific absorption coefficients of proteins, carbon-based constituents (CBC) and LMA. Adapted from Féret et al., 2021.

Figure 1.5: Leaf water and dry matter content's influence on spectral properties.

also present a lower area than usual, in part due to reduced turgor pressure (Tribulato et al., 2019). A common method in remote sensing to estimate vegetation water status is to estimate the EWT of the leaves, that is the ratio of the mass of water contained by the leaf to its area.

LMA is defined as the ratio of the leaf dry mass to its surface area. LMA has been linked to plant physiological processes, with leaves exhibiting different LMA values depending on the available sunlight. Most of LMA variations are due to differences in leaf anatomy and biochemical composition, which can vary greatly between plant species depending on their ecological strategies: for instance, slow growing plants that present a high LMA have thicker cell walls and more carbon and lignin (Arendonk and Poorter, 1994; Westoby et al., 2004). Overall, LMA can provide information on the vegetation photosynthetic capacity (Pierce, Running, and Walker, 1994). Leaf dry matter is a sum of proteins, cellulose, lignin, sugars and several other components that mostly affect the infrared region. The *Spectral Absorption Coefficient* (SAC) of LMA, carbon compounds and proteins are shown in Figure 1.5b: as visible, the SAC of LMA is very similar to the one of the structural carbohydrates, highlighting their more important contribution to the radiative behaviour compared to the proteins'.

Leaf spectral properties depend on their biochemistry and overall anatomy. However, additional parameters drive reflectance variations at the canopy level, as the entire canopy structure will affect the reflected radiation.

1.2.3 Tree and Canopy Structural Characteristics

The canopy fraction covered by a sensor's pixel constitutes hundreds to thousands of leaves distributed and orientated in specific manners, according to available sunlight, plant species and environmental conditions, all of which influence the canopy's spectral response. The probability that a ray of light does not hit vegetation elements when passing through the canopy is defined as the canopy gap fraction. This probability can be obtained from hemispherical photographies or specific optical instruments such as LAI-2000 (Welles and Cohen, 1996) or *Terrestrial Laser Scanner* (TLS) (Danson et al., 2007). It depends one multiple parameters, such as leaf orientation and distribution, and is frequently used to derive canopy traits such as LAI (Martens, Ustin, and Rousseau, 1993; Nilson and Kuusk, 2004). Overall, gap fraction controls the ratio of soil to canopy in a sensor's pixel and therefore the quantity of "useful" remote sensing vegetation signal, as radiation reflected by the understory is mixed with the soil reflectance.

The *Leaf Angle Distribution* (LAD) within crowns significantly affects light interception and scattering within the canopy (Wang, Li, and Su, 2007). It represents the probability for each leaf to have a specific inclination. Indeed, this distribution is not random and can belong to different main types of distributions (Figure 1.6). However, the spatial and temporal variability of LAD even within one species is high and difficult to determine, and can even vary depending on the canopy layer considered. Due to insufficient knowledge, it is often assumed to be spherical for temperate broadleaved species (Barclay, 2001; Pisek et al., 2013).

Light interception by the canopy is also determined by the LAI, the sum of upper surfaces

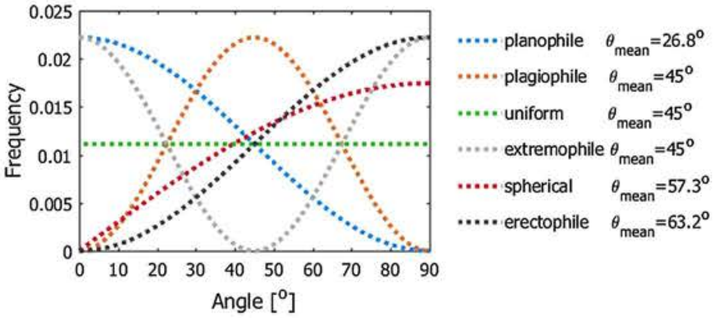


Figure 1.6: Mathematical functions used to approximate LAD and the corresponding average leaf inclination. From Liu et al., 2019.

of the leaves comprises within an area divided by the ground surface area the leaves are contained within. This index can naturally vary according to the seasons as deciduous species lose their leaves during their dormancy (see Figure 1.7). It is of prime interest to monitor carbon, water and energy exchanges between vegetation and the atmosphere (Misson et al., 2007). Depending on the LAI of a canopy, the ratio of overstory to understory visible to an airborne or satellite sensor will change, possibly making the estimation of leaves and canopy biochemistry more complex.

However, estimating LAI can be quite difficult, as non-destructive methods often rely on optical sensors. Attention should be paid to adjust for the clumping: at the crown scale, a single leaf could hide several others from the sensors, while at the canopy scale, depending on the *Canopy Cover* (CC), a crown could be partially hidden by another (Ryu et al., 2010, 2012). Moreover, it is in general hard to differentiate between leaves and woody elements, and their respective contributions often only allow for the estimation of *Plant Area Index* (PAI). Therefore, conversion factors from PAI to LAI often have to be considered to obtain a LAI estimate (Woodgate et al., 2016).

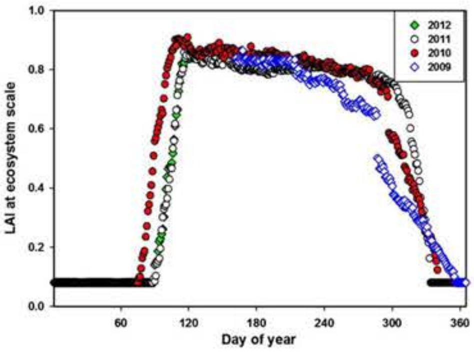


Figure 1.7: Time series of the LAI of a woodland savanna. From Ryu et al., 2012.

Non-photosynthetic vegetation in the overstory, such a woody elements, also influences canopy reflectance. Indeed, leaf distribution within a tree crown is highly heterogeneous (Jack and Long, 1992), so that parts of a crown can present a large proportion of woody elements and lead to important light scattering. For instance, in a young spruce forest, as much as 13%

of the canopy surface consisted of woody biomass (Malenovský et al., 2008a): consequently, the relative contribution of leaf biochemicals to the reflected radiation is reduced, which can limit estimates accuracy if not properly taken into account (Verrelst et al., 2010).

This section showed that the structural and biochemical properties of both the photosynthetic and non-photosynthetic constituents of a canopy drive its spectral properties in the visible and infrared regions. This allowed the development of physical models to simulate the interaction vegetation with electromagnetic radiation. An overview of the available RTM is given in the next section.

1.3 Radiative Transfer Models

Requiring some assumptions and simplifications regarding the environments they intend to simulate, several RTM describing light absorption and scattering due to vegetation have been developed and successfully validated at different scales.

1.3.1 At the Leaf Level

The PROSPECT (Jacquemoud and Baret, 1990) model is the most widely used leaf RTM. It simulates leaf reflectance and transmittance over the 0.4–2.5 µm spectral range. PROSPECT hypothesizes that leaves can be modeled as a stack of N plates whose optical properties are a mixture of those of leaves' biochemical components (Figure 1.8). While the plate model is not well adapted for conifer needles, PROSPECT has also been shown to yield more accurate estimates than the conifer leaf model LIBERTY (Dawson, Curran, and Plummer, 1998) provided the use of a normalization factor to account for the different leaf structure (Moorthy, Miller, and Noland, 2008).

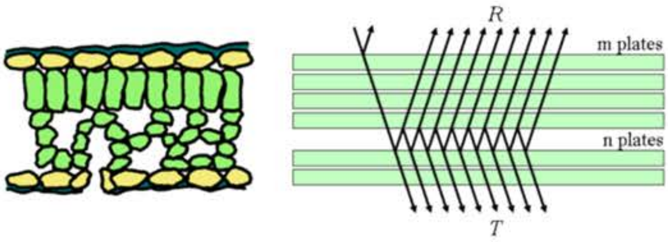


Figure 1.8: Schematic representation of a dicotyledon leaf (left) and the modelling within PROSPECT with a set of N = m + n plates (right). R: reflectance; T: transmittance. From `http://photobiology.info/Jacq_Ustin.html`.

PROSPECT originally only took into account *Chlorophyll Content* (C_{ab}), EWT and the leaf structural paramter N. Since then, several new versions have been developed, introducing *Carotenoid Content* (Car), LMA (Feret et al., 2008), fluorescence, and anthocyanins (Féret et al., 2017). More recently, LMA was even refined so as to distinguish proteins from other carbon based components (Féret et al., 2021). PROSPECT has been extensively validated and is the reference leaf RTM and a cornerstone of all canopy RTM.

1.3.2 At the Canopy Level

Several canopy RTM have been developed over the years, with various degrees of abstraction concerning canopy representation. Model intercomparison exercises, such as *RAdiation transfer Model Intercomparison* (RAMI), are regularly held to compare outputs (Pinty et al., 2004; Pinty et al., 2001; Widlowski et al., 2013; Widlowski et al., 2007, 2015). Two main families of canopy RTM exist: homogeneous and heterogeneous.

1.3.2.1 Homogeneous Radiative Transfer Models

Homogeneous, also designated as 1D, RTM approximate the canopy as an unlimited, horizontal turbid medium whose transmittance mostly depends on LAD and LAI following the Beer-Lambert law. The spectral signatures of leaves are determined by a leaf RTM such as PROSPECT. Thanks to these approximations, 1D RTM directly solve the radiative transfer equations to obtain the optical properties of a scene and are therefore fast in their execution (Kuusk, 2017).

The most popular canopy RTM is PROSAIL, resulting from the coupling of the SAIL (Verhoef, 1984) canopy RTM with PROSPECT (Figure 1.9). It has been extensively used due to its relative simplicity and accuracy (Berger et al., 2018; Jacquemoud et al., 2009), either in forward mode, to design vegetation indices (Delegido et al., 2014; Haboudane et al., 2002; Maire et al., 2008), or in inverse mode, to extract variables of interest from an input reflectance spectrum (Colombo et al., 2008; Darvishzadeh, Matkan, and Dashti Ahangar, 2012; Li et al., 2008; Weiss et al., 2000).

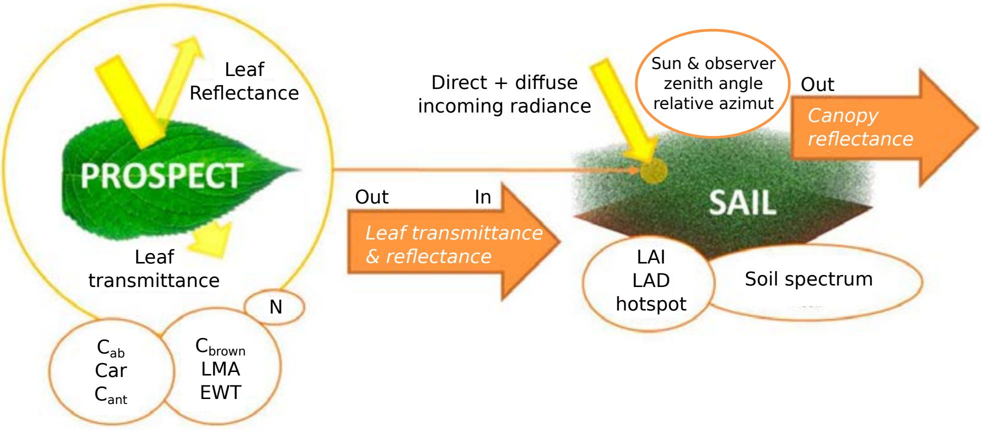


Figure 1.9: Scheme of the calculation of canopy reflectance with PROSAIL. Adapted from Berger et al., 2018.

1.3.2.2 Heterogeneous Radiative Transfer Models

Heterogeneous models, also designated as 3D models, are much more complex (Figure 1.10). The level of detail and complexity varies greatly between 3D RTM: the INFORM model (Atzberger, 2000) does not determine crown shape, and describes the canopy based on the probability that the sensor sees a tree crown or the understory; FLIGHT (North, 1996) describes the geometry of trunks and crowns and builds them according to a growth model, while foliage parameters are considered homogeneous within a crown; arbitrarily complex scenes can be modeled in raytran (Govaerts and Verstraete, 1998).

The third phase of the RAMI experiments found that a subset of 3D RTM consistently showed good agreement concerning the optical properties of various vegetation canopies. As a testament to their quality, their outputs are now used to serve as "surrogate truths" for the evaluation of other models (Widlowski et al., 2007). These models are: DART (Gastellu-Etchegorry et al., 2015), drat (Lewis, 1999), FLIGHT, rayspread (Graziani et al., 2006) and raytran.

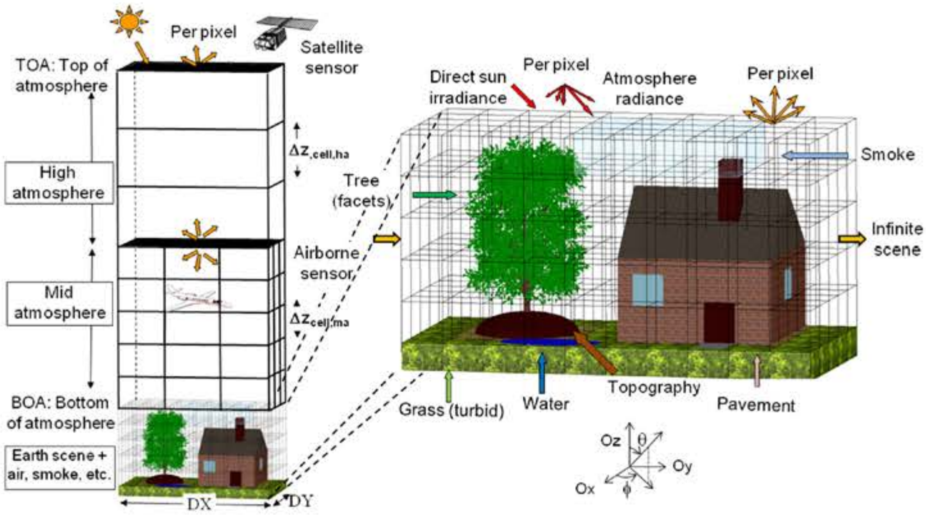


Figure 1.10: Representation of the Earth/atmosphere system in DART. Land surface elements can be modeled as facets or turbid cells. From Gastellu-Etchegorry et al., 2015.

These models have a high complexity, and expert knowledge is required to prepare input data and correctly set up the simulations. Improving their performances through more accurate or adapted modeling, by refining the representation of the understory (Melendo-Vega et al., 2018), of the tree crown (Widlowski, Côté, and Béland, 2014), or the overall representation (Janoutová et al., 2019), is the subject of constant research. Moreover, they are computationally demanding, which can further restrict their use. Still, they have been extensively employed to better understand the drivers of canopy optical properties (Malenovský et al., 2008b; Verrelst et al., 2010), the carbon and energy fluxes of heterogeneous woodlands (Kobayashi et al., 2012) or the texture of forest images (Bruniquel-Pinel and Gastellu-Etchegorry, 1998). Similar to 1D RTM, they are of course used to retrieve canopy biochemical contents from remote sensing imagery (Darvishzadeh et al., 2019; Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012; Malenovsky et al., 2007; Zarco-Tejada et al., 2013).

This section presented possible RTM developed for the estimation of leaf and canopy biochemistry, based on physical processes and modeling hypotheses. However, estimation methods are numerous and do not all rely on these models.

1.4 Methods for the Estimation of Vegetation Traits from Remote Sensing

Few retrieval methods from remote sensing imagery have been incorporated into operation processing chains. Notably, ARTMO (Verrelst, Romijn, and Kooistra, 2012) has been used to process CHRIS/PROBA (Verrelst, Romijn, and Kooistra, 2012), Sentinel-2 (Verrelst et al., 2015a) or HyMAP data, and various toolboxes (Delegido et al., 2015; Rivera-Caicedo et al., 2017; Verrelst et al., 2014) are available that easily implement retrieval methods. However, validation has only been done over a reduced number of ecosystems and sensors, and most methods are still in an experimental state due to the limited scope of study during their development. Four main retrieval method groups can be identified: parametric and non-parametric empirical-statistical methods, that fit functions or regression models between the estimates and leaf or canopy properties and relying on field data to do so; physical-based methods, making use of the RTM by comparing simulated and measured spectra to obtain the estimations; hybrid methods, using the outputs of RTM in direct mode to train empirical-statistical models (Verrelst et al., 2015b).

1.4.1 Empirical-Statistical Methods

Empirical-statistical methods were the first retrieval methods to be used to retrieve information from multispectral or hyperspectral images.

1.4.1.1 Parametric Methods

They are the oldest members of this group. They consider an explicit relationship between an index and the variable of interest. The most common approach is to obtain a relationship between a VI, a function of reflectance values at specific wavelengths, and the parameter to be estimated (an illustration of this process is given in Figure 1.11). While VI have been extensively studied because of (i) their simplicity and (ii) the historically strong prevalence of multispectral data, they are also widely used to process hyperspectral data. The most famous VI is the NDVI (Tucker, 1979), used to estimate LAI and C_{ab}, however a multitude of VI have been developed over the years, for instance to correct for soil influence (Haboudane et al., 2002, 2004; Huete, 1988), estimate Car (Gitelson et al., 2002), vegetation water content (Gao, 1996; Hunt and Rock, 1989) or LMA (Maire et al., 2008). Concerning open canopies, Zarco-Tejada et al., 2019 showed that retrieval of needle C_{ab} of a conifer forest could be retrieved from

Sentinel-2 imagery using a VI. The VI calibration step involved the reflectance of pure tree crowns was obtained using high-resolution hyperspectral data and field-measured C_{ab}.

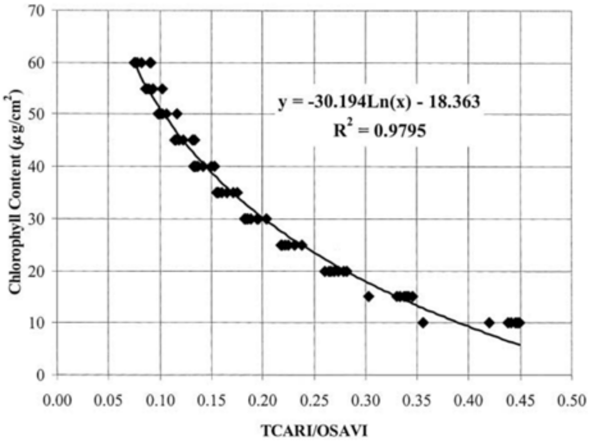


Figure 1.11: Predictive function for C_{ab} with regards to the TCARI/OSAVI vegetation index. From Haboudane et al., 2002.

However, VI only use a very limited set of spectral bands and do not take advantage of all the spectral information, which can severely limit their transportability to other sites and sensors: for instance, when retrieving needle carotenoids content over a conifer forest at high resolution, Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012 assessed the performances of VI's used in past studies and found they were considerably affected by the site's structural specificities, leading to the development of a new VI. Indexes relying on high spectral densities have also been developed, either through integration (Delegido et al., 2010; Malenovský et al., 2006) or derivation (Le Maire, François, and Dufrêne, 2004) of the reflectance spectra, although these indices still focused on specific subsets in the visible and infrared.

1.4.1.2 Non-Parametric Methods

These methods fit non-explicit models to the available data to obtain estimators. They can take advantage of the full reflectance spectrum in the training phase, as they will themselves assign weights to the various spectral bands. However, this can be problematic when working with hyperspectral data: Axelsson et al., 2013 showed that the colinearity of close spectral bands had a detrimental effect on the performances of *Support Vector Regression* (SVR). Spectral preprocessing can therefore be required to achieve acceptable performances. Still, non-parametric methods have had good success for vegetation traits estimation. Mutanga, Adam, and Cho, 2012 used *Random Forest Regression* (RFR) to estimate biomass in wetlands, while Siegmann and Jarmer, 2015 found that *Partial Least-Square Regression* (PLSR) and RFR both provided consistent results when estimating crop LAI. Overall, Asner et al., 2011a demonstrated the potential of PLSR for the estimation of canopy biochemicals at the global scale, and PLSR was subsequently used to estimate foliar traits from high-resolution imagery over tropical forests (Asner et al., 2015; Dana Chadwick and Asner, 2016).

Study	No. and type of data
Dana Chadwick and Asner, 2016	leaf collection from 254 trees
Asner et al., 2011b	leaf collection from 2873 trees
Mutanga, Adam, and Cho, 2012	biomass measurements from 82 400 m^2 plots
Darvishzadeh et al., 2011	LAI measurements from 41×4 (or 5) 1 m^2 plots
Wang, Huang, and Lou, 2011	LAI measurements from 123 plots
Roelofsen et al., 2014	various measurements from 379 plots

Table 1.1: Some studies based on non-parametric methods and their associated databases from field campaigns.

Non-parametric methods have two major downsides, which are (i) the expert knowledge required to adjust hyperparameters to optimize the model while avoiding overfitting and (ii) the large quantity of field measurements that must be undertaken to build the training and validation databases (see Table 1.1 for some examples). While the former is not a major issue, the latter is more problematic as obtaining field measurements in every ecosystem can not be expected: generalizing such a model over different ecosystems could be impossible.

1.4.2 Physical-Based Methods

Physical-based methods rely on the RTM presented in Section 1.3. Contrary to the previous methods, these can provide estimates with little to no knowledge of the imaged canopy. These methods rely on the inversion of the RTM, making the assumption that it is representative of the canopy. To do so, they either iteratively try to match a target spectrum by adjusting the RTM inputs, or directly select the best matches from large LUT previously generated. However, this inversion problem is ill-posed: different combinations of inputs may lead to similar outputs and there remains modeling uncertainties. Therefore, the solution is often not unique. Nevertheless, mitigation methods have been developed, such as adding noise in the generated data (Richter et al., 2011), selecting multiple "best" solutions (Weiss et al., 2000), or constraining the model using a priori knowledge (Atzberger and Richter, 2012; Combal et al., 2003). The vast majority of physical-based methods rely on LUT, as iterative optimization is often unable to deal with local minima.

Gascon et al., 2004 developed a LUT-based inversion method to estimate the LAI of a temperate broad-leaved forest from SPOT and Ikonos images. To generate the LUT, they used the DART model and modeled the canopy with a simple representation, with a repetition of identical trees with ellipsoidal crowns. Comparing forest-level LAI estimates on two different dates, they showed that it was possible to identify a LAI decrease due to a storm. However, the estimation method could not consider spatial variations of other vegetation traits and could not give accurate information over specific stands.

Darvishzadeh, Matkan, and Dashti Ahangar, 2012 demonstrated that LUT generated with the PROSAIL model could be used to retrieve canopy C_{ab} of rice crops from 10 m spatial resolution multispectral images containing 4 spectral bands. They averaged the C_{ab} values of the 100 entries that had the lowest *Root-Mean-Square Error* (RMSE) with measured spectra from a 100,000-entry LUT in order to realize their estimates, and showed that acceptable results could be obtained despite the limited band number.

Working with high-resolution hyperspectral imagery, Ferreira et al., 2018 simulated trees in the DART model with a simple tree architecture, defined by a spherical crown consisting of a mixed turbid medium of foliage and woody elements. They demonstrated that this simple representation was sufficient to acceptably simulate the reflectance of individual trees, and that DART outputs could subsequently be used for VI calibration. However, they did compare any estimate with data acquired *in situ* to validate the method in practice.

More recently, Darvishzadeh et al., 2019 investigated the use of a LUT-based inversion method to retrieve needle C_{ab} of spruce stands in a temperate forest. A 500,000-entry LUT was generated randomly using a random distribution for the model input parameters describing the canopy. They showed the potential of using the INFORM model with multispectral satellite data for mapping needle C_{ab} in temperate forests. However, their study highlighted the limitations of the modeling used in INFORM to properly simulate reflectance bands in the *Near InfraRed* (NIR) (see Figure 1.12), which the study had shown to be important for C_{ab} retrieval.

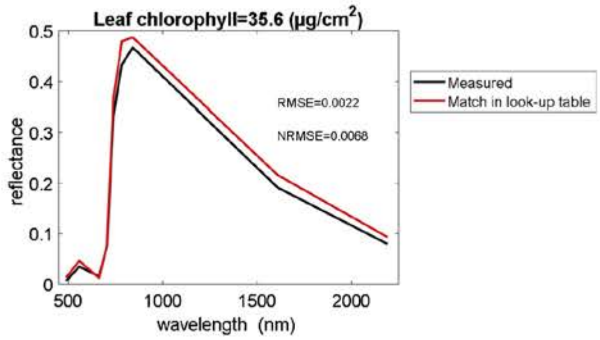


Figure 1.12: Reflectance spectrum from Sentinel-2 and the associated spectrum of the "best match" from the INFORM model. From Darvishzadeh et al., 2019.

Zarco-Tejada et al., 2019 also employed a LUT-based inversion method to map needle C_{ab} of a conifer forest from Sentinel-2 imagery with the INFORM model. They demonstrated that, provided sufficient knowledge about the structure of the sites was available, such an estimation was possible during both summer and winter seasons. This study also highlighted the retrieval difficulties arising when working with winter images, were grass was active and shadows more present due to a lower solar angle, possibly requiring finer tuning of the RTM.

Methods based on RTM inversion always rely on the minimizing of a cost-function between the RTM outputs and the measured reflectance spectra that can take several forms, from VI or RMSE to more complex ones based on information measure or M-estimates (Verrelst et al., 2012). Choosing a suitable cost function for a specific estimation requires *a priori* knowledge, as different ecosystems and sensors will not show the same spectral response.

1.4.3 Hybrid Methods

Hybrid methods merge physical-based and empirical statistical methods, by replacing the field measurements necessary to generate the training datasets used by non-parametric empirical methods with the outputs of a RTM. Therefore, the estimation procedure from physical-based models could benefit from the computational efficiency of regression methods and their ability to take advantage of the full reflectance spectrum. While hybrid methods originally mostly relied on *Artificial Neural Network* (ANN), other non-parametric methods started to be used successfully in the 2000s to estimate various vegetation traits.

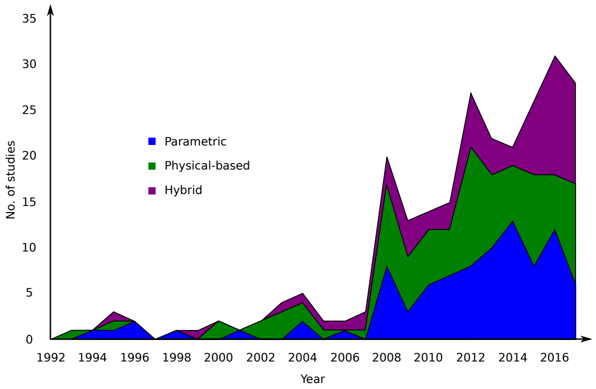


Figure 1.13: Temporal development of variable retrieval methods involving PROSAIL. "Parametric" refers to parametric methods calibrated on the outputs of the RTM. Adapted from Berger et al., 2018.

Trombetti et al., 2008 demonstrated that ANN could be trained with a multitude of VI computed from the outputs of PROSAIL to estimate canopy water content from MODIS images over the US, leading ecoregion-scale temporal estimates. While simulating vegetation quite roughly due to the use of a 1D RTM, these estimations showed the potential of hybrid methods to obtain large-scale estimates.

Malenovský et al., 2013 retrieved needle C_{ab} from sub-meter resolution hyperspectral images using an ANN trained over a dataset generated with DART. The ANN was shown to outperform all VI, even obtaining slightly more accurate estimates than one specifically designed for the study.

Focusing on the retrieval of LAI and forest C_{ab} from the images of a hyperspectral airborne sensor, Homolová, Janoutová, and Malenovský, 2016 fit a SVR on the outputs of the DART model. They found better performances when training the SVR with preprocessed (continuum-removal) subsets of the full reflectance spectra than when training over the full, unprocessed spectra. Their main conclusion was that SVR was faster and more robust than ANN, and therefore a good alternative.

More recently, Ali et al., 2020 compared the performances of LUT-based and various hybrid methods (ANN, *Gaussian Process Regression* (GPR), RFR, *Stepwise MultiLinear Regression* (SMLR), SVR) for the retrieval of C_{ab} and LAI of mixed forest stands from Sentinel-2 imagery.

They showed that all methods provided outputs statistically similar to *in situ* measurements for C_{ab}, while ANN, GPR, RFR and LUT-based inversion did the same for LAI. Overall, hybrid methods led to estimation accuracies similar to LUT-based inversion at a fraction of the time cost, highlighting the potential of these methods for leaf trait mapping from satellite imagery over large swathes.

While hybrid methods still suffer from the need to appropriately configure the RTM, the past decade saw a considerable increase in their use for the estimation of vegetation traits (see Figure 1.13 for an example concerning PROSAIL). As they are more suited to process high-dimensional data, they may prove to be the most adequate methods for future operational retrievals (Berger et al., 2018).

1.5 Summary & Selected Approach

From the state of the art, it appears that several vegetation traits related to tree and canopy stress status have a direct influence on vegetation reflectance spectrum both in the visible and infrared regions and can therefore be estimated by remote sensing.

Various methods have been developed to retrieve these traits from airborne and satellite imagery, relying either on empirical data collected in the field or on synthetic data from the outputs of a RTM. Empirical methods have been shown to provide good results, in particular when applied over images acquired via airborne sensors. On the other hand, physically-based and hybrid methods are being employed with more and more success. Indeed, as illustrated by the RAMI experiments, several RTM with consistent, realistic results are available and can be used to generate surrogate training data. These retrieval methods are therefore ideal candidates for operational purposes at the global scale, as they would not require extensive and expensive field campaigns.

Recent studies have shown that with the advent of the Sentinel-2 missions, global monitoring of the vegetation is possible and that some vegetation traits can be retrieved from satellite multispectral images using RTM-based methods. Methods taking advantage of hyperspectral data showed improved results for the estimation of a multitude of vegetation traits compared to those using hyperspectral images: Thenkabail, Smith, and De Pauw, 2002 have shown that LAI and wet biomass variations were better captured by narrowband than broadband VI, and more recently Marshall and Thenkabail, 2015 demonstrated that the advantage of a higher spectral resolution outweighed those of the higher spatial resolution associated with multispectral satellite sensors. However, studies using hyperspectral data have mostly been used with airborne sensors presenting a high spatial resolution. Moreover, they have mostly focused on closed canopies (the works of Zarco-Tejada et al., 2019 – concerning multispectral imagery – was published during the present thesis) and for a very limited number of traits: LAI and C_{ab} retrievals are the vast majority.

As the present thesis aims at developing methods for the estimation of vegetation traits with limited field knowledge in the context of hyperspectral satellite missions, it was necessary to focus on RTM-based estimation methods. Although multiple RTM were available, the highly heterogeneous nature of open canopies made 1D RTM a poor choice compared to the more complex 3D RTM. It was decided to use the DART model, which has been successfully used in previous studies and is among the six reference RTM identified during the RAMI experiments.

Chapter 2

Materials

Contents

2.1 Study Sites

The two open canopies that were considered in the present thesis are located in the lower foothills of the Sierra Nevada (Tonzi Ranch (TZ), latitude: 38.5 °N; longitude: 121.0 °W, San Joaquin Experimental Range (SJER), latitude: 37.1 °N; longitude: 119.7 °W). Locations and aerial views of the sites are given in Figure 2.1.

They are categorized as grass-oak-pine woodlands and both have a Mediterranean climate with hot, dry summers and mild, wet winters. The canopy of TZ is dominated by blue oaks (*Quercus douglasii* – QUDO) with a small population of grey pine (*Pinus sabiniana* – PISA), while the one of SJER mostly consists in QUDO, interior live oaks (*Quercus wislizeni* – QUWI) and PISA. The period of activity of QUDO starts in April and ends around November, while QUWI and PISA are evergreen species. Picture of both QUDO and PISA as well as average dimensions of QUDO are shown in Figures 2.2 and 2.1.

The understory of both sites is composed of annual grass species active from December to May and dry during the summer period. For TZ, the average LAI is 0.8 m^2/m^2 and the mean canopy cover 47% (Kobayashi et al., 2013), with mean annual tempatures and precipitations of 16.5 °C and 562 mm, respectively, and a soil classified as Auburn very rocky silt loam soil. For

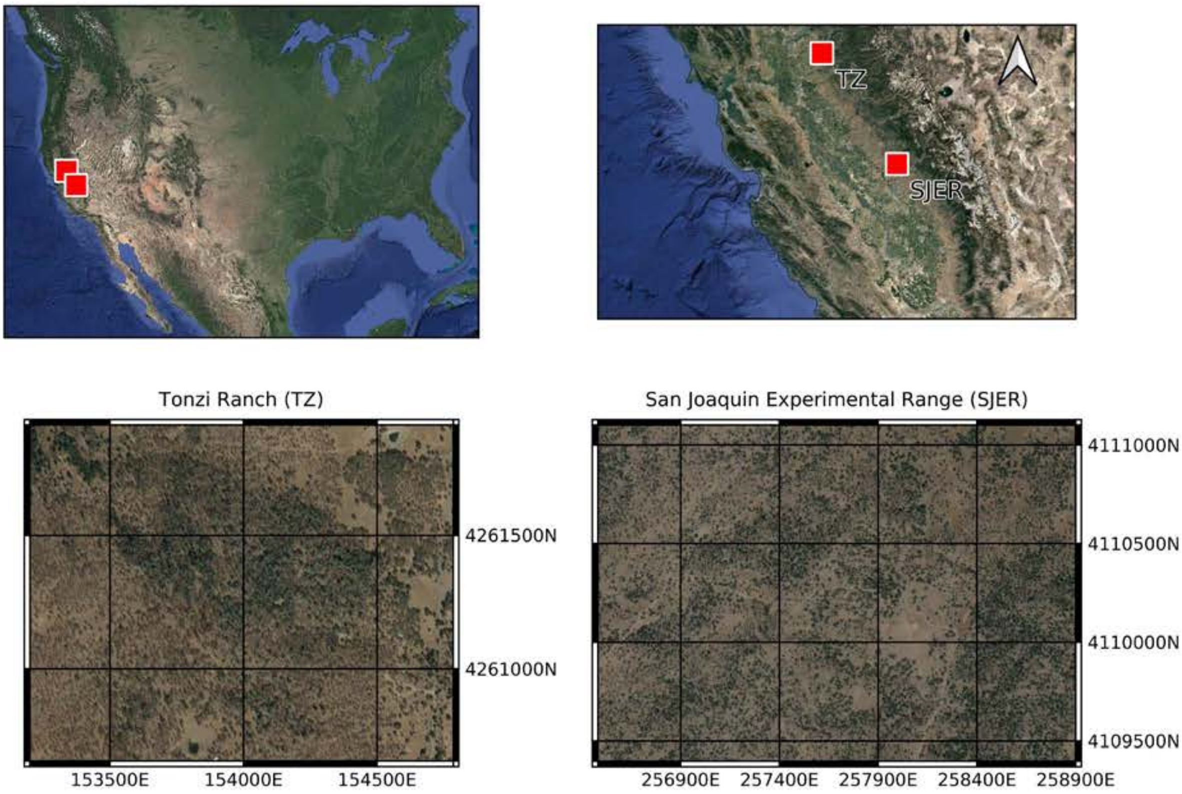


Figure 2.1: Location and aerial views of the two study sites: TZ and SJER.

Table 2.1: Crown characteristics of the QUDO and QUWI trees obtained from a field survey over the two sites. In parenthesis, the number of tree measurements used to compute the statistics.

Site	Crown height (m)	Total height (m)	Crown diameter (m)
TZ	10.4 (14)	14.5 (14)	8.2 (14)
SJER	7 (160)	8.6 (162)	10.1 (16)

SJER, the canopy cover is about 30%, with average temperatures and annual precipitations of 16.5 °C and 485 mm, respectively. The soil type at SJER is classified as Vista rocky coarse sandy loam (Tate et al., 2004).

Figure 2.2: A picture of both a Blue Oak (foreground) and a Grey Pine (background) at Tonzi Ranch, in summer.

2.2 Airborne Hyperspectral Remote Sensing Data

AVIRIS Classic (AVIRIS-C) hyperspectral data were processed and delivered by NASA Jet Propulsion Laboratory (JPL; `http://aviris.jpl.nasa.gov`). The sensor is a whiskbroom scanner with an instantaneous field of view of 1 mrad and a total field of view of 34°. It is composed of 224 contiguous bands with a *Full Width at Half Maximum* (FWHM) of 0.010 µm from 0.365 to 2.5 µm (Green et al., 1998). Airborne acquisitions took place in spring, summer, and fall (corresponding to the 3 phenological stages) from 2013 to 2015 and every summer from 2016 to 2018. Images were acquired at nadir. The ER-2 aircraft flew 20 km above the ground, around noon, in order to avoid spectral directional effects. Preprocessing steps provided by NASA JPL included radiometric calibration, geometrical orthorectification, nearest neighbor spatial resampling at 18 m, and atmospherical correction performed with ATREM (Bo-Cai Gao and Goetz, 1990), in order to retrieve surface reflectance. Additional flights over the study site were done during the summer 2014 using the *AVIRIS Next-Generation* (AVIRIS-NG) sensor. AVIRIS-NG images were acquired with a 2 m spatial resolution and 432 spectral bands, with a FWHM of 0.005 µm from 0.380 to 2.510 µm.

2.3 Field Data

Field data were collected coincidentally with NASA Hyperspectral Infrared Imager (HyspIRI) Mission Study Airborne Campaigns. They took place in fall 2013 and summers 2014 and 2016 (`https://hyspiri.jpl.nasa.gov/`). They consisted in *Digital Hemispherical Photography* (DHP), leaf collection, reflectance measurements and TLS acquisitions.

2.3.1 LAI and Gap Fraction

Several DHP were collected over multiple 60 m × 60 m plots across the study site covering all vegetation cover fraction and species composition, according to the sampling pattern shown in Figure 2.3. From each 20 m × 20 m subplot, three DHP were taken from a Nikon Coolpix 4300 camera post sunset when no direct sunlight was visible.

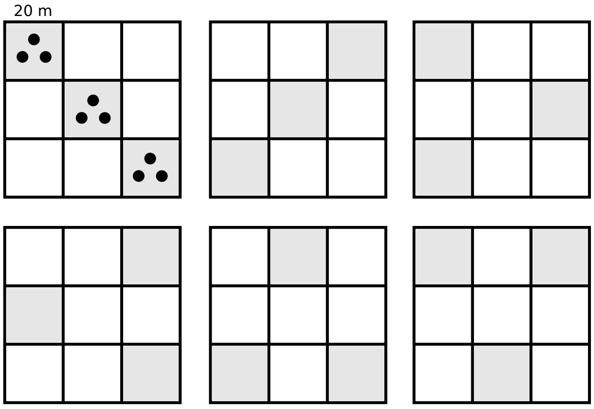


Figure 2.3: Possible sampling patterns of the gap fraction and LAI plots. For each plot, three DHP (black dots) were taken in a North–South East–Southwest pattern 10 m apart in each of the three randomly selected (shaded in gray).

In Chapter 3, it was decided to focus on LAI retrieval. The LAI of each subplot was obtained from the DHP using the Hemiview software (Rich et al., 1999). Hemiview calculates PAI from the fisheye images estimating image gap fraction using eight azimuth and seven zenith angles. The threshold was estimated for each image based on a visual interpretation. Hemiview calculates an effective PAI that has to be corrected by clumping factors. However, some studies have considered that in forests, effective PAI is almost identical to true LAI, as the underestimation of LAI due to clumping effects could be compensated by the overestimation due to woody structures (Fang et al., 2019). In Chapter 3, effective PAI estimated through Hemiview is therefore considered equal to true LAI and simply referred to as LAI. There was one LAI value per 20 m × 20 m subplot, obtained by averaging the LAI derived from the 3 DHP of the subplot.

In Chapters 4 and 6, it was decided to estimate gap fraction instead of LAI in order to avoid the uncertainties inherent to the assumption done in Chapter 3. From each 60 m × 60 m plot, the nine DHP were used and processed using CAN-EYE (`https://www6.paca.inrae.fr/can-eye`). CAN-EYE calculated the Gap Fraction with azimutal and zenithal resolutions of 2.5°, using a circle of interest of 65°. The theory behind CAN-EYE estimations is described by Weiss et al., 2004 and is also available at `https://www6.paca.inrae.fr/can-eye/Documentation/Documentation`. As the LAI of the site is very low, there is no significant risk of Gap Fraction saturation as this phenomenon happens for LAI higher than 5. There was one gap fraction value per 60 m × 60 m subplot.

2.3.2 Leaf Biochemistry

Sets of leaves from five QUDO and QUWI individuals were collected from the upper, sunlit portion of the canopy on both study sites. Sampling started within an hour of the timing of the overflight. Leaf samples were collected from open grown trees that were in full sunlight, as high into the canopy as possible to reach, and from branches on the east and west side of the tree. Attention was paid to ensure that collected leaves were healthy, and collection always occurred during dry days. Leaves were placed in foil packets so that light would not reach them and stored in a container with blue ice for cooling and transferred to a lab refrigerator at the field station until lab measurements could be made, no later than 48 hours after data collection. Leaf samples were frozen in liquid nitrogen for a short time until lyophilized.

The analysis extraction was done in 90% acetone (14 mL) for 48 hours, which provided enough solvent extract to make 3.5 mL samples for each individual. Lambda 25 UV/Vis spectrophotometer runs were done with standards with a concentration of 23 mg/mL for chlorophylls and 8 mg/mL for carotenoids. Absorbance A of the solution was measured at 0.470, 0.662, 0.645, and 0.710 μm. Chlorophylls and carotenoids concentrations of the solution (c_a, c_b, $c_{(x+c)}$) were computed using Equations (2.1)–(2.3) (Lichtenthaler, 1987; Lichtenthaler and Buschmann, 2001) with a dilution factor equal to 3.5, and further processed to get chlorophylls a + b and carotenoid concentration on the leaf area (μm/cm^2).

While leaves from fives trees were originally collected from summer 2013, biochemistry results from two trees were rejected as they showed lower C_{ab} and Car values than those from the same trees in fall 2013 (20 to 37 μg/cm^2 and 34 to 38 μg/cm^2 from summer to fall, respectively). This is contrary to the expected behaviour of these pigments, and it was assumed that the leaf samples from those trees suffered degradation between collection and laboratory analysis. They were flagged as outliers and not considered further in the thesis. Anthocyanins were not measured during the field campaigns and the leaf anthocyanin content was assumed to be either null or between 0 and 2 μg/cm^2 over the course of the thesis.

$$c_a(\mu\text{g/mL}) = (11.24\,(A_{0.662} - A_{0.710}) - 2.04\,(A_{0.645} - A_{0.710})) \times \text{dilution factor} \tag{2.1}$$

$$c_b(\mu\text{g/mL}) = (20.13\,(A_{0.645} - A_{0.710}) - 4.19\,(A_{0.662} - A_{0.710})) \times \text{dilution factor} \tag{2.2}$$

$$c_{(x+c)}(\mu\text{g/mL}) = \left(\frac{1000(A_{0.470} - A_{0.710}) - 1.90c_a - 63.140c_b}{214}\right) \times \text{dilution factor} \tag{2.3}$$

To retrieve EWT, LMA, plastic bags containing leaves were weighed with a mg precision scale before going to the field. In the lab, the bags with leaves inside were weighed, with the weight difference giving leaf fresh weight. After that, the thickness of each leaf was measured using a caliper and all leaves were scanned in TIF format with 150 dpi. Leaf area was estimated using the scanned image with the TOASTER software. Finally, all the leaves were put into a paper bag to dry at 65 degrees Celsius until the weight did not change when the leaves were reweighed (two to three days) to obtain leaf dry weight. Finally, EWT and LMA were calculated according to Equations 2.4 and 2.5.

$$\text{EWT} = \frac{\text{fresh weight} - \text{dry weight}}{\text{leaf area}} \tag{2.4}$$

$$\text{LMA} = \frac{\text{dry weight}}{\text{leaf area}} \tag{2.5}$$

2.3.3 Trunk Reflectances

Tree trunk reflectances were measured from trees that had been selected for leaf collection. Reflectance was measured over the 0.35–2.50 µm range with an Analytical Spectral Device (ASD; ASD Inc., Boulder, CO, USA) contact probe previously calibrated with a Spectralon panel before every acquisition. In order to do the measurements, pieces of bark from the trunks were removed and put on a horizontal surface.

2.3.4 Terrestrial Laser Scanning

Terrestrial lidar scans were acquired for a number of sites on each of the locations. The *Compact Biomass Lidar* (CBL) (Paynter et al., 2016), developed in-house at the University of Massachusetts Boston, was used to obtain 4 scans around each area. The lightweight, robust and rapidly scanning CBL utilizes a commercial 905 nm SICK LMS-151 lidar on a motorized rotary table, acquiring a point cloud of the 270 ° enveloping the instrument out to a maximum of 40 m, with a point cloud of first and second returns at a 0.25 ° resolution and with a 0.86 ° beam divergence.

2.4 Data overview

A timeline of the various data acquisitions that took place over both TZ and SJER is given in Figure 2.4, and Tables 2.2 and 2.3 give an overview of the number of leaf biochemistry, gap fraction, and LAI values that were available for this study and their statistics.

2.5 Radiative Transfer Code: DART

DART is a radiative transfer model able to simulate light interactions and multiple scattering effects within the modeling of a 3D scene, including the topography and the atmosphere. The

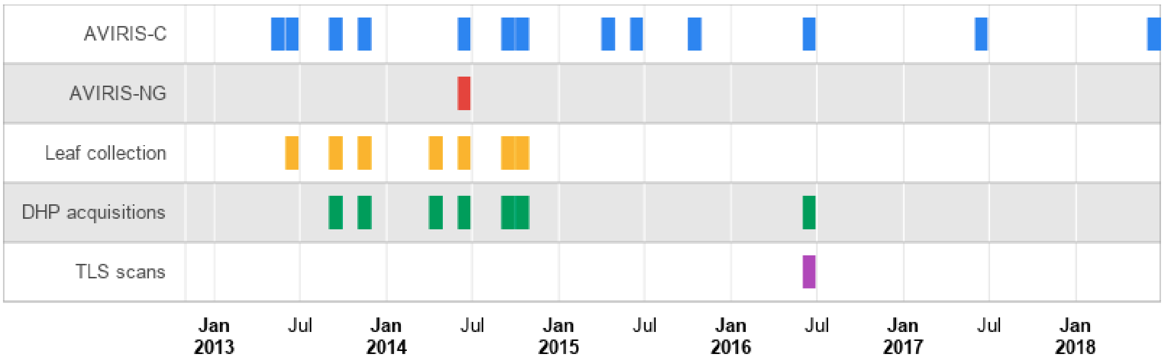


Figure 2.4: Timeline of the various data acquisitions available on TZ and SJER and used during this thesis.

Table 2.2: Statistics of the leaf biochemistry data acquired on TZ and SJER at the different dates.

	Sample	Cab (µg/cm^2)		Car (µg/cm^2)		EWT (g/cm^2)		LMA (g/cm^2)	
	No.	min.	max.	min.	max.	min.	max.	min.	max.
TZ									
summer 2013	3	37.9	52.6	8.9	12.6	0.009	0.013	0.008	0.011
fall 2013	5	25.2	38.62	6.9	10.3	0.011	0.013	0.008	0.014
spring 2014	4	31.2	40.2	7..5	9.2	0.010	0.012	0.011	0.012
summer 2014	5	40.66	49.9	9.2	11.0	0.011	0.014	0.009	0.012
fall 2014	5	07.2	34.5	7.3	6.5	0.011	0.014	0.008	0.009
SJER									
summer 2013	10	29.9	49.2	6.64	12.3	0.012	0.016	0.011	0.013
fall 2013	9	21.7	39.3	6.6	13.8	0.012	0.019	0.011	0.016
spring 2014	5	26.7	43.9	6.8	11.0	0.009	0.017	0.005	0.012
summer 2014	6	27.6	41.9	7.2	10.7	0.011	0.018	0.009	0.015
fall 2014	7	10.3	39.8	4.4	11.1	0.012	0.020	0.007	0.014
Total									
	59	10.3	52.6	4.4	13.8	0.009	0.02	0.005	0.016

Table 2.3: Statistics of the LAI and gap fraction data acquired on TZ and SJER at the different dates.

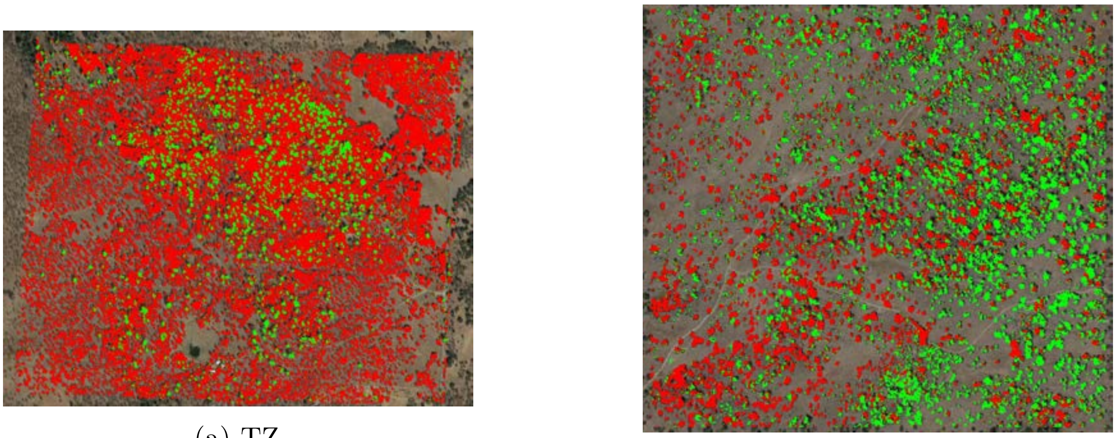
	Sample	LAI (m^2/m^2)		Gap fraction	
	No.	min.	max.	min.	max.
TZ					
fall 2013	12/2	0.36	0.48	0.40	0.63
summer 2014	19/7	0.34	1.9	0.21	0.86
fall 2016	21/7	0.37	1.6	0.33	0.90
SJER					
summer 2014	n.a./4			0.50	0.88
fall 2014	n.a./8			0.51	0.94
Total					
	52/28	0.34	1.9	0.21	0.94

3D scene is made of voxels whose size can be set by the user to simulate various scene elements. These elements can be represented through planar elements (triangles, allowing for precise modeling) and/or turbid voxels (statistical representation of fluids and vegetation). In the case of vegetation, these turbid voxel properties are determined by the canopy leaf area and its spatial distribution, the LAD, and leaf optical properties. A more precise description of the DART model can be found in Gastellu-Etchegorry et al., 1996 and Gastellu-Etchegorry et al., 2015. DART makes it possible to define trees with simplified representations such as turbid crowns of conic, spherical, or ellipsoidal shape and allows for the tuning of the proportion of voxels full of leaves within the crown as well as branches and twigs modeling. Therefore, trees are defined by structural parameters such as the shape and size of their crown and the distribution of their branches and leaf cells, as well as optical parameters with the optical properties of leaves and branches. The good performances of DART as a 3D RTM have been consistently demonstrated in several studies, notably during its participation in the RAMI experiments (Pinty et al., 2004; Pinty et al., 2001; Widlowski et al., 2013; Widlowski et al., 2007).

2.6 Classification of Conifers and Broad-Leaved Trees

Since this thesis only focused on broadleaved trees, a species classification was needed as a first step. A *Support Vector Machine* (SVM) method was used to classify the tree types (broad leaved and conifer) present on both sites in the AVIRIS-NG images. For TZ, 506 and 557 pixels were manually selected for the broad leaved and conifer classes, respectively, while for SJER 833 and 412 pixels were selected, based on field inventories, tree structural information derived from LiDAR data (crown height) collected in 2009 (PISA sample height assumed to be higher than 20 m for TZ, as per Kobayashi et al. (2012)) and visual inspection. Concerning SJER, more pixels were selected for broad-leaved than for conifers to account for both QUDO and

QUWI. The SVM was run with the radial basis function kernel (C: 6.5; γ: 0.0055). Over 20% of the data, used as test set, accuracy scores were 93% and 96% for TZ and SJER, respectively; it was decided not to optimize the hyperparameters further. The resulting classification maps (see Figure 2.5) were also visually compared to true-color aerial images of the site, as PISA have a visually distinct color from both QUDO and QUWI, in order to further confirm classification accuracy.



(a) TZ

(b) SJER

Figure 2.5: Classification maps: in red, broad-leaved trees; in green, conifers.

Chapter 3

Identifying an Estimation Method for the Retrieval of LAI and Foliar Pigment Contents of a Sparse Forest

---

Contents

3.1 Objectives

The main objective of the study presented in this chapter was to find an inversion strategy suited for the monitoring of a heterogeneous tree–grass ecosystem's health (e.g., resiliency to change, continuity in GPP, and nutrient capital (Trumbore, Brando, and Hartmann, 2015)), i.e., capable of estimating LAI and leaf C_{ab} and Car over time using multitemporal hyperspectral airborne acquisition with medium ground sampling distance.

Canopy LAI and biochemistry estimations using hyperspectral imagery have mostly been done over canopies with high LAI (for instance, see Banskota et al., 2015; Maire et al., 2008, Ali et al., 2016b; Darvishzadeh, Matkan, and Dashti Ahangar, 2012; Malenovský et al., 2008a). As presented in Chapter 1, various estimations have also specifically been done over open-canopy ecosystems (*e.g.* the works of Zarco-Tejada et al., 2004, Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012, Zarco-Tejada et al., 2019), and research concerning acceptable modeling methods within RTM for tree-grass ecosystems is still ongoing (Melendo-Vega et al., 2018; Widlowski, Côté, and Béland, 2014). The validity of a simplified representation of trees (*Simple Forest Representation* (SFR)) within the DART scene (such as the one done by Gascon et al., 2004) when working with medium-resolution hyperspectral images of very sparse forests was tested in this study, based on the work presented in Widlowski, Côté, and Béland, 2014. As acceptable representation of coniferous trees within RTM is still a difficult topic (Janoutová et al., 2019), this study only focused on broadleaved trees as a preliminary test case to study Mediterranean forests. Validation data were available at multiple dates, allowing for a more robust calibration of the inversion method. The validity of the results presented in this chapter was assessed using: (i) comparison between estimations and field measurements when and where available, and (ii) comparison between the site's expected and estimated LAI and foliar pigment contents variations over two years.

3.2 Materials and Methods

3.2.1 Materials

The present study focused on Tonzi Ranch and LAI and foliar pigments retrieval over multiple years and seasons. An aerial picture of the site and its location are given in Figure 3.1. Information concerning the number of plots and validation points for each date is given in Table 3.1. The set of airborne acquisitions used that were considered in given in Table 3.2.

Table 3.1: The number of validation points for LAI and leaf biochemistry and associated dates of collection. See Table 3.2 for more detailed information concerning the date of collection.

	Validation Data	
Date	**LAI**	**Biochemistry**
Summer 2013		3
Fall 2013	12	5
Summer 2014	19	5
Summer 2016	21	
Total	52	13

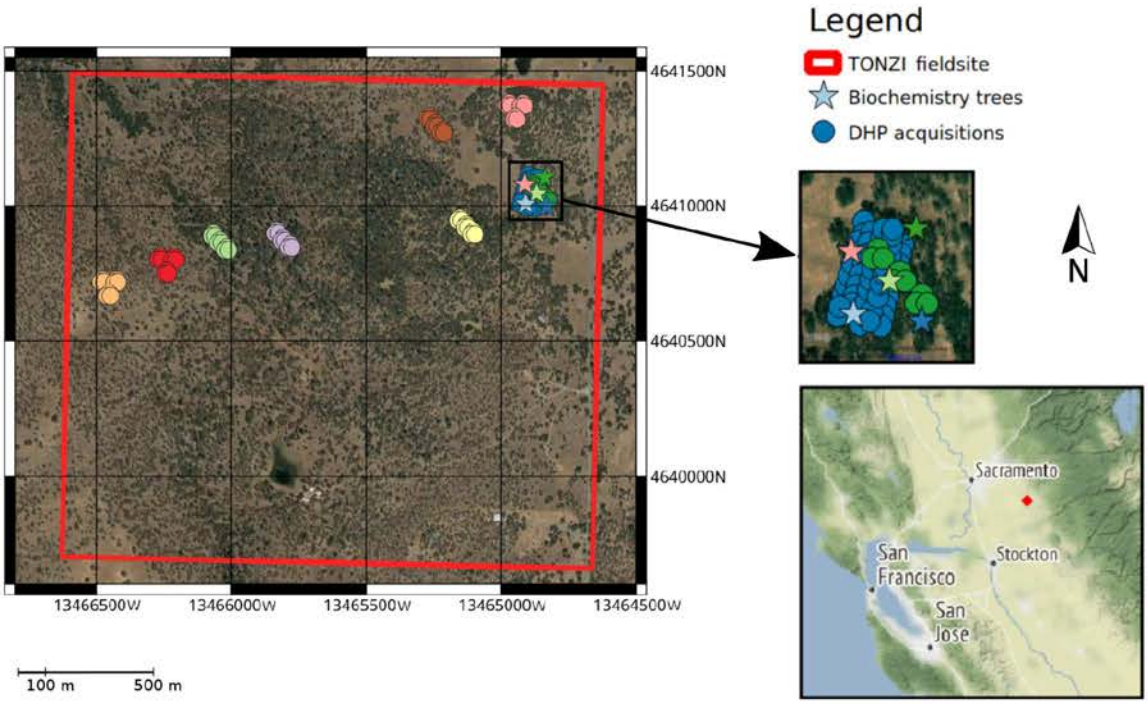


Figure 3.1: Location of the study site in California (bottom right) and location of field measurements. Discs: LAI measured with DHP; colors indicate plots differing in species composition and canopy cover. Stars: trees where leaves were sampled for laboratory biochemistry analysis; colors indicate individual trees.

Table 3.2: Description of AVIRIS-C and AVIRIS-NG airborne acquisitions.

		AVIRIS-C		**AVIRIS-NG**	
Year	**Season**	**Date (*Day Of Year* (DOY))**	**Time (*Pacific Daylight Time* (PDT))**	**Date (DOY)**	**Time (PDT)**
2013	Spring	2 May (122)	01h30 p.m.		
	Summer	4 June (155)	12h30 p.m.		
	Fall	19 September (262)	12h40 p.m.		
2014	Summer	2 June (153)	12h00 p.m.	6 June (157)	12h24 p.m.
2015	Spring	29 April (119)	11h40 a.m.		
	Summer	8 June (159)	01h40 p.m.		
	Fall	13 October (286)	12h30 p.m.		
2016	Summer	9 June (161)	12h30 p.m.		
2017	Summer	20 June (171)	01h00 p.m.		
2018	Summer	21 June (172)	01h00 p.m.		

3.2.2 General Methodology

In the present section, a description of each step of the methodology followed in this study is given. Figure 3.2 presents an outline of the steps described in the following subsections.

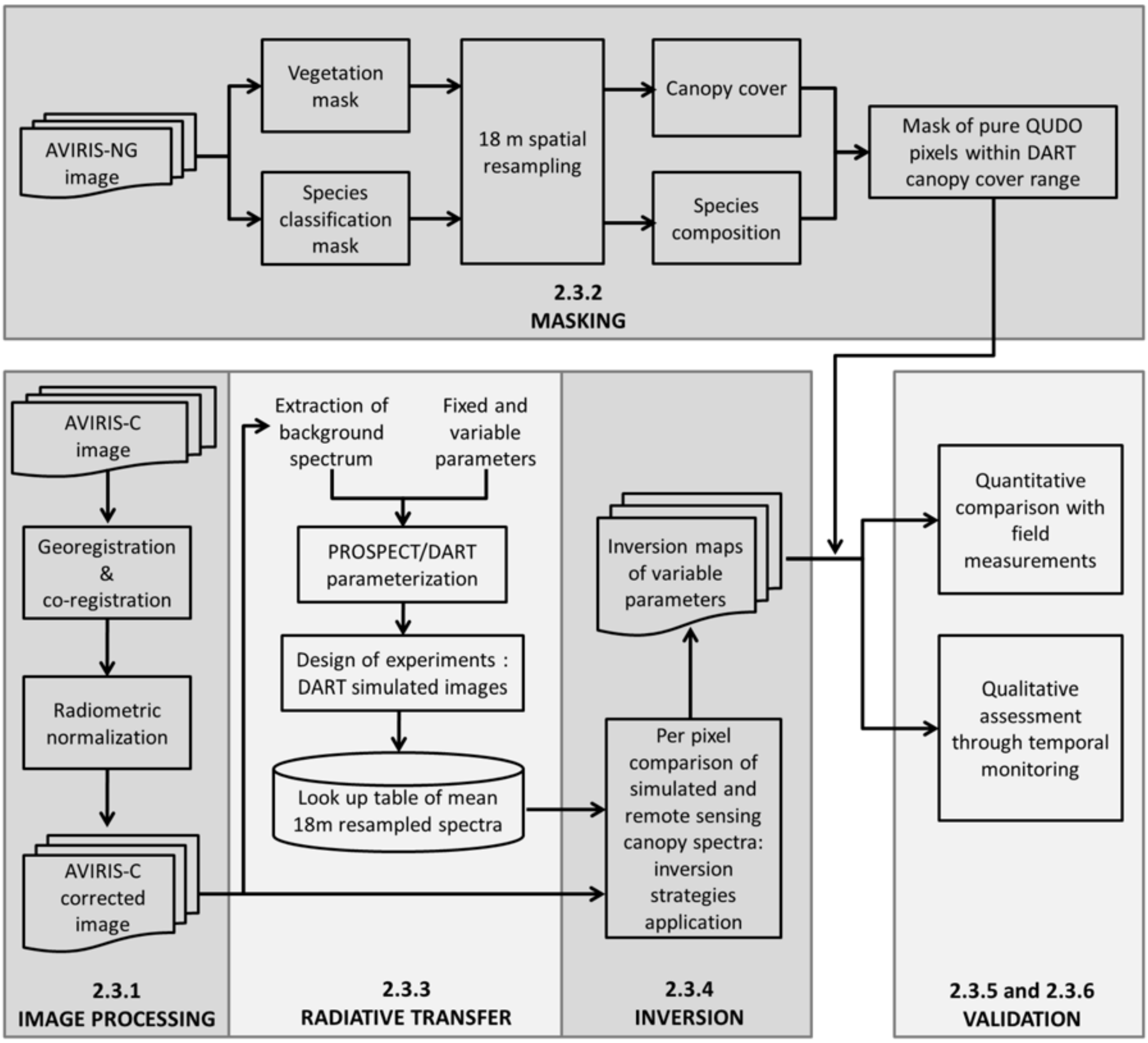


Figure 3.2: Methodology followed in the present study.

3.2.2.1 Image Processing for Multitemporal AVIRIS-C Data

In addition to the image processing protocol from NASA JPL (see Chapter 2), co-registration was further considered to improve the spatial matching between the AVIRIS-C images over the Tonzi site. The reference image was the summer 2014 AVIRIS-NG image, as the georeferencement seemed acceptable when overlaid on QGIS Microsoft Bing Maps satellite backgrounds. Co-registration was based on a nearest-neighbour resampling using around 25 ground control points per image. This amount could vary depending whether or not some features were identifiable on the AVIRIS-C images. For each image, the co-registration root mean square error (RMSE) was less than one pixel.

A temporal correction was also applied on the images, serving as an in-between flight radiometric normalization in order to compare the images on the same spectral baseline. An empirical line correction method (Hadjimitsis, Clayton, and Retalis, 2009; Karpouzli and Malthus, 2003; Schott, Salvaggio, and Volchok, 1988) was done using 9 invariant targets (3 black targets from the waters of Tahoe lake; 3 grey targets from flat sand/gravel areas; 3 white from different flat roofs with white painting) within the AVIRIS-C atmospherically corrected reflectance images. The reference image for the temporal correction was the spring 2013 AVIRIS-C image. The calibration targets were all located outside of the Tonzi site and spectral comparison over the various AVIRIS-C images showed that they were invariant in time.

3.2.2.2 Masking Using Canopy Cover and Species Composition Maps Derived From the AVIRIS-NG Image

The objective was to create a monospecific mask of QUDO. The final monospecific mask was obtained by (i) eliminating the areas with too many PISA in the images, (ii) eliminating pixels with too low tree presence (CC $< 10\%$) and (iii) removing the pixels containing water, roads, and manmade infrastructures. The AVIRIS-NG image at 2 m spatial resolution was processed to derive a 18 m CC map as well as a species composition map. Figure 3.3 gives an overview of the processing steps done over the AVIRIS-NG image.

To build a canopy cover map, a vegetation mask was built based on a simple histogram thresholding with the application of a vegetation index. Version 2 of the modified chlorophyll absorption ratio index (MCARI2, (Haboudane et al., 2004)) was chosen given that the two main components in the image can be easily separated in summer: a healthy green tree canopy (chlorophylls peak) and dry dead grass (no chlorophyll). Finally, a spatial resampling was applied using the pixel aggregate method to come back to the 18 m spatial resolution of AVIRIS. The statistics provided from the derived canopy cover map show that the mean canopy cover is 53% over the area of interest with a concave bell curve from 0% to 90%. In addition, a peak also appears around 100% in the distribution histogram (see Figure 3.3).

To build the species classification map, the supervised support vector machine (SVM) classification method was used through the program LibSVM (Chang and Lin, 2011). One class was created for each tree species (QUDO and PISA). A total of 506 and 557 pixels were respectively selected for each class as training samples based on fieldwork data, tree structural information derived from LiDAR data collected in 2009 (PISA sample height assumed to be higher than 20 m, (Kobayashi et al., 2012)) and visual inspection. A total of 339 spectral bands were selected to obtain the same spectral bands as AVIRIS-C. The SVM was run with the radial basis function kernel (C: 6.5; γ: 0.0055; accuracy: 93%; cross-validation correlation coefficient: 98.4). For verification purposes, the resulting classification map was also visually compared to a true-color image of the site, as the tree species have clearly distinct colors. Finally, an 18 m species composition map was created by applying the same spatial resampling

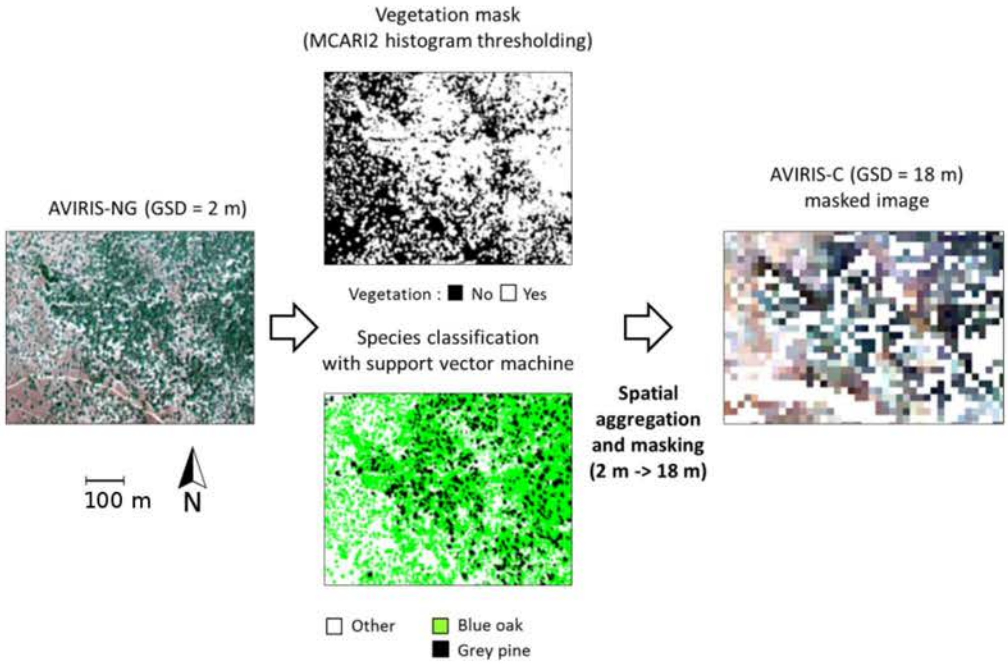


Figure 3.3: Steps used for the mask building using the AVIRIS-NG image and masking of the AVIRIS-C images.

as for the canopy cover map.

Pixels where QUDO crowns represented less than 80% of the total tree crowns were rejected as the PISA presence would be too high, as well as pixels with a CC below 10%. Finally, pixels containing water, roads, and manmade infrastructures were manually masked over each image.

3.2.2.3 DART and PROSPECT Radiative Transfer Models Parametrization

DART version 5.7.3v1078 and PROSPECT-5 were used to simulate canopy reflectances and build the LUT.

The scene modeling done in this study is based on a SFR (see Figure 3.4). A repetitive pattern of four trees with ellipsoidal crowns was used (Gascon et al., 2004). Crown size and height were set to the Tonzi site average values (Chen et al., 2008). Branches and twigs were neglected to reduce modeling complexity and introduction of errors if not correctly set. Trunk reflectance was set to the value measured in-field. Canopy cover variations were obtained by increasing the scene dimensions. The trees were placed so that the scene *Bidirectional Reflectance Factor* (BRF) is the closest to a forest BRF, following the pattern given in Gastellu-Etchegorry, Gascon, and Estève, 2003 and Gascon et al., 2004. Trees and scene characteristics are given in Table 3.3. The forest understory was represented as a lambertian surface whose optical properties were determined by pure soil pixels selected within the AVIRIS-C images in open parts of the site, at least one pixel (18 m) away from canopy pixels and several pixels wide. These open parts were visually chosen using QGIS high-resolution Microsoft Bing Maps satellite backgrounds. Leaf optical properties were simulated through the PROSPECT model implemented in DART, using the input parameters described in Table 3.4. The leaf structure parameter N was set to 1.8. Sun and atmosphere were the only radiation sources. The sun positions were set so that they match with the AVIRIS-C images based on site location and hours of acquisition. The simulations used a rural aerosol model with a visibility of 23 km corresponding to mid-latitude summer standard atmosphere. A total of 168 bands with a spectral bandwidth of 0.010 µm were simulated between 0.36 and 2.45 µm to cover both the visible and *Short-Wave InfraRed* (SWIR) spectral ranges.

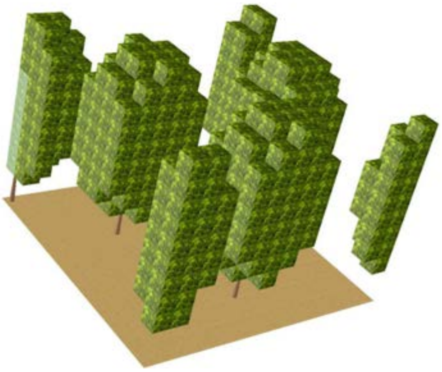


Figure 3.4: Illustration of the simplified forest representation.

Table 3.3: DART elementary scene fixed and variable input parameters used for the LUT generation.

Parameters	Values/Range	Step	No. of Values
General settings			
CC (%)	10–90	20	5
Scene dimensions $x \times y$ (m^2) for:			
CC 10 %	32.4 × 32.4		
CC 30 %	18.8 × 18.8		
CC 50 %	14.4 × 14.4		
CC 70 %	12.4 × 12.4		
CC 90 %	10.8 × 10.8		
Voxel size x, y, z (m)	0.4, 0.4, 0.4		
Tree characteristics			
Tree height (m)	9.4		
Crown shape	ellipsoidal		
Crown diameter (m)	5.8		
Crown height (m)	7.5		
Trunk height (below & within crown) (m)	1.9, 4.73		
Trunk dbh (m)	0.26		

Table 3.4: Leaf fixed and variable input parameters used for the LUT generation.

Parameters	Values/Range	Step	No. of Values
LAD	spherical		
LAI (m^2/m^2)	0.1–1.9	0.3	7
C_{ab} (µg/cm^2)	10–60	10	6
Car (µg/cm^2)	2–22	4	6
LMA (g/cm^2)	0.001–0.016	0.003	6
EWT (cm)	0.001–0.021	0.004	6

3.2.2.4 Inversion Stategies

This study follows a LUT-based approach consisting in finding the DART model parameters minimizing a cost function comparing measured y to computed $\hat{y}$ reflectances. Since the inversion problem is ill-posed, multiple sets of parameters can yield similar reflectances. A common estimation method is to consider the mean values from the q best sets of parameters as the final solution (Combal et al., 2003; Hill, Buddenbaum, and Townsend, 2019). Various cost functions were tested: RMSE (Equation (4.2)), *Spectral Angle Mapper* (SAM) (Equation (4.3)) and VI differences (D_{VI}, Equation (4.4)).

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3.1}$$

$$SAM(y,\hat{y}) = cos^{-1}\left(\frac{\Sigma_{i=0}^{N} y_i \hat{y}_i}{\sqrt{\Sigma_{i=1}^{N} \hat{y}_i \hat{y}_i}\sqrt{\Sigma_{i=1}^{N} y_i y_i}}\right) \tag{3.2}$$

$$D_{VI}(y,\hat{y}) = abs\left(VI(y) - VI(\hat{y})\right) \tag{3.3}$$

RMSE and SAM were computed using variable-specific spectral intervals. The NIR and SWIR were used for LAI (INT LAI, 0.8–2.45 µm), while parts of the visible range were used for C_{ab} and Car (INT CAB, 0.5–0.75 and INT CAR, 0.5–0.55 µm respectively). D_{VI} was computed using various indices. Concerning LAI, the NDVI was selected, as it is commonly used to estimate vegetation cover, as well as MSAVI2, which is a soil-adjusted index; for C_{ab}, Maccioni, GM_94b, gNDVI, MCARI2, and TCARI/OSAVI were considered, since they are soil-adjusted indices specifically designed for low-LAI environments, and they were successfully used in a sparse forest context (Haboudane et al., 2004; Qi et al., 1994; Verrelst et al., 2010; Zarco-Tejada et al., 2013); concerning Car, CRI and R515/R570 were chosen, as they were also successfully used in previous studies (Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012; Zarco-Tejada et al., 2013). Table 3.5 summarizes the selected methods for all investigated variables as well as the number of spectral bands used.

The number q of best-matching cases to use for the inversion was selected on a per-variable basis. The objective was to find an optimal value that would minimize the RMSE between field data and estimations when performing the LUT-based inversion. Therefore, inversions were done with increasing q values, from 1 to 400 (0.9 % of the LUT). If an optimal value existed across all validation dates, it would determine the number of solutions to keep for the inversion validation step (Section 4.2.5) and then the inversion of all remaining AVIRIS-C dates available.

Table 3.5: Spectral strategies applied for the inverse problem.

Parameter	Method	N$_{bands}$	VI formula (wavelengths in µm)	Reference
LAI	RMSE INT LAI	120		
	SAM INT LAI	120		
	D_{NDVI}	2	$\frac{\rho_{0.833} - \rho_{0.677}}{\rho_{0.833} + \rho_{0.677}}$	(Tucker, 1979)
	D_{MSAVI2}	2	$\frac{1}{2}(2\rho_{0.800} + 1 - \sqrt{(2\rho_{0.800} + 1)^2 - 8(\rho_{0.800} - \rho_{0.670})})$	(Qi et al., 1994)
Cab	RMSE INT CAB	27		
	SAM INT CAB	27		
	$D_{TCARI/OSAVI}$	4	$3\frac{(\rho_{0.700} - \rho_{0.670}) - 0.2(\rho_{0.700} - \rho_{0.550})\frac{\rho_{0.700}}{\rho_{0.670}}}{1.16\frac{\rho_{0.800} - \rho_{0.670}}{\rho_{0.800} + \rho_{0.670} + 0.16}}$	(Haboudane et al., 2002)
	$D_{MACCIONI}$	3	$\frac{\rho_{0.780} - \rho_{0.710}}{\rho_{0.780} - \rho_{0.680}}$	(Maccioni, Agati, and Mazzinghi, 2001)
	D_{gNDVI}	2	$\frac{\rho_{0.780} - \rho_{0.550}}{\rho_{0.780} + \rho_{0.550}}$	(Smith et al., 1995)
	D_{GM_94b}	2	$\frac{\rho_{0.750}}{\rho_{0.550}}$	(Gitelson and Merzlyak, 1994)
	D_{MCARI2}	3	$1.5\frac{2.5(\rho_{0.800} - \rho_{0.670}) - 1.3(\rho_{0.800} - \rho_{0.550}))}{\sqrt{(2\rho_{0.800} + 1)^2 - (6\rho_{0.800} - 5\sqrt{\rho_{0.670}}) - 0.5}}$	(Haboudane et al., 2004)
Car	RMSE INT CAR	6		
	SAM INT CAR	6		
	$D_{R515/R570}$	2	$\frac{\rho_{0.515}}{\rho_{0.550}}$	(Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012)
	D_{CRI}	2	$\frac{1}{\rho_{0.515}} - \frac{1}{\rho_{0.570}}$	(Gitelson et al., 2006)

3.2.2.5 Validation Metrics

Validation was performed by comparing LAI, chlorophylls, and carotenoids estimates with the field measurements available using RMSE, bias, the R^2 of the predicted-vs.-measured regression line and the *standard deviation of the difference between estimated points and the regression line* (STDB) as cost functions.

Concerning LAI, each validation point was the average of LAI values derived from 3 hemispherical pictures taken within a 20 × 20 m plot (see Figure **??**). Therefore, one LAI validation point does not necessarily only contain information about the LAI of the AVIRIS-C pixel it is associated with. Figure 3.5 illustrates the issue: Since the AVIRIS pixels and the grid used for DHP acquisitions have no reason to match, some DHP (dots) may have been acquired out of the pixel the final LAI has been associated with (red square). Direct comparison between pixel-estimated value and validation data can therefore be inappropriate, especially for spatially heterogeneous canopies. To address this, LAI validation points were compared to a spatially smoothed LAI inversion map. The spatial smoothing was done by computing the average LAI within a sliding window with an aperture size of 3 pixels. The aperture size was set to 3 as the LAI validation plot size (20 × 20 m) roughly matches the AVIRIS-C pixel size (18 × 18 m) (Figure 3.5).

Biochemistry validation data were obtained at the leaf scale, for a single tree in each validation pixel. There is therefore an unavoidable uncertainty, as validation is done by comparing field data at the leaf scale (single tree) to plot biochemistry (canopy scale, multiple trees). An assumption is made that tree biochemistry does not vary much spatially, and C_{ab} and Car estimations were directly extracted from the pixels associated to the acquisition positions.

3.2.2.6 Seasonal Monitoring

To provide a qualitative validation of LAI and pigment estimations, estimations using the best inversion strategies (Sections 3.3.3.2 and 3.3.3.3) were computed, and seasonal variations were monitored and compared to expected variations. The seasonal monitoring involved LAI and foliar C_{ab} and Car. Estimation maps were built using AVIRIS-C acquisitions done in the spring (DOY 122 and 119), summer (DOY 155 and 159), and fall (DOY 262 and 286) of 2013 and 2015. Each estimation's uncertainty is the STDB of the variables' estimations (STDB are presented in Sections 3.3.3.2 and 3.3.3.3). Changes in the site's mean LAI and foliar C_{ab} and Car were considered significant if they were greater than the STDB of their respective variables.

3.3 Results

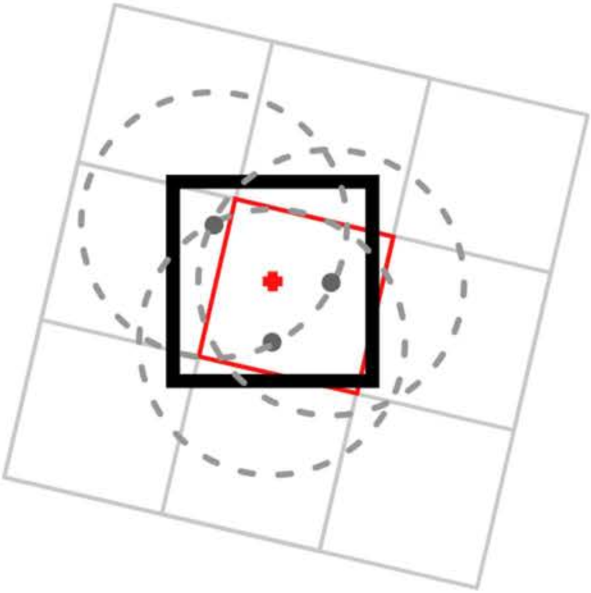


Figure 3.5: Diagram showing one LAI validation point (red dot) within its 20 × 20 m^2 plot (bold square). The validation point falls within the AVIRIS-C pixel in red. Other pixels are represented by the grey grid. LAI is derived from 3 DHP taken at various locations within the plot, each one covering an area approximated by a dashed circle.

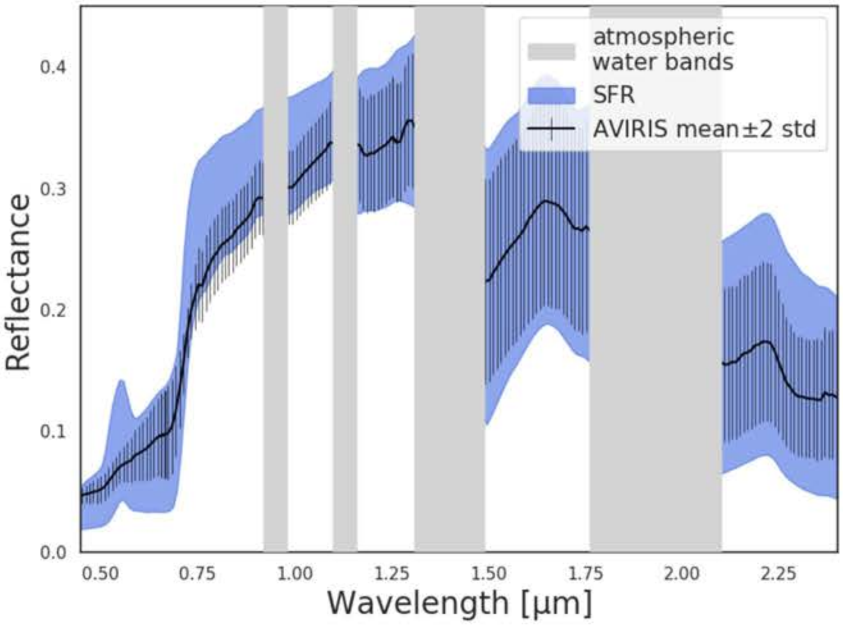


Figure 3.6: Reflectances computed by the DART simulations compared to reflectances of vegetation pixels within the AVIRIS-C image for summer 2014.

3.3.1 Comparison between Airborne and DART-Simulated Reflectances

The assessment of LUT reflectances acceptability with regard to those measured by AVIRIS-C was done in two steps. First, DART-simulated reflectances of each LUT were plotted and compared to the reflectances measured by AVIRIS-C at the associated date (see, for example, Figure 3.6). All LUTs consistently overestimated canopy reflectances between 0.75 and 1.1 µm , with the mean image reflectance being either barely within or below LUT reflectances near the red-edge. Image reflectances in the visible and SWIR regions were, however, within simulation ranges.

Secondly, the amount of pixels that could be explained by the LUT was evaluated. The percentages of pixels explained by the LUT within each image are given in Table 3.6. LUT reflectances encompass image reflectances at various degrees depending on the spectral intervals considered (INT LAI, INT CAB, INT CAR). On average, 54% of pixels' reflectances completely fall within the LUT reflectance boundaries for INT LAI. This goes up to 67% and 97% for INT CAB and INT CAR, respectively. When considering the vegetation indices computed from both images and LUT reflectances, pixel-derived VI are within LUT-derived VI boundaries from 90% to 100% of the time (excepting summer 2017, which has one occurrence at 83%), depending on the VI considered.

Table 3.6: Percentages of pixels within AVIRIS-C images which reflectance (for INT LAI, INT CAB, and INT CAR rows) or associated vegetation indices (other rows) are within those derived from the DART LUT.

		2013			**2014**	**2015**			**2016**	**2017**	**2018**
		Spr.	**Sum.**	**Fall**	**Sum.**	**Spr.**	**Sum.**	**Fall**	**Sum.**	**Sum.**	**Sum.**
LAI	INT LAI	32	58	67	62	67	65	53	25	56	55
	NDVI	99	99	99	100	99	98	97	99	99	97
	MSAVI2	96	98	94	99	93	97	90	97	98	96
C_{ab}	INT CAB	46	71	84	82	84	79	71	33	60	62
	MCARI2	98	98	91	99	97	98	90	98	98	97
	TCARI/OSAVI	100	100	97	99	100	98	98	93	98	97
	Maccioni	100	100	100	100	100	100	100	100	100	100
	gNDVI	100	100	100	100	100	100	100	100	100	100
	GM_94b	100	100	100	100	100	100	100	100	100	100
Car	INT CAR	100	100	100	100	100	99	100	100	94	78
	R515/R570	100	100	100	100	100	100	100	100	100	100
	CRI	100	99	100	100	100	92	100	93	83	68

3.3.2 Influence of the Canopy Cover on Vegetation Indices

Once the LUT was built, it was also necessary to assess the suitability of each VI to serve as a criterion for LAI and biochemistry estimations when using the LUT entries. Indeed, it is clear that the green vegetation signal is quite low. The reflectance of low CC pixels is close to that of pure ground, with only a slight red-edge increase (Figure 3.7). Even at high CC, the red-edge is not very steep, and the chlorophylls reflectance peak at 0.55 µm is barely visible: It could be possible that some VI would not be able to detect some of the variations of the variables of interest.

To do this assessment, vegetation indices were computed over the LUT reflectances. Figure 3.8 shows the evolution of three different VI over the variable they are associated with, on a per-CC basis (colors). Ideally, no matter the CC, the VI should present variations when their associated variable increases, i.e., behave as a bijective function.

It was found that the indices selected for LAI estimations all started saturating at 0.4 m^2/m^2 LAI for the 10% of CC cases (Figure 3.8a, red line). This behavior was not found for higher CC. Since the 10% CC category could only bring confusion due to this saturation when estimating LAI, it was decided to remove these cases from the LUT. Excluding these cases could also be done based on their lack of realism, as low CC–high LAI canopies are unrealistic. Vegetation indices dedicated to C_{ab} and Car estimation did not show such behavior (Figure 3.8b,c); therefore, all CC cases were considered during the inversion step for C_{ab} and Car.

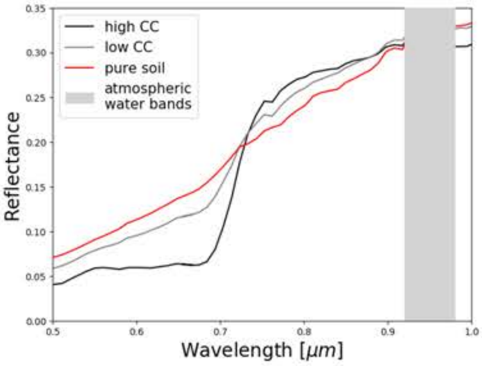
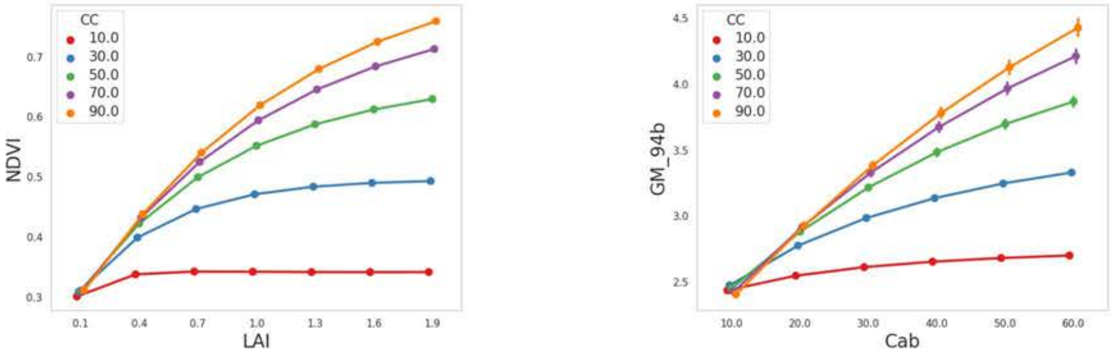
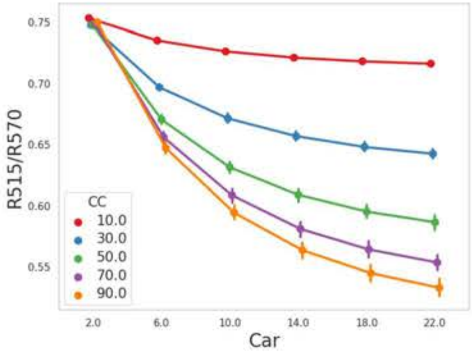


Figure 3.7: Reflectances of a low (grey) and a high (black) canopy cover QUDO pixel in summer 2014. In red, reflectance of a pure ground pixel.



(a) NDVI

(b) GM_94b



(c) R515/R570

Figure 3.8: Computed VI for each LUT entry over (**a**) LAI, (**b**) C_{ab}, and (**c**) Car. In color, distinction by CC. Only one example is given for each variable, as other VI behaved the same.

3.3.3 Selection of the Best Inversion Strategy

3.3.3.1 Determination of the Number q Achieving the Best Inversion Performances

For each variable, in order to assess the uncertainty associated with determining q using only validation data from a single date, q was determined both on a date-to-date basis and on a multidate basis.

For LAI, ignoring the RMSE INT criterion which performed very poorly, all criteria seemed to have reached a plateau by $q = 100$ at most (see Table 3.7 and Figure 3.9a). As there was no obvious optimal q which could minimize RMSE, it was decided to consider the 100 best performing solutions for LAI estimation. While for summers 2014 and 2016, SAM INT LAI, NDVI, and MSAVI2 performed similarly, for fall 2013, only the VI-based criteria achieved acceptable performances (around 0.6 m^2/m^2 for RMSE and SAM INT LAI, around 0.2 m^2/m^2 for NDVI and MSAVI2).

C_{ab} results showed a similar plateauing behavior (see Table 3.7 and Figure 3.9b). Therefore, it was decided to proceed the same way as the LAI and to consider the 300 best performing solutions for C_{ab} estimations, since the RMSE plateau was always reached at this point. The lowest RMSE for summer 2013 was obtained with TCARI/OSAVI (5.86 $\mu g/cm^2$). The lowest RMSEs for fall 2013 and summer 2014 were obtained with gNDVI (3.86 $\mu g/cm^2$) and GM_94b (2.89 $\mu g/cm^2$), respectively. MCARI2 performed very poorly for all dates, even more so than the RMSE and SAM criteria that at least showed acceptable performances for summer 2014.

For Car, an optimal q value was found when singling out the summer 2013 data and using the RMSE INT CAR criterion. However, at the other dates, the RMSE INT CAR criterion barely reached its plateauing value by $q = 400$ (see Table 3.7 and Figure 3.9c), so the 400 best performing solutions were considered for the multidate Car estimations. RMSE INT CAR presented the lowest RMSE for summer and fall 2013, while only presenting acceptable performances for summer 2014. Only RMSE INT CAR and CRI consistently showed acceptable performances, with a slight advantage for RMSE INT CAR. Both SAM INT CAR and R515/R570 had important RMSE variability over the three available dates.

3.3.3.2 LAI Retrieval Performance Comparison

Table 3.8 shows LAI retrieval performances when using the mean of the $q = 100$ best solutions. Inversions based on NDVI and MSAVI2 differences performed equivalently in terms of RMSE (less than 0.25 m^2/m^2), bias, STDB, and R^2 (above 0.80). RMSE INT LAI performed the worse for all validation metrics, followed by SAM INT LAI.

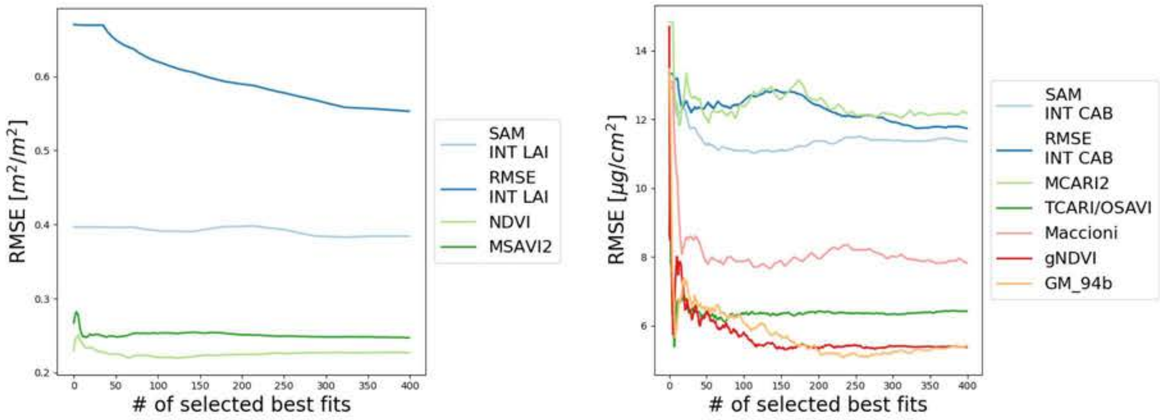
Figure 3.10 shows that the highest LAI are slightly underestimated and vice versa, and that estimation uncertainties increase with LAI. Concerning low LAI points, only the fall 2013 points (blue) can be overestimated. No clear distinction in terms of performances is observed between other seasonal measurements. The global performance shows a R^2 of 0.8 and a total

Table 3.7: RMSE of the various criteria at the retained q value for LAI, C_{ab} and Car estimations of each date.

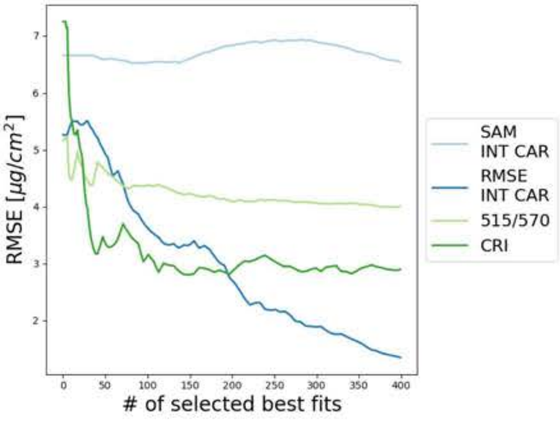
		Fall 2013	**Summer 2014**	**Summer 2016**	**All Dates**
	q	100	100	100	100
LAI (m^2/m^2)	RMSE INT LAI	0.61	0.61	0.63	0.62
	SAM INT LAI	0.66	0.21	0.31	0.39
	NDVI	0.17	0.23	0.24	0.22
	MSAVI2	0.18	0.24	0.29	0.25
		Summer 2013	**Fall 2013**	**Summer 2014**	**All Dates**
	q	300	300	300	300
C_{ab} ($\mu g/cm^2$)	RMSE INT CAB	12.45	15.36	6.36	11.92
	SAM INT CAB	9.1	15.91	5.8	11.37
	MCARI2	10.44	14.38	10.57	12.15
	TCARI/OSAVI	5.86	8.09	4.31	6.34
	Maccioni	8.38	9.34	6.12	8.02
	gNDVI	9.09	4.22	2.89	5.39
	GM_94b	8.62	3.86	3.39	5.21
	q	400	400	400	400
Car ($\mu g/cm^2$)	RMSE INT CAR	0.58	1.14	2.94	1.34
	SAM INT CAR	4.78	9.31	2.36	6.54
	R515/R570	5.74	4.32	2.74	4.01
	CRI	2.87	3.83	1.91	2.89

Table 3.8: Inversion methods performances for LAI estimation for $q = 100$ over all dates available.

Method	**RMSE (m^2/m^2)**	**bias (m^2/m^2)**	**STDB (m^2/m^2)**	R^2
RMSE INT LAI	0.62	−0.49	0.29	0.38
SAM INT LAI	0.39	−0.27	0.25	0.63
NDVI	0.22	0.07	0.18	0.80
MSAVI2	0.25	0.14	0.17	0.81



(a) LAI (b) C_{ab}



(c) Car

Figure 3.9: Influence of q on the estimation RMSE for LAI (**a**), C_{ab} (**b**) and Car (**c**) using validation data from every date.

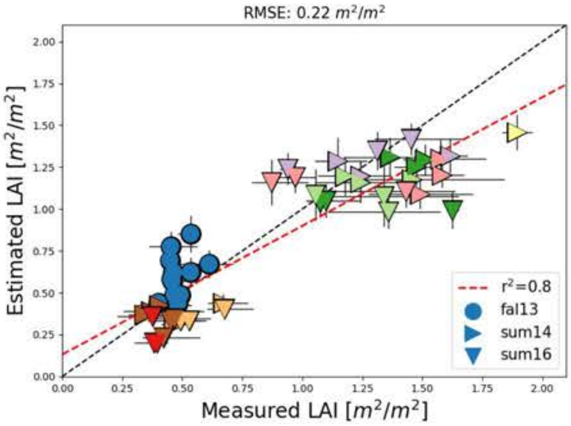


Figure 3.10: Estimated LAI using the D_{NDVI} method versus measured LAI. Colors indicates the location within the site, see Figure 3.1. Error bars for the y-axis correspond to the estimation standard deviation, computed as the standard deviation of the first $q = 100$ solutions. For the x-axis, as each plot's measured LAI is the average of the computed LAI from each of the DHP of the plot, error bars correspond to the minimum and maximum values of each plot's DHP set.

RMSE of 0.22 m^2/m^2.

3.3.3.3 C_{ab} and Car Retrieval Performance Comparison

For C_{ab} at $q = 300$, apart from D_{MCARI2}, VI differences performed better than methods based on RMSE and SAM (Table 3.9). GM_94b is the overall best-performing VI, with the lowest RMSE, highest R^2, and lowest STDB (5.21 µg/cm^2, 0.73, and 3.38 µg/cm^2, respectively), besting even soil-adjusted VI. When compared to field measurements from all dates, most GM_94b-estimated points are very close to the first bisector, and only one point (pink from summer 2013) is greatly underestimated (Figure 3.11a).

For Car, at $q = 400$, the best method is also clear: RMSE INT CAR is the only method to present a low RMSE, a low STDB, and a high R^2 (1.34 µg/cm^2, 1.06 µg/cm^2 and 0.59, respectively. See Table 3.9). The RMSE INT CAR method showed the best performances overall, with estimated values very close to the first bisector (Figure 3.11b) for all seasons.

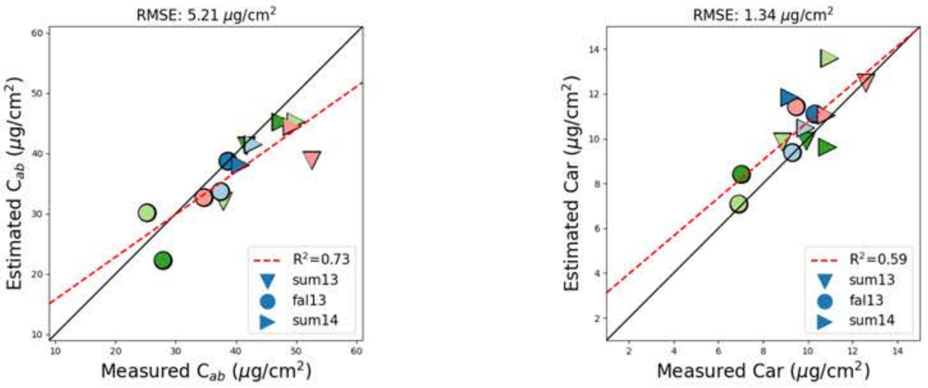
3.3.4 LAI, C_{ab} and Car Seasonal Monitoring

In Appendix A are given estimation maps of Tonzi summer LAI and canopy C_{ab} and Car (LAI × foliar pigment content) from 2013 to 2018 (Figures A.1–A.3, respectively). These maps were computed using the best VI identified in Sections 3.3.3.2 and 3.3.3.3.

Focusing on years 2013 and 2015, estimated LAI and foliar C_{ab} decreased from spring to fall (steady decrease from 0.9 m^2/m^2 and 38 µg/cm^2 to 0.4 m^2/m^2 and 30 µg/cm^2, respectively).

Table 3.9: Inversion method performances for C_{ab} ($q = 300$) and Car ($q = 400$) estimations over all dates available.

Method	**RMSE (μg/cm^2)**	**bias (μg/cm^2)**	**STDB (μg/cm^2)**	R^2
C_{ab}				
RMSE INT CAB	11.92	8.99	5.21	0.14
SAM INT CAB	11.37	7.46	6.05	0.08
MCARI2	12.15	−5.05	8.62	0.01
TCARI/OSAVI	6.34	2.75	4.15	0.48
Maccioni	8.02	4.36	5.03	0.32
gNDVI	5.39	−2.15	3.82	0.61
GM_94b	5.21	−3.21	3.38	0.73
Car				
RMSE INT CAR	1.34	0.79	1.06	0.59
SAM INT CAR	6.53	1.59	4.53	0.29
R515/R570	4.01	−1.26	3.74	0.26
CRI	2.89	-0.2	2.1	0.05



(a)

(b)

Figure 3.11: Measured versus estimated (**a**) leaf C_{ab} using the D_{GM_94b} method and (**b**) leaf Car using the RMSE INT CAR method. Colors indicate the location within the site, see Figure 3.1.

Site LAI variations from summer to fall in 2013 and 2015 are greater than the estimations' STDB (Figure 3.12), while from C_{ar} only, summer to fall variations in 2015 are greater than the STDB (Figure 3.13). Concerning carotenoids, estimated variations are below the STDB and are therefore not remarkable.

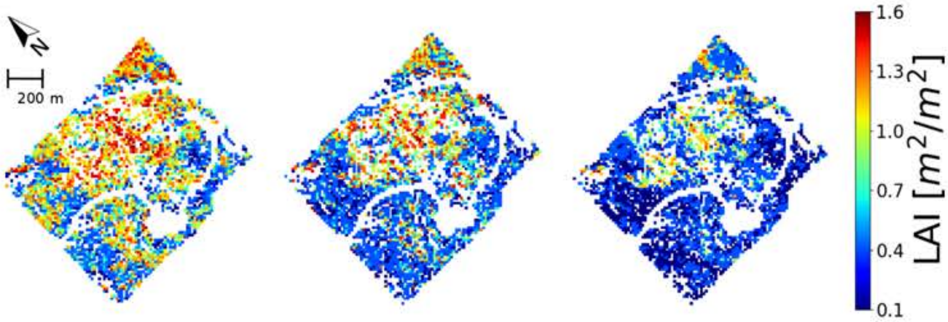


Figure 3.12: Estimated LAI over the year 2013. From left to right, spring, summer, and fall 2013.

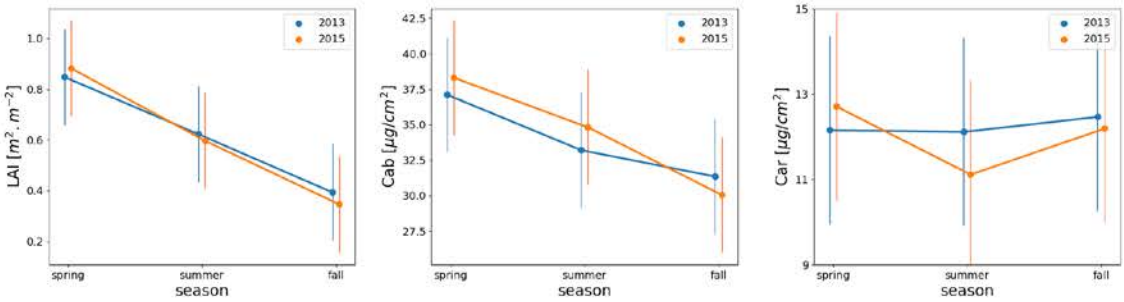


Figure 3.13: Mean site LAI, foliar C_{ab}, and foliar Car over the years 2013 and 2015. Bars indicate the STDB.

3.4 Discussion

3.4.1 Limitations of the SFR for Low-LAI Sparse Forests Modeling within DART

The forest representation chosen for this study aimed at minimizing the *a priori* knowledge necessary to monitor the Tonzi site. Trees modeled within the DART scenes had simplistic crown shapes and few nonphotosynthetic elements, as branches were not included. This has an impact on both LAI and C_{ab} estimations, as woody elements' influence on canopy reflectance can be important (Malenovský et al., 2008a; Verrelst et al., 2010) for heterogeneous and defoliated stands such as the Tonzi site. In particular, Malenovský et al., 2008a and Widlowski, Côté, and Béland, 2014 showed that inclusion of woody elements within the DART canopies affects NIR reflectances for the resolutions considered in this study, which could explain the overestimated reflectances of the LUT in this spectral region (Figure 3.6).

These mismatches between computed and measured reflectances are such that some of the inversion methods tested in this study are not appropriate. Indeed, Table 3.6 shows that when considering the INT LAI and the INT CAB intervals, a large proportion of the AVIRIS-C images' pixels could not be explained by the LUT. The consequence is that VI-based inversion criteria outperform those based on spectral intervals. While some VI may use reflectances in the NIR which are overestimated by DART (Figure 3.6), it appears that the overestimation effect is dampened as VI are less sensitive to differences in reflectance levels, and pixel-derived VI are almost always within the boundaries of LUT-derived VI.

The simplistic ground representation within the DART scenes, with a flat ground having a lambertian reflectance, may also be a source of errors, especially for the spring acquisitions, as important spatial variations of ground reflectance can be expected (see Sections 2.1 and 3.4.4). Unfortunately, such variation cannot be taken into account by the simple forest representation implemented in this study. A possible way to tackle this, though it would increase computation time, would be to introduce ground reflectance as another LUT parameter. However, the difficulty would be to make sure that most of the possible ground reflectances are captured by the LUT, which may not be an easy task when working with medium- to low-resolution images.

3.4.2 Time Dependency of the Best Performing Inversion Method

The objective was to be able to monitor the study site over multiple dates with or without available field data. As field data are usually scarce, inversion methods are often calibrated using data from only a single date. While this guarantees that the optimal method for this date has been found, the validity of extrapolations to images acquired at different times is not a given. Indeed, in Section 3.3.3's Table 3.7, TCARI/OSAVI, GM_94b and gNDVI indices could be identified as optimal depending on the date. However, when considering the complete dataset, which includes summer and fall data, GM_94b outperforms the others with a lower RMSE and higher R^2 (Table 3.9). This highlights the need for multiple validation dates to properly select the inversion method if one aims to monitor canopies over time.

The number of best matches q to keep when doing the LUT-based inversion also varied depending on the validation date, and a generalizable method to determine it was not found. There is still no clear methodology to determine this number. Banskota et al., 2015 used three different values for the inversion ($q = 50$ when using 6 bands from AVIRIS-C images, $q = 150$ when using their full spectrum, and $q = 200$ when using Landsat images; LUT size: 115,000); Atzberger and Richter, 2012 only considered the best match when using 8 bands from CHRIS images (LUT size: 1,056,000); Weiss et al., 2000 used $q = 50$ when trying to inverse digital data sets (LUT size: 280,000); and Ferreira et al., 2018 considered $q = 20$ when using AISA images (LUT size: 3500). This is unfortunately not satisfactory, as much room is left for subjectivity and overfitting of the inversion method.

3.4.3 Assessment of LAI and Pigment Estimations Accuracy

Even though the site LAI is very low (mean LAI is 0.8 m^2/m^2 and mean field points' LAI is 0.88 m^2/m^2), LAI estimations were encouraging. The LAI heterogeneity within the Tonzi site was captured, as for a given date, points from both low and high LAI parts of the site were correctly estimated. Seasonal variations were also visible, as LAI variations from summer to fall are clearly followed. Indeed, in Figure 3.10, blue and dark green points come from the same part of the site (see site map, Figure 3.1) but are from fall and summer, respectively. Estimation RMSE (0.22 m^2/m^2) is remarkably low, and nRMSE (RMSE divided by the mean field points' LAI) is 0.25, quite similar to what can be obtained over a higher-LAI site: when estimating LAI over a Mediterranean grassland with a mean LAI of 2.87 m^2/m^2, Darvishzadeh et al., 2011 obtained nRMSEs of 0.22 and 0.18 using statistical and RTM-inversion approaches, respectively.

C_{ab} estimation performances were in line with previous studies. Foliar chlorophylls content estimations (RMSE = 5.21 $\mu g/cm^2$, $R^2 = 0.73$) are as good as what could be obtained over higher-LAI coniferous and deciduous sites in various studies (RMSE = 4.34 $\mu g/cm^2$, $R^2 = 0.47$ (Zhang et al., 2008); RMSE = 4.2 $\mu g/cm^2$ (Delegido et al., 2010); RMSE = 8.1 $\mu g/cm^2$, $R^2 = 0.40$ (Zarco-Tejada et al., 2004)). Carotenoid estimations also performed well with a low RMSE and high R^2 (RMSE = 1.34 $\mu g/cm^2$, $R^2 = 0.59$). This is similar to the values obtained by Zarco-Tejada et al., 2013 using high-resolution imagery (50 cm) over vineyards(RMSE below 1.3 $\mu g/cm^2$ and R^2 of at most 0.46 when using the SAILH and the FLIGHT radiative transfer models).

A factor that could explain some estimation errors (and specifically the underestimation of the C_{ab} summer 2013 pink point) is the assumption that the biochemistry measured on one tree can be considered representative of (i) the biochemistry of leaves that are sensed by AVIRIS-C (i.e., mostly top of canopy leaves) and (ii) the leaf biochemical content of all the trees within a pixel. While to address (i), attention was given to only collecting leaves on the sunlit part of the crown (see Section 2.3.2), it is not possible to confirm (ii) with the available field data, and a leap of faith must be made. There is therefore always an uncertainty concerning how representative validation data are with regards to their associated pixels.

3.4.4 Challenges Concerning LAI, C_{ab} and Car Monitoring

Concerning seasonal monitoring, while the QUDO phenophase is such that the site LAI and foliar C_{ab} should have presented a maximum in summer (Demarez et al., 1999; Gara et al., 2019), Figure 3.13 showed a decrease from spring to fall. The most probable cause is that the C3 grass was not completely dead in May, as such a behavior is consistent with the results found by Stckli et al., 2011. This induces a combined contribution of both the green grass and the QUDO vegetation layers in the spring. Even though grass reflectance was taken into account within DART models so that its biochemistry would not influence results, the pure soil pixels manually selected from the AVIRIS-C image for ground reflectance input within the DART model may have contained less green grass than other, non-pure-ground pixels. This would lead to erroneous estimations where green grass is present. As for both summer and fall, the grass is completely dry, which was not an issue for these seasons, and a LAI and foliar C_{ab} decrease was detected. Foliar carotenoid content was correctly found to be increasing between summer and fall for the two years with available fall images, which allows for a secondary validation of carotenoid estimations for this period. Concerning the high foliar Car during spring 2015, a possible cause would be the still-dying grass during the AVIRIS-C acquisition, due to problems similar to those already mentioned for LAI. Estimated LAI and C_{ab} variations were greater than the estimation RMSE found in Sections 3.3.3.2 and 3.3.3.3, which gives confidence in the ability of the chosen methodology to monitor the site over summer and fall.

While this study used clear-sky images, clouds are often present in hyperspectral images. Obviously too thick clouds hide the ground, and cloud shadows can be hard to detect and correct for, but models to correct for thin cirrus presence in the 0.4–2.5 µm spectral range exist (Gao and Li, 2017; Gao et al., 1998; Salgado et al., 2019) and could be used in case of operational use.

3.5 Conclusion

Estimating canopy pigments and LAI using radiative transfer inversion is challenging in highly heterogeneous environments. This chapter investigated the ability of LUT-based inversion of AVIRIS-C medium resolution (18 m) hyperspectral images to estimate LAI and biochemical properties (C_{ab}, Car) of low-foliage open canopies (mean LAI: 0.82 m^2/m^2, mean CC: 47%) in a woodland savanna. The methodology followed during this study used almost no *a priori* knowledge, so as to be able to monitor the study site in its entirety over multiple dates and obtain more general results. Various inversion criteria were tested, relying either on spectral intervals or hyperspectral vegetation indices. Results from very different site locations in terms of LAI, canopy cover, and tree structure were consistent and showed good accuracy for LAI and leaf C_{ab} and Car retrieval. These results are further strengthened by the temporal aspect of the estimations, with validation done using both summer and fall data from various years. Estimated LAI also showed a temporal pattern coherent with expected seasonal variations for dates when field data were not available.

The results demonstrated that medium resolution (18 m) hyperspectral imagery, combined with a 3D RTM only using a simple geometric representation of the canopy, is appropriate for the monitoring of vegetation bio-physicochemical parameters for low-LAI open broadleaved forests such as woodland savannas.

However, challenges remain: while the LUT generated in this study were appropriate to estimate LAI and leaf pigments, they were not as suitable in the event of heterogeneous understory reflectance (as for Spring 2013 in this chapter). Indeed, understory reflectance was considered to be fixed during LUT building, and accounting for the spatial heterogeneity would require it to be a variable parameter in DART. However, the CPU time required to generate the LUT used in the present chapter was already extremely limiting, so much so that the sampling schemes used to generate the LUT were rather coarse. The addition of more variables in the LUT either for generalization or estimation purposes (if EWT and LMA were to be retrieved), and in general increasing the number of LUT entries, would first require to significantly reduce the LUT building computational effort.

Chapter 4

Densifying LUT Sampling by Developing a Method to Accelerate LUT Building through Joint Use of two RTM

Contents

4.1 Objectives

In many cases, data are insufficiently available to calibrate an empirical-statistical method, and it is necessary to turn to physical or hybrid approaches and rely on RTM. This can prove to be computationally demanding, as a consequent number of simulations could be necessary so that acceptable accuracy in vegetation traits retrieval is achieved. As detailed in Chapter 1, 1D RTM are fast but not well adapted to model heterogeneous canopies. A realistic modeling, using 3D RTM, is of prime importance for the modeling of sparse forests and tree–grass ecosystems that are widely distributed on Earth Hanan and Hill, 2012, as the spectral contribution of the canopy to the total scene reflectance is limited and ground and shadows are more visible to the sensors. However, 3D RTM most commonly rely on ray tracing methods and their complexity can lead to a dramatic increase in computation times, which limits the sampling schemes that can realistically be considered for each variable of interest.

Unfortunately, no single best sampling scheme has been identified so far: Ali et al. (2016a) considered a uniform distribution of the variables over their respective ground-truth ranges when working with INFORM (Atzberger, 2000); Weiss et al. (2000) drew each parameter's values according to a distribution law which was proportional to the reflectance's sensitivity to the parameter; Ali et al. (2016b) used multivariate normal distributions and covariance matrices produced from ground truth data with INFORM ; Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada (2012) used both monovariate and multivariate random samplings with DART. Due to the computation times of 3D RTM, testing multiple sampling schemes when building LUT with tens of thousands of entries is not realistically feasible. Being able to consistently optimize the LUT sampling scheme at low time cost, either for direct LUT-based inversion or subsequent training of a machine-learning model, could therefore prove beneficial.

The aim of this study was to combine the realism of 3D RTM with the speed of 1D RTM to be able to quickly generate LUTs with several variable parameters and arbitrary sampling for estimation purposes. To do so, the PROSAIL (1D) and DART (3D) canopy RTM were considered. Both PROSAIL and DART were used to calibrate a model that approximates DART reflectance outputs from PROSAIL's. The performances of this model (named PROSAIL2DART) were assessed by comparing its outputs with DART reference values. PROSAIL2DART was then used to estimate gap fraction, oak leaf pigment content, and oak LMA and leaf EWT over a woodland savanna. Estimation accuracy was assessed by confronting estimations with field measurements done at various stands and dates.

4.2 Materials and Methods

4.2.1 Materials

The present study focused on Tonzi Ranch and gap fraction, foliar pigments, and EWT and LMA retrievals over multiple years and seasons. Figure 4.1 shows an aerial view of the site and the locations of the various plots used in this study.

As PISA only represent 10% of the overstory, the present study gap fraction plots contained either only QUDO or a QUDO-PISA mix with a QUDO majority. Leaf collection for EWT, LMA and leaf pigment content estimation were done in a pure-QUDO part of the site. Only QUDO were modeled in the RTM, as correctly modeling coniferous trees in 3D RTM can be challenging (Janoutová et al., 2019) and no PISA-dominant plots were included in the study (three stands are mixed and PISA canopy cover represents only 37% of the total canopy cover in the most mixed stand).

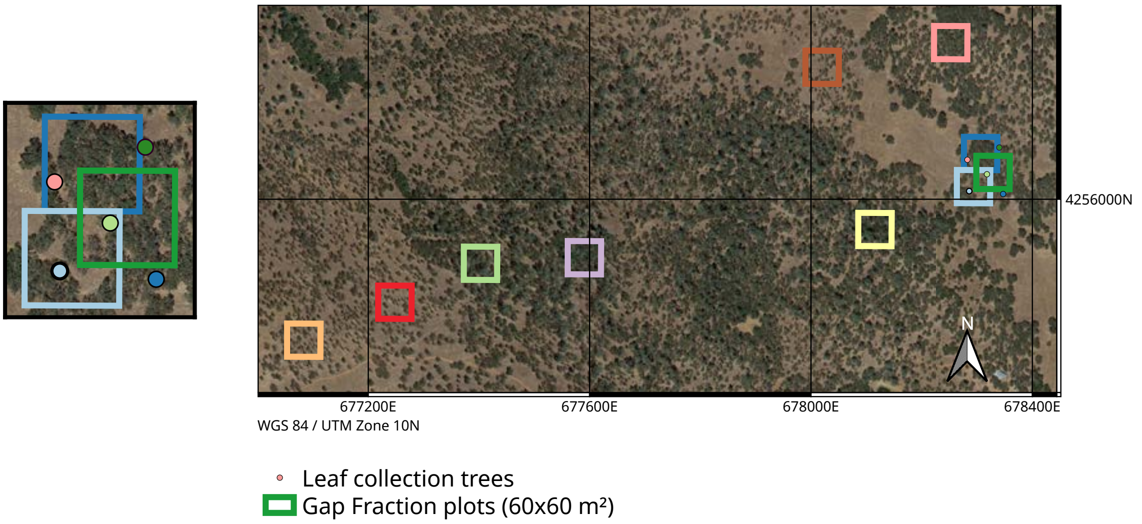


Figure 4.1: Aerial view (right) of the study site and zoom-in on the leaf collection trees (left). Different gap fraction plots and leaf collection trees are identified with different colors.

Information concerning the number of plots and validation points for each date is given in Table 4.1. The set of airborne acquisitions used that were considered in given in Table 4.2.

Table 4.1: Field data used in this study for gap fraction and leaf biochemistry and associated dates of collection. In bold, data used for validation in the present study.

	Field Data	
Date	**gap fraction Plots**	**Biochemistry Trees**
June 2013		**3**
September 2013	**2**	**5**
April 2014		5
June 2014	**8**	**5**
October 2014		5
June 2016	**7**	
Validation data	17	13

Table 4.2: Description of AVIRIS-C airborne acquisitions.

Year	Date (DOY)	Time (PDT)
2013	4 June (155)	12h30 p.m.
	19 September (262)	12h40 p.m.
2014	2 June (153)	12h00 p.m.
2016	9 June (161)	12h30 p.m.

4.2.2 PROSAIL2DART

4.2.2.1 Methodology

Both PROSAIL and DART were used to compute scene reflectances over a calibration (LAI, C_{ab}, Car, LMA, EWT) grid hereafter designated as *cal*. Band reflectance ratios $\frac{\rho_{DART}}{\rho_{PROSAIL}}$ were computed for each point of the grid and used to calibrate linear 5-D interpolators. These interpolators are used to transform any reflectance obtained with PROSAIL into a reflectance similar to what DART would have obtained. For each date, one 5-D interpolator was computed for each (CC, understory reflectance, *Leaf Anthocyanins Content* (C_{ant})) triplet as these parameters were either not possible to model in PROSAIL (CC) or to limit the uncertainties (understory reflectance, C_{ant}). This calibration/transformation method is hereafter called PROSAIL2DART (P2D). A diagram of the methodology is shown in Figure 4.2.

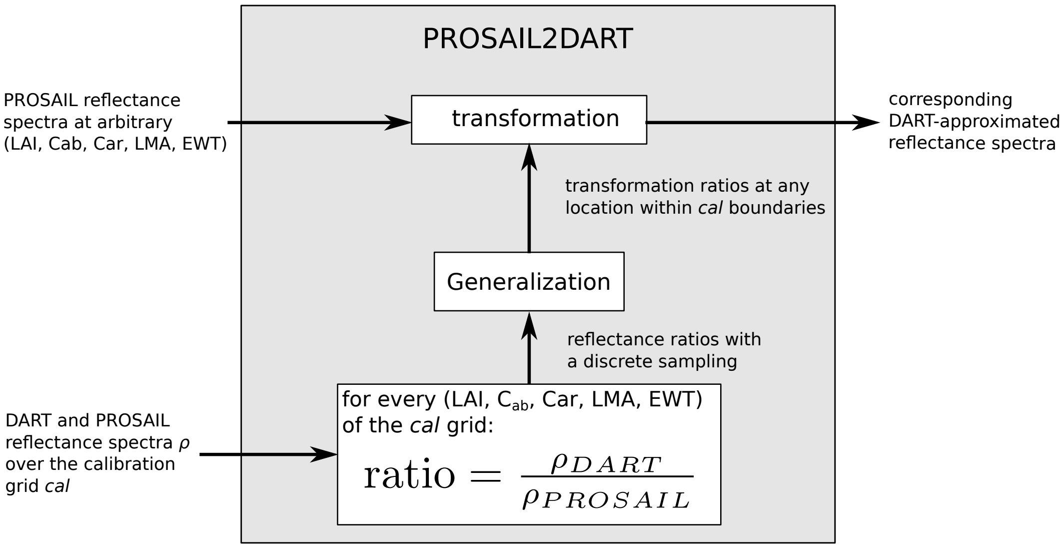


Figure 4.2: PROSAIL2DART methodology.

4.2.2.2 RTM Parametrization

DART DART version 5.7.3v1078 was used to simulate canopy reflectances.

The scene modeling done within DART in this study is based on the same simplified forest representation as the one done in Chapter 3: canopy is represented with 4 lollipop trees and the ground is modeled as a lambertian surface, the reflectance of which was extracted from AVIRIS-C images, as this previously proved sufficient to estimate both LAI and leaf pigment content at the AVIRIS-C spatial resolution. Different soil reflectance were considered: from the sets of pure soil pixels extracted from open parts of the site, mean and mean ± standard deviation reflectances were used to build the LUT in order to better take into account possible ground reflectance variations over the site. For simplicity purposes and to ensure that (i) the (LAI, C_{ab}, Car, LMA, EWT) space that could be covered by P2D was an hyperrectangle and (ii) the density of calibration samples was uniform over this space, *cal* followed a regular sampling scheme. Tables 4.3 and 4.4 describe the various inputs used to create the DART scenes.

The Gap Fractions of the DART scenes were computed for all combinations of CC and LAI. It was retrieved using the DART 3D Radiative Budget tool, by considering the percentage of diffuse illumination intercepted by the ground. The specific DART parameters to obtain these results are

- illumination using a single wavelength,
- no radiative transfer in the atmosphere,
- SKYL (atmospheric scattering of sun radiance) set to 1,
- number of iterations set to 0, and
- smaller mesh size of irradiance sources set to 0.005 m

Gap fraction can be considered a function of CC an LAI, i.e., gap fraction $= f(\text{CC}, \text{LAI})$. Therefore, when generating the P2D, fine LUT Gap Fractions were derived from the (CC, LAI) values by linear interpolation, using the *test* and *cal* values as reference.

Table 4.3: DART-fixed scene parameters used in this study for the calibration and test LUT. PROSAIL used the same LAD and sun zenith/azimuth as DART.

Parameter	**Value**
Voxel size x, y, z (m)	0.4, 0.4, 0.4
Tree height (m)	9.4
Crown shape	ellipsoidal
Crown diameter (m)	5.8
Crown height (m)	7.5
Trunk height (m)	6.63
Trunk dbh (m)	0.26
LAD	spherical
Sun zenith/azimut	according to acquisition date

Table 4.4: DART, PROSAIL, and PROSPECT variable parameters used for the calibration (*cal*) and test (*test*) LUT. For C_{ant}, values between parentheses concern the September 2013 LUT, other values concern all June LUT.

Parameter	**Values/Range**		**Step**	
	cal	***test***	***cal***	***test***
CC (%)	10–90	10–90	20	
LAI (m^2/m^2)	0.1–1.9	0.25–1.75	0.3	
C_{ab} ($\mu g/cm^2$)	10–60	15–55	10	
Car ($\mu g/cm^2$)	2–14	4–12	4	
LMA (g/cm^2)	7–16 $\times 10^{-3}$	8.5–14.5 $\times 10^{-3}$	3×10^{-3}	
EWT (cm)	5–17 $\times 10^{-3}$	7–15 $\times 10^{-3}$	4×10^{-3}	
C_{ant} ($\mu g/cm^2$)	0(0–2)	0	0(2)	0
Ground Reflec.	mean, mean ± Std	June 2016 mean		

PROSAIL The 4SAIL version of PROSAIL was used in the present study, using a Python wrapper (`https://github.com/jgomezdans/prosail`, DOI: 10.5281/zenodo.2574925). PROSAIL combines the leaf model PROSPECT with the 1-D turbid canopy RTM SAIL. A thorough description and history of the PROSAIL model can be found in Berger et al., 2018. LAD is not a direct input of PROSAIL and both the average leaf slope LIDFa and the associated distribution bimodality LIDFb must be given. A spherical LAD is obtained with LIDFa $= -0.35$ and LIDFb $= -0.15$. Ground reflectance was the same reflectance as that given to DART, as were the solar zenith and azimuth angles. LAI variations were the same as those given to DART.

PROSPECT Leaf optical properties were simulated using the PROSPECT model, which is implemented in DART and PROSAIL. PROSPECT-D was used in this study, and a small C_{ant} was introduced as a possible case for September to take into account possible leaf senescence. The leaf structure parameter N was set to 1.8. The PROSPECT specific parameters considered in this study are given in Table 4.4.

4.2.2.3 Error Assessment

As the P2D linear interpolators were calibrated on a regular grid, the maximum differences between P2D and DART reflectances are located at the centers of the hypercubes defined by the *cal* grid. The P2D approach will be validated if the difference between P2D and DART reflectances at the hypercube center are negligible.

Let *test* designate the set of the hypercubes centers, which correspond to the *test* grid. Let $P2D^{test}$ and D^{test} be the reflectances computed by P2D and DART on *test* points and D^{cal} the reflectances computed by DART on *cal* points, which correspond to the hypercubes corners. The P2D approximation was evaluated using the E ratio, computed as

$$E = \underset{i}{\text{median}} \frac{\left|P2D_i^{test} - D_i^{test}\right|}{\underset{j}{\min}\left|D_i^{test} - D_{i,j}^{cal}\right|_{>0.001}} \times 100 \tag{4.1}$$

with i the hypercube identifier and j the hypercube corner identifier. This ratio was designed to compare the reflectance distance between P2D and DART at the hypercube center (maximum error) with the reflectance distance between the hypercube center and corners obtained through DART. A value close to 100 indicates that the P2D error is similar to the smallest difference between the hypercube's center and corners, while a value close to 0 indicates that the P2D error is negligible. An illustration of the P2D validation methodology is presented Figure 4.3. An E value lower than 50 indicates that the P2D approximation is closer to the hypercube center than the corners. A condition that $\left|D_i^{test} - D_{i,j}^{cal}\right|$ should be non-negligible (>0.001 when reflectance range is 0 to 1) was used as variables do not necessarily have an influence at all wavelengths, which could lead to close to zero differences between some corners and the center and make E diverge erroneously.

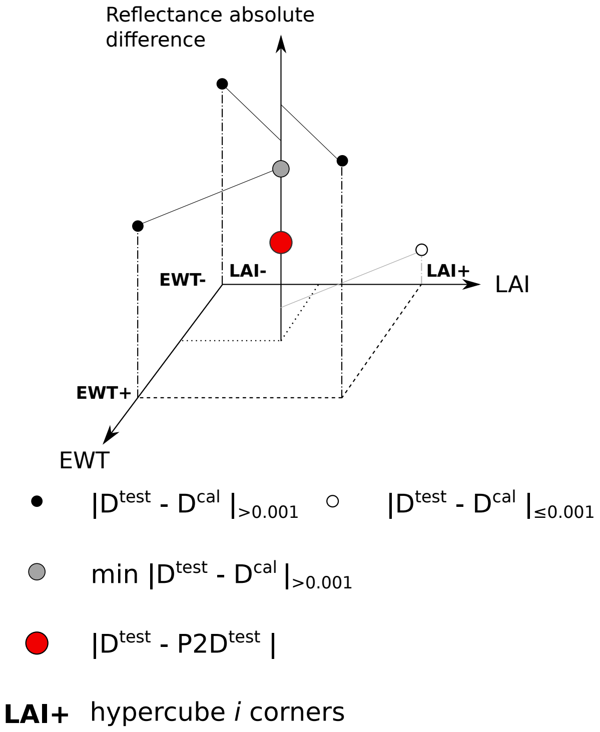


Figure 4.3: Illustration of the P2D validation methodology on a 2D grid, with calculation of the various differences necessary to compute E for one hypercube.

4.2.3 Fine LUT Building

P2D was subsequently used to generate a fine DART-like LUT for each AVIRIS-C image, following a Latin Hypercube sampling. The correlation between C_{ab} and Car visible in the field data was taken into account by constraining the Car values around 2.5 times the standard deviation between field data and the regression line (see Figure 4.4). For LAI, C_{ab}, LMA and EWT are the boundaries of the Latin Hypercube corresponding to those of the calibration LUT presented in Table 4.4. Only CC 30 to 90% were considered, and for every ground reflectance and C_{ant} value, 50,000 cases were generated and distributed equally among the CC. Therefore, June 2013, 2014, and 2016 fine LUT each have 150,000 entries, and September 2013 LUT has 300,000 entries.

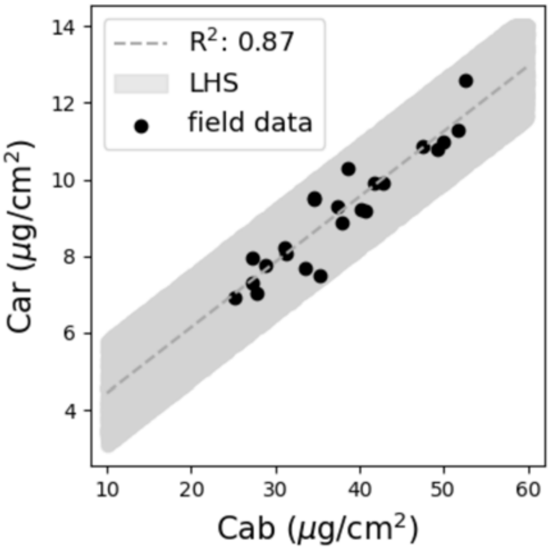


Figure 4.4: Relationship between Car and C_{ab} in the field data (dots) and as implemented (shaded area) with the Latin Hypercube sampling.

4.2.4 LUT-Based Inversions

To assess the performance improvement when using P2D instead of simply using DART with a regular sampling scheme, both DART *cal* and P2D fine LUT were used to retrieve gap fraction and leaf biochemistry. LUT-based approaches consist in finding the simulated reflectance $\hat{y}$ that is the most similar to the measured one, y, according to a cost function. Several cost functions were selected for this study: RMSE (Equation (4.2)), SAM (Equation (4.3)), and VI differences (D_{VI}; Equation (4.4)).

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4.2}$$

$$SAM(y,\hat{y}) = cos^{-1}\left(\frac{\Sigma_{i=1}^{N} y_i \hat{y}_i}{\sqrt{\Sigma_{i=1}^{N} \hat{y}_i \hat{y}_i}\sqrt{\Sigma_{i=1}^{N} y_i y_i}}\right) \tag{4.3}$$

$$D_{VI}(y,\hat{y}) = abs\left(VI(y) - VI(\hat{y})\right) \tag{4.4}$$

RMSE and SAM were computed using variable-specific spectral intervals. The gap fraction interval covered the NIR and SWIR (INT GAP, 0.8–2.45 μm). Intervals for C_{ab} and Car were parts of the visible range (INT CAB, 0.5–0.75; INT CAR, 0.5–0.55 μm), while those used for LMA and EWT were parts of the NIR and SWIR (INT LMA, 0.8–1.3; INT EWT, 1.3–2.45 μm). The spectral intervals were chosen based on their sensitivity to the variables of interest according to the results of a Sobol sensitivity analysis on the DART *cal* LUT (not shown).

Before LUT-based inversion, VI capabilities to estimate the variables of interest were assessed by fitting a function between VI and variables ($VI = v^a + b$ for biochemistry and $VI = a \times v + b$ for gap fraction, with v the variable's value). If no relationship could be found, the VI was not considered for the inversion.

Table 4.5 shows all the cost functions tested for this study for each variable of interest, as well as the goodness of the fit of each VI when applicable. Estimation results were computed as the mean of multiple best solutions. The number of solutions considered for each LUT was 0.5% of the LUT size.

4.2.5 Validation Metrics

Gap fraction, leaf pigments content, EWT, and LMA estimates were compared with the field measurements available using the following criteria; total RMSE, systematic and unsystematic RMSE (Willmott, 1981), the model performance index d_r (Willmott, Robeson, and Matsuura, 2012), and R^2 of the predicted-vs.-measured regression line.

Concerning gap fraction, each validation point was the average of gap fraction values derived from hemispherical pictures taken within a 60 m × 60 m plot (see Section 2.3.1). Direct comparison between pixel-estimated value and validation data can be inappropriate as the area covered by the DHP of a plot is wider than the AVIRIS-C pixel (18 m × 18 m). Therefore, validation values were compared to the average of the Gap Fractions estimated over a 3 × 3 pixels windows centered on the pixel corresponding to the plot center, similarly to the method employed in Chapter 3.

Biochemistry validation data were obtained at the leaf scale, for one tree in each validation pixel. It was assumed that biochemistry estimations could be directly extracted from the pixels associated with the acquisition positions.

Table 4.5: Cost functions used in this study, best R^2 obtained across all LUT for the VI and associated performances for the LUT-based (DART and P2D) inversions (when applicable).

	VI Source	**VI R^2**	**d_r**	
			DART *cal*	**P2D fine LUT**
Gap fraction				
RMSE INT GAP			0.75	0.75
SAM INT GAP			0.76	0.76
D_{NDVI}	Tucker, 1979	0.96	0.78	0.77
D_{MSAVI2}	Qi et al., 1994	0.9	0.76	0.74
C_{ab}				
RMSE INT CAB			0.49	0.31
SAM INT CAB			0.12	0.23
$D_{TCARI/OSAVI}$	Haboudane et al., 2002	0.81	0.25	0.23
$D_{Maccioni}$	Maccioni, Agati, and Mazzinghi, 2001	0.91	0.57	0.69
D_{gNDVI}	Smith et al., 1995	0.76	0.65	0.75
D_{GM_94b}	Gitelson and Merzlyak, 1994	0.6	0.65	0.77
D_{MCARI2}	Haboudane et al., 2004	0.01		
Car				
RMSE INT CAR			-0.08	0.2
SAM INT CAR			-0.37	0.53
D_{R515_R570}	Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012	0.29		
D_{CRI550}	Gitelson et al., 2006	0.09		
$D_{TCARI/OSAVI}$	Haboudane et al., 2002	0.72	0.04	0.27
$D_{Maccioni}$	Maccioni, Agati, and Mazzinghi, 2001	0.81	0.12	0.59
D_{gNDVI}	Smith et al., 1995	0.68	0.11	0.64
D_{GM_94b}	Gitelson and Merzlyak, 1994	0.61	0.34	0.65
D_{MCARI2}	Haboudane et al., 2004	0.01		
LMA				
RMSE INT LMA			0.03	0.19
SAM INT LMA			0.29	0.24
D_{lma_ND}	Maire et al., 2008	0		
D_{lma_D}	Maire et al., 2008	0.5	-0.34	0.1
D_{NDNI}	Serrano, Peñuelas, and Ustin, 2002	0		
D_{NDLI}	Serrano, Peñuelas, and Ustin, 2002	0		
EWT				
RMSE INT EWT			-0.32	0.04
SAM INT EWT			-0.49	0.29
D_{EVI}	Huete et al., 2002	0		
D_{NDWI}	Gao, 1996	0.01		
D_{SIWSI}	Fensholt and Sandholt, 2003	0.02		
D_{NDI7}	Trombetti et al., 2008	0.01		
D_{NDII}	Hardisky, Klemas, and Smart, 1983	0		
D_{SRWI}	Zarco-Tejada, Rueda, and Ustin, 2003	0.01		
D_{MSI}	Hunt and Rock, 1989	0		
D_{SWIRR}	Trombetti et al., 2008	0.01		
D_{WI}	Penuelas et al., 1993	0.01		

4.3 Results

4.3.1 Comparison between AVIRIS-C and DART Reflectances

Figure 4.5 shows the validation pixels' reflectances and compares them to the reflectance extrema found in their corresponding LUT. For June 2013, September 2013 and June 2016, all pure pixel reflectances fall within the extrema of the LUT whatever the wavelength. One reflectance from a mixed plot is severely out of the boundaries of the LUT, while up to 8 other mixed plot reflectances are slightly below the LUT minima between 0.9 and 1.6 μm, for a total of at most 9 reflectances below the LUT minima at some wavelengths out of 63 for June 2016. For June 2014, almost all reflectances including those from mixed pixels also fall within the extrema: of the 68 pixel reflectances, 6 are below the LUT minima around 1.24 μm and 1.6 μm, with a maximum difference of 0.02.

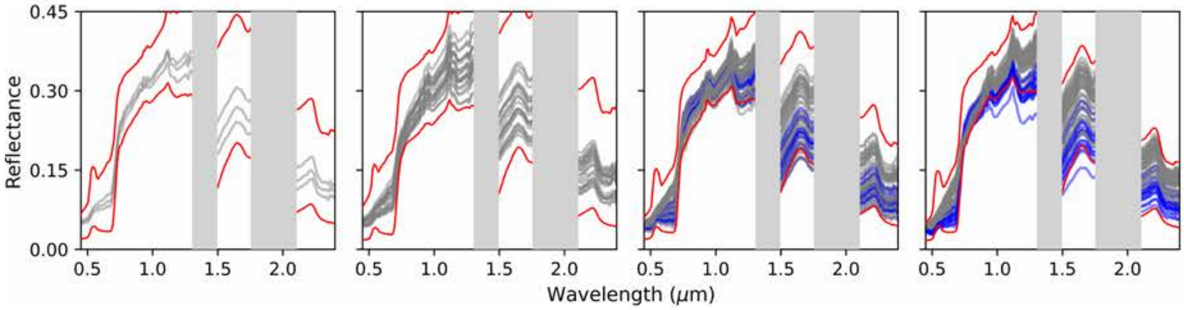


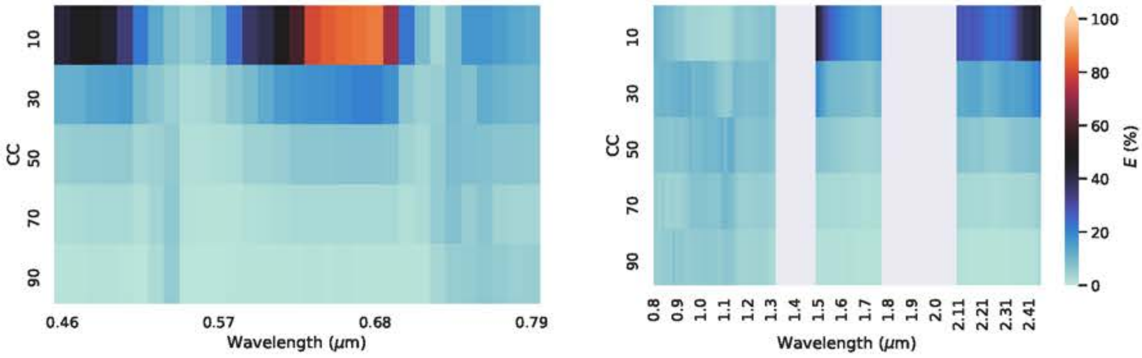
Figure 4.5: LUT reflectances and AVIRIS-C reflectances at validation pixels for each date (from left to right: June 2013, September 2013, June 2014, June 2016). In red, reflectance boundaries of the associated LUT, in gray, AVIRIS-C reflectances. In blue, AVIRIS-C reflectances from mixed QUDO-PISA validation pixels.

4.3.2 PROSAIL2DART Errors

Figure 4.6 and Table 4.6 show the evolution of the E ratio over the CC and wavelengths. In the visible, all wavelengths are well approximated by P2D for CC $\geq 30\%$, with the highest E value being 21% at 0.68 μm and 30% CC. While for 10% CC, the green and NIR are also well approximated ($E < 50\%$), this is not the case for the blue and red regions where E can be above 50%. In the SWIR, for 10% CC E values are considerably below 50% for $\lambda < 1.8$ μm. However, higher values are found at higher wavelengths and the maximum, 46%, is obtained at 1.49 μm. Estimations with either DART *cal* or P2D LUT only used the CC $\geq 30\%$ cases to avoid uncertainties.

4.3.3 PROSAIL2DART Fine LUT Generation

It took 12,666 h (total CPU time of a server equipped with Broadwell Intel^{ő} Xeon^{ő} CPU E5-2650 v4 @ 2.20 GHz) to generate the 21,840 reflectances required to build the DART *cal*



(a) (b)

Figure 4.6: P2D error for each wavelength and CC over the (**a**) visible and (**b**) NIR spectral regions. Panels (**a**,**b**) share the same color bar.

Table 4.6: Maximum value of the E ratio and associated wavelength for each CC over the visible (VIS) and NIR spectral ranges.

CC	**Maximum E (%)**		**Wavelength (μm)**	
	VIS	**NIR**	**VIS**	**NIR**
10	87	46	0.68	1.49
30	21	21	0.68	2.45
50	9	13	0.73	1.13
70	8	10	0.73	1.10
90	6	8	0.76	0.88

LUT dedicated to September 2013 (the most extensive LUT, as two values of anthocyanins are also considered). For comparison, once P2D was calibrated (the calibration time is negligible), it took 1.5 h (total CPU time on a computer equipped with an Intel$^{\text{ő}}$ Core™ i5-6300HQ CPU @ 2.30 GHz) to generate the 300,000 entries of the P2D fine LUT.

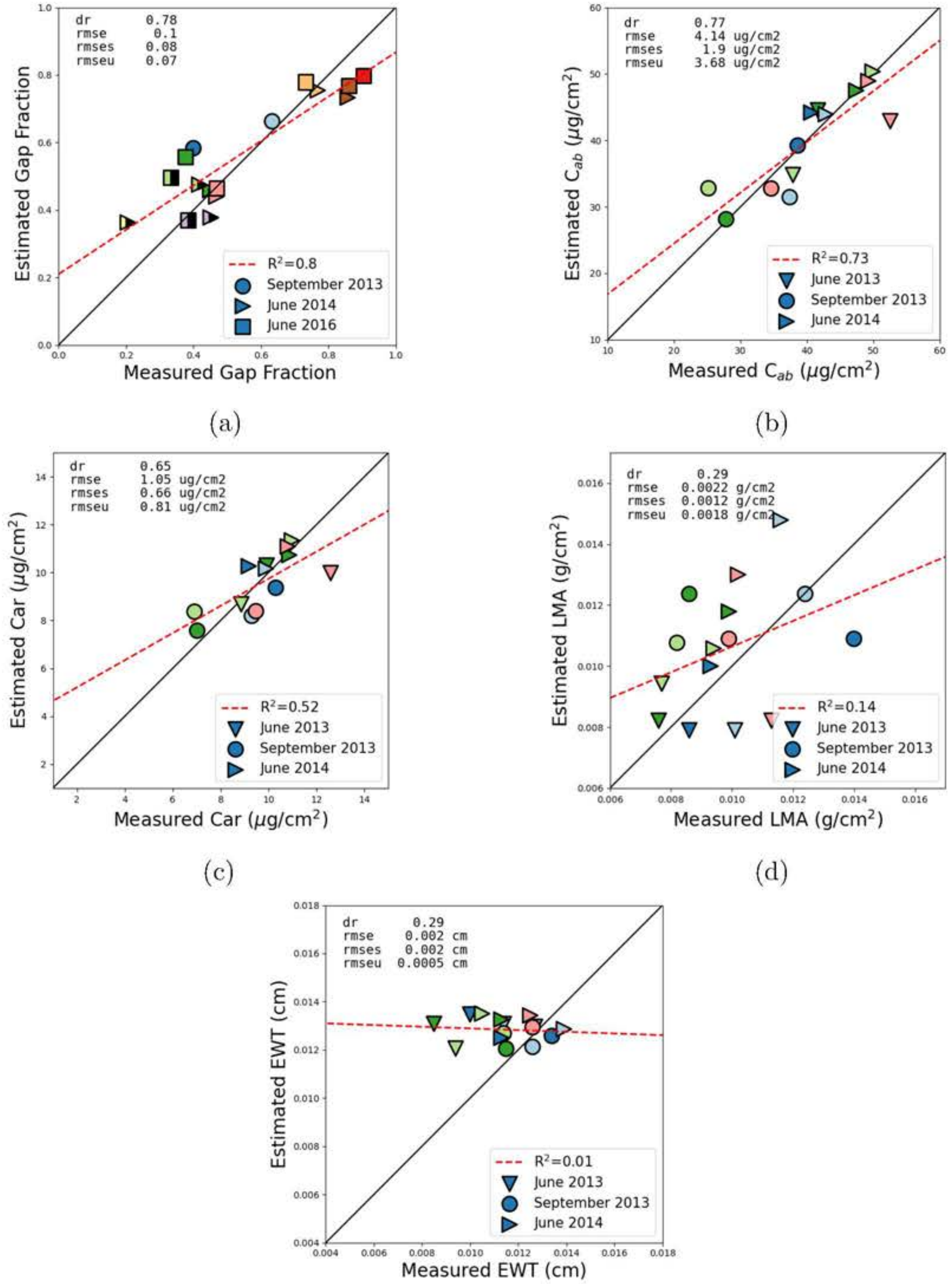
4.3.4 Estimation Performances

Table 4.5 shows the best R^2 achieved by the various VI when fitted over each LUT. Gap fraction was very well measured by its VI, and overall so were C_{ab} and Car, with only MCARI2, R515_570, and CRI550 presenting no relation with the estimates' values. Concerning LMA and EWT, no satisfying relation could be found, and only lma_D could find a slight relation with a R^2 of 0.5. Only the inversion methods with R^2 higher than 0.5 were considered suitable candidates for LUT-based inversions.

Concerning gap fraction, both LUT perform in equivalent manner, with good performances whatever the method (maximum d_r is 0.78 for D_{NDVI} applied on DART *cal* and 0.77 when applied on the P2D LUT). C_{ab} estimations show improved performances with the P2D LUT for all cost functions except RMSE INT CAB and $D_{TCARI/OSAVI}$, which have slightly lower d_r. D_{GM_94b} offers the best performances, with $d_r = 0.77$ when applied on the P2D LUT. For Car, the best d_r is also obtained with the P2D LUT with D_{GM_94b}, and P2D consistently improved the d_r. Concerning LMA and EWT, the P2D fine LUT appear to slightly improve performances for the selected cost functions (at the exception of SAM INT LMA that decreases slightly); however, d_r remains low. Their respective best-performing cost functions are SAM INT LMA with DART *cal* and SAM INT EWT with P2D.

4.3.5 Estimation Plots

Figure 4.7 compares estimated and field values for the various estimates of this study when using the best performing methods identified in Section 4.3.4. Gap fraction estimations present both a high R^2 and a low RMSE (0.78 and 0.1, respectively). While it appears that one of the June 2014 mixed plots (yellow) was overestimated, other mixed plots present a similar behavior as pure QUDO plots. Similar behaviors are found for C_{ab} and Car: most of the points seem to follow the first bisector, and the point with the highest C_{ab} and Car values is slightly underestimated. RMSE errors in both cases remain low (4.14 and 1.05 µg/cm^2, respectively). No trend between estimations and field data could be found for either LMA and EWT (their respective R^2 are 0.14 and 0.01) and RMSE of 2.2×10^{-3} g/cm^2 and 2×10^{-3} cm are obtained.



(a) (b)

(c) (d)

(e)

Figure 4.7: Comparison between estimated and measured parameters using the best-performing methods identified in Table 4.5. Panels (**b**,**c**,**e**) were obtained with the P2D fine LUT, and panels (**a**,**d**) were obtained with the DART calibration LUT. Marker color identifies the location within the study site, and for gap fraction QUDO-PISA mixed stands are identified with bicolored markers.

4.4 Discussion

4.4.1 Prosail2dart Performances

Choosing a sampling method for the variables' space for LUT generation, either for LUT-based inversions or calibration of machine-learning methods, is a critical step. No ideal sampling has yet be determined: Darvishzadeh et al., 2011 and Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada, 2012 used uniform distributions over the variables' intervals; Malenovsky et al., 2007 and Leonenko, Los, and North, 2013 considered a regular grid for LAI; and Richter et al., 2011 and Verrelst et al., 2012 used both uniform and Gaussian distribution for the parameters of interest. These methods can lead to different estimation accuracies, as they may put an emphasis on unadapted ranges or allow for the existence of duplicate or near-duplicate cases. However, as LUT-building is very time-consuming when using 3D RTM, it is difficult to test different sampling methods.

As shown in Figure 4.3, the intrinsic error of the P2D approximation is very low for all wavelengths at CC 30% and gets lower when the CC increases, as the higher the CC the closer a DART scene is to a completely turbid medium. The error E calculated in this study is also relative to the *cal* LUT's sampling, so lower absolute errors could be obtained by simply refining the calibration sampling. This low intrinsic error made it possible to approximate with minimal error DART outputs using the PROSAIL model, which has considerably shorter computation times. Each interpolator necessitated a total of 784 (672 in the visible and 112 in the NIR and SWIR) DART reflectances. As specified in Section 4.3.3, the most extensive calibration LUT, encompassing 21,840 cases, necessitated 12,666 CPU hours to generate with DART 5.7.3v1078 and allowed for calculating 300,000 entries in 1.5 h, which is a significant decrease of the computation time. Even though the latest DART versions are significantly faster than the version used in this study, the execution time remains considerably longer than P2D as P2D is basically as fast as PROSAIL.

This short execution time makes it possible to test various sampling schemes and either use the output reflectances as is, as the P2D error is small, or determine an optimal sampling scheme to use for final LUT generation with the 3D RTM.

4.4.2 Gap fraction Estimations

Table 4.5 and Figure 4.7a show that gap fraction estimations are good, with high d_r and low RMSE, and that this is true no matter the estimation strategy. Indeed, all methods presented a d_r higher than 0.7, a RMSE lower than 0.11, and a R^2 higher than 0.73 (not shown), and using the P2D LUT did not lead to significant improvement of the estimations. The method presenting the highest d_r was D_{NDVI}. When estimating scene LAI over the site, the study done in Chapter 3 also identified this method as the best performing. The LAI had been obtained from the same DHP pictures, with the assumption that effective plant area index (PAI) was equivalent to true LAI (Fang et al., 2019). As effective PAI is derived from the

DHP gap fraction, the fact that the same inversion strategy yielded good estimation results for both gap fraction and true LAI may confirm that true LAI can be considered equivalent to effective PAI for sparse broad-leaved forests.

4.4.3 Pigment Estimations

C_{ab} and Car estimations (RMSE 4.14 and 1.85 µg/cm^2 and R^2 of 0.73 and 0.52, respectively) are in line with what can be found in the literature: Zarco-Tejada et al., 2004 obtained a RMSE of 8.1 µg/cm^2 for C_{ab} over open-canopy tree crops; Maire et al., 2008 obtained a RMSE of 8.2 µg/cm^2 estimating leaf C_{ab} of broadleaved forests; Darvishzadeh et al., 2019 had an RMSE of 8.6 µg/cm^2 when estimating leaf C_{ab} of spruce stands; Zarco-Tejada et al., 2013 obtained a 1.3 µg/cm^2 RMSE for Car estimation in vineyards with high-resolution imagery; Huang et al., 2018 had a 2.02 µg/cm^2 RMSE when monitoring crop Car.

While final C_{ab} estimation performances obtained in this study are similar or better than those obtained in Chapter 3, there are two main differences between the two methodologies: The first one is that the previous study used PROSPECT-5 and not PROSPECT-D: the principal differences between PROSPECT-D and -5 are the introduction of anthocyanins and the modification of the pigments specific absorption coefficients (SAC) at various wavelengths (Féret et al., 2017): the chlorophylls SAC were lowered over the 0.45–0.65 µm interval and increased over 0.65–0.7 µm, and the carotenoids SAC were increased over the 0.45–0.55 µm interval. The second difference is the sampling scheme, which was previously a regular grid over the variables' variation ranges, meaning that no correlation between C_{ab} and Car, such as the one visible Figure 4.4, could be modeled. While this was not critical for C_{ab} estimations, as the high d_r values in Table 4.5 show, this limited the number of possible VI to use for Car estimations. Introducing the relationship between C_{ab} and Car when building the LUT allowed to use C_{ab} VI for Car estimations, which proved to be more adequate. This relationship also helps to reduce the unnecessary cases in the LUT, as low C_{ab}-high Car cases (and vice versa) are unrealistic and could bring confusion when doing the inversions.

4.4.4 LMA and EWT Estimations

The RMSE obtained for LMA was 0.0022 g/cm^2, more than two times higher than the 9.1×10^{-4} g/cm^2 value obtained by Maire et al., 2008 over a higher LAI broadleaved forest. EWT estimations present a 0.002 cm RMSE. Neither LMA nor EWT present good R^2. As visible in Table 4.5, almost none of the VI tested for these variables showed a relationship with them. This is in line with the results found by Yanfang Xiao et al., 2013, who showed using PROSAIL that for LAI lower than 3 it was not possible to estimate EWT as its contribution to the signal was too low in comparison to LAI's and ground's. It is possible that higher resolution hyperspectral images would be needed for EWT and LMA estimation, as this would make it possible to locate pure-vegetation pixels where EWT and LMA spectral signature would be more visible.

4.5 Conclusion

The results obtained in this chapter demonstrated the possibility to approximate with minimal error the reflectance outputs of DART with those of PROSAIL even at low (30%) CC. For higher CC, it was shown that approximation errors were negligible. The approximation model was further used to generate extensive LUT to estimate gap fraction of mixed oak and pine stands as well as leaf C_{ab}, Car, EWT, and LMA of oak stands in a low-foliage woodland savanna. Gap fraction and leaf pigment content estimations presented similar or improved performances when taking advantage of the proposed model instead of only relying on DART. EWT and LMA could not be retrieved using either models.

In summary, the findings showed that acceptably approximating DART results from PROSAIL was possible and that the subsequent reflectances could be successfully used for estimation purposes of even very sparse oak stands, although conclusions should also be applicable to other broadleaved stands due to the elementary modeling used in the 3D RTM. This is valuable, as 1D RTM are dramatically faster than 3D RTM. In the exploration phase, this allows for the testing of various sampling schemes at a negligible cost for either the training of machine learning methods, that require extensive training databases, or the generation of more complex LUT. Approximated reflectances can also directly be used as-is to retrieve canopy structural and biochemical parameters with acceptable accuracy.

Due to the tree distribution within the study site and the ground sampling distance of AVIRIS-C, no pine-dominant stands could be considered for gap fraction and leaf biochemistry estimations and this study focused mainly on pure-oak stands. Further work is necessary to possibly extend them to coniferous trees or mixed stands. More work is also necessary to acceptably estimate EWT and LMA of tree–grass ecosystems. This chapter showed that the LUT sampling was not at cause for the inability of the proposed method to estimate EWT and LMA. It was therefore probable that the SFR modeling, while appropriate to retrieve LAI, gap fraction and foliar pigment contents, was not adapted for EWT and LMA retrievals. Further refinements of the DART modeling may be required, possibly by improving the understory representation by modeling the grass layer Melendo-Vega et al., 2018 or the tree representation with the inclusion of detailed trunk structures Widlowski et al., 2013 within the 3D RTM.

Chapter 5

Identifying an Appropriate Method to Estimate EWT and LMA from DART Modeling

---

Contents

5.1 Objectives

As explained in Chapter 1, EWT and LMA are important traits for the description of ecosystem processes and functions that can be used to infer essential biodiversity variables (Jetz et al., 2019; Skidmore, 2013) such as photosynthetic activity and growth rate, and these are recognized indicators of plant health status (Penuelas et al., 1993, 1997; Schimel et al., 2015; Serbin et al., 2019; Westoby et al., 2004). These traits are also important in fire prevention efforts, to determine the FMC, a common indicator used in fire risk assessment (Roberts et al., 2006; Yebra et al., 2013).

The various RTM available for physically-based and hybrid methods can be roughly categorized in two families. 1D homogeneous models, describing the vegetation as a turbid medium, are the most extensively used (Berger et al., 2018). They are easily configurable and fast, at the cost of realism. 3D heterogeneous models employ ray-tracing methods. These 3D models are highly complex and come at a high computational cost, but are much more realistic. These RTM are more appropriate to model heterogeneous canopies, which present a complex architecture that is all the more important when working at high spatial resolutions. However, due to insufficient field knownledge, trees are often modeled through abstractions, which may lead to inappropriate RTM simulations as these simplifications may affect scene and crown reflectance: Widlowski, Côté, and Béland (2014) assessed the effects of omitting or simplifying the woody structure of the trees within DART, and found that this could result in a significant bias in the BRF that increased as the spatial resolution decreased; Ali et al. (2016a) found that stand height and crown diameter were two of several parameters significantly affecting canopy reflectance; Janoutová et al. (2019) compared the adequacy of various modeling methods in a RTM to lead to realistic spruce crown reflectance. Wocher et al. (2018) used a hybrid method to estimate canopy EWT, calibrating their model over PROSAIL outputs: they found that their initial model had to be empirically adjusted to account for radiative effects not modelled in the RTM before proving acceptable results. Overall, making simplifications to model trees or leaves within a RTM may lead to significant reflectance variations over some part of the reflectance spectrum. To what extent does this affect resultant EWT and LMA estimates? Are there spectral ranges that remain mostly unaffected by the abstractions? These issues are of high importance to design methods to estimate EWT and LMA with few *a priori* information about the sites of study and that could be generalized to sites belonging to the same ecosystem type.

The objective of the present chapter was to evaluate the effect tree modeling abstractions within a RTM on crown reflectance to identify spectral ranges that are not very sensitive, and to assess these findings by training machine learning models on RTM-generated data to estimate EWT and LMA. The DART model (Gastellu-Etchegorry et al., 2015; Sensing, Environment, and Iii, 2016) is a typical 3D RTM that allows for various levels of detail in the modeling, and was therefore chosen for the present chapter. To do so, a preliminary analysis was done using synthetic images to compare the performances of various modelings within a RTM and serve as reference, followed by a real-case study involving two different Mediterranean-climate sparse forests. EWT and LMA estimation accuracies were assessed by comparing estimated values to field measurements, and their consistency was validated using the results from the synthetic analysis.

5.2 Materials and Methods

5.2.1 Materials

The study detailed in this chapter considered both Tonzi Ranch and San Joaquin. Figure 5.1 shows aerial views of the sites as well as the location of the field measurements used in this chapter. It focused specifically on broad-leaved trees, using the classification maps described in Section 2.6. Statistics concerning the field measurements used in the present chapter are given in Table 6.2. Airborne acquisitions with AVIRIS-NG took place on June 6th 2014 at 11h24 *Pacific Standard Time* (PST) over TZ, and on June 11th 2014 at 13h00 PST over SJER.

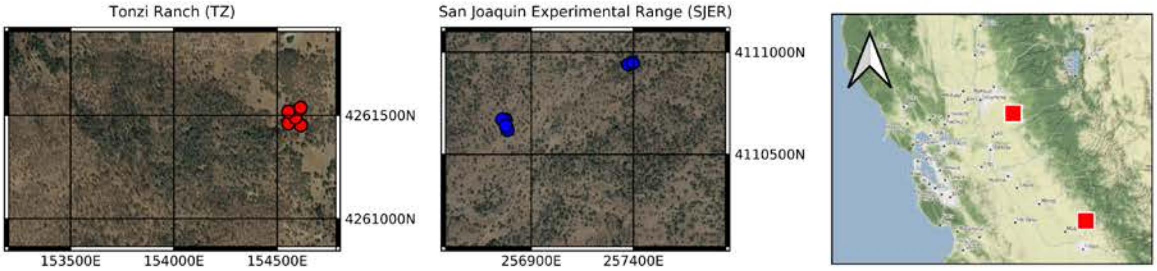
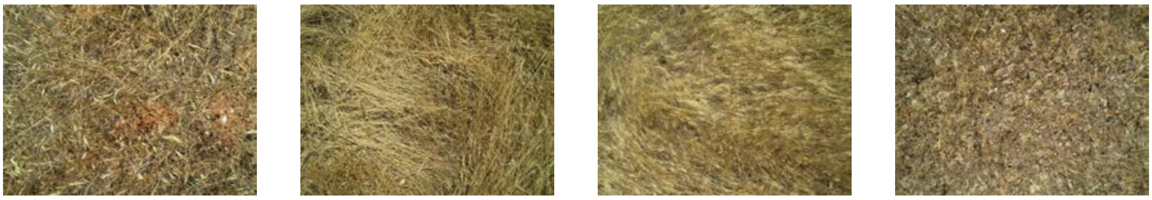


Figure 5.1: Aerial view of Tonzi Ranch (red markers) and San Joaquin Experimental Range (blue markers). Location of the trees where leaf collection took place is indicated by the colored markers.

Table 5.1: Field data collected over TZ and SJER at the time of the AVIRIS-NG flights for QUDO and QUWI.

	Sample No.	Min.	Max.	Mean.
EWT (g/cm^2)	12	0.011	0.018	0.013
LMA (g/cm^2)	12	0.009	0.015	0.011

Reflectance measurements for ground types and tree trunks took place on both TZ and SJER with an Analytical Spectral Device (ASD; ASD Inc., Boulder, CO, USA). A spectralon panel served for calibration before each acquisition. Different soil types were measured so as to ensure that the spatial variability of both sites was taken into account. These ground types included, but were not limited to: PISA litter, QUDO litter, mixed dry herbaceous layer (see Figure 5.2 for illustrations of the various ground types considered in this chapter). Trunk reflectances were obtained from portions of the trunks collected and put on a horizontal surface to facilitate the measurements. All reflectances were obtained over the 0.350–2.500 µm spectral range.



(a) *Avena* (b) *Brachypodium* (c) *Bromus* (d) Low species

(e) QUDO litter (f) QUWI litter (g) PISA litter (h) Senesced grass

Figure 5.2: Illustration of the various ground types that were encountered and used for background spectra measurements. Figures (**a**), (**b**), (**c**), (**d**) were taken on TZ, while Figures (**e**), (**f**), (**g**), (**h**) were taken on SJER.

5.2.2 Trunk and Branches Structure

The lidar point clouds acquired on the site (see Section 2.3.4) were coregistered using CloudCompare 2.6.3.1 (`www.cloudcompare.org`, accessed 23 July 2019) and several trees were isolated and trimmed so that the majority of the remaining points originated from the woody structure. Trimming was done in two steps: first, a Statistical Outlier Removal filter was applied, as the points corresponding to foliage followed no structure; then, connected components were segmented and those visually corresponding to foliated parts of the tree were manually removed. The point clouds were then processed using TreeQSM v2.3.0 (Raumonen et al. (2013), DOI: 10.5281/zenodo.844626) to generate a 3D woody structure through fitting of multiple circular cylinders, up to the fourth order of branches when possible. Six different woody structures were reconstructed for QUDO and were assumed to be representative of the *Non Photosynthetic Vegetation* (NPV) of both QUDO and QUWI trees. One of these woody structures is shown in Figure 5.3 as an example.

5.2.3 Radiative Transfer Modeling

DART 5.7.8 model was used to simulate canopy reflectances. Leaf optical properties were simulated within DART using the PROSPECT-D (Féret et al., 2017) version of PROSPECT (Jacquemoud and Baret, 1990). Unless stated otherwise, the sampling scheme for every DART and PROSPECT parameter was done according to a latin hypercube and the DART pixel size was 40 cm.

DART was used to generate a synthetic hyperspectral image of a sparse forest, and multiple databases were subsequently used to trained random forest regressors (RFR). These databases


(a) front

(b) side

Figure 5.3: Illustration of one of the tree non-photosynthetic structure obtained from the lidar acquisitions.

are detailed in the following sections:

- Section 5.2.5, for the sensitivity analysis:
  - **SFR–Reference** database, with a simplified forest representation (SFR) serving as reference;
  - **variation** database, with the variation cases.
- Section 5.2.6, to train the RFR dedicated to the synthetic image:
  - **SFR** database, with an SFR modeling;
  - **DETAIL** database, with a detailed modeling.
- Section 5.2.6, to train the RFR dedicated to the AVIRIS-NG images:
  - **TZ** database, with an SFR modeling, dedicated to TZ;
  - **SJER** database, with an SFR modeling, dedicated to SJER.

5.2.4 Synthetic Images Generation

The use of synthetic images for this chapter was driven by two main reasons: (i) to overcome the insufficient field validation data available for this chapter by adding synthetic validation data and (ii) to give an idea of the estimation performances in controlled conditions, e.g. when all the field data is available to properly configure the RTM, and how the performance degrades with incomplete modeling.

The synthetic forest scenes were procedurally generated in DART. To do so, for each scene, up to a thousand trees whose characteristics (tree height, crown diameter, LAI, leaf biochemical content, trunk and branchs 3D model) were randomly determined were placed over a 240 × 240 m^2 area such that the tree crowns would not overlap. To avoid crown overlap due to the scene repetition done by DART, no tree was present in a 20 m band on the outskirts of the scenes. Tree crowns were modelled as semi-ellipsoids, with 60% empty crown voxels. The tree LAI (ratio of the sum of the upper surfaces of the leaves over the projected surface of the crown) distribution within the crown depended on elevation: crowns were divided into three equal-sized parts (bottom, middle and top) and each part's LAI was multiplied by a weighting factor depending on the average height of each part. These weighting factors were obtained by computing a leaf area density for each part following the relationship found by Béland et al. (2014) for oak trees, and normalizing these densities so that their sum equals 1. Other overstory characteristics corresponded to those that can be found in the literature concerning the blue oaks of Tonzi (Karlik and McKay, 2002; Ryu et al., 2012). The understory was modelled using a heterogeneous lambertian surface for the soil, and a 3D dry grass layer with a LAI of 0.7 m^2/m^2, similar to what can be found over the Tonzi Ranch site (Kobayashi et al., 2012; Misson et al., 2007). Sun elevation was set to 75 °, the atmosphere was configured for mid-latitude summers and aerosols were set to rural, summer mid-latitudes with a visibility of 23 km. Table 5.2 gives a summary of both overstory and understory characteristics.

Table 5.2: Overstory and understory characteristics used for the procedural generation of the synthetic scenes within DART

	Value	Comment
Tree characteristics		
Crown diameter (D_{crown}, m)	6 (μ_{crown}) 2 (σ_{crown})	normal distribution μ: mean; σ: scale
Height below crown (m)	$1.8 \times \frac{D_{crown}}{\mu_{crown}}$	
Crown height (m)	$7.6 \times \frac{D_{crown}}{\mu_{crown}}$	
Crown shape	semi-ellipsoid	
Empty voxels (%)	60	
3D NPV	modeled	from lidar data
Tree LAI (m^2/m^2)	2.6–7.7	uniform distribution
Leaf characteristics		
ALA (°)	55–65	uniform distribution
EWT (g/cm^2)	0.002–0.025	uniform distribution
LMA (g/cm^2)	0.002–0.025	uniform distribution
Understory characteristics		
LAI (m^2/m^2)	0.7	
LAD	spherical	
EWT (g/cm^2)	0	
LMA (g/cm^2)	0.01	
soil reflectance	brown loam	from DART mineral database

After generation of reflectance image, it was downsampled to 2 m by pixel aggregation, to attain the same GSD as AVIRIS-NG. Figure 5.4 shows the synthetic scene as generated by DART, with a 20 cm spatial resolution, and after aggregation to a 2 m spatial resolution.

(a) modeled tree and ground (b) original (20 cm GSD) (c) downsampled (2 m GSD)

Figure 5.4: Example of the scene modeling and colored compositions of the procedurally generated 240 × 240 m^2 scene. The spatial downsampling was done by spatial aggregation and averaging of the pixels.

5.2.5 Reflectance Sensitivity to Structural Elements

Before generating the training databases for the random forest regressor (RFR), a monovariant sensitivity analysis was done in order to assess reflectance differences between the tree crowns modeled in DART with and without various detailed structural elements. To do so, reflectance values were extracted from the within-crown pixels and averaged. The reference DART modeling for the structural sensitivity was a SFR, with four lollipop trees placed over a lambertian ground surface so that the scene SFR is the closest to a forest (Gastellu-Etchegorry, Gascon, and Estève, 2003). This simple modeling is based on the work done in Banskota et al. (2015) and Gascon et al. (2004) and in Chapter 3. The leaf angle distribution was set to ellipsoidal, so that the average leaf angle was a variable parameter. The solar elevation angle was set to 75 °, which roughly corresponds to the solar elevation at noon at the latitudes of TZ and SJER in June (and the time of the AVIRIS-NG overpasses).

The reflectance changes due to the introduction of five refined structural elements in the SFR scene were evaluated over the 0.8–2.4 µm range, for which EWT and LMA have the most influence on the reflectance Yanfang Xiao et al., 2013. The tested structural elements are indicated in Table 5.3. To ensure that results were not specific to a single combination of LAI, *Average Leaf Angle* (ALA) and leaf biochemical content, reflectance changes were measured for 1,000 combinations (see Table 5.4 for the variation ranges of each variable).

5.2.6 Random Forest Regressors

In order to assess the best performances that could be obtained by RFR when estimating EWT and LMA over high resolution hyperspectral images, and to understand how these performances were affected by the RTM parametrization, two different training databases were built for estimation purposes over the synthetic images.

Table 5.3: Structural elements and their variation that were considered as model inputs for the sensitivity analysis.

Structural elements	SFR – reference	Variation
3D NPV	none	imported from lidar data
canopy height (m)	9.3	13.6
ground modeling	lambertian surface (soil + herbaceous layer)	lambertian surface (soil) + 3D herbaceous layer
crown shape, leaf distribution	ellipsoidal, homogeneous	semi-ellipsoidal, heterogeneous
empty voxels (%)	0	60

The first database, designated as the SFR database, used a simplified forest representation using DART. Therefore, only the mean tree height, height below crown and crown diameter structural components were considered for modeling purposes. The LAD was set to ellipsoidal, and the lambertian ground reflectance was extracted from the open parts of the synthetic images. This is a model that is realistically achievable and has been successfully used in the past to estimate biophysical and biochemical parameters (*e.g.* by Gascon et al., 2004 and in the previous chapters).

The second database, designated as the DETAIL database, used a very detailed 3D modeling, including all the structural elements used during the building of the synthetic scenes. Ground and tree crowns were modeled as in the synthetic images, using detailed 3D NPV. Three different tree heights were considered, following the height distribution within the images. Such a precise parametrization, while not realistically feasible on real cases due to insufficient field information, was accomplished to serve as a reference for the best performance that could be expected from RFR.

For estimations over the two AVIRIS-NG images, two different training databases were built following the specificities of each site. A simplified forest representation was used in each case, taking into account the results presented section 5.3.2 for the synthetic images. To take into account the various understory types present on the sites, it was decided that the ground would be modeled as a lambertian surface but with varying reflectance values derived from the background spectra obtained in the field (see Section 5.2.1). Since the specific crown shape and leaf distribution within the crown was unknown, crowns were modeled as homogeneous ellipsoids with varying ALA, in a similar fashion as what was done for the synthetic images.

Each database comprised 5000 entries, generated according to Table 5.4. The variation ranges considered for EWT and LMA were set such that they would encompass more than the field values given in Table **??** and reduce the overfitting when training the RFR for TZ and SJER. Similarly, the LAI variation range was set so as to encompass more than the tree LAI field measured values by Karlik and McKay (2002). Indeed, all DART scenes were designed so that the CC was 40%, the relationship between LAI and tree LAI therefore being $\text{LAI} = \text{LAI}_{\text{tree}} \times 0.4$. Therefore, a LAI range of 0.3–4 m^2/m^2 leads to a tree LAI range of 0.75–10 m^2/m^2, wider than the 2.5–7.7 m^2/m^2 reported by Karlik and McKay (2002). Variable parameters were sampled following a Latin hypercube sampling. For each case, reflectances from the pixels comprised within the tree crowns were extracted from the DART images, averaged, and linked with their respective EWT and LMA values to build the various databases. All reflectances were noised using a multiplicative wavelength-independant noise $\epsilon_m \sim \mathcal{N}(1, 0.03)$ and an additive wavelength-independant noise $\epsilon_a \sim \mathcal{N}(0, 0.005)$ as such:

$$R' = R \times \epsilon_m + \epsilon_a \tag{5.1}$$

with R' the noised reflectance, and R the RTM-computed reflectance.

Table 5.4: DART and PROSPECT fixed and variables parameters used to generate the various databases used in this chapter.

	SFR	**DETAIL**	**TZ**	**SJER**
Scene parameters				
Ground reflectance (Lambertian)	brown loam + herbaceous layer	brown loam	*Avena* *Brachypodium* *Bromus* low stature species	QUDO litter QUWI litter PISA litter senescent grass
3D herbaceous layer	no	yes	no	no
Cell dimensions (m^2)	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4
Scene dimensions (m^2)	9 × 9	9 × 9	13 × 13	16 × 16
Solar elevation (°)	75	75	adapted to image acquisition time	
Tree structural parameters				
Crown shape	ellipsoidal	as per Figure 5.4a	ellipsoidal	ellipsoidal
Crown diameter (m)	6	6	8.2	10.1
Tree height (m)	9.4	6.3; 9.4; 12.6	14.5	8.6
Crown height (m)	7.6	5.1; 7.6; 10.1	10.4	7
LAI (m^2/m^2)	0.3–4	0.3–4	0.3–4	0.3–4
ALA (°)	55–65	55–65	55–65	55–65
Empty voxels (%)	0	60	0	0
3D NPV	no	yes	no	no
Leaf biochemical parameters				
EWT (g/cm^2)	0–0.025	0–0.025	0–0.025	0–0.025
LMA (g/cm^2)	0–0.025	0–0.025	0–0.025	0–0.025
Structural parameter N	1.5–2.1	1.5–2.1	1.5–2.1	1.5–2.1

Random forest regressors require a number of hyperparameters to tune them, such as the number of trees, features to consider at each split, number of levels, etc. For all the RFR used in this chapter, bootstrapping was used and all features were considered at each split. The tuning of the other hyperparameters was done through a randomized search with 150 iterations and 3-fold cross validation over the grid defined in Table 5.5. RFR were trained using different spectral intervals. Two cases were tested concerning training interval: using information from the full 0.8–2.4 µm range, corresponding to the spectral ranges affected by EWT and LMA (Yanfang Xiao et al., 2013), or only using the spectral information from the 1.5–2.4 µm range. 70% of the databases were used for training, the remaining 30% serving as test data. RFR trained on the average tree crown reflectances and their associated EWT or LMA values.

Table 5.5: Values of the RFR hyperparameters considered in the randomized search for the optimal combination.

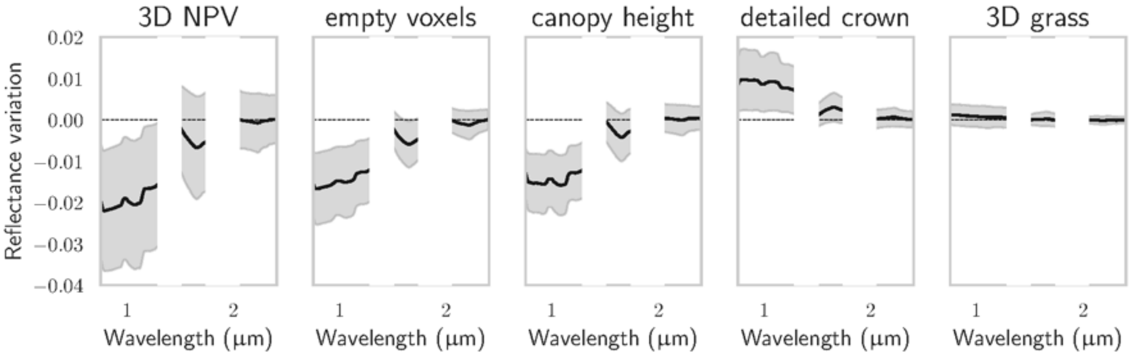
Hyperparameter	Values
minimizing function	mean squared error; mean absolute error
number of estimators	50; 112; 175; 238; 250
maximal depth	10; 20; 30; 40; 50; 60
min. samples for a split	2; 5; 10
min. samples for a leaf (input %)	10^{-3}; 10^{-2}; 10^{-1}

For the synthetic images, the position of trees and leaf biochemical content was perfectly known. Therefore, the estimated EWT and LMA contents of each vegetation pixel were directly compared to the true values by averaging estimates done over each crown. Performance estimates were assessed using both RMSE and R^2. On the other hand, for the AVIRIS-NG images, because of some degree of uncertainty inherent to GPS measurements, registration of the images, and the difficulty in delineating exact tree crowns, the exact crown pixels to associate with the validation data were not known. To overcome this, it was decided to compare the validation values with the average of estimated values over a window of 5×5 pixels centered on the GPS locations of the trees. Estimation performances were also assessed using RMSE and R^2, although R^2 results are difficult to discuss as the number of validation data and their variation ranges are limited.

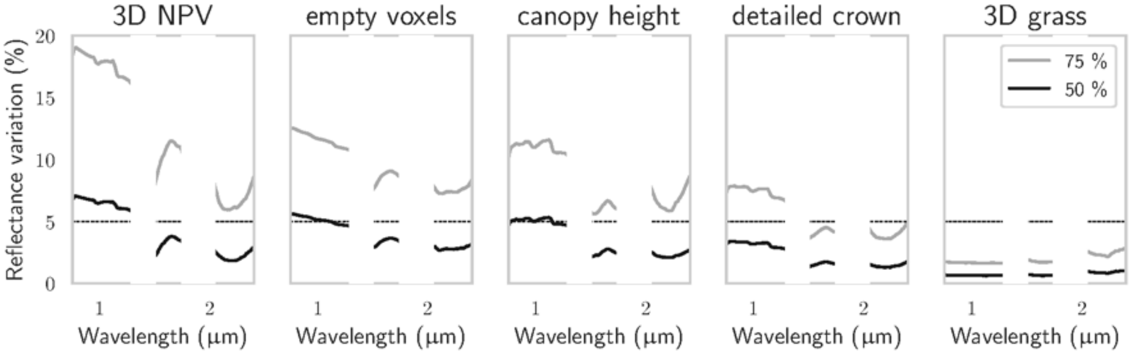
5.3 Results

5.3.1 Effects of Structural Elements on Crown Reflectance

The sensitivity analysis shed light on the influence of the various structural modeling parameters on tree crown reflectance (see Figure 5.5a). 3D NPV, empty voxels proportion, and canopy height affected crown reflectance in a similar fashion, lowering reflectance by about 0.02 over the 0.75–1.3 µm spectral range while barely affecting the 1.5–2.4 µm range. Similarly, changing crown shape and LAI distribution within the crown (detailed crown scenario) led to an increase of reflectance over the 0.75–1.3 µm range, but barely any change at higher wavelengths. Understory modeling did not affect reflectance significantly. Overall, the introduction of 3D NPV had the most effect on crown reflectance: more than 50% of the cases had a variations of more than 0.02, with a standard deviation of about 0.015, the highest among the structural elements. Analyzing the relative variation of reflectance, shown in Figure 5.5b, it appears that for wavelengths above 1.5 µm, more than half of the sensitivity cases present reflectance variations below 5 %, a typical uncertainty associated to high calibration efforts (Widlowski, Côté, and Béland, 2014). Still, the 3D NPV, empty voxels, and canopy height scenario all presented relative variations of more than 5 % for at least 25 % of the sensitivity database, highlighting the important effect of the various tree structural elements on the crown reflectance computed by DART.



(a) The black line corresponds to the median reflectance variation, while the shaded area corresponds to ± one standard deviation.



(b) The black line corresponds to the median relative reflectance variation, while the gray line corresponds to the 75th percentile.

Figure 5.5: Crown reflectance variations caused by the introduction of more detailed structural elements during the monovariant sensitivity analysis over the 1,000 tested cases.

To further analyze the results for the 3D NPV case, an *Analysis Of VAriance* (ANOVA) of crown reflectance variations was done at 1.1 µm, 1.7 µm and 2.05 µm to assess which of the traits (LAI, LMA, EWT, N, ALA) were responsible for them. The results of the ANOVA are presented in Table 5.6. At 1.1 µm, only LMA values were driving the variance, with p values clearly below 0.05 ($< 10^{-10}$). At 1.7 µm, LMA was still the main driver of variance, although at this wavelengths the effects of LAI and EWT were also non-negligibles, with p values of 10^{-15}, 10^{-9} and 10^{-12}, respectively. It is at the 2.05 µm that all vegetation traits except N had a significant influence, LAI and EWT being the most important drivers.

Table 5.6: p values of each vegetation trait obtained from the ANOVA on crown reflectance variations.

wavelength	LAI	LMA	EWT	ALA	N
1.1 µm	0.13	4.8×10^{-11}	0.52	0.90	0.12
1.7 µm	7.7×10^{-9}	4.8×10^{-15}	1.6×10^{-12}	0.43	0.59
2.05 µm	1.2×10^{-127}	1.3×10^{-5}	4.1×10^{-21}	1.8×10^{-6}	0.10

5.3.2 EWT and LMA Estimations

Figure 5.6 shows how the databases generated with DART in forward mode compare with the crown reflectances from the hyperspectral images. For the synthetic image, crown reflectances were within databases boundaries for both SFR and DETAIL over the whole 0.75–2.4 µm. The effect of the detailed structure on database reflectances was clearly visible over the 0.75–1.3 µm spectral range, with reflectances from the DETAIL database being noticeably lower than those from the SFR database and showing overall a larger dispersion. For the TZ and SJER databases, dedicated to the AVIRIS-NG images and generated with a SFR modeling, it appeared that the spectral behaviour of the tree crowns over the 0.75–1.2 µm range was not well captured, with an overestimation of the reflectance by DART. However, over the 1.5–2.4 µm range, crown reflectances were well within database boundaries and could therefore be considered appropriate to train the RFR.

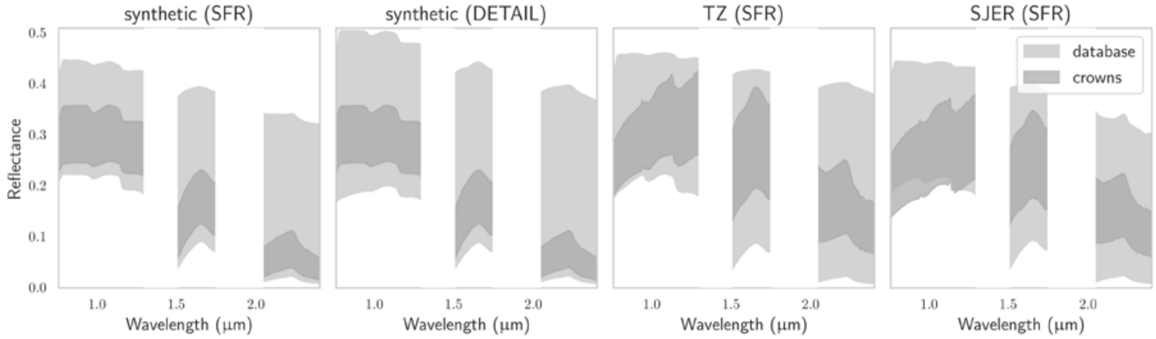


Figure 5.6: Reflectance spectra contained within the various databases, and the reflectance spectra of the broad-leaved trees present within the associated hyperspectral images.

Table 5.7 presents the RMSE and R^2 values of the trained RFR predictions when applied over the train, test, and application sets for the synthetic forest image. For all databases, dedicated to the synthetic or AVIRIS-NG images, training and testing accuracies were consistent.

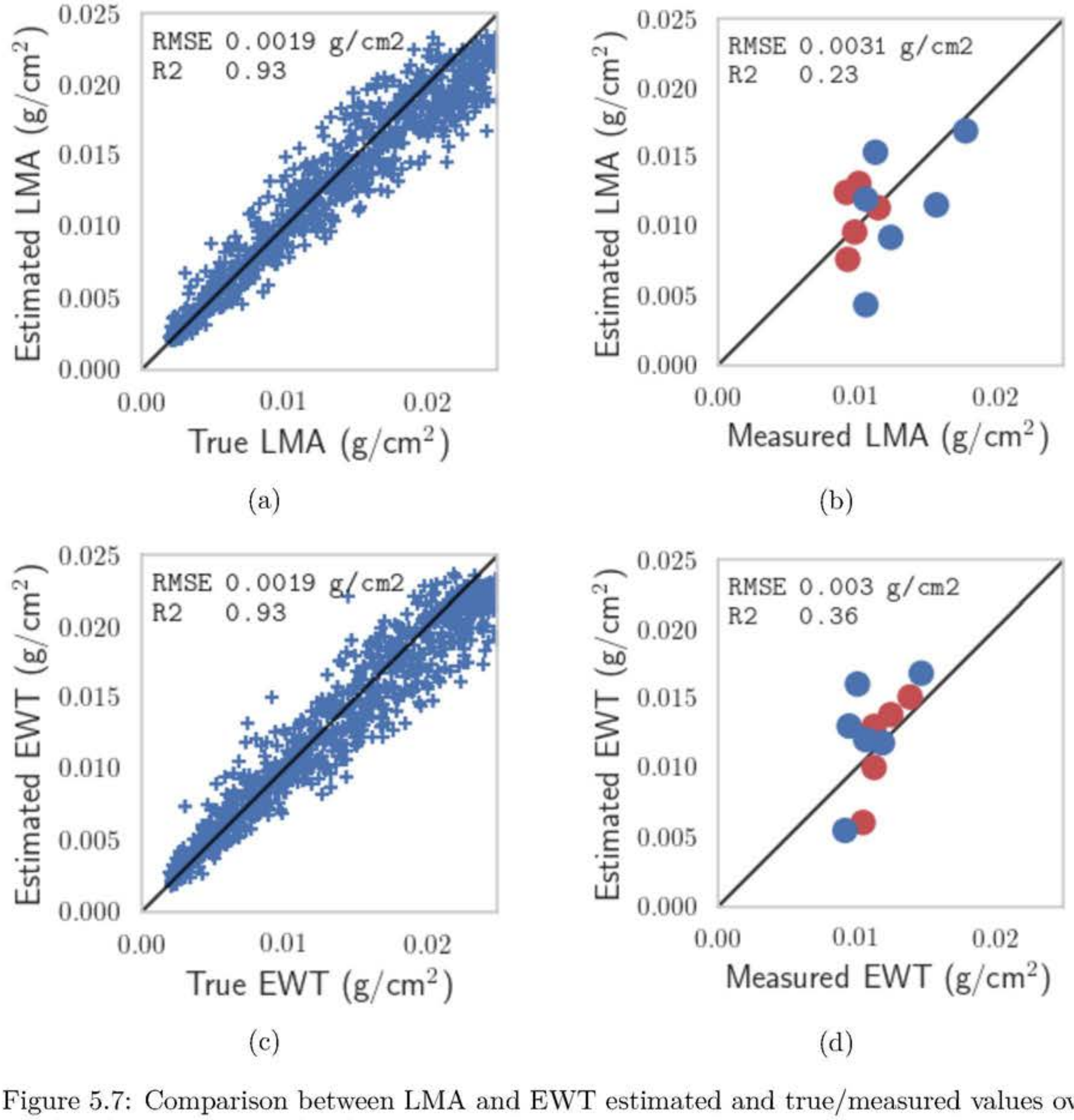
Over the training datasets, RMSE were about 0.0011 g/cm^2 and $R^2 > 0.95$ for both EWT and LMA, while over the testing datasets performances slightly degraded, with RMSE around 0.0025 g/cm^2 and R^2 close to 0.90. Concerning the synthetic images, it appeared that the spectral intervals used to train the RFR, either 0.75–2.4 µm or 1.5–2.4 µm, did not affect performances on the train and test databases. The feature importances identified by each RFR after training are presented in Figures B.1, B.2 and B.3 in Appendix B.

Table 5.7: Performances of the RFR dedicated to the synthetic scene when applied over the trained and test sets as well as over the hyperspectral images.

$\times 10^{-4}$ g/cm^2 (RMSE)			**train**		**test**		**application**	
			RMSE	**R^2**	**RMSE**	**R^2**	**RMSE**	**R^2**
synthetic image								
0.75–2.4 µm	LMA	SFR	11	0.98	25	0.88	35	0.81
		DETAIL	12	0.97	26	0.87	21	0.91
	EWT	SFR	10	0.98	22	0.91	19	0.92
		DETAIL	10	0.98	21	0.91	19	0.93
1.5–2.4 µm	LMA	SFR	11	0.97	22	0.91	19	0.93
		DETAIL	11	0.98	24	0.90	19	0.94
	EWT	SFR	11	0.98	22	0.90	19	0.93
		DETAIL	11	0.98	22	0.91	19	0.93
AVIRIS-NG images								
1.5–2.4 µm	LMA	TZ	12	0.97	27	0.87	31	0.23
		SJER	13	0.97	25	0.88		
	EWT	TZ	13	0.97	27	0.86	30	0.36
		SJER	14	0.96	29	0.84		

Over the hyperspectral images however, performances could differ depending on the spectral interval. Indeed, when trained over the whole spectral range, LMA estimates were significantly less accurate with SFR modeling, with a RMSE of 35 g/cm^2 compared to DETAIL's 21 g/cm^2 RMSE. Accuracy concerning EWT was unaffected by the modeling. Restricting the training from to wavelengths above 1.5 µm made the SFR scenario perform significantly better, with a RMSE of 19 g/cm^2, equal to the one of the DETAIL scenario. For the synthetic image, estimates accuracy were in line with those obtained over the test sets. The estimators dedicated to the AVIRIS-NG images also presented performances similar to those obtained over the test datasets, with the former having RMSE of 31 and 30 g/cm^2 for LMA and EWT, and the latter having RMSE comprised between 25 and 29 g/cm^2 for these traits. Figure 5.7 presents the estimated EWT and LMA values and how they compare with the reference values over the synthetic and AVIRIS-NG images.

Figure 5.8 shows EWT and LMA estimation maps over the broad-leaved trees of the two study sites. Estimated values within tree crowns did not vary much, indicating that they are not random. For both sites, it seemed that the zones with the highest tree density were also the zones with the highest EWT and LMA.



(a) (b)

(c) (d)

Figure 5.7: Comparison between LMA and EWT estimated and true/measured values over the (**a**, **c**) synthetic and (**b**,**d**) AVIRIS-NG images.

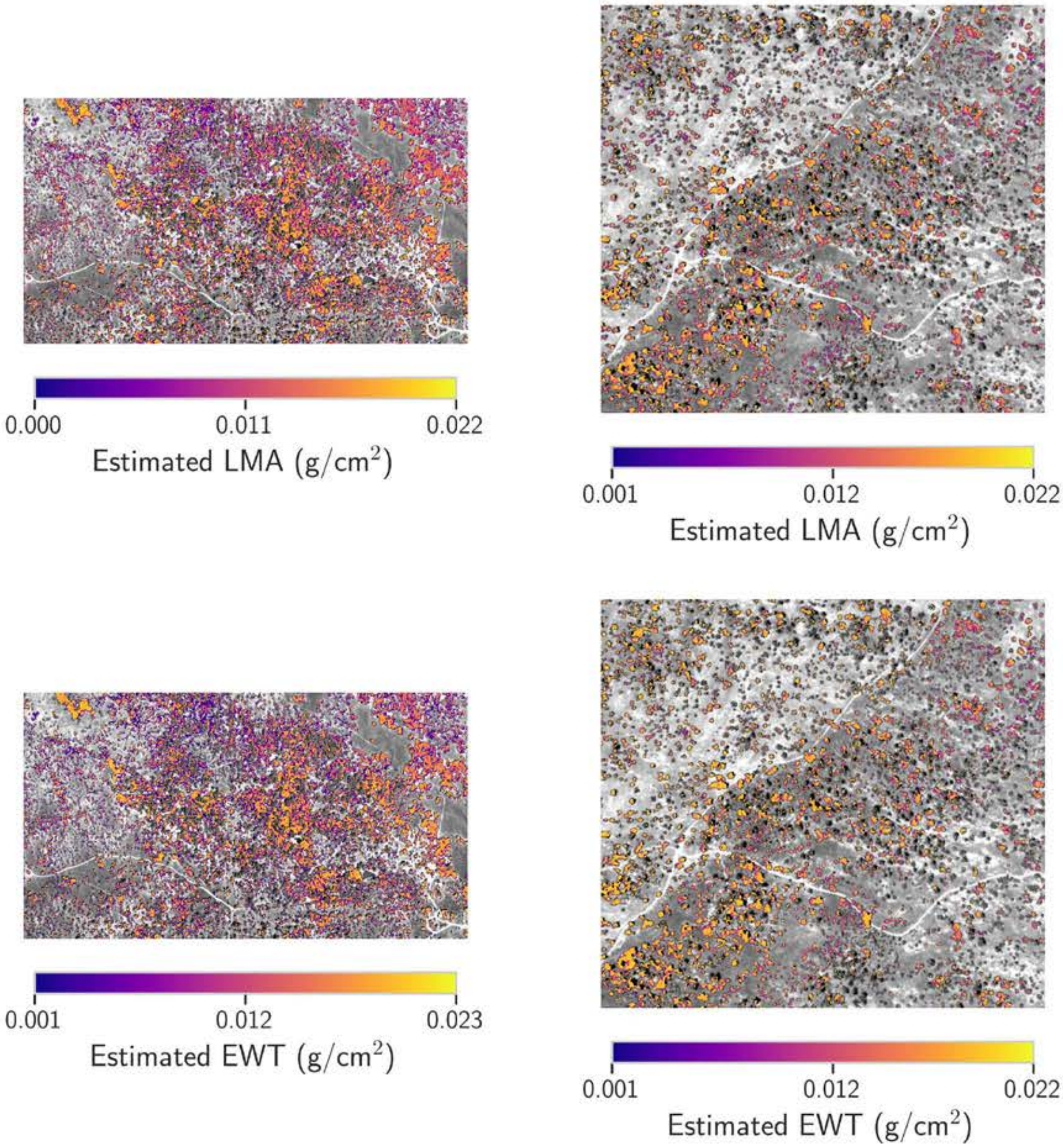


Figure 5.8: EWT and LMA estimation maps from the AVIRIS-NG hyperspectral images acquired over the two study sites (left: TZ; right: SJER).

5.4 Discussion

5.4.1 On the Use of Synthetic Images

This chapter complemented the limited field measurement dataset with synthetic forest scenes generated within DART. The DART model took part in the third phase of the RAMI experiments (Widlowski et al., 2007), which showed that the simulation results of six 3D Monte-Carlo models (including DART) were consistently in good agreement, and could therefore be used to generate surrogate truth data for later RTM performance assessment. Synthetic scenes generated with DART have previously been used to calibrate vegetation indexes (Malenovský et al., 2013), to assess the sensitivity of a biomass prediction method (Proisy et al., 2012), and DART is regularly used for the estimation of various biophysical and biochemical parameters. While obviously, synthetic scenes do not fully replace real hyperspectral acquisitions, the consistency that the truth data produce as compared to field data should make them acceptable surrogates for validation purposes, provided sufficient modeling realism has been undertaken.

5.4.2 Influence of the Structural Parameters on Crown Reflectance

The sensitivity analysis, whose results are presented in Section 5.3.1, demonstrated that most abstractions done as part of the SFR scenario could significantly influence crown reflectance over the 0.75–2.45 μ spectral range, in particular over the 0.75–1.5 μm range, with more than 50 % of the cases of the 3D NPV scenario exceeding the 5% uncertainty threshold considered by Widlowski, Côté, and Béland (2014). Some of the variations of crown reflectance found in the sensitivity analysis were in line with what was previously obtained in the literature. Ferreira et al. (2018) and Malenovský et al. (2008b) have shown that the presence of NPV within the tree crown (in the form of a turbid medium with branch optical properties) significantly affected reflectance over 0.7–0.9 μm and dampened the red-edge, especially for the sunlit portions of the crowns. The results concerning the proportion of empty voxels within tree crowns also corroborated what was obtained by Ferreira et al. (2018) at lower wavelengths, which was that completely full crowns presented higher reflectance values than those without, due to increased within-crown scattering leading to more radiance leaving the crown. It appears these modeling parameters still significantly affect crown reflectance in the short-wave infrared.

Even for wavelengths above 1.5 μm, this 5% uncertainty threshold was exceeded for the scenarios 3D NPV, empty voxels, and canopy height for more than 25% of the (LAI, EWT, LMA, N, ALA) combinations tested. However, the ANOVA results presented Table 5.6 showed that the reflectance variations in the various scenarios depended on the vegetation traits' values. Moreover, the relative quantity of variance explained by each trait varied depending on the spectral band. This is significant: if reflectance variations due to a certain level of abstraction depend on vegetation traits in forward mode, it means that the abstraction will lead to uncertainties regarding the trait estimations in inverse mode.

Therefore, considering wavelengths below 1.5 µm to estimate LMA may lead to great inaccuracies when working with images of trees crowns with a low GSD if crown are not modeled realistically, even though wavelengths in this spectral region are mostly sensitive to LMA (Yanfang Xiao et al., 2013). As visible in Figure 5.6, the reflectance range covered by the DETAIL database was much larger than the one covered by the SFR database for these spectral bands with, in particular, an upward slope from 0.75 to 1.0 µm not present in the SFR database. Such a slope could be found in the TZ and SJER AVIRIS-NG images. While the TZ crown reflectances were actually within the database reflectance ranges, the shape difference at these wavelengths may make the latter inadequate candidates for the training of an estimator. In the case of the SJER AVIRIS-NG images, considering the shape difference between the SFR and DETAIL datanases, it appears plausible that if the trees within the RTM had been modeled without any abstraction, crown reflectances for wavelengths below 1.5 µm would have been comprised within the RTM outputs' ranges.

Above 1.5 µm, the majority of the tested combinations' variations were below the 5% threshold, and the variance did not depend on a single vegetation traits anymore, tampering the possible retrieval inaccuracies. Brede et al. (2020) assessed the sensitivity of spectral bands in this region to various vegetation traits, for two RTM (DART and PROSAIL) and found that that EWT and LMA were the main reflectance driver for both RTM, even though PROSAIL scenes were of course considerably more abstracted than even the DART SFR modeling done in their study. It would therefore appear that the spectral bands in this region are more fit for EWT and LMA estimation purposes when tree crowns are not accurately modeled.

Modeling the understory as a 3D turbid layer instead of a simple flat lambertian surface had almost no effect on the scene reflectance, provided the reflectance used for the lambertian was representative of the effective understory. Melendo-Vega et al. (2018) found that, despite refining the modeling of the understory by coupling PROSAIL with FLIGHT, scene reflectance were systematically overestimated in the near-infrared, possibily due to the inability of PROSPECT to properly model senescent and decomposing grass material. This could be adressed in the future: while more validation studies are necessary to properly understand the limitations of their proposed PROSPECT version, Lu, Proctor, and He (2021) obtained promising results for the modeling of senescent grass material at the leaf and canopy scale. Meanwhile, provided the spatial resolution of the hyperspectral images allows it, it appears that directly extracting understory reflectance from the image to use it as an input in the RTM is sufficient and barely affects the final crown reflectance even for low-LAI trees.

5.4.3 EWT and LMA Estimations

Because the DART scene that was used to generate the synthetic hyperspectral image was completely known, direct comparison between EWT and LMA estimates for crowns and the true values was possible, ensuring that the estimation errors were only due to limitations inherent to the RFR or to the DART modeling used to generate the databases. In the present chapter, from each tree crown, all pixels were used when applying the RFR, as the final estimate was computed as the average of the estimates from all the pixels of a crown. While this was not shown, this averaging helped to improve the accuracy of the estimates, possibly because otherwise pixels from the shaded parts of the crown could be detrimental to the estimations due to their lower SNR (Malenovský et al., 2013). For this reason, multiple studies had only considered sunlit crown pixels for estimation purposes. Ferreira et al. (2018) rejected all pixels with a NIR reflectance below 25 %; Malenovský et al. (2008b) separated sunlit and shaded parts using a maximum likelihood classifier; Asner et al. (2015) and Dana Chadwick and Asner (2016) both employed a lidar-based mask to identify and reject shaded parts of the crowns. The distinction between sunlit and shaded foliage was not done in the present chapter, as the 2 m GSD would have been rather coarse with regards to the tree crowns of the studied sites.

Over the synthetic image, it was shown that LMA estimation accuracy varied significantly depending on the spectral range and the database used to train the RFR. Indeed, when including the spectral bands below 1.5 µm, they were rightfully identified by the RFR to be sensitive to LMA. However, the reflectance bias due to the modeling abstractions, discussed in the previous section, led to poorer results with the SFR than with the DETAIL database. Forcing the RFR to train only on a subset of the short-wave infrared (above 1.5 µm) allowed to reject the most biased bands and translated into the SFR database being equivalent to the DETAIL database for estimation purposes. EWT estimation however were unaffected by tree crown abstractions: indeed, the wavelengths identified as critical by the RFR were close to the atmospheric water vapor bands (as shown in Appendix B), and thus not very sensible to the actual tree structural characteristics, as visible in Figure 5.5.

Estimates of EWT and LMA for broad-leaved trees over the AVIRIS-NG images showed slightly poorer performances to those obtained over the synthetic images (see Table 5.7), with RMSE around 30 instead of 20 g/cm^2. Nevertheless, these RMSE were very close to those obtained over the test datasets after training of the RFR, and better performances could hardly have been expected in this situation. While the variation range of the *in situ* measurements was limited, results concerning the estimation of EWT and LMA were still encouraging considering the multiple uncertainties inherent to field validation such as small inaccuracies in the image registration, incorrect identification of some pixels as belonging to the crown of a broad-leaved tree, or imprecisions in the estimation of the reference EWT and LMA values. At the leaf level, Féret et al. (2019) obtained RMSE of 0.0015 g/cm^2 for both traits when using PROSPECT inversions, and suggested to reject the 0.9–1.3 µm range due to possible modeling inaccuracies of the leaf optical properties in this domain. At the canopy level Dana Chadwick and Asner (2016) and Asner et al. (2015) estimated LMA with an RMSE of 0.0020 g/cm^2 and 0.0023 g/cm^2, respectively, when fitting a partial least-square regression model on some of the data acquired in the field for a tropical forest from images acquired with a 2 m spatial resolution. Maire et al. (2008) fitted vegetation indexes on databases generated with PROSAIL and obtained RMSE of 0.0009 g/cm^2. However, this also required a calibration of the spectral indices on *in situ* reflectance measurements, which can't possibly be expected if regular image acquisitions over large areas are to be processed in an operational context. Few studies have focused on the retrieval of EWT over forests: Zarco-Tejada, Rueda, and Ustin (2003) focused on chapparal vegetation and demonstrated a relationship between estimated EWT and measured fuel moisture, while studies more generally focus on canopy water content estimation (EWT × LAI) for agricultural purposes (Clevers, Kooistra, and Schaepman, 2008; Wocher et al., 2018). Still, EWT estimations over a variety of communities have been accomplished, and Li et al. (2008) obtained encouraging results when training a partial least-square regressor with PROSAIL outputs.

Further work is needed to determine how well LMA and EWT could be estimated from hyperspectral images acquired with a lower spatial resolution. Indeed, the present and future satellite hyperspectral missions, such as PRISMA (Stefano et al., 2013), SBG (inheritor mission of HyspIRI (Lee et al., 2015)) or Biodiversity (inheritor mission of HYPXIM (Carrere et al., 2013)), will have spatial resolution ranging from 8 to 30 m. Indeed, difficulties should arise with 30 m-resolution images acquired over open canopies, as spectral mixing would occur. However, 8 m-resolution images could in general allow to isolate tree crowns quite well, and the results obtained in the present chapter concerning the sensitivity of crown reflectance to the modeling using within the RTM, as well as the conclusions regarding the best spectral range to use for the training of the estimator, should still be valid.

5.5 Conclusion

In this paper, we assessed how the level of abstraction used to represent a scene within a RTM affects crown reflectance over the 0.75–2.4 μm spectral range, and what it means concerning EWT and LMA estimation accuracies in the context of hybrid retrieval methods. The hybrid method consisted in training a RFR over databases generated with the DART model, considering two different spectral intervals. The estimators were then applied on (i) a synthetic forest image and (ii) two AVIRIS-NG hyperspectral images of sparse woodlands.

Our results showed that using an abstracted tree representation in the RTM, in the form of homogeneous ellipsoidal crowns with no NPV, significantly affected reflectance over the 0.75–1.3 μm spectral range. The impact of this simplified modeling over the 1.5–2.4 μm range was less significant, and we showed that LMA estimates of the RFR trained over this range were not affected by the crown representation. EWT estimates were overall not affected by the DART modeling. Application of the RFR on the two AVIRIS-NG images yield RMSE for LMA and EWT in line with what had been obtained on the test sets in the training phase (0.0031 and 0.0030 g/cm^2, respectively), with spatially coherent estimation maps, illustrating the potential of hybrid methods for retrieval of these vegetation traits over large swathes. Further work is needed to assess the transferability of the present conclusions to hyperspectral images that could be acquired by satellite hyperspectral sensors.

Chapter 6

Assessment of Vegetation Traits Retrieval Capabilities from Synthetic Hyperspectral Satellite Images

---

Contents

6.1 Objectives

Due to their ability to cover large swathes almost continuously, satellite-based remote sensing methods are expected to have an important role in future global monitoring efforts, and the concept of satellite-specific Essential Biodiversity Variables has recently been defined (Pettorelli et al., 2016). Several hyperspectral missions have recently been launched or are in preparatory phase, prefiguring a new era for hyperspectral remote sensing of vegetation. While most satellite sensors should have a GSD of 30 m, SHALOM and Biodiversity are expected to have a GSD of 10 and 8 m, respectively, which would make it possible to better identify canopy structure, especially for open forests, at the cost of a lower SNR. Assessing retrieval performances of vegetation parameters from images taken by 10 m and 30 m GSD satellite sensors over open and closed canopies would help to understand how they complement each other as well as their respective limitations.

The goal of the present study was to assess the estimation accuracies that could be expected from hyperspectral images acquired by SBG and Biodiversity, whose characteristics are presented in Table 6.1, over heterogeneous canopies such as those present in Mediterranean ecosystems and the evolution of these estimations with the openness of the canopy. To do so, synthetic SBG and Biodiversity images were generated from AVIRIS-C and AVIRIS-NG images acquired over two woodland savannas during different years and seasons. PLSR models were fitted on the outputs of the DART model to retrieve leaf traits (leaf C_{ab} and Car contents, EWT, LMA) and, when possible, canopy gap fraction. Estimates of accuracy were obtained by comparison with values measured *in situ*, and behaviours over the different CC were cross-compared with the help of CC maps derived from the AVIRIS-NG images.

Table 6.1: Specifications of the SBG and Biodiversity sensors' characteristics (Transon et al., 2018).

	SBG	Biodiversity
Swath width (km)	145–600	15
Spectral range (µm)	0.38–2.51	0.4–2.5
Spectral bands	214	210
Spatial resolution (m)	30	8
Spectral resolution (nm)	10	10
Temporal resolution (day)	5–16	3–5
SNR (30% albedo)	560:1 to 236:1	200:1 to 100:1

6.2 Materials and Methods

6.2.1 Materials

The study detailed in this chapter considered both Tonzi Ranch and San Joaquin. Aerial views of the sites as well as location of the field measurements considered in this study are given Figure 6.1. This study focused on broad-leaved- dominant pixels, using the classification described in Section 2.6 to identify them. Statistics concerning the field measurements used in the present chapter are given in Table **??**, while Table 6.3 presents the acquisition dates of the hyperspectral images for each sensor.

SBG preparatory hyperspectral data were directly delivered by NASA JPL, in the same

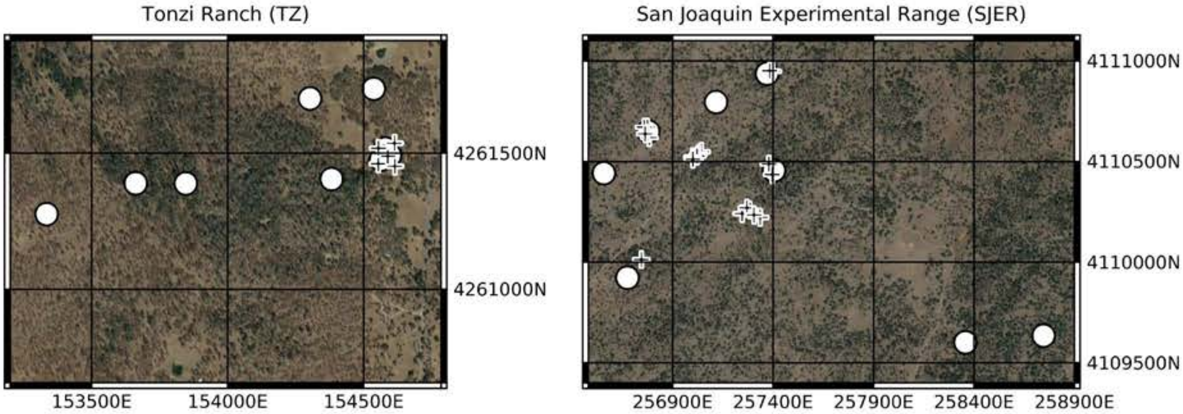


Figure 6.1: Aerial view of Tonzi Ranch and the San Joaquin Experimental Range. Circles and crosses indicate the locations where DHP and leaf collection took place, respectively.

Table 6.2: Data collected over the study sites for gap fraction and leaf biochemistry and minimum, maximum and average values.

	No. of field data										
	TZ			SJER							
(MM/YY)	06/13	09/13	06/14	06/13	10/13	06/14	11/14	Min.	Max.	Mean	
Gap Fraction		2	7			4	8	0.21	0.94	0.61	
C_{ab}	3	5	5	10	9	6	7	10.3	52.6	35.2	$\mu g/cm^2$
Car	3	5	5	10	9	6	7	4.4	13.8	9.2	$\mu g/cm^2$
EWT	3	5	5	10	9	6	7	0.008	0.020	0.016	g/cm^2
LMA	3	5	5	10	9	6	7	0.007	0.016	0.012	g/cm^2

Table 6.3: Dates (DOY/YY) of the overflights of AVIRIS-C and AVIRIS-NG over the study sites.

	TZ			SJER			
AVIRIS-C	155/13	262/13	153/14	163/13	309/13	154/14	279/14
AVIRIS-NG			157/14			162/14	

fahion as the AVIRIS-C images. SBG images were obtained by resampling the AVIRIS-C images using Gaussian weighted sampling with a 30 m FWHM over a 90 m by 90 m area, and a noise approximating a SBG VSWIR noise function was added to the radiance data before obtention of the reflectance images.

Synthetic Biodiversity images were not directly available, and had to be generated from the AVIRIS-NG reflectance images for the present study. First, images were spectrally resampled as to match to the spectral bands of AVIRIS-C/SBG, since Biodiversity is expected to have a similar spectral resolution. Then, a Gaussian weighted sampling with a 8 m FWHM over a 24 m by 24 m area was undertaken to arrive at Biodiversity's GSD. A noise approximating Biodiversity's was added to the reflectance image following the protocol described thereafter, since only a *Noise-Equivalent Delta Radiance* (NEdL) function (with a and b the noise parameters) was available instead of a *Noise-Equivalent Delta Reflectance* (NEdR). This NEdL therefore had to be converted into a NEdR before application to the hyperspectral image.

$$\text{NEdL} = \sqrt{a + b \times L} \tag{6.1}$$

The total radiance L can be divided into its direct, atmospheric and diffuse components (L_{direct}, $L_{\text{atmosphere}}$ and L_{diffuse}, respectively):

$$L = L_{\text{direct}} + L_{\text{atmosphere}} + L_{\text{diffuse}} \tag{6.2}$$

L_{diffuse} can be considered negligible with regards to the other components. L_{direct} can be rewritten as a function of the reflectance R, with E the solar irradiance and τ the atmospheric transmission:

$$L = R \times \frac{E \times \tau}{\pi} + L_{\text{atmosphere}} \tag{6.3}$$

leading to

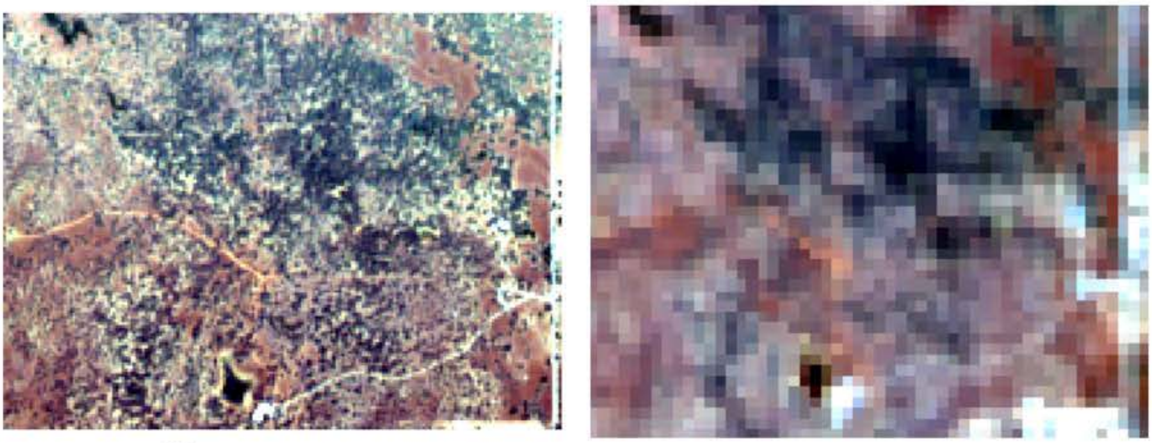
$$\frac{dL}{dR} = \frac{E \times \tau}{\pi} \tag{6.4}$$

Injecting equations 6.3 and 6.4 into equation 6.1, one obtains:

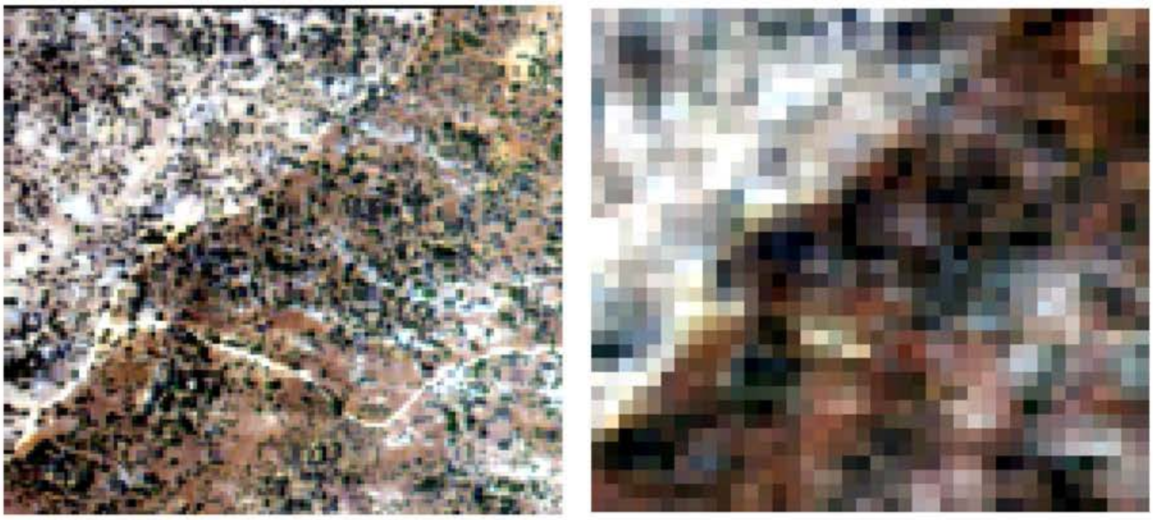
$$\text{NEdR} = \frac{\pi}{E \times \tau} \sqrt{a + b \times \left(L_{\text{atmosphere}} + R \times \frac{E \times \tau}{\pi} \right)} \tag{6.5}$$

Which allowed the noise to be added to the reflectance images. $L_{\text{atmosphere}}$, E and τ were obtained for each AVIRIS-NG image using the COCHISE (Poutier et al., 2002) in-house atmospheric correction code. Examples of synthetic Biodiversity and SBG images are given in

Figure 6.2.



(a) TZ – Biodiversity (b) TZ – SBG



(c) SJER – Biodiversity (d) SJER – SBG

Figure 6.2: Color compositions of the synthetic hyperspectral images over TZ and SJER for June 2014.

6.2.2 Radiative Transfer Modeling

DART 5.7.8v1173 was used for the present study. The optical properties of the leaves are obtained with the leaf RTM PROSPECT-D (Féret et al., 2017).

Trees were modeled within DART with an ellipsoidal crown, following the average proportions of the principal dimensions (width, height, size) of the broad-leaved trees measured on the sites, and distributed along the DART scenes so that the BRF was closest to a forest BRF, in a similar fashion as what was done by Gastellu-Etchegorry, Gascon, and Estève (2003) and Gascon et al. (2004). The trunk and branch 3D models were modeled from the lidar cloud points that were imported within DART to represent the non-photosynthetic elements of the trees. Table 6.4 presents the characteristics of the scenes. DART was used in the Lux mode, with a pixel size of 20 cm and a target ray density per pixel of 200. The spectral bands simulated spanned the 0.5–2.4 µm range, with a 10 nm spectral resolution and a 10 nm bandwidth, for a total of 140 spectral bands after removal of the noisy bands. The ground was modeled as a flat lambertian surface whose reflectance was set to the average value of pixels selected in the open parts of the AVIRIS-C images through a criterion based on their NDVI: all non-water pixels with a NDVI below 0.3 were considered to correspond to the understory.

The various CC were obtained by keeping the same tree dimensions and adjusting the scene dimensions and tree positions. The leaf parameter values were randomly picked following a latin hypercube sampling. In total, 1,600 combinations were generated, leading to 400 combinations for each CC. From the lab measurements, it appeared that a relationship could be established between Car and C_{ab} contents regardless of the site, season and tree species. As such, Car contents within 2.5 standard deviation of the linear relationship presented Figure 6.3 were used to keep cases realistic.

Table 6.4: Values and ranges used to model the forest scenes within the DART model.

Canopy structural parameters	Values & ranges
Canopy Closure (CC)	30; 50; 70; 90 %
Crown shape	ellipsoidal
Crown diameter	8.2 m
Crown height	10.4 m
Height below crown	4.1 m
Ground parameters	
Type	lambertian surface
Soil reflectance	from AVIRIS-C images
Leaf parameters	
LAD	ellipsoidal
ALA	55 – 65 °
LAI	1 – 4 m^2/m^2
C_{ab}	5 – 70 $\mu g/cm^2$
Car	following the Car/C_{ab} relationship
EWT	0.001 – 0.025 g/cm^2
LMA	0.001 – 0.025 g/cm^2
Structural parameter (N)	1.5 – 2.1

DART computes the gap fraction of forest scenes, as described in Section 4.2.2.2. The resultant range in variation of the gap fraction of the DART outputs was 19 – 55 %, admittedly lower than some *in situ* measurements. However, as LAI and CC were both already very low, it was decided not to reduce them further to avoid confusion of the estimator when training on scenes with too little foliage, and instead let the PLSR extrapolate if needed. Finally, all reflectances resulting from the DART simulations were noised using a multiplicative wavelength-independant noise $\epsilon_m \sim \mathcal{N}(1, 0.02)$.

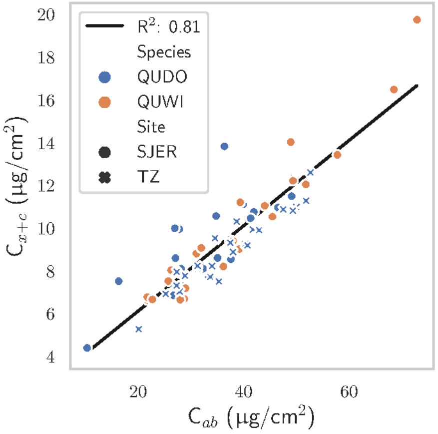


Figure 6.3: Relationship found between Car and C$_{ab}$ in SJER and TZ for the two broad-leaved species after the laboratory measurements.

Different subsets of the 0.5–2.4 µm range were used to estimate each of the vegetation traits, based on their respective spectral contribution to canopy reflectance (Yanfang Xiao et al., 2013). For gap fraction, the whole range was used; for C$_{ab}$ and Car, only bands over the 0.5–0.8 µm; for EWT and LMA, following the findings presented in Chapter 5, bands over the 1.5–2.4 µm range were used.

6.2.3 Partial-Least Square Regressors

PLSR is an estimation method appropriate when model input variables present strong multi-collinearities, as they are projected in a new space so that the variance is maximized between the response and the variables. It models a relationship between p variables $\mathbf{X} \in \Re^{n \times p}$ and l outputs $\mathbf{Y} \in \Re^{n \times l}$, with n samples, as per Equation 6.6,

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \\ \mathbf{Y} &= \mathbf{T}\mathbf{Q}^{\mathbf{T}} + \mathbf{F} \\ \mathbf{T} &= \mathbf{X}\mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1} \end{aligned} \tag{6.6}$$

with $\mathbf{T} \in \Re^{n \times A}$, $\mathbf{P} \in \Re^{A \times p}$, and $\mathbf{E} \in \Re^{n \times p}$ the score, loadings, and residual matrices of $\mathbf{X}$, respectively; $\mathbf{Q} \in \Re^{A \times l}$, and $\mathbf{F} \in \Re^{n \times l}$ are the loadings and residual matrices of $\mathbf{Y}$, respectively; $\mathbf{W} \in \Re^{A \times p}$ is the weight matrix; A is the number of latent variables in the PLSR. While mostly used in chemometrics, it has been successfully applied to remote sensing on several occasions (Asner et al., 2015; Axelsson et al., 2013; Feilhauer, Asner, and Martin, 2015; Meacham-Hensold et al., 2019).

Before processing, reflectance spectra, gap fraction, C$_{ab}$ and Car, EWT and LMA were mean-centered and scaled. As Mevik, Segtnan, and Næs (2004) showed that bagged PLSR were less sensitive to overfitting, an ensemble of 300 bagged PLSR was used in this study to improve the consistency of the results. For each vegetation trait (gap fraction, C_{ab}, Car, EWT, LMA), the ensemble was trained on 75 % of the databases (hereafter refered to as the *training* set), the remaining 25 % serving as *testing* set.

The optimal number of latent variables in the PLSR, n_{opt}, was determined through Monte-Carlo resampling, as described in Kvalheim et al. (2018): the *training* data was repeatedly divided into calibration and validation sets, and for each set the RMSE obtained over the validation sets was computed for a=1, 2, ...,A latent variables to define distributions of RMSE for each number of latent variables. The number A_{min} with the lowest median RMSE is identified, and the fraction p of the RMSE values of A_{min} that exceeded the median RMSE for the preceding latent variables is determined. p is corresponds to a probability measure and can be used to automatically determine if the RMSE with A_{min} latent variables is significantly lower than the RMSE for the preceding number of latent variables (A_{min}-1) for a preselected threshold (p_{upper}). If so, it defines the optimum number of PLS variables as A_{min}. If not, A_{min}=A_{min}-1 and the process is repeated until significance is achieved. For this study, 500 repetitions were made to compute the RMSE distributions, p_{upper} was set to 0.401 and the calibration/validation ratio was 50/50.

Hyperspectral data contain multiple variables (*i.e.* information over several spectral bands), and prediction performance can be improved by a preliminary selection of the relevant variables. Concerning PLSR, a common method, which has been shown to be one of the most consistent by Wang, He, and Wang (2015), is to select the variables of interest based on their importance in the projection (*Variable Importance in Projection* (VIP)). The VIP of the j^{th} variable in a PLSR with A components is scored according to Equation 6.7,

$$\mathrm{VIP}_j = \sqrt{\frac{p\Sigma_{a=1}^{A} SS(q_a t_a)(w_{ja}/||w_a||)^2}{\Sigma_{a=1}^{A} SS(q_a t_a)}} \tag{6.7}$$

with $SS(q_a t_a) = q_a^2 t_a^T t_a$; t_a the a^{th} column vector of $\mathbf{T}$; q_a the a^{th} element of the regression vector q of $\mathbf{T}$; w_a is the a^{th} column vector of $\mathbf{W}$; p the number of variables in $\mathbf{X}$. Once the VIP scores have been obtained for each variable, only variables with a score greater than one are retained and the others are deemed as not significant.

For each variable (gap fraction, C_{ab}, Car, EWT, LMA) and each hyperspectral image, one bagged PLSR model was trained.

6.2.4 PLSR Application and Estimates Accuracy Assessment

Once trained, the bagged PLSR were used to estimate C_{ab}, Car, EWT, LMA and, when possible, gap fraction from the AVIRIS-C, SBG and Biodiversity images, providing estimation maps.

As the GSD of SBG is quite large with regards to the CC of the sites, estimates accuracy was first assessed by comparing the traits' values measured *in situ* to the values estimated in the pixels closest to the GPS positions for both AVIRIS-C and SBG and evaluating the RMSE and R^2. Using the CC maps derived from the AVIRIS-NG images, pixels could be distinguished by the values of the underlying canopy, potentially allowing to identify behaviours depending on its closure. Then, the AVIRIS-C estimation map was spatially resampled to 30 m with a 30 m FWHM averaging filter in order to study the coherency of AVIRIS-C and SBG estimations. For each date, RMSE and R^2 between AVIRIS-C and SBG estimations were computed, a low RMSE and high R^2 indicating good consistency of the estimations across sensors.

Concerning Biodiversity data, only one date was available for each site, greatly reducing the number of validation points. Still, *in situ* measurements were once again compared to the pixels of the estimation map. Only C_{ab} and Car were estimated, as no field data equivalent to a tree gap fraction was available. To complete the accuracy assessment and overcome the small quantity of validation points, the Biodiversity images were, in a similar fashion as in the previous subsection, spatially resampled to 30 m to compare estimates of behaviours across CC and cross-validate the estimated values.

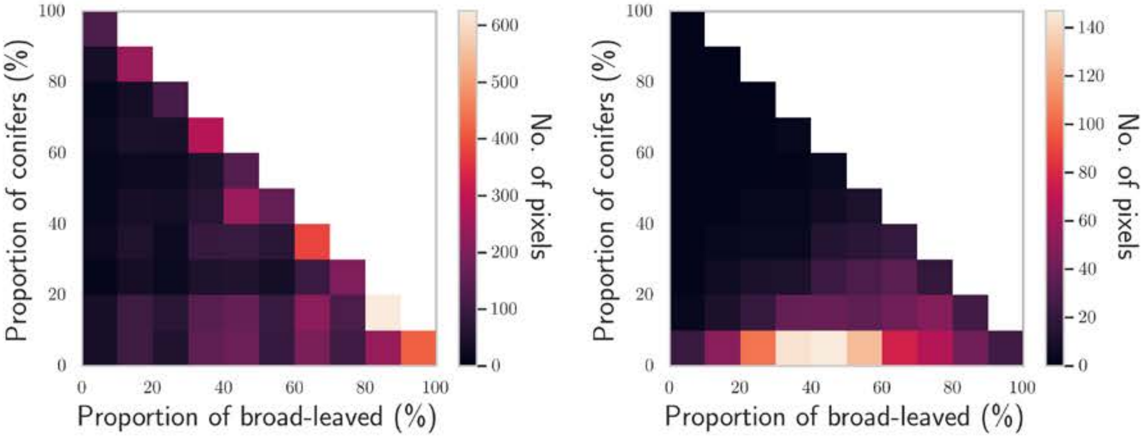
6.3 Results

6.3.1 Canopy Cover within the Sensors' Pixels

Figure 6.4 shows the pixel canopy content of the Biodiversity and SBG images of TZ, derived from the tree-type classification maps. Unsurprisingly, it appears that the vegetation pixels of Biodiversity contain a high proportion of vegetation. Most of the time, the canopy cover of vegetation pixels is around 100%, with a preponderance of QUDO: from the histogram, it appears that QUDO crowns represent 85, 95, and 65% of pixels' contents from the three most common categories. Overall, most vegetation pixels of the Biodiversity image are on the diagonal, *i.e.* have almost 100% CC. Comparatively, most vegetation pixels in the SBG image only had a CC of about 50%, a result consistent with the field measurements presented in Section 6.2.1.

6.3.2 DART-Generated Spectral Databases

First, the databases generated with DART to train the PLSR were compared to the AVIRIS-C, SBG and Biodiversity hyperspectral images in order to assess their adequacy to retrieve vegetation traits. Vegetation pixels within the image were identified by looking for pixels whose NDVI was above that of the reflectance used to model the understory within the DART model. Figures 6.5 and 6.6 illustrate how well the databases encompass the hyperspectral images' vegetation reflectances. Concerning TZ, almost all vegetation pixels were within the databases boundaries over the whole spectral range for both AVIRIS-C and SBG. Only a few pixel reflectances were below those in the database, around 0.8 µm, giving confidence in the ability of PLSR trained from these databases to yield accurate predictions. For SJER,



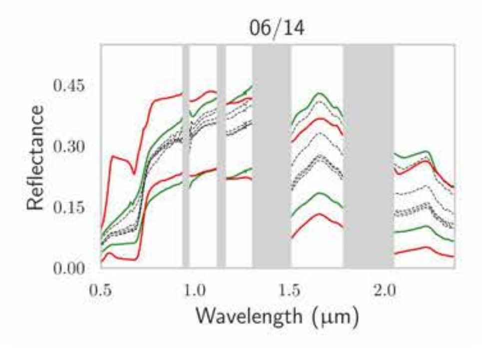
(a) Biodiversity (b) SBG

Figure 6.4: Histograms of the CC of conifers and broad-leaved trees of TZ within the pixels of the hyperspectral images.

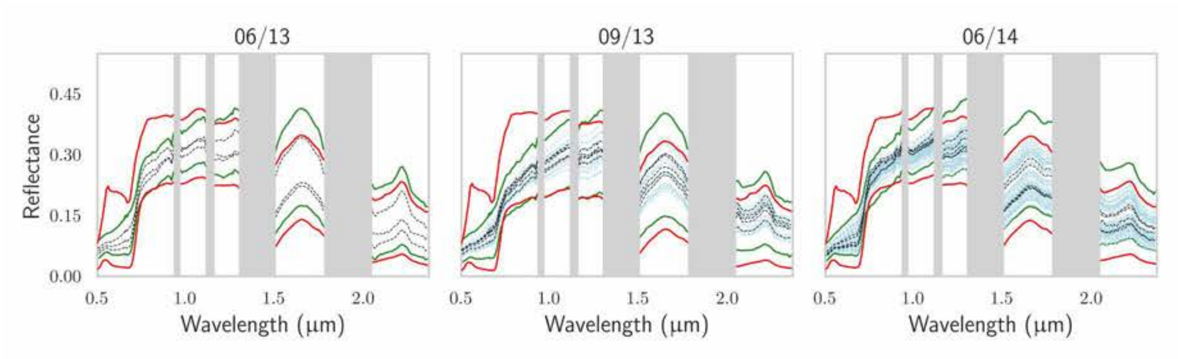
the results were not as good, with a recurrent overestimation of the reflectance in over the 0.8–1.5 µm range, especially visible for the two images acquired during the fall (10/13 and 11/14). Moreover, it appears that some pixels that correspond to *in situ* measurements are not detected as vegetation pixels: for SJER images corresponding to days 06/13 and 10/13, from AVIRIS-C to SBG, some spectra disappear as the pixel's NDVI becomes too low. For the two Biodiversity images, not all vegetation pixels were comprised within the databases' mapped boundaries. It appears that for both SJER and TZ, some pixel reflectances may be lower than those in the databases over the 0.8–1.5 µm region. Moreover, concerning TZ, some reflectances were also above those of the database over the 1.5–2.4 µm region. Overall, 71 % of the vegetation pixels were completely within database boundaries for the Biodiversity image of TZ, while it was 55 % for SJER, probably due to the higher CC at TZ (48%) compared to SJER (30%).

6.3.3 Bagged PLSR Predictions over the Hyperspectral Images

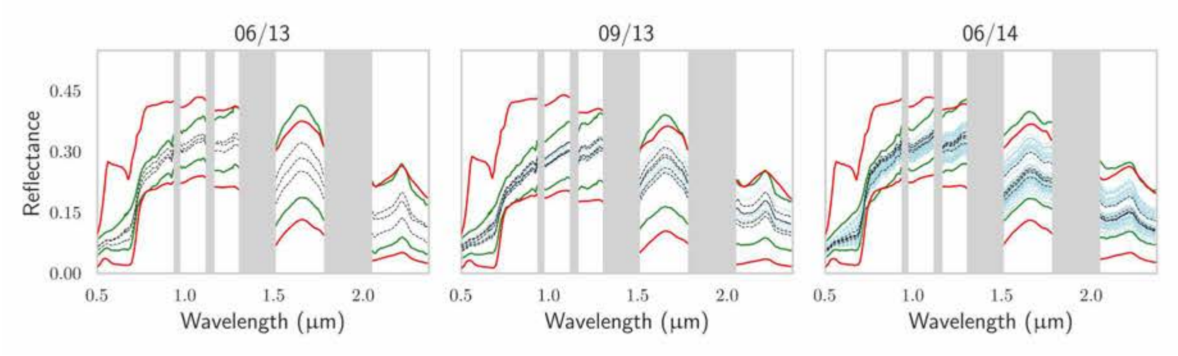
Table 6.5 presents the main statistics of the bagged PLSR models after their fit on the *training* datasets from each date. Concerning the *training* datasets, RMSE and R^2 values are consistent across sites and dates, with similar values for each vegetation trait. RMSE are about 0.03, 8.5 µg/cm^2, 1.5 µg/cm^2, 0.0035 g/cm^2 and 0.004 g/cm^2 for gap fraction, C_{ab}, Car, EWT and LMA, respectively. Goodness of fit is high ($R^2 \geq 0.69$) for every case. Over the *test* sets, performances remained similar, with little change to both RMSE and R^2 values.



(a) Biodiversity

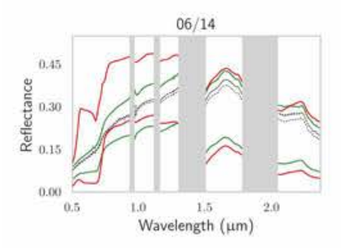


(b) AVIRIS-C

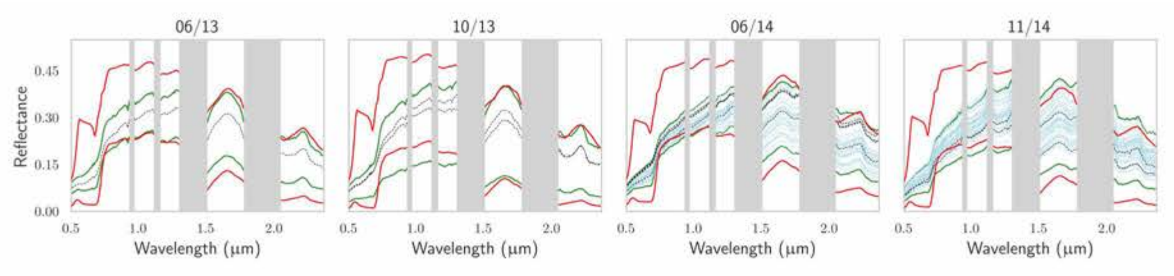


(c) SBG

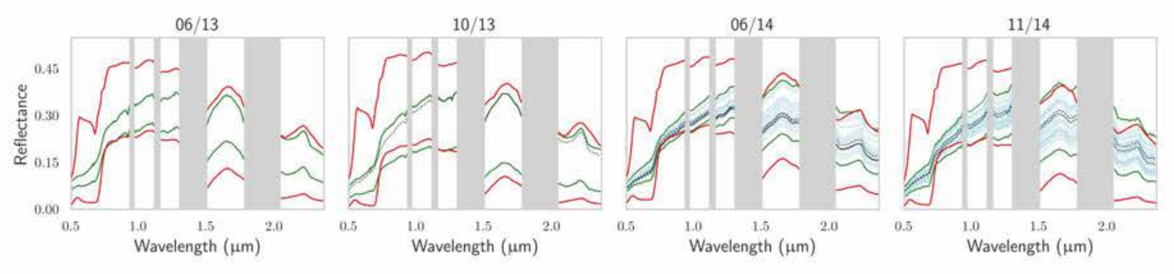
Figure 6.5: For TZ, comparison of (in red) the extrema of the DART-generated databases for Biodiversity, AVIRIS-C and SBG with: (in green) the extrema of the vegetation pixels of their respective images; the reflectances of the pixels associated with *in situ* (in black) biochemistry and (in blue) gap fraction measurements.



(a) Biodiversity



(b) AVIRIS-C



(c) SBG

Figure 6.6: For SJER, comparison of (in red) the extrema of the DART-generated databases for Biodiversity, AVIRIS-C and SBG with: (in green) the extrema of the vegetation pixels of their respective images; the reflectances of the pixels associated with *in situ* (in black) biochemistry and (in blue) gap fraction measurements.

Table 6.5: Statistics of the bagged PLSR models when predicting the vegetation traits over the *training* and *test* databases and over the hyperspectral images. MM/YY corresponds to the month and year of the image acquisition and N to the number of latent component of the PLSR.

		Gap fraction			C_{ab}			Car			EWT			LMA		
	MM/YY	n_{opt}	RMSE	R^2	n_{opt}	RMSE (µg/cm^2)	R^2	n_{opt}	RMSE (µg/cm^2)	R^2	n_{opt}	RMSE (g/cm^2)	R^2	n_{opt}	RMSE (g/cm^2)	R^2
training																
TZ	06/13				6	8.79	0.78	6	1.55	0.77	4	0.0036	0.74	4	0.004	0.69
	09/13				6	9.21	0.76	6	1.62	0.75	4	0.0036	0.74	4	0.0039	0.71
	06/14	7	0.03	0.96	6	8.5	0.79	7	1.55	0.77	4	0.0034	0.76	2	0.0038	0.72
SJER	06/13				4	8.71	0.78	4	1.55	0.77	4	0.0035	0.76	4	0.0037	0.73
	10/13				6	8.66	0.79	6	1.48	0.79	4	0.0035	0.75	4	0.0036	0.74
	06/14	6	0.03	0.95	4	8.54	0.79	4	1.59	0.76	4	0.0035	0.76	4	0.0036	0.74
	11/14	7	0.02	0.97	6	8.91	0.77	6	1.54	0.78	4	0.0036	0.74	4	0.0038	0.72
test																
TZ	06/13					8.77	0.79		1.54	0.79		0.0035	0.75		0.0039	0.71
	09/13					9.47	0.76		1.61	0.78		0.0036	0.74		0.0041	0.69
	06/14		0.03	0.95		8.67	0.8		1.51	0.8		0.0035	0.76		0.0036	0.75
SJER	06/13					8.81	0.79		1.55	0.8		0.0035	0.75		0.0038	0.73
	10/13					9	0.79		1.47	0.82		0.0035	0.75		0.0036	0.76
	06/14		0.03	0.94		9.07	0.78		1.58	0.79		0.0036	0.74		0.0035	0.77
	11/14		0.02	0.96		8.9	0.79		1.49	0.81		0.0038	0.72		0.0038	0.72
Application (all dates)																
	Biodiversity					8.5	0.03		2.24	0.18		0.0071	0.11		0.0049	0.61
	AVIRIS-C		0.09	0.74		7.91	0.26		1.73	0.05		0.0079	0.		0.0053	0.07
	SBG		0.1	0.59		7.44	0.33		1.6	0.17		0.004	0.11		0.0035	0.2

For the hyperspectral images, predictions for all dates were pooled together to compute the statistics. Estimates accuracy for C_{ab} and Car remained similar to what was obtained over the synthetic databases (RMSE are about 8 and 1.75 µg/cm^2, respectively) for all sensors. Gap fraction was also well estimated with a RMSE of 0.1, higher than in the testing phase, and R^2 values of 0.74 and 0.59 for AVIRIS-C and SBG, respectively. EWT and LMA estimations over the hyperspectral images were quite poor: ignoring results for Biodiversity, which only had seven *in situ* points to compute the statistics, R^2 were always very low and no relation between estimates and field measurements were found. RMSE values for EWT and LMA similar for Biodiversity and AVIRIS-C ($\sim$0.0075 and $\sim$0.0050 g/cm^2 for EWT and LMA, respectively), but appeared to be lower for SBG (0.004 and 0.0035 g/cm^2).

6.3.4 Impact of the GSD on Estimates

For illustration purposes, estimation maps of all the vegetation traits, obtained from the synthetic SBG and Biodiversity images for June 2014 are presented in Appendix C.

As visible in Figures 6.7 and 6.8, gap fraction was well estimated no matter the CC of the hyperspectral pixels for both sensors. Concerning C_{ab} and Car estimation, and especially for AVIRIS-C, it appeared that high-CC pixels (CC $\geq$ 90 %) led to more accurate estimations than those with lower CC. Overall, estimates were coherent between AVIRIS-C and SBG. Table 6.6 shows the results of the comparison of these estimates, considering not only RMSE and R^2 but also systematic and unsystematic RMSE (RMSEs and RMSEu, respectively (Willmott, 1981)) over different CC (low, around 40 %, and high, above 80 %). For all vegetation traits, coherency was only slightly improved going from the low to the high CC areas, and a good fit was always obtained (R$^2 > 0.3$ for gap fraction and C_{ab}, and R$^2 > 0.5$ for other traits). After AVIRIS-C estimates were aggregated to 30 m, RMSE between AVIRIS-C and SBG C_{ab}, EWT and LMA values were lower than those obtained when comparing the estimates with the *in situ* measurements. For gap fraction and Car, these RMSE values were of the same order.

Figures 6.9 and 6.10 show a comparison between the estimates obtained from the Biodiversity and SBG images after Biodiversity predictions were aggregated at a 30 m spatial resolution. Two different behaviors are visible for C_{ab} and Car estimations between the low and high CC zones. While the most significant change happened for C_{ab}, whose RMSE was improved from 5.59 to 3.56 µg/cm^2 from one CC category to the other, estimates for all vegetation traits (C_{ab}, Car, EWT, LMA) presented a clearly lower RMSEu between Biodiversity and SBG over the closed parts of the canopy than the open parts. It appears that over closed canopies estimates from Biodiversity and SBG images were well correlated ($0.48 \geq$ R$^2 \geq 0.32$), even though a constant bias could be present for some traits. This was not the case over the open parts (CC $\sim$ 40%) where no real relationship could be found between SBG and Biodiversity estimates. In particular for C_{ab}, R^2 was only 0.2 with a large dispersion of the points around 40 µg/cm^2.

Finally, to complement the results presented in Table 6.6 and Figures 6.9 and 6.10, is shown in Figure 6.11 the correlation between C_{ab} estimations from Biodiversity and SBG over

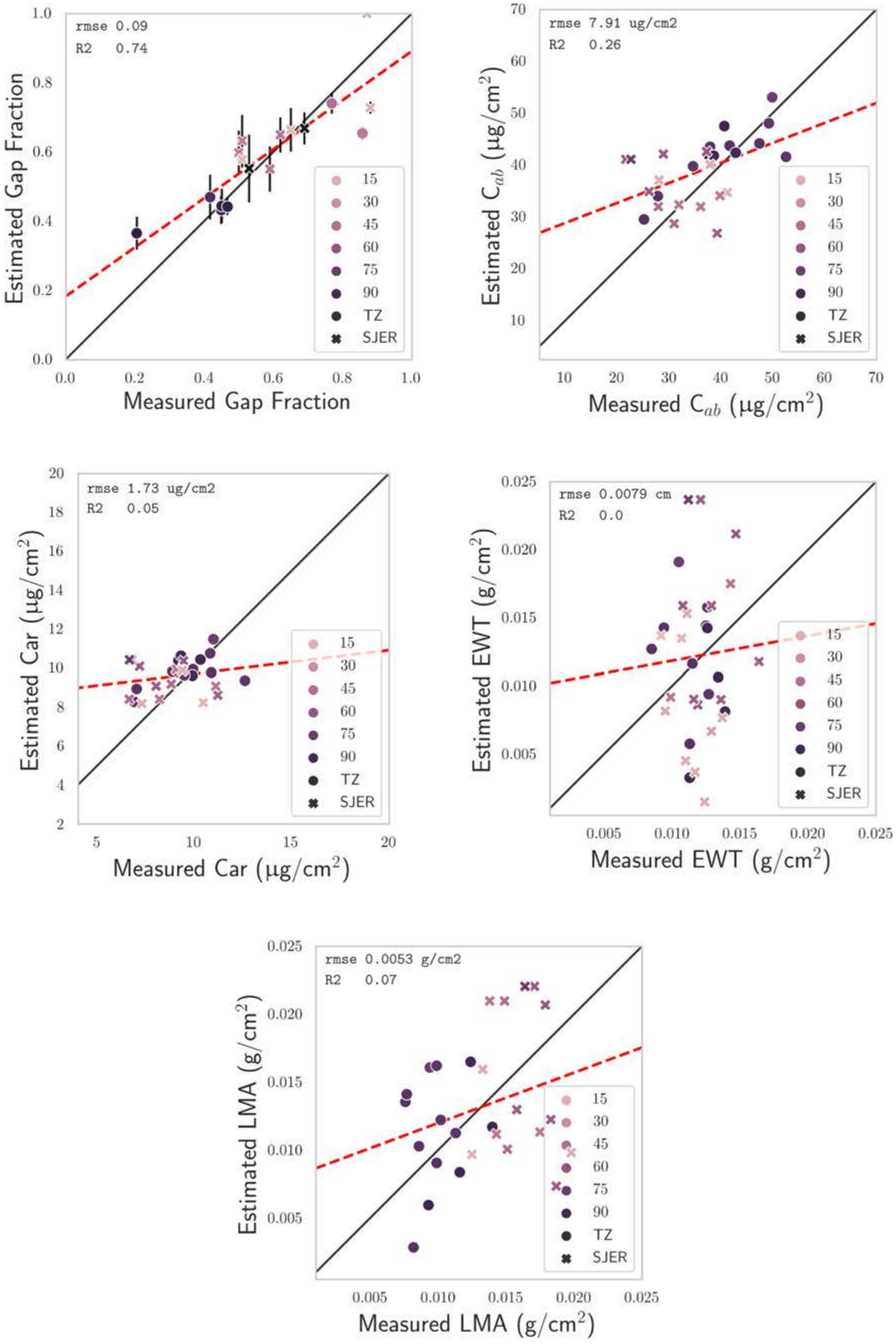


Figure 6.7: Comparison between vegetation traits as estimated by the PLSR and the *in situ* measurements for the AVIRIS-C images: gap fraction, C_{ab}, Car, EWT, LMA. The hue of the markers depends on the CC of the hyperspectral pixels that were considered. Black markers correspond to zones where CC information was not available.

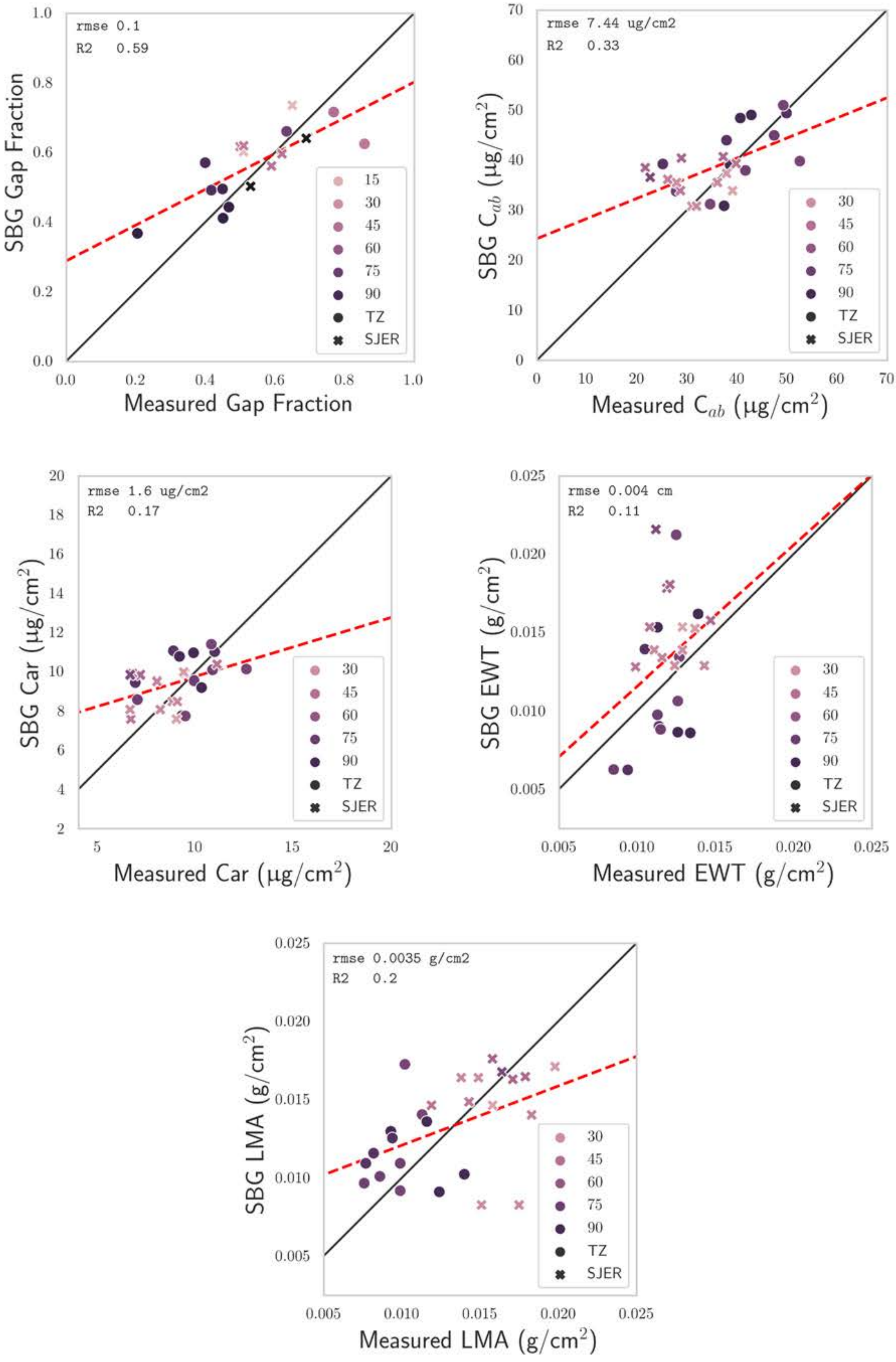


Figure 6.8: Same as Figure 6.7, but for SBG images.

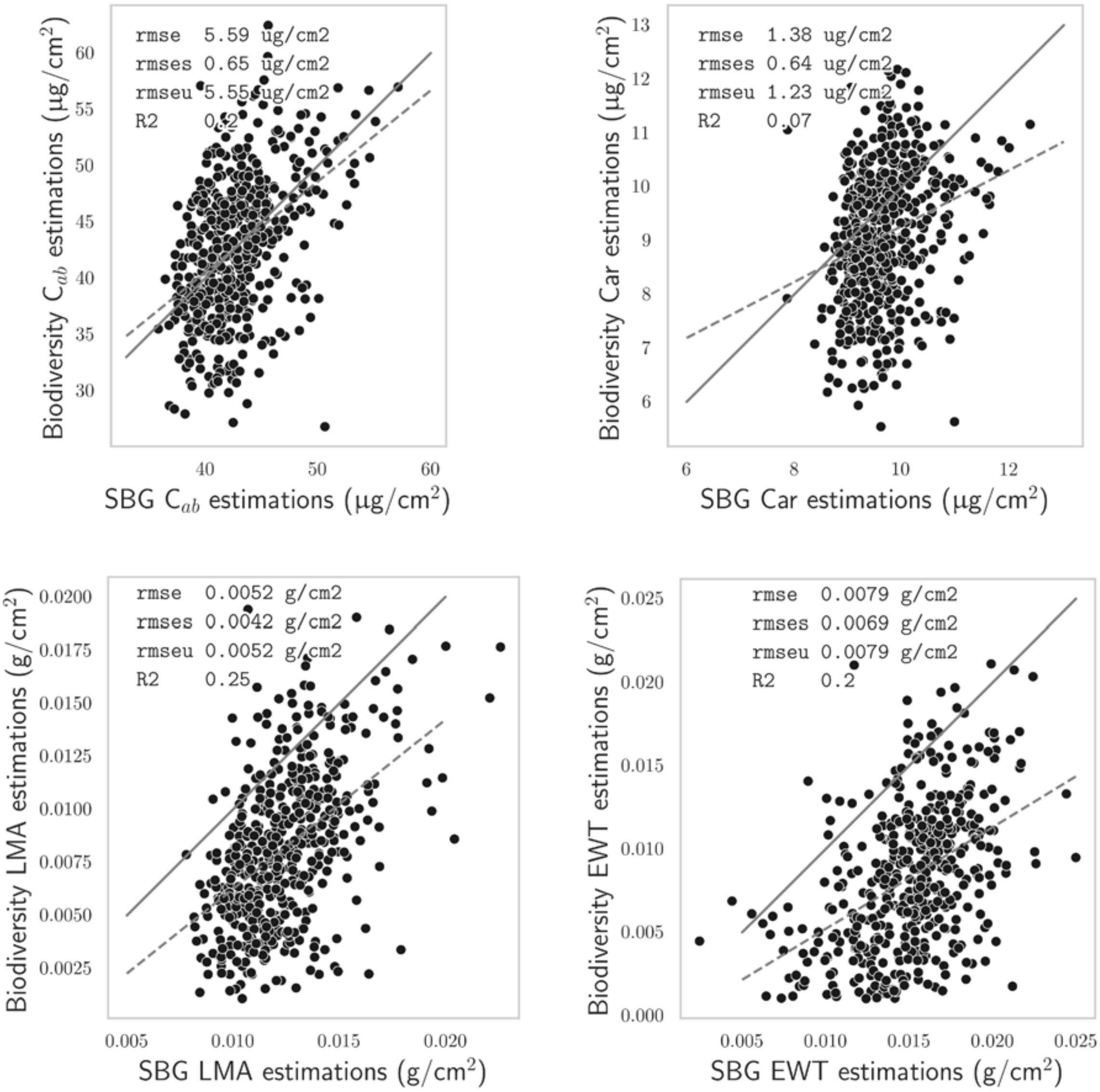


Figure 6.9: Comparison between estimates obtained from the Biodiversity and SBG images for pixels with 30% $\leq$ CC $\leq$ 50%.

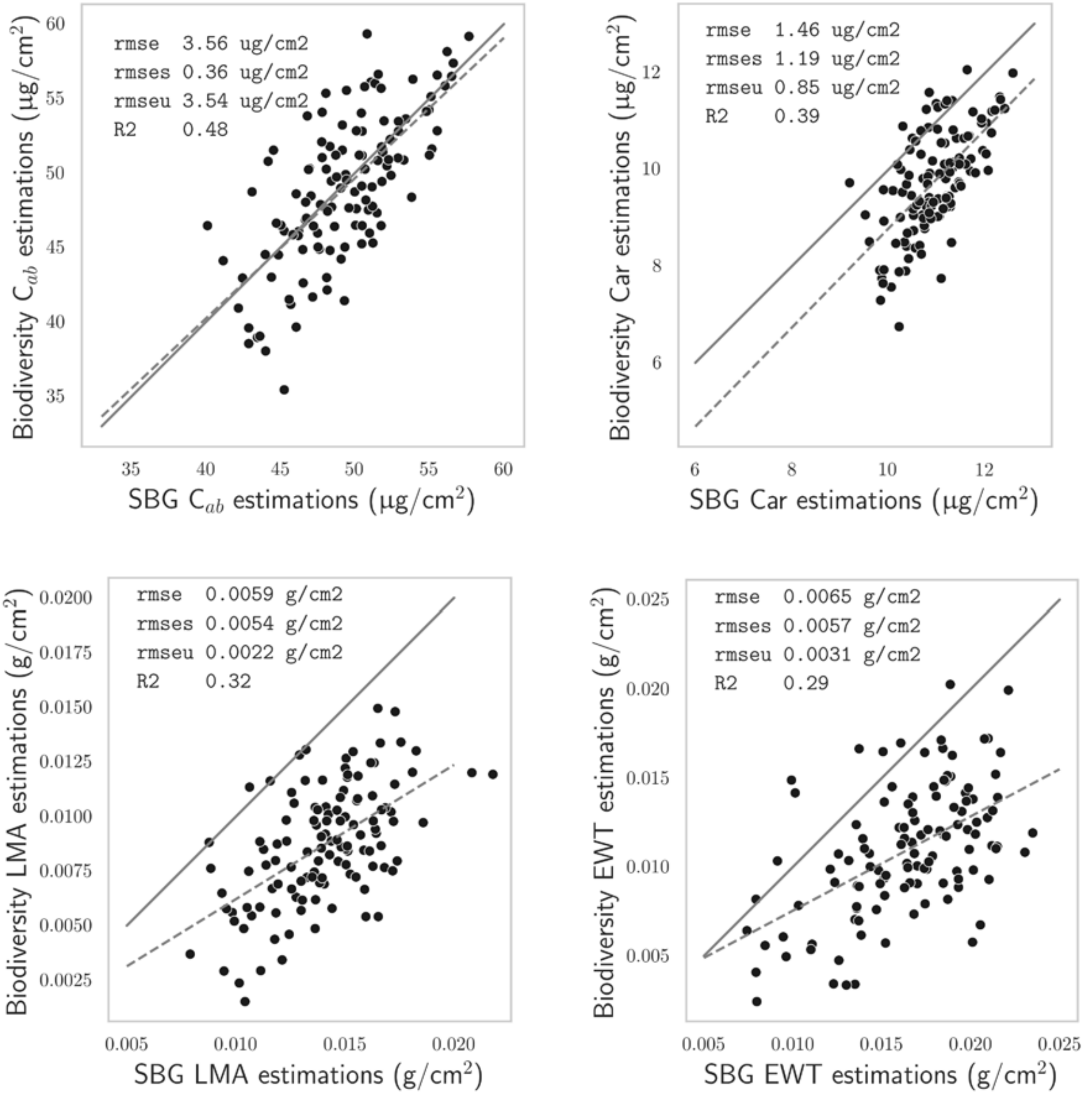


Figure 6.10: Same as Figure 6.9 but for pixels with 80% $\leq$ CC.

Table 6.6: Comparison between the values estimated from different sensors over two CC zones of TZ.

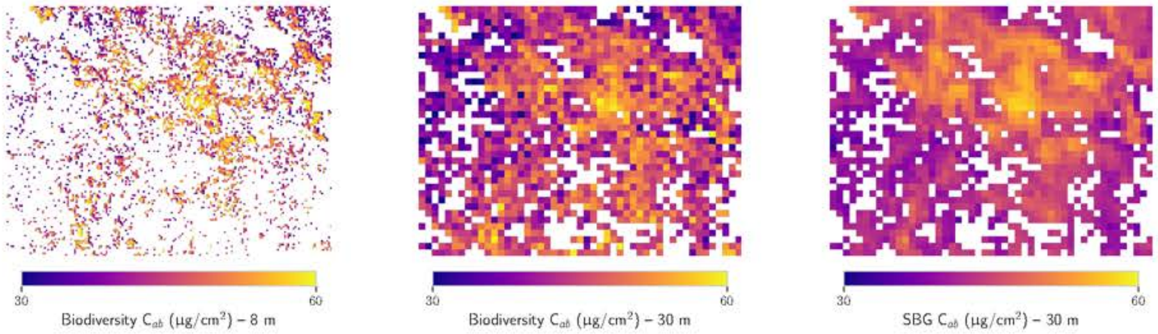
	30 % ≤ CC ≤ 50 %					80 % ≤ CC				
	Gap fraction	Cab	Car	EWT	LMA	Gap fraction	Cab	Car	EWT	LMA
Comparison between SBG and AVIRIS-C estimates										
RMSE	0.1	2.86	2.4	0.0034	0.0021	0.12	3.65	3.22	0.0030	0.0021
RMSEs	0.09	2.21	2.3	0.0009	0.0001	0.11	3.24	3.17	0.0001	0.0004
RMSEu	0.1	1.81	0.66	0.0034	0.0021	0.05	1.69	0.52	0.0030	0.002
R^2	0.35	0.34	0.52	0.5	0.55	0.42	0.52	0.66	0.59	0.6
Comparison between SBG and Biodiversity estimates										
RMSE		5.59	1.38	0.0079	0.0052		3.56	1.43	0.0065	0.0059
RMSEs		0.65	0.64	0.0069	0.0042		0.36	1.19	0.0057	0.0054
RMSEu		5.55	1.23	0.0079	0.0052		3.54	0.85	0.0031	0.0022
R^2		0.2	0.07	0.2	0.25		0.48	0.39	0.29	0.32

TZ for June 2014. The estimation map from Biodiversity (Figure 6.11a) is spatially averaged to attain a 30 m resolution (Figure 6.11b) and allow comparison with the SBG-derived map (Figure 6.11c). For the CC categories 20–40%, 40–60%, 60–80% and 80–100%, the R^2 between Biodiversity and SBG estimates were computed. As visible in Figure 6.11d, the lower the CC, the lower the correlation. For the savanna portion of TZ, with a CC $\leq$ 60%, the R^2 is lower than 0.24. Estimates from the sensors are mostly in agreement over the CC $\geq$ 80% areas (with a R^2 of 0.48), that are spatially very limited.

6.4 Discussion

6.4.1 Adequacy of the Training Databases to the Hyperspectral Images

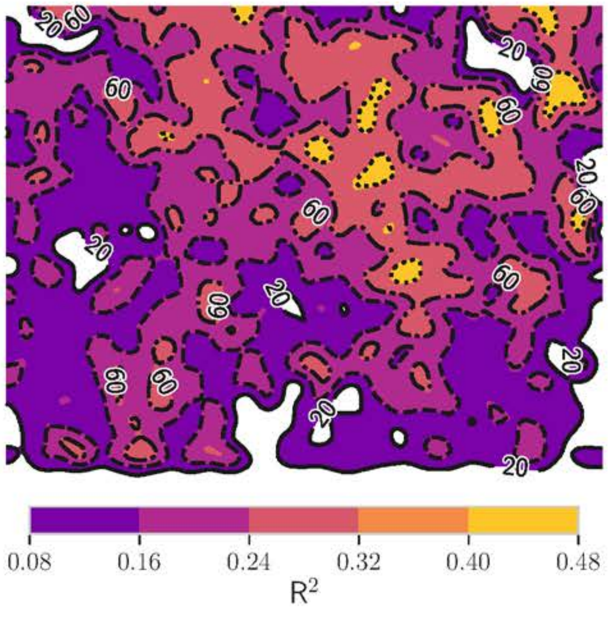
As visible in Figures 6.5 and 6.6, over parts of the 0.5–2.4 µm range the vegetation reflectance was not within the synthetic databases' boundaries, especially over the 0.8–1.5 µm range. As seen in the Chapter 5, this may partly be due to the lack of detail in the modeling of the trees within DART. The findings were in agreement with the results of Widlowski, Côté, and Béland, 2014 who showed that, in both the red and NIR regions, a simplified forest representation could lead to high biases (> 10 %) for the spatial resolutions considered in the present study. Moreover, Malenovský et al., 2008a demonstrated that introducing small woody components in the DART tree crowns would lead to lower reflectance in the NIR. As a consequence the present study only focused on the 1.5–2.4 µm spectral interval to for EWT and LMA retrieval. In a more general manner, the present study considered an ellipsoidal LAD with a free ALA varying from 55 to 65 °, broadly encompassing the spherical and erectophil LAD. Indeed, from 1200 leaf inclination measurements, Ryu et al., 2010 had found that, in summer, TZ tree leaves



(a) C_{ab} estimations from synthetic Biodiversity imagery

(b) Same as Figure 6.11a, but aggregated to a 30 m resolution

(c) C_{ab} estimations from synthetic SBG imagery



(d) Correlation between Biodiversity and SBG estimates of various CC zones. The lines indicate iso-CC (%).

Figure 6.11: Comparison between C_{ab} estimates from Biodiversity and SBG imagery over TZ.

presented an ALA of 62 ° and an erectophil LAD. Erectophilic leaves have indeed been observed in arid environments, and associated with drought-tolerance adaptation to reduce midday heating and maximize the photosynthetic activity during morning and evening, when the leaf water potential is highest (Crawford and Gibson, 1997). However LAD may vary spatially and temporally, and these results may no have been applicable to other seasons and sites. A spherical LAD is often assumed in the absence of *in situ* measurements (Bonan et al., 2011; Ryu et al., 2011). This assumption is questionable: both Pisek et al., 2013 and Liu et al., 2019 found, for temperate broad-leaved forests, that planophil or plagiophil distributions, whose ALA are closer to 30 and 45 °, respectively (Chianucci et al., 2018), were more representative of the LAD of mature stands and may be more adequate for modeling purposes.

Another possible limitation in the DART modeling was the ground reflectance, that was directly set as the average of understory pixels from the hyperspectral image. This was done as it was one of the most simple and straightforward methods, easily implementable in an operational context. Chapter 5 showed that using a flat ground with a lambertian reflectance extracted from the hyperspectral images did not lead to any significant bias concerning tree crown reflectance. However, within-site variability of ground reflectance was not taken into account, and the database was built considering only the median value of the pure ground pixels identified within the images. Bach and Verhoef, 2003 and Darvishzadeh et al., 2008 showed that soil backgrounds strongly affected canopy reflectance, especially at low LAI. The acquisitions done in the present study always took place during dry days, limiting the effects of soil moisture affecting the reflected variation. However, Eriksson et al., 2006 demonstrated that the understory vegetation had a significant effect on the red and NIR spectral ranges: spatial heterogeneity of the understory could have led to inadequacies of the SJER databases for some parts of the site, as SJER canopies are considerably more open than TZ on average (30 % and 50 %, respectively).

A solution to improve the adequacy of the synthetic databases for inversion purposes could be either to take into account the heterogeneity of the ground reflectance, by making it a free-varying parameter, or to build one database per ground reflectance and use the more adequate one based on the spatial information from the image. The first solution, while in theory leading to very generalisable databases, may lead to suboptimal fits when training machine-learning methods, but could be employed no matter the GSD of the sensor that acquired the images. The second method would require a high spatial resolution to clearly distinguish the ground surrounding the tree crowns and choose the best-corresponding database: while not realistic for SBG, this could be done with images acquired with sensors with a finer spatial resolution such as Biodiversity.

6.4.2 Vegetation Trait Estimates from Airborne and Satellite Hyperspectral Imagery

The results displayed in Figure 6.7 showed that, for the more closed parts of the canopy, a PLSR fit on synthetic data was able to retrieve C_{ab} and Car with high accuracy. While the dispersion was higher in the open parts, RMSE remained low overall when considering the whole dataset, and the PLSR estimates remained close to the *in situ* values. Moreover, gap fraction estimates are good no matter the CC and, evidently, related to it: it is therefore possible to get an idea of the estimates' accuracy by assessing their respective gap fractions. The good results for gap fraction are particularly encouraging as, as specified in Section 6.2.2, the gap fraction variation range was only 19–55 % in the *training* database: values outside of this range were also well predicted. Overall, estimated gap fraction, C_{ab} and Car values were in line with what was obtained in Chapter 5 and in previous studies. In Chapter 5, a LUT-based inversion with less variable parameters had been used, reducing the possible generalization of the method, and being more time-consuming in the inversion phase. The accuracy obtained here for gap fraction estimates (RMSE: 0.09; R^2: 0.74) was in line with what had obtained Baret, Clevers, and Steven, 1995 when using a neural network trained with synthetic data over crop fields (RMSE: 0.09; R^2: 0.9). The present study shows that equivalent accuracies are attainable with a PLSR trained on more general synthetic data.

Similar performances were obtained with the SBG hyperspectral images. As indicated in Table 6.5, it seems that C_{ab}, Car, EWT and LMA estimations are all improved for this GSD, as all RMSE values are lower when comparing estimated values with *in situ* measurements. However, it appears that AVIRIS-C and SBG images are overall equivalent when it comes to estimations over TZ, as shown Table 6.6. The slightly coarser spatial resolution (30 m instead of 18 m) does not seem to affect much retrieval accuracy, which is encouraging: results obtained with, and methods developed for, AVIRIS-C images may be directly transportable to SBG.

Over the high-CC areas, Biodiversity and SBG images led to similar C_{ab} and Car estimation. For those areas the influence of understory reflectance is minimal (CC $\geq$ 80%), and as discussed above estimate accuracy from the SBG images was good. This correlation between Biodiversity and SBG estimations further confirms that acceptable retrieval is possible from Biodiversity images, despite its lower SNR. Moreover, as shown in Figure 6.4, the majority of vegetation pixels in the Biodiversity images have CC higher than 80%: Biodiversity is barely affected by the openness of canopies as tree crowns can be, most of the time, isolated. Thus, the lack of relationship between Biodiversity and SBG estimates for low-CC zones is illustrative of the advantage of a finer spatial resolution for mapping vegetation traits over open canopies. These results are consistent with what Carter et al., 2009 and Jafari and Lewis, 2012 found: the spatial resolution of the hyperspectral images is a limiting factor, especially for open canopies: identifying spectrally pure pixels used for estimation or classification purposes is all the more difficult. Heldens et al., 2011 suggested to use data fusion to make EnMAP acquisitions fit for use in an urban context, as their original resolution would not allow to distinguish the various objects. In a similar vein, Siegmann et al., 2015 showed the potential of pan-sharpening for improving LAI retrieval accuracy of wheat fields when working with EnMAP images. The high spatial resolution of Biodiversity could therefore prove useful either for direct use in the retrieval of vegetation traits, or indirectly through data fusion with images acquired by other sensors. More recently, working with Sentinel-2A images, Zarco-Tejada et al., 2019 have shown that being able to account from the specificities of each pixel components was critical to obtain acceptable C_{ab} estimations over sparse canopies. Overall, there seems to be a clear potential in taking advantage of the smaller GSD of Biodiversity to improve vegetation traits' retrieval, either directly from the sensor's acquisitions or indirectly through data fusion, despite its lower expected SNR.

The present study did not have the *in situ* measurements to directly map LAI from either the SBG or Biodiversity images. While performances for the gap fraction estimations using SBG images hint at potentially good accuracy for LAI, this would necessitate further works to confirm. Validation of Biodiversity LAI estimations would require the field measurement of isolated trees' LAI, possibly requiring lidar measurements to obtain these values (Béland et al., 2014; Klingberg et al., 2015).

No clear conclusion can yet be drawn for EWT and LMA: for the high-CC areas, the variation range of the *in situ* measurements is rather limited, so that it is not possible to say whether or not estimations from the SBG image (Figure 6.8) are really accurate. However, it must be noted from Table 6.6 that EWT and LMA estimates from the high-CC areas are similar between Biodiversity and SBG. While RMSE is high, it is mostly due to a systematic bias, and the unsystematic RMSE is only 0.0022 g/cm^2 for both EWT and LMA, while the variation range is roughly 0.007–0.023 g/cm^2. That estimations from hyperspectral images with so different spatial resolutions show similar behaviors may indicate that, while some bias is present, EWT and LMA can be retrieved from SBG images for closed canopies and *a fortiori* from Biodiversity images.

6.5 Conclusion

This study assessed the potential of the future hyperspectral missions SBG (30 m GSD) and Biodiversity (8 m GSD) for the mapping of vegetation traits (gap fraction, C_{ab}, Car, EWT, LMA) of broad-leaved trees in two woodland savannas in California, using synthetic data generated from AVIRIS-NG and AVIRIS-C images. The accuracy of the traits' estimates from an hybrid method, using a PLSR trained on DART-generated data, was assessed through (i) direct comparison with *in situ* measurements and (ii) intercomparison between the values obtained from AVIRIS-C, Biodiversity and SBG images. SBG and AVIRIS-C gap fraction, C_{ab}, and Car estimates showed great correspondence with ground-based values (RMSE $\sim$ 0.1, R^2 $\sim$ 0.65; RMSE $\sim$ 7.7 µg/cm^2, R^2 $\sim$ 0.3; RMSE $\sim$ 1.7 µg/cm^2, R^2 $\sim$ 0.1, respectively), however concerning pigments this was more true for the densest parts of the canopy. At the landscape scale, there was in general good agreement between retrievals from SBG and AVIRIS-C imagery for all traits (R^2 ranging from 0.34 to 0.66). Biodiversity estimates were also coherent with SBG estimates over the closed parts of the canopy (R^2 ranging from 0.32 to 0.48), giving confidence in the accuracy of the estimates, but exhibited different behaviors over the open parts, probably owing to increased spectral mixing in the SBG pixels affecting the estimation.

This work not only demonstrates that gap fraction, C_{ab} and Car estimations are possible using 30 m GSD satellite data over the more closed parts of canopies without detailed information concerning the imaged landscape, but also shows the potential of $\sim$ 8 m GSD images to map these vegetation traits considered in this study over theoretically any canopy. Further studies, focusing on different ecosystems, are still necessary to confirm these results. Despite their lower SNR, hyperspectral satellite sensors with finer spatial resolution could prove to be very beneficial to the global mapping of vegetation traits by complementing sensors with coarser GSD.

General conclusion and perspectives

General conclusion

Due to increasing climatic and anthropic pressures, the Mediterranean ecoregions, that are recognized plant biodiversity hotspots, are some of the most endangered ecosystems. While these plants species are present in a multitude of habitats, most of the vegetation in Mediterranean-climate regions is distributed over open canopies as the summer aridity limits the development of true forests. Monitoring trees' health status is crucial to guide conservation policies: airborne and satellite remote sensing methods, which can provide information about large landscapes in a regular fashion, are most adapted for a future global monitoring. However, difficulties arise when retrieving information from open forests, as the contribution of tree crowns to the measured radiative signal may be limited compared to the understory's.

This thesis aimed at identifying potential methods for the estimation of various vegetation traits (LAI, gap fraction, leaf pigment contents, EWT, LMA) from passive hyperspectral remote sensing over open canopies. In the context of the recent and future hyperspectral satellite missions and the monitoring at the landscape scale, these methods had to rely on little to no field knowledge in the calibration phase. To do so, this study focused on two oak-dominant woodland savannas where *in situ* measurements and airborne hyperspectral acquisitions had been regularly undertaken over several years. Previous studies had shown the potential of physically-based and hybrid methods for the monitoring of traits at the landscape scale, however mostly for closed canopies or from images with a high spatial resolution. The approach chosen for this thesis was to first identify appropriate retrieval methods from airborne hyperspectral imagery, as this was considered a prerequisite before moving to synthetic satellite images. Initially, the research focused on the retrieval of LAI and leaf chlorophyll content, the two most commonly estimated traits, as well as leaf carotenoid content. Then, exploratory work was undertaken to improve estimation accuracy and identify potential methods for the retrieval of EWT and LMA. Finally, the results were consolidated and their potential for the retrieval of vegetation traits in an operational context was evaluated using synthetic Biodiversity and SBG imagery.

First, an inversion strategy suitable for LAI and leaf pigments retrievals over open canopies had to be identified. A physically-based method, relying on a 3D RTM and LUT-based inversion, was proposed. It relied on an abstract modeling of the forest scene, designated as SFR, the trees being modeled with a cylindrical trunk and an ellipsoidal crown and the ground as a flat, lambertian surface. Multiple cost-functions, either relying on vegetation indexes or on spectral ranges, were tested. It appeared that, while over the 0.45–0.75 µm and 1.1–2.4 µm intervals, the reflectances computed by DART had a behavior similar to the AVIRIS-C images, they were overestimated over the 0.75–1.1 µm region, probably indicating that the canopy cover was less than modeled. The VI, less sensitive to differences in reflectance levels, were found to be more appropriate for cost functions. The proposed method successfully estimated the vegetation traits over hyperspectral images acquired in both summer and fall (for LAI, RMSE: 0.22 m^2/m^2, R^2: 0.8; for C_{ab}, RMSE: 5.21 µm/cm^2, R^2: 0.73; for Car, RMSE: 1.34 µm/cm^2 R^2: 0.59), and was shown to correctly catch the LAI decrease between fall and summer for two years at the landscape scale.

To be able to estimate EWT and LMA, the analysis was twofold. The first idea was to improve the LUT sampling scheme and, so as to have access to more precise estimates. In order to reduce the time-cost of extensive LUT building when using a 3D RTM, an approximation method designated P2D, relying on both PROSAIL and DART, was implemented. This method was designed so that the reflectance of open canopies modeled within DART could be approximated from PROSAIL outputs, greatly reducing the time-cost of generating an extensive LUT. After calibration, the approximation error, calculated relatively to the calibration database's sampling scheme, was shown to be negligible even when predicting DART outputs for canopies with a 30% CC. Very large P2D LUT with more than 100,000 entries were then generated for each hyperspectral image, and it appeared that, while estimate accuracy was indeed improved for gap fraction and leaf pigments compared to those obtained with the initial DART LUT, EWT and LMA could still not be estimated with either VI-based cost-functions or RMSE or SAM-based cost-functions.

The second idea was to study the influence of the modeling within DART on EWT and LMA retrievals, focusing on high spatial resolution (2 m) imagery, which should be the most affected. Starting from the SFR scenario, various structural details were introduced in a monovariant fashion, and their effect on tree crown reflectance was assessed. It appeared that most of the spectral variations caused by these details occured over the 0.8–1.3 µm range, the spectral range already identified as problematic in the first studies, the 1.5–2.4 µm being considerably less affected. A hybrid method was then used for estimation purposes. Random forest regressors were trained to estimate EWT and LMA with two different databases, generated using the SFR scenario and a detailed scenario (DETAIL) matching the structural characteristics of the hyperspectral image of study, over both the (i) 0.8–2.4 and (ii) 1.5–2.4 µm ranges. When applied over synthetic hyperspectral images of realistic forests, it appeared that while estimate accuracy was significantly lower for SFR than for DETAIL in scenario (i), the databases led to similar accuracies in scenario (ii). A RFR trained with the SFR modeling over the 1.5–2.4 µm range was then applied over 18 m spatial resolution AVIRIS-NG hyperspectral images of two woodland savannas. Although the variation range of the *in situ* measurements was limited, estimate accuracy was encouraging (for LMA, RMSE: 0.0031 g/cm^2, R^2: 0.23; for EWT, RMSE: 0.0030 g/cm^2, R^2: 0.36), illustrating the interest of the 1.5–2.4 µm range for EWT and LMA retrievals when abstractions are made in the RTM modeling.

Estimation of all the vegetation traits from airborne imagery was shown to be possible using only an abstract modeling in the RTM, realistically achievable in an operational context. As hybrid methods are expected to be the most adequate for future monitoring efforts, the accuracy of a hybrid method using a SFR modeling employed over synthetic 8 m (Biodiversity) and 30 m (SBG) spatial resolution hyperspectral satellite data was assessed. PLSR, that increasingly used to link reflectance to vegetation traits (Meacham-Hensold et al., 2019), were trained with databases generated using DART and a SFR modeling and applied over AVIRIS-C and the synthetic satellite images to estimate gap fraction, leaf C_{ab}, Car, EWT and LMA. It appeared that estimations from SBG and AVIRIS-C matched *in situ* measurements for gap fraction, C_{ab} and Car. Between the sensors, it was found that estimates were well correlated (R^2 ranging from 0.34 to 0.66, depending on the variable) over both open and closed canopies, demonstrating that SBG could be considered equivalent to AVIRIS-C concerning estimation performances. Similarly, estimates from Biodiversity were well correlated with those from SBG over the high-CC parts (R^2 ranging from 0.32 to 0.49) of the canopy, indicating good retrieval performances from Biodiversity images too. Over the low-CC parts however, correlation was low. As Biodiversity images can mostly isolate tree crowns, they are not affected by the openness of the canopy, contrary to SBG's: this absence of correlation indicates the great potential of Biodiversity for the mapping of very open tree-grass ecosystems.

This thesis demonstrated that physically-based and, more importantly, hybrid methods, were suitable for the estimation of a multitude of vegetation traits of open broad-leaved canopies from medium resolution (8 m to 30 m) hyperspectral images in an operational context, necessitating little to no knowledge about the imaged landscapes: an abstracted forest representation within a 3D RTM, using ground reflectance directly extracted from the images, is sufficient for accurate estimates. The proposed methods should be directly applicable over any broad-leaved canopy where pixels representative of the understory can be identified. The methods were validated using images acquired over different sites and at different dates, however validation could benefit from more *in situ* measurements over other sites, in order to cover a broader range of variation of the estimated traits. This would require additional airborne acquisitions and field campaigns to confirm the findings in a more general context, and identify possible limitations.

Perspectives

Taking into account the Spatial Heterogeneity of the Landscape

As discussed in Chapter 6, a major drawback of the 30 m spatial resolution of SBG was the inability to isolate the elements of a forest scene. In particular, while pure-ground pixels could still be found in a given SBG image of an open canopy, its use in a DART scene relies on the assumption that its reflectance is representative of that of the understory, which may not be the case. This is problematic, as the understory can have a significant effect on pixel reflectance, especially for open canopies (Darvishzadeh et al., 2008). Spectral unmixing may be a solution: from suitable endmember families, it could be possible to preliminarily identify the different materials (crown and types of understory in the present case) in each pixel. Over a landscape, a fine classification of the understory could then allow for the modeling of more adequate DART scenes by using the most appropriate ground reflectance for each understory group. Thus, more specialized – one could say, more correct – estimators could be trained and would, in theory, be more accurate.

As seen in Chapter 6, SBG and Biodiversity could complement each other in an operational context. On one hand, Biodiversity, with its more frequent revisit time (5 days) and better spatial resolution, could help to calibrate classification methods employed over SBG images, that present a significantly larger swathe. Moreover, thanks to the better spatial resolution, it could be possible to identify of areas where estimates derived from SBG images are expected to be less accurate due to the presence of various object not resolved at 30 m. Over landscapes presenting a strong gradient, Biodiversity images could help interpret SBG-derived estimations. On the other hand, as shown in Section 6.3.4 and Figure **??**, estimates from SBG and Biodiversity are coherent over homogeneous areas, although some bias was present when estimating traits with a low spectral signature such as Car. Because of its higher SNR, it is possible that SBG images would be more adequate to estimate low-signature traits over homogeneous zones. Therefore, it could be possible to calibrate estimation methods dedicated to Biodiversity images on the SBG estimates. This would allow for precise estimation of traits over homogeneous landscapes using SBG images, and over highly heterogeneous landscapes using Biodiversity images and calibrated retrieval methods. In a more general manner, these synergies should also be applicable when considering SHALOM in lieu of Biodiversity, or EnMAP or PRISMA in lieu of SBG.

The methods proposed during this thesis all trained over databases included all canopy closures from 30 to 90%: while the estimators were more generic, they were necessarily less accurate as a drawback. Training estimators for each CC class, and use them over the appropriate areas, would certainly help improve estimates accuracy. This problem can be tackled in various ways. A first solution could be to use the global high-resolution panchromatic images already available to classify forest CC at the target spatial resolution. However, difficulties could arise if the canopy cannot be distinguished well from the understory with the few spectral bands available, as could be the case in particular for tree-grass ecosystems for periods when the grass is active. Another solution could be to train a classifier on the outputs of a RTM and apply it on the hyperspectral images. However, this would rely on the assumption that the modeling is representative of the landscape which, as discussed above, is not a given depending on the spatial resolution.

Extending the Scope of the Findings to more Diverse Forests and Traits

In order to completely validate the results obtained in this thesis, it is necessary to assess the proposed methods' performances over environments presenting more diversity, specifically in terms of leaf traits. Relatively few *in situ* measurements were available during this thesis, covering only a limited range of the possible leaf traits' values especially for EWT and LMA: larger ranges would help give confidence in the findings. Nitrogen, which was not estimated in this work, and LMA are important traits whose estimation from satellite sensors would prove very beneficial for the understanding of plant response to external stress (Schimel et al., 2015; Westoby et al., 2004). EWT is also of great interest, as it has important implications in both agriculture, giving information on crop water status (Hunt and Rock, 1989), and forest fire prevention, since it gives an indication of how much the fire fuel is susceptible to burn (Yebra, Chuvieco, and Riaño, 2008). In particular, accurately estimating both EWT and LMA from SBG or Biodiversity imagery would allow for the precise estimation of FMC, one of the primary variables affecting wildfire ignition and spread (Yebra et al., 2013). This would prove invaluable in a context where wildfire frequency and extent is expected to increase significantly.

Moreover, over the course of this thesis, it was decided to move from LAI to gap fraction estimations: indeed, the field measurements used DHP, that cannot differentiate between leaves and NPV and is affected by leaf clumping. This could lead to inaccuracies concerning what was considered the ground truth, and more accurate measurement methods, such as lidar acquisitions, would have been necessary to properly assess the capacity of the proposed methods to retrieve LAI. In particular, no *in situ* measurement was available to assess how accurately LAI could be estimated from Biodiversity images, which would be an information of great interest.

The methods developed in this thesis were employed under the assumption that most of the imaged canopies consisted in broad-leaved species, that were dominant in the sites of interest. Indeed, while DART and PROSPECT are suitable to model broad-leaved trees, they would require extensive, fine tuning to be adapted to conifers. Malenovský et al., 2006 showed that recalibrating the specific absorption coefficients of chlorophylls and dry matter improved PROSPECT performances when applied to needle spectra, while Moorthy, Miller, and Noland, 2008 further showed that, while PROSPECT was already performing better than LIBERTY in the absence of *a priori* information on needle properties, normalization factors could help improve estimates accuracy. There is at the moment no reason to believe that the proposed methods should be suited for coniferous canopies, and the development of RTM satisfactory for needle and canopy modeling would help significantly. Coniferous forests constitute a major portion of the global land cover (Friedl et al., 2010), and in Europe, 16% of forests consist in coniferous open canopies (Zarco-Tejada et al., 2019), while conifers dominate boreal ecosystems (Caldwell et al., 2001; Majasalmi and Rautiainen, 2021). Identifying a retrieval method appropriate for conifers is therefore critical for the monitoring of vegetation traits at the global scale.

Finally, the tree-grass ecosystems present in both TZ and SJER had clearly distinct canopy and understories for the periods considered (summer and fall), which corresponded to the period of senescence/inactivity of the grass species. In Chapter 3, the presence of active grass patches were shown to have detrimental effects on the estimations, however this was most probably because the reflectance used for the ground in the RTM corresponded to a senescent understory. In the event of a photosynthetically active understory, spectral mixing may be all the more problematic when tree crowns cannot be isolated in the hyperspectral images, as reflectance of the understory and canopy layers would be very similar. Uncertainties in canopy traits retrieval would most certainly increase, as the results of Zarco-Tejada et al., 2019 seem to indicate. Tackling this issue would allow to extend the monitoring of open canopies over the entire annual phenological cycle.

Appendix A

Estimated vegetation traits over TZ through LUT-based inversion using a SFR modeling

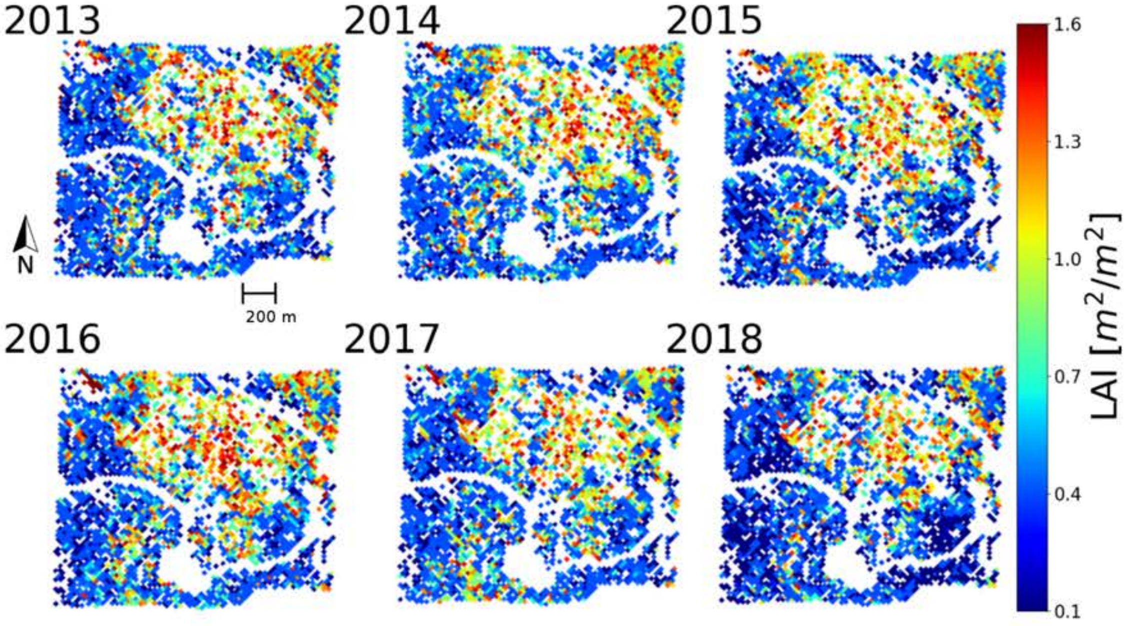


Figure A.1: Estimated canopy LAI from 2013 to 2018. Areas in white were masked as they contained roads, buildings, water, too few QUDO and PISA trees or too many PISA trees.

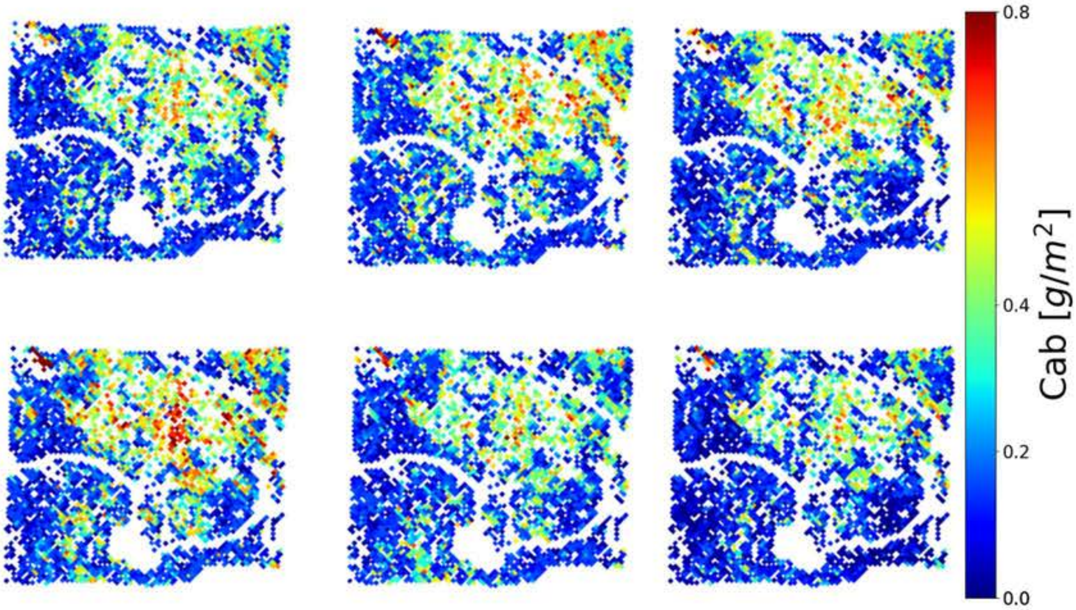


Figure A.2: Estimated canopy C_{ab} from 2013 to 2018. Areas in white were masked as they contained roads, buildings, water, too few QUDO and PISA trees or too many PISA trees.

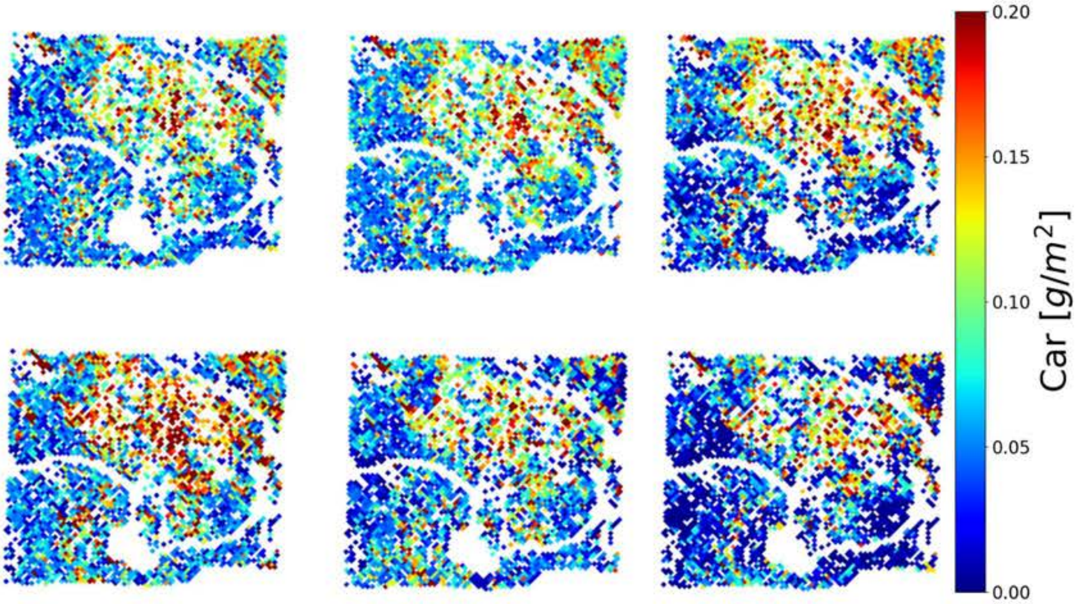
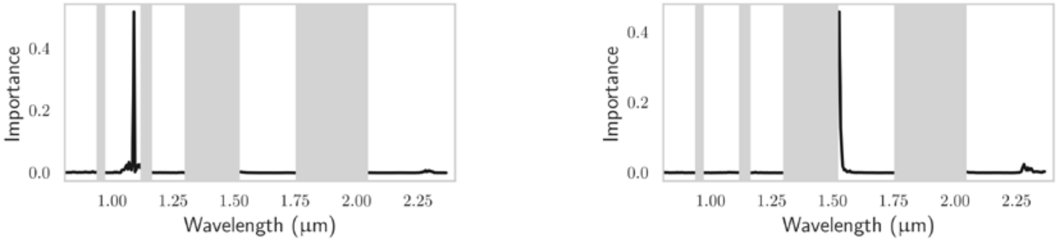


Figure A.3: Estimated canopy Car from 2013 to 2018. Areas in white were masked as they contained roads, buildings, water, too few QUDO and PISA trees or too many PISA trees.

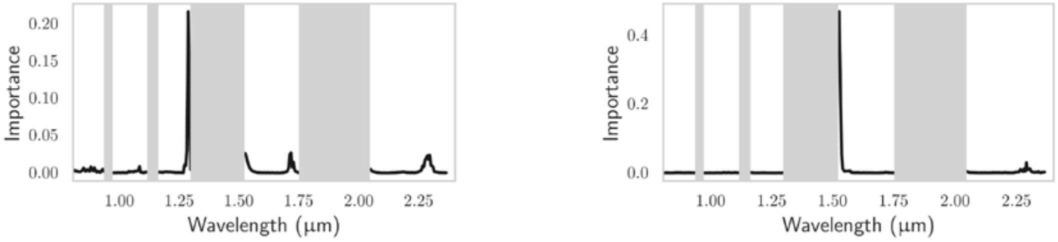
Appendix B

Feature importances of the RFR



(a) LMA – SFR

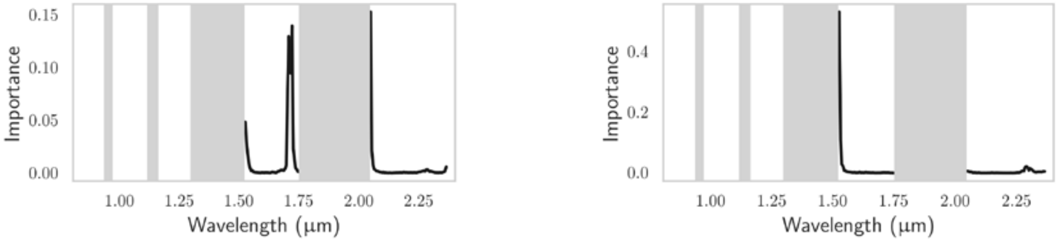
(b) EWT – SFR



(c) LMA – DETAIL

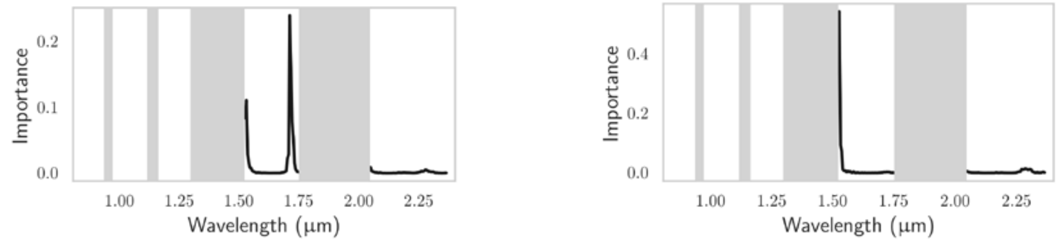
(d) EWT – DETAIL

Figure B.1: Feature importances of the RFR trained for the synthetic hyperspectral image over the 0.8–2.4 μm interval.



(a) LMA – SFR

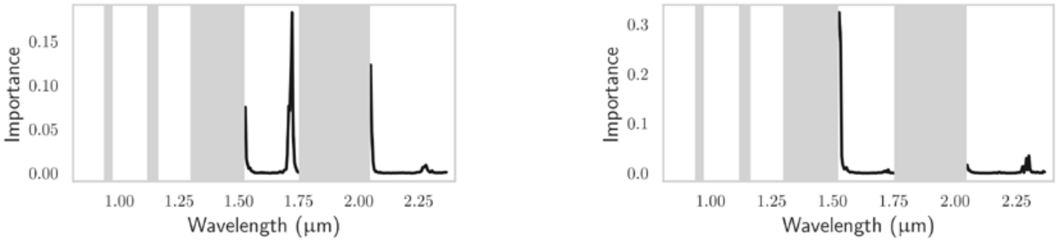
(b) EWT – SFR



(c) LMA – DETAIL

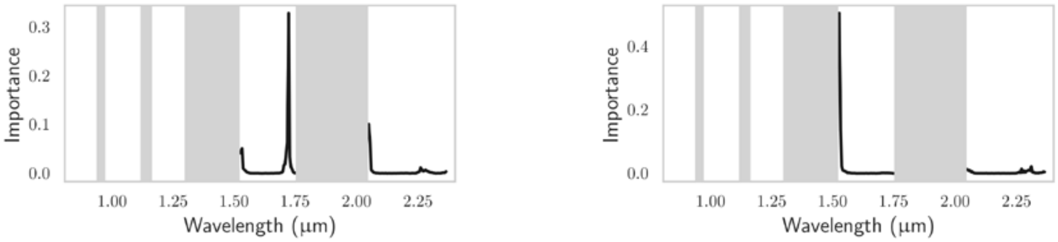
(d) LMA – DETAIL

Figure B.2: Feature importances of the RFR trained for the synthetic hyperspectral image over the 1.5–2.4 µm interval.



(a) LMA – SFR

(b) EWT – SFR



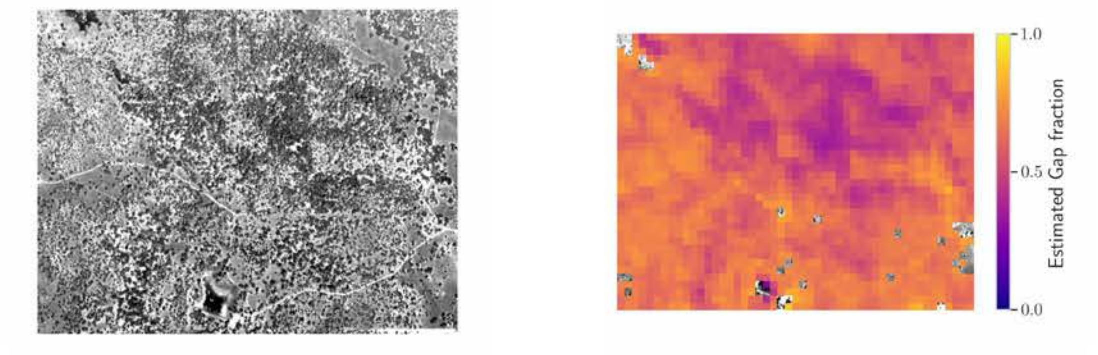
(c) LMA – SFR

(d) EWT – SFR

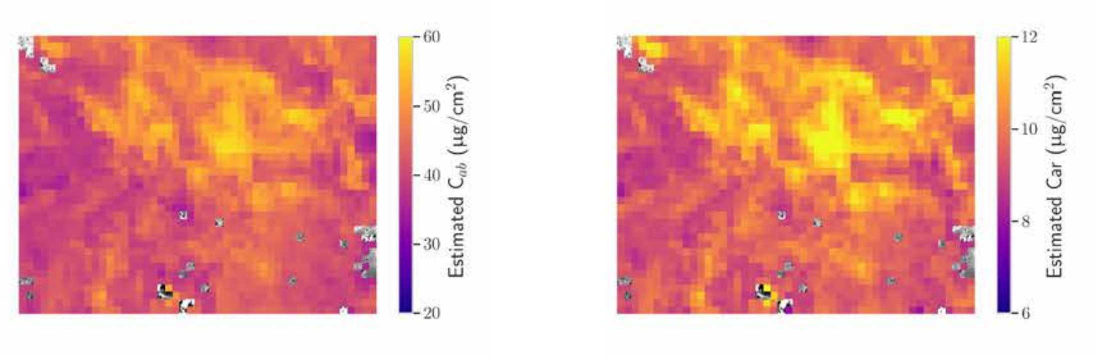
Figure B.3: Feature importances of the RFR trained for the AVIRIS-NG images over the 1.5–2.4 µm for (**a**,**b**) TZ and (**c**,**d**) SJER.

Appendix C

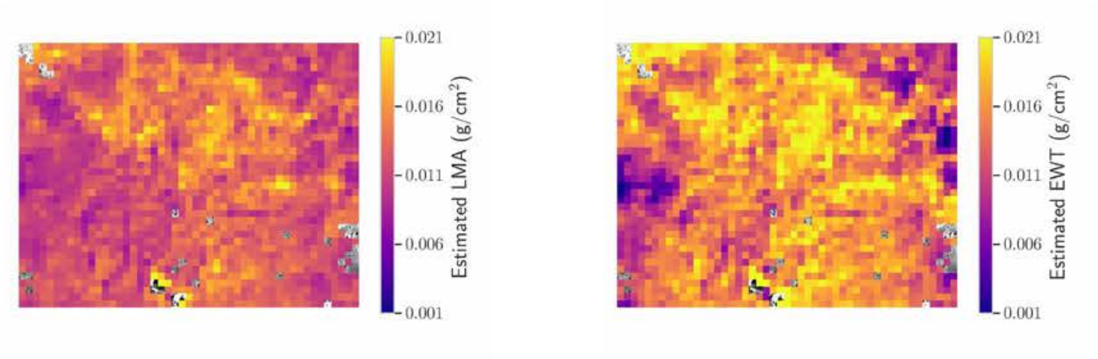
SBG and Biodiversity estimations



(a) Site map (b) Gap fraction

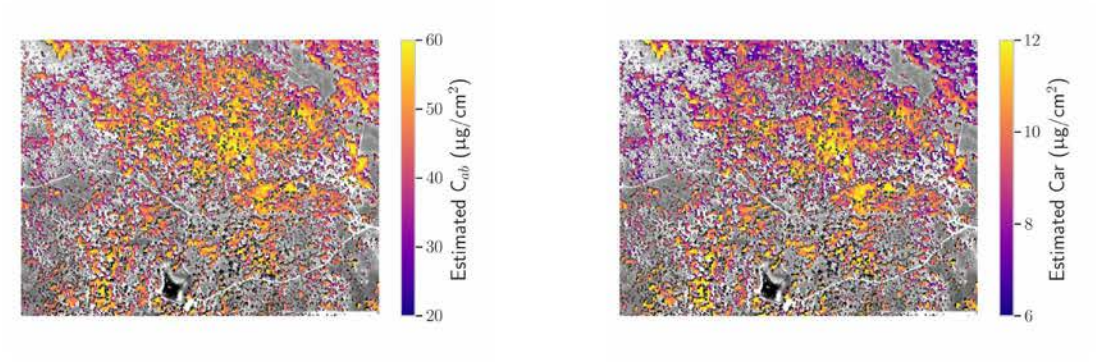


(c) C_{ab} (d) Car

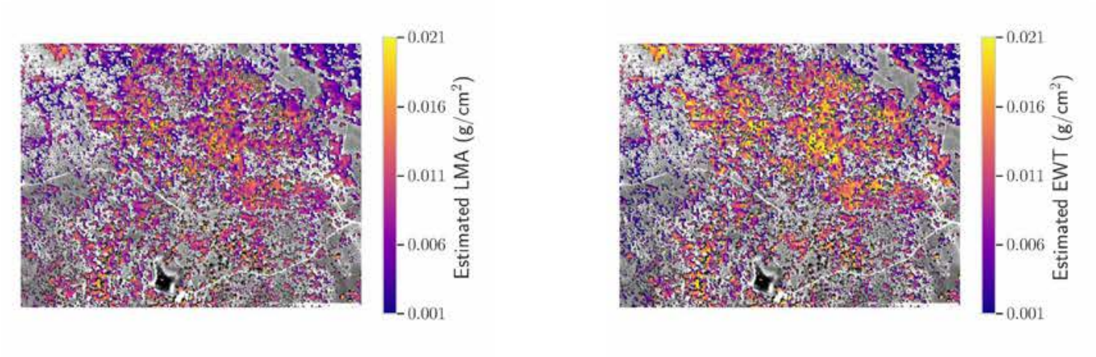


(e) LMA (f) EWT

Figure C.1: Estimation obtained over TZ for June 2014 from synthetic SBG imagery.

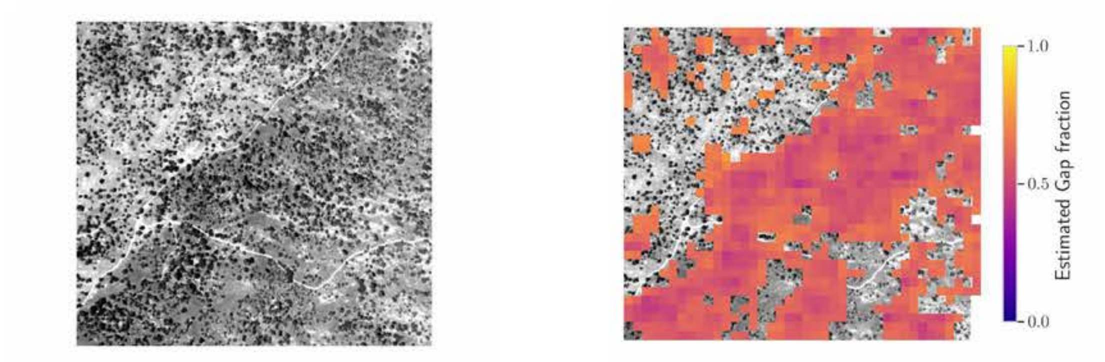


(a) C_{ab} (b) Car

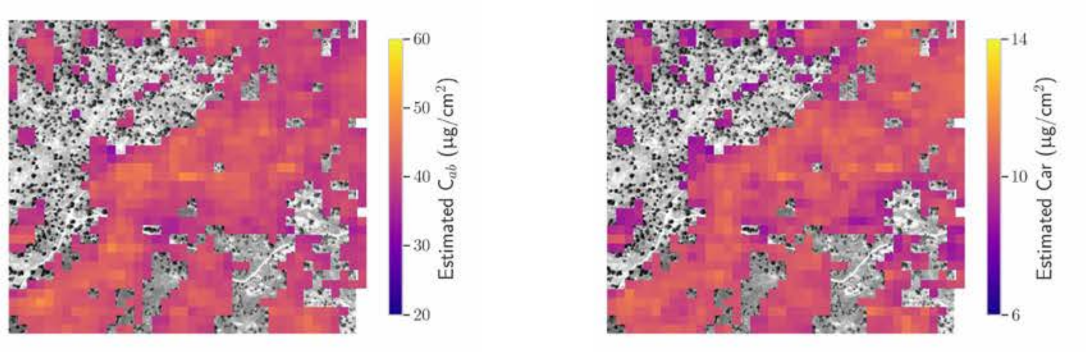


(c) LMA (d) EWT

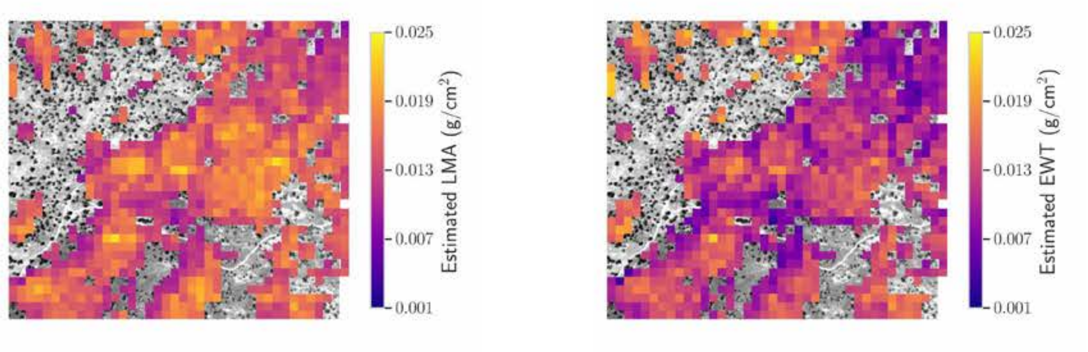
Figure C.2: Estimation obtained over TZ for June 2014 from synthetic Biodiversity imagery.



(a) Site map (b) Gap fraction

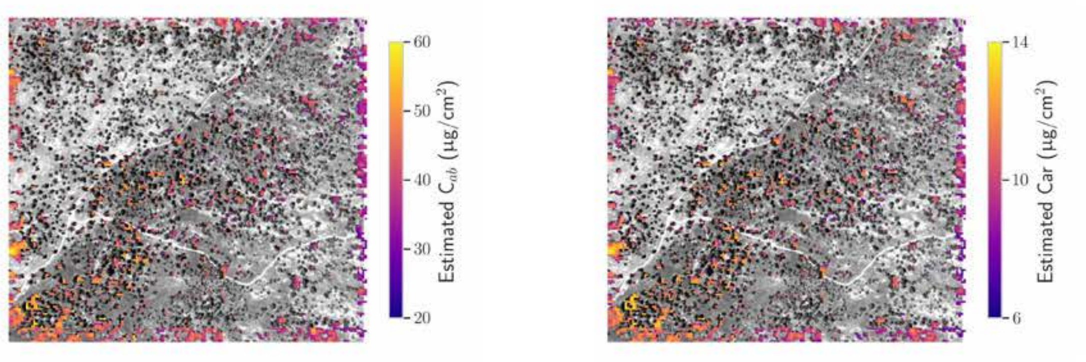


(c) C_{ab} (d) Car

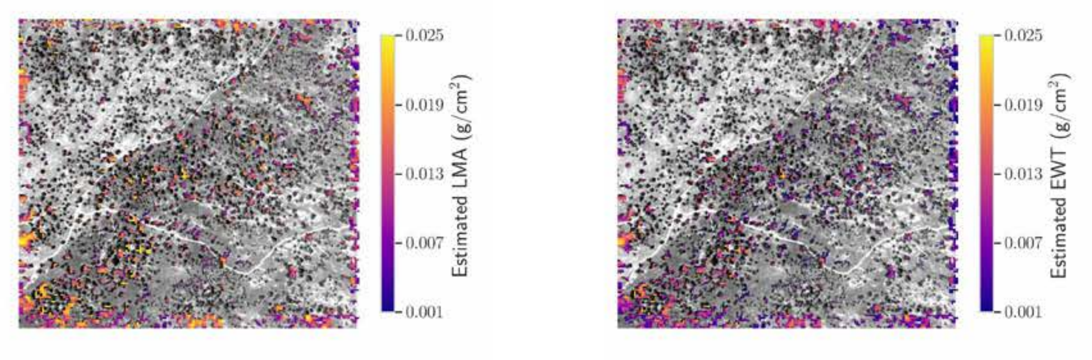


(e) LMA (f) EWT

Figure C.3: Estimation obtained over SJER for June 2014 from synthetic SBG imagery.



(a) C_{ab} (b) Car



(c) LMA (d) EWT

Figure C.4: Estimation obtained over SJER for June 2014 from synthetic Biodiversity imagery.

Publications and Communications

Peer-Reviewed Publications

Miraglio, T.; Adeline, K; Huesca, M.; Ustin, S.; Briottet, X. Monitoring LAI, Chlorophylls, and Carotenoids Content of a Woodland Savanna Using Hyperspectral Imagery and 3D Radiative Transfer Modeling. *Remote Sens.* **2020**, 12, 28. https://doi.org/10.3390/rs12010028.

Miraglio, T.; Adeline, K.; Huesca, M.; Ustin, S.; Briottet, X. Joint Use of PROSAIL and DART for Fast LUT Building: Application to Gap Fraction and Leaf Biochemistry Estimations over Sparse Oak Stands. *Remote Sens. 2020*, 12, 2925. https://doi.org/10.3390/rs12182925

Miraglio, T.; Huesca, M.; Gastellu-Etchegorry, J.-P.; Schaaf, C.; Adeline, K.R.M.; Ustin, S. L.; Briottet, X. Impact of Modeling Abstractions When Estimating Leaf Mass per Area and Equivalent Water Thickness over Sparse Forests Using a Hybrid Method. *Remote Sens. 2021*, 13, 3235. https://doi.org/10.3390/rs13163235

Communications

Miraglio, T.; Adeline, K., Gastellu-Etchegorry, J.-P. Huesca, M.; Ustin, S.; Briottet, X. Influence of forest modeling on LAI and chlorophyll content estimation with radiative transfer inversion for a woodland savanna. *SFPT-GH* **2019**, Toulouse, France

Miraglio, T.; Adeline, K., Gastellu-Etchegorry, J.-P. Huesca, M.; Ustin, S.; Briottet, X. Fast LUT building by combining PROSAIL and DART: application to Gap Fraction and Leaf Pigments estimations over a woodland savanna. *American Geophysical Union Fall Meeting* **2020**, San Francisco, USA

Miraglio, T.; Adeline, K.; Huesca, M.; Ustin, S.; Briottet, X. Assessing The Vegetation Traits Retrieval Capabilities Of The Biodiversity And SBG Hyperspectral Satellite Missions Over Open Canopies. *40^{th} EARSeL symposium* **2021**, Warsaw, Poland

Bibliography

Ali, Abebe Mohammed et al. (2016a). “Effects of Canopy Structural Variables on Retrieval of Leaf Dry Matter Content and Specific Leaf Area from Remotely Sensed Data.” In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cit. on pp. 66, 84).

Ali, Abebe Mohammed et al. (2016b). “Retrieval of forest leaf functional traits from HySpex imagery using radiative transfer models and continuous wavelet analysis.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* (cit. on pp. 37, 66).

Ali, Abebe Mohammed et al. (2020). “Machine learning methods’ performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest.” In: *International Journal of Digital Earth*, pp. 1–15 (cit. on p. 23).

Arendonk, J. J. C. M. and H. Poorter (1994). “The chemical composition and anatomical structure of leaves of grass species differing in relative growth rate.” In: *Plant, Cell and Environment* 17.8, pp. 963–970 (cit. on p. 14).

Asner, Gregory P. et al. (2011a). “Spectroscopy of canopy chemicals in humid tropical forests.” In: *Remote Sensing of Environment* 115.12, pp. 3587–3598 (cit. on pp. 5, 20).

Asner, Gregory P et al. (2011b). “Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests.” In: *Ecological Applications* 21.1, pp. 85–98 (cit. on pp. 6, 21).

Asner, Gregory P. et al. (2015). “Quantifying forest canopy traits: Imaging spectroscopy versus field survey.” In: *Remote Sensing of Environment* 158, pp. 15–27 (cit. on pp. 5, 20, 101, 109).

Atzberger, Clement (2000). “Development of an invertible forest reflectance model The INFORM-Model.” In: *A decade of trans-European remote sensing cooperation, Proceedings of the 20th EARSeL Symposium*, pp. 39–44 (cit. on pp. 18, 66).

Atzberger, Clement and Katja Richter (2012). “Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery.” In: *Remote Sensing of Environment* 120, pp. 208–218 (cit. on pp. 21, 60).

Axelsson, Christoffer et al. (2013). “Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression.” In: *International Journal of Remote Sensing* 34.5, pp. 1724–1743 (cit. on pp. 20, 109).

Bach, Heike and Wout Verhoef (2003). “Sensitivity studies on the effect of surface soil moisture on canopy reflectance using the radiative transfer model geoSAIL.” In: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 3. IEEE, pp. 1679–1681 (cit. on p. 123).

Banskota, Asim et al. (2015). “An LUT-Based Inversion of DART Model to Estimate Forest LAI from Hyperspectral Data.” In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cit. on pp. 37, 60, 89).

Barclay, H. J. (2001). “Distribution of leaf orientations in six conifer species.” In: *Canadian Journal of Botany* 79.4, pp. 389–397 (cit. on p. 14).

Baret, Frédéric, J. G.P.W. Clevers, and M D Steven (1995). “The robustness of canopy gap fraction estimates from red and near-infrared reflectances: A comparison of approaches.” In: *Remote Sensing of Environment* 54.2, pp. 141–151 (cit. on p. 124).

Barnes, William L., X. Jack Xiong, and Vincent V Salomonson (2002). “Status of Terra MODIS and Aqua MODIS.” In: *International Geoscience and Remote Sensing Symposium (IGARSS).* Vol. 2, pp. 970–972 (cit. on p. 6).

Beard, J S (1975). “The vegetation survey of Western Australia.” In: *Vegetatio* 30.3, pp. 179–187 (cit. on p. 2).

Béland, Martin et al. (2014). “On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR.” In: *Agricultural and Forest Meteorology* 184, pp. 82–97 (cit. on pp. 88, 125).

Berger, Katja et al. (2018). *Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study* (cit. on pp. 6, 17, 23, 24, 70, 84).

Bo-Cai Gao and A. F.H. Goetz (1990). “Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data.” In: *Journal of Geophysical Research* 95.D4, pp. 3549–3564 (cit. on p. 29).

Bonan, Gordon B. et al. (2011). “Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data.” In: *Journal of Geophysical Research* 116.G2 (cit. on p. 123).

Brede, Benjamin et al. (2020). “Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, landsat 7 ETM+ and Landsat 8 OLI.” In: *Remote Sensing* 12.6 (cit. on p. 100).

Bruniquel-Pinel, V and J P Gastellu-Etchegorry (1998). “Sensitivity of texture of high resolution images of forest to biophysical and acquisition parameters.” In: *Remote Sensing of Environment* 65.1, pp. 61–85 (cit. on p. 18).

Caldwell, MM et al. (1995). *Mediterranean-Type Ecosystems.* Ed. by George W. Davis and David M. Richardson. Vol. 109. Ecological Studies. Berlin, Heidelberg: Springer Berlin Heidelberg (cit. on pp. 1, 3).

Caldwell, MM et al. (2001). *Global Biodiversity in a Changing Environment.* Ed. by F. Stuart Chapin, Osvaldo E. Sala, and Elisabeth Huber-Sannwald. Vol. 152. Ecological Studies. New York, NY: Springer New York (cit. on p. 131).

Carrere, Veronique et al. (2013). “HYPXIM: A second generation high spatial resolution hyperspectral satellite for dual applications.” In: *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing.* Vol. 2013-June. IEEE Computer Society (cit. on pp. 6, 102).

Carter, Gregory A. et al. (2009). “Remote sensing and mapping of tamarisk along the Colorado river, USA: A comparative use of summer-acquired hyperion, thematic mapper and quickbird data.” In: *Remote Sensing* 1.3, pp. 318–329 (cit. on p. 124).

Cavender-Bares, Jeannine et al. (2017). “Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales.” In: *American Journal of Botany* 104.7, pp. 966–969 (cit. on p. 5).

Chalker-Scott, Linda (1999). *Environmental significance of anthocyanins in plant stress responses* (cit. on p. 11).

Chang, Chih Chung and Chih Jen Lin (2011). “LIBSVM: A Library for support vector machines.” In: *ACM Transactions on Intelligent Systems and Technology* 2.3 (cit. on p. 42).

Chen, Qi et al. (2008). “Modeling radiation and photosynthesis of a heterogeneous savanna woodland landscape with a hierarchy of model complexities.” In: *Agricultural and Forest Meteorology* 148.6-7, pp. 1005–1020 (cit. on p. 43).

Chianucci, Francesco et al. (2018). “A dataset of leaf inclination angles for temperate and boreal broadleaf woody species.” In: *Annals of Forest Science* 75.2, p. 50 (cit. on p. 123).

Choat, Brendan et al. (2012). “Global convergence in the vulnerability of forests to drought.” In: *Nature* 491.7426, pp. 752–755 (cit. on p. 3).

Chow, W S (2003). *Photosynthesis: From Natural Towards Artificial.* Tech. rep., pp. 447–459 (cit. on p. 11).

Clevers, J. G.P.W., L. Kooistra, and M. E. Schaepman (2008). “Using spectral information from the NIR water absorption features for the retrieval of canopy water content.” In: *International Journal of Applied Earth Observation and Geoinformation* 10.3, pp. 388–397 (cit. on p. 101).

Coblentz, W W (1913). “The diffuse reflecting power of various substances.” In: *Journal of the Franklin Institute* 175.1, p. 68 (cit. on p. 10).

Colombo, R. et al. (2008). “Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling.” In: *Remote Sensing of Environment* (cit. on p. 17).

Combal, Bruno et al. (2003). “Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem.” In: *Remote Sensing of Environment* (cit. on pp. 21, 45).

Cowling, Richard M et al. (1996). “Plant diversity in mediterranean-climate regions.” In: *Trends in Ecology & Evolution* 11.9, pp. 362–366 (cit. on p. 1).

Crawford, R. M. M. and A. C. Gibson (1997). “Structure-Function Relations of Warm Desert Plants.” In: *The Journal of Ecology* 85.5, p. 735 (cit. on p. 123).

Croft, H. and J. M. Chen (2017). “Leaf pigment content.” In: *Comprehensive Remote Sensing.* Vol. 1-9. Elsevier, pp. 117–142 (cit. on p. 11).

Dana Chadwick, K. and Gregory P. Asner (2016). “Organismic-scale remote sensing of canopy foliar traits in lowland tropical forests.” In: *Remote Sensing* 8.2 (cit. on pp. 6, 20, 21, 101).

Danson, F. Mark et al. (2007). “Forest canopy gap fraction from terrestrial laser scanning.” In: *IEEE Geoscience and Remote Sensing Letters* 4.1, pp. 157–160 (cit. on p. 14).

Darvishzadeh, Roshanak, Ali A. Matkan, and Abdolhamid Dashti Ahangar (2012). “Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach.” In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cit. on pp. 17, 21, 38).

Darvishzadeh, Roshanak et al. (2008). “Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture.” In: *International Journal of Applied Earth Observation and Geoinformation* 10.3, pp. 358–373 (cit. on pp. 123, 129).

Darvishzadeh, Roshanak et al. (2011). “Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* (cit. on pp. 21, 61, 79).

Darvishzadeh, Roshanak et al. (2019). “Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model.” In: *International Journal of Applied Earth Observation and Geoinformation* 79, pp. 58–70 (cit. on pp. 6, 18, 22, 80).

Dawson, Terence P, Paul J Curran, and Stephen E Plummer (1998). “LIBERTY - Modeling the effects of Leaf Biochemical Concentration on Reflectance Spectra.” In: *Remote Sensing of Environment* 65.1, pp. 50–60 (cit. on p. 16).

Delegido, Jesús et al. (2010). “Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC).” In: *International Journal of Applied Earth Observation and Geoinformation* 12.3, pp. 165–174 (cit. on pp. 20, 61).

Delegido, Jesús et al. (2014). “Chlorophyll content mapping of urban vegetation in the city of Valencia based on the hyperspectral NAOC index.” In: *Ecological Indicators* 40, pp. 34–42 (cit. on p. 17).

Delegido, Jesús et al. (2015). “Brown and green LAI mapping through spectral indices.” In: *International Journal of Applied Earth Observation and Geoinformation* 35.PB, pp. 350–358 (cit. on p. 19).

Demarez, V. et al. (1999). “Seasonal variation of leaf chlorophyll content of a temperate forest. Inversion of the prospect model.” In: *International Journal of Remote Sensing* 20.5, pp. 879–894 (cit. on p. 61).

Demmig-Adams, Barbara and William W Adams (1996). *The role of xanthophyll cycle carotenoids in the protection of photosynthesis* (cit. on p. 11).

Edmonds, Robert L., James K. Agee, and Robert I. Gara (2000). *Forest health and protection.* 1. New York: McGraw-Hill, pp. 285–299 (cit. on p. 4).

Eriksson, Helena Margaretha et al. (2006). “Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates.” In: *Remote Sensing of Environment* 103.4, pp. 408–418 (cit. on p. 123).

Fang, Hongliang et al. (2019). “An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications.” In: *Reviews of Geophysics* 57.3, pp. 739–799 (cit. on pp. 30, 79).

Fath, Brian (2018). *Encyclopedia of ecology.* Vol. 105. 3, pp. 1–663 (cit. on pp. 1, 3).

Feilhauer, Hannes, Gregory P. Asner, and Roberta E. Martin (2015). “Multi-method ensemble selection of spectral bands related to leaf biochemistry.” In: *Remote Sensing of Environment* 164, pp. 57–65 (cit. on p. 109).

Fensholt, Rasmus and Inge Sandholt (2003). “Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment.” In: *Remote Sensing of Environment* 87.1, pp. 111–121 (cit. on p. 74).

Féret, J. B. et al. (2017). “PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle.” In: *Remote Sensing of Environment* (cit. on pp. 16, 80, 86, 107).

Féret, J. B. et al. (2019). “Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning.” In: *Remote Sensing of Environment* 231 (cit. on p. 101).

Feret, Jean Baptiste et al. (2008). “PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments.” In: *Remote Sensing of Environment* (cit. on p. 16).

Féret, Jean Baptiste et al. (2021). “PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents.” In: *Remote Sensing of Environment* 252 (cit. on pp. 13, 17).

Ferreira, Matheus Pinheiro et al. (2018). “Retrieving structural and chemical properties of individual tree crowns in a highly diverse tropical forest with 3D radiative transfer modeling and imaging spectroscopy.” In: *Remote Sensing of Environment* (cit. on pp. 22, 60, 99, 101).

Flint, Stephan D, Peter W Jordan, and Martyn M Caldwell (1985). “Plant Protective Response to Enhanced UV-B Radiation Under Field Conditions: Leaf Optical Properties and Photosynthesis.” In: *Photochemistry and Photobiology* 41.1, pp. 95–99 (cit. on p. 11).

Frampton, William James et al. (2013). “Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 82, pp. 83–92 (cit. on p. 6).

Friedl, Mark A. et al. (2010). “MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets.” In: *Remote Sensing of Environment* 114.1, pp. 168–182 (cit. on p. 131).

Füssel, Hans-Martin et al. (2012). *Climate change, impacts and vulnerability in Europe 2012: an indicator-based report.* 12, p. 304 (cit. on p. 2).

Gamon, J. A. et al. (2019). *Assessing Vegetation Function with Imaging Spectroscopy* (cit. on p. 5).

Gao, Bo-Cai (1996). “NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space.” In: *Remote Sensing of Environment* 7212, pp. 257–266 (cit. on pp. 19, 74).

Gao, Bo Cai and Rong Rong Li (2017). “Removal of thin cirrus scattering effects in landsat 8 OLI images using the cirrus detecting channel.” In: *Remote Sensing* 9.8 (cit. on p. 62).

Gao, Bo Cai et al. (1998). “Correction of thin cirrus path radiances in the 0.4-1.0 μm spectral region using the sensitive 1.375 μm cirrus detecting channel.” In: *Journal of Geophysical Research Atmospheres* 103.D24, pp. 32169–32176 (cit. on p. 62).

Gara, Tawanda W. et al. (2019). “Evaluating the performance of PROSPECT in the retrieval of leaf traits across canopy throughout the growing season.” In: *International Journal of Applied Earth Observation and Geoinformation* 83, p. 101919 (cit. on p. 61).

Gascon, F. et al. (2004). “Retrieval of forest biophysical variables by inverting a 3-D radiative transfer model and using high and very high resolution imagery.” In: *International Journal of Remote Sensing* (cit. on pp. 21, 38, 43, 89, 90, 107).

Gastellu-Etchegorry, J. P., F. Gascon, and P. Estève (2003). “An interpolation procedure for generalizing a look-up table inversion method.” In: *Remote Sensing of Environment* (cit. on pp. 43, 89, 107).

Gastellu-Etchegorry, J. P. et al. (1996). “Modeling radiative transfer in heterogeneous 3-D vegetation canopies.” In: *Remote Sensing of Environment* 58.2, pp. 131–156 (cit. on p. 34).

Gastellu-Etchegorry, Jean Philippe et al. (2015). “Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes.” In: *Remote Sensing* (cit. on pp. 18, 34, 84).

Giorgi, Filippo and Piero Lionello (2008). “Climate change projections for the Mediterranean region.” In: *Global and Planetary Change* 63.2-3, pp. 90–104 (cit. on p. 2).

Gitelson, Anatoly and Mark N. Merzlyak (1994). “Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation.” In: *Journal of Plant Physiology* (cit. on pp. 47, 74).

Gitelson, Anatoly A et al. (2002). “Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy.” In: *Photochemistry and Photobiology* 75.3, p. 272 (cit. on p. 19).

Gitelson, Anatoly A. et al. (2006). “Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity.” In: *Journal of Geophysical Research Atmospheres* 111.8 (cit. on pp. 47, 74).

Goulden, M. L. and R. C. Bales (2019). “California forest die-off linked to multi-year deep soil drying in 2012–2015 drought.” In: *Nature Geoscience* (cit. on p. 3).

Govaerts, Yves M and Michel M Verstraete (1998). “Raytran: a monte carlo raytracing model to compute light scattering in threedimensional heterogeneous media.” In: *IEEE Transactions on Geoscience and Remote Sensing* 36.2, p. 493505 (cit. on p. 18).

Graziani, Frank et al. (2006). *Computational Methods in Transport* (cit. on p. 18).

Green, Robert O. et al. (1998). “Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS).” In: *Remote Sensing of Environment* 65.3, pp. 227–248 (cit. on p. 29).

Guanter, Luis et al. (2015). *The EnMAP spaceborne imaging spectroscopy mission for earth observation* (cit. on p. 6).

Haboudane, Driss et al. (2002). “Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture.” In: *Remote Sensing of Environment.* arXiv: `NIHMS150003` (cit. on pp. 17, 19, 20, 47, 74).

Haboudane, Driss et al. (2004). “Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture.” In: *Remote Sensing of Environment* (cit. on pp. 19, 42, 45, 47, 74).

Hadjimitsis, D. G., C. R.I. Clayton, and A. Retalis (2009). “The use of selected pseudo-invariant targets for the application of atmospheric correction in multi-temporal studies using satellite remotely sensed imagery.” In: *International Journal of Applied Earth Observation and Geoinformation* 11.3, pp. 192–200 (cit. on p. 42).

Hanan, Niall and Michael J Hill (2012). “Savannas in a Changing Earth System: The NASA Terrestrial Ecology Tree-Grass Project.” In: p. 56 (cit. on p. 66).

Hardisky, M A, V. Klemas, and R. M. Smart (1983). “The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina alterniflora canopies.” In: *Photogrammetric Engineering & Remote Sensing* 49.1, pp. 77–83 (cit. on p. 74).

Heldens, Wieke et al. (2011). *Can the Future EnMAP mission contribute to urban applications? A literature survey* (cit. on p. 124).

Hernández-Clemente, Rocío, Rafael M. Navarro-Cerrillo, and Pablo J. Zarco-Tejada (2012). “Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT+DART simulations.” In: *Remote Sensing of Environment* (cit. on pp. 18, 20, 38, 45, 47, 66, 74, 79).

Hill, Joachim, Henning Buddenbaum, and Philip A. Townsend (2019). “Imaging Spectroscopy of Forest Ecosystems: Perspectives for the Use of Space-borne Hyperspectral Earth Observation Systems.” In: *Surveys in Geophysics* 40.3, pp. 553–588 (cit. on pp. 6, 45).

Homolová, L., R. Janoutová, and Z. Malenovský (2016). “Evaluation of various spectral inputs for estimation of forest biochemical and structural properties from airborne imaging spectroscopy data.” In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* (cit. on p. 23).

Hopper, Stephen D (1979). “Biogeographical Aspects of Speciation in the Southwest Australian Flora.” In: *Annual Review of Ecology and Systematics* 10.1, pp. 399–422 (cit. on p. 2).

Huang, Wenjiang et al. (2018). “Monitoring Crop Carotenoids Concentration by Remote Sensing.” In: *Progress in Carotenoid Research* (cit. on p. 80).

Huete, A et al. (2002). “Overview of the radiometric and biophysical performance of the MODIS vegetation indices.” In: *Remote Sensing of Environment* 83.1-2, pp. 195–213 (cit. on p. 74).

Huete, A.R (1988). “A soil-adjusted vegetation index (SAVI).” In: *Remote Sensing of Environment* 25.3, pp. 295–309 (cit. on p. 19).

Hunt, E. Raymond and Barrett N Rock (1989). “Detection of changes in leaf water content using Near- and Middle-Infrared reflectances.” In: *Remote Sensing of Environment* 30.1, pp. 43–54 (cit. on pp. 19, 74, 130).

Jack, Steven B. and James N. Long (1992). “Forest production and the organization of foliage within crowns and canopies.” In: *Forest Ecology and Management* (cit. on p. 15).

Jacquemoud, S and F Baret (1990). “PROSPECT: A model of leaf optical properties spectra.” In: *Remote Sensing of Environment* 34.2, pp. 75–91 (cit. on pp. 16, 86).

Jacquemoud, Stéphane and Susan Ustin (2019). *Leaf optical properties.* Cambridge University Press, pp. 1–556 (cit. on pp. 11, 12).

Jacquemoud, Stéphane et al. (2009). “PROSPECT + SAIL models: A review of use for vegetation characterization.” In: *Remote Sensing of Environment* 113.SUPPL. 1 (cit. on p. 17).

Jafari, R. and M. M. Lewis (2012). “Arid land characterisation with EO-1 hyperion hyperspectral data.” In: *International Journal of Applied Earth Observation and Geoinformation* 19.1, pp. 298–307 (cit. on p. 124).

Janoutová, Růžena et al. (2019). “Influence of 3D spruce tree representation on accuracy of airborne and satellite forest reflectance simulated in DART.” In: *Forests* 10.3 (cit. on pp. 18, 38, 67, 84).

Jetz, Walter et al. (2019). “Essential biodiversity variables for mapping and monitoring species populations.” In: *Nature Ecology and Evolution* 3.4, pp. 539–551 (cit. on p. 83).

Jiménez-Ruano, Adrián, Marcos Rodrigues Mimbrero, and Juan de la Riva Fernández (2009). *Earth Observation of Wildland Fires in Mediterranean Ecosystems.* Ed. by Emilio Chuvieco. Vol. 89. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 100–111 (cit. on pp. 1, 3).

Junker, Laura Verena and Ingo Ensminger (2016). “Relationship between leaf optical properties, chlorophyll fluorescence and pigment changes in senescing Acer saccharum leaves.” In: *Tree Physiology* 36.6, pp. 694–711 (cit. on p. 13).

Karlik, John F and a H McKay (2002). “Leaf Area Index, Leaf Mass Density, and Allometric Relationships Derived From Harvest of Blue Oaks in a California Oak Savanna.” In: *USDA Forest Service Gen. Tech. Rep.* PSW-GTR-18, pp. 719–729 (cit. on pp. 88, 90).

Karpouzli, E. and T. Malthus (2003). “The empirical line method for the atmospheric correction of IKONOS imagery.” In: *International Journal of Remote Sensing* 24.5, pp. 1143–1150 (cit. on p. 42).

Klausmeyer, Kirk R. and M. Rebecca Shaw (2009). “Climate change, habitat loss, protected areas and the climate adaptation potential of species in mediterranean ecosystems worldwide.” In: *PLoS ONE* 4.7 (cit. on p. 3).

Klingberg, Jenny et al. (2015). “Measured and modelled leaf area of urban woodlands , parks and trees in Gothenburg , Sweden.” In: *ICUC9 - 9th International Conference on Urban Climate jointly with 12th Symposium on the Urban Environment*, pp. 1–6 (cit. on p. 125).

Kobayashi, Hideki et al. (2012). “Modeling energy and carbon fluxes in a heterogeneous oak woodland: A three-dimensional approach.” In: *Agricultural and Forest Meteorology* 152.1, pp. 83–100 (cit. on pp. 18, 34, 42, 88).

Kobayashi, Hideki et al. (2013). “On the correct estimation of gap fraction: How to remove scattered radiation in gap fraction measurements?” In: *Agricultural and Forest Meteorology* (cit. on p. 27).

Krause, G H and E Weis (1991). “Chlorophyll Fluorescence and Photosynthesis: The Basics.” In: *Annual Review of Plant Physiology and Plant Molecular Biology* 42.1, pp. 313–349 (cit. on p. 10).

Kuusk, Andres (2017). “Canopy radiative transfer modeling.” In: *Comprehensive Remote Sensing*. Vol. 1-9. Elsevier, pp. 9–22 (cit. on p. 17).

Kvalheim, Olav Martin et al. (2018). “Determination of optimum number of components in partial least squares regression from distributions of the root-mean-squared error obtained by Monte Carlo resampling.” In: *Journal of Chemometrics* 32.4 (cit. on p. 110).

Lambers, Hans, F. Stuart Chapin, and Thijs L. Pons (2008). *Plant physiological ecology: Second edition*. Springer New York, pp. 1–604 (cit. on pp. 10, 12).

Le Maire, G., C. François, and E. Dufrêne (2004). “Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements.” In: *Remote Sensing of Environment* 89.1, pp. 1–28 (cit. on p. 20).

Lee, Christine M. et al. (2015). “An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities.” In: *Remote Sensing of Environment* 167, pp. 6–19 (cit. on pp. 6, 102).

Leonenko, Ganna, Sietse O. Los, and Peter R.J. North (2013). “Statistical distances and their applications to biophysical parameter estimation: Information measures, m-estimates, and minimum contrast methods.” In: *Remote Sensing* 5.3, pp. 1355–1388 (cit. on p. 79).

Lewis, Philip (1999). “Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System.” In: *Agronomie* 19.3-4, pp. 185–210 (cit. on p. 18).

Li, L. et al. (2008). “Retrieval of vegetation equivalent water thickness from reflectance using genetic algorithm (GA)-partial least squares (PLS) regression.” In: *Advances in Space Research* (cit. on pp. 17, 102).

Lichtenthaler, Hartmut K. (1987). “Chlorophylls and Carotenoids: Pigments of Photosynthetic Biomembranes.” In: *Methods in Enzymology* 148.C, pp. 350–382 (cit. on p. 31).

Lichtenthaler, Hartmut K. and Claus Buschmann (2001). “Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy.” In: *Current Protocols in Food Analytical Chemistry* 1.1, F4.3.1–F4.3.8 (cit. on p. 31).

Lionello, Piero et al. (2014). *The climate of the Mediterranean region: research progress and climate change impacts* (cit. on p. 2).

Liu, Jing et al. (2019). “Variation of leaf angle distribution quantified by terrestrial LiDAR in natural European beech forest.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 148, pp. 208–220 (cit. on pp. 15, 123).

Lu, Bing, Cameron Proctor, and Yuhong He (2021). “Investigating different versions of PROSPECT and PROSAIL for estimating spectral and biophysical properties of photosynthetic and non-photosynthetic vegetation in mixed grasslands.” In: *GIScience and Remote Sensing* (cit. on p. 100).

Maccioni, Andrea, Giovanni Agati, and Piero Mazzinghi (2001). “New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra.” In: *Journal of Photochemistry and Photobiology B: Biology* 61.1-2, pp. 52–61 (cit. on pp. 47, 74).

Maire, Guerric le et al. (2008). “Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass.” In: *Remote Sensing of Environment* (cit. on pp. 17, 19, 37, 74, 80, 101).

Majasalmi, Titta and Miina Rautiainen (2021). “Representation of tree cover in global land cover products: Finland as a case study area.” In: *Environmental Monitoring and Assessment* 193.3 (cit. on p. 131).

Makino, Amane and Barry Osmond (1991). “Effects of nitrogen nutrition on nitrogen partitioning between chloroplasts and mitochondria in pea and wheat.” In: *Plant Physiology* 96.2, pp. 355–362 (cit. on p. 11).

Malenovský, Z. et al. (2006). “Applicability of the PROSPECT model for Norway spruce needles.” In: *International Journal of Remote Sensing* 27.24, pp. 5315–5340 (cit. on pp. 20, 131).

Malenovsky, Z. et al. (2007). “Retrieval of coniferous canopy chlorophyll content from high spatial resolution hyperspectral data.” In: *Proceedings of the 10th International Symposium on Physical Measurements and Spectral Signatures in Remote Sensing (ISPMSRS’07).* Vol. XXXVI, pp. 108–113 (cit. on pp. 19, 79).

Malenovský, Z. et al. (2008a). “Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution.” In: *Remote Sensing of Environment* 112.1, pp. 1–18 (cit. on pp. 16, 38, 59, 121).

Malenovský, Zbyněk et al. (2008b). “Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution.” In: *Remote Sensing of Environment* (cit. on pp. 18, 99, 101).

Malenovský, Zbyněk et al. (2013). “Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer.” In: *Remote Sensing of Environment* 131, pp. 85–102 (cit. on pp. 23, 99, 101).

Markham, Brian L. et al. (2004). *Landsat sensor performance: History and current status* (cit. on p. 6).

Marshall, Michael and Prasad Thenkabail (2015). “Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 108, pp. 205–218 (cit. on p. 24).

Martens, Scott N, Susan L Ustin, and Robert A Rousseau (1993). *Estimation of tree canopy leaf area index by gap fraction analysis.* Tech. rep., pp. 91–108 (cit. on p. 14).

Meacham-Hensold, Katherine et al. (2019). “High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity.” In: *Remote Sensing of Environment* 231 (cit. on pp. 109, 128).

Médail, Frédéric (2017). *The specific vulnerability of plant biodiversity and vegetation on Mediterranean islands in the face of global change* (cit. on p. 3).

Medail, Frederic and Pierre Quezel (1997). “Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean Basin.” In: *Annals of the Missouri Botanical Garden* 84.1, p. 112 (cit. on p. 1).

Melendo-Vega, José Ramón et al. (2018). “Improving the performance of 3-D radiative transfer model FLIGHT to simulate optical properties of a tree-grass ecosystem.” In: *Remote Sensing* (cit. on pp. 18, 38, 81, 100).

Merzlyak, Mark N. et al. (2008). “Light absorption by anthocyanins in juvenile, stressed, and senescing leaves.” In: *Journal of Experimental Botany* 59.14, pp. 3903–3911 (cit. on p. 11).

Mevik, Bjørn Helge, Vegard H. Segtnan, and Tormod Næs (2004). “Ensemble methods and partial least squares regression.” In: *Journal of Chemometrics* 18.11, pp. 498–507 (cit. on p. 110).

Migita, C., Y. Chiba, and T. Tange (2007). “Seasonal and spatial variations in leaf nitrogen content and resorption in a Quercus serrata canopy.” In: *Tree Physiology* 27.1, pp. 63–70 (cit. on p. 11).

Misson, L et al. (2007). “Partitioning forest carbon fluxes with overstory and understory eddy-covariance measurements: A synthesis based on FLUXNET data.” In: *Agricultural and Forest Meteorology* 144.2, pp. 14–31 (cit. on pp. 15, 88).

Mooney, Harold A. and E. Lloyd Dunn (1970). “Convergent Evolution of Mediterranean-Climate Evergreen Sclerophyll Shrubs.” In: *Evolution* 24.2, p. 292 (cit. on p. 3).

Moorthy, Inian, John R. Miller, and Thomas L. Noland (2008). “Estimating chlorophyll concentration in conifer needles with hyperspectral data: An assessment at the needle and canopy level.” In: *Remote Sensing of Environment* 112.6, pp. 2824–2838 (cit. on pp. 16, 131).

Moriondo, Marco et al. (2006). “Potential impact of climate change on fire risk in the Mediterranean area.” In: *Climate Research* 31.1, pp. 85–95 (cit. on p. 3).

Mutanga, Onisimo, Elhadi Adam, and Moses Azong Cho (2012). “High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm.” In: *International Journal of Applied Earth Observation and Geoinformation* 18.1, pp. 399–406 (cit. on pp. 20, 21).

Myers, Norman et al. (2000). “Biodiversity hotspots for conservation priorities.” In: *Nature* 403.6772, pp. 853–858 (cit. on p. 1).

Nilson, Tiit and Andres Kuusk (2004). “Improved algorithm for estimating canopy indices from gap fraction data in forest canopies.” In: *Agricultural and Forest Meteorology* 124.3-4, pp. 157–169 (cit. on p. 14).

North, Peter R J (1996). *Three-Dimensional Forest Light Interaction Model Using a Monte Carlo*. Tech. rep. 4 (cit. on p. 18).

Paynter, Ian et al. (2016). “Observing ecosystems with lightweight, rapid-scanning terrestrial lidar scanners.” In: *Remote Sensing in Ecology and Conservation* 2.4, pp. 174–189 (cit. on p. 32).

Pearlman, Jay S. et al. (2003). “Hyperion, a space-based imaging spectrometer.” In: *IEEE Transactions on Geoscience and Remote Sensing* 41.6 PART I, pp. 1160–1173 (cit. on p. 6).

Penuelas, J. et al. (1993). “The reflectance at the 950-970 nm region as an indicator of plant water status.” In: *International Journal of Remote Sensing* 14.10, pp. 1887–1905 (cit. on pp. 74, 83).

Penuelas, J. et al. (1997). “Estimation of plant water concentration by the reflectance Water Index WI (R900/R970).” In: *International Journal of Remote Sensing* 18.13, pp. 2869–2875 (cit. on p. 83).

Pereira, H. M. et al. (2013). “Essential biodiversity variables.” In: *Science* 339.6117, pp. 277–278 (cit. on p. 4).

Pettorelli, Nathalie et al. (2016). “Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions.” In: *Remote Sensing in Ecology and Conservation* 2.3, pp. 122–131 (cit. on pp. 4, 103).

Pierce, Lars L, Steven W Running, and Joe Walker (1994). “Regional-Scale Relationships of Leaf Area Index to Specific Leaf Area and Leaf Nitrogen Content.” In: *Ecological Applications* 4.2, pp. 313–321 (cit. on p. 14).

Pinty, B. et al. (2004). “Radiation Transfer Model Intercomparison (RAMI) exercise: Results from the second phase.” In: *Journal of Geophysical Research D: Atmospheres* 109.6 (cit. on pp. 17, 34).

Pinty, Bernard et al. (2001). “Radiation transfer model intercomparison (RAMI) exercise.” In: *Journal of Geophysical Research Atmospheres* 106.D11, pp. 11937–11956 (cit. on pp. 17, 34).

Pisek, Jan et al. (2013). “Is the spherical leaf inclination angle distribution a valid assumption for temperate and boreal broadleaf tree species?” In: *Agricultural and Forest Meteorology* 169, pp. 186–194 (cit. on pp. 14, 123).

Poutier, Laurent et al. (2002). “COMANCHE and COCHISE: two reciprocal atmospheric codes for hyperspectral remote sensing.” In: *2002 AVIRIS Earth Science and Applications Workshop Proceedings*. 2 (cit. on p. 106).

Proisy, Christophe et al. (2012). *Biomass Prediction in Tropical Forests: The Canopy Grain Approach* (cit. on p. 99).

Qi, J. et al. (1994). “A modified soil adjusted vegetation index.” In: *Remote Sensing of Environment* 48.2, pp. 119–126 (cit. on pp. 45, 47, 74).

Ramsfield, T. D. et al. (2016). “Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts.” In: *Forestry* 89.3, pp. 245–252 (cit. on p. 4).

Rast, Michael et al. (2019). *Copernicus hyperspectral imaging mission for the environment: Mission requirements document.* English. 2.1. Mission Requirements Document (MRD) ESA-EOPSM-CHIM-MRD-3216. European Space Agency (ESA) (cit. on p. 6).

Raumonen, Pasi et al. (2013). "Fast automatic precision tree models from terrestrial laser scanner data." In: *Remote Sensing* 5.2, pp. 491–520 (cit. on p. 86).

Rich, P M et al. (1999). "Guide to HemiView: software for analysis of hemispherical photography." In: *Delta-T Devices, Ltd., Cambridge, England* (cit. on p. 30).

Richter, Katja et al. (2011). "Evaluation of Sentinel-2 Spectral Sampling for Radiative Transfer Model Based LAI Estimation of Wheat, Sugar Beet, and Maize." In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.2, pp. 458–464 (cit. on pp. 21, 79).

Ritz, Thorsten et al. (2000). *Efficient light harvesting through carotenoids* (cit. on p. 11).

Rivera-Caicedo, Juan Pablo et al. (2017). "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 132, pp. 88–101 (cit. on p. 19).

Roberts, D. A. et al. (2006). "Evaluation of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Moderate Resolution Imaging Spectrometer (MODIS) measures of live fuel moisture and fuel condition in a shrubland ecosystem in southern California." In: *Journal of Geophysical Research: Biogeosciences* 111.4 (cit. on p. 84).

Roelofsen, Hans D. et al. (2014). "Mapping a priori defined plant associations using remotely sensed vegetation characteristics." In: *Remote Sensing of Environment* 140, pp. 639–651 (cit. on p. 21).

Ryu, Youngryel et al. (2010). "How to quantify tree leaf area index in an open savanna ecosystem: A multi-instrument and multi-model approach." In: *Agricultural and Forest Meteorology* (cit. on pp. 15, 121).

Ryu, Youngryel et al. (2011). "Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales." In: *Global Biogeochemical Cycles* 25.4 (cit. on p. 123).

Ryu, Youngryel et al. (2012). "Continuous observation of tree leaf area index at ecosystem scale using upward-pointing digital cameras." In: *Remote Sensing of Environment* (cit. on pp. 15, 88).

Sala, Osvaldo E. et al. (2000). "Global biodiversity scenarios for the year 2100." In: *Science* 287.5459, pp. 1770–1774 (cit. on p. 3).

Salgado, Sandra et al. (2019). "Validation of an empirical method for thin cirrus correction with Sentinel-2 data." In: SPIE-Intl Soc Optical Eng, p. 16 (cit. on p. 62).

San, B. Taner and M. Lütfi Süzen (2011). "Evaluation of cross-track illumination in EO-1 hyperion imagery for lithological mapping." In: *International Journal of Remote Sensing* 32.22, pp. 7873–7889 (cit. on p. 12).

Sanger, Jon E. (1971). "Quantitative Investigations of Leaf Pigments From Their Inception in Buds Through Autumn Coloration to Decomposition in Falling Leaves." In: *Ecology* 52.6, pp. 1075–1089 (cit. on p. 11).

Sardans, J. and J. Peñuelas (2013). "Plant-soil interactions in Mediterranean forest and shrublands: Impacts of climatic change." In: *Plant and Soil* 365.1-2, pp. 1–33 (cit. on p. 3).

Scarascia-Mugnozza, Giseppe et al. (2000). “Forests of the Mediterranean region: Gaps in knowledge and research needs.” In: *Forest Ecology and Management* 132.1, pp. 97–109 (cit. on p. 1).

Schaepman, Michael E. et al. (2015). “Advanced radiometry measurements and Earth science applicationswith the Airborne Prism Experiment (APEX).” In: *Remote Sensing of Environment* 158.1, pp. 207–219 (cit. on p. 6).

Schimel, David et al. (2015). *Observing terrestrial ecosystems and the carbon cycle from space* (cit. on pp. 83, 130).

Schott, John R, Carl Salvaggio, and William J Volchok (1988). *Schott_1988*. Tech. rep., pp. 1–16 (cit. on p. 42).

Sensing, Remote, O F Environment, and Toulouse Iii (2016). “Modeling Radiative Transfer in Heterogeneous 3-D Vegetation Canopies Modeling Radiative Transfer in Heterogeneous 3-D Vegetation Canopies.” In: *Science* 4257.February, pp. 131–156 (cit. on p. 84).

Serbin, Shawn P. et al. (2019). “From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance.” In: *New Phytologist* 224.4, pp. 1557–1568 (cit. on p. 83).

Serrano, Lydia, Josep Peñuelas, and Susan L Ustin (2002). “Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data.” In: *Remote Sensing of Environment* 81.2-3, pp. 355–364 (cit. on p. 74).

Shaw, Gary A and Hsiao-hua K Burke (2003). *Spectral Imaging for Remote Sensing*. Tech. rep. (cit. on p. 5).

Siegmann, Bastian and Thomas Jarmer (2015). “Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data.” In: *International Journal of Remote Sensing* 36.18, pp. 4519–4534 (cit. on p. 20).

Siegmann, Bastian et al. (2015). “The potential of pan-sharpened EnMAP data for the assessment of wheat LAI.” In: *Remote Sensing* 7.10, pp. 12737–12762 (cit. on p. 124).

Skidmore, A.K. (2013). “Essential biodiversity variables (EBV) and plant functional traits from earth observation and image spectroscopy : powerpoint.” In: *AGU Fall Meeting 2013, 9-13 December 2013, San Fransisco, United States of America.* 13 slides (cit. on p. 83).

Smith, R C G et al. (1995). “Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite.” In: *Australian Journal of Agricultural Research* 46.1, pp. 113–125 (cit. on pp. 47, 74).

Somot, S. et al. (2008). “21st century climate change scenario for the Mediterranean using a coupled atmosphere-ocean regional climate model.” In: *Global and Planetary Change* 63.2-3, pp. 112–126 (cit. on p. 2).

Stckli, R. et al. (2011). “A global reanalysis of vegetation phenology.” In: *Journal of Geophysical Research: Biogeosciences* 116.3 (cit. on p. 61).

Stefano, Pignatti et al. (2013). “The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring.” In: *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4558–4561 (cit. on pp. 6, 102).

Syphard, Alexandra D. et al. (2009). “Conservation threats due to human-caused increases in fire frequency in mediterranean-climate ecosystems.” In: *Conservation Biology* 23.3, pp. 758–769 (cit. on p. 3).

Tate, Kenneth W et al. (2004). “Effect of Canopy and Grazing on Soil Bulk Density.” In: *Journal of Range Management* 57.4, p. 411 (cit. on p. 29).

Tessler, Naama et al. (2016). “Recovery of Mediterranean Vegetation after Recurrent Forest Fires: Insight from the 2010 Forest Fire on Mount Carmel, Israel.” In: *Land Degradation & Development* 27.5, pp. 1424–1431 (cit. on p. 3).

Thayer, Susan S and Olle Björkman (1990). “Leaf Xanthophyll content and composition in sun and shade determined by HPLC.” In: *Photosynthesis Research* 23.3, pp. 331–343 (cit. on p. 11).

Thenkabail, Prasad S., John G. Lyon, and Alfredo Huete, eds. (2018). *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation.* CRC Press (cit. on p. 12).

Thenkabail, Prasad S, Ronald B Smith, and Eddy De Pauw (2002). *Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization* (cit. on p. 24).

Transon, Julie et al. (2018). “Survey of hyperspectral Earth Observation applications from space in the Sentinel-2 context.” In: *Remote Sensing* 10.2 (cit. on p. 104).

Tribulato, Alessandro et al. (2019). “Effects of water stress on gas exchange, water relations and leaf structure in two ornamental shrubs in the Mediterranean area.” In: *Agronomy* 9.7 (cit. on p. 14).

Trombetti, M. et al. (2008). “Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA.” In: *Remote Sensing of Environment* 112.1, pp. 203–215 (cit. on pp. 23, 74).

Trumbore, S, P Brando, and H Hartmann (2015). “Forest health and global change.” In: *Science* 349.6250, pp. 814–818 (cit. on pp. 4, 37).

Tucker, Compton J. (1979). “Red and photographic infrared linear combinations for monitoring vegetation.” In: *Remote Sensing of Environment* 8.2, pp. 127–150 (cit. on pp. 19, 47, 74).

Underwood, Emma C. et al. (2009). “Threats and biodiversity in the mediterranean biome.” In: *Diversity and Distributions* (cit. on p. 3).

Ustin, Susan L. and John A. Gamon (2010). *Remote sensing of plant functional types* (cit. on p. 5).

Vane, Gregg et al. (1993). “The airborne visible/infrared imaging spectrometer (AVIRIS).” In: *Remote Sensing of Environment* 44.2-3, pp. 127–143 (cit. on p. 6).

Vautard, Robert et al. (2007). “Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit.” In: *Geophysical Research Letters* 34.7, pp. 1–5 (cit. on p. 2).

Verhoef, W (1984). “Light Scattering by Leaf layers with Application to Canopy Reflectance Modeling: the SAIL Model.” In: *Remote Sensing Of Environment* 141, pp. 125–141 (cit. on p. 17).

Verrelst, J. et al. (2012). “Optimizing LUT-based radiative transfer model inversion for retrieval of biophysical parameters using hyperspectral data.” In: *International Geoscience and Remote Sensing Symposium (IGARSS)* (cit. on pp. 22, 79).

Verrelst, Jochem, Erika Romijn, and Lammert Kooistra (2012). “Mapping Vegetation Density in a Heterogeneous River Floodplain Ecosystem Using Pointable CHRIS/PROBA Data.” In: *Remote Sensing* 4.9, pp. 2866–2889 (cit. on p. 19).

Verrelst, Jochem et al. (2010). “Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval.” In: *Remote Sensing of Environment* (cit. on pp. 16, 18, 45, 59).

Verrelst, Jochem et al. (2014). “Optimizing LUT-based RTM inversion for semiautomatic mapping of crop biophysical parameters from sentinel-2 and -3 data: Role of cost functions.” In: *IEEE Transactions on Geoscience and Remote Sensing* 52.1, pp. 257–269 (cit. on p. 19).

Verrelst, Jochem et al. (2015a). “Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods - A comparison.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 108, pp. 260–272 (cit. on p. 19).

Verrelst, Jochem et al. (2015b). “Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review.” In: *ISPRS Journal of Photogrammetry and Remote Sensing* 108, pp. 273–290 (cit. on p. 19).

Wang, Fu min, Jing feng Huang, and Zhang hua Lou (2011). “A comparison of three methods for estimating leaf area index of paddy rice from optimal hyperspectral bands.” In: *Precision Agriculture* 12.3, pp. 439–447 (cit. on p. 21).

Wang, W. M., Z. L. Li, and H. B. Su (2007). “Comparison of leaf angle distribution functions: Effects on extinction coefficient and fraction of sunlit foliage.” In: *Agricultural and Forest Meteorology* 143.1-2, pp. 106–122 (cit. on p. 14).

Wang, Zi Xiu, Q. Peter He, and Jin Wang (2015). *Comparison of variable selection methods for PLS-based soft sensor modeling* (cit. on p. 110).

Weiss, M. et al. (2004). “Review of methods for in situ leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling.” In: *Agricultural and Forest Meteorology* 121.1-2, pp. 37–53 (cit. on p. 30).

Weiss, Marie et al. (2000). “Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data.” In: *Agronomie* (cit. on pp. 17, 21, 60, 66).

Welles, Jon M and Shabtai Cohen (1996). *Canopy structure measurement by gap fraction analysis using commercial instrumentation.* Tech. rep. 302, pp. 1335–1342 (cit. on p. 14).

Westoby, Mark et al. (2004). “The worldwide leaf economics spectrum.” In: *Nature* 428, pp. 821–827 (cit. on pp. 14, 83, 130).

White, Joanne C. et al. (2016). *Remote Sensing Technologies for Enhancing Forest Inventories: A Review* (cit. on p. 5).

Widlowski, J. L. et al. (2013). “The fourth radiation transfer model intercomparison (RAMI-IV): Proficiency testing of canopy reflectance models with ISO-13528.” In: *Journal of Geophysical Research Atmospheres* 118.13, pp. 6869–6890 (cit. on pp. 17, 34, 81).

Widlowski, Jean Luc, Jean François Côté, and Martin Béland (2014). “Abstract tree crowns in 3D radiative transfer models: Impact on simulated open-canopy reflectances.” In: *Remote Sensing of Environment* (cit. on pp. 18, 38, 59, 84, 94, 99, 121).

Widlowski, Jean Luc et al. (2007). “Third Radiation Transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance models.” In: *Journal of Geophysical Research Atmospheres* 112.9 (cit. on pp. 17, 18, 34, 99).

Widlowski, Jean Luc et al. (2015). “The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing.” In: *Remote Sensing of Environment* 169, pp. 418–437 (cit. on p. 17).

Wieneke, S. et al. (2016). “Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence: Implications for improved estimates of gross primary productivity.” In: *Remote Sensing of Environment* 184, pp. 654–667 (cit. on p. 6).

Willmott, Cort J. (1981). “On the validation of models.” In: *Physical Geography* 2.2, pp. 184–194 (cit. on pp. 73, 116).

Willmott, Cort J., Scott M. Robeson, and Kenji Matsuura (2012). “A refined index of model performance.” In: *International Journal of Climatology* 32.13, pp. 2088–2094 (cit. on p. 73).

Wocher, Matthias et al. (2018). “Physically-based retrieval of canopy equivalent water thickness using hyperspectral data.” In: *Remote Sensing* 10.12 (cit. on pp. 84, 101).

Woodgate, William et al. (2016). “Quantifying the impact of woody material on leaf area index estimation from hemispherical photography using 3D canopy simulations.” In: *Agricultural and Forest Meteorology* (cit. on p. 15).

Wulder, Michael A. et al. (2012). *Lidar sampling for large-area forest characterization: A review* (cit. on p. 5).

Yanfang Xiao et al. (2013). “Sensitivity Analysis of Vegetation Reflectance to Biochemical and Biophysical Variables at Leaf, Canopy, and Regional Scales.” In: *IEEE Transactions on Geoscience and Remote Sensing* (cit. on pp. 80, 89, 93, 100, 109).

Yebra, Marta, Emilio Chuvieco, and David Riaño (2008). “Estimation of live fuel moisture content from MODIS images for fire risk assessment.” In: *Agricultural and Forest Meteorology* 148.4, pp. 523–536 (cit. on p. 130).

Yebra, Marta et al. (2013). *A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products* (cit. on pp. 84, 131).

Zarco-Tejada, P. J., C. A. Rueda, and S. L. Ustin (2003). “Water content estimation in vegetation with MODIS reflectance data and model inversion methods.” In: *Remote Sensing of Environment* (cit. on pp. 74, 101).

Zarco-Tejada, P. J. et al. (2004). “Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops.” In: *Remote Sensing of Environment* (cit. on pp. 38, 61, 80).

Zarco-Tejada, P. J. et al. (2013). “Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV).” In: *Agricultural and Forest Meteorology* (cit. on pp. 19, 45, 61, 80).

Zarco-Tejada, P. J. et al. (2019). “Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline.” In: *Remote Sensing of Environment* 223, pp. 320–335 (cit. on pp. 6, 19, 22, 24, 38, 124, 131).

Zhang, Yongqin et al. (2008). “Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery.” In: *Remote Sensing of Environment* 112.7, pp. 3234–3247 (cit. on p. 61).

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut Supérieur de l'Aéronautique et de l'Espace

Présentée et soutenue par

Thomas MIRAGLIO

Le 7 octobre 2021

Estimation de traits de végétation de canopées ouvertes méditerranéennes par télédétection hyperspectrale

Ecole doctorale : **GEETS - Génie Electrique Electronique,Télécommunications et Santé : du système au nanosystème**

Spécialité : **Photonique et Systèmes Optoélectronique**

Unité de recherche :
ISAE-ONERA OLIMPES Optronique, Laser, Imagerie Physique et Environnement Spatial

Thèse dirigée par
Xavier BRIOTTET et Susan USTIN

Jury
M. Stéphane JACQUEMOUD, Rapporteur
M. Kamel SOUDANI, Rapporteur
Mme Agnès BÉGUÉ, Examinatrice
M. Clement ATZBERGER, Examinateur
M. Jochem VERRELST, Examinateur
Mme Karine ADELINE, Examinatrice
M. Xavier BRIOTTET, Directeur de thèse
Mme Susan USTIN, Co-directrice de thèse

Table des matières

Table des figures

Table des acronymes

1D	*One Dimensional*
3D	*Three Dimensional*
ALA	*Average Leaf Angle*
ANN	*Artificial Neural Network*
ANOVA	*Analysis Of VAriance*
AVIRIS-C	*AVIRIS Classic*
AVIRIS-NG	*AVIRIS Next-Generation*
BRF	*Bidirectional Reflectance Factor*
$\mathbf{C}_{ab}$	*Chlorophyll Content*
Car	*Carotenoid Content*
$\mathbf{C}_{ant}$	*Leaf Anthocyanins Content*
CBL	*Compact Biomass Lidar*
CC	*Canopy Cover*
DHP	*Digital Hemispherical Photography*
DOY	*Day Of Year*
EBV	*Essential Biodiversity Variables*
EWT	*Equivalent Water Thickness*
FMC	*Fuel Moisture Content*
FWHM	*Full Width at Half Maximum*
GEO BON	*Group on Earth Observations Biodiversity Observation Network*
GPR	*Gaussian Process Regression*
GSD	*Ground Sampling Distance*
LAD	*Leaf Angle Distribution*
LAI	*Leaf Area Index*
LMA	*Leaf Mass per Area*
LUT	*Look-Up Table*
NEdL	*Noise-Equivalent Delta Radiance*
NEdR	*Noise-Equivalent Delta Reflectance*
NIR	*Near InfraRed*
NPV	*Non Photosynthetic Vegetation*
PAI	*Plant Area Index*
PDT	*Pacific Daylight Time*

PST	*Pacific Standard Time*
PLSR	*Partial Least-Square Regression*
RAMI	*RAdiation transfer Model Intercomparison*
RFR	*Random Forest Regression*
RMSE	*Root-Mean-Square Error*
RTM	*Radiative Transfer Model*
SAC	*Spectral Absorption Coefficient*
SAM	*Spectral Angle Mapper*
SFR	*Simple Forest Representation*
SMLR	*Stepwise MultiLinear Regression*
SNR	*Signal-to-Noise Ratio*
STDB	*standard deviation of the difference between estimated points and the regression line*
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*
SWIR	*Short-Wave InfraRed*
TLS	*Terrestrial Laser Scanner*
VI	*Vegetation Index*
VIP	*Variable Importance in Projection*

Introduction

Les écorégions méditerranéennes sont présentes en Californie, au Chili, en Afrique du sud, en Australie, et bien entendu sur le pourtour Méditerranéen (Figure 1). Elles sont caractérisées par des étés chauds et secs, des hivers doux et humides, des évènements de précipitations extrêmes et des feux de forêt récurrents (FATH, 2018). Malgré ces conditions climatiques rudes et une pression anthropique historiquement forte, 10% de la biodiversité végétale globale est endémique au biome méditerranéen, et ces régions sont considérées comme étant des bassins de biodiversité (MEDAIL et QUEZEL, 1997 ; MYERS et al., 2000). La végétation méditerranéenne est principalement présente au sein de forêts ouvertes telles que des savanes, et les forêts denses sont rares du fait des conditions climatiques.

FIGURE 1 – Emplacement des écorégions méditerranéennes (en marron). COWLING et al., 1996.

Les projections actuelles pour le changement climatique global prévoient une augmentation significative des températures ainsi qu'un assèchement des sols d'ici 2100 pour les écorégions méditerranéennes (FÜSSEL et al., 2012 ; GIORGI et LIONELLO, 2008 ; LIONELLO et al., 2014 ; SOMOT et al., 2008 ; VAUTARD et al., 2007). Cette pression climatique, couplée à une augmentation plus rapide que la moyenne de la population dans ces régions (UNDERWOOD et al., 2009), est telle que d'ici 2100 les environnements méditerranéens devraient être ceux ayant subi la plus importante perte de biodiversité (SALA et al., 2000). Il est donc crucial de pouvoir suivre l'état de santé de ces écosystèmes afin de pouvoir guider les futures politiques de conservation.

Cette nécessité de suivre l'état de santé de la végétation à l'échelle globale a nécessité de définir des *Essential Biodiversity Variables* (EBV) pour unifier les efforts de recherche (PEREIRA et al., 2013). Bien qu'un certain nombre de ces variables ne soient uniquement accessibles que grâce à des mesures *in situ*, coûteuses et spatialement limitées, d'autres sont potentiellement mesurables par télédétection aéroportée et, surtout, satellitaire. La télédétection satellitaire représente une fantastique opportunité d'accroître spatialement et temporellement les jeux de données associés aux EBV en fournissant une couverture globale et régulière des écosystèmes (Pettorelli et al., 2016).

La plupart des méthodes de télédétection passive aéroportée ou satellite fournissent des informations sur les surfaces terrestres en mesurant le rayonnement lumineux réfléchi sur l'intervalle spectral 0.4–2.5 µm (Figure 2) à l'aide de capteurs multispectraux ou hyperspectraux En effet, de nombreux traits de végétation influent sur les propriétés optiques des plantes, à l'échelle de la feuille comme de la canopée (Cavender-Bares et al., 2017 ; Ustin et Gamon, 2010), ce qui rend possible leur estimation à partir de l'analyse du spectre lumineux.

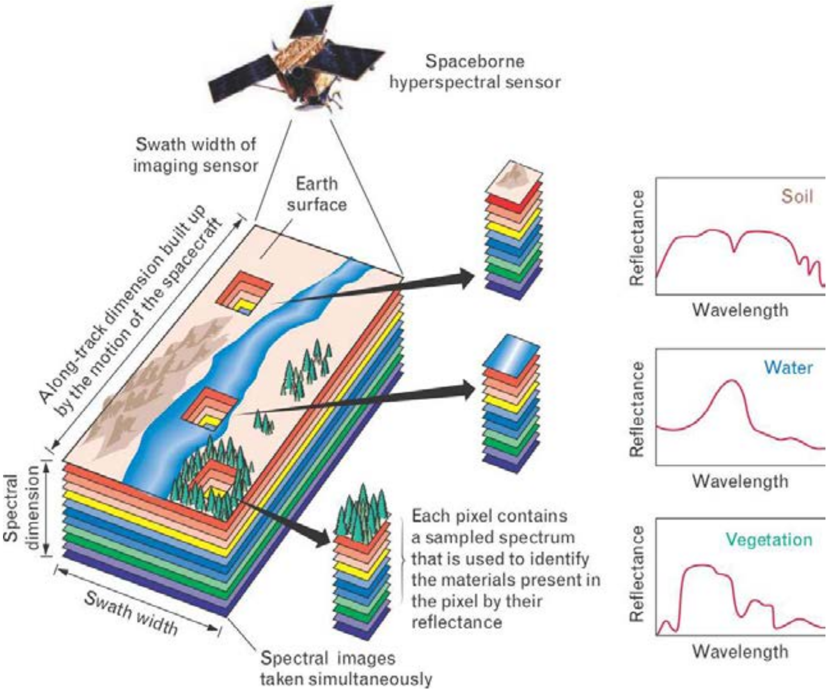


Figure 2 – Principe de la télédétection passive des surfaces dans le domaine réflectif : un capteur mesure la réflectance spectrale à différentes longueurs d'onde sur de larges zones. Le signal capturé dans chaque pixel est ensuite traitée pour obtenir des informations sur la scène. De Shaw et Burke, 2003.

La plupart des imageurs hyperspectraux actuellement utilisés sont aéroportés et ont de hautes résolutions spectrales (Schaepman et al., 2015 ; Vane et al., 1993 ; Wieneke et al., 2016) et, en général, un *Ground Sampling Distance* (GSD) allant du submétrique au décamétrique. Les images acquises par ces systèmes permettent en général de résoudre les couronnes d'arbres, et d'estimer de nombreux traits de végétation (Asner et al., 2011b ; Dana Chadwick et Asner, 2016). Cependant, les acquisitions aéroportées sont temporellement et spatialement limitées, avec au mieux quelques survols sur certaines zones d'intérêt.

La télédétection par satellite permet de s'affranchir de ces deux limitations. Historiquement, la plupart des capteurs satellitaires ont été multispectraux, *i.e.* ne mesuraient que quelques larges bandes spectrales, ce qui limitait considérablement le nombre de traits de végétation pouvant être estimés (Barnes, Xiong et Salomonson, 2002 ; Frampton et al., 2013 ;

Markham et al., 2004). Les capteurs hyperspectraux sont plus récents (EO-1/Hyperion a été lancé en 2000 (Pearlman et al., 2003)) et n'ont pas été très nombreux. Ces dernières années, de nouvelles missions hyperspectrales (PRISMA (Stefano et al., 2013), EnMAP (Guanter et al., 2015)) ont commencé et d'autres, comme SBG (Lee et al., 2015), CHIME (Rast et al., 2019) ou Biodiversity (Carrere et al., 2013), sont en préparation. Cependant, les résolutions spatiales des capteurs satellitaires hyperspectraux devraient varier de 8 à 30 m, ne permettant pas de résoudre complètement les couronnes d'arbres et augmentant grandement les difficultés d'estimations.

D'une façon générale, il a été démontré que les capteurs hyperspectraux étaient très prometteurs pour le suivi de la végétation, et les traits les plus souvents estimés sont le *Leaf Area Index* (lai) et le contenu en chlorophylle des feuilles (Berger et al., 2018). Cependant la variété de traits estimables à l'aide d'instruments satellitaires, et les méthodes pour ce faire, doivent encore être identifiées (Hill, Buddenbaum et Townsend, 2019). Ceci est particulièrement vrai pour les forêts éparses, dominantes dans les biomes méditerranéens, car le signal végétal peut être particulièrement faible aux résolutions spatiales des capteurs satellitaires.

Objectifs de la thèse

Dans le contexte des missions satellitaires hyperspectrales en cours et futures (spécifiquement SBG (NASA) et Biodiversity (CNES)), la présente thèse cherche à développer des méthodes d'estimation de traits de végétation de canopées éparses à partir d'imagerie hyperspectrale moyenne résolution ($8 \leq$ GSD ≤ 30 m), quand l'insuffisance de données terrain ne permet pas de calibrer directement des modèles de régression.

Dans le cadre de la dérogation accordée par l'ISAE, ce document complète le manuscrit de thèse, rédigé en anglais, et présente une version synthétisée du document rédigée en français.

Chapitre 1

État de l'art de la télédétection hyperspectrale de la végétation et méthodes d'estimations associées

---

1.1 Introduction

Ce chapitre présente un état de l'art des méthodes actuellement utilisées pour la télédétection des traits de végétation. Tout d'abord, les principes physiques de l'interaction entre la végétation et le rayonnement lumineux sont donnés, de l'échelle de la feuille à celle de la canopée. Ensuite, les différents outils pour modéliser ces interactions sont présentés. Enfin, la dernière section présente les principales familles de méthodes d'estimation et dans quel contexte elles ont été employées.

1.2 Propriétés optiques des traits de feuilles et de canopées

Les propriétés optiques des canopées dépendent de nombreux facteurs, les radiations étant affectées par le sol, le sous-bois, les éléments non-photosynthétiques (troncs et branches) de la végétation, et les feuilles. Tous les éléments constitutifs des feuilles influencent la signature spectrale de la végétation dans le visible (0.4–0.75 µm) et l'infrarouge (0.75–2.45 µm), comme illustré dans la Figure 1.1. Cette section présente l'influence de divers traits de feuilles et de canopées sur le signal réfléchi.

1.2.1 Pigments foliaires

Les plantes sont des organismes photoautotrophes, capturant des photons et du carbone pour créer des éléments organiques complexes. Les photons sont capturés dans les chloroplastes grâce à deux pigments foliaires : les chlorophylles *a* et *b*, dont les pics d'absorption sont situés dans les régions du violet, bleu et rouge. Ces photons augmentent les niveaux d'énergie des chlorophylles, qui vont ensuite transférer cette énergie pour son utilisation photochimique, la convertir en chaleur, ou bien émettre un photon au cours du processus de fluorescence (Krause et Weis, 1991).

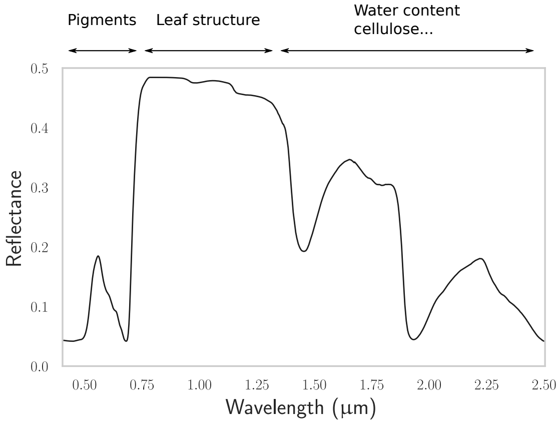


FIGURE 1.1 – Intervalles d'influence de quelques traits de feuilles sur un spectre de réflectance de feuille typique.

Les chloroplastes contiennent la majorité des nutriments de la plante, et si les conditions sont favorables, les chlorophylles sont constamment renouvellées pendant la période d'activité de la plante. À l'automne, les chlorophylles des arbres décidus ne sont plus remplacées lorsqu'elles sont détruites, afin d'assurer la relocation des nutriments dans les organes de stockage de la plante avant la perte des feuilles (MIGITA, CHIBA et TANGE, 2007). Du fait de leur importance dans le processus photosynthétique et leur forte signature spectrale liée à leur abondance, les chlorophylles sont les pigments foliaires les plus estimés.

Les caroténoïdes représentent une famille de pigments foliaires secondaires divisable en deux sous-groupes, les carotènes et les xantophylles. Ces pigments ont une double fonction : photosynthétique et photoprotective (DEMMIG-ADAMS et ADAMS, 1996 ; RITZ et al., 2000). En effet, ils complémentent la capture de photons des chlorophylles en les absorbant à d'autres longueurs d'onde, et préviennent les dommages que pourrait amener un excès d'énergie en la dissipant sous forme de chaleur via le cycle des xantophylles. Une illustration des différents transferts d'énergie entre les pigments foliaires est donnée Figure 1.2. Les chlorophylles étant détruites avant les caroténoïdes, ces derniers deviennent visibles pendant la période de sénescence des feuilles (SANGER, 1971).

Les anthocyanes sont un ensemble de pigments principalement responsables de la coloration rouge des feuilles. Si leur rôle exact est encore mal compris, il est considéré qu'elles sont créées durant les périodes de stress afin de protéger les photosystèmes d'un rayonnement excessif (FLINT, JORDAN et CALDWELL, 1985). En dehors des quelques espèces de plantes à feuilles rouges, les anthocyanes sont seulement présentes dans les feuilles juvéniles et sénescentes, ou en réponse à des sécheresses ou des basses températures (CHALKER-SCOTT, 1999 ; MERZLYAK et al., 2008).

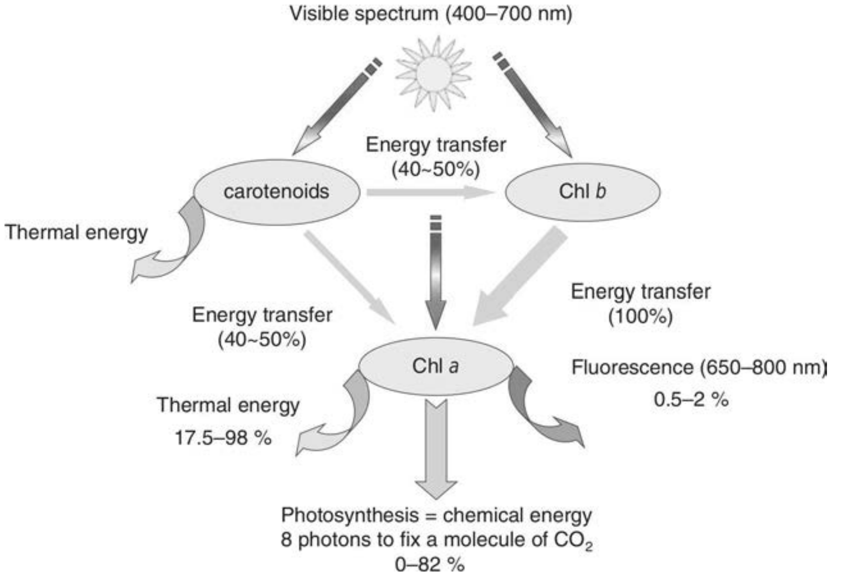


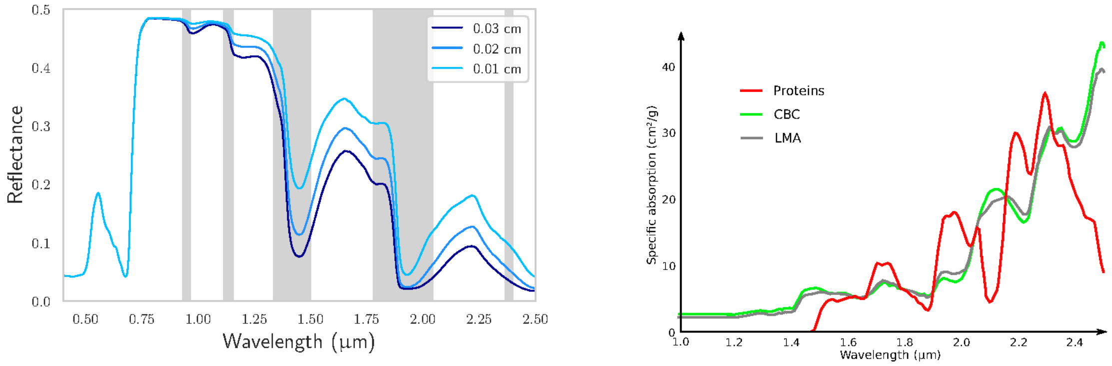
FIGURE 1.2 – Transferts d'énergie entre les différents pigments photosynthétiques des feuilles. De JACQUEMOUD et USTIN, 2019.

1.2.2 Contenus en eau et matière sèche

L'eau représente entre 70 et 95% de la masse des feuilles, et permet le transport de divers éléments au sein des végétaux. Elle assure également la rigidité des plantes, en exerçant une pression positive sur les parois cellulaires (la turgescence). Les plantes nécessitent un apport conséquent en eau : 99% de l'eau absorbée par les plantes est utilisée pour la thermorégulation du végétal, par transpiration via les stomates des feuilles, qui capturent le CO_2 nécessaire à la photosynthèse (LAMBERS, CHAPIN et PONS, 2008). Du fait de son abondance dans les feuilles, sa signature spectrale est clairement visible au niveau de la feuille (Figure 1.3a). À l'échelle aéroportée, l'estimation de l'épaisseur équivalent en eau des feuilles (*Equivalent Water Thickness* (EWT)) est plus complexe car nombre des bandes d'absorptions sont également affectées par la vapeur d'eau atmosphérique (SAN et SÜZEN, 2011 ; THENKABAIL, LYON et HUETE, 2018).

La masse surfacique foliaire (*Leaf Mass per Area* (LMA)) est définie comme le ratio de la masse de la feuille sèche sur sa surface. Le LMA est lié aux processus physiologiques des végétaux, les feuilles présentant différents LMA selon leur exposition au rayonnement solaire. La plupart des variations de LMA sont dues à des différences anatomiques et biochimiques entre les espèces, selon leurs stratégies écologiques : des plantes à croissance lente avec un haut LMA ont des parois cellulaires plus épaisses et plus de lignine (ARENDONK et POORTER, 1994 ; WESTOBY et al., 2004). La matière sèches des feuilles est la somme de ses protéines, de la cellulose, de la lignine, des sucres et de nombreux autres composants affectant principalement le rayonnement infrarouge (Figure 1.3b).

Les propriétés optiques des feuilles dépendent principalement de leur biochimie et de leur anatomie. Cependant, d'autres paramètres influent sur la réflectance à l'échelle de la canopée,



(a) Spectres de réflectance de feuilles à différentes valeurs d'épaisseur équivalent en eau, simulés par PROSPECT. Les zones grisées correspondent aux bandes d'absorption de la vapeur d'eau atmosphérique.

(b) Coefficients spécifiques d'absorption des protéines, composants carbonés et de la masse foliaire surfacique. Adapté de Féret et al., 2021.

Figure 1.3 – Influence du contenu foliaire en eau et en matière sèche sur les propriétés optiques.

car l'intégralité de sa structure peut affecter le rayonnement réfléchi.

1.2.3 Caractéristiques structurelles de l'arbre et de la canopée

La fraction de canopée couverte par le pixel d'un capteur est constituée de centaines ou de milliers de feuilles disposées et orientées de différentes façons selon l'éclairement solaire, les espèces présentes et les conditions environnementales, ce qui influe sur sa réponse spectrale. La probabilité qu'un photon ne soit pas intercepté par la végétation lors de son passage à travers la canopée est défini comme la fraction de trou de la canopée. Cette grandeur peut être obtenue à partir de photographies hémisphériques ou d'instruments spécifiques comme le LAI-2000 (Welles et Cohen, 1996) ou bien des lidars 3D (Danson et al., 2007). La fraction de trou est fréquemment utilisée pour obtenir différents traits de canopée comme le LAI (Martens, Ustin et Rousseau, 1993 ; Nilson et Kuusk, 2004). D'une façon générale, la fraction de trous détermine la proportion de sol "visible" par un imageur, et donc la quantité de signal utile pour la télédétection de la végétation, car le rayonnement mesuré est un mélange du signal du sol, du sous-bois et de la canopée.

La distribution angulaire des feuilles influe sur l'interception de la lumière et sa diffusion au sein de la canopée (Wang, Li et Su, 2007). Elle représente la probabilité que chaque feuille ait une inclinaison spécifique : en effet, cette distribution n'est pas aléatoire et peut appartenir à différentes familles (Figure 1.4). Cependant, la variabilité spatiale et temporelle du LAD, même au sein d'une même espèce, est élevée et difficile à déterminer. Du fait de connaissances insuffisantes, le LAD est souvent supposé sphérique pour les espèces de feuillus de régions tempérées (Barclay, 2001 ; Pisek et al., 2013).

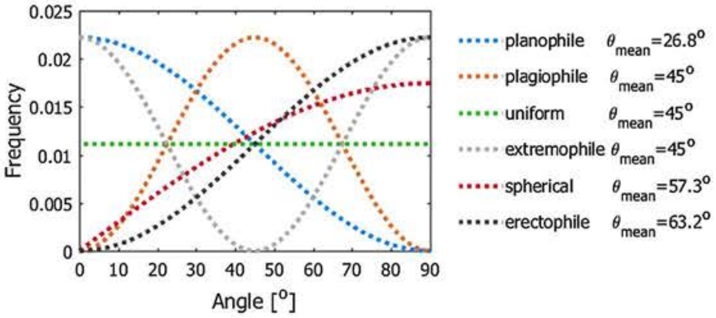


FIGURE 1.4 – Fonctions mathématiques utilisées pour approximer le LAD, et l'inclinaison foliaire correspondante. De LIU et al., 2019.

L'interception de la lumière par la canopée est également déterminée par le LAI, le ratio entre la somme de la surface supérieure des feuilles présentent au sein d'une zone sur la surface de cette zone. Cet indice varie naturellement selon les saisons, les décidus perdant leurs feuilles pendant leur période de dormance (Figure 1.5), et est d'un grand intérêt pour le suivi des échanges de carbone, d'eau et d'énergie entre la végétation et l'atmosphère (MISSON et al., 2007). Cependant, l'estimation du LAI est difficile, et les méthodes non destructives nécessitent souvent des capteurs optiques. Il est généralement ardu de distinguer les feuilles des éléments non photosynthétiques lors du traitement des images que ces capteur acquièrent, si bien que seul le *Plant Area Index* (PAI) peut souvent être estimé sans hypothèses fortes, bien que des facteurs de conversion puissent être utilisés pour passer du PAI au LAI (WOODGATE et al., 2016).

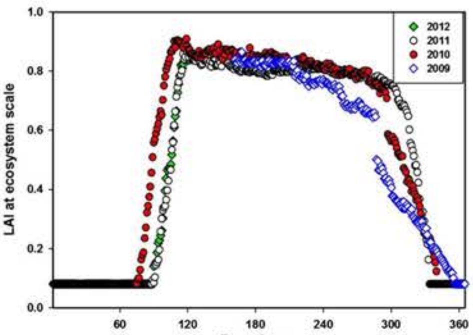


FIGURE 1.5 – Série temporelle du LAI d'une savane boisée. De RYU et al., 2012.

Les éléments non photosynthétiques de la végétation, comme les éléments ligneux, influencent également la réflectance de la canopée. En effet, la distribution des feuilles dans la canopée étant fortement hétérogène, certaines parties de la couronne peuvent présenter une large proportion d'élément ligneux et mener à une importante dispersion du rayonnement lumineux. Ainsi, MALENOVSKÝ et al., 2008a ont montré que dans une jeune forêt de sapins, jusqu'à 13% de la surface de la canopée était constituée de bois : en conséquence, la contribution relative des constituants biochimiques des feuilles est réduite, limitant la précision des estimations si le bois n'est pas correctement pris en compte (VERRELST et al., 2010).

Cette section a montré que les propriétés structurelles et biochimiques de la canopée influent sur ses propriétés optiques dans le visible et l'infrarouge. Cela a permis le développement de modèles physiques simulant les interactions entre la végétation et le rayonnement électromagnétique. Une vue d'ensemble des modèles de transfert radiatifs (*Radiative Transfer Model* (RTM)) disponibles est présentée dans la section suivante.

1.3 Modèles de transfert radiatif

À l'aide de quelques hypothèses simplificatrices sur les objets qu'ils simulent, plusieurs RTM décrivant l'absorption et la dispersion de la lumière par la végétation ont été développés et validés à différentes échelles.

1.3.1 À l'échelle de la feuille

Le modèle PROSPECT (Jacquemoud et Baret, 1990) est le RTM de feuille le plus utilisé. Il simule la réflectance et la transmittance des feuilles sur l'intervalle 0.4–2.5 μm. PROSPECT fait l'hypothèse que les feuilles peuvent être modélisées comme un empilement de N couches donc les propriétés optiques sont un mélange de celles des différents composants biochimiques des feuilles (Figure 1.6). Bien que ce modèle ne soit pas bien adapté pour les aiguilles des conifères, il a été montré que PROSPECT pouvait mener à des estimations plus précises que le modèle LIBERTY (Dawson, Curran et Plummer, 1998), consacré aux aiguilles, au prix de facteurs de normalisation pour prendre en compte la différence de structure des feuilles (Moorthy, Miller et Noland, 2008).

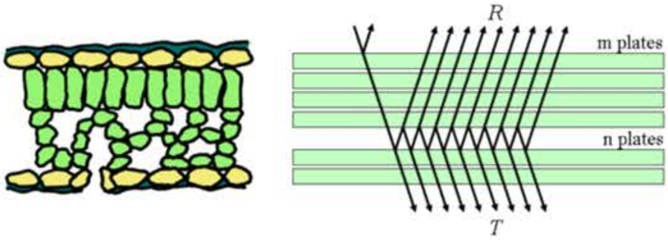


Figure 1.6 – Représentation schématique d'une feuille de dicotylédone (à gauche) et la modélisation au sein de PROSPECT avec N=m+n couches (à droite). R : réflectance ; T : transmittance. De `http://photobiology.info/Jacq_Ustin.html`.

Bien que PROSPECT ne prenait initialement en compte que *Chlorophyll Content* (C_{ab}), EWT et le paramètre structurel N, de nombreuses versions ont depuis été développées, introduisant *Carotenoid Content* (Car), LMA (Feret et al., 2008), la fluorescence, et les anthocyanes (Féret et al., 2017). Plus récemment, le LMA a été raffiné afin de distinguer les protéines des autres éléments carbonés (Féret et al., 2021). PROSPECT a été largement validé et est le RTM de feuille de référence, et la pierre angulaire de la plupart des RTM de canopée.

1.3.2 À l'échelle de la canopée

Une multitude de RTM de canopée ont été développés, représentant la canopée de façon plus ou moins abstraite. Des exercices de comparaisons inter-modèles (comme les exercices RAMI), ont régulièrement lieu pour comparer les sorties de chaque modèle (Pinty et al., 2004 ; Pinty et al., 2001 ; Widlowski et al., 2013 ; Widlowski et al., 2007, 2015). Deux grandes familles de RTM de canopée existent : les modèles homogènes et hétérogènes.

1.3.2.1 Modèles de transfert radiatif homogènes

Les RTM homogènes, également nommés 1D, modélisent la canopée comme une couche turbide homogène infinie dont la transmittance dépend principalement du LAI, selon une loi de Beer-Lambert. La signature spectrale des feuilles est déterminée par PROSPECT. Grâce à ces approximations, les modèles 1D résolvent directement les équations du transfert radiatif pour obtenir les propriétés optiques des scènes, et sont donc très rapides d'exécution (Kuusk, 2017).

Le RTM de cette famille le plus populaire est PROSAIL, résultant du couplage du RTM de canopée SAIL (Verhoef, 1984) avec PROSPECT. Il est largement utilisé du fait de sa simplicité et de sa précision (Berger et al., 2018 ; Jacquemoud et al., 2009), pour créer des indices de végétation (*Vegetation Index* (VI)) (Delegido et al., 2014 ; Haboudane et al., 2002 ; Maire et al., 2008) ou pour réaliser des estimations à partir de spectres de réflectance (Colombo et al., 2008 ; Darvishzadeh, Matkan et Dashti Ahangar, 2012 ; Li et al., 2008 ; Weiss et al., 2000).

1.3.2.2 Modèles de transfert radiatif hétérogènes

Les modèles hétérogènes, également dits 3D, sont bien plus complexes (Figure 1.7). Le niveau de détail et de complexité peut grandement varier entre les RTM 3D : le modèle INFORM (Atzberger, 2000) ne se soucie pas de la forme des couronnes et décrit la canopée comme une probabilité que le capteur voie une couronne d'arbre ou le sol ; FLIGHT (North, 1996) décrit la géométrie des troncs et des couronnes et les crée ensuite suivant un modèles de croissance, et considère que le feuillage est homogène au sein des couronnes ; des scènes arbitrairement complexes peuvent être modélisées dans raytran (Govaerts et Verstraete, 1998).

La troisième phase des exercices RAMI a montré qu'un certain nombre de RTM 3D étaient régulièrement en accord concernant les propriétés optiques de différentes canopées. Du fait de leur qualité, leurs sorties peuvent à présent servir de "vérité de substitution" pour l'évaluation d'autres RTM (Widlowski et al., 2007). Ces modèles sont : DART (Gastellu-Etchegorry et al., 2015), drat (Lewis, 1999), FLIGHT, Rayspread (Graziani et al., 2006) et raytran.

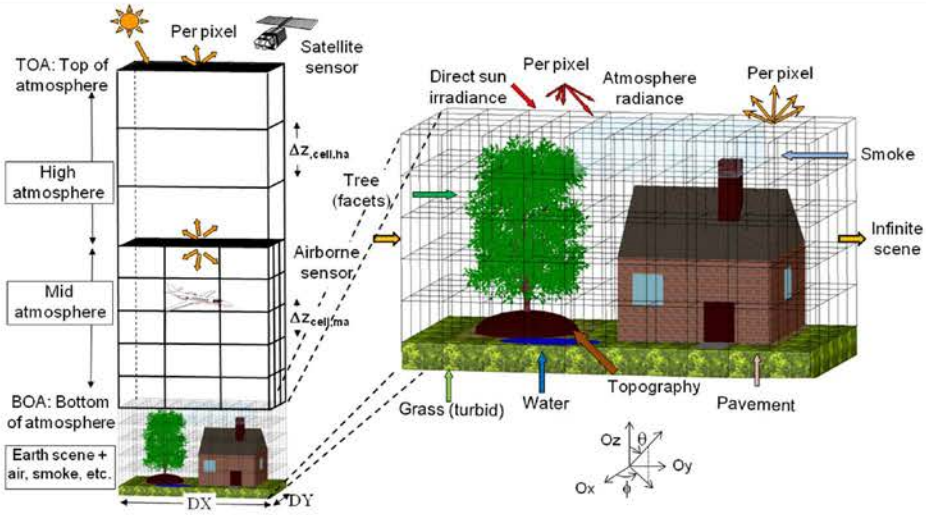


Figure 1.7 – Représentation du système Terre/atmosphère au sein de DART. Les différents éléments peuvent être modélisés à l'aide de facettes ou de cellules turbides. De Gastellu-Etchegorry et al., 2015.

Ces modèles sont très complexes et requièrent une connaissance experte pour préparer les données d'entrée et correctement mettre en place les simulations. L'amélioration de leurs performances via des représentations plus réalistes ou plus adaptées, en améliorant la modélisation du sous bois (Melendo-Vega et al., 2018), de la couronne des arbres (Widlowski, Côté et Béland, 2014), ou la représentation générale (Janoutová et al., 2019), est un sujet de recherche constant. De plus, ces RTM sont gourmands en ressources, ce qui peut plus encore restreindre leur utilisation. Néanmoins, ils ont été largement utilisés pour mieux comprendre les paramètres influençant les propriétés optiques des canopées (Malenovský et al., 2008b ; Verrelst et al., 2010) ou les flux d'énergie et de carbone des forêts hétérogènes (Kobayashi et al., 2012). De même que les modèles 1D, ils sont également utilisés dans le cadre de l'estimation des contenus biochimiques des canopées par imagerie hyperspectrale (Darvishzadeh et al., 2019 ; Hernández-Clemente, Navarro-Cerrillo et Zarco-Tejada, 2012 ; Malenovsky et al., 2007 ; Zarco-Tejada et al., 2013).

Cette section a présenté différents outils développés pour modéliser et étudier la structure et la biochimie des feuilles et des couverts. Cependant, les méthodes d'estimations des traits de végétation sont nombreuses et ne requièrent pas toutes des RTM.

1.4 Méthodes d'estimation des traits de végétation par télédétection

Peu de méthodes d'estimation par télédétection ont été effectivement mises en pratique dans des chaines de traitement de données opérationnelles, et la grande majorité sont encore au stade expérimental du fait du cadre limité de l'étude ayant mené à leur développement.

Trois grand groupes peuvent être identifiés : les méthodes empiriques-statistiques, créant un modèle entre les variables à estimer et les propriétés optiques des canopées et requérant des mesures terrains pour la calibration ; les méthodes physiques, utilisant des RTM pour résoudre le problème inverse qu'est l'estimation par télédétection ; les méthodes hybrides, utilisant les sorties des RTM utilisés en mode direct pour entraîner des modèles empiriques-statistiques (Verrelst et al., 2015).

1.4.1 Méthodes empiriques-statistiques

1.4.1.1 Méthodes paramétriques

Les méthodes paramétriques sont les plus anciennes de ce groupe. Elles font l'hypothèse qu'une relation explicite peut être établie entre une grandeur et la variable d'intérêt. L'approche la plus commune est d'établir VI, une fonction de valeurs de la réflectance à différentes longueurs d'ondes, et de le relier au trait de végétation à estimer (voir la Figure 1.8 pour un exemple). Initialement très utilisés du fait de leur simplicité et de la prévalence historique des données multispectrales, ils sont toujours utilisés pour traiter les images hyperspectrales. Une multitude de VI ont été développés, pour corriger l'influence du sol, estimer les caroténoïdes, le contenu en eau de la végétation ou encore le LMA (Gao, 1996 ; Gitelson et al., 2002 ; Haboudane et al., 2002 ; Maire et al., 2008).

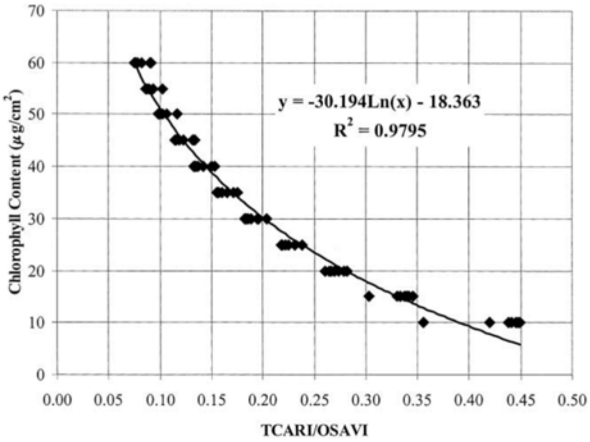


Figure 1.8 – Fonction prédictive de C_{ab} par rapport au VI TCARI/OSAVI. De Haboudane et al., 2002.

Les VI n'utilisent cependant qu'un nombre limité de bandes spectrales et ne tirent pas pleinement avantage de toute l'information spectrale, ce qui peut limiter leur transfert à d'autres sites ou capteurs. Des indices tirant parti de hautes densités spectrales ont également été développés, mais ils se concentrent toujours uniquement sur certaines parties du visible ou de l'infrarouge.

1.4.1.2 Méthodes non paramétriques

Ces méthodes calibrent un modèle non explicite sur les données pour obtenir des estimateurs. Elles peuvent tirer parti de tout le spectre de réflectance lors de leur entrainement, assignant elles même un poids à chaque bande spectrale. Cependant, cela peut être problématique lors du traitement d'images hyperspectrales : par exemple, Axelsson et al., 2013 a montré que la colinéarité des bandes spectrales proches affectait négativement les performances des machines à vecteurs de support (*Support Vector Machine* (SVM)). Un prétraitement spectral peut donc être nécessaire pour atteindre de bonnes performances. Néanmoins, les méthodes non paramétriques ont rencontré du succès dans le cadre de l'estimation de traits de végétation. Ainsi, Asner et al., 2011a a démontré le potentiel de la régression des moindres carrés partiels (*Partial Least-Square Regression* (PLSR)) pour l'estimation de la biochimie des canopées à l'échelle globale, et la PLSR a été utilisée pour estimer de nombreux traits foliaires à partir d'images haute-résolution acquises au dessus de forêts tropicales (Asner et al., 2015 ; Dana Chadwick et Asner, 2016).

Les méthodes non paramétriques ont deux contraintes principales : (i) les connaissances poussées requisent pour ajuster les hyperparamètres optimisant les modèles tout en évitant l'overfit et (ii) la grande quantité de mesures terrain nécessaires pour construire les bases de données d'entrainement et de validation. Bien que (i) ne soit pas un problème majeur, (ii) est plus contraignant car obtenir des données *in situ* pour chaque écosystème est inenvisageable.

1.4.2 Méthodes physiques

Les méthodes physiques utilisent les RTM présentés Section 1.3. Contrairement aux méthodes précédentes, elles peuvent fournir des estimations avec peu ou pas de connaissances sur la canopée imagée. Ces méthodes reposent sur l'inversion d'un RTM, en faisant l'hypothèse que la modélisation de la canopée au sein de ce dernier est représentative de la réalité. Pour ce faire, la méthode la plus commune est de générer une table de référence (*Look-Up Table* (LUT)) et de sélectionner les entrées dont la réflectance correspond le plus à celle du pixel de l'image. Cependant, le problème d'inversion est mal posé : différentes combinaisons des paramètres d'entrée des RTM peuvent conduire à des sorties similaires. Pour palier cela, différentes méthodes ont été développées, comme ajouter du bruit sur les données générées (Richter et al., 2011), sélectionner plusieurs "meilleures" solutions (Weiss et al., 2000), ou contraindre le modèle en utilisant des connaissances *a priori* (Atzberger et Richter, 2012 ; Combal et al., 2003).

Les méthodes reposant sur l'inversion d'un RTM utilisent toujours une fonction de coût pour trouver les meilleures correspondances, que celle-ci prenne la forme d'un VI, d'un *Root-Mean-Square Error* (RMSE), ou de M-estimateurs (Verrelst et al., 2012). Choisir une fonction de coût adaptée peut nécessiter des connaissances *a priori*, car la réponse spectrale à des variations de traits de végétation sera mesurée différemment selon les écosystèmes et les capteurs.

1.4.3 Méthodes hybrides

Les méthodes hybrides combinent les méthodes physiques et empiriques-statistiques, en remplaçant les mesures terrain nécessaires pour construire les jeux de données d'entrainement par les sorties d'un RTM. Ainsi, les méthodes hybrides tirent parti à la fois de l'efficacité des méthodes de régression et de leur capacité à tirer parti de tout le spectre de réflectance, tout en ayant l'avantage de se passer au maximum de données acquises *in situ*. Bien que les méthodes hybrides aient initialement principalement utilisé des *Artificial Neural Network* (ANN), d'autres méthodes non paramétriques ont commencé à être utilisées avec succès dans les années 2000 pour estimer différents traits de végétation.

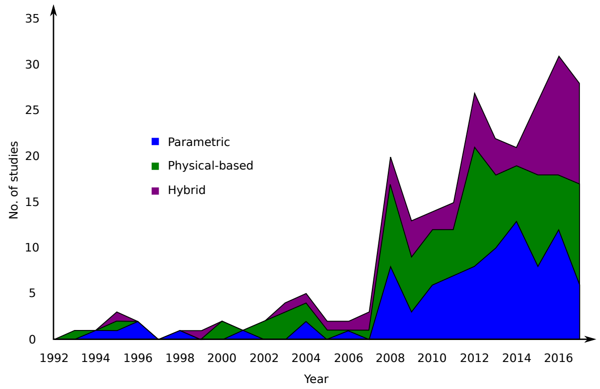


FIGURE 1.9 – Évolution des différentes méthodes d'estimation utilisant PROSAIL. "Paramétrique" fait ici référence à des méthodes paramétriques calibrées sur les sorties du RTM. Adapté de BERGER et al., 2018.

Bien que les méthodes hybrides nécessitent toujours de correctement configurer le RTM, ces dernières années leur utilisation a considérablement augmenté (voir Figure 1.9 pour un exemple avec PROSAIL). Étant plus appropriées pour le traitement de données ayant de grandes dimensions, elles pourraient être les méthodes les plus appropriées pour des estimations dans un cadre opérationnel (BERGER et al., 2018).

1.5 Synthèse et approche choisie

De l'état de l'art, il apparaît que nombre de traits de végétation ont une influence claire sur la réflectance de la canopée dans le visible et l'infrarouge, et peuvent donc être estimés par télédétection.

Plusieurs méthodes ont été développées pour estimer ces traits à partir d'imagerie aéroportée et satellitaire, utilisant soit des données empiriques collectées sur le terrain ou des données synthétiques issues de RTM. Si il a été montré que les méthodes empiriques obtenaient de bons résultats, les méthodes physiques ou hybrides sont employées avec de plus en plus de succès. Ces méthodes semblent donc être des candidats idéaux pour des utilisations dans un cadre opérationnel, car elles ne requerraient pas de longues et coûteuses campagnes terrain.

Avec l'arrivée des missions Sentinel-2, il a été démontré qu'un suivi global de la végétation était possible et que des traits de végétation pouvaient être estimés par imagerie multispectrale satellitaire. Des méthodes tirant parti de données hyperspectrales, bien que trop souvent cantonnées à l'imagerie aéroportée, ont montré le fort potentiel de ce type d'imagerie. Cependant, ces études se sont le plus souvent concentrées sur des canopées fermées et principalement sur l'estimation du LAI et de C_{ab}.

Comme cette thèse cherche à développer des méthodes d'estimation de traits de végétation avec peu de données terrain, il a été nécessaire de se concentrer sur des méthodes d'estimation reposant sur des RTM. Du fait de l'hétérogénéité des canopées éparses, il a été décidé d'utiliser le RTM 3D DART, qui fait parti des 6 modèles de référence identifiés lors des exercices RAMI.

Chapitre 2

Identification d'une méthode d'estimation du LAI et des pigments foliaires de couverts épars

2.1 Objectifs

Le principal objectif de cette étude était d'identifier une stratégie d'inversion adaptée au suivi de la santé d'un écosystème herbe-arbre, c'est à dire capable d'estimer le LAI et le contenu foliaire en C_{ab} et Car des arbres dans le temps à partir d'images hyperspectrales aéroportées acquises avec un GSD de 18 m.

Les estimations de LAI de la canopée et des pigments foliaires ont principalement été effectuées sur des forêts présentant des LAI élevés (par exemple, avec les travaux de Banskota et al., 2015, Maire et al., 2008, Ali et al., 2016b ou Darvishzadeh, Matkan et Dashti Ahangar, 2012). Bien que quelques travaux aient porté sur les écosystèmes ayant des canopées ouvertes (comme ceux de Zarco-Tejada et al., 2004, Hernández-Clemente, Navarro-Cerrillo et Zarco-Tejada, 2012, ou de Zarco-Tejada et al., 2019), les recherches concernant des méthodes de modélisation adaptées au sein des RTM sont toujours en cours (Melendo-Vega et al., 2018 ; Widlowski, Côté et Béland, 2014). Dans cette étude, la validité d'une représentation simplifiée des arbres au sein de scènes DART (telles que celles faites par Gascon et al., 2004) dans le cadre de travaux sur des images hyperspectrales de moyenne-résolution spatiale a été testée. Ces travaux ont uniquement porté sur des feuillus, car la représentation des conifères au sein des RTM est encore un sujet difficile (Janoutová et al., 2019). Les données de validation étaient disponibles à plusieurs dates, ce qui a permis de calibrer de façon plus robuste la méthode d'inversion. La validité des résultats présentés dans cette partie a été faite à l'aide de : (i) une comparaison entre les estimations et les mesures terrain quand et où elles étaient disponibles, et (ii) en comparant les variations attendues et estimées du LAI et des pigments foliaires des arbres du site sur deux ans.

2.2 Principaux résultats

La représentation de la forêt choisie pour cette étude avait pour objectif de minimiser les connaissances *a priori* nécessaires pour suivre le site de Tonzi. Les arbres modélisés au sein de

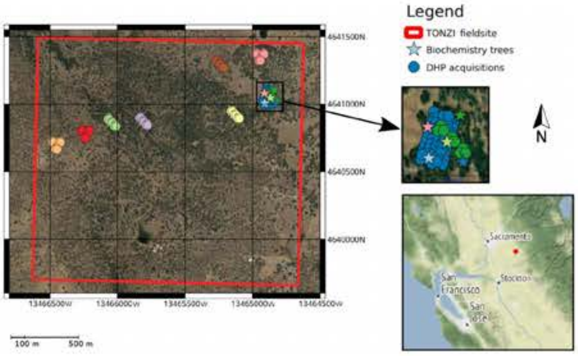
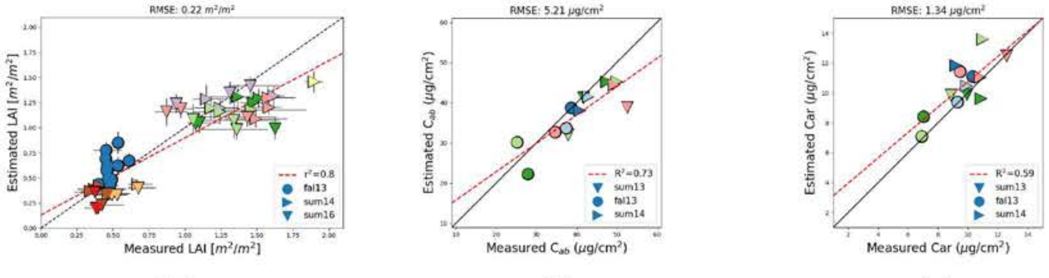


FIGURE 2.1 – Localisation du site d'étude et des mesures *in situ*. Les disques correspondent aux emplacements des photographies hémisphériques, les étoiles aux positions des arbres où des feuilles ont été collectées pour les mesures en laboratoire.

DART avaient des couronnes simplistes et peu d'éléments non photosynthétiques, les branches n'ayant pas été incluses. Cette modélisation sera désignée tout au long du document comme la modélisation *Simple Forest Representation* (SFR). Ceci affecte les estimations du LAI et de C_{ab}, car l'influence des éléments ligneux sur la canopée peut être importante (MALENOVSKÝ et al., 2008a ; VERRELST et al., 2010) pour des couverts hétérogènes et peu feuillus comme Tonzi. En particulier, MALENOVSKÝ et al., 2008a et WIDLOWSKI, CÔTÉ et BÉLAND, 2014 ont montré que l'inclusion d'élémens ligneux dans les canopées de DART affectait les réflectances dans l'infrarouge aux résolutions utilisées lors de cette étude.

Bien que le LAI du site soit très bas (LAI moyen de 0.8 m^2/m^2) et que le LAI moyen des mesures *in situ* soit de 0.88 m^2/m^2, les estimation étaient encourageantes. L'hétérogénéité du LAI sur Tonzi a été capturée : pour une date donnée, les LAI élevés comme faibles étaient correctement estimés. Les variations saisonnières étaient également visibles, la diminution du LAI entre l'été et l'automne étant clairement mesurée. En effet, sur la Figure 2.2a, les points bleu et vert foncé viennent des même parties du site (voir la carte du site, Figure 2.1) mais ont respectivement été mesurés en automne et en été.



(a) (b) (c)

FIGURE 2.2 – LAI (**a**) et contenus foliaires en C_{ab} (**b**) et Car (**c**) mesurés et estimés. Les couleurs indiquent la localisation dans le site.

Les performances d'estimation de C_{ab} étaient similaires à celles d'études précédentes. Les estimations de ces pigments foliaires (RMSE = 5.21 $\mu g/cm^2$, R^2 = 0.73) sont aussi bonnes que ce qui a pu être obtenu sur des sites présentant de plus forts LAI (RMSE = 4.34 µg/cm^2, R^2 = 0.47 (ZHANG et al., 2008) ; RMSE = 4.2 µg/cm^2 (DELEGIDO et al., 2010) ; RMSE = 8.1 µg/cm^2, R^2 = 0.40 (ZARCO-TEJADA et al., 2004)). Les estimation des caroténoïdes étaient également correctes, avec un faible RMSE et un R^2 élevé (RMSE = 1.34 µg/cm^2, R^2 = 0.59). Ces valeurs sont similaires à ce qu'ont pu obtenir ZARCO-TEJADA et al., 2013 en utilisant des images à haute résolution spatiale au dessus de champs de vignes (RMSE inférieur à 1.3 µg/cm^2 et R^2 de 0.46 au plus).

Concernant le suivi saisonnier, bien que les phénophases des arbres décidus soient telles que le LAI et le contenu en C_{ab} des arbres du site auraient dû présenter un maximum en été (DEMAREZ et al., 1999 ; GARA et al., 2019), les estimations indiquaient plutôt une décroissance du printemps à l'automne. La cause la plus probable est que l'herbe présente sur le site n'était pas complètement morte en mai : les contributions de l'herbe verte et des couronnes d'arbres étaient combinées au printemps. La réflectance de l'herbe soit normalement prise en compte dans DART, de sorte que sa biochimie ne devrait pas influencer les résultats. Cependant, les pixels de sol pur sélectionnés manuellement à partir des images AVIRIS-C pour la modélisation dans DART pourraient avoir contenu moins d'herbe verte que d'autres pixels, ce qui aurait conduit à des estimations incorrectes là où de l'herbe verte était présente. Concernant l'été et l'automne, l'herbe était complètement sèche, et les décroissances du LAI et de C_{ab} étaient correctement détectées. Ces résultats contribuent à donner de la confiance en la capacité de la méthodologie proposée à réaliser des estimations correctes sur le site à l'été et à l'automne.

2.3 Conclusion

L'estimation du LAI et des pigments foliaires à l'aide d'inversion de logiciels de transfert radiatif est difficile pour les environnements très hétérogènes. Ce chapitre a étudié la capacité d'une méthode d'inversion par LUT d'images AVIRIS-C à estimer le LAI et les contenus en pigments foliaires de canopées ouvertes. La métholodogie suivie durant cette étude n'utilisait presque aucune connaissance *a priori*, afin de pouvoir suivre le site d'étude dans son intégralité sur plusieurs dates et d'obtenir des résultats plus généraux. Les estimations faites au dessus de parties du site très différentes en terme de structure étaient cohérentes et montraient une bonne précision pour le LAI, C_{ab} et Car.

Ces résultats démontrent que l'imagerie hyperspectrale à 18 m de résolution, combinée à un RTM 3D n'utilisant qu'une représentation simplifiée de la canopée, est appropriée pour le suivi des traits de végétation de forêts ouvertes de faible LAI comme les savannes boisées. Cependant, des difficultés sont toujours présentes : bien que les LUT générées pour cette étude aient été appropriées pour estimer le LAI et les pigments, elles n'étaient pas adaptées au cas où la réflectance du sol et du sous-bois était hétérogène. En effet, cette réflectance a été fixée lors de la génération de la LUT, et il pourrait être nécessaire d'en faire un paramètre variable dans les scènes DART. Cependant, les temps de calculs requis pour générer les LUT étant déjà conséquents, rajouter des paramètres variables serait très coûteux : il serait d'abord nécessaire de pouvoir générer des LUT à moindre coût.

Chapitre 3

Raffinement de l'échantillonnage des LUT par développement d'une méthode d'accélération par utilisation conjointe de deux RTM

3.1 Objectifs

Quand les données *in situ* sont insuffisantes pour calibrer une méthode empirique-statistique, il est nécessaire d'utiliser un RTM dans le cadre de méthodes d'estimation physiques ou hybriques. Ces méthodes peuvent être coûteuses en temps, car elles requièrent des capacités de calcul importantes et un grand nombre de simulations doit être effectué afin d'arriver à des performances d'estimation acceptables. De nombreux RTM, modélisant la canopée de façon homogène (1D) ou hétérogène (3D) sont disponibles. Les modèles 1D sont adaptés aux scènes homogènes et n'ont par conception qu'une quantité limitée de paramètres : bien que peu réalistes, ils sont très rapides d'utilisation et faciles à inverser dans le cadre de l'estimation de traits de végétation. À l'inverse, les modèles 3D permettent une modélisation détaillée des différentes composantes de la canopée, à l'aide de nombreux paramètres pouvant être fixés ou laissés libres. Ces modèles sont particulièrement adaptés à la modélisation des couverts épars comme les écosystèmes arbres-herbe, très présents sur les terres émergées (Hanan et Hill, 2012), car la contribution radiative de la canopée y est limitée, le sol, le sous-bois et les ombres étant plus visibles par les capteurs. Cependant, cette complexité amène une augmentation conséquente des temps de calcul, ce qui limite l'échantillonage des paramètres variables pouvant être considéré de façon réaliste.

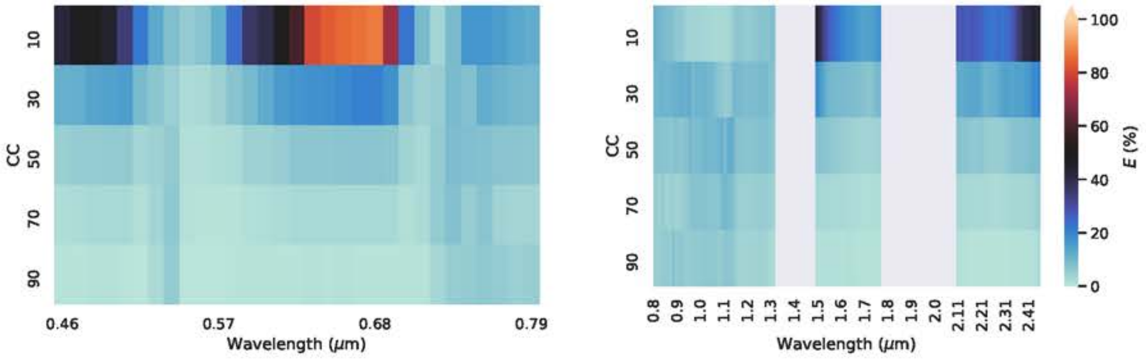
Aucun schéma d'échantillonage idéal n'a jusqu'ici été identifié : Ali et al. (2016a) a utilisé une distribution uniforme des variables sur les plages de variation mesurées *in situ* ; Weiss et al. (2000) a tiré les valeurs de chaque paramètre selon une loi de distribution proportionelle à la sensibilité de la réflectance à ce paramètre ; Ali et al. (2016b) a utilisé des distributions normales multivariées et des matrices de covariance tirées des mesures *in situ* ; Hernández-Clemente, Navarro-Cerrillo et Zarco-Tejada (2012) a utilisé des échantillonages mono et multivariés. Du fait des temps de calcul, tester différents échantillonages pour construire des LUT de dizaines de milliers d'entrées n'est pas réaliste. Être capable d'optimiser aisément l'échantillonage ayant servi à construire les LUT pourrait être grandement bénéfique.

Le but de cette étude était de pouvoir générer rapidement des LUT ayant de nombreux paramètres variables et un échantillonage arbitraire. Pour ce faire, les RTM PROSAIL (1D) et DART (3D) ont été combinés, en calibrant un modèle approximant les sorties de DART à partir de celles de PROSAIL. Les performances de ce modèle (nommé PROSAIL2DART (P2D)) ont été estimées en comparant ses sorties à des valeurs de référence de DART. PROSAIL2DART a ensuite été utilisé pour estimer le gap fraction, le contenu en pigments foliaires, et le LMA et EWT des chênes d'une savanne boisée. La précision des estimations a été vérifiée en comparant les estimations réalisées aux valeurs obtenues *in situ* à différents endroits du site et à différentes dates.

3.2 Principaux résultats

Choisir une méthode d'échantillonnage appropriée pour la génération de LUT est critique. De nombreuses méthodes d'échantillonage sont disponibles dans la littérature, et peuvent mener à des précisions d'estimation différentes en mettant en avant des plages de variation inadaptées et mener à des cas quasi-dupliqués. Comme montré dans la Figure 3.1, l'erreur intrinsèque des approximations de P2D était très faible pour toutes les longueurs d'onde à partir de CC 30%, et décroissait quand le CC augmentait : en effet, plus le CC est haut, plus une scène 3D se rapproche d'une scène 1D car l'hétérogénéité diminue. Il était ainsi possible d'approximer avec un minimum d'erreur les sorties de DART à partir du modèle PROSAIL, qui a des temps de calcul significativement moins longs. La construction du modèle de passage pour chaque LUT a nécessité au plus 12 666 heures CPU à DART pour effectuer 784 simulations, et a permis par la suite d'approximer 300 000 spectres de réflectances en 1,5 heure CPU : P2D a entrainé une réduction drastique du temps de calcul. Bien que les dernières versions de DART soient plus rapides que celle utilisée lors de cette étude, les temps d'exécution restent nettement plus longs que P2D, qui était, après calibration, presque aussi rapide que PROSAIL. Ce faible temps de calcul rendait donc possible de générer rapidement des LUT ayant des échantillonages très fins.

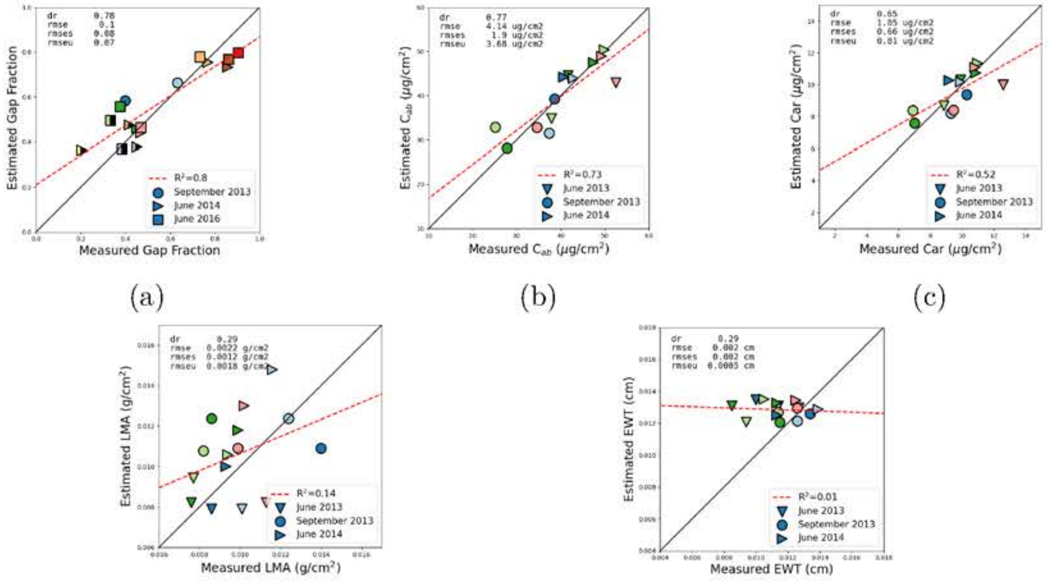
Comme visible dans les Figures 3.2a, 3.2bb et 3.2c, les estimations de gap fraction et de contenu en C_{ab} et Car obtenues grâce aux LUT générées par P2D étaient bonnes et semblables à ce qu'il était possible de trouver dans la littérature (Darvishzadeh et al., 2019 ; Maire et al., 2008 ; Zarco-Tejada et al., 2004, 2013). Les performances d'estimations de C_{ab} et Car obtenues lors de cette étude étaient similaires ou meilleures que celles obtenues dans le chapitre précédent, probablement du fait de deux raisons. (i) Dans cette étude, ce n'était pas PROSPECT-5 mais PROSPECT-D qui a été utilisé pour modéliser le comportement radiatif des feuilles. Ce dernier introduit les anthocyanes dans la modélisation et considère des coefficients d'absorption spécifique différents pour les chlorophylles et les caroténoïdes. (ii) L'échantillonnage n'était pour cette étude plus effectué selon une grille régulière, et prenait en compte la corrélation entre contenus foliaires en C_{ab} et Car. Si, en dehors de la finesse de l'échantilllonage, C_{ab} n'était pas affecté, la prise en compte de cette relation a amélioré les



(a) (b)

FIGURE 3.1 – Erreur de P2D à chaque longueur d'onde et CC dans (**a**) le visible et (**b**) le NIR. Les figures partagent la même échelle de couleur.

estimations de Car en supprimant les cas irréalistes (tels que ceux avec un fort taux de C_{ab} et un faible taux de Car), qui pouvaient créer de la confusion lors de la phase d'inversion par LUT.



(a) (b) (c)

(d) (e)

FIGURE 3.2 – Comparaison entre traits de végétation mesurés *in situ* et estimés.

La précision des estimations n'était en revanche bonne ni pour le LMA (plus de deux fois la valeur obtenue par MAIRE et al., 2008), ni pour le EWT pour lequel aucune relation n'a été trouvée entre les valeurs mesurées *in situ* et estimées. Ces résultats étaient cohérents avec ce qu'avaient obtenu YANFANG XIAO et al., 2013 : leurs résultats indiquaient que pour des LAI inférieurs à 3 il n'était pas possible d'estimer le EWT car sa contribution au signal lumineux était trop faible au regard de l'influence du LAI. Il était probable qu'une meilleure résolution spatiale des images ait été nécessaire pour estimer correctement EWT et LMA, car il aurait alors été possible de localiser les pixels de végétation pure, où les signatures spectrales de EWT et LMA sont bien plus visibles.

3.3 Conclusion

Les résultats obtenus dans ce chapitre ont démontré la possibilité d'approximer avec une erreur minimale les sorties de DART à partir de celles de PROSAIL, et ce même pour des CC de 30%. Le modèle d'approximation (P2D) a été utilisé pour générer des LUT ayant ensuite servi à estimer le gap fraction, C_{ab}, Car, EWT et LMA d'une savanne boisée. Les estimations de gap fraction et de contenus en pigments obtenues grâce à P2D étaient plus précises que celles obtenues en se servant uniquement de DART. Le temps de contruction d'une LUT à l'aide de P2D était négligeable une fois la phase de calibration effectuée, ce qui est un résultat très intéressant : d'une part, lors de la phase d'exploration, il est ainsi possible de tester différents échantillonages à un coût négligeable, et d'autre part, les sorties de P2D peuvent ensuite directement être utilisées pour estimer des traits de végétation avec une précision acceptable.

Plus de travaux étaient nécessaires pour pouvoir estimer EWT et LMA pour des écosystèmes arbres-herbe. Ce chapitre a montré que la grossièreté de l'échantillonage de la LUT n'était pas la cause de l'inaptude de la méthode à estimer correctement ces deux traits. Il était ainsi possible que la modélisation SFR utilisée dans ce chapitre n'ait pas été adaptée aux méthodes d'estimation utilisées pour ces traits de végétation. Plusieurs raffinements dans la modélisation de l'écosystème au sein de DART pourraient être nécessaires pour arriver à des résultats acceptables.

Chapitre 4

Identification d'un méthode appropriée pour l'estimation de EWT et LMA à partir de modélisations DART

4.1 Objectifs

Les RTM 3D sont les plus appropriés pour modéliser les canopées hétérogènes, qui présentent une architecture complexe. Cependant, du fait de connaissances du terrain insuffisantes, les arbres sont souvent modélisés de façon abstraite, ce qui peut mener à des simulations inadaptées car la réflectance des scènes et des couronnes peut être affectée : Widlowski, Côté et Béland, 2014 ont évalué l'influence de l'omission ou de la simplification de la structure ligneuse des arbres dans DART, et ont trouvé que cela pouvait engendrer un biais spectral significatif augmentant quand le GSD diminuait ; Ali et al., 2016a ont trouvé que la hauteur de la canopée et le diamètre des couronnes étaient deux paramètres affectant la réflectance des canopées ; Janoutová et al., 2019 ont comparé différentes modélisations de conifères dans DART pour identifier les plus adaptées. Si Wocher et al., 2018 ont utilisé une méthode hybride se basant sur PROSAIL pour estimer EWT, ils durent ajuster empiriquement les sorties de leur modèle pour prendre en compte les effets radiatifs non modélisés dans le RTM avant d'obtenir des estimations acceptables. En général, modéliser arbres et feuilles de façon abstraite dans un RTM peut mener à des variations de réflectance importantes. Dans quelle mesure cela affecte-t-il les estimations de EWT et LMA ? Y a-t-il des intervalles spectraux peu affectés par ces abstractions ? Ces questions sont d'une grande importance pour la mise en place de méthodes d'estimation de EWT et LMA avec peu d'information *a priori* sur les sites d'études, méthodes qui pourraient être aisément généralisables à des écosystèmes similaires.

L'objectif des travaux présentés dans ce chapitre était d'évaluer l'influence de diverses abstractions faites lors de la modélisation des arbres au sein d'un RTM sur la réflectance des couronnes, pour identifier des intervalles spectraux y étant peu sensibles, et de tester les intervalles identifiés en entrainant une méthode hybride (RFR) pour estimer EWT et LMA. Pour ce faire, une étude préliminaire fut effectuée sur une image synthétique afin de comparer les performances d'estimation à partir de différentes modélisations dans DART et d'obtenir des valeurs de référence. Ensuite, une étude sur des images hyperspectrales AVIRIS-NG acquises au dessus de deux sites a été effectuée (voir la Figure 4.1). La précisions des valeurs de EWT et LMA estimées a été vérifiée en les comparant aux données récoltées *in situ*, et leur cohérence a été validée en utilisant les résultats issus de l'étude sur des données synthétiques.

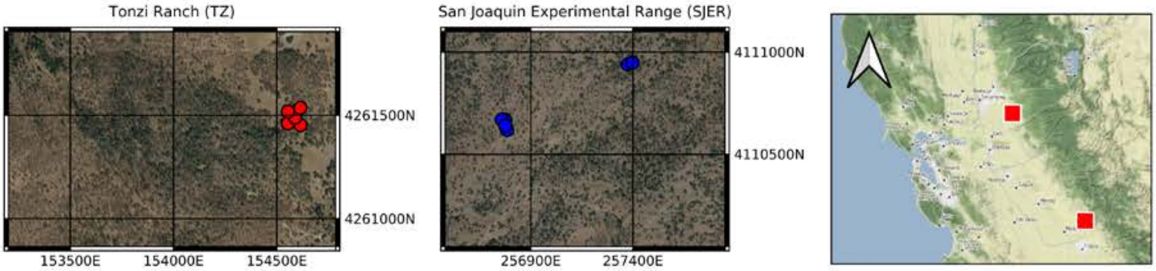


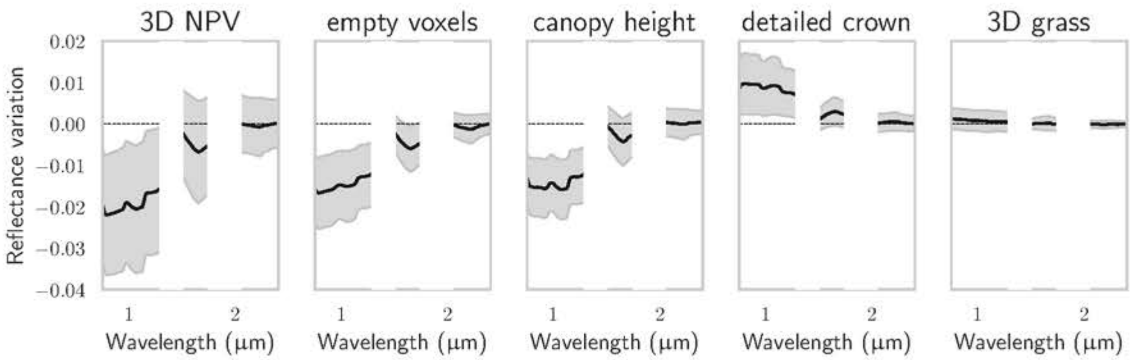
Figure 4.1 – Localisation des sites d'étude et des mesures *in situ*.

4.2 Principaux résultats

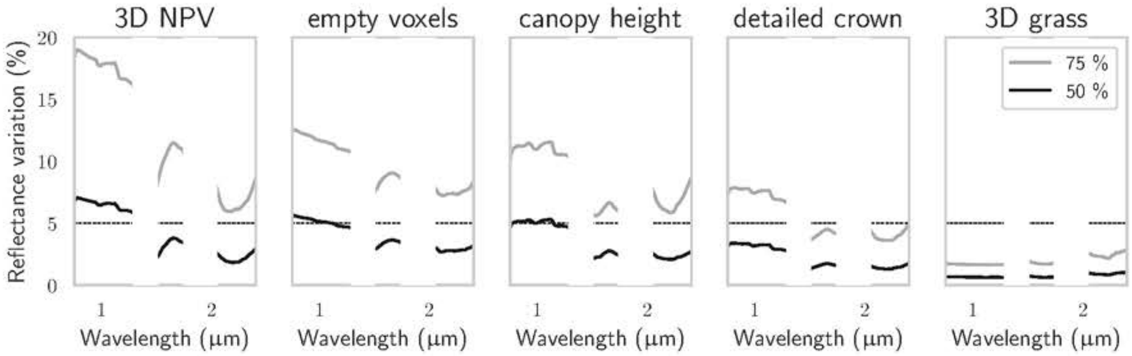
Une étude de sensibilité a démontré que la plupart des abstractions faites dans le cadre de la modélisation SFR pouvaient affecter de façon significative la réflectance des couronnes sur l'intervalle 0.75–2.45 µm, et particulièrement sur l'intervalle 0.75–1.5 µm : 50% des cas du scénario 3D NPV, correspondant à l'introduction d'une structure ligneuse détaillée, dépassaient le seuil de 5% utilisé par Widlowski, Côté et Béland, 2014 (voir la Figure 4.2). Les résultats sur l'influence des abstractions sur la réflectance des couronnes étaient corroborés par la littérature : Ferreira et al., 2018 et Malenovský et al., 2008b avaient trouvé que la présence de végétation non photosynthétiques (*Non Photosynthetic Vegetation* (npv)) dans les couronnes affectait la réflectance sur 0.7–0.9 µm en réduisant le red-edge, et Ferreira et al., 2018 avaient également montré que les couronnes sans voxels vides présentaient une réflectance plus élevée que celles en ayant. Même pour les longueurs d'onde au delà de 1.5 µm, ce seuil de 5% était excédé pour les scénarios 3D NPV, voxels vides, et hauteur de canopée pour plus de 25% des cas testés.

Ainsi, considérer les longueurs d'onde inférieures à 1.5 µm pour estimer le LMA peut amener à de grandes imprécisions si les couronnes ne sont pas modélisées de façon réalistes, même si cette région spectrale est la plus sensible au LMA (Yanfang Xiao et al., 2013). Au dessus de 1.5 µm, la majorité des combinaisons testées étaient sous le seuil de 5%. Brede et al., 2020 ont estimé la sensibilité des bandes spectrales de ces régions à différents traits de végétation pour deux RTM (DART et PROSAIL) et ont trouvé que EWT et LMA étaient les responsables principaux de la réflectance, bien que les scènes PROSAIL étaient bien entendu considérablement plus abstraites que la modélisation SFR de DART utilisée lors de l'étude. Il semblerait donc que les bandes spectrales de ces régions soient particulièrement adéquates pour estimer EWT et LMA quand les couronnes ne peuvent pas être précisément modélisées.

Sur les images synthétiques, des RFR ont été entrainés sur différents intervalles spectraux et bases de données. Une base de données avait été générée avec la modélisation SFR, la second avait été générée en prenant tous les éléments structurels précédemment testés (mo-

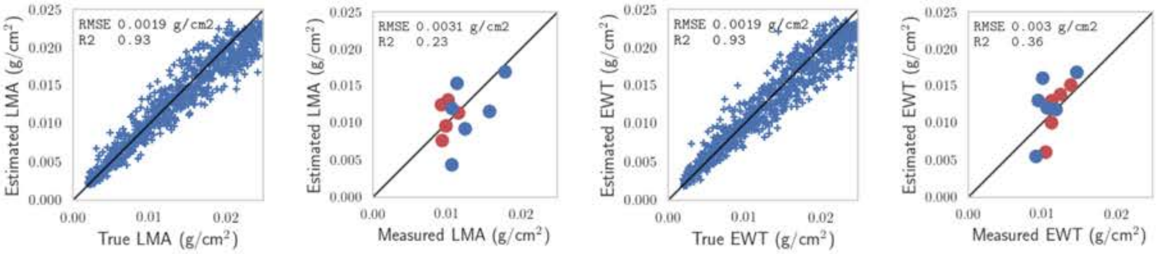


(a) La ligne noire correspond à la médiane des variations de réflectance, et la zone grisée correspond à ± une déviation standard.



(b) La ligne noire correspond à la médiane des variations relatives de réflectance, et la ligne grise au 75$^{\text{ème}}$ percentile.

FIGURE 4.2 – Variations des réflectances de couronnes causées par l'introduction d'éléments structuraux plus détaillés de façon monovariante pour 1000 cas test.



(a) (b) (c) (d)

FIGURE 4.3 – Comparaison entre les valeurs de LMA et EWT estimées et mesurées *in situ* pour les images (**a**, **c**) synthétiques et (**b**,**d**) AVIRIS-NG.

délisation DETAIL). Il a été montré que les estimations de LMA variaient significativement selon l'intervalle spectral et la base de données (voir la Figure 4.3). Si les différences de performance entre SFR et DETAIL étaient claires quand les bandes inférieures à 1.5 µm étaient incluses, forcer le RFR à ne considérer que les bandes comprises dans l'intervalle 1.5–2.4µm a permis de rejeter les bandes les plus biaisées et a rendu la précision des estimations de LMA se basant sur SFR et DETAIL équivalentes. Les estimations de EWT n'étaient, quant à elles, pas affectées par les abstractions : les bandes identifiées comme critiques par le RFR étaient, naturellement, les bandes de l'eau, peu sensibles aux abstractions pouvant être faites lors de la modélisation.

Les estimations de LMA et EWT à partir des images AVIRIS-NG présentaient des performances légèrement moins bonnes que celles obtenues sur les images synthétiques, avec des RMSE autour de 0.003 au lieu de 0.002 g/cm^2. Cependant, les RMSE étaient proches de ceux obtenus lors de l'entrainement du RFR, et donc de meilleures performances auraient été surprenantes. Bien que la plage de variation des mesures *in situ* était limitée, les résultats concernant l'estimation de EWT et LMA étaient encourageants, du fait de la multitude d'incertitudes (registration des images, identification incorrecte des pixels correspondant aux couronnes d'arbres, potentielle faible représentativité des feuilles utilisées pour obtenir les valeurs de référence pour EWT et LMA) inhérentes à la validation via des mesures terrain.

4.3 Conclusion

Ces travaux ont montré qu'une représentation abstraite des arbres au sein d'un RTM, sous la forme de couronnes ellipsoïdales sans structure ligneuse détaillée, affectait significativement l'intervalle spectral 0.75–1.5 µm. Cependant, il a été montré que cet effet était moindre au delà de 1.5 µm, et que les longeurs d'onde de l'intervalle 1.5–2.4 µm pouvaient être utilisées avec succès pour entrainer un RFR et estimer le LMA. Les estimations de EWT n'étaient, quant à elles, pas affectées par les abstractions. L'application du RFR à deux images AVIRIS-NG a permis d'obtenir des précisions d'estimation similaires à celles obtenues sur le jeu de données d'entrainement (0.0031 et 0.0030 g/cm^2 pour EWT et LMA, respectivement). Bien que les résultats étaient encourageants, plus de travaux étaient nécessaires pour vérifier la généralisation des conclusions au traitement d'images hyperspectrales qui pourraient être acquises par des capteurs satellitaires à des résolution spatiales plus grossières.

Chapitre 5

Évaluation des performances d'estimation de traits de végétation à partir d'images hyperspectrales satellitaires synthétiques

5.1 Objectifs

Du fait de leur capacité à imager de larges zones de façon récurrente, les capteurs satellitaires devraient jouer un rôle important pour les futures études environnentales, et le concept de variable de biodiversité essentielle spécifique aux satellites a récemment été établi (Pettorelli et al., 2016). Plusieurs missions hyperspectrales, comme EnMAP (Guanter et al., 2015), PRISMA (Stefano et al., 2013) ou HISUI (Iwasaki et al., 2011), et d'autres comme SHALOM (Feingersh et Dor, 2015), SBG (héritière de HyspIRI (Lee et al., 2015)), CHIME (Rast et al., 2019) ou Biodiversity (héritière de hyPXIM/HYPEX-2 (Briottet et al., 2017 ; Carrere et al., 2013)) sont en préparation, préfigurant une nouvelle ère pour la télédétection hyperspectrale de la végétation. Alors que la plupart des capteurs satellitaires devrait avoir un GSD de 30 m, SHALOM et Biodiversity devraient eux avoir un GSD de 10 et 8 m, respectivement, ce qui permettrait de mieux identifier la structure de la canopée, en particulier pour les forêts ouvertes, au prix d'un *Signal-to-Noise Ratio* (SNR) plus faible. Évaluer les performances d'estimation de traits de végétation à partir d'images acquises par des capteurs satellitaires ayant $\sim$ 10 m et $\sim$ 30 m de GSD au dessus de canopées ouvertes et fermées permettrait de mieux comprendre comment ces capteurs se complètent, et d'identifier leur principales limitations.

Le but de la présente étude était d'évaluer les précisions d'estimation pouvant être attendues à partir d'images hyperspectrales acquises par SBG et Biodiversity au dessus de canopées hétérogènes comme celles présentes dans les écosystèmes méditerranéens, et l'évolution de ces estimations avec l'ouverture de la canopée. Pour ce faire, des images synthétiques SBG et Biodiversity ont été générées à partir d'images AVIRIS-C et AVIRIS-NG acquises au dessus de deux forêts ouvertes (voit la Figure 5.1). Des modèles de PLSR ont été entrainés sur les sorties du RTM DART pour estimer des traits de végétation (C_{ab}, Car, EWT, LMA) et, si possible, le gap fraction. La précision des estimations a été obtenue par comparaison avec les valeurs mesurées *in situ*, et les estimations au dessus de zones des sites ayant différents CC ont été intercomparées avec l'aide de cartes de CC obtenues à partir des images AVIRIS-NG.

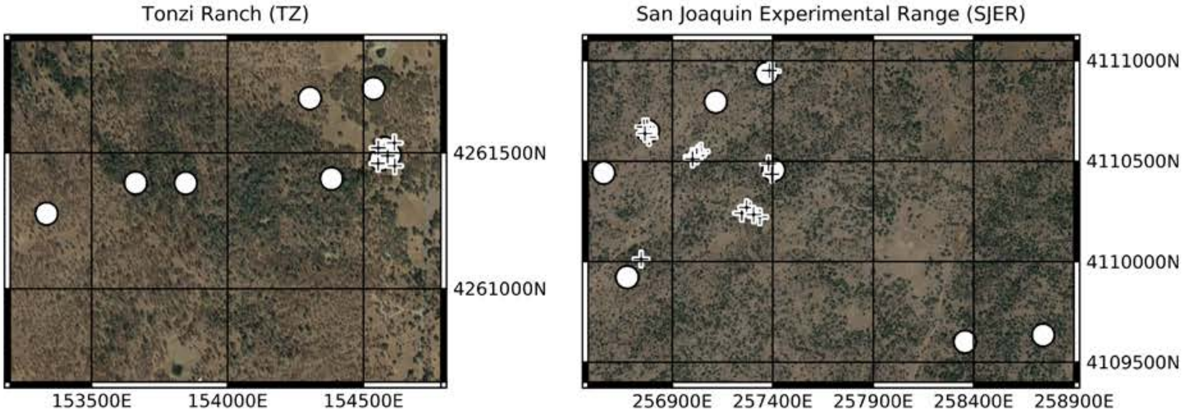


FIGURE 5.1 – Sites d'étude et localisation des mesures *in situ*. Les disques correspondent aux emplacements des acquisitions de DHP, les croix aux arbres retenus pour les mesures foliaires.

5.2 Principaux résultats

Il est apparu que, pour les parties les plus fermées de la canopée, un modèle de PLSR entrainé sur des données synthétiques était capable d'estimer C_{ab} et Car avec une grande précision aussi bien à partir d'images AVIRIS-C que SBG. De plus, les estimations de gap fraction étaient bonnes quel que soit le CC, et une relation de proportionalité entre gap fraction et CC a sans suprise été trouvée. Les bons résultats pour l'estimation du gap fraction étaient particulièrement encourageants car la plage d'entrainement couvrait uniquement 19–55% : des valeurs hors de cette plages ont été correctement estimées, suggérant une bonne généralisabilité. Il est apparu que les images AVIRIS-C et SBG étaient équivalentes concernant les estimations au dessus de TZ : la résolution spatiale légèrement plus grossière (30 m au lieu de 18 m) n'a pas semblé affecter significativement la précision d'estimation, ce qui est encourageant. Il est probable que les méthodes d'estimation développées pour AVIRIS-C pourraient être directement transposables à SBG. La Figure 5.2 montre les estimations obtenues à partir des images SBG pour les différents traits de végétation.

Pour les zones où la canopée était fermée, les images Biodiversity et SBG ont conduit à des estimations de C_{ab} et Car similaires. Pour ces parties, l'influence du sous-bois est minimale ($CC \geq 80\%$), et les estimations à partir des images SBG étaient bonnes. Cette corrélation entre estimations Biodiversity et SBG confirmait que des estimations précises sont également correctes à partir d'images Biodiversity. De plus, la majorité des pixels de végétation des images Biodiversity résolvaient presque les couronnes d'arbres, et Biodiversity n'était *a priori* que très peu affecté par l'ouverture de la canopée. Le fait que les estimations issues de Biodiversity et SBG n'étaient que peu corrélées au dessus des zones de faible CC (voir la Figure 5.3) illustrait l'avantage d'une résolution spatiale plus fine pour l'imagerie et le suivi des traits de végétation au dessus des canopées ouvertes. Il est ressorti que, grâce à sa meilleure résolution spatiale, Biodiversity présente un fort potentiel pour améliorer l'estimation des

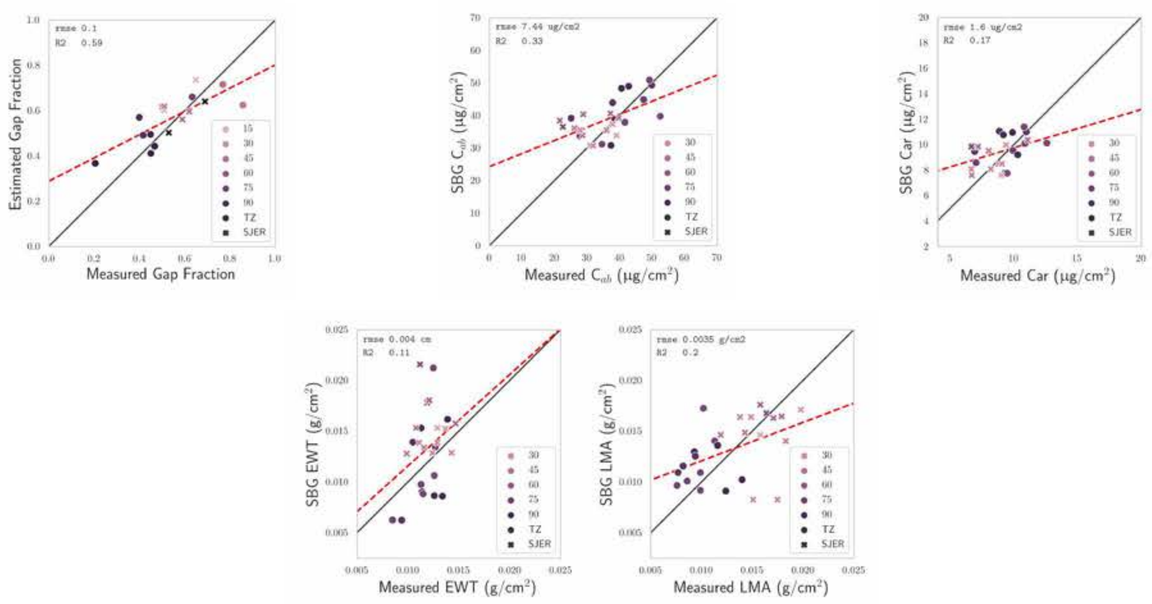
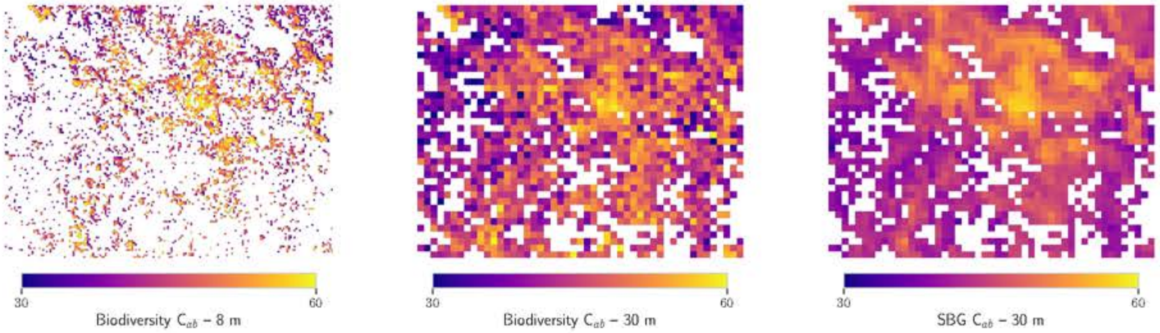


FIGURE 5.2 – Comparaison entre les traits de végétation estimés par la PLSR et les mesures *in situ* avec les images SBG. La coloration des marqueurs dépend du CC des pixels considérés. Les marqueurs noirs correspondent aux zones où l'information sur le CC n'était pas disponible.

traits de végétation, soit directement à partir des acquisitions du capteur hyperspectral, ou indirectement via fusion des données avec celles issues d'autres capteurs comme celui de SBG, qui auraient l'avantage d'avoir un SNR plus élevé. Il est également apparu que les estimations de EWT et LMA, si elles n'étaient que peu corrélées aux mesures *in situ*, étaient cohérentes entre les capteurs SBG et Biodiversity au dessus des zones de fort CC : cela pourrait indiquer que EWT et LMA peuvent effectivement être estimés avec succès, en dépit de ce que la comparaison aux mesures *in situ* semblait indiquer.

5.3 Conclusion

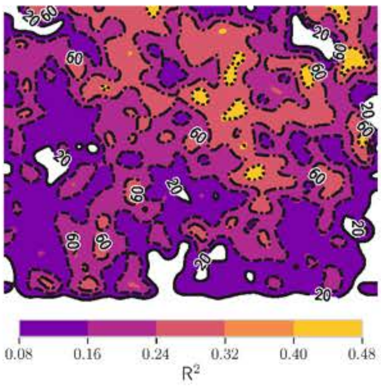
Cette étude visait à estimer le potentiel des futures missions hyperspectrales SBG et Biodiversity pour l'estimation de traits de végétation d'arbres feuillus de forêts éparses à partir d'une méthode hybride. Les estimations de C_{ab}, Car et gap fraction issues d'images SBG et AVIRIS-C étaient cohérentes non seulement avec les mesures *in situ*, mais également entre elles à l'échelle du paysage. Les estimations issues des images Biodiversity étaient également cohérentes avec les estimations SBG pour les zones de canopée fermée, donnant confiance en leur précision, mais exhibaient un comportement différent pour les zones de canopée ouverte. Ces travaux démontrent (i) que C_{ab}, Car et gap fraction peuvent être estimés à partir d'images satellitaires ayant un GSD de 30 m pour les canopées fermées avec peu de connaissances *a priori* du paysage imagé et (ii) que des images ayant un résolution spatiale de $\sim$ 10 m ont un fort potentiel pour le suivi des traits de végétation considérés dans cette étude pour tout type de canopée. Malgré leur plus faible SNR, des capteur satellitaires ayant un GSD similaire



(a) Estimations de C_{ab} à partir d'une image Biodiversity

(b) Comme la Figure 5.3a, mais aggrégée à 30 m de résolution

(c) Estimations de C_{ab} à partir d'une image SBG



(d) Corrélation entre les estimations à partir de Biodiversity et SBG pour différentes zones de CC. Les lignes indiquent les iso-CC.

FIGURE 5.3 – Comparaison entre les estimations de C_{ab} obtenues à partir d'images Biodiversity et SBG sur TZ

pourraient être très utile pour le suivi global des traits de végétation, en complémentant les capteurs avec un GSD plus élevé.

Conclusion et perspectives

Conclusion générale

Cette thèse avait pour objectif d'identifier des méthodes pouvant servir à l'estimation de différents traits de végétation à partir d'images hyperspectrales acquises au dessus de canopée ouvertes. Dans le contexte des récentes et futures missions hyperspectrales satellitaires. Ces méthodes devaient se servir le moins possible de connaissances *a priori* du terrain. Cette thèse s'est concentrée sur deux savannes boisées dominées par des feuillus, où des données *in situ* et des acquisitions aéroportées hyperspectrales ont été effectuées régulièrement sur plusieurs années.

Tout d'abord, les travaux se sont concentrés sur des images aéroportées acquises à une résolution spatiale de 18 m. Une stratégie d'inversion permettant d'estimer le LAI et les pigments foliaires (C_{ab} et Car) a été identifiée. Une méthode physique d'inversion a été utilisée, et il a été montré qu'une modélisation simpliste des arbres au sein d'un logiciel de tranfert radiatif 3D, avec des couronnes ellipsoidales et un sol plat lambertien, était suffisante pour atteindre des précisions satisfaisantes. Ensuite, l'étude a porté sur l'estimation de EWT et LMA : il a initialement été question de raffiner le pas d'échantillonage des LUT utilisées à moindre coût, en mettant en place un émulateur de DART. Il s'est avéré que si les estimations de la fraction de trous et des pigments foliaires étaient améliorées, EWT et LMA ne pouvaient toujours pas être estimés. Ensuite, des images à 2 m de résolution spatiales ont été utilisées pour étudier l'influence de la modélisation du couvert au sein de DART sur les estimations de EWT et LMA. Il est apparu que l'intervalle spectral 1.5–2.4 µm était le plus adapté pour les estimations de EWT et LMA, car il était moins sensible aux abstractions structurelles faites lors de la modélisation au sein du RTM. Enfin, la précision des estimations des traits de végétation (fraction de trous, C_{ab}, Car, EWT, LMA) à partir d'images synthétiques SBG (GSD : 30 m) et Biodiversity (GSD : 8 m) a été évaluée. Il a été montré qu'une méthode hybride, entrainant un PLSR sur les sorties de DART, permettait de réaliser ces estimations pour les deux imageurs. Cette étude a montré que des images de SBG et Biodiversity permettaient d'obtenir des estimations précises au dessus des canopées fermées, et que Biodiversity pourrait également servir à l'estimation des traits de végétation de canopées très ouvertes comme les savannes boisées.

Cette thèse a démontré que des méthodes physiques et hybrides étaient adéquates pour l'estimation de plusieurs traits de végétation de canopée ouvertes de feuilles à partir d'images hyperspectrales ayant des résolutions de 8 à 30 m dans un contexte opérationnel. Ces méthodes ne nécessitaient que peu de connaissances sur les ecosystèmes imagés, et n'utilisaient qu'une représentation abstraite de la forêt au sein d'un RTM 3D (couronne ellipsoidales et tronc cylindrique), avec une réflectance du sol directement extraite des images hyperspectrales.

Perspectives

Une des principales limitations de la résolution spatiale de 30 m de SBG était l'impossibilité d'isoler les différents éléments des scènes. En particulier, bien que des pixels de sol pur aient toujours pu être identifiés, leur usage dans les scènes DART reposait sur l'hypothèse que leur réflectance était représentative du sous-bois, ce qui pouvait ne pas être le cas du fait de la forte hétérogénéité spatiale. Cela pourrait être problématique dans les cas où l'hypothèse est inexacte : le sous-bois peut avoir un effet significatif sur la réflectance des pixels, en particulier pour les canopées ouvertes (Darvishzadeh et al., 2008). Il serait possible de considérer des méthodes de démélange spectral pour identifier les différents éléments présents dans les pixels, de façon à effectuer une classification préalable des différentes familles de pixels et de générer des bases de données d'entrainement adaptées à chacune.

De plus, il a été montré que SBG et Biodiversity pouvaient se compléter dans un contexte opérationnel. D'une part, Biodiversity, avec ses temps de revisite plus fréquents et une meilleur résolution spatiale, pourrait aider à la calibration de méthodes de classification dédiées aux images SBG, qui ont une couverture spatiale plus étendue. De plus, il pourrait être possible d'identifier les zones où les estimations effectuées à partir des images SBG sont incertaines du fait d'éléments non résolus à 30 m, en s'appuyant sur ses images Biodiversity. D'autre part, les estimations issues des images Biodiversity et SBG sont cohérentes sur les zones homogènes, bien que des biais puissent être présents entre elles pour les traits de végétation ayant de faibles signatures spectrales (caroténoïdes, EWT, LMA). Du fait de leur SNR plus élevé, il est probable que les images SBG soient plus adaptées à l'estimation de ces traits de végétation : il serait donc envisageable de calibrer les méthodes d'estimation dédiées à Biodiversity sur les estimations obtenues par SBG sur les couverts homogènes, avant de les appliquer sur les couvertes hétérogènes, qui peuvent être mieux étudiés à partir des images Biodiversity. D'une façon plus générale, ces synergies devraient également être présentes avec SHALOM, EnMAP, PRISMA ou CHIME.

Afin de complètement valider les résultats obtenus durant cette thèse, il est nécessaire d'évaluer les performances des méthodes proposées pour d'autres environnements, afin de considérer une plus grande diversité biologique, notamment en terme de traits de feuilles. Relativement peu de mesures *in situ* étaient disponibles pour cette thèse, couvrant seulement un intervalle limité des traits de feuilles, en particulier pour EWT et LMA. Ces trois traits de feuilles mériteraient une étude plus approfondie, car leur estimation depuis des capteurs satellitaires serait très intéressante pour mieux comprendre la réponse de la végétation aux différents stress (Schimel et al., 2015 ; Westoby et al., 2004). De plus, pouvoir estimer précisément le FMC, qui combine LMA et EWT, serait extrêmement bénéfique dans un contexte d'augmentation de la fréquence des incendies.

Les méthodes développées durant cette thèses étaient dédiées aux canopées composées principalement de feuillus, qui étaient majoritaires sur les sites d'étude. En effet, DART et PROSPECT sont pleinement adaptés pour modéliser les feuilles, mais requerraient de nombreux ajustements pour pouvoir reproduire fidèlement le comportement radiatif des conifères. Le développement de RTM adaptés aux conifères aiderait grandement la mise en place de méthodes d'estimation permettant de se passer de données terrain. Les forêts de conifères représentent une grande portion des canopées : les forêts de conifères constituent 16% des forêts Européennes, et l'essentiel des forêts boréales. Identifier des méthodes d'estimation leur étant adaptées est donc critique pour le suivi des traits de végétation à l'échelle globale. Enfin, les écosystèmes présent sur Tonzi et San Joaquin avaient durant l'été et l'automne des canopées et des sous-bois clairement distincts, l'herbe étant sèche à ces périodes. Dans une configuration où le sous-bois serait photosynthétiquement actif, le mélange spectral des contributions de l'herbe et des couronnes d'arbres serait d'autant plus problématique si la résolution spatiale des images ne permettait pas d'isoler les arbres. En effet, les réflectances de l'herbe et de la canopée seraient extrêmement similaires, menant à des incertitudes bien plus élevées pour les estimations. Développer une méthode adaptée à ces conditions permettrait d'étendre le suivi des canopées ouvertes à l'intégralité du cycle phénologique annuel.

Bibliographie

Ali, Abebe Mohammed et al. (2016a). « Effects of Canopy Structural Variables on Retrieval of Leaf Dry Matter Content and Specific Leaf Area from Remotely Sensed Data ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cf. p. 21, 25).

Ali, Abebe Mohammed et al. (2016b). « Retrieval of forest leaf functional traits from HySpex imagery using radiative transfer models and continuous wavelet analysis ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* (cf. p. 17, 21).

Arendonk, J. J. C. M. et H. Poorter (1994). « The chemical composition and anatomical structure of leaves of grass species differing in relative growth rate ». In : *Plant, Cell and Environment* 17.8, p. 963-970 (cf. p. 7).

Asner, Gregory P. et al. (2011a). « Spectroscopy of canopy chemicals in humid tropical forests ». In : *Remote Sensing of Environment* 115.12, p. 3587-3598 (cf. p. 14).

Asner, Gregory P et al. (2011b). « Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests ». In : *Ecological Applications* 21.1, p. 85-98 (cf. p. 2).

Asner, Gregory P. et al. (2015). « Quantifying forest canopy traits : Imaging spectroscopy versus field survey ». In : *Remote Sensing of Environment* 158, p. 15-27 (cf. p. 14).

Atzberger, Clement (2000). « Development of an invertible forest reflectance model The INFORM-Model ». In : *A decade of trans-European remote sensing cooperation, Proceedings of the 20th EARSeL Symposium*, p. 39-44 (cf. p. 11).

Atzberger, Clement et Katja Richter (2012). « Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery ». In : *Remote Sensing of Environment* 120, p. 208-218 (cf. p. 14).

Axelsson, Christoffer et al. (2013). « Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression ». In : *International Journal of Remote Sensing* 34.5, p. 1724-1743 (cf. p. 14).

Banskota, Asim et al. (2015). « An LUT-Based Inversion of DART Model to Estimate Forest LAI from Hyperspectral Data ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cf. p. 17).

Barclay, H. J. (2001). « Distribution of leaf orientations in six conifer species ». In : *Canadian Journal of Botany* 79.4, p. 389-397 (cf. p. 8).

Barnes, William L., X. Jack Xiong et Vincent V Salomonson (2002). « Status of Terra MODIS and Aqua MODIS ». In : *International Geoscience and Remote Sensing Symposium (IGARSS).* T. 2, p. 970-972 (cf. p. 2).

Berger, Katja et al. (2018). *Evaluation of the PROSAIL model capabilities for future hyperspectral model environments : A review study* (cf. p. 3, 11, 15).

Brede, Benjamin et al. (2020). « Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, landsat 7 ETM+ and Landsat 8 OLI ». In : *Remote Sensing* 12.6 (cf. p. 26).

Briottet, X et al. (2017). « European Hyperspectral Explorer : HYPEX-2—a new space mission for vegetation biodiversity, bare continental surfaces, coastal zones and urban area ecosystems ». In : *Abstract 10th EARSeL SIG imaging spectroscopy workshop*, p. 19-21 (cf. p. 29).

Carrere, Veronique et al. (2013). « HYPXIM : A second generation high spatial resolution hyperspectral satellite for dual applications ». In : *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing.* T. 2013-June. IEEE Computer Society (cf. p. 3, 29).

Cavender-Bares, Jeannine et al. (2017). « Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales ». In : *American Journal of Botany* 104.7, p. 966-969 (cf. p. 2).

Chalker-Scott, Linda (1999). *Environmental significance of anthocyanins in plant stress responses* (cf. p. 6).

Colombo, R. et al. (2008). « Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling ». In : *Remote Sensing of Environment* (cf. p. 11).

Combal, Bruno et al. (2003). « Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem ». In : *Remote Sensing of Environment* (cf. p. 14).

Cowling, Richard M et al. (1996). « Plant diversity in mediterranean-climate regions ». In : *Trends in Ecology & Evolution* 11.9, p. 362-366 (cf. p. 1).

Dana Chadwick, K. et Gregory P. Asner (2016). « Organismic-scale remote sensing of canopy foliar traits in lowland tropical forests ». In : *Remote Sensing* 8.2 (cf. p. 2, 14).

Danson, F. Mark et al. (2007). « Forest canopy gap fraction from terrestrial laser scanning ». In : *IEEE Geoscience and Remote Sensing Letters* 4.1, p. 157-160 (cf. p. 8).

Darvishzadeh, Roshanak, Ali A. Matkan et Abdolhamid Dashti Ahangar (2012). « Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (cf. p. 11, 17).

Darvishzadeh, Roshanak et al. (2008). « Estimation of vegetation LAI from hyperspectral reflectance data : Effects of soil type and plant architecture ». In : *International Journal of Applied Earth Observation and Geoinformation* 10.3, p. 358-373 (cf. p. 36).

Darvishzadeh, Roshanak et al. (2019). « Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model ». In : *International Journal of Applied Earth Observation and Geoinformation* 79, p. 58-70 (cf. p. 12, 22).

Dawson, Terence P, Paul J Curran et Stephen E Plummer (1998). « LIBERTY - Modeling the effects of Leaf Biochemical Concentration on Reflectance Spectra ». In : *Remote Sensing of Environment* 65.1, p. 50-60 (cf. p. 10).

Delegido, Jesús et al. (2010). « Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC) ». In : *International Journal of Applied Earth Observation and Geoinformation* 12.3, p. 165-174 (cf. p. 19).

Delegido, Jesús et al. (2014). « Chlorophyll content mapping of urban vegetation in the city of Valencia based on the hyperspectral NAOC index ». In : *Ecological Indicators* 40, p. 34-42 (cf. p. 11).

Demarez, V. et al. (1999). « Seasonal variation of leaf chlorophyll content of a temperate forest. Inversion of the prospect model ». In : *International Journal of Remote Sensing* 20.5, p. 879-894 (cf. p. 19).

Demmig-Adams, Barbara et William W Adams (1996). *The role of xanthophyll cycle carotenoids in the protection of photosynthesis* (cf. p. 6).

Fath, Brian (2018). *Encyclopedia of ecology.* T. 105. 3, p. 1-663 (cf. p. 1).

Feingersh, Tal et Eyal Ben Dor (2015). « SHALOM - A Commercial Hyperspectral Space Mission ». In : *Optical Payloads for Space Missions.* Chichester, UK : John Wiley & Sons, Ltd, p. 247-263 (cf. p. 29).

Féret, J. B. et al. (2017). « PROSPECT-D : Towards modeling leaf optical properties through a complete lifecycle ». In : *Remote Sensing of Environment* (cf. p. 10).

Feret, Jean Baptiste et al. (2008). « PROSPECT-4 and 5 : Advances in the leaf optical properties model separating photosynthetic pigments ». In : *Remote Sensing of Environment* (cf. p. 10).

Féret, Jean Baptiste et al. (2021). « PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents ». In : *Remote Sensing of Environment* 252 (cf. p. 8, 10).

Ferreira, Matheus Pinheiro et al. (2018). « Retrieving structural and chemical properties of individual tree crowns in a highly diverse tropical forest with 3D radiative transfer modeling and imaging spectroscopy ». In : *Remote Sensing of Environment* (cf. p. 26).

Flint, Stephan D, Peter W Jordan et Martyn M Caldwell (1985). « Plant Protective Response to Enhanced UV-B Radiation Under Field Conditions : Leaf Optical Properties and Photosynthesis ». In : *Photochemistry and Photobiology* 41.1, p. 95-99 (cf. p. 6).

Frampton, William James et al. (2013). « Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* 82, p. 83-92 (cf. p. 2).

Füssel, Hans-Martin et al. (2012). *Climate change, impacts and vulnerability in Europe 2012 : an indicator-based report.* 12, p. 304 (cf. p. 1).

Gao, Bo-Cai (1996). « NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space ». In : *Remote Sensing of Environment* 7212, p. 257-266 (cf. p. 13).

Gara, Tawanda W. et al. (2019). « Evaluating the performance of PROSPECT in the retrieval of leaf traits across canopy throughout the growing season ». In : *International Journal of Applied Earth Observation and Geoinformation* 83, p. 101919 (cf. p. 19).

Gascon, F. et al. (2004). « Retrieval of forest biophysical variables by inverting a 3-D radiative transfer model and using high and very high resolution imagery ». In : *International Journal of Remote Sensing* (cf. p. 17).

Gastellu-Etchegorry, Jean Philippe et al. (2015). « Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes ». In : *Remote Sensing* (cf. p. 11, 12).

Giorgi, Filippo et Piero Lionello (2008). « Climate change projections for the Mediterranean region ». In : *Global and Planetary Change* 63.2-3, p. 90-104 (cf. p. 1).

Gitelson, Anatoly A et al. (2002). « Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy ». In : *Photochemistry and Photobiology* 75.3, p. 272 (cf. p. 13).

Govaerts, Yves M et Michel M Verstraete (1998). « Raytran : a monte carlo raytracing model to compute light scattering in threedimensional heterogeneous media ». In : *IEEE Transactions on Geoscience and Remote Sensing* 36.2, p. 493505 (cf. p. 11).

Graziani, Frank et al. (2006). *Computational Methods in Transport* (cf. p. 11).

Guanter, Luis et al. (2015). *The EnMAP spaceborne imaging spectroscopy mission for earth observation* (cf. p. 3, 29).

Haboudane, Driss et al. (2002). « Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture ». In : *Remote Sensing of Environment*. arXiv : `NIHMS150003` (cf. p. 11, 13).

Hanan, Niall et Michael J Hill (2012). « Savannas in a Changing Earth System : The NASA Terrestrial Ecology Tree-Grass Project ». In : p. 56 (cf. p. 21).

Hernández-Clemente, Rocío, Rafael M. Navarro-Cerrillo et Pablo J. Zarco-Tejada (2012). « Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT+DART simulations ». In : *Remote Sensing of Environment* (cf. p. 12, 17, 21).

Hill, Joachim, Henning Buddenbaum et Philip A. Townsend (2019). « Imaging Spectroscopy of Forest Ecosystems : Perspectives for the Use of Space-borne Hyperspectral Earth Observation Systems ». In : *Surveys in Geophysics* 40.3, p. 553-588 (cf. p. 3).

Iwasaki, Akira et al. (2011). « Hyperspectral Imager Suite (HISUI)-Japanese hyper-multi spectral radiometer ». In : *International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 1025-1028 (cf. p. 29).

Jacquemoud, S et F Baret (1990). « PROSPECT : A model of leaf optical properties spectra ». In : *Remote Sensing of Environment* 34.2, p. 75-91 (cf. p. 10).

Jacquemoud, Stéphane et Susan Ustin (2019). *Leaf optical properties*. Cambridge University Press, p. 1-556 (cf. p. 7).

Jacquemoud, Stéphane et al. (2009). « PROSPECT + SAIL models : A review of use for vegetation characterization ». In : *Remote Sensing of Environment* 113.SUPPL. 1 (cf. p. 11).

Janoutová, Růžena et al. (2019). « Influence of 3D spruce tree representation on accuracy of airborne and satellite forest reflectance simulated in DART ». In : *Forests* 10.3 (cf. p. 12, 17, 25).

Kobayashi, Hideki et al. (2012). « Modeling energy and carbon fluxes in a heterogeneous oak woodland : A three-dimensional approach ». In : *Agricultural and Forest Meteorology* 152.1, p. 83-100 (cf. p. 12).

Krause, G H et E Weis (1991). « Chlorophyll Fluorescence and Photosynthesis : The Basics ». In : *Annual Review of Plant Physiology and Plant Molecular Biology* 42.1, p. 313-349 (cf. p. 5).

Kuusk, Andres (2017). « Canopy radiative transfer modeling ». In : *Comprehensive Remote Sensing*. T. 1-9. Elsevier, p. 9-22 (cf. p. 11).

Lambers, Hans, F. Stuart Chapin et Thijs L. Pons (2008). *Plant physiological ecology : Second edition*. Springer New York, p. 1-604 (cf. p. 7).

Lee, Christine M. et al. (2015). « An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities ». In : *Remote Sensing of Environment* 167, p. 6-19 (cf. p. 3, 29).

LEWIS, Philip (1999). « Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System ». In : *Agronomie* 19.3-4, p. 185-210 (cf. p. 11).

LI, L. et al. (2008). « Retrieval of vegetation equivalent water thickness from reflectance using genetic algorithm (GA)-partial least squares (PLS) regression ». In : *Advances in Space Research* (cf. p. 11).

LIONELLO, Piero et al. (2014). *The climate of the Mediterranean region : research progress and climate change impacts* (cf. p. 1).

LIU, Jing et al. (2019). « Variation of leaf angle distribution quantified by terrestrial LiDAR in natural European beech forest ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* 148, p. 208-220 (cf. p. 9).

MAIRE, Guerric le et al. (2008). « Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass ». In : *Remote Sensing of Environment* (cf. p. 11, 13, 17, 22, 23).

MALENOVSKY, Z. et al. (2007). « Retrieval of coniferous canopy chlorophyll content from high spatial resolution hyperspectral data ». In : *Proceedings of the 10th International Symposium on Physical Measurements and Spectral Signatures in Remote Sensing (ISPM-SRS'07).* T. XXXVI, p. 108-113 (cf. p. 12).

MALENOVSKÝ, Z. et al. (2008a). « Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution ». In : *Remote Sensing of Environment* 112.1, p. 1-18 (cf. p. 9, 18).

MALENOVSKÝ, Zbyněk et al. (2008b). « Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution ». In : *Remote Sensing of Environment* (cf. p. 12, 26).

MARKHAM, Brian L. et al. (2004). *Landsat sensor performance : History and current status* (cf. p. 2).

MARTENS, Scott N, Susan L USTIN et Robert A ROUSSEAU (1993). *Estimation of tree canopy leaf area index by gap fraction analysis.* Rapp. tech., p. 91-108 (cf. p. 8).

MEDAIL, Frederic et Pierre QUEZEL (1997). « Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean Basin ». In : *Annals of the Missouri Botanical Garden* 84.1, p. 112 (cf. p. 1).

MELENDO-VEGA, José Ramón et al. (2018). « Improving the performance of 3-D radiative transfer model FLIGHT to simulate optical properties of a tree-grass ecosystem ». In : *Remote Sensing* (cf. p. 12, 17).

MERZLYAK, Mark N. et al. (2008). « Light absorption by anthocyanins in juvenile, stressed, and senescing leaves ». In : *Journal of Experimental Botany* 59.14, p. 3903-3911 (cf. p. 6).

MIGITA, C., Y. CHIBA et T. TANGE (2007). « Seasonal and spatial variations in leaf nitrogen content and resorption in a Quercus serrata canopy ». In : *Tree Physiology* 27.1, p. 63-70 (cf. p. 6).

MISSON, L et al. (2007). « Partitioning forest carbon fluxes with overstory and understory eddy-covariance measurements : A synthesis based on FLUXNET data ». In : *Agricultural and Forest Meteorology* 144.2, p. 14-31 (cf. p. 9).

Moorthy, Inian, John R. Miller et Thomas L. Noland (2008). « Estimating chlorophyll concentration in conifer needles with hyperspectral data : An assessment at the needle and canopy level ». In : *Remote Sensing of Environment* 112.6, p. 2824-2838 (cf. p. 10).

Myers, Norman et al. (2000). « Biodiversity hotspots for conservation priorities ». In : *Nature* 403.6772, p. 853-858 (cf. p. 1).

Nilson, Tiit et Andres Kuusk (2004). « Improved algorithm for estimating canopy indices from gap fraction data in forest canopies ». In : *Agricultural and Forest Meteorology* 124.3-4, p. 157-169 (cf. p. 8).

North, Peter R J (1996). *Three-Dimensional Forest Light Interaction Model Using a Monte Carlo*. Rapp. tech. 4 (cf. p. 11).

Pearlman, Jay S. et al. (2003). « Hyperion, a space-based imaging spectrometer ». In : *IEEE Transactions on Geoscience and Remote Sensing* 41.6 PART I, p. 1160-1173 (cf. p. 3).

Pereira, H. M. et al. (2013). « Essential biodiversity variables ». In : *Science* 339.6117, p. 277-278 (cf. p. 1).

Pettorelli, Nathalie et al. (2016). « Framing the concept of satellite remote sensing essential biodiversity variables : challenges and future directions ». In : *Remote Sensing in Ecology and Conservation* 2.3, p. 122-131 (cf. p. 2, 29).

Pinty, B. et al. (2004). « Radiation Transfer Model Intercomparison (RAMI) exercise : Results from the second phase ». In : *Journal of Geophysical Research D : Atmospheres* 109.6 (cf. p. 11).

Pinty, Bernard et al. (2001). « Radiation transfer model intercomparison (RAMI) exercise ». In : *Journal of Geophysical Research Atmospheres* 106.D11, p. 11937-11956 (cf. p. 11).

Pisek, Jan et al. (2013). « Is the spherical leaf inclination angle distribution a valid assumption for temperate and boreal broadleaf tree species ? » In : *Agricultural and Forest Meteorology* 169, p. 186-194 (cf. p. 8).

Rast, Michael et al. (2019). *Copernicus hyperspectral imaging mission for the environment : Mission requirements document*. English. 2.1. Mission Requirements Document (MRD) ESA-EOPSM-CHIM-MRD-3216. European Space Agency (ESA) (cf. p. 3, 29).

Richter, Katja et al. (2011). « Evaluation of Sentinel-2 Spectral Sampling for Radiative Transfer Model Based LAI Estimation of Wheat, Sugar Beet, and Maize ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.2, p. 458-464 (cf. p. 14).

Ritz, Thorsten et al. (2000). *Efficient light harvesting through carotenoids* (cf. p. 6).

Ryu, Youngryel et al. (2012). « Continuous observation of tree leaf area index at ecosystem scale using upward-pointing digital cameras ». In : *Remote Sensing of Environment* (cf. p. 9).

Sala, Osvaldo E. et al. (2000). « Global biodiversity scenarios for the year 2100 ». In : *Science* 287.5459, p. 1770-1774 (cf. p. 1).

San, B. Taner et M. Lütfi Süzen (2011). « Evaluation of cross-track illumination in EO-1 hyperion imagery for lithological mapping ». In : *International Journal of Remote Sensing* 32.22, p. 7873-7889 (cf. p. 7).

Sanger, Jon E. (1971). « Quantitative Investigations of Leaf Pigments From Their Inception in Buds Through Autumn Coloration to Decomposition in Falling Leaves ». In : *Ecology* 52.6, p. 1075-1089 (cf. p. 6).

Schaepman, Michael E. et al. (2015). « Advanced radiometry measurements and Earth science applicationswith the Airborne Prism Experiment (APEX) ». In : *Remote Sensing of Environment* 158.1, p. 207-219 (cf. p. 2).

Schimel, David et al. (2015). *Observing terrestrial ecosystems and the carbon cycle from space* (cf. p. 36).

Shaw, Gary A et Hsiao-hua K Burke (2003). *Spectral Imaging for Remote Sensing.* Rapp. tech. (cf. p. 2).

Somot, S. et al. (2008). « 21st century climate change scenario for the Mediterranean using a coupled atmosphere-ocean regional climate model ». In : *Global and Planetary Change* 63.2-3, p. 112-126 (cf. p. 1).

Stefano, Pignatti et al. (2013). « The PRISMA hyperspectral mission : Science activities and opportunities for agriculture and land monitoring ». In : *International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 4558-4561 (cf. p. 3, 29).

Thenkabail, Prasad S., John G. Lyon et Alfredo Huete, éd. (2018). *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation.* CRC Press (cf. p. 7).

Underwood, Emma C. et al. (2009). « Threats and biodiversity in the mediterranean biome ». In : *Diversity and Distributions* (cf. p. 1).

Ustin, Susan L. et John A. Gamon (2010). *Remote sensing of plant functional types* (cf. p. 2).

Vane, Gregg et al. (1993). « The airborne visible/infrared imaging spectrometer (AVIRIS) ». In : *Remote Sensing of Environment* 44.2-3, p. 127-143 (cf. p. 2).

Vautard, Robert et al. (2007). « Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit ». In : *Geophysical Research Letters* 34.7, p. 1-5 (cf. p. 1).

Verhoef, W (1984). « Light Scattering by Leaf layers with Application to Canopy Reflectance Modeling : the SAIL Model ». In : *Remote Sensing Of Environment* 141, p. 125-141 (cf. p. 11).

Verrelst, J. et al. (2012). « Optimizing LUT-based radiative transfer model inversion for retrieval of biophysical parameters using hyperspectral data ». In : *International Geoscience and Remote Sensing Symposium (IGARSS)* (cf. p. 14).

Verrelst, Jochem et al. (2010). « Effects of woody elements on simulated canopy reflectance : Implications for forest chlorophyll content retrieval ». In : *Remote Sensing of Environment* (cf. p. 9, 12, 18).

Verrelst, Jochem et al. (2015). « Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* 108, p. 273-290 (cf. p. 13).

Wang, W. M., Z. L. Li et H. B. Su (2007). « Comparison of leaf angle distribution functions : Effects on extinction coefficient and fraction of sunlit foliage ». In : *Agricultural and Forest Meteorology* 143.1-2, p. 106-122 (cf. p. 8).

Weiss, Marie et al. (2000). « Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data ». In : *Agronomie* (cf. p. 11, 14, 21).

Welles, Jon M et Shabtai Cohen (1996). *Canopy structure measurement by gap fraction analysis using commercial instrumentation.* Rapp. tech. 302, p. 1335-1342 (cf. p. 8).

WESTOBY, Mark et al. (2004). « The worldwide leaf economics spectrum ». In : *Nature* 428, p. 821-827 (cf. p. 7, 36).

WIDLOWSKI, J. L. et al. (2013). « The fourth radiation transfer model intercomparison (RAMI-IV) : Proficiency testing of canopy reflectance models with ISO-13528 ». In : *Journal of Geophysical Research Atmospheres* 118.13, p. 6869-6890 (cf. p. 11).

WIDLOWSKI, Jean Luc, Jean François CÔTÉ et Martin BÉLAND (2014). « Abstract tree crowns in 3D radiative transfer models : Impact on simulated open-canopy reflectances ». In : *Remote Sensing of Environment* (cf. p. 12, 17, 18, 25, 26).

WIDLOWSKI, Jean Luc et al. (2007). « Third Radiation Transfer Model Intercomparison (RAMI) exercise : Documenting progress in canopy reflectance models ». In : *Journal of Geophysical Research Atmospheres* 112.9 (cf. p. 11).

WIDLOWSKI, Jean Luc et al. (2015). « The fourth phase of the radiative transfer model intercomparison (RAMI) exercise : Actual canopy scenarios and conformity testing ». In : *Remote Sensing of Environment* 169, p. 418-437 (cf. p. 11).

WIENEKE, S. et al. (2016). « Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence : Implications for improved estimates of gross primary productivity ». In : *Remote Sensing of Environment* 184, p. 654-667 (cf. p. 2).

WOCHER, Matthias et al. (2018). « Physically-based retrieval of canopy equivalent water thickness using hyperspectral data ». In : *Remote Sensing* 10.12 (cf. p. 25).

WOODGATE, William et al. (2016). « Quantifying the impact of woody material on leaf area index estimation from hemispherical photography using 3D canopy simulations ». In : *Agricultural and Forest Meteorology* (cf. p. 9).

YANFANG XIAO et al. (2013). « Sensitivity Analysis of Vegetation Reflectance to Biochemical and Biophysical Variables at Leaf, Canopy, and Regional Scales ». In : *IEEE Transactions on Geoscience and Remote Sensing* (cf. p. 23, 26).

ZARCO-TEJADA, P. J. et al. (2004). « Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops ». In : *Remote Sensing of Environment* (cf. p. 17, 19, 22).

ZARCO-TEJADA, P. J. et al. (2013). « Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV) ». In : *Agricultural and Forest Meteorology* (cf. p. 12, 19, 22).

ZARCO-TEJADA, P. J. et al. (2019). « Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline ». In : *Remote Sensing of Environment* 223, p. 320-335 (cf. p. 17).

ZHANG, Yongqin et al. (2008). « Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery ». In : *Remote Sensing of Environment* 112.7, p. 3234-3247 (cf. p. 19).

Abstract — Due to increasing climatic and anthropic pressures, the Mediterranean ecoregions, that are recognized plant biodiversity hotspots, are some of the most endangered ecosystems. Airborne and satellite remote sensing methods, that can provide information about large landscapes in a regular fashion, are most adapted for a future global monitoring. However, difficulties arise when retrieving information from open forests, largely distributed in the Mediterranean-climate regions, as the contribution of tree crowns to the measured radiative signal is limited. This thesis aims at developing method for the estimation of vegetation traits of open canopies, when field knowledge is insufficient to calibrate regression models. Initially, 18 m GSD airborne hyperspectral images were considered. First, using the DART model, a simplified modeling, with a flat lambertian ground and ellipsoidal tree crowns, was identified as suitable for physically-based estimations of LAI and leaf pigment contents. Then, exploratory works were undertaken to identify a method to estimate EWT and LMA with acceptable accuracy, first by considering refinements in the sampling schemes, then by assessing the effects of the 3D modeling on trees' radiative behavior. Finally, the findings were used to estimate multiple vegetation traits (gap fraction, leaf chlorophyll and carotenoid contents, EWT, LMA) from synthetic hyperspectral satellite images with 8 and 30 m GSD using a hybrid method. This thesis demonstrated that physically-based and hybrid methods were adequate for the estimation of multiple canopy and leaf traits from hyperspectral satellite images of open canopies in an operational context, using little if any *a priori* knowledge. To consolidate the results, efforts are required to test the methods over more various sites that would present a higher diversity in terms of traits, trait variation ranges, or species. Moreover, identifying methods that would work for periods during which the understory is photosynthetically active would be necessary to allow for a global monitoring over the complete annual phenological cycle.

Keywords : Hyperspectral remote sensing ; Vegetation ; Radiative transfer model ; Hybrid method ; Satellite imagery

---

Résumé — Les pressions climatiques et anthropiques croissantes font des écorégions méditerranéennes, des réservoirs exceptionnels de biodiversité, certains des écosystèmes les plus menacés. Les méthodes de télédétection aéroportée et satellitaire, qui peuvent permettre d'obtenir des estimations sur de vastes étendues de façon régulière, sont particulièrement adaptées aux futurs efforts globaux de suivi de la biodiversité végétale. Cependant, de nombreuses difficultés surviennent quand les écosystèmes imagés sont composés de canopées ouvertes, particulièrement présentes dans les régions au climat méditerranéen. Cette thèse a pour objectif de développer des méthodes d'estimation de traits de végétation de canopées ouvertes, quand les connaissances du terrain sont insuffisantes pour directement calibrer des modèles de régression. Initialement, des images aéroportées acquises avec un GSD de 18 m ont été utilisées. En utilisant le modèle DART, une modélisation simplifiée des forêts, avec des couronnes ellipsoïdales et un sol plat lambertien, a été démontrée suffisante pour réaliser des estimations de LAI et de contenus en pigments foliaires par méthode physique. Ensuite, des travaux exploratoires ont été menés pour identifier une méthode permettant d'estimer EWT et LMA de façon satisfaisante, tout d'abord en considérant des raffinements dans les pas d'échantillonnage utilisés pour générer les bases de données, puis en évaluant l'influence de la modélisation en 3D au sein de DART sur les propriétés radiatives des arbres. Enfin, les différents résultats ont été utilisés pour estimer plusieurs traits de végétation (fraction de trous, contenus foliaires en chlorophylles et caroténoïdes, EWT, LMA) à partir d'images hyperspectrales satellitaires synthétiques ayant des GSD de 8 m et 30 m en utilisant une méthode hybride. Cette thèse a démontré que les méthodes physiques et hybrides sont adéquates pour l'estimation de plusieurs traits de végétation à partir d'images hyperspectrales satellitaires ayant des GSD de 8 à 30 m dans un contexte opérationnel, n'utilisant que peu ou pas de connaissances *a priori*. Pour consolider les résultats, de plus amples travaux sont nécessaires pour tester ces méthodes sur des écosystèmes qui présenteraient plus de diversité végétale, et pour d'autres traits de végétation. De plus, identifier des méthodes d'estimation pour les périodes où le sous-bois est photosynthétiquement actif serait nécessaire pour suivre les forêts tout au long du cycle phénologique annuel.

Mots clés : Télédétection hyperspectrale ; Végétation ; Modèle de transfert radiatif ; Méthode hybride ; Imagerie satellitaire

---

ONERA/DOTA, Université Fédérale de Toulouse
2 avenue Edouard Belin, 31055 Toulouse, France