The EnMAP Managed Vegetation Scientific Processor

Towards an Automatic Retrieval of Canopy Structure and Leaf Biochemical Properties from Hyperspectral Images

Dissertation zur Erlangung des Doktorgrades

an der Fakultät für Geowissenschaften

der Ludwig-Maximilians-Universität München

vorgelegt von

Martin Danner

München, 15.07.2019

Erstgutachter:	Prof. Dr. Wolfram Mauser
Zweitgutachter:	PD Dr. Tobias B. Hank
Tag der mündlichen Prüfung:	16. Dezember 2019

In freudiger und dankbarer Erinnerung an meinen Erdkundelehrer

Hans Schappert
† 20. Oktober 2016

Jetzt, da wir wissen, wo das Schweinchen im Pfeffer liegt, wollen wir dem Kind mal einen Namen geben …

Zusammenfassung

Nach jahrelanger wissenschaftlicher und technischer Vorbereitungszeit wird voraussichtlich Ende des Jahres 2020 der Start der orbitalen Phase einer unbemannten deutschen Weltraum-Mission initiiert. Das Environmental Mapping and Analysis Program (EnMAP) wird an Bord des gleichnamigen Satelliten einen hyperspektralen Sensor zur Erfassung terrestrischer Oberflächen tragen. In den Umweltdisziplinen zur Erforschung von Ökosystemen, landwirtschaftlicher, forstwirtschaftlicher und urbaner Flächen, im Bereich der Küsten- und Inlandsgewässer sowie der Geologie und Bodenkunde bereitete man sich im Vorfeld des Starts auf die kommenden Daten vor. Zwar existiert bereits eine Vielzahl an Algorithmen zur wissenschaftlichen Analyse von spektralen Daten, allerdings ergeben sich auch neue Herausforderungen, da die EnMAP-Mission bislang im weltweiten Kontext der Fernerkundung einzigartig ist. Die Abdeckung des vollen optischen Spektrums (420 nm – 2450 nm) in Verbindung mit einer moderaten räumlichen Auflösung von 30 m und einem hohen Signal-Rausch-Verhältnis von mindestens 180 im kurzwelligen Infrarot und über 400 im sichtbaren Spektrum, ermöglichen eine Aufnahmequalität, die bislang nur von flugzeuggestützten Systemen erreicht werden konnte.

Die Bemühungen in dieser Dissertation umfassen Aktivitäten in der wissenschaftlichen Vorbereitungsphase zu agrargeographischen Fragestellungen. Algorithmen und Tools zur Analyse der hyperspektralen Daten werden kostenlos im QGIS-Plugin EnMAP-Box 3 zur Verfügung gestellt. Die drängenden Fragen im Agrarsektor drehen sich hierbei um die Ableitung biochemischer und biophysikalischer Parameter aus Fernerkundungsdaten, weshalb die übergeordnete Problemstellung des Promotionsvorhabens die Entwicklung eines wissenschaftsbasierten EnMAP-Tools für bewirtschaftete Vegetationsflächen (EnMAP Managed Vegetation Scientific Processor) darstellt.

Zu Beginn wurde eine umfassende Feldkampagne geplant, welche ab April 2014 umgesetzt wurde. Neben der spektralen Erfassung von Blatt-, Bestands- und Bodensignaturen in einem Winterweizen- und einem Maisfeld erfolgte auch die Messung wesentlicher Pflanzenparameter an den exakt gleichen Positionen. Hierzu zählt die non-destruktive Ableitung des Blattflächenindex (LAI), des Blattchlorophyllgehalts (C_{cab}), des Blattwassergehalts (EWT oder C_w), des relativen Blatttrockengewichts (LMA oder C_m), des

mittleren Blattneigungswinkels im Bestand (ALIA) sowie weiterer sekundärer Parameter wie Wuchshöhe, das phänologisches Stadium und der Sonnenvektor. Um die Fähigkeit des späteren EnMAP-Satelliten sich um bis zu 30° orthogonal zur Flugrichtung zu kippen nachzustellen, wurden die spektralen Aufnahmen aus verschiedenen Betrachtungswinkeln erstellt, die dieser Aufnahme-Geometrien nachempfunden sind.

Ein gängiges Verfahren zur Ableitung der relevanten Pflanzenparameter ist die Verwendung des Strahlungstransfermodells PROSAIL, welches das spektrale Signal einer Vegetationsfläche auf Basis der zugrundeliegenden biophysikalischen und biochemischen Parameter simuliert. Bei der Umkehr dieses Prozesses können ebendiese Variablen von gemessenen spektralen Daten abgeleitet werden. Hierzu wurde eine Datenbank (Look-Up-Table, LUT) aus PROSAIL-Modellläufen aufgebaut und die in den Feldkampagnen gemessenen Spektren mit dieser abgeglichen. Mit dieser Methode der LUT-Invertierung aus unterschiedlichen Aufnahmewinkeln konnten Genauigkeiten bei der LAI-Schätzung von 18 % und bei Blattchlorophyll von 20 % erzielt werden. Eine starke Anisotropie, also eine Reflexionsabhängigkeit von der Beleuchtungs- und Aufnahmerichtung, wurde bei Winterweizen vor allem für frühe Entwicklungsstadien festgestellt.

Bei einer anschließenden Studie zur Unsicherheitsanalyse des Spektralmodells wurden PROSAIL-Ergebnisse, bei denen real gemessene Pflanzenparameter als Input dienten, den zugehörigen Reflektanzspektren gegenübergestellt. Es zeigten sich hierbei mitunter starke Abweichungen zwischen gemessenen und modellierten Spektren, die im Falle des Winterweizens einen saisonalen Verlauf zeichneten. Vor allem während frühen Wachstumsstadien tendierte das Modell dazu die Reflektanz im nahen Infrarot zu überschätzen, während es gegen Ende der Wachstumsperiode eher eine Unterschätzung aufwies. Als Unsicherheitsfaktor wurde die Parametrisierung des Modells ausgemacht, wenn der ALIA-Parameter als echter physikalische Blattwinkel interpretiert wird. Es wurde geschlussfolgert, dass eine Separierung von LAI und ALIA bei der Invertierung von PROSAIL eine korrekte Abschätzung der weniger sensitiven Parameter behindert.

Die Erstellung des Vegetations-Prozessors erforderte die Verwendung von Regressions-Algorithmen des maschinellen Lernens (MLRA), da eine Verteilung von großen LUTs an die User nicht praktikabel wäre. Die MLRAs wurden an synthetischen Datensätzen trainiert, wobei zunächst die Optimierung der Hyperparameter im Vordergrund stand, bevor die Anwendung an echten Spektraldaten unternommen wurde. Es konnten dabei erst

aussagekräftige Ergebnisse produziert werden, als die Trainingsdaten mit einem künstlichen Rauschen belegt wurden, da die Algorithmen unter einer Überanpassung an die Modellumgebung litten. Mithilfe des Prozessors konnten schließlich LAI, ALIA, C_{cab} und C_w aus hyperspektralen Daten abgeleitet werden. Künstliche neuronale Netze dienen dabei als Blackbox-Modelle, die in kurzer Zeit große Datenmengen verarbeiten können und somit einen entscheidenden Beitrag zur modernen angewandten Fernerkundung für eine breite User-Community leisten.

Summary

After years of scientific and technical preparation, the launch of an unmanned German space-mission is planned to be initiated in 2020. The Environmental Mapping and Analysis Program (EnMAP) is going to provide an equally named hyperspectral imager to map land surfaces. Scientists of environmental disciplines of monitoring of ecosystems, agricultural, forestry and urban areas as well as coastal and inland waters, geology and soils prepared themselves for the upcoming data prior to the actual launch. Although there already exists a variety of useful algorithms for a profound analysis of spectral data, new challenges will arise given the uniqueness of the EnMAP-mission in the global context of remote sensing; i.e. coverage of the full range of the optical spectrum (420 nm – 2450 nm) in combination with a moderate spatial resolution of 30 m and a high signal-to-noise ratio of at least 180 in the shortwave infrared and above 400 in the visible spectrum. This enables an imaging quality which to this date has only been reached by airborne systems.

The efforts of this dissertation comprise activities in the scientific preparation phase for agro-geographical tasks. Algorithms and tools for an analysis of hyperspectral data are being provided for free in the QGIS-plugin EnMAP-Box 3. Urgent questions in the agricultural sector revolve around the derivation of biochemical and biophysical parameters from remote sensing data. For this reason, the overarching objective of this promotion is the development of a scientific EnMAP-tool for managed areas of vegetation (EnMAP Managed Vegetation Scientific Processor).

At first, an extensive field campaign was planned and then started in April, 2014. Apart from spectral observations of leaves, canopies and soils in a winter wheat and a maize field, also relevant plant parameters were acquired at the exact same spots. Namely, they are the Leaf Area Index (LAI), leaf chlorophyll content (C_{cab}), leaf water content (EWT or C_w), relative dry leaf weight (LMA or C_m), Average Leaf Inclination Angle (ALIA) as well as other secondary parameters like canopy height, phenological stage and the solar vector. Spectral measurements were captured from different observation angles to match ground data with the sensing geometry of the future EnMAP-satellite, which can be tilted up to 30° orthogonal to its direction of flight.

A common procedure to derive relevant crop parameters is to make use of the radiative transfer model PROSAIL, which simulates the spectral signal of a vegetated surface based on biophysical and biochemical input parameters. If this process is reverted, said parameters can be derived from measured spectral data. To do so, a Look-Up-Table (LUT) is built containing model runs of PROSAIL and then subsequently compared against spectra from the field campaigns. With this approach of LUT-inversions from different observation angles, an accuracy of 18 % could be achieved for LAI and 20 % for C_{cab}. Strong anisotropic effects, i.e. dependence on illumination geometry and sensor orientation, were identified for winter wheat mainly in the early stages of plant development.

In a consecutive study about uncertainties of the spectral model, PROSAIL results fed with in situ measured crop parameters as input, were opposed to their associated reflectance signatures. A strong deviation between measured and modelled spectra was observed, which – in the case of winter wheat – showed a seasonal behavior. The model tended to overestimate reflectances in the near infrared for early phenological stages and to underestimate them at end of the growing period. The parametrization of the model was identified as an uncertainty factor if the ALIA parameter is interpreted as true physical leaf inclinations. It was concluded that a separation of LAI and ALIA at inversion of PROSAIL prevents an adequate estimation of the less sensitive parameters.

The development of the vegetation processor required the use of Machine Learning Regression Algorithms (MLRA), since distribution of large LUTs to the user would be impracticable. The MLRAs were trained with synthetic datasets with primary importance to optimize their hyperparameters, before attempting to apply the algorithms to real spectral data. Significant results could not be obtained until training data were altered with artificial noise, because algorithms suffered from overfitting to the model environment. Executing the processor allowed to derive LAI, ALIA, C_{cab} and C_w from hyperspectral data. Artificial neural networks served as black box models, which digest great amount of data in a short period of time and thus make a decisive contribution to modern applied remote sensing with relevance for a broad user-community.

Acknowledgement

My deepest appreciation is expressed firstly to Prof. Dr. Wolfram Mauser for giving me the opportunity to conduct my research under his kind supervision. I am very grateful for the excellent working conditions he provided as head of the chair of Geography and Remote Sensing at the Department of Geography, LMU Munich. His trust in this dissertation and his exemplary dedication to environmental sciences have been and continue to be a major source of inspiration to me.

PD Dr. Tobias Hank deserves my deepest and sincere gratitude. Not only for guiding me as a second supervisor, but also for being an outstanding source of motivation throughout all the years. Under no circumstances I could have managed to gain ground in research the way I did without his professional expertise and his convincing support. I further owe him a number of fantastic memories at the department as well as at conferences and meetings. Thanks to him for always listening and always having an answer to whatever issue comes up!

I want to further express thankfulness to my project collaborators and co-authors: Matthias Wocher, with his sharp mind, his supportiveness and the joy he brings to everyone around. Dr. Katja Berger, for her valuable contributions to my studies, her helping insights and practical recommendations and for her devoted contribution to the EnMAP project in general. Dr. Matthias Locherer, for helping me mastering my first steps at the department and for all his preliminary work with which he paved the way for this dissertation.

The massive amount of field data could not have been acquired without the helping hands of an incredible number of student researchers: Adrian J., Adrian K., Alexander M., Alexander W., Andi, Anja, Ann-Cathrin, Christian, Elin, Hannah, Johannes, Joshua, Katharina A., Katharina W., Lisa, Luis, Lukas, Luzie, Magdalena, Manou, Marie, Marina, Markus, Marlis, Matthias, Miriam, Mona, Sarah, Sebastian, Stefanie L., Stefanie W. and Werner. Throughout the four seasons of field campaigns, all of you contributed to build up this massive data set and therefore deserve my greatest acknowledgement.

I further want to thank my other colleagues for all the inspiring discussions on the balcony, for enjoying the haut cuisine at LMU Mensa together and for helping me out in all sorts of critical situations. Thank you especially, Vroni and Andrea, for your patience and for always sharing your thoughts with me!

Table of Contents

List of Figures

List of Tables

1. Introduction

Knowledge about the state of soils and crops have been essential for farming practices at all times. However, the way how information is obtained, has drastically changed in the past decades. Visual inspection and random sampling are far from being outdated, but they have become only an additional component in practical farming, since they are often laborious, time-consuming, imprecise and rely on destructive methods (Khan et al. 2018). The assimilation of remote sensing data into computer-aided farming caused a major step forward for site-specific management (Pinter Jr et al. 2003). Sensing with tractor-mounted devices and UAVs represent the closest form of crop monitoring (Adão et al. 2017), other sources embrace airborne systems and increasingly also various spaceborne imagery (Hank et al. 2018). The additionally acquired information contributes to a smart management of fertilization, plant protection, irrigation and harvesting (Mulla 2013). The diversity of biophysical and biochemical variables with immediate relevance for modern farming demands practical methods to retrieve the desired parameters with decent accuracy.

The end of hunger and the achievement of food security are essential parts of the second Sustainable Development Goal. Their success, however, is endangered by environmental degradation and climate change (FAO 2016). These global issues necessitate global solution strategies, which clearly emphasizes the suitability of satellite data in this context (Mauser et al. 2012).

1.1. Hyperspectral Remote Sensing

The literal meaning of remote sensing is recording information about a desired target without the requirement of actually being present at the exact location during that time (Sabins 2007). To this end, remote sensing is opposed to laboratory methods or contact-bound measurements in situ. The scientific need for data of large or far-off areas had enforced early technical inventions in the field of aerial photography at the dawn of the 20th century (Campbell and Wynne 2011). In the following decades, sensors made remarkable progress in terms of data quality and accuracy. But even if today's micro sensors seem to have little in common with former analog cameras, they still share the basic physical principles in making use of the complex interactions between mass and energy.

Electromagnetic waves pass energy from a source to the target without relying on physical matter as a transport medium. The frequency at which their magnetic and electric field alternate determines how they interact with molecules of different size and structure (Rees 2013). Remote Sensing means measuring radiation that has been emitted, attenuated, scattered or reflected from the target of interest and interpret the result to gain knowledge about its physical or chemical nature (Jones and Vaughan 2010).

There are several ways to categorize sensors depending on their measurement principle, field of application, technical specifications or mounting system (Schaepman-Strub et al. 2009). In each category of sensors, there is a possible further classification into different spatial, spectral, temporal and radiometric resolutions. Sensors which record radiation in many contiguous wavebands are called hyperspectral. They are opposed to multispectral systems with a broader spectral resolution. Hyperspectral sensors have first been used in laboratories and on site for proximal sensing, before NASA's Johnson Space Center became the first to mount an imaging spectrometer on an airborne platform (van der Meer et al. 2001). The transition of hyperspectral remote sensing into space, however, took several decades with the main issue being that spectral irradiance decreases with smaller bandwidths (Ortenberg et al. 2011). Splitting the incoming energy into sections of 10 nm and less comes to the expense of a lower signal-to-noise ratio if the spatial resolution is to keep up with that of other environmental satellites in the low earth orbit. Nevertheless, as of 2019, space agencies are making a collective effort in launching satellites with hyperspectral sensors given the reported benefits in remote sensing of geology (Ramakrishnan and Bharti 2015), forests and grassland (Lopatin et al. 2017; Thenkabail et al. 2004) or agriculture (Hank et al. 2018; Liu et al. 2015) arising with that technology.

A descriptive way of illustrating the quantity of hyperspectral data is by visualizing them as a cube. The two-dimensional image (rows, columns) is extended into a spectral axis with several hundred or thousand discrete wavelengths. As technical progress advances, slices into one or the other direction become finer, increasing not only the informative content but also the required memory to transmit, process and store data (Figure 1.1.1).

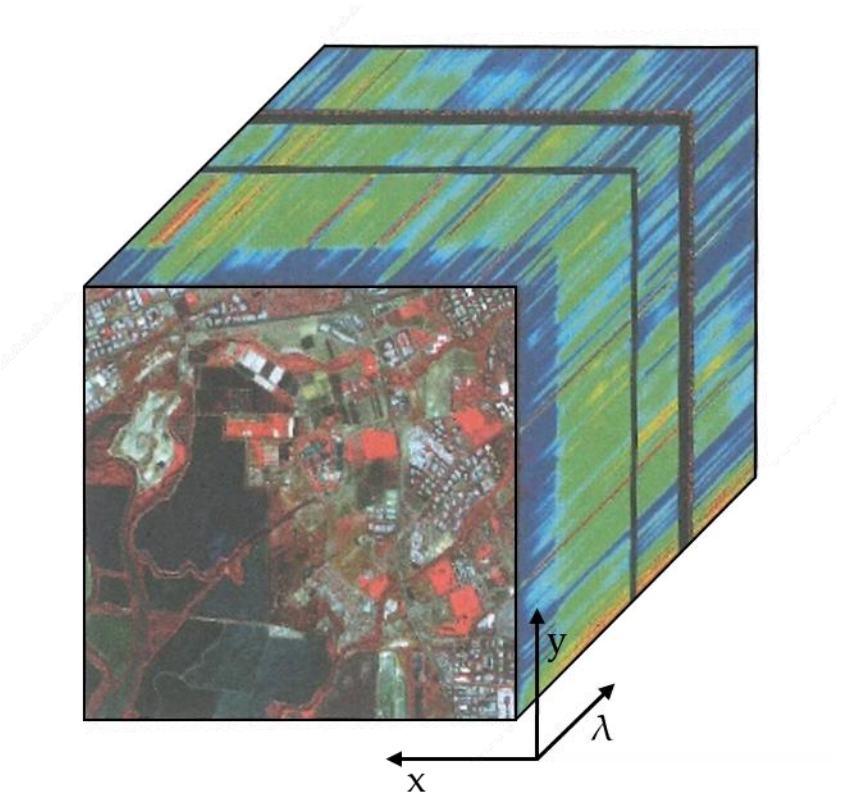


Figure 1.1.1. Concept of the hyperspectral data cube. A hyperspectral image can be envisaged as an image with a spatial domain of columns (x) and rows (y) with sensing wavelengths (λ) as an added third dimension. The scene shows an AVIRIS image acquired from Moffett Field, San Francisco Bay, California on June 20, 1997 as a false-color composite with NIR/Red/Green. Figure adapted from Pu 2017.

The first hyperspectral spaceborne sensor was Hyperion equipped on EO-1 (Folkman et al. 2001), launched in November, 2000. Suffering from a low signal-to-noise ratio, consecutive programs sought a better tradeoff between data quality and adequate spatial and spectral resolution. While the Italian PRecursore IperSpettrale della Missione Applicativa (PRISMA; Candela et al. 2016) has recently been launched, other missions like SHALOM (Feingersh and Dor 2015), NASA's Hyperspectral InfraRed Imager (HyspIRI; Lee et al. 2015) and ESA's Copernicus Hyperspectral Imaging Mission for the Environment (CHIME; Nieke and Rast 2018) are still in planning phase. In the near future, the German Aerospace Center (DLR e.V.) will finish development and production phase of the Environmental Mapping and Analysis Program satellite and target its launch for 2020 (EOC 2018a).

1.2. An overview of the EnMAP mission

The Environmental Mapping and Analysis Program (EnMAP) is a German project with multinational involvement. It is organized in four project parts (Figure 1.2.1, EOC 2018b):

- **Project Management** – led by DLR Space Administration
- **Scientific Principal Investigator** and the **EnMAP Science Team** – led by the Deutsches GeoForschungsZentrum (GFZ) with contributions of the EnMAP Science Advisory Group
- **Space Segment** – led by the OHB System AG
- **Ground Segment** – led by DLR's Remote Sensing Technology Institute

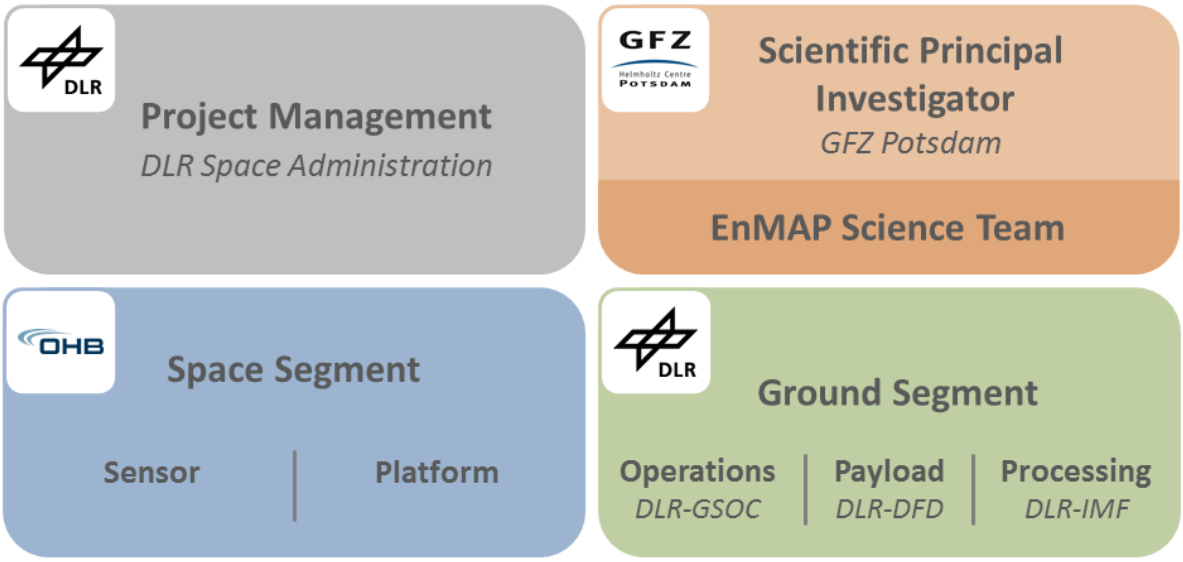


Figure 1.2.1. Organigram of the project management of EnMAP (EOC 2018b).

The EnMAP Science Teams were formed to cover the most important environmental issues and their respective fields of research (Guanter et al. 2016):

- Natural Ecosystems and Ecological Gradients
- Forests
- Agricultural Land
- Geological Exploitation
- Digital Soil Mapping
- Coastal and Inland Waters
- Urban Areas

Each science team contributes to the general scope of hyperspectral remote sensing with particular focus on future EnMAP applications. They are assisted by members of the EnMAP Science Advisory Group, which is a consolidation of experts in remote sensing and environmental subjects.

The instrument of the mission accommodates one spectrometer for the visible and near infrared (VNIR) and another for the shortwave infrared (SWIR), both operating in push-broom mode and together covering the full spectral domain from 420 nm – 2450 nm (Guanter et al. 2015). The system is conceptualized to minimize effects of smile and keystone and offers an across-track pointing capability of ± 30° allowing to decrease revisit times down to e.g. 2.3 days at 45° latitude. The ground sampling distance (GSD) – often referred to as spatial resolution – is 30 m. The mean spectral sampling distance (SSD), i.e. spectral resolution, is 6.5 nm for the VNIR and 10 nm for the SWIR sensor (Guanter et al. 2015). In-flight calibration will ensure quality criteria such as spectral accuracy < 0.5 nm in the VNIR and < 1.0 nm in the SWIR throughout the mission lifetime (Storch et al. 2014). A full sheet of all major characteristics of EnMAP is listed in Table 1.2.1.

Table 1.2.1. Specifications of the planned EnMAP mission and instrument. Adapted from (Guanter et al. 2015).

Mission Requirements	
Spectral range	420–2450 nm
Ground sampling distance	30 m
Swath width	30 km
Swath length	up to 1000 km/orbit
Coverage	Global in near-nadir mode (OZA ≤5°)
Orbit	Sun-synchronous, 11:00 local time descending node
Daily coverage	5000 km
Target revisit time	4 days with 30°across-track pointing
Instrument Requirements	
Imaging principle	Push-broom-prism
Spectral range	VNIR: 420–1000 nm/SWIR: 900–2450 nm
Mean spectral sampling distance	VNIR: 6.5 nm/SWIR: 10 nm
Spectral oversampling	1.2
SNR at reference radiance	>400:1 at 495 nm/>180:1 at 2200 nm
Spectral calibration accuracy	VNIR: 0.5 nm/SWIR: 1 nm
Radiometric calibration accuracy	<5 %
Radiometric stability	<2.5 %
Radiometric resolution	14 bit, dual gain in VNIR
Sensitivity to polarization	<5 %
Spectral smile/keystone effect	<20 % of a pixel
Co-registration VNIR-SWIR	<20 % of a pixel

In the years before its launch, EnMAP data need to be available to build and test applications. Hyperspectral data of airborne or field based sensors can be processed to represent expected future EnMAP-HSI spectra in terms of spectral, spatial and radiometric resolution. The correct SSD is obtained by aggregating narrow-band reflectances according to the spectral response function, which represents the sensitivities decreasing with distance to the central wavelengths of each sensor band (Figure 1.2.2). Since usually original data from airplanes or field sensors have a GSD of < 30 m, they have to be spatially aggregated.

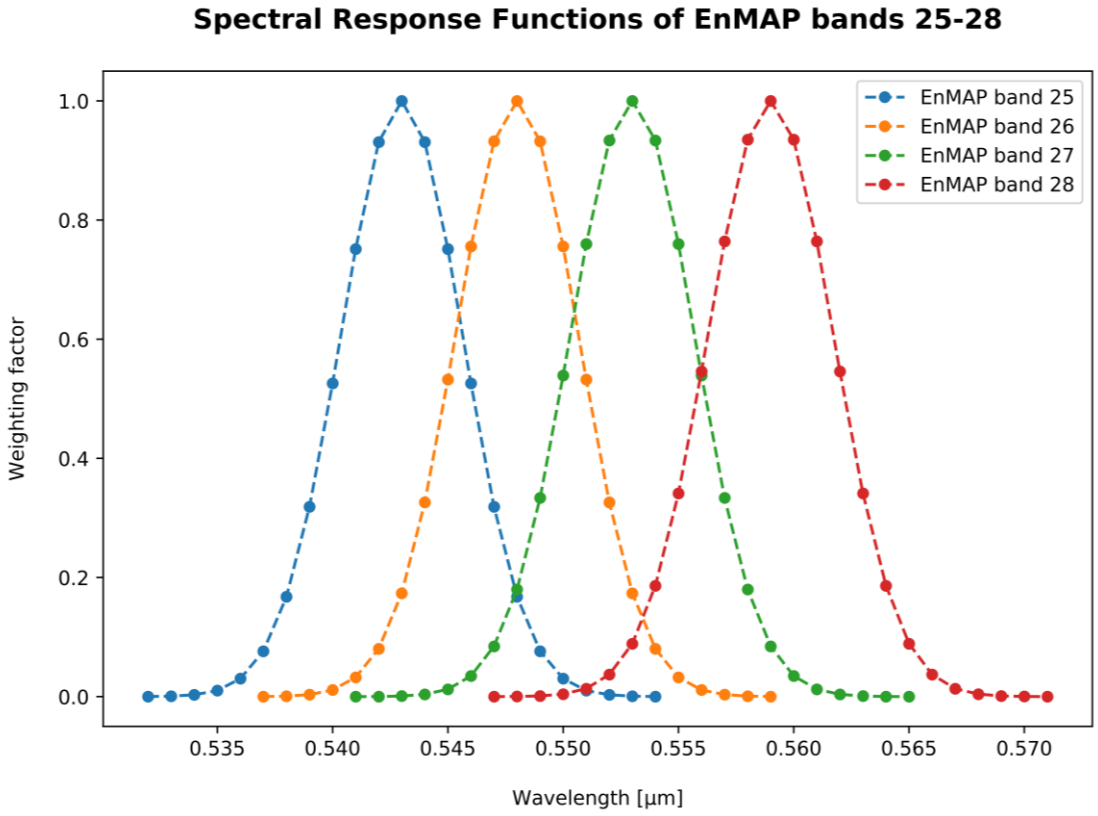


Figure 1.2.2. Illustration of the Spectral Response Functions (SRF) of the EnMAP VNIR-sensor with the examples of bands 25 to 28. The respective central wavelengths are 543 nm, 548 nm, 553 nm and 559 nm. This information can be used for a non-linear spectral resampling with weighted average of the spectrally adjacent wavelengths. *Data source:* GFZ Potsdam.

An even more realistic approach to simulate future EnMAP data is by taking into account atmospheric disturbances, sensor inaccuracies like keystone and smile as well as random sensing errors of the detectors. The EnMAP end-to-end simulator (EeteS; Segl et al. 2012) provides exactly these features and has been used in several studies that assess the future impact and capabilities of EnMAP.

Another demand of the mission is the development of an open-source toolbox that contains applications to work with hyperspectral data. Although it is designed to be usable also with hyperspectral images of other sensors, the conceptualization is intentionally geared towards EnMAP data and the toolbox is therefore called EnMAP-Box (van der Linden et al. 2015). After two major versions running on IDL code, the current EnMAP-Box 3 (Rabe et al. 2018) is embedded in QGIS as a free plug-in. This yields the advantage that all applications have access to most python libraries of the QGIS environment, such as PyQt for GUI development (Summerfield 2007) or the Geospatial Data Abstraction Library (GDAL; Warmerdam 2008).

All EnMAP Science Teams contribute to the EnMAP-Box, providing a final mixture of applications to analyze images in the context of geology, soils, coastal and inland water, and vegetation. This set is complemented by tools to adjust meta-data, perform pixel algebra and a range of other image processing algorithms.

1.3. Canopy Reflectance Modelling

Canopy reflectance modelling is the simulation of the reflectance signature of vegetation with multiple scatterers. Different aspects of the structure of canopies are considered, depending on the type of model used.

Geometric-optical models were designed for canopies with singular clusters of vegetation, which can be described by a fixed array (Jones and Vaughan 2010). This type of model has been mainly used in forestry applications, e.g. by assuming coniferous trees as Lambertian cones aligned in a raster grid (Li and Strahler 1985).

Monte-Carlo ray-tracing models, on the other hand, require a more detailed description of the architecture of vegetation and their interaction with landscapes (Lewis 1999). Each simulation of a ray is randomly initialized at the light source and interacts with canopy elements according to individual probability density functions (Govaerts and Verstraete 1998). This task is computationally very demanding, which is why often diffuse irradiance is neglected in ray-tracing models. The probably most famous example of a 3D canopy reflectance model is the Discrete Anisotropic Radiation-Transfer Model (DART; Martin et al. 2003), which uses kernel functions to model radiative interactions between earth and its atmosphere.

The most widely used approach to model canopy reflectance is based on the turbid medium assumption. Radiative transfer in vegetation is thus described following the theory of radiation propagation in gases (Kubelka and Munk 1931). This approach differs from the former mentioned models, as there is no real localization of rays in the medium, but rather a statistical description of direct radiant fluxes by distribution functions (Jones and Vaughan 2010). This principle was extended in the Suits model (Suits 1971) by the directional solar radiation and the flux towards the observer to form the first classical four-stream radiative transfer model. In the Suits model, two major approximations are used to simplify the complex properties of real canopies:

- Homogeneity – all components that interact with radiation are evenly distributed and infinitesimally small
- Horizontality – canopies consist of horizontal layers with non-overlapping leaves

This somewhat unsatisfying simplification was improved in the first version of Scattering of Arbitrarily Inclined Leaves (SAIL; Verhoef 1984), which incorporates a specific leaf angle distribution into the modelling approach. In SAIL, leaves follow a random azimuthal and a statistical inclination distribution function. The shortcoming of treating leaves as point scatterers was overcome by introduction of a hot spot size parameter. There exist several adaptations and advancements in the family of SAIL-models, e.g. a multilayer structure approach 2M-SAIL (Weiss et al. 2001), a hybrid approach to incorporate discontinuous canopies (GeoSAIL) and its speed-optimized two-layer version 4SAIL2 (Verhoef and Bach 2003).

Radiative transfer in leaves was originally described in the plate model (Allen et al. 1969). Treating leaves as single layer absorbers, however, was soon considered a weak approximation of real biochemical processes in plant tissues. The PROSPECT leaf optical properties model (Jacquemoud and Baret 1990) describes scattering by a spectral refractive index and an arbitrary parameter indicating the number of absorber plates and the air-filled spaces between them. Absorption follows the Lambert-Beer law of attenuated radiation with wavelength-dependent absorption coefficients for all major biochemical constituents of leaves. All versions share the consideration of Equivalent Water Thickness (EWT, or C_w) and Leaf

Mass per Area (LMA, or C_m). Leaf pigments were first characterized by a single parameter (C_{cab}) and later split into leaf chlorophyll, xanthophylls/carotenoids (C_{car}; Feret et al. 2008) and later anthocyanins (C_{anth}; Féret et al. 2017). Brown pigments (C_{brown}) are summarized into an arbitrary parameter to describe the influence of polyphenols in the red-edge region (Baret and Fourty 1997).

The demand for a coupled PROSAIL model containing both the leaf optical properties and a description of the canopy architecture, arose from the need of a model to retrieve biophysical variables from canopy reflectance spectra under consideration of a given sun-sensor-target geometry. Since the inversion of SAIL leads to an under-determined system of equations, this issue is solved with providing the directional and hemispherical radiation from leaf reflectance and leaf transmittance by PROSPECT (Jacquemoud et al. 2009b). A description of input variables for both models is shown in Table 1.3.1, the coupling is illustrated in Figure 1.3.1.

Table 1.3.1. Overview of the PROSAIL parameters and their according dimensions. Some parameters, e.g. the leaf chlorophyll content, are used in all PROSPECT versions, whereas other parameters were included in newer releases.

Parameter	Description	Unit	Model versions
N	Leaf structure parameter	-	Prospect (all)
C_{cab}	Leaf Chlorophyll$_{a+b}$ content	µg cm^{-2}	Prospect (all)
C_w	Leaf Equivalent Water Thickness (EWT)	cm	Prospect (all)
C_m	Leaf Mass per Area	g cm^{-2}	Prospect (all)
C_{car}	Leaf Carotenoids content	µg cm^{-2}	Prospect 5
C_{brown}	Leaf Brown Pigments parameter	-	Prospect Xb
C_{anth}	Leaf Anthocyanins content	µg cm^{-2}	Prospect D
LAI	Leaf Area Index	m^2 m^{-2}	SAIL
LIDF or ALIA	Leaf Inclination Distribution Function or Average Leaf Inclination Angle	- or Deg	SAIL
Hspot	Hot Spot size parameter	-	SAIL
ρ_{soil}	Soil Reflectance	-	SAIL
P_{soil}	Soil Brightness Parameter	-	SAIL
SZA	Sun Zenith Angle	Deg	SAIL
OZA	Observer Zenith Angle	Deg	SAIL
rAA	relative Azimuth Angle	Deg	SAIL
skyl	Ratio of diffuse to total incident radiation	-	SAIL

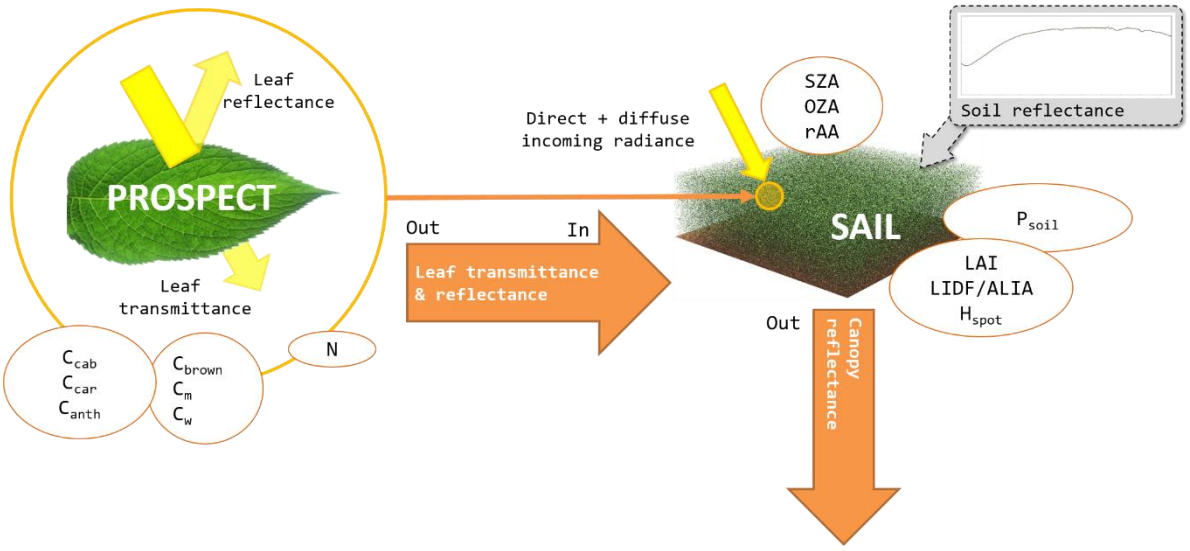


Figure 1.3.1. Flowchart of the coupling between PROSPECT and SAIL to form the PROSAIL model.

Despite known shortcomings in practical applications, its limited demand in input variables and quick processing time renders PROSAIL the most widely used radiative transfer model for canopy reflectance modelling in the past decades (Berger et al. 2018). Even today, model development prevails on leaf scale, e.g. PROSPECT-DyN (Wang et al. 2015), and canopy scale, e.g. Soil-Leaf-Canopy (SLC; Verhoef and Bach 2007) or SCOPE (van der Tol et al. 2009).

1.4. Machine Learning Regression Algorithms

Machine learning is a broad term for algorithms and statistical models that perform a given task without relying on instructions about the required steps. The idea is merely to let the machine find, learn and reproduce certain patterns within the data (Mitchell et al. 1990). A software or programming script is needed to fit a mathematical relationship between a vector of predictor variables X and target variables Y from training data. This model can then be used to predict target variables of unseen data.

When each X can be mapped to an associated Y in the training process, the task is called supervised learning. Semi-supervised methods refer to learning data sets with incomplete labels. Unsupervised algorithms finally perform self-organized modelling according to the probability densities of the inputs (Murphy 2012). Another way of separating machine learning is whether it is used to solve a classification or regression problem. For classification, the target variables are called labels and describe the discrete expression of that variable. A typical classification problem would be the production of a land use map from remote sensing

data, separating between e.g. water, bare soil, built-up area or vegetation cover (Abburu and Golla 2015). Regression tasks can be seen as the special cases of a classifications, for which the classes become so small that they create a continuous range. Hence, a Machine Learning Regression Algorithm (MLRA) is a computer script that learns the non-linear relationship between numerical variables by applying a routine of mathematical calculations (Smola and Vishwanathan 2008).

Machine Learning has a long tradition in remote sensing, starting off with automated mapping of features identified in optical or radar data (e.g. Conway et al. 1991; Greenspan et al. 1994; Kubat et al. 1998) before evolving into the computationally more demanding field of parameter estimation (Ali et al. 2015; Mountrakis et al. 2011; Verrelst et al. 2012; Wang et al. 2016). Even though it may be appealing to use machine learning in all sorts of tasks, it remains problematic that most approaches work like black boxes and that only the results but not the predictions themselves can be assessed afterwards. So firstly, caution is advised to check for causality along with correlation. Secondly, ways must be found to prevent the algorithms from overfitting, which happens when they become overspecialized in the training process (Dietterich 1995). On the upside, Machine Learning has proven to be particularly useful in geosciences by either speeding up deterministic models or by generating empirical relationships if there is no deterministic description at all (Lary 2010).

2. Publications

2.1. Framework

The incentive for this thesis was to analyze a large amount of field data and to use it to improve the estimation of biochemical and biophysical variables from future EnMAP images. The resulting algorithms will be made available to a broad user community of experts and higher-educated audiences. The dissertation started in 2014 at EnMAP project phase II, entitled "Monitoring the Phenological Development of Agricultural Crops", and was continued in project phase III "Developing the EnMAP Managed Vegetation Scientific Processor". A first foundation for this dissertation was laid in the author's master thesis:

Danner, M. *(2013): Development and demonstration of a campaign layout for the field measurement of reflectance relevant biophysical parameters with respect to the observation geometry of the future EnMAP-HSI. Unpublished master thesis. Munich, Germany.*

In that work, a review of the methodological orientated literature identified a set of instruments and approaches for a large-scale measurement campaign. Given the practical nature of that document, emphasis was placed on the optimization of the relation between effort and benefit on site. The preparation of the thesis included finding optimal methods to measure Leaf Area Index, Leaf Chlorophyll Content, Leaf Carotenoid Content, Brown Pigment Content, Leaf Water Content, Leaf Dry Matter Content and Average Leaf Inclination Angle. It further described the derivation of the Leaf Inclination Distribution Function and the Solar Vector as well as the acquisitions of spectra of the leaves, the canopy and of the bare soil. Basically, this list has not changed since that date and the parameters as well as most of the measurement techniques were still the same at the tentative end of the field campaigns in 2018. A 3x3 raster with 10 m grid size was chosen for the preliminary field work, which had been done according to a step-by-step instruction (Action Plan) for varying team sizes. Finally, field protocols and a spreadsheet pre-print had been conceptualized and tested before the actual campaign started.

With beginning of the actual Ph.D. studies, the first two years of field campaigns were accompanied by a constant improvement of methodology for the data analysis. The achieved progress in creating and inquiring Look-Up-Tables was published as

*Locherer, M., Hank, T., **Danner, M.** & Mauser, W. (2015): Retrieval of Seasonal Leaf Area Index from simulated EnMAP data through optimized LUT-based Inversion of the PROSAIL Model. Remote Sensing, Vol. 7(8).*

In the following years, the research was geared towards an analysis of the model uncertainties and an assessment of the potential synergies between EnMAP and Sentinel-2. The results were presented at international conferences and published as proceeding papers afterwards (see chapter 2.2). In 2017,

***Danner, M.**, Berger, K., Wocher, M., Mauser, W. & Hank, T. (2017): Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy. Remote Sensing, Vol. 9(7)*

was published, in which again a Look-Up-Table approach was successfully applied for the retrieval of LAI and C_{cab}, but this time from multi-angular data on EnMAP resolution. Over the years, the scope of this dissertation had spread out to additionally cover water retrieval methods, e.g.

*Wocher, M., Berger, K., **Danner, M.**, Mauser, W. & Hank, T. (2018): Physically-Based Retrieval of Canopy Equivalent Water Thickness Using Hyperspectral Data. Remote Sensing, Vol. 10(12),*

advances in hyperspectral data preprocessing, e.g.

*Berger, K., Atzberger, C., **Danner, M.**, Wocher, M., Mauser, W. & Hank, T. (2018): Model-Based Optimization of Spectral Sampling for the Retrieval of Crop Variables with the PROSAIL Model. Remote Sensing, Vol. 10(12)*

and a long-sought updated review study of the canopy reflectance model PROSAIL, which was used for spectral modelling of the gathered field data throughout the thesis:

*Berger, K., Atzberger, C., **Danner, M.**, D'Urso, G., Mauser, W., Vuolo, F. & Hank, T. (2017): Evaluation of the PROSAIL model capabilities for the future EnMAP model environment: a review study. Remote Sensing, Vol. 10(1).*

In the next period, the main focus was placed on building the final applications for the EnMAP-Box. The tools, their structure and exemplary results were demonstrated on international conferences and published as proceedings, before a second scientific paper with contribution as first author,

Danner, M., *Berger, K., Wocher, M., Mauser, W. & Hank, T. (2019): Fitted PROSAIL Parameterization of Leaf Inclinations, Water Content and Brown Pigment Content for Winter Wheat and Maize Canopies. Remote Sensing, Vol. 11(10),*

was published in 2019. It shed light on the performance of PROSAIL to model crop biophysical and biochemical parameters. Leaf inclinations were identified as a main source for modelling uncertainties with major influence also to the accuracy of PROSAIL inversions. The third paper,

Danner, M., *Wocher, M., Berger, K., Mauser, W. & Hank, T. (2019): Training Machine Learning Regression Algorithms to predict biophysical & biochemical variables from resampled hyperspectral EnMAP data using PROSAIL. ISPRS J. Photogramm. Remote Sens., under review,*

was submitted to the ISPRS Journal of Photogrammetry and Remote Sensing and is currently under review. In this concluding publication, appropriate methods towards a scientific vegetation processor were evaluated with special focus on optimization of the hyperparameters of Machine Learning Regression Algorithms.

2.2. Complete list of Scientific Dissemination

Research Articles

- *Locherer, M., Hank, T., **Danner, M.** & Mauser, W. (2015): Retrieval of Seasonal Leaf Area Index from simulated EnMAP data through optimized LUT-based Inversion of the PROSAIL Model. Remote Sensing, Vol. 7(8).*
- ***Danner, M.**, Berger, K., Wocher, M., Mauser, W. & Hank, T. (2017): Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy. Remote Sensing, Vol. 9(7).*
- *Wocher, M., Berger, K., **Danner, M.**, Mauser, W. & Hank, T. (2018): Physically-Based Retrieval of Canopy Equivalent Water Thickness Using Hyperspectral Data. Remote Sensing, Vol. 10(12).*
- *Berger, K., Atzberger, C., **Danner, M.**, Wocher, M., Mauser, W. & Hank, T. (2018): Model-Based Optimization of Spectral Sampling for the Retrieval of Crop Variables with the PROSAIL Model. Remote Sensing, Vol. 10(12).*
- *Berger, K., Atzberger, C., **Danner, M.**, D'Urso, G., Mauser, W., Vuolo, F. & Hank, T. (2017): Evaluation of the PROSAIL model capabilities for the future EnMAP model environment: a review study. Remote Sensing, Vol. 10(1).*
- ***Danner, M.**, Berger, K., Wocher, M., Mauser, W. & Hank, T. (2019): Fitted PROSAIL Parameterization of Leaf Inclinations, Water Content and Brown Pigment Content for Winter Wheat and Maize Canopies. Remote Sensing, Vol. 11(10).*
- ***Danner, M.**, Wocher, M., Berger, K., Mauser, W. & Hank, T. (2019): Training Machine Learning Regression Algorithms to predict biophysical & biochemical variables from resampled hyperspectral EnMAP data using PROSAIL. ISPRS J. Photogramm. Remote Sens., under review.*

Technical Articles

- ***Danner, M.**, Locherer, M. & Hank, T. (2015): Defining Campaign Layouts & Sampling Strategies – Theory – Principles – Problems – Practice. An EnMAP Field Guide. Technical Report, GFZ Data Services. Potsdam.*
- ***Danner, M.**, Locherer, M., Hank, T. & Richter, K. (2015): Determining Leaf Area Index (LAI) with the LI-Cor LAI 2200c or LAI-2200 (+ 2200Clear Kit) – Theory, Measurement, Problems, Interpretation. An EnMAP Field Guide. Technical Report, GFZ Data Services. Potsdam.*
- ***Danner, M.**, Locherer, M., Hank, T. & Richter, K. (2015): Spectral Sampling with the ASD FieldSpec 4 – Theory, Measurement, Problems, Interpretation. An EnMAP Field Guide. Technical Report, GFZ Data Services. Potsdam.*

- *Dotzler, S., **Danner, M.**, Locherer, M., Hank, T. & Richter, K. (2015): Measuring Soil Moisture with TDR-Probes – Theory, Measurement, Problems, Interpretation. An EnMAP Field Guide. Technical Report, GFZ Data Services. Potsdam.*
- *Süß, A., Obster, C., **Danner, M.**, Locherer, M., Hank, T. & Richter, K. (2015): Determining Leaf Chlorophyll Content with the Konica Minolta SPAD-502Plus – Theory, Measurement, Problems, Interpretation. An EnMAP Field Guide. Technical Report, GFZ Data Services. Potsdam.*

Conference Proceedings

- ***Danner, M.**, Locherer, M., Hank, T. & Mauser, W. (2015): Analyzing Uncertainties in simulated Canopy Reflectance through exhaustive Comparison with in-situ measured Optical Properties. International Geoscience and Remote Sensing Symposium (IGARSS), Milan (Italy).*
- *Locherer, M., Hank, T., **Danner, M.** & Mauser, W. (2015): Systematic Analysis of the LUT-based Inversion of Prosail using full range Hyperspectral Data for the Retrieval of Leaf Area Index in View of the future EnMAP Mission. International Geoscience and Remote Sensing Symposium (IGARSS), Milan (Italy).*
- ***Danner, M.**, Hank, T. & Mauser, W. (2016): Comparing the Potential of the Sentinel-2 MSI and the future EnMAP HSI for the Retrieval of Winter Wheat Crop Parameters in Southern Germany. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Hank, T., Bach, H., **Danner, M.**, Hodrius, M., Goetz, H. & Mauser, W. (2016): Tracing Crop Nitrogen Dynamics on the Field-Scale by combining Multisensoral EO data with an integrated Process Model – A Validation Experiment for Cereals in Southern Germany. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Klug, P., Schlenz, F., Hank, T., Migdall, S., Weiss, I., **Danner, M.**, Bach, H. & Mauser, W. (2016): Implementation of Sentinel-2 data in the M4Land system for the generation of continuous information products in agriculture. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Berger, K., Wang, Z., **Danner, M.**, Wocher, M., Mauser, W. & Hank, T. (2018): Simulation of Spaceborne Hyperspectral Remote Sensing to Assist Crop Nitrogen Content Monitoring in Agricultural Crops. International Geoscience and Remote Sensing Symposium (IGARSS), Valencia (Spain).*
- ***Danner, M.**, Wocher, M., Berger, K., Mauser, W. & Hank, T. (2018): Developing a Sandbox Environment for Prosail, Suitable for Education and Research. International Geoscience and Remote Sensing Symposium (IGARSS), Valencia (Spain).*

- *Wocher, M., Berger, K., **Danner, M.**, Mauser, W. & Hank, T. (2018): Hyperspectral Retrieval of Canopy Water Content Through Inversion of the Beer-Lambert Law. International Geoscience and Remote Sensing Symposium (IGARSS), Valencia (Spain).*

Oral Presentations

- ***Danner, M.**, Locherer, M., Hank, T. & Mauser, W. (2015): Analyzing Uncertainties in simulated Canopy Reflectance through exhaustive Comparison with in-situ measured Optical Properties. International Geoscience and Remote Sensing Symposium (IGARSS), Milan (Italy).*
- *Locherer, M., Hank, T., **Danner, M.** & Mauser, W. (2015): Systematic Analysis of the LUT-based Inversion of Prosail using full range Hyperspectral Data for the Retrieval of Leaf Area Index in View of the future EnMAP Mission. International Geoscience and Remote Sensing Symposium (IGARSS), Milan (Italy).*
- ***Danner, M.**, Locherer, M., Hank, T. & Mauser, W. (2015): Describing the Quality of simulated Canopy Reflectance by comparing modelled Spectra with in-situ measured Optical Properties. 9th EARSeL SIG Imaging Spectroscopy Workshop, Luxembourg.*
- ***Danner, M.**, Hank, T., Berger, K. & Mauser, W. (2016): Die Relevanz hyperspektraler Fernerkundung zur Unterstützung einer globalen Ernährungssicherung. Ringvorlesung Universität Osnabrück.*
- ***Danner, M.** (2016 & 2017): Anwendung von Raumfahrtantrieben: Fernerkundung. DLR_Summer_School Raumfahrtantriebe, Lampoldshausen.*
- ***Danner, M.**, Hank, T. & Mauser, W. (2016): Comparing the Potential of the Sentinel-2 MSI and the future EnMAP HSI for the Retrieval of Winter Wheat Crop Parameters in Southern Germany. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Hank, T., Bach, H., **Danner, M.**, Hodrius, M., Goetz, H. & Mauser, W. (2016): Tracing Crop Nitrogen Dynamics on the Field-Scale by combining Multisensoral EO data with an integrated Process Model – A Validation Experiment for Cereals in Southern Germany. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Klug, P., Schlenz, F., Hank, T., Migdall, S., Weiss, I., **Danner, M.**, Bach, H. & Mauser, W. (2016): Implementation of Sentinel-2 data in the M4Land system for the generation of continuous information products in agriculture. ESA Living Planet Symposium, Prague (Czech Republic).*
- *Berger, K., **Danner, M.**, Hank, T. & Mauser, W. (2017): Towards Globally Applicable Retrieval of Bio-Geophysical Variables: Developing an ENMAP Canopy Reflectance Model Environment. EARSeL SIG Imaging Spectroscopy Workshop, Zurich.*

- *Berger, K., **Danner, M.**, Wocher, M., Wang, Z., Mauser, W. & Hank, T. (2019): Monitoring of Crop Nitrogen Status Using a Hybrid Inversion Scheme in the Context of the Future Hyperspectral EnMAP Mission. EARSeL SIG Imaging Spectroscopy Workshop, Brno.*
- *Berger, K., **Danner, M.**, Wocher, M., Wang, Z., Mauser, W. & Hank, T. (2018): Schätzung von pflanzengebundenem Stickstoff landwirtschaftlicher Kulturen aus hyperspektralen Erdbeobachtungsdaten. DLR Symposium, KOMED Köln.*
- *Wocher, M., Berger, K., **Danner, M.**, Mauser, W. & Hank, T. (2018): Hyperspectral Retrieval of Canopy Water Content Through Inversion of the Beer-Lambert Law. International Geoscience and Remote Sensing Symposium (IGARSS), Valencia (Spain).*

Poster Presentations

- ***Danner, M.**, Hank, T., Locherer, M. & Mauser, W. (2015): EnSAG – EnMAP Scientific Advisory Group – Phase II: Monitoring The Phenological Development of Agricultural Crops. EnMAP School 2015, Lauenburg.*
- ***Danner, M.**, Hank, T. & Mauser, W. (2016): Comparing The Potential of the Future EnMAP HSI and Sentinel-2 MSI for the Retrieval of Winter Wheat Crop Parameters in Southern Germany. EnMAP School 2016, Trier.*
- ***Danner, M.**, Wocher, M., Berger, K., Mauser, W. & Hank, T. (2018): Developing a Sandbox Environment for Prosail, Suitable for Education and Research. International Geoscience and Remote Sensing Symposium (IGARSS), Valencia (Spain).*
- *Wocher, M., **Danner, M.**, Berger, K., Mauser, W. & Hank, T. (2018): Environmental Mapping And Analysis Program – A German Scientific Precursor For Hyperspectral Environmental Monitoring. EnviroInfo 2018 Conference - LRZ, Garching.*
- *Wocher, M., Berger, K., **Danner, M.**, Mauser, W. & Hank, T. (2019): Monitoring Crop Water Content from Multi-angular data in the context of the future EnMAP hyperspectral mission. ESA Living Planet Symposium, Milan (Italy).*
- *Berger, K., Verrelst, J., Wocher, M., **Danner, M.**, Wang, Z., Mauser, W. & Hank, T. (2019): Opportunities & Limitations Of Crop Nitrogen Products Retrieval From Future Imaging Spectroscopy Data. Workshop On International Cooperation In Spaceborne Imaging Spectroscopy, Frascati (Italy).*

2.3. Preparation of Scientific Publication I

Before initialization of EnMAP project phase II, scientific objectives for this period had to be defined. In the grant proposal, a knowledge gap about the added value of hyperspectral remote sensing was exposed regarding the identification of the phenological state of agricultural crops from single-look observations (Mauser and Hank 2012).

The leaf structure parameter (N) in PROSAIL is supposed to be closely correlated with the phenological development of agricultural crops, as the leaf mesophyll structure builds up more layers of chlorophyll-containing tissue until senescence (Jones and Vaughan 2010). In theory, an inversion of N is possible if all other parameters are known and fixed in the process. The consecutive field campaign had to be designed to provide all measureable inputs of PROSAIL (see Table 1.3.1). Retrieving N by inversion of a Look-Up-table (LUT) with spectra of winter wheat gave slight indications of a correlation with the phenology (Danner et al. 2015), but the link was too weak and could not be confirmed for maize. Instead, this topic was resolved by a multivariate regression to obtain phenology from the development of N together with LAI and Leaf Chlorophyll Content (Hank et al. 2017).

Instead, subsequent project work was continued to gain better insight in LUT inversion techniques. Complimentary to existing studies of that topic, the approach was tested with special focus on non-nadir observations in constellations that simulate sun-sensor-target geometries with observation angles of the future EnMAP platform. Field data of viewer zenith angles with ± 30° deflection from nadir were analyzed to calculate the anisotropy factors for forward scatter and backscatter. Furthermore, the performance of parameter retrieval from PROSAIL was tested for both viewing directions. The results show that winter wheat crops reveal stronger anisotropic effects at early growth stages and that LAI inversions on average were achieved with highest accuracy from near-nadir observations. In contrast, leaf chlorophyll estimations had lower errors when the image of predominant forward scatter was used.

2.4. Scientific Publication I

Article

Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy

Martin Danner *, Katja Berger, Matthias Wocher, Wolfram Mauser and Tobias Hank

Department of Geography, Ludwig-Maximilians-Universität München, Luisenstraße 37, D-80333 Munich, Germany; katja.berger@iggf.geo.uni-muenchen.de (K.B.); m.wocher@iggf.geo.uni-muenchen.de (M.W.); w.mauser@lmu.de (W.M.); tobias.hank@lmu.de (T.H.)

* Correspondence: martin.danner@iggf.geo.uni-muenchen.de; Tel.: +49-89-2180-6695

Received: 13 June 2017; Accepted: 12 July 2017; Published: date

Abstract: The future German Environmental Mapping and Analysis Program (EnMAP) mission, due to launch in late 2019, will deliver high resolution hyperspectral data from space and will thus contribute to a better monitoring of the dynamic surface of the earth. Exploiting the satellite's ±30° across-track pointing capabilities will allow for the collection of hyperspectral time-series of homogeneous quality. Various studies have shown the possibility to retrieve geo-biophysical plant variables, like leaf area index (LAI) or leaf chlorophyll content (LCC), from narrowband observations with fixed viewing geometry by inversion of radiative transfer models (RTM). In this study we assess the capability of the well-known PROSPECT 5B + 4SAIL (Scattering by Arbitrarily Inclined Leaves) RTM to estimate these variables from off-nadir observations obtained during a field campaign with respect to EnMAP-like sun–target–sensor-geometries. A novel approach for multiple inquiries of a large look-up-table (LUT) in hierarchical steps is introduced that accounts for the varying instances of all variables of interest. Results show that anisotropic effects are strongest for early growth stages of the winter wheat canopy which influences also the retrieval of the variables. RTM inversions from off-nadir spectra lead to a decreased accuracy for the retrieval of LAI with a relative root mean squared error (rRMSE) of 18% at nadir vs. 25% (backscatter) and 24% (forward scatter) at off-nadir. For LCC estimations, however, off-nadir observations yield improvements, i.e., rRMSE (nadir) = 24% vs. rRMSE (forward scatter) = 20%. It follows that for a variable retrieval through RTM inversion, the final user will benefit from EnMAP time-series for biophysical studies regardless of the acquisition angle and will thus be able to exploit the maximum revisit capability of the mission.

Keywords: EnMAP; hyperspectral; PROSAIL; multi-angle; canopy; biophysical variables; agriculture; spectroscopy

1. Introduction

The retrieval of biophysical plant variables from optical imagery has been playing an important role in remote sensing and ecosystem modelling for more than 30 years. With ongoing technical progress of the sensors, there is also a steady demand for improved extraction of information from the gathered data. Especially in the agricultural context, many studies have pointed out the suitability of multispectral data (e.g., (Atzberger and Richter 2012; Bontemps et al. 2015; Campos-Taberner et al. 2016; Verrelst et al. 2014; Verrelst et al. 2015b)), hyperspectral data (e.g., (Atzberger et al. 2003; Burkart et al. 2015; Duan et al. 2014; Honkavaara et al. 2012; Locherer et al. 2015; Verger et al. 2011)) and a combination

of both (e.g., (Gevaert et al. 2015; Richter et al. 2009; Yang et al. 2013)) for an assessment of crop characteristics. In order to make these benefits available to modern farming, scientific tools and algorithms need to be directly applicable for a broader user community. Variables like the leaf area index (LAI) or leaf chlorophyll content (LCC) are of prime importance for a proper characterization of the canopy and plant biochemistry (Schueller 1992).

Several approaches are known to successfully retrieve hyperspectral canopy variables from measured spectra. The approaches can be divided into empirical and generic methods. The former builds up a statistical relationship between vegetation spectral signatures and in situ measured variables as parametric or non-parametric regressions (Verrelst et al. 2015b). Due to this site-specific linkage, empirical methods are not transferable in space or time (Baret and Buis 2008; Danner et al. 2016). To become independent of in situ measurements, more generic approaches often make use of radiative transfer models (RTMs). They are the intermediate link between biophysical characteristics of the canopy and its geometry, radiometric interaction and the reflected radiation (Verhoef 1984). Location, intensity and quality of the radiation source, atmosphere, vegetation / canopy, soil as well as position and properties of the sensor are important subsystems for the remote sensing of vegetation (Goel 1988). RTMs separate exterior parameters from the influence of the target itself, allowing quantitative analysis and the establishment of distinct relationships between signal and object variables (Verhoef 1984). One of the major improvements in RTMs was the incorporation of arbitrarily inclined leaves instead of a representation by plates. The resulting SAIL model (Verhoef 1984) (Scattering by Arbitrarily Inclined Leaves) was later coupled with the leaf optical properties model PROSPECT (Jacquemoud and Baret 1990) to form the new fusion model PROSAIL (Jacquemoud et al. 2009b).

In the direct or forward mode, PROSAIL simulates synthetic spectra from input variables that describe plant physiology and canopy architecture. In the indirect or inverse mode, these variables are obtained from spectral signatures. Inversion techniques are either based on optimization methods, artificial neural networks (ANN), machine learning algorithms (MLA) or look-up-tables (LUT) (see (Kimes et al. 2000) for overview). Their advantages and drawbacks vary with purpose of use. Optimization methods aim at minimizing deviations between modelled and measured spectra (Jacquemoud et al. 1994). Such minimization algorithms continuously change the input variables of the RTM until the modelled result matches the observation as closely as possible, leading to comparatively long computation times (Combal et al. 2003). ANNs and MLAs on the other hand are quicker in training and execution, but they require a priori information, calibration and lack of mathematical transparency (Liang 2007; Richter et al. 2009). Look-up-tables are databases of modelled spectra and their associated input parameter configurations. LUTs are known to be fast and robust methods producing reasonable results

(e.g., (Baret et al. 1999; Darvishzadeh et al. 2008; Locherer et al. 2015; Verrelst et al. 2014; Weiss et al. 2000)). In a first step, the LUT is built up in forward mode before it can be browsed in inverse mode. For the compilation of the LUT, the user has the choice of size (number of simulations), artificial noise type and noise level of the spectral model output as well as distribution type and constraints for all input parameters. Inversion of RTMs is impeded by the fact that more than one combination of variables can lead to the same model result. This effect has become known as equifinality or ill-posed problem and is dealt with either by restriction of the input range or by inclusion of the n-best performing results rather than just considering the number one fit (Combal et al. 2003).

For an ideal analysis, communication between sensor and model must be optimal. Since spectral models have been developed in the laboratory with the help of ground-based spectrometers, they basically are of hyperspectral nature. In order to use the models in combination with multispectral data, their spectral resolution normally is toned down using the spectral response functions of the respective instruments. Using hyperspectral data as input allows for making full use of the quasi-contiguous narrowband output of the RTM in forward mode. The retrieval methods tested in this article thus are intended to be applied on hyperspectral data of the Environmental Mapping and Analysis Program (EnMAP). EnMAP is a German spaceborne imaging spectroscopy scientific mission carrying the EnMAP Hyperspectral Imager (HSI) instrument (Guanter et al. 2015a). Currently under development, EnMAP-HSI will deliver data at high spectral resolution of 6.5 nm in the VNIR and 10 nm in the SWIR

domain which together cover the full spectral range of 420 to 2450 nm (Guanter et al. 2016a). Competition for actual data is expected to be intense, since the data take capacity of EnMAP is limited and—as of today—only the Italian hyperspectral mission PRISMA (Loizzo et al. 2016) may be going to record comparable data by the estimated time of launch in 2019. Repeat cycles of 23 days in quasi-nadir mode will limit the availability of cloud-free scenes (Storch et al. 2013). To mitigate this problem, the satellite platform will be capable of a max. ±30° across-track tilt, allowing side looks upon the target with revisit times of up to 4 days (Stuffler et al. 2009) near earth's equator or even less for latitudes of central Europe, e.g., 2.5 days for Munich, Germany (Storch et al. 2013). The effects of this off-nadir pointing for the retrieval of biophysical variables have not been tested in the EnMAP context. Therefore, the objectives of this study are (1) to demonstrate the expected impact of the EnMAP-specific sun–target–sensor-geometry (s–t–s-geometry) on reflectance spectra, (2) to quantify the effect on agriculturally relevant variable retrievals, such as LAI and leaf pigments and (3) to introduce a new hierarchical LUT approach for an optimized retrieval of these parameters.

2. Materials and Methods

2.1. Study Area & Sampling Layout

The study area is located in the North of Munich, Bavaria, in Southern Germany. Two study sites at 48°17′31.25″N, 11°42′21.53″E (field 517) and 48°14′51.46″N, 11°42′24.10″E (field 509) were visited regularly during two field campaigns. Both fields are part of communal farmland belonging to the city of Munich. Each was cultivated with winter wheat (triticum aestivum), representing the dominant cereal crop in the area and situated within 1.5 km distance to the Isar river. The average cloud cover in the Munich-North Isar (MNI) region was a bit higher than usual (5.68 instead of 5.44 okta). This indicates the difficulty of recording spectra on a frequent basis, since adequate clear sky conditions occurred only on few occasions. Information on the site management was provided via personal communication by the farm managers; i.e., dates of seeding, fertilization methods and quantities, harvesting dates, etc.

During the first campaign, data were collected at almost weekly intervals from 17 April to 25 July 2014 (14 sampling dates). The second campaign period already started in autumn. Measurements were conducted from 28 November 2014, to 21 July 2015 (13 sampling dates). In this way, the complete growing cycle of the crop from seeding to harvest could be observed. One elementary sampling unit (ESU) was defined as a 10 m × 10 m pixel size. The measurements were then related to a 3 × 3 ESU raster with equal distances of 10 m. All nine ESUs were marked with sticks and revisited for each sampling date. The row azimuth direction of the winter wheat crops was 170°/350° for 2014 and 150°/330° for the 2014/15 season, with the angular definition of 0° = N. Table 2.4.1 shows the complete list of sampling dates.

Table 2.4.1. Dates of field visits and corresponding availability of spectral observations nadir/angles as well as crop variable measurements, indicated with check marks. Crop variables were measured weekly for both growing cycles of 2014 (left) and 2014/2015 (right). The frequency of spectral observations was subjected to weather conditions, since reflectance measurements require a cloud-free sky.

Date	Spectral Nadir	Spectral Angles	Crop Variables	Date	Spectral Nadir	Spectral Angles	Crop Variables
17 April 2014	√	√	√	28 November 2014			√
23 April 2014	√	√	√	12 December 2014	√	√	√
30 April 2014			√	19 March 2015	√		√
6 May 2014			√	10 April 2015	√	√	√
14 May 2014			√	22 April 2015	√	√	√
9 May 2014	√	√	√	5 May 2015			√
26 May 2014			√	8 May 2015	√	√	√
2 June 2014	√	√	√	3 June 2015	√	√	√
6 June 2014	√	√	√	12 June 2015	√		√
18 June 2014	√	√	√	1 July 2015	√	√	√
26 June 2014	√	√	√	10 July 2015	√	√	√
3 July 2014	√		√	16 July 2015	√		√
17 July 2014	√		√	21 July 2015	√	√	√
25 July 2014	√	√	√				
Total observations	10	8	14	Total observations	11	8	1

2.2. In Situ Measurements

2.2.1. Spectral Data

Spectral data were collected with an Analytical Spectral Devices Inc. (ASD) FieldSpec 3 Jr. Five separate measurements were carried out per ESU and per observation angle to obtain representative values. Outliers were removed and the remaining spectra were averaged and subjected to further processing. The post-processing included splice correction, radiometric calibration to absolute reflectance values and smoothing with a moving Savitzky-Golay-filter (Savitzky and Golay 1964). Apart from nadir measurements, the canopy was also measured under observer zenith angles (OZA) of +30° and −30° regarding the solar plane: a sensor inclination towards the sun is defined as a positive OZA, whereas an inclination away from the sun is described as a negative OZA (Figure 2.4.1). Due to the *backscatter* effect, spectra with positive OZA are noticeably brighter than nadir views or negative zenith angles, as they draw nearer to the spot of increased backscatter, also known as the hot spot (Hapke 1986). The opposite direction shall accordingly be called cold spot or *forward scatter* and usually leads to reduced reflectances and darker images. For the angular spectral measurements, a microphone stand was modified to hold the ASD glass fiber optic. The horizontal rod of the stand could be raised or lowered to adjust the viewing angle with help of an attached inclinometer. The observer azimuth angles (OAA) matched up with the row azimuth angle of the canopy stands (170° for field 517 and 150° for field 509). EnMAP will operate on a sun-synchronous orbit with 97.96° satellite inclination angle descending node (Kaufmann et al. 2012) which corresponds to an OAA of 187.96°. The angular effects measured in the presented campaign therefore are assumed to adequately represent the angular effects expected from future EnMAP data.

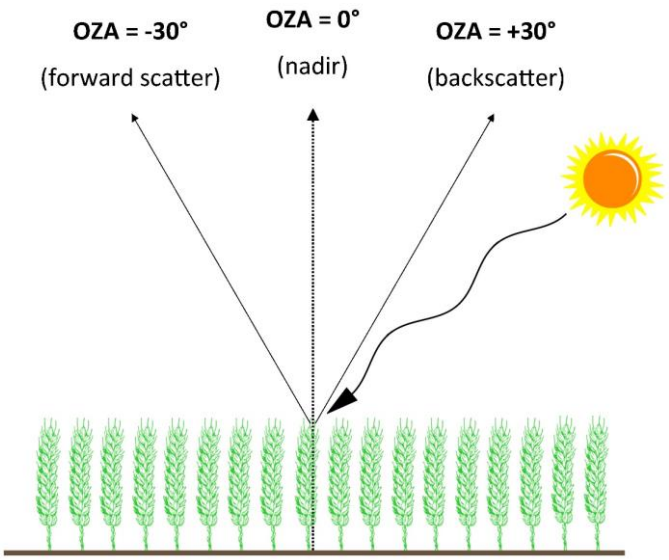


Figure 2.4.1. Sun–target–sensor-geometry. The three arrows illustrate the three different observer zenith angles (OZA). A positive OZA is associated with backscatter and commonly shows higher rates of reflectance than negative OZAs (forward scatter).

Spectral information from each of the 3 × 3 ESUs was compiled to pseudo-images with a ground sampling distance of 10 m. Furthermore, the processed signatures were converted into simulated EnMAP spectra via the EnMAP-end-to-end-Simulator (EeteS) (Segl et al. 2012a). In this process, the sensor-specific radiometric and spectral properties were adapted. The spatial resolution in this case was retained at 10 m to preserve the data population.

The gap fraction is a measure for the probability of a ray of light to penetrate through the canopy undisturbedly (Baret et al. 1995). Accordingly, this parameter decreases with density and/or height of a canopy. Canopy height can also be seen as a path length on which energy can interact with plant traits.

Assuming identical canopy height, the path length is shortest for nadir views and increases with OZA > 0°. For a 30° deflection from nadir, the travelled path is longer by factor cos (30°)$^{-1}$ which is 15.5%. This leads to a weaker influence of soil background and to an apparently higher portion of visible leaf surface.

The anisotropy factor (ANIF) (Sandmeier et al. 1998) yields useful information about the sensitivity of different wavebands towards changes in illumination geometry. It is simply defined as Equation (1):

$$\mathrm{ANIF} = \frac{\mathrm{R_{off-nadir}}}{\mathrm{R_{nadir}}} \quad (1)$$

Since the experimental setup covers two different viewing directions, two ANIFs were obtained: one for forward scatter (ANIFfs) and one for backscatter (ANIFbs). Additionally, a third ANIFfs/bs was calculated as the ratio between reflectances per waveband in forward and backscatter direction.

If spectral information of the same target is available for multiple angles, it was found useful to combine them and thus raise predictability. This has been done with surface-near spectrometers that are handheld (Wang et al. 2013), mounted on a tower (Hilker et al. 2011), on hemispherical devices (Tripathi et al. 2012) or UAVs (Burkart et al. 2015).

A prominent example for multi-angular optical remote sensing from space is CHRIS/Proba which allows to record narrowband spectra in the VNIR-domain from five different viewing angles

(e.g., (Verhoef and Bach 2007)). EnMAP will be able to perform an across-track satellite tilt, but will keep up this slanting position for longer time than its view upon the target. If the same target shall be observed under different zenith angles, more than just one acquisition will have to be made with a time gap of at least several days or possibly several weeks or months. Canopy parameters that are strongly influencing the bidirectional reflectance distribution function (BRDF) change diurnally as well as during the seasonal growth cycle. For this reason, in this study we concentrated on single looks only, regardless of the possible improvement of results for a *combined* multi-angular approach.

2.2.2. Biophysical Variables

Agricultural crop variables were measured at the exact same location where spectral signatures were recorded shortly before. The time offset between the variable sampling and the spectral sampling was 45 min on average and 60 min at maximum.

Average Leaf Inclination Angle (ALIA) was measured with a Suunto PM-5/360 inclinometer held along the leaf petiole to display its slope against the horizontal plane (Gratani and Ghia 2002). The measurement was repeated at different positions of the leaf and for different leaves within the canopy. Additionally the Leaf Inclination Distribution Function (LIDF) was noted down for a more detailed description of the canopy architecture (Goel and Strebel 1984). Leaf chlorophyll content (LCC) was measured with a Konica-Minolta SPAD-502 handheld device at different heights with focus on the upper canopy layer. The chlorophyll meter had been individually calibrated in a preceding field campaign against destructive measurements of winter wheat leaf chlorophyll content from different senescence states. Coefficients of (Lichtenthaler 1987) were used to derive LCC from the samples. Leaf senescence (C_{br}) was estimated as the fraction of brown leaf parts in the foliage. This variable varies between zero (no brown spots = 100% fresh vegetation) and one (no green spots = 100% senescent vegetation). For a proper estimation of C_{br}, the approach of (Verhoef and Bach 2003) was slightly adapted to incorporate the non-linearity of the vertical distribution of brown leaves. The dissociation factor between upper and lower layer thus created consistent results. This was achieved by applying a cosine function of the brownness in the upper layer to the power of two. C_{br} can be written as Equation (2):

$$\mathrm{C_{br}} = \frac{\mathrm{br_u} + \mathrm{br_l}}{2} - \frac{|\mathrm{br_u} - \mathrm{br_l}|}{2} \cdot \cos^2(90° \cdot \mathrm{br_u}) \quad (2)$$

with br_u as the fraction of brown leaf parts in the upper and br_l in the lower layer of the canopy. For LAI measurements, a LI-COR Biosciences LAI-2200 instrument was used that had been upgraded

with the ClearSky Kit to obtain functionalities of the advanced LAI-2200C. Equipped with a GPS sensor and a white diffuser cap, the device allows for nondestructive measurements of leaf area index under sunlight conditions. To obtain green LAI, the measured LAI value was multiplied with the factor $1 - C_{br}$ to exclude the impact of non-photosynthetic vegetation on LAI measurements. Multiplication of leaf variables with the LAI value allows their interpretation on canopy level, e.g., canopy chlorophyll content (CCC).

2.3. Radiative Transfer Modelling

For this study, PRO4SAIL-5B (PROSPECT 5B + 4SAIL) was used which operates based on the input parameters listed in Table 2.4.2 and described in Section 2.2.2:

Table 2.4.2. Overview of the PROSAIL parameter ranges for creation of the look-up-table. Biophysical parameters were varied via uniform distribution, with typical min & max values. Parameters that control the sun–target–sensor-geometry were varied in discrete steps, covering all conditions that were observed in the field.

Model	Parameter	Description	Unit	Min	Max	
PROSPECT	N	Leaf structure parameter	-	1.0	2.5	
	LCC	Leaf Chlorophyll$_{a+b}$ content	µg cm^{-2}	0.0	80	
	LCarC	Leaf Carotenoids content	µg cm^{-2}	0.0	20	
	EWT	Leaf Equivalent Water content	cm	0.001	0.05	
	LMA	Leaf Mass per Area	g cm^{-2}	0.001	0.02	
	C_{br}	Fraction of brown leaves	-	0.0	1.0	
SAIL	LAI	Leaf Area Index	m^2 m^{-2}	0.0	8.0	
	ALIA	Average Leaf Inclination Angle	deg	20	90	
	Hspot	Hot Spot size parameter	-	0.01	0.5	
	Skyl	Ratio of diffuse and total incident radiation	-	0.1	0.1	
	γ	Soil Brightness Parameter	-	0.0	1.0	
Model	Parameter	Description	Unit	Min	Max	Divisions
SAIL	SZA	Sun Zenith Angle	deg	30	55	6
	OZA	Observer Zenith Angle	deg	−30	30	3
	rAA	relative Azimuth Angle	deg	0	65	14

Following the suggestion of (Atzberger and Richter 2012), the Skyl-parameter was kept stable at 0.1. The soil brightness parameter scales the dominance of the bright and dark canopy background in the output signal. By default, standard literature soils are used for this. In this study they were replaced by the brightest and the darkest soil spectrum of the campaign, measured directly at the study fields for each date. The background signal gains more weight in the simulated reflectance for vegetation that is sparse in terms of green LAI. It is important to note that spectral signatures of senescent canopies differ from those of small plants that cover the soil only partially, although situations might result in the same low value for green LAI. For this reason, another background type is introduced that was calculated as the mean senescence signal for ripe wheat crops of both seasons. All other leaf and canopy parameters were randomly drawn from uniform distributions with min and max values adjusted according to Table 2.4.2. Input parameters regulating the s–t–s-geometry uniformly covered all field scenarios. For example, the minimum SZA observed in the field was 29.14° and the maximum was 52.63°. As a result, SZAs of 30°, 35°, 40°, 45°, 50° and 55° were used for generating the LUT. Variations in the OZA of −30°, 0° and +30° took account of the three experiments of the simulated EnMAP platform tilt. For winter wheat crops it was suggested setting the leaf structure parameter N to a mean of 2.0 with a SD of 0.34 (Atzberger et al. 2003). Since our study data covered the complete vegetation cycle from seeding to harvest, these values were slightly adapted to a wider range of 1.0 to 2.5.

The size of each LUT (n_{lut}) can be understood as the number of variations of the parameters (n_{para}) multiplied by the number of variations of the s–t–s-geometry (n_{angles}). n_{lut} has a linear influence on the calculation time for the generation and the inversion of the LUT. On the other hand, larger LUT sizes yield more possible parameter constellations, which may improve the quality of the retrieval. Many

authors suggest n_{para} = 100,000 as the best trade-off between calculation time and inversion accuracy (e.g., (Darvishzadeh et al. 2012; Duan et al. 2014; Richter et al. 2009)). In each of these studies, however, angles of sun and observer were fixed. As described in Table 2.4.2, n_{angles} here needed to cover 252 different geometric constellations which would result in n_{lut} = 25,200,000 for each soil and senescence background. A LUT-size of 12,600,000 (n_{para} = 50,000) turned out to perform equally well (Δ_{RMSE} < 1%), while allowing a quicker inversion and thus the conduction of more experiments in the same period of time. Accounting for the different potential background signals (soil or senescent material respectively) the LUTs are duplicated and only varied by a different background signal. This method, therefore, shall be called duplex LUT.

Finally, artificial noise can be applied to make simulated spectra more realistic and to improve the inversion accuracy (overview given by (Locherer et al. 2015)). The best performing LUT settings have been varied in noise type (Gaussian additive/Gaussian inverse multiplicative) and noise level (0.0%, 0.1%, 1.0%, 2.0%, 5.0%, 10.0%) respectively.

2.4. Step-Wise Inversion of the LUT

Inverting a LUT means comparing measured spectra with all PROSAIL model results and selecting the parameters that led to the best performing LUT members. Different cost functions can be used to quantify the agreement between measurement and model result. Most authors use the Root Mean Squared Error cost function type ($RMSE_{cft}$), defined as Equation (3):

$$\mathrm{RMSE_{cft}} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\mathrm{R_{measured}}(\lambda_i) - \mathrm{R_{simulated}}(\lambda_i)\right)^2} \quad (3)$$

By squaring the distances before extracting the root, larger deviations gain more influence in this term. Consequently, the $RMSE_{cft}$ favors results for which both spectral signatures match rather closely for all wavelengths. An alternative cost function tested is the Nash-Sutcliffe-Efficiency (NSE_{cft} (Wainwright and Mulligan 2005)) as defined in Equation (4),

$$\mathrm{NSE_{cft}} = 1 - \frac{\sum_{i=1}^{n}\left(\mathrm{R_{measured}}(\lambda_i) - \mathrm{R_{simulated}}(\lambda_i)\right)^2}{\sum_{i=1}^{n}(\mathrm{R_{measured}}(\lambda_i) - \overline{\mathrm{R}}_{\mathrm{measured}})^2} \quad (4)$$

Weighing the squared sum of distances between measured and simulated reflectance against the squared sum of distances between measured reflectance and the average measured reflectance of the complete spectrum. The mathematically simplest approach is the mean absolute error (MAE) as defined in Equation (5):

$$\mathrm{MAE} = \sum_{i=1}^{n} \left|\mathrm{R_{measured}}(\lambda_i) - \mathrm{R_{simulated}}(\lambda_i)\right| \quad (5)$$

In all three cases, $R_{measured}(\lambda_i)$ is the measured and $R_{simulated}(\lambda_i)$ is the modelled reflectance at wavelength λ for the i^{th} spectral sensor band, whereas n corresponds to the total number of bands used for the optimization.

For each sampling date, the s–t–s-geometry must be known. Prior to application of the cost function, the correct sub-LUT must be selected. At first, by analysis of the observed spectrum, the corresponding LUT is inquired, depending on the expected canopy background. Senescent vegetation does not only show distinct absorption features by leaf pigments, but also significant features in the SWIR domain. A new index that has been optimized for the EnMAP spectral configuration, the $NPVI_{EnMAP}$ (Equation (6)) is introduced. $NPVI_{EnMAP}$ is used to classify the background of a pixel as either soil (type A) or senescent vegetation (type B) based on a simple threshold.

$$\mathrm{NPVI_{EnMAP}} = \frac{\mathrm{R_{2218}}}{\mathrm{R_{671}}} \quad (6)$$

If $NPVI_{EnMAP} < 1.4$, the spectrum is classified as type B and classified as type A for all other cases. Based on the angular constellation for each pixel a decision is made, which of the remaining 252 sub-LUTs shall finally be used for the inversion.

The ill-posedness can be mitigated by narrowing the parameter constraints for the generation of the LUT. In this case, the user needs to have access to a priori information about the expected data range. These constraints make the approach more empirical and thus inconsistent with the proposed generic conviction of the study. For this reason, the ill-posed problem was dealt with by considering more than just the one best performing LUT member and its according parameter configuration (Verrelst et al. 2014). The final results vary with the number of considered best fits (n_{bf}). A tradeoff between singular (ill-posed) and multiple (over-balanced) solutions needed to be found for an optimal retrieval setup ($n_{bf} \in \{1, 20, 50, 100, 200, 500, 1000\}$). For $n_{bf} > 1$, the median is used to get the final parameter value. Figure 2.4.2 illustrates the necessity to include an adequate amount of LUT-members for the variable retrieval. Parameter constraints for the creation of the LUT can be narrowed down to further increase inversion performance (e.g., (Tripathi et al. 2012)). In doing so, the model is calibrated to site-specific characteristics and might not be able to help retrieve variables for other fields, crop types or phenological stages.

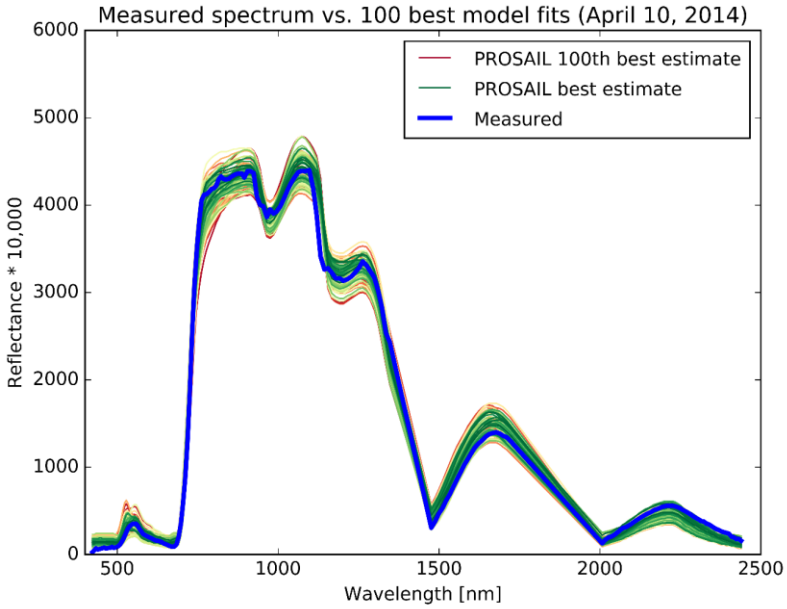


Figure 2.4.2. Impact of the choice of number of best fits for the retrieval accuracy. The measured winter wheat spectrum was obtained on 10 April 2014, with an Analytical Spectral Devices (ASD) FieldSpec 3Jr and then converted into pseudo-EnMAP reflectances. The other signatures are the closest 100 members of the LUT, as simulated by the PROSAIL model. The best estimate, i.e., the model run with least distance to the measured spectrum, is drawn in green. With increasing statistical distance the colors fade from green to yellow until the 100th best estimate is finally plotted in red.

The step-wise hierarchical variable retrieval was achieved by several consecutive complete LUT-inversions. A motivation for this approach is the dominance of some parameters (e.g., LAI) that may suppress the signal of others (e.g., LCC) affecting similar spectral domains. For the first inversion run, all available spectral bands were included except for those influenced by the atmospheric water vapor absorptions (1359 nm – 1465 nm and 1731 nm – 1998 nm) and the VNIR-bands in the detector overlap of the EnMAP-HSI (911 nm – 985 nm). Although all variables were obtained in this first step, LAI was the only one of interest at that time. The average inclination of leaves is an important regulation parameter that describes the visibility of photosynthetically active parts of the vegetation for the sensor. Erectophile canopies reveal larger parts of the underlying soil, especially for low SZAs. Planophile and plagiophile canopies on the other hand cover more of the background and lead to stronger signals just like an increased LAI would. ALIA and LAI therefore counterbalance each other. In an attempt to

separate their influences on the measured spectra, another pre-selection is investigated for the first inversion run, selecting only those LUT members with an ALIA close to the one estimated in the field.

For the second inversion run, the LAI values resulting from the first run were fixed. A pre-analysis selected only those LUT members containing the retrieved LAI ± an absolute tolerance of 0.01 (m^2 m^{-2}). If this pre-selection left fewer members than two times the size of n_{bf}, the valid tolerance was expanded by increments of 0.01 until the minimum condition was met. Only then the second run was started during which LCC was retrieved by separately applying the cost function to the variable-specific sensitive wavelengths (LCC @ 423–705 nm).

According to the authors of (Richter et al. 2012) the performance of an inversion setting shall ideally be assessed by multiple statistical quality criterions when comparing retrieved model parameters to in situ measured values. Most importantly, the relative Root Mean Squared Error (rRMSE) and the slope of the regression line (m) were considered in this study. The coefficient of determination (R^2) played a minor role, since it measures the strength of the correlation according to the linear regression rather than the 1:1 relationship between model parameter and in situ variable. A regression model was calculated nonetheless and its slope served as an indicator of the inversion accuracy. A slope of 1.0 suggests a perfectly outbalanced relationship. Slopes > 1.0 reveal an underestimation for low and overestimation for high values. The reverse relationship applies for slopes < 1.0. For all following analyses, model runs with a slope < 0.7 or > 1.3 were not considered in the final results.

3. Results

3.1. Impact of the Observer Zenith on Reflectance Spectra

Analyzing all 3 × 3 images separately would be time consuming and impractical. For this reason, the spatial dimension was partially sacrificed in favor of an additional temporal dimension by mosaicking all images of one growing season below each other. This principle is explained in Figure 2.4.3. The combination of several sub-images in one mosaic allows the application of algorithms for all observations at once and visualizes seasonal changes in reflectance.

A strong dependency of the anisotropy factor (ANIF, Equation (1)) towards wavelength can be observed (see Figure 2.4.4). A striking anisotropic behavior of the canopy is evident in the short wave visible range for both illumination settings, with decreasing impact towards longer wavelengths. The experiment was repeated for three different growth stages: (1) Day of year (DOY) 128: stem elongation (flag leaf visible, but still rolled), (2) DOY 182: development of fruit (late milk) & (3) DOY 202: Ripening (fully ripe). Each of these three growth stages is represented by an average of all adjacent canopy spectral signatures of that specific date. For both forward and backscatter the ANIF is highest for phenological stage (1). This indicates that during earlier growth stages angular effects in the observer zenith have the strongest influence on the spectral signal. An ANIF of value 1.0 means identical reflectance for nadir and off-nadir observation. Reflectances of the fully ripe canopy (3) draw closest to this equilibrium line for both viewing directions. $ANIF_{fs}$ marginally drops below 1.0 for the VIS-domain and from 1500 nm to 2500 nm at growth stage (2), due to lower reflectances in the off-nadir compared to nadir observations. For backscatter, $ANIF_{bs}$ of (2) is situated mostly between (1) and (3). Lowest impact of angular variations can be assumed for the visible range for forward scatter and for the NIR plateau for backscatter observations. For green vegetation, there is a local minimum in $ANIF_{bs}$ around 550 nm. On the other hand, maximum anisotropy is observed for shortest (450 nm) and longest (680 nm) visible wavelengths. $ANIF_{fs}$ shows an anisotropic behavior that is exactly reverse to $ANIF_{bs}$.

In the special case of $ANIF_{fs/bs}$, differences between negative and positive observation angles can be assessed directly. As expected, backscatter angles generally lead to higher reflectances. The 1.0 line is closest for the NIR plateau of green vegetation signatures. For all three settings, the senescent vegetation had a constant ANIF that was less sensitive to wavelength, but more sensitive to viewing direction.

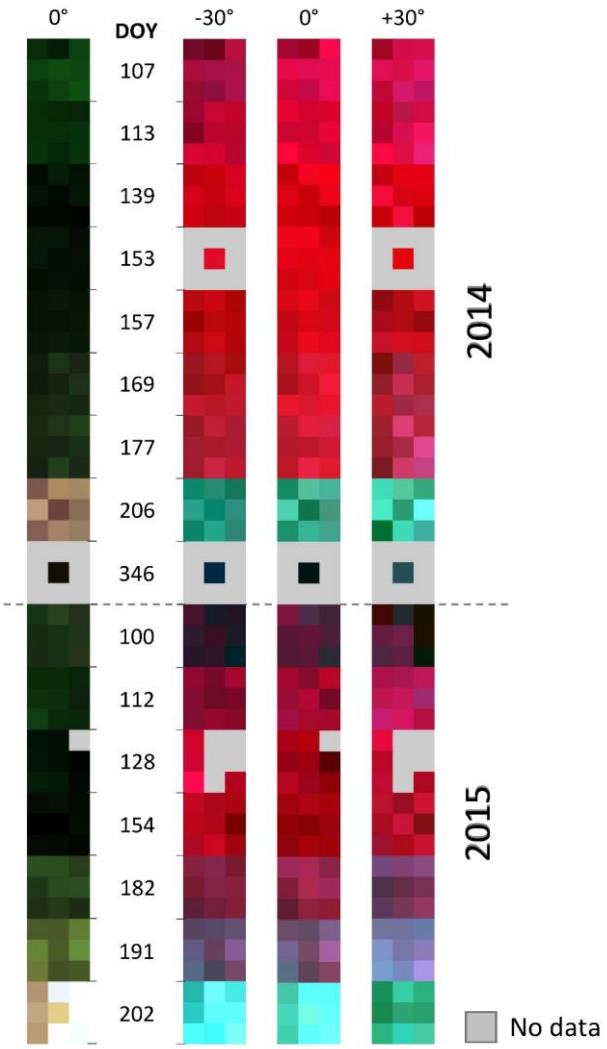
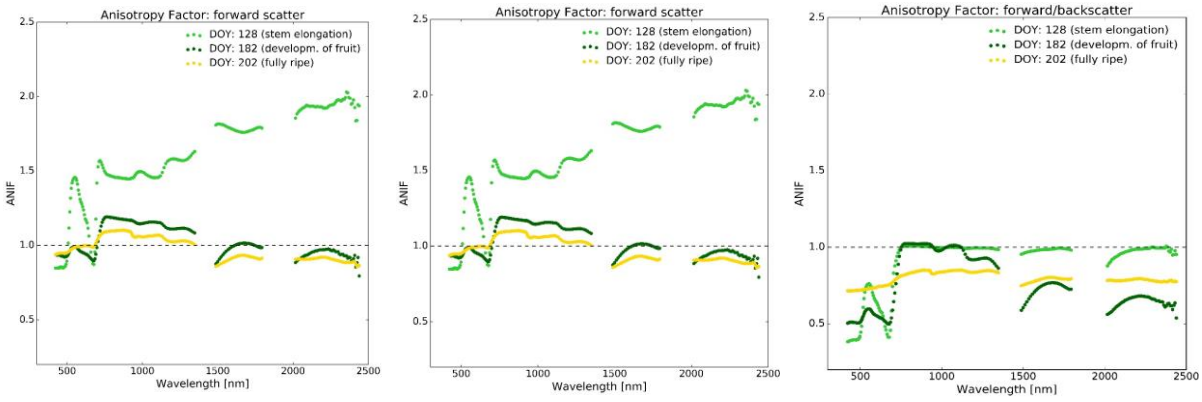


Figure 2.4.3. Red-green-blue (RGB) composite imagery (left) and colored infrared (right) illustration of the spectral image mosaic (standard deviation stretch n = 3.0). Each of the stripes represents the same area of interest under a different observer zenith angle (OZA). OZA = −30° is associated with forward scatter, OZA = 0° with nadir and OZA = +30° with backscatter observations. The stripes are composed of 16 sub-images of 3 × 3 pixels, each representing a different field date (nine in 2014 and seven in 2015), as indicated by the Julian day of year (DOY).



(**a**) (**b**) (**c**)

Figure 2.4.4. Illustration of the Anisotropy Factor (ANIF) for three different phenological stages of winter wheat (early: bright green, medium: dark green, late: yellow) and observation angles: ANIF for forward scatter ((**a**) $ANIF_{fs}$), backscatter ((**b**) $ANIF_{bs}$) and the off-nadir ratio ((**c**) $ANIF_{fs/bs}$).

3.2. Impact of the Observer Zenith on the Retrieval of Crop Parameters

Results for the retrieval of LAI and LCC from different observation angles and two different cost functions are shown in Table 2.4.3. For LAI, the lowest rRMSE of 0.18 was obtained for nadir observation with MAE cost function. For OZA = −30° (forward scatter) rRMSE was 0.24 and 0.25 for OZA = +30° which means an error increase of 30%. The associated scatterplots are shown in Figure 2.4.5. Subplot (c) illustrates the instable model inversion for LAI from backscatter spectra. Especially for medium LAI values, PROSAIL suggested a widespread variable range and better results for nadir inversion. LCC retrieval on the other hand improved for negative observation angles (rRMSE = 0.20), as the clusters of high and low chlorophyll content moved closer to the 1:1 line. For the lower reflectances in opposite viewing direction, rRMSE = 0.27 was the best possible result. Slope and intercept of the regression line nearly reached the optimum of f(x) = x, but the scattergram shows a non-linear behavior and a reduced rRMSE.

Table 2.4.3. Influence of angular spectral measurements, representing the tilt of the EnMAP satellite platform of ±30° across track (OZA). Model inversion was conducted with relative root mean squared error $(RMSE)_{cft}$ and mean absolute error (MAE) and different sizes of n_{bf}. Best results for leaf area index (LAI) and leaf chlorophyll content (LCC) for each angle and cost function type are shown.

OZA	RMSE Cost Function				MAE Cost Function			
	LAI		LCC		LAI		LCC	
(deg)	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE
−30	0.81	0.27	0.95	0.20	0.84	0.24	0.82	0.22
0	0.94	0.19	0.87	0.24	0.92	0.18	0.89	0.26
+30	0.83	0.25	0.94	0.27	0.82	0.27	0.79	0.28

Figure 2.4.6 compares the map of in situ variables as measured at the study fields with the results of the inversion for LAI and LCC. Discrete steps in hue and saturation can be found in both in situ values and model results, indicating that the seasonal dynamics were captured in general. Within the 3 × 3 pixels for each field date, variations could not be fully reproduced by the inversion. For this, model uncertainties would have to be lower than the lateral disparity for the 30 × 30 m sub-plots.

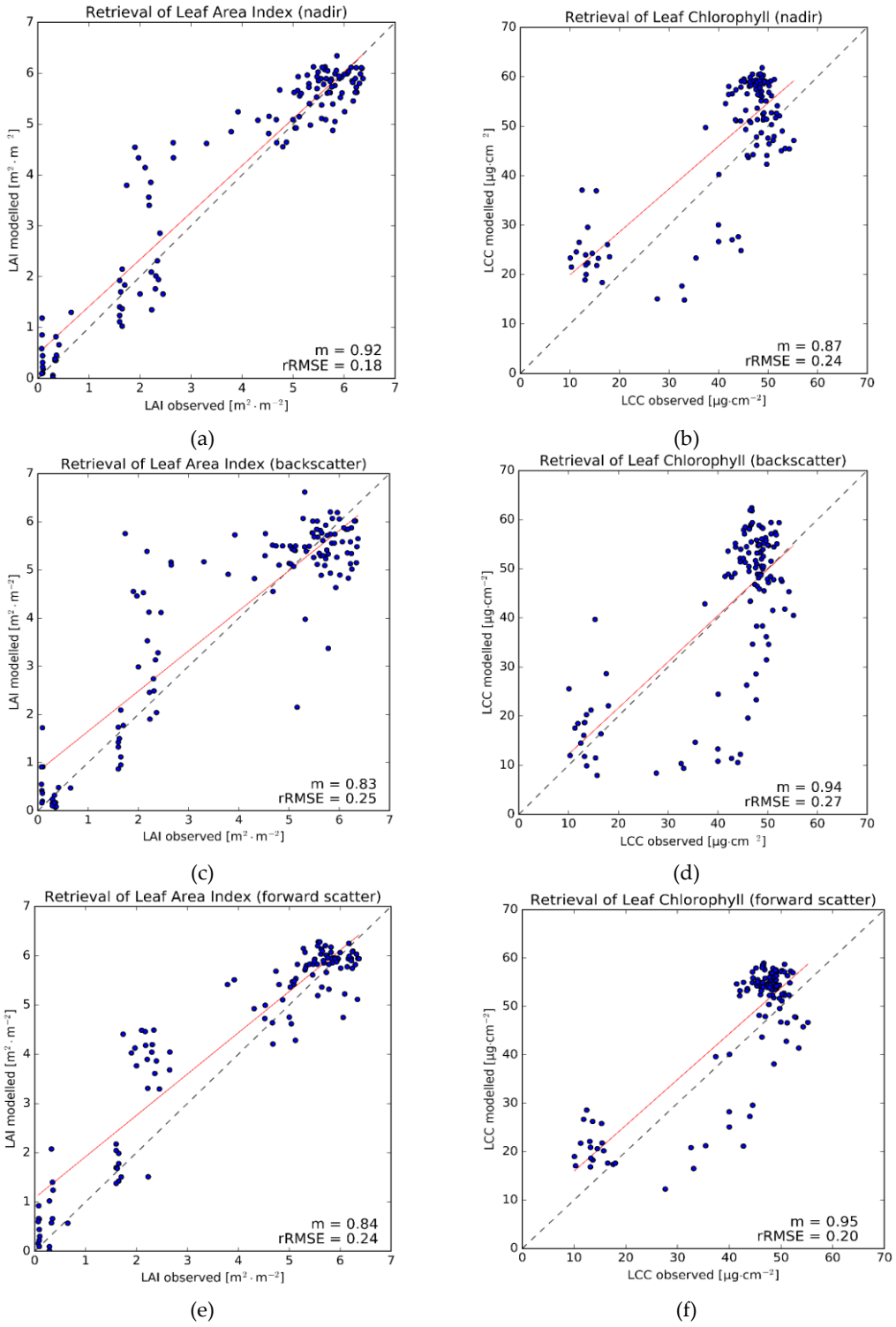


Figure 2.4.5. (**a–f**) Evaluation of best inversion results for LAI (left column) and LCC (right column). Nadir is displayed in the top row, backscatter (OZA = +30°) in the middle, forward scatter (OZA = −30°) in the bottom row. The slope of the regression line is indicated as *m*.

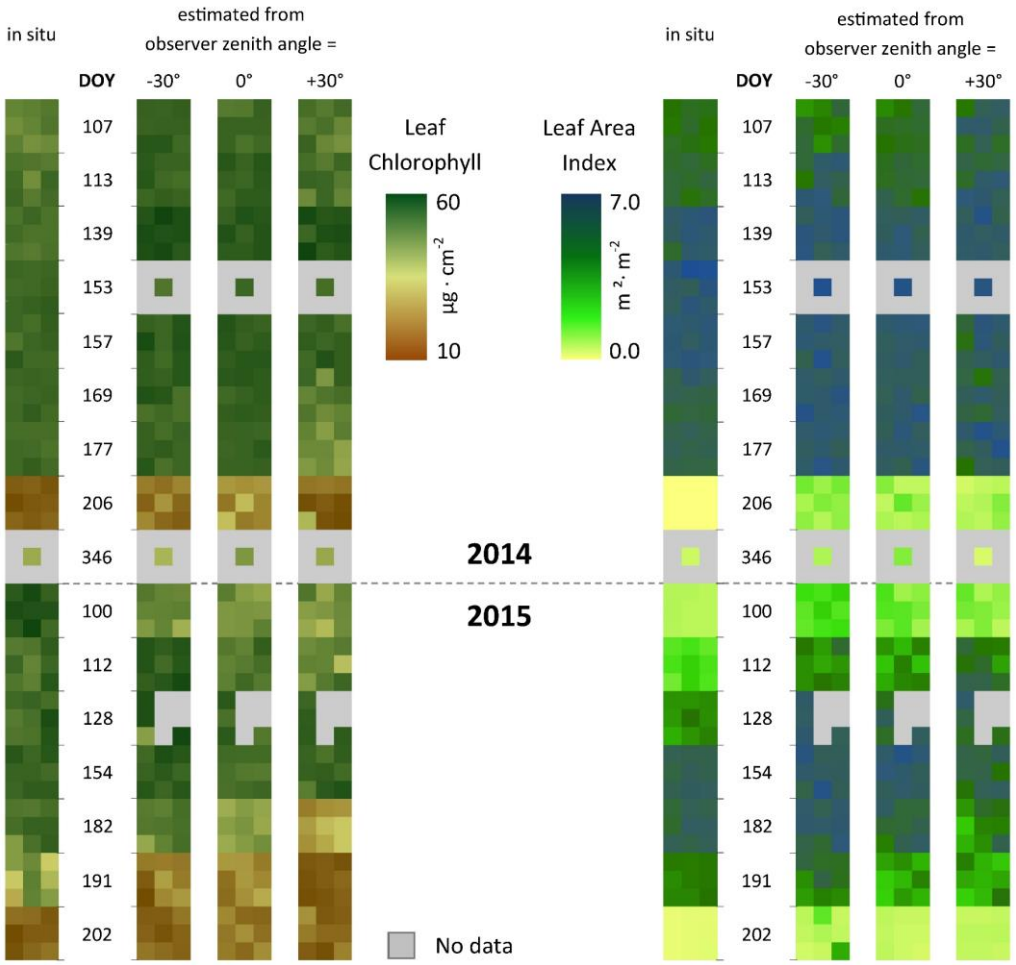


Figure 2.4.6. Spatial distribution of measured and estimated LCC (**left**) and LAI (**right**) for the two growing seasons of 2014 and 2015 under different observation angles.

By subtracting parameter estimations from in situ measurements, residuals for the inversion of LAI and LCC are obtained. Figure 2.4.7 is an illustration of these residuals as a map. For the first season in 2014, the model overestimated LCC most of the time for all angular settings, especially for the beginning of the season in April and May. In April and for one sampling date in July 2015, higher chlorophyll concentrations were measured than predicted by the model. The latter is especially striking, because estimations in nadir were quite close the week before (ΔLCC_{DOY182} = 2.56 μg cm^{-2}) and after (ΔLCC_{DOY202} = 4.24 μg cm^{-2}). Interannual changes in model predictability are represented by the mean standard deviation over all 9 pixels and per season. The results are listed in Table 2.4.4. Both LAI and LCC were more homogeneously predicted in 2014, whereas in 2015 residuals tended to be more variable throughout the growing season. Model inversions from nadir observations were 38% less prone to seasonal effects than those from angular observations. The pattern of the residuals for LAI and LCC do not show any relation and statistics suggest that they are independent of each other (R^2 = 0.10). Residuals of LAI followed a normal distribution (p > 0.05 for all OZA), but those of the LCC estimation did not (p < 0.01 for all OZA). A shift to negative residuals, i.e., an overestimation of LCC by the model, indicated a slight systematic bias.

Retrieval of CCC proved quite successful (see Figure 2.4.8). Concentrations of LCC obtained from the second inversion run were multiplied with estimated LAI values and compared to in situ measured CCC (i.e., measured LAI × measured LCC). PROSAIL tended to overestimate CCC with an intercept of 0.31 g cm^{-2} for nadir spectra. The relative rRMSE = 0.37 was higher than for both, LAI and LCC. From forward scatter observations, however, CCC was estimated with an rRMSE of 0.33 meaning an improvement of 12%. Retrieval from backscatter spectra suffered from a weaker estimation of LAI and LCC, leading to a decrease of relative RMSE of 0.40 accordingly.

Table 2.4.4. Mean deviations between estimated and in situ measured LAI and LCC, separated by season and angular setting (OZA).

OZA	Season 2014		Season 2014/15	
	LAI	LCC	LAI	LCC
(deg)	(m^2 m^{-2})	(µg cm^{-2})	(m^2 m^{-2})	(µg cm^{-2})
−30°	0.62	8.43	0.99	7.44
0°	0.47	11.86	0.82	7.17
+30°	0.59	7.38	1.08	11.22

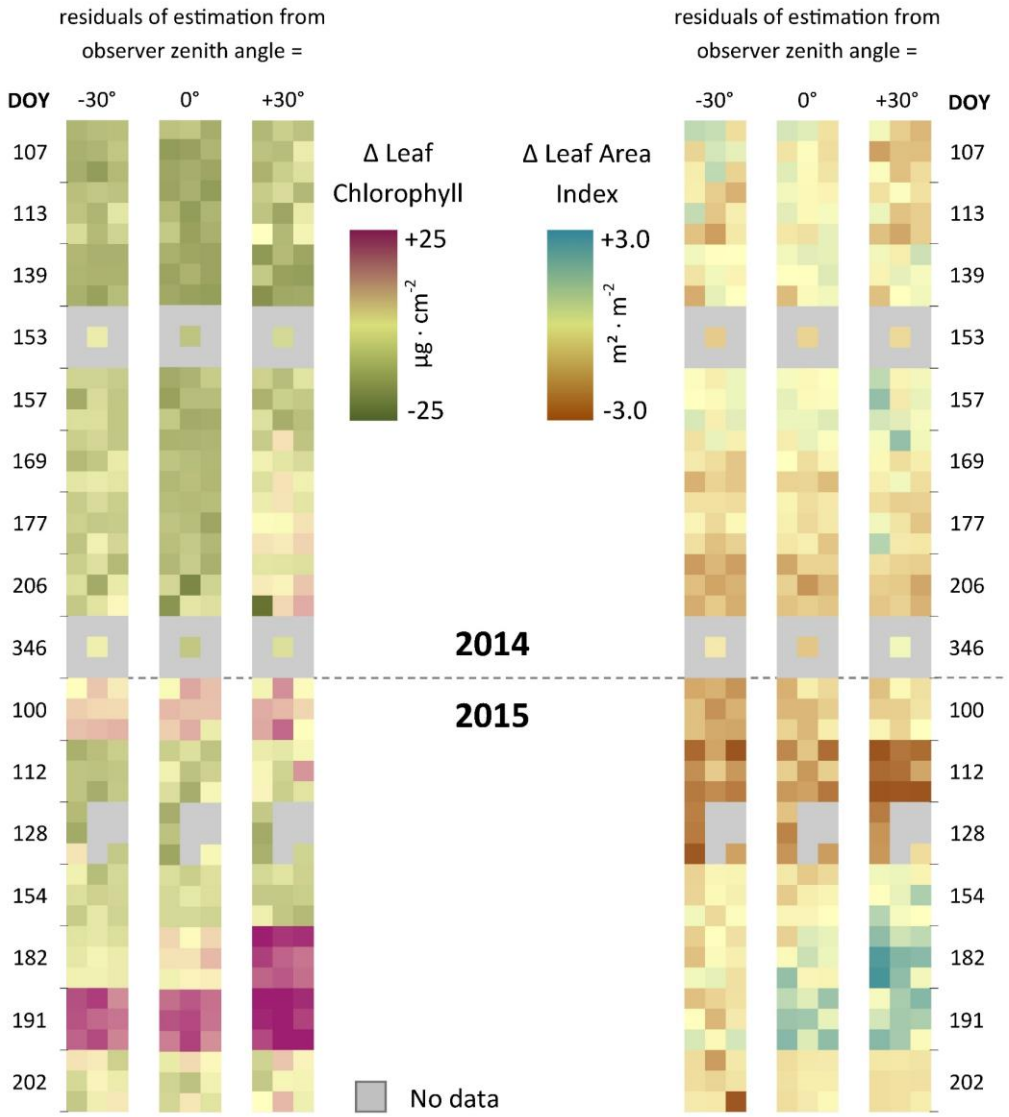


Figure 2.4.7. Visualization of the residuals, i.e., in situ measurements minus parameter estimations. For LCC, purple pixels show an underestimation of the model results, green pixels indicate overestimation. For LAI, green to blue hues show model underestimations and brown pixels model overestimations. Pastel yellow shades indicate a good model agreement with in situ observations.

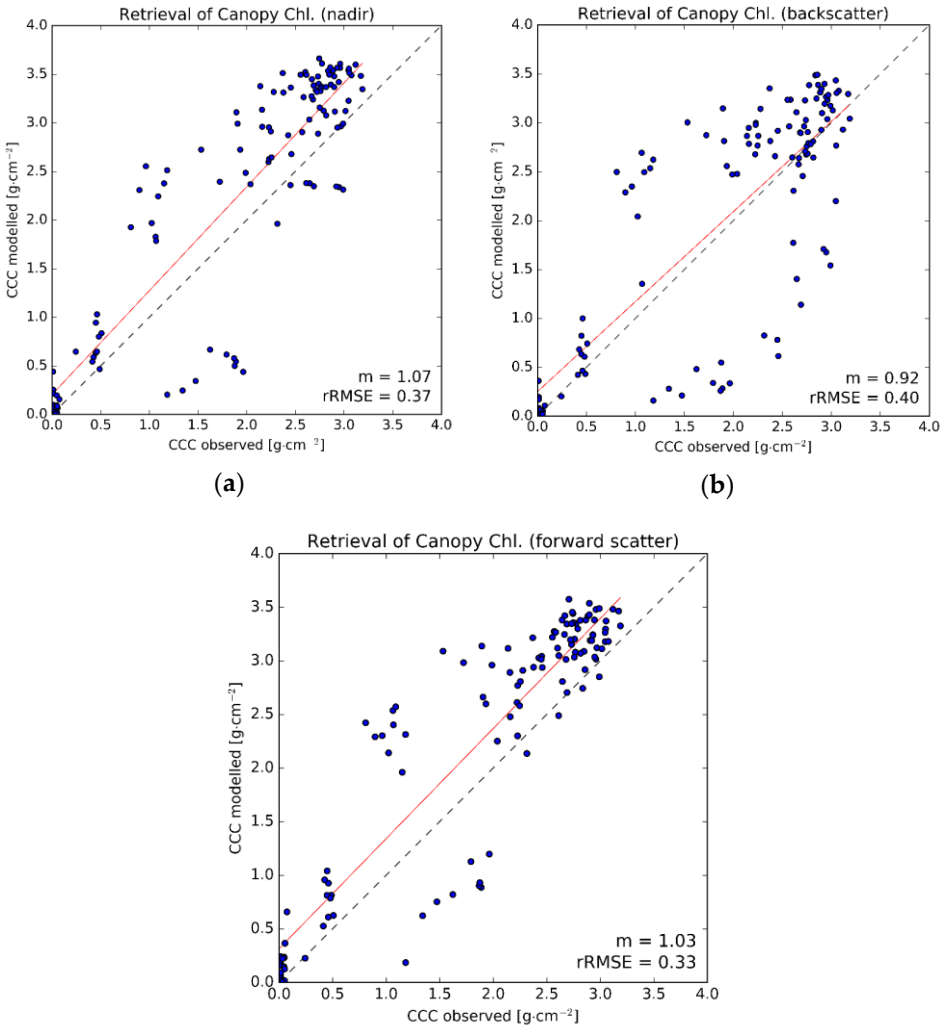


Figure 2.4.8. Canopy Chlorophyll Content, as a multiplication of LAI and LCC, combines the performance of the two underlying parameters. Results are shown for nadir (**a**), backscatter (**b**) and forward scatter (**c**) observations. The slope of the regression line is indicated as *m*.

3.3. Improved Look-Up-Table Inversions

Different techniques were explored to find an optimal setting for the inversion of the PROSAIL RTM. At first, the number of best fits (n_{bf}) was fixed to 100 and artificial noise was added to the modelled spectra (Table 2.4.5). For the LAI retrieval, performance was better, the lower the level of additive noise was assumed (rRMSE = 0.19 for σ = 0.0%). LCC on the other hand could be better retrieved with noise levels of up to 5% (best rRMSE = 0.27). Inverse multiplicative noise created similar results that were almost independent of the absolute level of σ. Interestingly, in contrast to LCC, for LAI inverse multiplicative noise yielded slightly better results than the additive noise type. A parameter-specific tendency can be assumed. Therefore, by default an inverse multiplicative noise level of σ = 2.0% was used for the first and additive noise of the same level for the second inversion run for all analyses.

In the second experiment, n_{bf} and the cost function type (cft) were varied (Table 2.4.6). For LAI, the overall best result was achieved for MAE and $50 \leq n_{bf} \leq 200$ (rRMSE = 0.18). The MAE generated 3% lower rRMSE in average. NSE as a cost function did not pass the slope threshold test for any of the experiments and was omitted in the analysis of the results. Also, for LCC, the choice of a merit function seemed to be of minor importance in comparison to the number of best fits taken into account for the retrieval. The more LUT members were considered, the better the performance in terms of rRMSE at the

expense of precision at the extreme ranges, leading to a reduced regression slope. $RMSE_{cft}$ performed slightly better than MAE, again suggesting a parameter-specific behavior of the LUT inversion.

Table 2.4.5. Experimental results for look-up-table (LUT)-based inversion of LAI and LCC with application of artificial noise (additive and inverse multiplicative) on modelled spectra for n_{bf} set to 100.

Noise Level	Additive Noise				Inverse Multiplicative Noise			
	LAI		LCC		LAI		LCC	
σ (%)	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE
0.0	0.89	0.19	1.50	0.34	0.89	0.19	1.50	0.34
0.1	0.89	0.19	1.49	0.33	0.89	0.19	1.33	0.30
1.0	0.89	0.20	1.40	0.29	0.89	0.19	1.33	0.30
2.0	0.88	0.22	1.30	0.29	0.89	0.19	1.34	0.30
5.0	0.80	0.29	1.24	0.28	0.90	0.19	1.32	0.30
10.0	0.64	0.38	1.68	0.31	0.88	0.20	1.27	0.27

Table 2.4.6. Experimental results for minimization by different cost function types. Deviations between modelled and measured spectra were quantified by either the $RMSE_{cft}$ or the MAE. Additionally, the number of best fits to be averaged was increased from 1 to 1000 to find the optimal setup. Artificial inverse multiplicative noise is set to a level of 2.0% for all results.

Number of Best Fits	RMSE Cost Function				MAE Cost Function			
	LAI		LCC		LAI		LCC	
	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE	Slope	rRMSE
1	0.88	0.28	1.61	0.44	0.88	0.28	1.55	0.42
50	0.94	0.19	1.44	0.32	0.93	0.18	1.36	0.32
100	0.89	0.19	1.30	0.29	0.92	0.18	1.18	0.28
200	0.90	0.20	1.26	0.28	0.89	0.18	1.13	0.28
500	0.86	0.21	1.03	0.26	0.87	0.20	0.89	0.26
1000	0.83	0.22	0.87	0.24	0.84	0.20	0.77	0.26

4. Discussion

For the interpretation of directional angular effects on spectral reflectance, the anisotropy factor (ANIF) can be consulted: due to self-shading effects, forward scatter images appear darker than nadir or backscatter images. In the latter case, a greater fraction of incident sunlight is reflected back to the direction of its origin and is consequently missing on the opposite viewing direction. This so-called hot spot effect leads to a spectral saturation and superimposes parts of the signal of leaf constituents. Moreover, ANIF is highly correlated with the magnitude of reflectance itself. If the canopy reflectance is higher, discrepancies increase between nadir and forward scatter but decrease between nadir and backscatter. On the other hand, if more radiation is absorbed or transmitted by the canopy, anisotropy decreases for forward scatter and increases for backscatter. Spectrally, high anisotropy occurs for blue and red portions of the solar spectrum from which leaf chlorophyll mainly absorbs radiation to photosynthesize. This was also found by (Dorigo 2012) for both directions, but in our study this could only be confirmed for backscatter mechanisms. This phenomenon may be the reason why it was more difficult for the PROSAIL model to reproduce the measured spectra from this direction, leading to a weaker estimation of LCC from backscatter in comparison to forward scatter spectra.

Our main study objective was to assess the effect of off-nadir observations on the prediction accuracy for leaf and canopy variables, namely LAI and leaf chlorophyll content, as it will have major implications for the user community of future EnMAP data. Generally, for both off-nadir observations, accuracy decreased when estimating LAI: rRMSE = 18% at nadir vs. rRMSE = 25% (backscatter) and rRMSE = 24% (forward scatter). For LCC and CCC, the off-nadir mode with forward scatter yielded highest accuracies with rRMSE = 20% and rRMSE = 33% respectively. Once again, the complex structure of the canopy plays an important role for the output of PROSAIL. Turbid medium assumptions are best met for homogeneous crops with least possible complexity in plant structural traits. Winter wheat is

thought to be particularly well suited for a representation through RTMs (Goel 1988). Nevertheless, the leaf surface of wheat exhibits anisotropic reflectance that is mathematically described by the BRDF [40]. Backscattering leads to glare effects and thus complicates the retrieval of LAI. In the opposite direction, i.e., forward scatter OZAs, the canopy appears darker which seems slightly better suited for the retrieval of LAI. If EnMAP data is only available for backscatter observations, an inversion will still be successful, but the user will encounter larger uncertainties.

Leaf glint generally leads to a reduced accuracy for the inversion of LCC. Senescent plant material absorbs less of the incident radiation and is more subjected to hot spot effects (Asner et al. 1998). This could be confirmed (see Figure 2.4.5), as LCC > 40 μg cm^{-2} was poorly estimated from backscatter spectral images, but well inverted from forward scatter observations. The findings suggest that LCC is best retrieved, the greater the difference of the zenith angle between sun and observer becomes. CCC acts like a linear combination of LAI and LCC in any statistical analysis. The rRMSE for nadir appears to be comparatively high and the intercept of the linear model is t = 0.20 which is 10.6% of the data average. This constant overestimation is caused by the before mentioned overestimation of LAI and LCC from nadir spectra reducing the models accuracy. An improvement of the retrieval of these two parameters also yields an improvement for CCC, as can be seen for results from the forward scatter observations.

Another focus of this work was to test different LUT-based strategies, while keeping a special focus on the zenithal-angular effects for spectral observations from the future EnMAP sensor. Most commonly, the $RMSE_{cft}$ is used as a merit function to find the best matching LUT members. In fact, there is only a 3.0% mean difference between $RMSE_{cft}$ and MAE in the resulting parameter estimation. The n_{bf} to incorporate in the parameter retrieval has a much stronger impact on the success of the inversion. If we assess only the rRMSE as a statistical measure for the inversion performance, it could be concluded that for LCC there is a steady improvement in predictability for larger n_{bf}. It should be noted, however, that the slope responds conversely, decreasing for larger n_{bf} and moving away from the optimal value of 1.0. Lower regression slopes indicate that the range of predicted variables becomes more level, cutting off lowest and highest inversion results. Sehgal et al. (Sehgal et al. 2016) found an optimal inversion routine with n_{bf} = 10% of the LUT-size which, in their case, was 5400 members. For larger n_{bf} the inversion approaches the expected value of the variable as specified before the creation of the LUT, so RMSE is bound to decrease if field observations served as a reference for the original parameter distribution. For all angular settings, statistics deteriorated for n_{bf} > 100 or 0.2% of the compared LUT-members. This suggests that the LUT composition was optimally set. Accordingly, n_{bf} = 100 is considered as the optimal setting in the case of this study.

A comparison of the performance with other studies is generally difficult, due to the exploration of different sensor data, LUT-compositions, inversion techniques, crop types and measurement ranges. However, as example, Atzberger et al. (2003) retrieved LAI with an RMSE of 0.83 (m^2 m^{-2}) and CCC with RMSE 0.66 (g m^{-2}) from winter wheat spectra by training artificial neural networks on PROSAIL which is roughly in the same accuracy range as our findings for nadir observations.

Different sources of errors and uncertainties in the whole inversion process must be considered as limitations to this study: in situ measurements of biophysical variables, spectral measurements, simulation of EnMAP data, model representation and the inversion scheme. For most variables, in situ errors can be reduced by choosing an adequate sampling scheme with multiple repetitions. The median standard deviation for LAI measurements of two seasons was 0.22 (m^2 m^{-2}) and 3.16 (μg cm^{-2}) for LCC. Repetitions of the ALIA estimation in the field revealed a mean error of ±7°. Senescent canopies yielded higher uncertainties for the measurement of most variables. Standard deviation of all EnMAP-end-to-end simulations was σ = 0.013 (Refl.) at the NIR-plateau which is 0.28% of the mean reflectance at this wavelength. In comparison to other error sources, this uncertainty played only a minor role. LAI acts as a scaling factor for the leaf constituents. The reflectance signal is ambiguous for substances of lower concentrations within a dense canopy or substances of higher concentrations in a sparse canopy respectively. The hierarchical approach estimates LAI first, fixates it and then finds the other parameters in consecutive inversion steps. This proved to work well for LCC, but not yet for other parameters. For instance, ALIA could not be estimated despite its high sensitivity throughout the covered spectral range

(Verrelst et al. 2015). Early in the development of PROSAIL it was stated that ALIA and LAI are highly correlated and therefore can hardly be separated in the inversion (Jacquemoud 1993). In fact, if LAI is inverted with an accuracy of 20%, there is no autocorrelation ($R^2 < 0.01$) of the ALIA residuals, although the parameter itself could not be retrieved in any acceptable way. For an improvement of leaf pigment estimations, the new version of PROSPECT (Prospect-D (Féret et al. 2017a)) is eagerly awaited for a more detailed representation of the leaf-biochemistry, namely the consideration of anthocyanins.

5. Conclusions

Spectral differences from a change in observation geometry depend on the optical properties of foliage and canopy. Complex architecture does not necessarily lead to an increased anisotropy, but the photochemistry of the leaf does very much so. With a step-wise hierarchical variable retrieval based on the PROSAIL model it was shown that longer path lengths do not lead to a higher accuracy in LAI estimation, but still allow a retrieval of this variable with satisfying accuracy. For LCC, on the other hand, the retrieval accuracy did increase when using off-nadir observations. Overall, accuracies are still in the range of about 20% for LAI and LCC. The look-up-table approach was improved when parameters were inverted hierarchically with educated a priori knowledge about the considered wavelengths. The implementation of a non-photosynthetic vegetation background improved the estimation of biophysical parameters especially for senescent phenology states. We agree with the authors of Schlerf and Atzberger (2012) in the assertion that there is no universal convention about the zenith view angles best suited to retrieve canopy structure from. For the final user, the following findings can be synthesized for winter wheat crops:

- Effects of anisotropy are strongest for early phenological stages and backscatter observations;
- LAI is best estimated from near-nadir observations;
- Optimal results for a retrieval of leaf chlorophyll content is achieved for an observer zenith angle opposite to the sun (forward scatter);
- For both variables (LAI and LCC) feasible results are obtained for all considered EnMAP geometrical constellations.

In summary, the off-nadir capability of the future EnMAP sensor will increase the number of available scenes for the user as well as the probability of achieving continuous time-series acquisitions.

Acknowledgments: The research presented in this article was conducted at the Chair of Geography and Remote Sensing, Department of Geography, Ludwig-Maximilians-Universität Munich. The financial support through the Space Administration of the German Aerospace Center (DLR) in the frame of the project "EnMAP Scientific Advisory Group Phase II—Monitoring the Phenological Development of Agricultural Crops" through funding by the German Ministry of Economics and Technology under the grant code 50EE1308 is gratefully acknowledged. We want to thank the communal farms of Munich for granting access to the sampling plots and for providing auxiliary data.

Author Contributions: Martin Danner wrote the code for the model inversion, conducted the analyses and wrote most of the manuscript. Katja Berger and Tobias Hank assisted with the development of the inversion strategy and with the definition of the field sampling campaigns, Matthias Wocher helped with the field measurements, Wolfram Mauser and Tobias Hank contributed as supervisors. All authors contributed to the preparation of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Atzberger, C.; Richter, K. Spatially constrained inversion of radiative transfer models for improved lai mapping from future sentinel-2 imagery. *Remote Sens. Environ.* **2012**, *120*, 208–218.
2. Campos-Taberner, M.; García-Haro, F.J.; Camps-Valls, G.; Grau-Muedra, G.; Nutini, F.; Crema, A.; Boschetti, M. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. *Remote Sens. Environ.* **2016**, *187*, 102–118.
3. Verrelst, J.; Rivera, J.P.; Leonenko, G.; Alonso, L.; Moreno, J. Optimizing lut-based rtm inversion for semiautomatic mapping of crop biophysical parameters from sentinel-2 and-3 data: Role of cost functions. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 257–269.
4. Verrelst, J.; Rivera, J.P.; Veroustraete, F.; Muñoz-Marí, J.; Clevers, J.G.; Camps-Valls, G.; Moreno, J. Experimental sentinel-2 lai estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 260–272.
5. Bontemps, S.; Arias, M.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Matton, N.; Morin, D.; Popescu, R. Building a data set over 12 globally distributed sites to support the development of agriculture monitoring applications with sentinel-2. *Remote Sens.* **2015**, *7*, 16062–16090.
6. Locherer, M.; Hank, T.; Danner, M.; Mauser, W. Retrieval of seasonal leaf area index from simulated enmap data through optimized lut-based inversion of the prosail model. *Remote Sens.* **2015**, *7*, 10321–10346.
7. Atzberger, C.; Jarmer, T.; Schlerf, M.; Kötz, B.; Werner, W. Retrieval of wheat bio-physical attributes from hyperspectral data and sailh+ prospect radiative transfer model. In Proceedings of the 3rd EARSeL Workshop on imaging spectroscopy, Herrsching, Germany, 13–16 May 2003; pp. 473–482.
8. Duan, S.-B.; Li, Z.-L.; Wu, H.; Tang, B.-H.; Ma, L.; Zhao, E.; Li, C. Inversion of the prosail model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 12–20.
9. Verger, A.; Baret, F.; Camacho, F. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with chris/proba observations. *Remote Sens. Environ.* **2011**, *115*, 415–426.
10. Honkavaara, E.; Kaivosoja, J.; Mäkynen, J.; Pellikka, I.; Pesonen, L.; Saari, H.; Salo, H.; Hakala, T.; Marklelin, L.; Rosnell, T. Hyperspectral reflectance signatures and point clouds for precision agriculture by light weight uav imaging system. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *I-7*, 353–358.
11. Burkart, A.; Aasen, H.; Alonso, L.; Menz, G.; Bareth, G.; Rascher, U. Angular dependency of hyperspectral measurements over wheat characterized by a novel uav based goniometer. *Remote Sens.* **2015**, *7*, 725–746.
12. Richter, K.; Atzberger, C.; Vuolo, F.; Weihs, P.; d'Urso, G. Experimental assessment of the sentinel-2 band setting for rtm-based lai retrieval of sugar beet and maize. *Can. J. Remote Sens.* **2009**, *35*, 230–247.
13. Gevaert, C.M.; Suomalainen, J.; Tang, J.; Kooistra, L. Generation of spectral-temporal response surfaces by combining multispectral satellite and hyperspectral uav imagery for precision agriculture applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3140–3146.
14. Yang, C.; Everitt, J.H.; Du, Q.; Luo, B.; Chanussot, J. Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture. *Proc. IEEE* **2013**, *101*, 582–592.
15. Schueller, J.K. A review and integrating analysis of spatially-variable control of crop production. *Fertil. Res.* **1992**, *33*, 1–34.
16. Baret, F.; Buis, S. Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. *Adv. Land Remote Sens. Syst. Model. Invers. Appl.* **2008**, 173–201, doi:10.1007/978-1-4020-6450-0_7.
17. Danner, M.; Hank, T.; Mauser, W. Comparing the potential of the sentinel-2 msi and the future enmap hsi for the retrieval of winter wheat crop parameters in southern germany. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; p. 93.
18. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: The sail model. *Remote Sens. Environ.* **1984**, *16*, 125–141.

19. Goel, N.S. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sens. Rev.* **1988**, *4*, 1–212.
20. Jacquemoud, S.; Baret, F. Prospect: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91.
21. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. Prospect+ sail models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66.
22. Kimes, D.; Knyazikhin, Y.; Privette, J.; Abuelgasim, A.; Gao, F. Inversion methods for physically-based models. *Remote Sens. Rev.* **2000**, *18*, 381–439.
23. Jacquemoud, S.; Flasse, S.; Verdebout, J.; Schmuck, G. Comparison of several optimization methods to extract canopy biophysical parameters-application to caesar data. In Proceedings of the 6th International Symposium on Physical Measurements and Signatures in Remote Sensing, Val D'Isere, France, 17–21 January 1994; pp. 291–298.
24. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Mace, D.; Pragnere, A.; Myneni, R.; Knyazikhin, Y.; Wang, L. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* **2003**, *84*, 1–15.
25. Liang, S. Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Prog. Phys. Geogr.* **2007**, *31*, 501–516.
26. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation lai and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604.
27. Weiss, M.; Baret, F.; Myneni, R.; Pragnère, A.; Knyazikhin, Y. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* **2000**, *20*, 3–22.
28. Baret, F.; Knyazikhin, Y.; Weiss, M.; Pragnère, A.; Myneni, R. Overview of retrieval techniques for lai and fapar. In Proceedings of the ALPS99 Workshop, Meribel, France, 18–22 January 1999.
29. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C. The enmap spaceborne imaging spectroscopy mission for earth observation. *Remote Sens.* **2015**, *7*, 8830–8857.
30. Guanter, L.; Kaufmann, H.; Foerster, S.; Brosinsky, A.; Wulf, H.; Bochow, M.; Boesche, N.; Brell, M.; Buddenbaum, H.; Chabrillat, S.; et al. *Enmap Science Plan*; GFZ: Potsdam, Germany, 2016.
31. Loizzo, R.; Ananasso, C.; Guarini, R.; Lopinto, E.; Candela, L.; Pisani, A. The prisma hyperspectral mission. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; p. 415.
32. Storch, T.; Bachmann, M.; Eberle, S.; Habermeyer, M.; Makasy, C.; de Miguel, A.; Mühle, H.; Müller, R. Enmap ground segment design: An overview and its hyperspectral image processing chain. In *Earth Observation of Global Changes*; Springer: Berlin, Germany, 2013; pp. 49–62.
33. Stuffler, T.; Hofer, S.; Leipold, M.; Förster, K.; Sang, B.; Schubert, J.; Penné, B.; Kaufmann, H.; Müller, A.; Chlebek, C. Enmap—Space segment—Instrument and mission parameters. In Proceedings of the EARSeL SIG-IS Workshop, Tel Aviv, Israel, 16–19 March 2009.
34. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
35. Hapke, B. Bidirectional reflectance spectroscopy: 4. The extinction coefficient and the opposition effect. *Icarus* **1986**, *67*, 264–280.
36. Kaufmann, H.; Hill, J.; Hostert, P.; Krasemann, H.; Mauser, W.; Muller, A. *Science Plan of the Environmental Mapping and Analysis Program (ENMAP)*; Deutsches GeoForschungsZentrum GFZ: Potsdam, Germany, 2012.
37. Segl, K.; Guanter, L.; Rogass, C.; Kuester, T.; Roessner, S.; Kaufmann, H.; Sang, B.; Mogulsky, V.; Hofer, S. Eetes—The enmap end-to-end simulation tool. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 522–530.
38. Baret, F.; Clevers, J.; Steven, M. The robustness of canopy gap fraction estimates from red and near-infrared reflectances: A comparison of approaches. *Remote Sens. Environ.* **1995**, *54*, 141–151.
39. Sandmeier, S.; Müller, C.; Hosgood, B.; Andreoli, G. Physical mechanisms in hyperspectral brdf data of grass and watercress. *Remote Sens. Environ.* **1998**, *66*, 222–233.

40. Wang, L.; Dong, T.; Zhang, G.; Niu, Z. Lai retrieval using prosail model and optimal angle combination of multi-angular data in wheat. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 1730–1736.
41. Hilker, T.; Gitelson, A.; Coops, N.C.; Hall, F.G.; Black, T.A. Tracking plant physiological properties from multi-angular tower-based remote sensing. *Oecologia* **2011**, *165*, 865–876.
42. Tripathi, R.; Sahoo, R.N.; Sehgal, V.K.; Tomar, R.K.; Chakraborty, D.; Nagarajan, S. Inversion of prosail model for retrieval of plant biophysical parameters. *J. Indian Soc. Remote Sens.* **2012**, *40*, 19–28.
43. Verhoef, W.; Bach, H. Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and toa radiance data. *Remote Sens. Environ.* **2007**, *109*, 166–182.
44. Vuolo, F.; Dini, L.; D'urso, G. Retrieval of leaf area index from chris/proba data: An analysis of the directional and spectral information content. *Int. J. Remote Sens.* **2008**, *29*, 5063–5072.
45. Gratani, L.; Ghia, E. Changes in morphological and physiological traits during leaf expansion of arbutus unedo. *Environ. Exp. Bot.* **2002**, *48*, 51–60.
46. Goel, N.S.; Strebel, D.E. Simple beta distribution representation of leaf orientation in vegetation canopies. *Agron. J.* **1984**, *76*, 800–802.
47. Lichtenthaler, H.K. [34] chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol.* **1987**, *148*, 350–382.
48. Verhoef, W.; Bach, H. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sens. Environ.* **2003**, *87*, 23–41.
49. Darvishzadeh, R.; Matkan, A.A.; Ahangar, A.D. Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1222–1230.
50. Wainwright, J.; Mulligan, M. *Environmental Modelling: Finding Simplicity in Complexity*; John Wiley & Sons: New York, NY, USA, 2005.
51. Richter, K.; Atzberger, C.; Hank, T.B.; Mauser, W. Derivation of biophysical variables from earth observation data: Validation and statistical measures. *J. Appl. Remote Sens.* **2012**, *6*, doi:10.1117/1.JRS.6.063557.
52. Dorigo, W.A. Improving the robustness of cotton status characterisation by radiative transfer model inversion of multi-angular chris/proba data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 18–29.
53. Asner, G.P.; Wessman, C.A.; Schimel, D.S.; Archer, S. Variability in leaf and litter optical properties: Implications for brdf model inversions using avhrr, modis, and misr. *Remote Sens. Environ.* **1998**, *63*, 243–257.
54. Sehgal, V.K.; Chakraborty, D.; Sahoo, R.N. Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. *Inf. Process. Agric.* **2016**, *3*, 107–118.
55. Verrelst, J.; Rivera, J.; Moreno, J. Artmo's global sensitivity analysis (gsa) toolbox to quantify driving variables of leaf and canopy radiative transfer models. *EARSeL eProc.* **2015**, *14*, 1–11.
56. Jacquemoud, S. Inversion of the prospect+ sail canopy reflectance model from aviris equivalent spectra: Theoretical study. *Remote Sens. Environ.* **1993**, *44*, 281–292.
57. Féret, J.-B.; Gitelson, A.; Noble, S.; Jacquemoud, S. Prospect-d: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sens. Environ.* **2017**, *193*, 204–215.
58. Schlerf, M.; Atzberger, C. Vegetation structure retrieval in beech and spruce forests using spectrodirectional satellite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 8–17.

2.5. Preparation of Scientific Publication II

Leaf Inclinations, unlike LAI and C_{cab}, could not be retrieved successfully with the LUT approach of Publication I. This led to initial thoughts about the role of canopy structure in PROSAIL inversions in general. With field campaigns continuing in parallel to their scientific exploitation, a growing set of field data became available. Spectral observations and measurements of biochemical and biophysical variables of four seasons of winter wheat and three seasons of silage maize were used for further investigations about the capability of PROSAIL in retracing field conditions if all major input parameters are filled with actual in situ data. A comparison between PROSAIL simulated spectra and crop reflectances was facilitated by creating an EnMAP-Box application that allows ad-hoc calculation of any combination of PROSPECT and SAIL according to user-defined sets of input parameters. Its graphical user interface (GUI) contains sliders representing the values of all input parameters and was later extended by the possibility to upload a static spectrum, resample model outputs to the desired spectral resolution and calculate error statistics between the graphs on the fly. The Interactive Visualization of Vegetation Reflectance Models (IVVRM, see Figure 2.5.1; Danner et al. 2018) could be used for a model inversion via manual curve fitting.

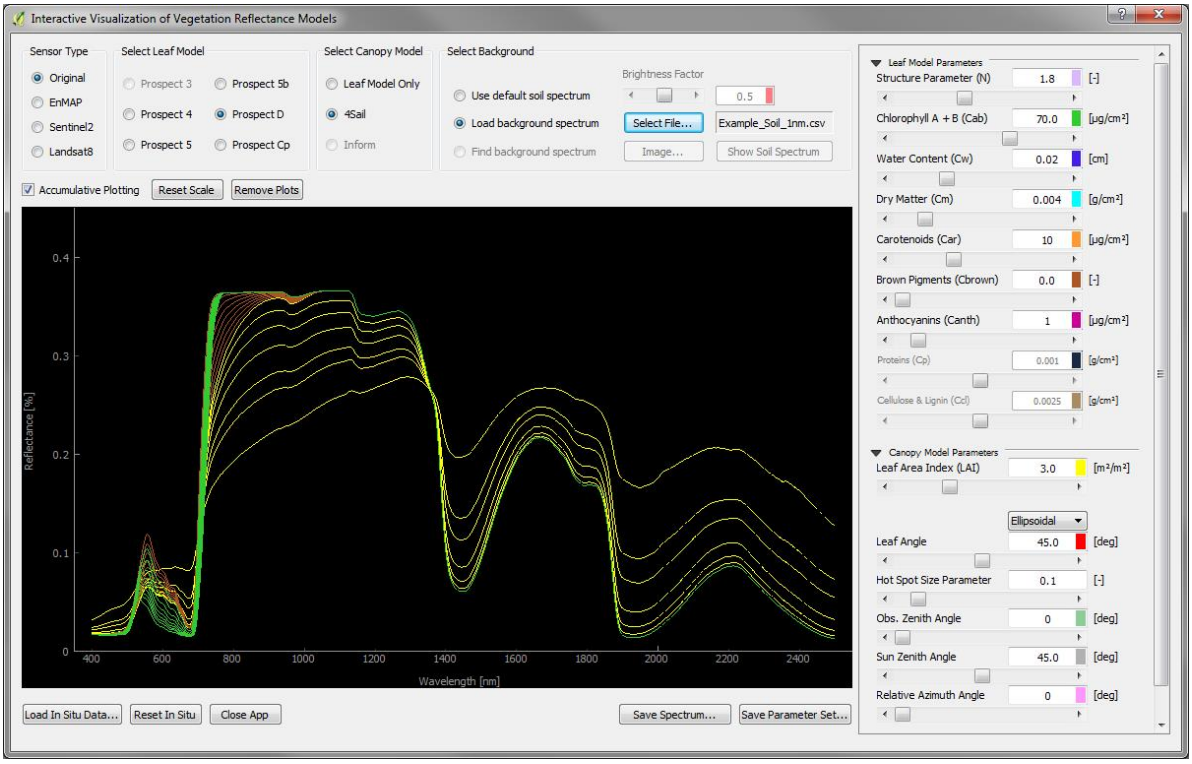


Figure 2.5.1. Illustration of the Interactive Visualization of Vegetation Reflectance Models (IVVRM) main window. Accumulative plotting is activated to draw multiple spectra into the same plotting canvas, which enables the studying of local sensitivity of the PROSAIL input parameters for the selected setting.

LAI, ALIA, C_{cab}, C_w, C_m, information about the sun-sensor-target geometry and the hotspot size served as input for 4SAIL + PROSPECT-5B. They were complemented by the leaf structure parameter N derived from C_m, the fraction of brown leaf area as an initial guess for the arbitrary C_{brown} parameter and the ratio between direct and diffuse irradiance (skyl) derived from the sun zenith angle. Despite the solid base of in situ data, PROSAIL output revealed quite distinct deviations from their associated spectral measurement. These deviations follow a seasonal trend for winter wheat with model underestimations until emergence of inflorescence where the direction of the gap reverts. The absolute error between modelled and measured spectra correlated positively with the share of visible fruit ears as obtained from automated segmentation of nadir RGB images.

The final conclusion of this publication is that the ALIA as considered in SAIL cannot be treated as a physical representation of leaf inclinations. With its strong sensitivity mainly in the NIR domain of vegetation spectra, the author suggests treating it as a free parameter and to couple its inversion to that of the LAI.

2.6. Scientific Publication II

Article

Fitted PROSAIL Parameterization of Leaf Inclinations, Water Content and Brown Pigment Content for Winter Wheat and Maize Canopies

Martin Danner *, Katja Berger, Matthias Wocher, Wolfram Mauser and Tobias Hank

Department of Geography, Ludwig-Maximilians-Universität München, Luisenstraße 37, D-80333 Munich, Germany; katja.berger@iggf.geo.uni-muenchen.de (K.B.); m.wocher@iggf.geo.uni-muenchen.de (M.W.); w.mauser@lmu.de (W.M.); tobias.hank@lmu.de (T.H.)

* Correspondence: m.danner@lmu.de; Tel.: +49-89-2180-6673

Received: 05 April 2019; Accepted: 09 May 2019; Published: 12 May 2019

Abstract: Decades after release of the first PROSPECT + SAIL (commonly called PROSAIL) versions, the model is still the most famous representative in the field of canopy reflectance modelling and has been widely used to obtain plant biochemical and structural variables, particularly in the agricultural context. The performance of the retrieval is usually assessed by quantifying the distance between the estimated and the in situ measured variables. While this has worked for hundreds of studies that obtained canopy density as a one-sided Leaf Area Index (LAI) or pigment content, little is known about the role of the canopy geometrical properties specified as the Average Leaf Inclination Angle (ALIA). In this study, we exploit an extensive field dataset, including narrow-band field spectra, leaf variables and canopy properties recorded in seven individual campaigns for winter wheat (4x) and silage maize (3x). PROSAIL outputs generally did not represent field spectra well, when in situ variables served as input for the model. A manual fitting of ALIA and leaf water (EWT) revealed significant deviations for both variables (RMSE = 14.5°, 0.020 cm) and an additional fitting of the brown leaf pigments (C_{brown}) was necessary to obtain matching spectra at the near infrared (NIR) shoulder. Wheat spectra tend to be underestimated by the model until the emergence of inflorescence when PROSAIL begins to overestimate crop reflectance. This seasonal pattern could be attributed to an attenuated development of $ALIA_{opt}$ compared to in situ measured ALIA. Segmentation of nadir images of wheat was further used to separate spectral contributors into dark background, ears and leaves + stalks. It could be shown that the share of visible fruit ears from nadir view correlates positively with the deviations between field spectral measurement and PROSAIL spectral outputs ($R^2 = 0.78$ for aggregation by phenological stages), indicating that retrieval errors increase for ripening stages. An appropriate model parameterization is recommended to assure accurate retrievals of biophysical and biochemical products of interest. The interpretation of inverted ALIA as physical leaf inclinations is considered unfeasible and we argue in favour of treating it as a free calibration parameter.

Keywords: reflectance modelling; hyperspectral remote sensing; radiative transfer model; PROSAIL; agriculture

1. Introduction

Estimation of plant biophysical characteristics is a key factor for agricultural science and applications [1]. Knowledge about type and proportions of the constituents in vegetation allows for a dedicated analysis of its state of health [2–6], potential photosynthetic activity [7–11] or yield potential [12–15]. Plant pigments can be optically measured in vitro with spectrophotometers [16]. Similarly,

water content and non-photosynthetic organic compounds like cellulose are obtained in laboratory analysis [17,18]. Even though these methods are important for a quantitative characterization of plants, they fail to cover larger areas, as they represent the state of individual plants or phyto-elements rather than provide an integrative assessment of canopies. A synoptic view of the Earth's surface can be achieved by remote sensing, which makes use of the complex interactions between radiation and environment [19]. The amount of transmitted, reflected and absorbed energy from the target of interest yields useful information about its structure and inherent substances. Indirect remote measurements of these variables from airborne platforms or even from space thus are preferred over time-consuming laboratory studies. One way of deriving canopy variables is to create empirical relationships between reflectances and known variables. This can be achieved either by parametric regression approaches using spectral indices [20–23] or directly via non-parametric approaches by obtaining these variables from reflectance data using machine learning regression algorithms (e.g., [24–27]). A comprehensive overview of this topic is provided by Verrelst et al. [28]. Although those methods can create reasonable results on the training data, they are prone to overfitting and the relationships found are rarely transferable in space, time or crop type [29]. Numerical models, on the other hand, allow a generic representation of vegetation as 3D-objects via ray tracing Monte Carlo models [30–33] or 1D turbid medium layers with intrinsic canopy architecture.

The most prominent 1D turbid medium Radiative Transfer Model (RTM) for vegetation is PROSAIL [34]. It consists of two separate simulation cores: a) one version of the PROSPECT leaf optical properties model [35–37] and b) a four-stream Scattering of Arbitrarily Inclined Leaves (e.g., 4SAIL) canopy architecture model [38,39]. The inputs of those two models in their respective versions are listed in Table 2.6.1.

Table 2.6.1. Overview of the PROSAIL parameters and their according dimensions. Some parameters, for example, the leaf chlorophyll content, are used in all PROSPECT versions, whereas other parameters were included in newer releases.

Parameter	Description	Unit	Model versions
N	Leaf structure parameter	-	Prospect (all)
C_{cab}	Leaf Chlorophyll$_{a+b}$ content	µg cm^{-2}	Prospect (all)
C_w	Leaf Equivalent Water Thickness (EWT)	cm	Prospect (all)
C_m	Leaf Mass per Area	g cm^{-2}	Prospect (all)
C_{car}	Leaf Carotenoids content	µg cm^{-2}	Prospect 5
C_{brown}	Leaf Brown Pigments parameter	-	Prospect 5b
C_{anth}	Leaf Anthocyanins content	µg cm^{-2}	Prospect D
LAI	Leaf Area Index	m^2 m^{-2}	4SAIL
LIDF	Leaf Inclination Distribution Function	-	
or	or	or	4SAIL
ALIA	Average Leaf Inclination Angle	Deg	
Hspot	Hot Spot size parameter	-	4SAIL
ρ_{soil}	Soil Reflectance	-	4SAIL
P_{soil}	Soil Brightness Parameter	-	4SAIL
SZA	Sun Zenith Angle	Deg	4SAIL
OZA	Observer Zenith Angle	Deg	4SAIL
rAA	relative Azimuth Angle	Deg	4SAIL
skyl	Ratio of diffuse to total incident radiation	-	4SAIL

PROSPECT calculates radiative interactions on leaf level with regard to the absorption coefficients of leaf constituents, producing continuous leaf reflectance and transmittance spectra over the optical domain (400–2500 nm). The line of published versions of PROSPECT differs in featured parameters, their absorption coefficients and the refractive indices. 4SAIL assimilates the output of PROSPECT and calculates inner-canopy scattering processes, which mainly depend on plant density, leaf orientation and the relative angles of observer and illumination source [34].

The procedure of generating reflectance spectra from a set of predefined input parameters is called forward or direct mode. Inverting the model means estimating those parameters from a vegetation spectrum that has been measured for example, with a spectrometer. The PROSAIL spectral sampling width is 1 nm, so its output normally requires spectral resampling to become comparable with specific Earth Observation sensor data. The more bands a sensor provides the less information is lost in aggregation. Such hyperspectral data with contiguous bands can be collected either from airborne, spaceborne or field-based sensors [40]. Spaceborne hyperspectral sensors, combining high resolution spectral and temporal sampling to cover dynamic processes, are especially promising for agricultural purposes [41]. Amongst the currently planned hyperspectral spaceborne missions are EnMAP [42], PRISMA [43], SHALOM [44], HyspIRI [45] and CHIME [46].

The inversion of hyperspectral data via RTMs is a key application of agricultural remote sensing [34,47]. For this purpose, the model inputs are changed such that the output spectrum matches the real-world reflectances as closely as possible. The final parameter set used for that optimum model result then is considered as an estimate for the variables of interest of the observed canopy. Common methods involve subsequent comparison of the measured spectra with a Look-Up-Table (LUT; for example, References [48–52]), systematic modifications of input parameters using iterative minimization – so called optimization algorithms (see References [53,54] for overview) – or manual fitting.

Many studies have used PROSAIL outputs as a substitute for real spectral data, either due to a lack of field measurements or to examine spectral responses of canopies of different structure and biochemistry [55–58]. When methods are trained on such synthetic data but applied to real data, we often observe a systematic bias. Even though PROSAIL is the most prominent canopy reflectance model for the inversion of vegetation spectral data [47], only few studies have focused on its capability to reproduce field spectral observations [51,59,60]. These observations are instead commonly used as validation data for retrieval methods assuming that errors of the inversion indicate a weak model performance.

The parameterization of crop architecture is upscaled to canopy level. Most of this process is driven by the well understood Leaf Area Index (LAI) and the Average Leaf Inclination Angle (ALIA, also known as Mean Leaf Angle MLA or Average Leaf Angle ALA), which in contrast has been scarcely discussed in the literature. Upon adjusting measured ALIA, the Equivalent Water Thickness (EWT) was also identified as deviating from the expected behaviour. The objective of our study is to analyse the seasonal development of these parameters, their fitted representations and the resulting deviations between the modelled and measured spectra of winter wheat and silage maize.

2. Materials and Methods

2.1. Study Site

A database of ground and remotely-sensed field data was obtained from the study area of Munich-North Isar (MNI), which is located in Bavaria, southern Germany. In the years of 2014, 2015, 2017 and 2018 the study sites (Figure 2.6.1) were visited regularly during the growing periods from March to September. Data were collected on winter wheat and silage maize fields belonging to the communal farmlands of Munich, east of the river Isar. The exact sampling sites varied from year to year due to crop rotation but all fields are located within a circle of 5 km in diameter, centred around 48°16′04″N, 11°42′45″E. Soil sampling provided no evidence of significant micro-locational (dis)advantages. An overview of all included field campaigns is given in Table 2.6.2.

Figure 2.6.1. Overview of the study site Munich North Isar with its test fields for winter wheat and silage maize in the years from 2014 to 2018 (left). The image on the right shows the layout of the measurement design for the nine sampling units per field. Reference system: WGS84 (EPSG 4326).

Table 2.6.2. List of field campaigns at the MNI study site in southern Germany.

Year	Crop	No. of Field Dates
2014	Winter wheat	10
2014	Silage maize	11
2015	Winter wheat	11
2017	Winter wheat	12
2017	Silage maize	8
2018	Winter wheat	7
2018	Silage maize	7

Each site was confined to a 30 × 30 meters area, corresponding to an average pixel size expected from hyperspectral satellite sensors and divided into nine elementary sampling units (ESU). Data was then aggregated back to the original 30 m resolution to obtain a stable average for that date. The ESUs were revisited at each field date to guarantee solid time series.

2.2. In Situ Measurements

Canopy reflectance spectra were obtained from nadir view with an ASD FieldSpec 3 Jr. (Boulder, CO, USA [61,62]) under clear-sky conditions in the range of 350 nm to 2500 nm. Nine measurements were carried out at the same spot, averaged, splice-corrected, converted into absolute reflectance values using a white reference panel and were slightly smoothed with a Savitzky-Golay-filter [63]. At last, noisy wavelengths, that is, atmospheric water absorption bands (1360 nm–1470 nm and 1790 nm – 2000 nm), were masked for any further analysis. Apart from top-of-canopy reflectances, one representative soil spectrum was recorded for each field date, serving as background information in the modelling process. In doing so, the default literature soil spectra of SAIL were temporarily replaced with in situ information and P_{soil} became obsolete. With nadir observations only (observation zenith angle = 0°), the sun-sensor-target-geometry is solely represented by the sun zenith angle (SZA), which was calculated from the geographical location, date and time of the observation.

Nadir RGB-photographic images were taken on winter wheat canopy scale with a Panasonic Lumix digital camera and a distance of 1.5 m above the ground. Due to the large canopy height of maize crops, no continuous series of these photographs were available for analysis of the maize site.

Average Leaf Inclination Angles (ALIA) were estimated from random samples of at least 10 leaves in vivo, using a Suunto PM-5/360 clinometer [64], which was moved along the flat leaf axis. Inclination angles of the single leaves were calculated as length-weighted averages (see Equation (1)) of those measurements and then aggregated to a mean ALIA value for the respective sampling unit.

$$\mathrm{ALIA} = \frac{1}{l_{tot}} \cdot \sum_{i=1}^{n} l_i \cdot \theta_i \tag{1}$$

with l as the length of the i-th of n leaf parts, θ_i as its associated inclination angle and total leaf length l_{tot}. The conformity of that method was estimated by repeated measurements of leaf inclinations by different field workers and was found to be in the range of ±8° for the MNI campaigns. In PROSAIL, the Leaf Angular Distribution is then calculated from single ALIA values according to Campbell's ellipsoidal distribution [65]. Leaf Area Index (LAI) was measured with a LI-COR Biosciences LAI-2200C [66] device as average of 14 measurements from the same location. The suggested sampling procedures for row crops were taken into account [67]. For early growth stages, single plants were sampled destructively and projected onto one square meter of ground area by multiplication with the factors plants per meter and seeding rows per meter.

Brown pigment (polyphenols; C_{brown}), as a PROSPECT input, is rather poorly analysed in the literature. Brown leaves are usually either removed before analysis (e.g., [37,68]) or set to fixed values [69]. In some studies, it is also labelled as Content of senescent material (C_s). The variable lacks of a proper physically meaning and cannot be measured with field instruments [37]. For a start, we used C_{brown} as the average "browning" of leaves between 0 (fully green leaves) and 1 (fully brown leaves) as initial guess. This is in line with Jiang et al. 2018 [70] but ignores the fact that the accumulation of brown pigments can happen without visible change of colour, which in contrast is more linked to the breakdown of leaf chlorophyll. Total LAI is reduced to green LAI by multiplication with a greenness factor $1 - C_{brown}$.

A Konica-Minolta SPAD-502 handheld device [71] served to measure internal SPAD-units, which were converted into leaf chlorophyll$_{a+b}$ contents (C_{cab}) via an instrument-specific calibration curve. Reference data on in vitro C_{cab} was obtained by applying the coefficients of Lichtenthaler [72] to spectrophotometric measurements of the extinction of chlorophyll solutions. From C_{cab}, the carotenoids content C_{car} was derived by way of a linear regression model, since C_{cab} and C_{car} showed a stable relationship for healthy green vegetation [73]. When the linear relation breaks down for senescent stages, errors for C_{car} carry only a marginal weight, because the overall biomass – as represented by the LAI – is low and has little impact on the reflectance signatures. Equivalent Water Thickness (EWT) was determined from the mass difference of sample leaves per unit leaf size before and after dehydration to constant weight (minimum 24 hours) in a laboratory oven. Leaf Mass per Area (C_m) is obtained by putting the mass of the dried leaves in relation to their one-sided surface area. The structure parameter (N) is calculated from C_m according to the inverse of Equation (2), given by Jacquemoud and Baret (1990) [37]:

$$\mathrm{LSA} = \frac{0.1 \cdot N + 0.025}{N - 0.9} \tag{2}$$

which (for $N \neq 0.9$) yields

$$\mathrm{N} = \frac{LSA \cdot 0.9 + 0.025}{LSA - 0.1} \tag{3}$$

with LSA as the Leaf Specific Area, which is the inverse of C_m (note the conversion of dimensions):

$$\mathrm{LSA}\left[\frac{cm^2}{mg}\right] = C_m\left[\frac{g}{cm^2}\right]^{-1} \div 1000\left[\frac{mg}{g}\right] \tag{4}$$

The variable distributions of the in situ measured parameters of wheat and maize for all field campaigns are summarized in Table 2.6.3.

Table 2.6.3. Statistics of in-situ measured/estimated canopy variables of the four field campaigns at the MNI test sites.

		Winter wheat			**Silage Maize**		
Variable	Year	Range	Mean	Std.	Range	Mean	Std.
LAI (-)	2014	0.08–6.27	4.82	1.85	0.09–4.03	2.21	1.58
	2015	0.33–6.20	2.82	2.10			
	2017	0.76–6.20	4.34	1.79	0.21–3.86	2.29	1.28
	2018	0.01–5.98	3.88	1.98	1.79–3.61	3.05	0.60
ALIA (deg)	2014	25–75	52	19	36–75	50	11
	2015	35–77	60	13			
	2017	45–78	68	9	49–71	63	8
	2018	45–76	64	10	49–75	59	8
C_{cab} ($\mu g\ cm^{-2}$)	2014	13.4–49.1	42.7	10.5	27.3–61.8	48.1	11.9
	2015	14.3–53.3	43.2	12.8			
	2017	18.2–59.5	50.0	10.7	38.4–55.2	48.8	5.6
	2018	11.6–53.2	43.2	14.3	48.2–60.8	56.8	4.1
C_{brown} (-)	2014	0.0–0.98	0.19	0.30	0.0–0.81	0.08	0.23
	2015	0.0–0.90	0.22	0.34			
	2017	0.0–0.80	0.09	0.22	0.0–0.05	0.01	0.02
	2018	0.0–1.0	0.18	0.37	0.0–0.01	<0.00	<0.00
EWT (cm)	2014	0.012–0.035	0.027	0.006	0.011–0.031	0.027	0.005
	2015	0.008–0.034	0.026	0.007			
	2017	0.003–0.020	0.015	0.004	0.012–0.021	0.016	0.003
	2018	0.001–0.019	0.013	0.006	0.020–0.025	0.023	0.002
C_m ($g\ cm^{-2}$)	2014	0.0047–0.0075	0.0063	0.0010	0.0032–0.0056	0.0046	0.0007
	2015	0.0036–0.0061	0.0046	0.0007			
	2017	0.0031–0.0059	0.0047	0.0008	0.0027–0.0049	0.0040	0.0007
	2018	0.0043–0.0066	0.0049	0.0008	0.0045–0.0070	0.0058	0.0008

The N parameter, as derived from C_m, ranges between 1.4 and 4.4 for wheat and between 1.3 and 3.6 for maize, with an average of 2.2 and 2.0, respectively.

Quantification of the phenological stages is taken from the BBCH Monograph [74], which is based on the perhaps better-known Zadoks scale [75]. No small-scale differences in plant development were observed within the 30 × 30 meter pixels. The code attribution is listed in Table 2.6.4.

Table 2.6.4. Macro stages of the BBCH-scale. Adapted from Table 1 in Bleiholder et al. [74].

BBCH-code	**Associated macro stage**
0	Germination / sprouting / bud development
1	Leaf development
2*	Tillering / Formation of side shoots
3	Stem elongation or rosette growth / shoot development
4*	Development of harvestable vegetative plant parts / booting
5	Inflorescence emergence / heading
6	Flowering
7	Development of fruit
8	Ripening or maturity of fruit and seed
9	Senescence, beginning of dormancy

* BBCH 2 and 4 are skipped in the classification of maize.

2.3. PROSAIL Environment

All data analysis was done with Python 3.6. For this reason, PROSAIL was first translated into Python code and improved in terms of computational performance by making use of C-based numerical python arrays [76] and substitution of recurring function calls by fixed local variables and look-up-tables.

For this study, we introduce the Interactive Visualization of Vegetation Reflectance Models (IVVRM) tool [77], which is an application in the open source software package EnMAP-Box 3 [78] and serves as a graphical user interface to work with data of multiple constellations of PROSPECT and SAIL. It was originally designed to offer ad hoc visual response to interactions with the model for educational purposes. Sliders can be accessed to change the value of the structural and biochemical input variables with each interaction causing an immediate re-calculation of the model output, which is displayed in a plot canvas. The tool was further extended with the possibility to overwrite the PROSAIL default background spectrum by a user-defined data vector and with an import function for any other in situ spectral signature. Those reflectances are then plotted in the same graph and—given an identical set of bands—multiple distance measures are instantaneously calculated and returned in the plot. These functions combined allow for a manual fitting between model output and in situ spectra, when model input is altered such that there is a minimal distance between the two spectra. The success of the spectral fitting can either be assessed with statistical measures like the root mean squared distance (RMSD, see Equation (5)) or visually with the degree of congruence of the curves in certain wavelength regions.

$$\mathrm{RMSD} = \sqrt{\frac{1}{\mathrm{n}} \cdot \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{R}_{\mathrm{measured}}(\lambda_i) - \mathrm{R}_{\mathrm{simulated}}(\lambda_i)\right)^2} \quad (5)$$

In Equation (5), $R_{measured}(\lambda_i)$ is the measured and $R_{simulated}(\lambda_i)$ the modelled reflectance at wavelength λ for the i^{th} spectral sensor band. n is the total number of bands analysed. The same equation works for the Root Mean Squared Error (RMSE), which was used to denote errors in the estimation of variables. By squaring the distances, information is lost about which of the terms is larger, that is, whether the model overestimates or underestimates the spectra measured in the field. In such cases, a simple mean deviation (Equation (6)) can be calculated alongside to the RMSD.

$$\text{Mean Deviation} = \frac{1}{\mathrm{n}} \cdot \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{R}_{\mathrm{measured}}(\lambda_i) - \mathrm{R}_{\mathrm{simulated}}(\lambda_i)\right) \quad (6)$$

The SAIL parameter *skyl* controls the ratio of diffuse to total solar radiation incident on the target. It is calculated in dependence of the sun zenith angle according to the approach of François et al. (2002) [79], which considers an average state of atmospheric conditions aligned to mid-latitudes:

$$\mathrm{skyl} = 0.847 - 1.61 \cdot \sin(90° - SZA) + 1.04 \cdot \sin^2(90° - SZA) \quad (7)$$

The dependency between skyl and SZA is non-linear. Lowest SZA (27.8°) results in skyl = 0.24, highest SZA (70.5°) in skyl = 0.43. The mean SZA of all dates was 41.1° which corresponds to skyl = 0.23. In PROSAIL, skyl is considered uniform over the optical domain, neglecting effects of stronger Rayleigh scattering for shorter wavelengths.

2.4. Variable Fitting

In situ measured variables that serve as input parameters for PROSAIL were manually adapted to match the model output with field spectral data. The altered variables are then called *optimize*— or *opt*— in the sense that they are better suited to represent the spectral behaviour of the analysed crops in the model environment of PROSPECT and SAIL. Even though PROSPECT-D has been shown to outperform its predecessors [36], in this study we used the previous version of PROSPECT-5b. The main reason for this is that anthocyanins (C_{anth}) were not measured in situ and deriving C_{anth} from spectral indices retrospectively without the possibility to validate these findings would induce another error source. Three main variables were identified for optimization to achieve proper agreement between measured

and modelled reflectance signatures: ALIA, EWT and C_{brown}. The tuning of PROSAIL canopy geometry was necessary to cope with the large deviations that occur when comparing real and synthetic data. A global sensitivity analysis (GSA) was performed with the Fourier Amplitude Sensitivity Test (FAST) using the GSAT tool [80] in Matlab. This way, the impact on modelled plant reflectivity was assessed in the range of 400 nm to 2500 nm with a five nm sampling width for all input variables of the used PROSAIL version (see Figure 2.6.2).

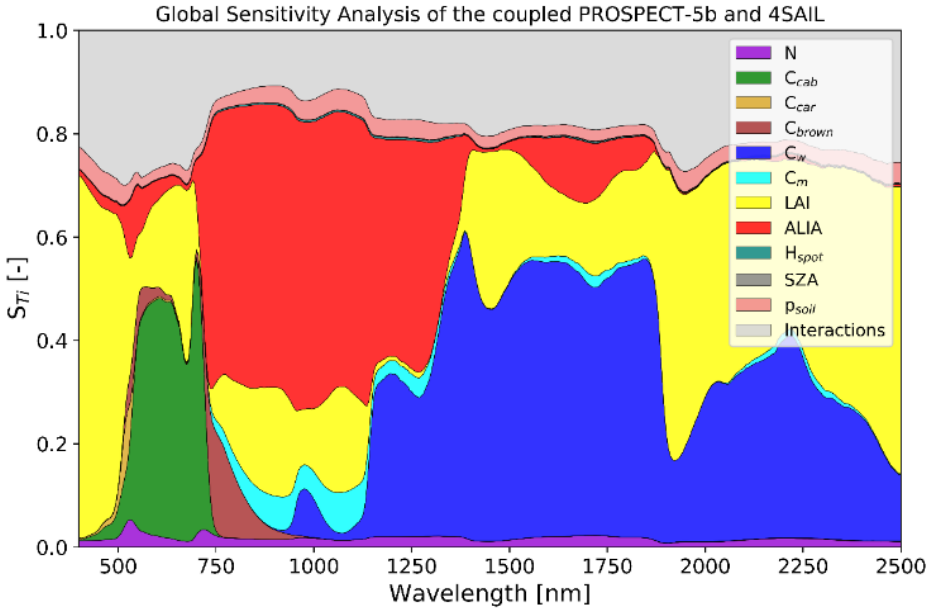


Figure 2.6.2. Global Sensitivity Analysis of the coupled PROSPECT-5b and 4SAIL models. The dimensions of sensitivity refer to the Sobol score and denote the relative contribution (S_{Ti}) of each input variable, and their interactions, to the variance of the model output. Parameter ranges: N: 1.0–2.5; C_{ab}: 0.0–80.0 µg cm^{-2}; C_{car}: 0.0–15.0 µg cm^{-2}; C_{brown}: 0.0–1.0; C_w: 0.0–0.07 cm; C_m: 0.0–0.02 g cm^{-2}; LAI: 0.0–8.0; ALIA: 0.0–88.0; H_{spot}: 0.0–0.1; p_{soil}: 0.0–1.0; SZA: 30, 35, 40, 45, 50, 55°; OZA: 0°; rAA: 0°.

ALIA was changed to obtain minimal deviations between modelled and measured spectra by using the interactive toolbox IVVRM (see Section 2.3). The success of this fitting process was monitored in the NIR spectral domain from 760 nm to 1300 nm, where ALIA shows the highest sensitivity and relative changes within SAIL. $ALIA_{opt}$ is thus obtained but the exact shape of the measured spectrum could only be represented when EWT was changed into EWT_{opt} at the same time. Again, the NIR range was chosen for this task, since inversions from the SWIR domain are known for saturation tendencies whereas the 970 nm feature considers highest radiation penetration depths for stacked leaves in a canopy due to strong vegetation reflectance and transmittance [81–83]. Finally, C_{brown} was subjected to the same calibration process. This does not have an impact on the results of EWT_{opt}, because $C_{brown,opt}$ was determined by signature matching in the red edge (<900 nm). Manual fitting may seem unusual, given the vast majority of studies that use look-up-table inversions or numerical solutions for inversion tasks. However, the drawback of possibly iterating into a local minimum is automatically avoided, since ALIA, EWT and C_{brown} have different spectral responses while all other parameters remain fixed for each date. With numerical solutions, on the other hand, slight shifts of the spectra may result in quite large deviations, whereas manual fitting yielded the advantage of a combined quantitative (error measures) and qualitative (shape) assessment of the agreement between the two spectral curves.

2.5. RGB Image Segmentation

Nadir RGB-images of winter wheat were processed to reveal visible fractions of the canopy (background, ears and leaves/stalks) as observed from a sensor-like position. The image processing was done with Python's scikit-image package [84]. Images were loaded, transformed into a value range of [0,1] and split into red, green and blue bands.

For the analysis of the background fraction, threshold values were found for each image, separating the data into dark and bright areas. Contour lines then were drawn around the darker pixel clusters and compared to the original image. At the end of an iterative process, an optimal threshold value is found and used to obtain the share of dark pixels in the image. These include not only soil pixels but also other parts of the canopy that show marginal reflectance in the visible spectrum and thus do not contribute to the canopy signature (see Figure 2.6.3c).

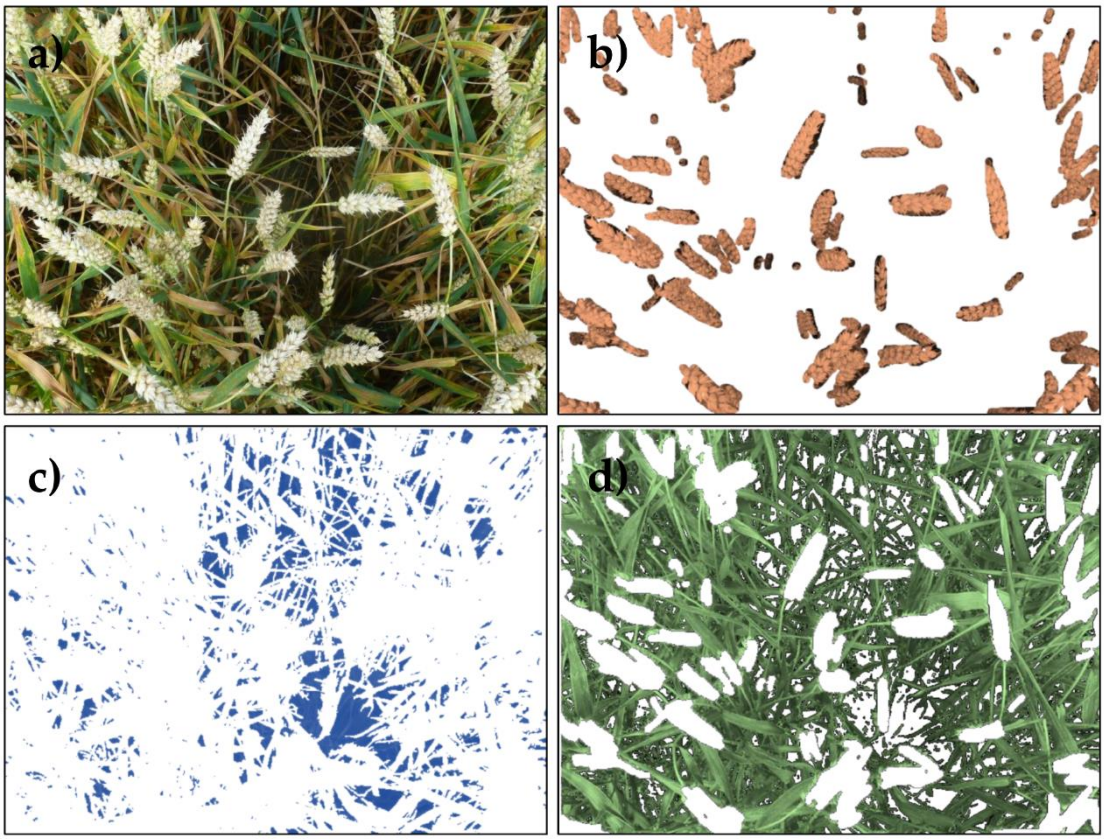


Figure 2.6.3. Illustration of the RGB image segmentation of a winter wheat canopy from 16 July 2014. From the original image (**a**), fruit ears (**b**), dark background (**c**) and leaves and stalks (**d**) are extracted.

Determining the fraction of visible area of ears requires a preceding step, in which the ears of the winter wheat canopy are manually masked using standard CAD software. Their share was again calculated with the help of scikit-image (see Figure 2.6.3b). All pixels that are classified neither as dark background nor as ears are considered leaves and stalks (Figure 2.6.3d). They usually make up the greater share of the picture. It should be noted though that these fractions could not be directly considered as linear contributors to the total spectral range of the sensor, because the RGB camera only captures wavelengths from 400 nm–750 nm, while vegetation transmissivity especially in the NIR domain can be higher compared to the VIS part. The share of vegetation pixels from nadir view is often referred to as fractional cover (f_{cover}) and the background fraction $f_{background} = 1 - f_{cover}$, as the gap fraction.

3. Results

3.1. Deviations between Model and Measurement

In situ measured variables served as input for the coupled PROSPECT-5b and 4SAIL. The spectral output of the model is contrasted with the spectral signatures recorded at the exact same spots. The results are shown in Figure 2.6.4 for winter wheat (2014) and silage maize (2017). Results of the other seasons can be found in the Supplement (wheat: S1; maize: S2).

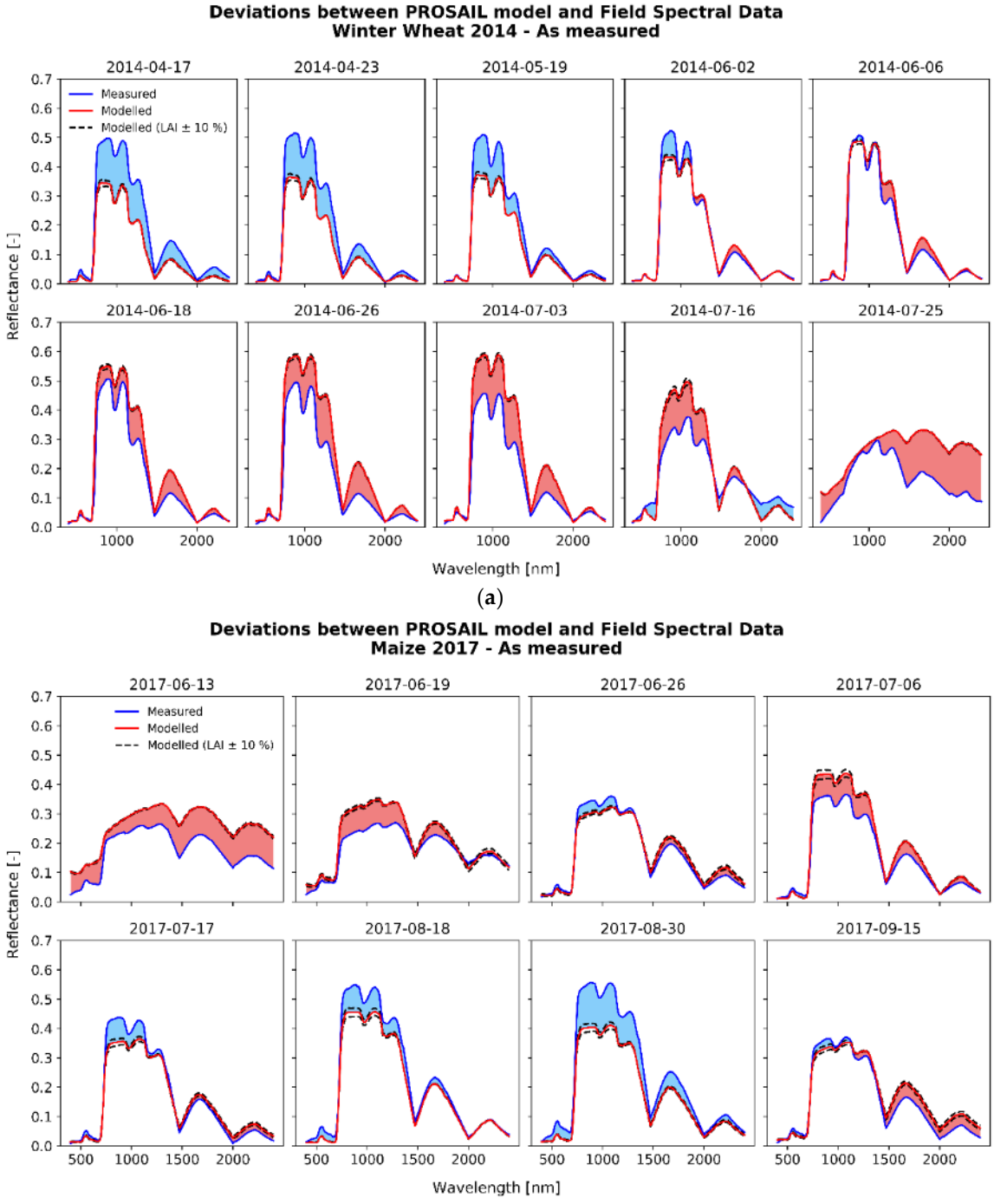


Figure 2.6.4. Spectral progression of winter wheat (**a**) and silage maize (**b**) canopies as shown for the seasons of 2014 and 2017, respectively. The measured spectra are drawn in blue, the PROSAIL output fed with in situ measured variables in red. The black dashed lines illustrate the model response to a ± 10% uncertainty of LAI.

In 2014, the winter wheat field campaign started in mid-April, when green biomass had already developed. In the first four field dates spanning one month, measured NIR reflectances increased, before revealing a decreasing tendency from June on. The SWIR range in contrast drops continuously until beginning senescence. PROSAIL output on the other hand did not capture this pattern. Modelled spectra in the NIR rise from date to date, reaching a maximum in early July (3 July 2014). In the SWIR, modelled spectra behave exactly opposite to the measured signal with a peak at 1750 nm successively

climbing from 0.08 at the beginning of the series to 0.22 at fruit development stage. In this regard, the underestimation of spectral signatures by PROSAIL gradually gave way to an overestimation.

Looking at the data of the other seasons (see Supplement S1), this pattern proves to be recurring. In 2015 and 2017, spectral and in situ data was recorded already at early growth stages of leaf development and formation of side shoots. At that time, modelled reflectances were constantly higher than those measured on site. In summary, two tipping points are observed: (1) from overestimation to underestimation (end of March; shoot development) and (2) from underestimation to overestimation (beginning of June; inflorescence emergence). Figure 2.6.5 illustrates this pattern by plotting mean deviations as PROSAIL output subtracted from field spectra. Aggregation into phenological stages allows a comparison independent of the growing conditions of each year. Patterns are similar for deviations in the NIR and SWIR, only the absolute values in the NIR were generally higher, thus so are the deviations.

Maize reflectances for 2017 were overestimated by the model for early growth stages, when fractional cover was low, and LAI ranged below 0.7. These stages were followed by a period of model-underestimation in the NIR (with exception of the aberration at the 6 July 2017 date) and signal increase in the SWIR that is not present in the model output. In the other seasons, there was no general trend in the deviation between model and measurement (see Supplement S2). For 2014, SWIR reflectances are captured well by the model but the NIR plateau was overestimated throughout the season. In 2018, a massive model-underestimation was recognized for 21 June 2018, which seems to be a single event or measurement outlier. The seasonality of deviations is analysed by calculating a simple mean of $R_{measured}$ minus $R_{modelled}$ (see Figure 2.6.5). This approach may cancel out a non-uniform behaviour of different bands of the same spectrum, so the results are separately shown for NIR and SWIR where deviations show a generally high coherence. Said deviations are less distinct for maize than for winter wheat. For this crop type, a predominant model underestimation is observed in all phenological states instead.

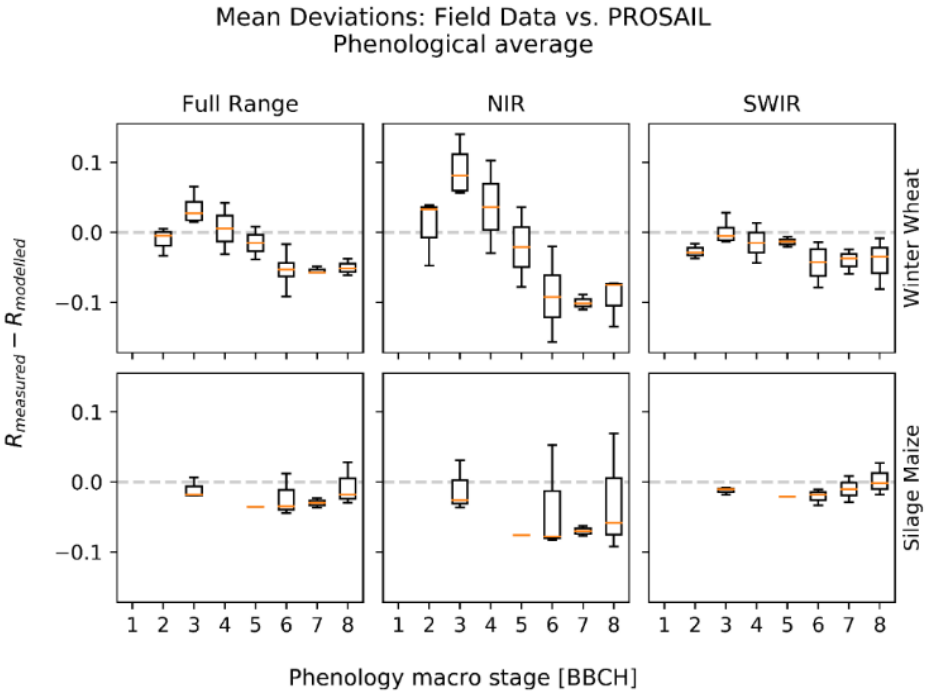
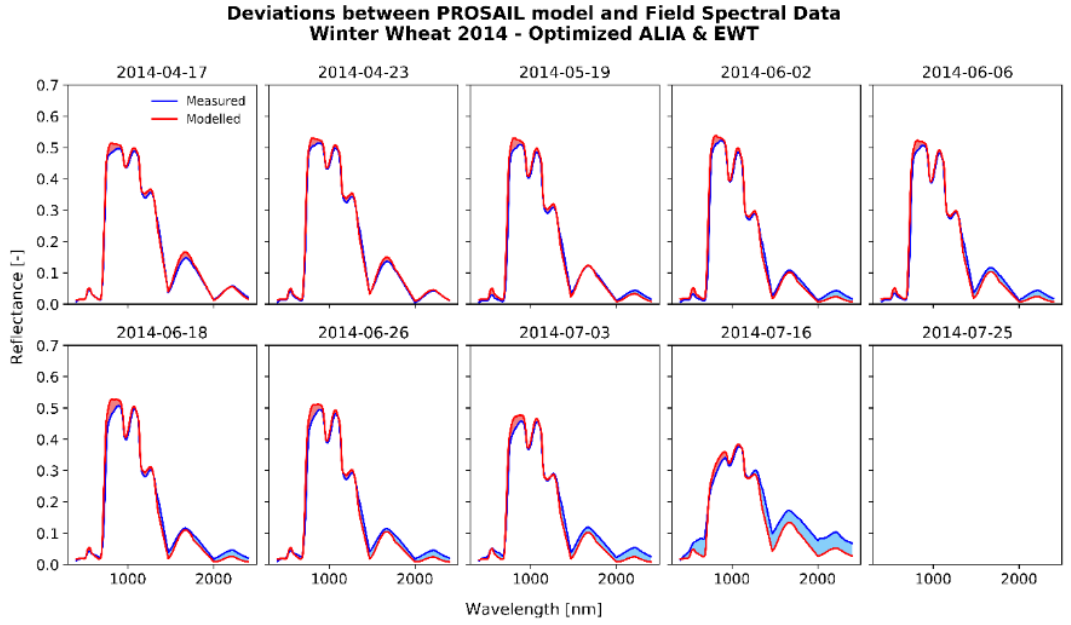


Figure 2.6.5. Mean deviations as the difference between field spectral measurements and PROSAIL model output, aggregated into BBCH growth stages. Positive values indicate an underestimation, negative values an overestimation of the model. Seasonal patterns are more distinct for winter wheat than for silage maize with emphasis on deviations in the NIR region.

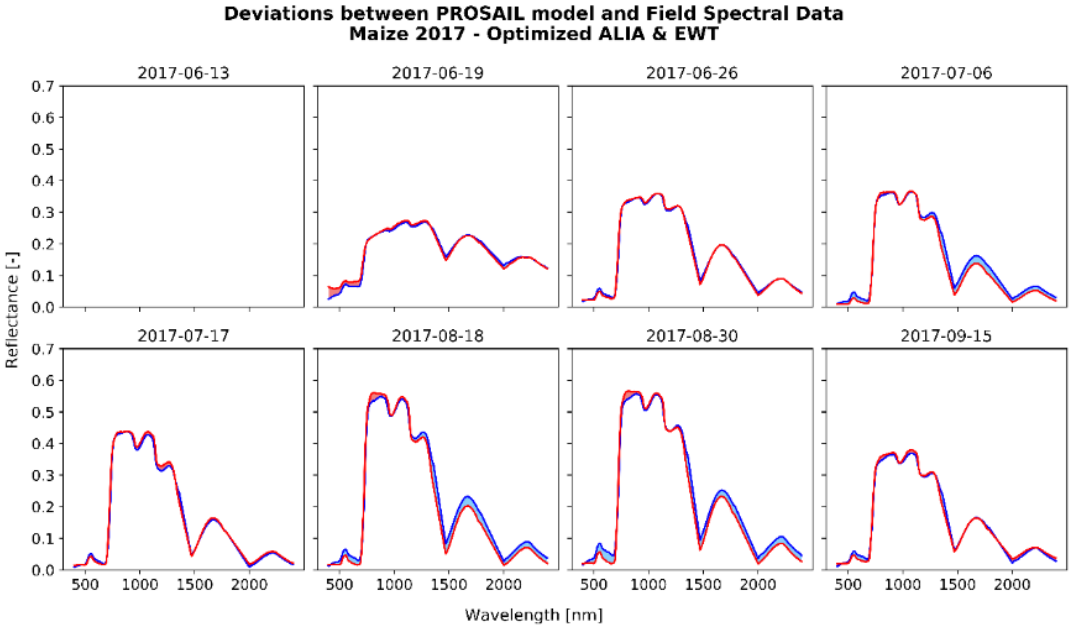
3.2. Optimized Parameter Sets

3.2.1. The Fitting Process

The fitting of ALIA and EWT led to a better agreement for example, for the 2014 winter wheat (Figure 2.6.6a) and the 2017 silage maize (Figure 2.6.6b) dataset – illustrations of all seasons are found in the Supplement (wheat: S3; maize: S4). But even after the first manual optimization, both crops show an overestimation from the red edge to the NIR shoulder. More precisely, PROSAIL models a plateau-like shape in this region with a distinct convex edge, which is not present in the field spectra. A beneficial solution for this occurrence is the fitting of C_{brown} into $C_{brown,opt}$. This second step is demonstrated for the same example years for wheat (2014, Figure 2.6.7a) and maize (2017, Figure 2.6.7b) The results for the final optimization of the other years are found in the Supplement (wheat: S5; maize: S6).

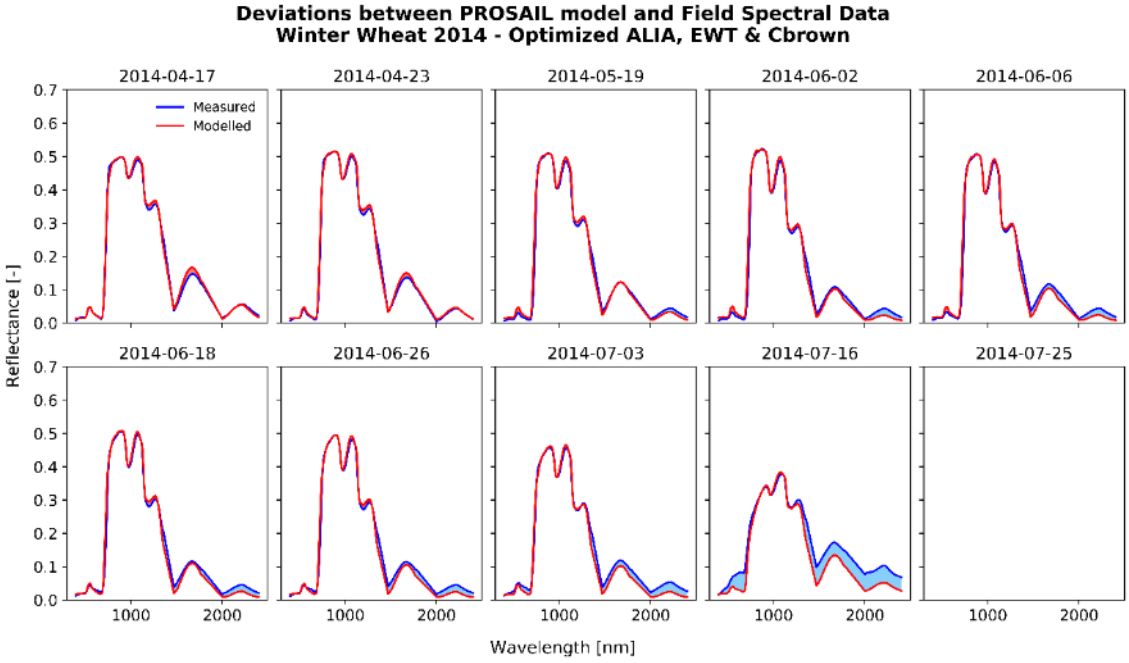


(**a**)

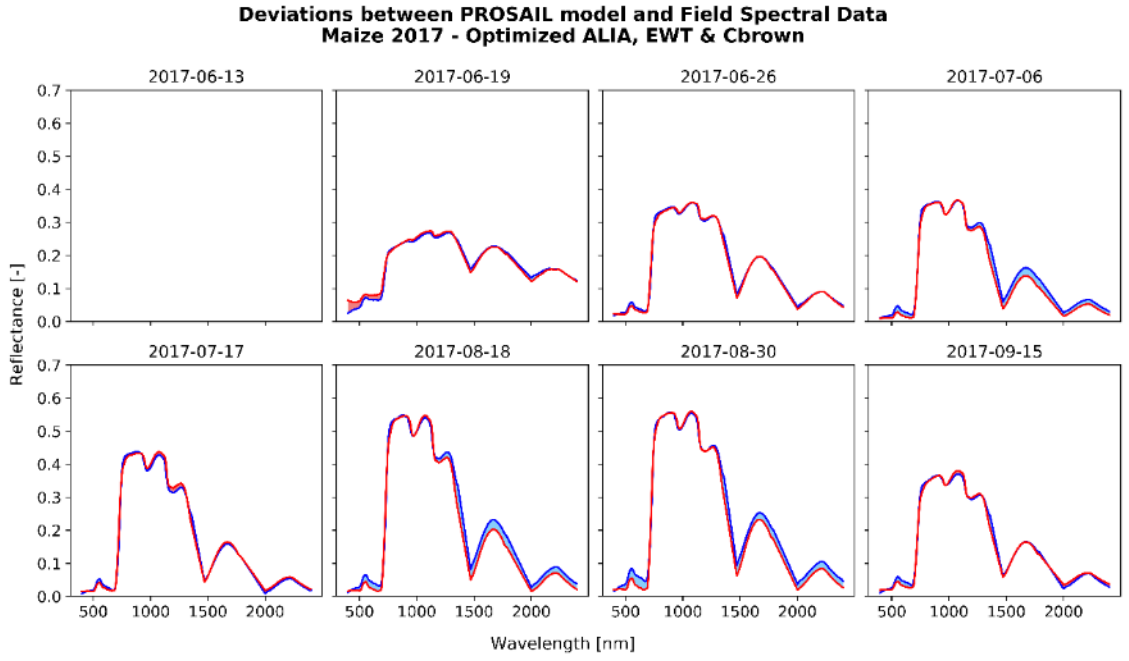


(**b**)

Figure 2.6.6. In the first step of the optimization, ALIA and EWT were fitted in the NIR region. This is demonstrated for winter wheat season 2014 (**a**) and silage maize season 2017 (**b**).



(**a**)



(**b**)

Figure 2.6.7. In the second step of the optimization, C_{brown} was fitted in the red edge region. Examples demonstrate the final fitting results for winter wheat 2014 (**a**) and silage maize 2017 (**b**).

Even though there were individual differences in the development of deviations between measured and modelled spectra for the four growing seasons, there is a greater pattern recognizable. To make these independent of the Julian days, those deviations were averaged for all macro stages according to the BBCH-scale (Figure 2.6.8) as RMSD values. Deviations for winter wheat are largest for the original data with elevated errors in the stages of stem elongation and booting as well as for fruit development and ripening (Figure 2.6.8a). After adapting to the measured spectra, deviations stay within a narrow boundary below 0.02 reflectance with smaller values for stages of intensive production of fresh green biomass and greater ones at crop maturity (Figure 2.6.8b).

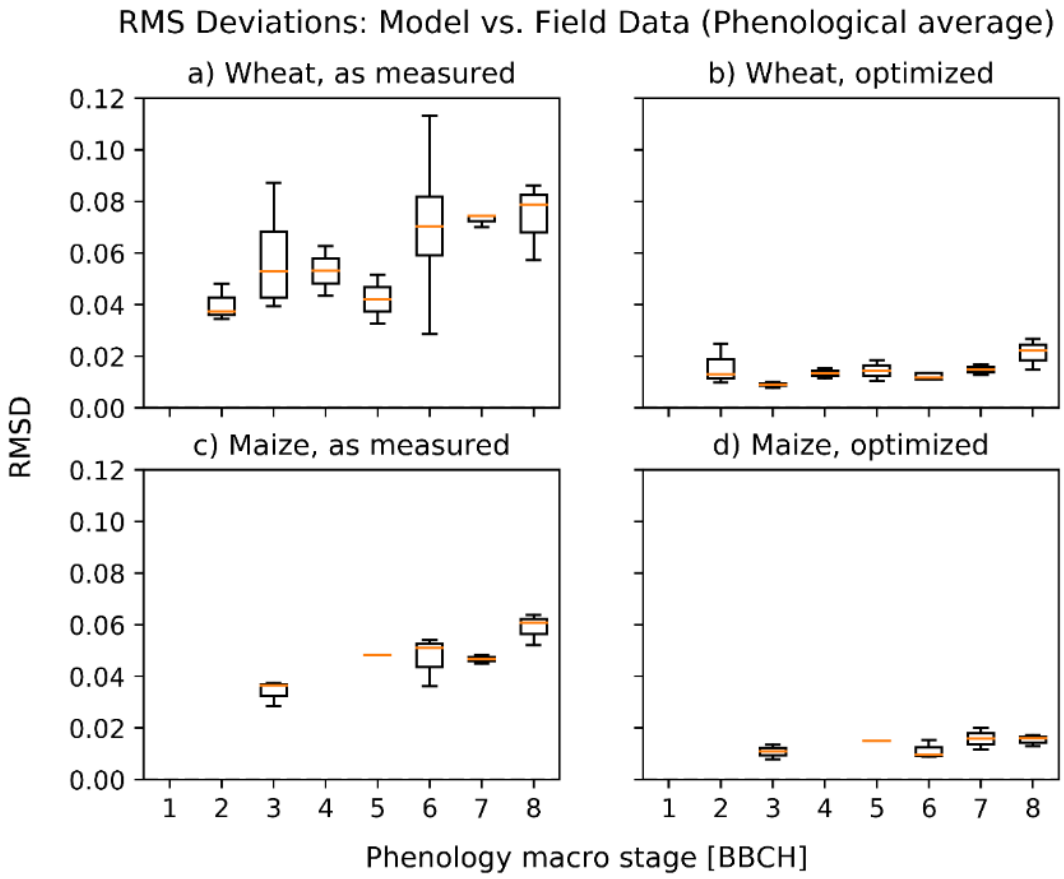


Figure 2.6.8. Results of the spectral fitting aggregated into BBCH growth stages. RMSD values were first calculated for the full range of the spectrum without adaptation (wheat: **a**, maize: **c**). A higher accuracy was obtained after fitting the spectral curves in the NIR range by changing ALIA, EWT and C_{brown} (wheat: **b**, maize: **d**).

In contrast to winter wheat, deviations between modelled and measured spectra of silage maize showed no seasonal fluctuation neither before nor after ALIA, EWT and C_{brown} were altered (Figure 2.6.8c,d). RMSDs only increase with ongoing phenological development and are on average 18% lower, indicating a better predictability for maize canopies than for winter wheat. Please note that macro stages 2 and 4 do not exist in the BBCH phenological categorization for maize.

3.2.2. Analysis of the Optimized Variables for Winter Wheat

To analyse the correspondence between measured and optimized variables, their seasonal progress is plotted in Figure 2.6.9 and their correlations are listed in Table 2.6.5. A full comparison between ALIA, EWT and C_{brown} as observed in situ versus the results of the two-step manual optimization process is listed in the Supplement for all field dates of the four seasons of winter wheat field campaigns (see Supplement S7).

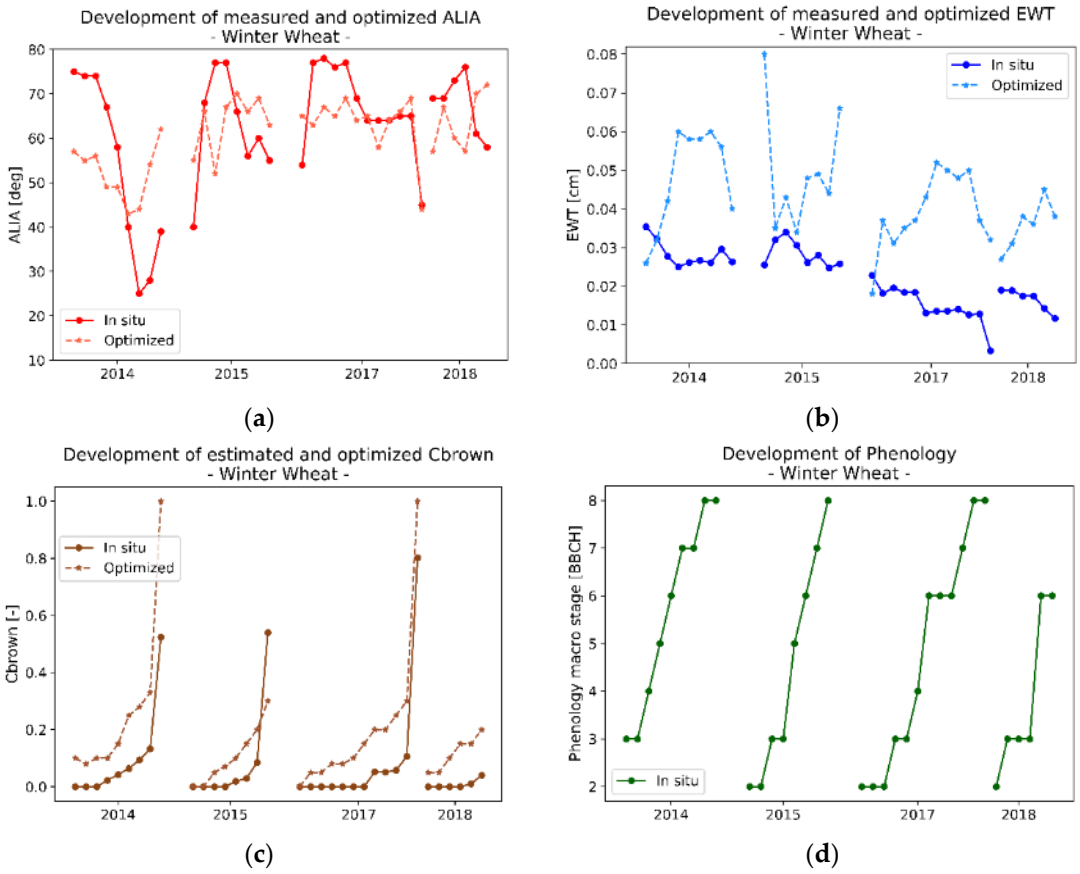


Figure 2.6.9. All values of ALIA (**a**), EWT (**b**), C_{brown} (**c**) and Phenology (**d**) for the four winter wheat field campaigns of 2014, 2015, 2017 and 2018. In situ measurements (**a** & **b**) and estimations (**c**) are shown as solid lines; optimized parameters are drawn with a dashed line style.

Table 2.6.5. Distance measures between in situ observations of ALIA, EWT and C_{brown} vs. optimized values in the manual fitting process for winter wheat.

Variable	Season	RMSE	rRMSE	R^2
ALIA	2014	18.2°	0.34	0.12
	2015	12.3°	0.20	0.02
	2017	7.7°	0.12	0.47
	2018	12.6°	0.19	0.77
	All	12.9°	0.21	0.18
EWT	2014	0.025 cm	0.87	0.65
	2015	0.027 cm	0.96	0.37
	2017	0.027 cm	1.8	0.16
	2018	0.021 cm	1.26	0.47
	All	0.026 cm	1.18	0.02
C_{brown}	2014	0.21	2.10	0.99
	2015	0.11	1.33	0.69
	2017	0.13	1.48	0.96
	2018	0.12	14.1	0.57
	All	0.15	1.94	0.79

The development of ALIA is quite stable in all seasons of wheat. After high inclinations at the beginning of the growing period, leaves tend to bend down and finally surpass the 0° horizontal line to point towards the soil, which again results in increasing angles for senescent stages. These tendencies are only partly illustrated in Figure 2.6.9a, because field dates with low LAI were not optimized and are

omitted in the graph. Still, winter wheat reveals notable deviations between these findings and the calibrated ALIA values. RMSEs range from 8° to 18°, which corresponds to relative RMSEs of 0.19 and 0.34. Inter-seasonal variation was 29%. A negative slope for the 2018 regression ($R^2 = 0.77$, slope = −0.85) and the low R^2 of 0.18 for the complete time series affirm the impression that $ALIA_{opt}$ is independent of the original ALIA.

For EWT, even stronger seasonal patterns are visible. Young wheat leaves hold highest water contents but then continuously desiccate with only rare disturbances, for example, in the early growth cycle of 2015, when the canopy first had to recover from a particularly dry spring. When approaching late senescence, leaf water drops <0.01 cm. After fitting the spectral signatures in the 970 nm and 1060 nm region and comparing EWT with the fitted EWT_{opt}, relative errors between the two variables range between 0.87 and 1.8 and are 1.18 for the complete time series of four years. This means that the distance of the optimized EWT to its original field representation on average ranges beyond 100%. EWT_{opt} follows a different seasonal pattern connected to the green biomass. In fact, correlation between EWT_{opt} and LAI range between $R^2 = 0.27$ (2014) and $R^2 = 0.6$ (2017), confirming the bias present in the data.

In situ estimations of C_{brown} were 0.0 during the vegetative stages of winter wheat. Canopies showed no senescent spots on leaves until beginning of ripening. Once senescence was initiated also in the top layer of the canopy at the beginning of July, C_{brown} skyrocketed within a few days' time. With these in situ measured values, however, no sufficient fitting was possible in the red edge region. A better match with modelled spectra was obtained when $C_{brown,opt}$ was continuously increased throughout the season. Deviations are in a constant range from 0.1 to 0.2. Similar to EWT, the relative distances to the increased optimized variable set reach disproportional magnitudes due to absolute values below 0.1 (see Figure 2.6.9c). It is worth noting, though, that C_{brown} is correlated with $C_{brown,opt}$. The slope of the regression for all four seasons is 0.25 and the intercept is 0.13, which can be considered as a significant bias ($p = 0.017$) that occurs right from the beginning of each season and could be accounted for with a linear model.

3.2.3. Analysis of the Optimized Variables for Silage Maize

In the same manner as it was done for winter wheat, the illustration of the development of the measured and adapted maize variables is shown in Figure 2.6.10 and the secondary statistics are summarized in Table 2.6.6. In situ records in full are listed in Supplement S8.

According to the in situ measured variables, ALIA of maize shows an erratic seasonal trend towards increased inclinations at ripening stages. Intact leaf tips of younger maize plants pitch over but get more and more rigid in their reproductive states. Shortly before vegetation dieback, leaves sag down in parallel direction to the stalks resulting in high inclination angles. This pattern is observable for all three seasons and partially reflected by $ALIA_{opt}$. Large deviations in the modelled $ALIA_{opt}$ occur as singular events at the end of 2017 and the beginning of 2018.

Inter-seasonal variability of EWT for maize was higher than the inner-seasonal changes. In contrast to winter wheat, water content in maize leaves was stable in all seasons. An intense loss of plant water happened only at the very end of the maturity stage beyond the time series used for fitting the spectra (not shown). The generally lower EWT in the 2017 data suggests a reduced water availability in that year for the respective soil condition of that field despite good meteorological conditions. EWT_{opt} shows significant concordance with in situ values for 2018 but the large relative errors indicate the overall weak predictive power of a PROSAIL simulation even if based on measured ground data. A bias towards biomass is similar to the findings for winter wheat: R^2 with LAI range between 0.33 in 2014 and 0.67 in 2018.

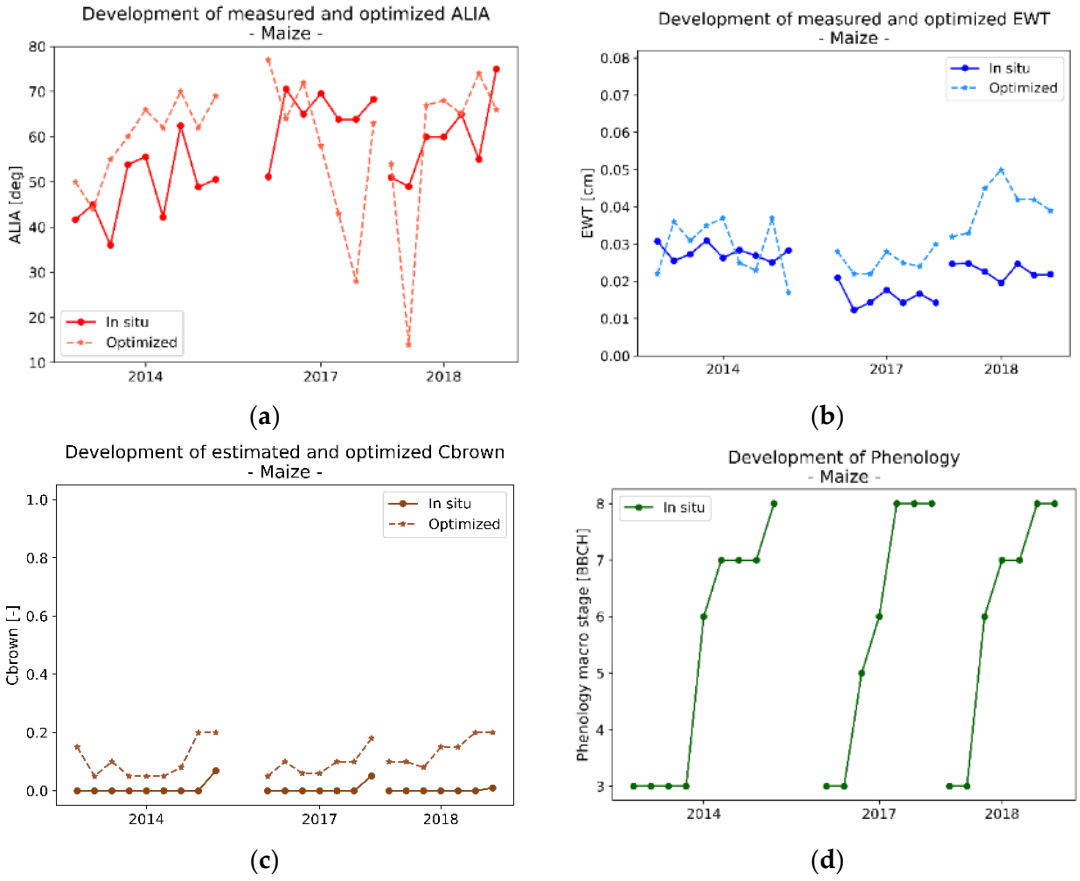


Figure 2.6.10. All values ALIA (**a**), EWT (**b**) and C_{brown} (**c**) and Phenology (**d**) for the three silage maize field campaigns of 2014, 2017 and 2018. In situ measurements (**a** & **b**) and estimations (**c**) are shown as solid lines; optimized parameters are drawn with a dashed line style.

Table 2.6.6. Distance measures between in situ observations of ALIA, EWT and C_{brown} vs. optimized values in the manual fitting process for silage maize.

Variable	Season	RMSE	rRMSE	R^2
ALIA	2014	13.1°	0.27	0.44
	2017	19.4°	0.30	0.06
	2018	16.0°	0.27	0.30
	All	16.1°	0.28	0.04
EWT	2014	0.008 cm	0.30	0.19
	2017	0.010 cm	0.64	0.25
	2018	0.019 cm	0.83	0.62
	All	0.013 cm	0.58	0.01
C_{brown}	2014	0.11	14.16	0.32
	2017	0.09	12.37	0.76
	2018	0.15	101.58	0.30
	All	0.12	20.58	0.24

C_{brown} was 0.0 for all phenological stages of maize until cob ripeness. At senescence, water content dropped, and little brown spots became visible. Unfortunately, higher values of C_{brown} had to be clipped, because they coincide with low green LAI and thus cannot be reasonably optimized. Despite this fact, $C_{brown,opt}$—as parameterized in PROSPECT—needs to be increased right away at stages of leaf development and booting to account for non-visible accumulation of brown pigments within the leaf. This is confirmed by the volatile trend of $C_{brown,opt}$ particularly in the 2014 season which could not be explained by brown leaf spots alone.

When comparing the two different crops, deviations from the in situ measured leaf inclinations were found to be 20% lower for winter wheat when compared to silage maize. A main reason for that can be found in the two striking outliers of 2017 and the one in 2018. When they are excluded from the analysis, $RMSE_{ALIA}$ for maize drops by 10% to 14.5°. $RMSE_{EWT}$ was twice as high for wheat than for maize.

3.3. Seasonal Development of Winter Wheat Canopy Fractions in Sensor View

The mean distribution of plant fractions of the winter wheat canopies – as seen from nadir view – is shown in Figure 2.6.11. Early observations in March and of pre-winter sprouting (only covered in 2014/15) reveal a high share of background pixels. Wheat plants are small and aligned in rows at that time. After some weeks, the vegetation forms a closed canopy of leaves with only little gaps in between. The fraction of dark background $f_{background}$ then remains at a constant level of 10–20%, even when ears begin to grow. Together they extend their visible fraction up to 50% in the years of 2014, 2015 and 2017 and even 70% in the season of 2018.

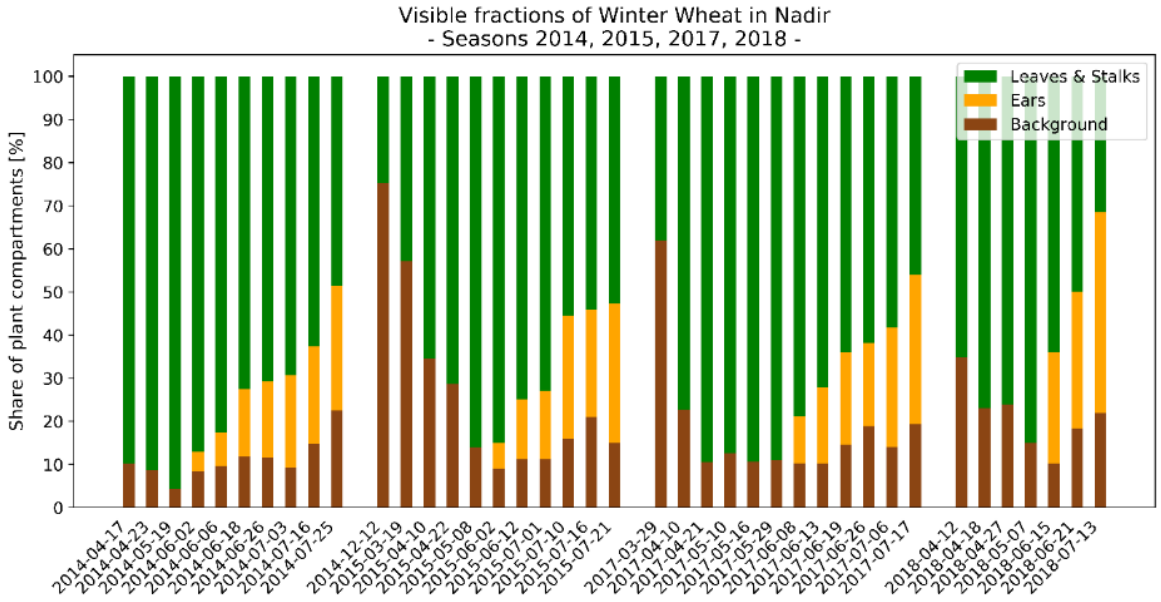


Figure 2.6.11. Seasonal development of plant fractions of winter wheat canopies as they become visible to a sensor that is observing the respective field in nadir view, obtained from nadir RGB image segmentation for four seasons (2014, 2015, 2017, 2018).

The meteorological and micro-ecological conditions vary between the four seasons, which limits the comparability even for the same days of the year. Similarly to the statistical analysis of the optimized sets of variables described in Section 2.4 visible fractions of plant compartments were aggregated in their phenological macro stages to derive a representative seasonal pattern (see Figure 2.6.12). A strengthening of the trends recognized in Figure 2.6.11 can be observed. Lowest values for $f_{background}$ are found in the stage of booting, when the green canopy is dense and covers most of the underlying soil. When the wheat ears are established, their visible influence grows mainly to the expense of leaves and stalks, whereas $f_{background}$ increases only slightly. The agreement between the estimations of $f_{background}$ was highest for low values, for example, an error of 0.3% at stage of emerging inflorescence and 3.4% at booting stage and lowest for the tillering stage with an error of 18%. Errors of the visible fraction of fruit ears were lowest at emergence of the flowers (σ = 0.8%) and fruit development (σ = 1.5%) and highest for flowering (σ = 7.9%) and ripening stages (σ = 7.2%).

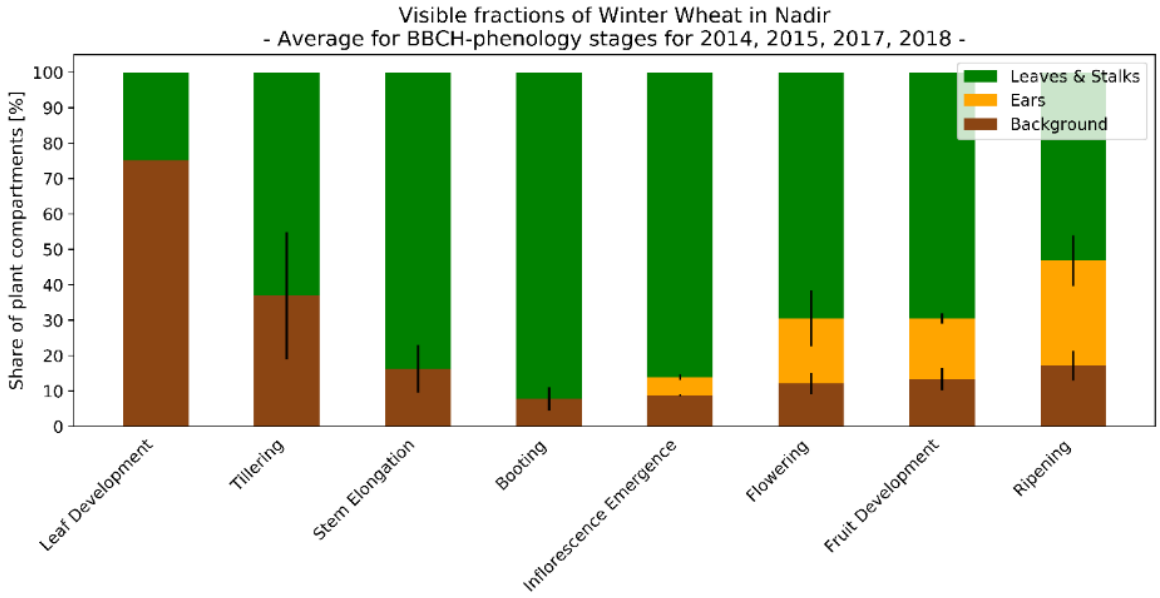


Figure 2.6.12. Development of plant fractions of winter wheat canopies as they become visible to a sensor that is observing the respective field in nadir view at different phenological stages, obtained from nadir RGB image segmentation. Black lines within the bars indicate the standard errors of background and ears.

Combining these results with the model deviations, the relationship between RMSD and the fraction of visible ears is analysed. As shown in Figure 2.6.13, the R^2 of that comparison is a mere 0.01 with a dispersed scatter plot for all data points in which wheat ears were visible in the RGB photography (Figure 2.6.13a). After aggregating the data into phenological stages, their representative values form a linear model with $R^2 = 0.78$ (Figure 2.6.13b). A correlation between the fraction of visible soil and the RMSD by contrast could not be found. R^2 for was 0.02 for all data and 0.1 after aggregation.

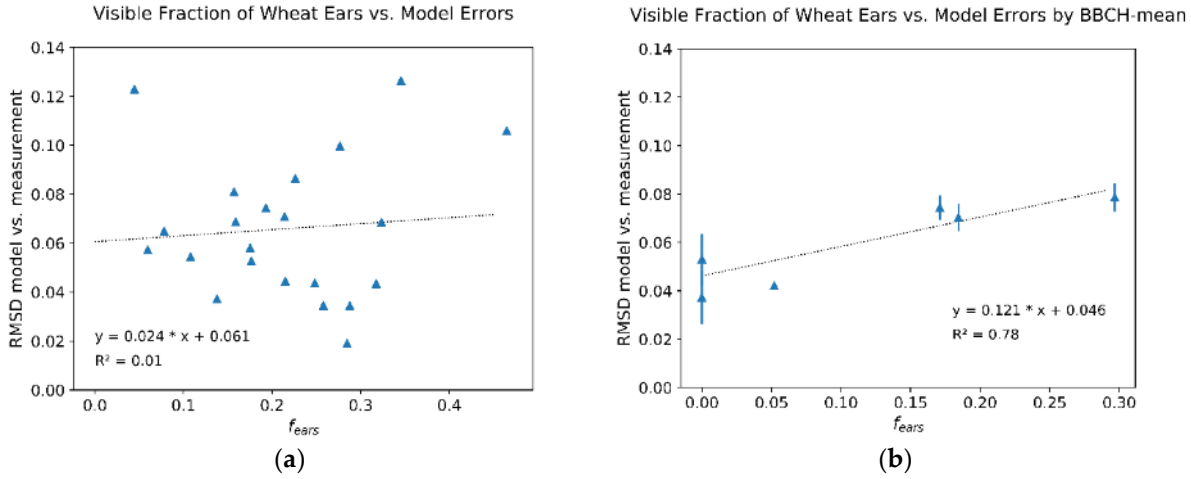


Figure 2.6.13. Dependency between the fraction of visible ears (f_{ears}) and the RMSD of spectral measurement vs. PROSAIL output for all winter wheat data (**a**) and aggregated into phenological macro stages (**b**). Standard deviations of the BBCH-aggregation in **b** are symbolized by vertical error bars.

4. Discussion

The performance of the parameter retrieval of canopy reflectance models is usually evaluated by means of descriptive statistics, comparing measured and estimated variables and assuming that low errors indicate high accuracies [85]. This requires that model parameters and in situ variables are representing the same natural property. Our study demonstrates that this prerequisite does not apply to every case when working with PROSAIL. Model results were significantly different to measured spectra, even when in situ measured variables define the input.

Nevertheless, even reasonable measurement errors of for example, LAI could not explain the observed deviations. Remaining errors are expected to be intrinsic uncertainties of the model design linked with the degree of abstraction. RMSD between spectral field observations of the winter wheat study sites and their model representations reveal that PROSAIL describes canopy properties better for early growth stages. Consequently, they are better suited for retrieving canopy variables than later stages. Within the non-reproductive stages (BBCH 1 – 5) there is an additional local RMSD minimum observed when vegetation initiates the heading of the ears (BBCH 5). This phenomenon is not seen in the maize dataset but the tendency towards increased errors with passing phenological stages here comes with an overall better agreement between model results and spectral measurements.

In PROSAIL, the ALIA plays an important role among structural model parameters with high total scores in the global sensitivity analysis, particularly in the NIR. The biomass density of the canopy is quantified via LAI and manifests in the spectrum through stronger reflectances in the NIR and stronger absorbance in the SWIR. For increased LAI, a saturation in the SWIR by water absorption in optically thick canopies was mentioned by, for example, Datt et al. [86]. In PROSAIL, high reflectance levels of the NIR plateau can only be simulated when assuming low ALIAs. Take note that SAIL is not a geometrical but a radiative transfer model. The incorporation of leaf inclination into the preceding Suits model [87] allowed accounting for the scattering processes that happen within the canopy. Leaf angle densities are calculated according to Campbell [65] representing frequencies of leaf inclinations as discrete classes. For each class, the volume scattering is calculated resulting in the Suits system coefficients, which denote contributions for each inclination class to the basic radiation transfer processes of extinction, attenuation and backscattering from the canopy. The LAI later serves as a scaling factor of these processes. This means that in PROSAIL the ALIA is used to estimate *probabilities* for radiation to be absorbed, attenuated or reflected. In reality, wheat fields are densely seeded, and the complex canopy structure appears closed from tillering stage (BBCH 3) onwards, even though ALIA at that time shows values above 70°. Any kind of minimization between model results and measured spectra therefore suggests a lower ALIA to capture high reflectances in the NIR caused by multiple scattering in thick canopies, especially in stages of intensive biomass production and growth [88,89]. After this period of stretching wheat leaves, the ALIA decreases and is soon overestimated by the model. This could be an indication that $ALIA_{opt}$ aims to reproduce vertical canopy structures, that is, stalks or fruit ears with high inclinations that make up 18% (flowering) to 30% (ripening) of the nadir view. In this regard, winter wheat seems to conflict with the basic turbid medium assumptions of scattering objects of infinitesimal size, disallowing shadowing within the canopy. Also in 4SAIL, the finiteness of leaves is only accounted for by the hot spot effect [90], while consideration of precise geometrical structures is still confined to 3D ray tracing models. Nevertheless, it was conversely expected that 4SAIL presumptions would rather fail for maize. It consists of much larger leaves aligned in a more heterogeneous canopy than wheat fields. Single plants grow larger and form a distinct row structure. It seems that overlapping leaves in the wheat field become stacks of green biomass within the canopy, even for erected leaves. Maize plants stand more isolated and allow radiation to penetrate deeper into the stand. The consideration of using 4SAIL to retrieve vegetation parameters independently of the vegetation type accordingly does not fully hold true.

The estimation of LAI from hyperspectral data has been sufficiently described by other authors, unlike the retrieval of ALIA, which is often treated as a free parameter without final assessment. Few studies took up this issue, like Casa et al. [91] who could not achieve adequate estimates of ALIA for maize data from PROSAIL optimizations but found lower deviations when leaf inclinations were

obtained from measured gap fraction data. The visible fraction of soil as an alternative to the gap fraction in this study did not correlate with optimized ALIA and thus turned out to be inefficient for an improvement of PROSAIL inversions. Botha et al. [92] retrieved canopy structure variables from PROSAIL as a side product to leaf chlorophyll content estimation of wheat canopies, struggling with low correlations between measured and estimated ALIA. They further experienced an overestimation of LAI except for the first growth stage. This supports our finding that wheat reflectances in the NIR are higher than modelled for given LAI in situ data in the essential growing period.

In the first PROSAIL review by Jacquemoud et al. [34], it was stated that spectral responses of LAI and ALIA were closely correlated with each other, making an independent inversion of those parameters problematic. Other studies have tried to overcome this deficit by providing statistical information as a priori boundaries [55], classifying ALIA within a narrow parameter range [93], transforming two-band-indices [94,95] or incorporating the information of pixels in spatial vicinity to obtain an object-based result [96]. Our proposed manual calibration of ALIA could potentially serve as another separation approach when the LAI has been reliably estimated. In all cases, a successful decoupling of canopy structure parameters is aided by integrating multi-angular observations in the inversion process [93,97] for example, from CHRIS/PROBA or future EnMAP.

In situ measured EWT did not agree with the values obtained from the optimization process. One reason is the complex canopy structure linked to the type and amount of biomass observed from the sensor. In the spectral signal, there is no differentiation between phyto-elements of the crop, whereas the model considers EWT as a pure leaf parameter. Figure 2.6.14 shows that wheat stalks contain more water than the leaves per unit fresh mass from June onwards (BBCH 4). Thick maize stalks contained more water than the leaves at all times in the growth cycle. Errors between EWT and EWT_{opt} were conversely higher for wheat than for maize, indicating that EWT_{opt} carries even more residual information of canopy features, which is in line with Clevers et al. (2010) [98]. The interleaving of foliage may lead to an additional vertical stacking of biomass and exaggeration of biophysical and biochemical features in the model. Inversion of PROSPECT alone was shown to produce adequate results for estimation of EWT [99–101], whereas the proper retrieval from complex canopy spectra remains a difficult task [102,103]. Another solution to this problem could be crop- and phenology-specific calibration curves if further research confirms the relationship between EWT and EWT_{opt} found for wheat in the four seasons presented.

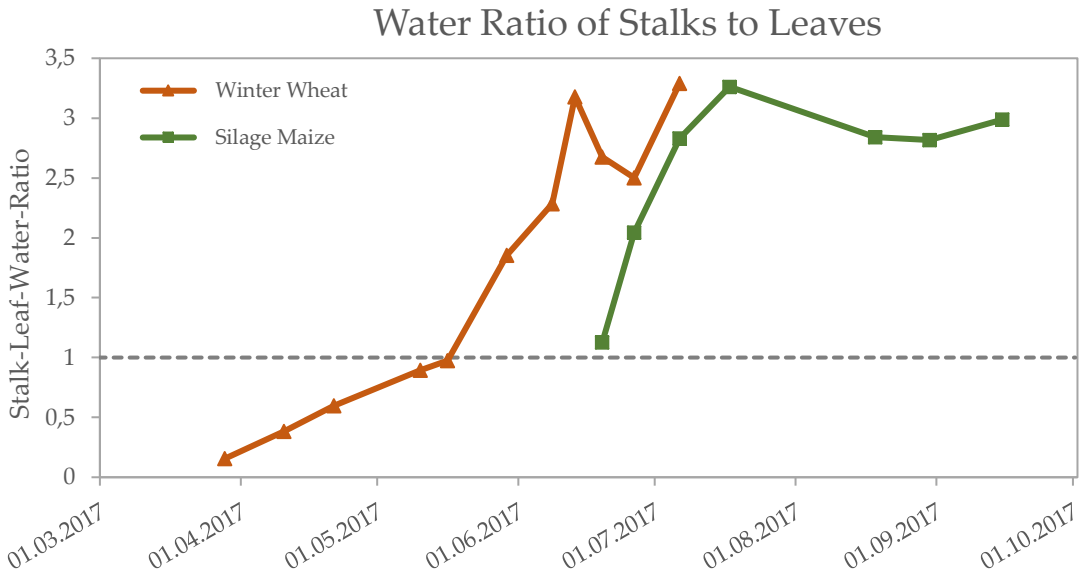


Figure 2.6.14. Ratio between water in the stalks compared with water in the leaves for winter wheat (orange) and maize (green). Water content is standardized to water loss per fresh mass. The grey line illustrates the 1:1 ratio between phyto-elements. Data was recorded at the 2017 MNI campaign.

The need for a proper parameterization of C_{brown} becomes apparent when observing the residual deviations in the red edge region and the absence of a distinct shoulder in the field spectra (Figure 2.6.6). The same spectral shape is found in other publications that analyse dense crops like winter wheat [104] or grassland [51]. Solid assumptions about the biochemical influence of brown leaf pigments are necessary to improve the quality of the retrieval of other plant pigments as well as of LAI and ALIA, which are all sensitive in the far red and NIR. The variable C_{brown} generally lacks a solid scientific study basis. There is no evidence of a successful inversion of C_{brown} in the literature and for forward mode simulations it was often used as a calibration parameter for the spectral region from 500 nm to 1000 nm. The authors of PROSPECT state that the disturbance around the NIR plateau occurs at senescence and for dry leaves [35]. Results of our study show that brown pigments play an important role even for vital green canopies, particularly for maize. Further research should be directed towards proper estimation and interpretation of C_{brown}, identifying valid ranges for different crop types and the relation of C_{brown} with crop phenology to assist the estimation process. Treating C_{brown} as visible leaf browning proved to be insufficient.

Further improvement of retrieval accuracy could of course be achieved by the use of newer versions and more sophisticated modelling approaches. Anthocyanins (C_{anth}) play an important role in the photo-protection of plants [105] and their incorporation into PROSPECT-D was shown to particularly improve the retrieval of carotenoids [36]. But even if C_{anth} had been considered for this study, impact on the calibration of C_{brown} is expected to be low, considering that the latter was adjusted through spectral matching in the red-edge region (750–900 nm) in which C_{car} is not sensitive (see Figure 2.6.2). The same holds true for the deviations occurring in the RED region, which indicate a potential issue in the retrieval of C_{cab} at crop ripening. Running such fitting procedures with different versions of PROSPECT could further identify the role of parameterization of the absorption coefficients and the refractive indices used in each model. Implementing the Leaf Inclination Distribution Function (LIDF) directly could pose a superior alternative to deriving it from ALIA with the Campbell approach, as it allows a more comprehensive description of the canopy geometry. For instance, an ALIA of 45° would result in one specific LIDF when using the ellipsoidal model but could be described by several different distributions when using for example, Verhoef's algorithm with the parameters $LIDF_a$ and $LIDF_b$ (please refer to the review of Wang et al. 2007 for an overview of that topic [106]). Splitting the information into two parameters increases the probability to encounter ill-posed solutions but could also potentially improve the retrieval process, especially in the case of multi-angular observations. Another limitation of this study concerns the choice of the parameter set for optimization. The restriction to ALIA, EWT and C_{brown} leads to an affiliation with the errors of other parameters if their sensitive spectral ranges overlap. Apart from LAI, this could also be the case for C_m, which—despite a good retrievability from PROSPECT model inversions (e.g., [107,108])—is expected to suffer from similar scaling problems like the EWT [109]. On the other hand, it remains problematic that more simultaneously adjusted parameters also result in a more severe ill-posedness. An increase in C_m leads to a decrease of the reflectance in the NIR, just like an increased ALIA does. Consequently, $ALIA_{opt}$, as the result of a calibration process, will also carry a residual signal about the uncertainties of other, less sensitive parameters like C_m and the derived leaf structure parameter N.

Uncertainties in the quantification of in situ variables vary with experience of the field workers, chosen sampling layout, technical equipment and methodology. Matching model output with real spectral observations would assume that the illumination is optimally represented in PROSAIL. The ratio of diffuse to direct radiation, for instance, is controlled by the *skyl* parameter, which was applied as a flat spectrum instead of a wavelength-dependent data vector. A full description of lighting conditions upon the canopy would have to incorporate atmospheric modelling, taking into account the aerosol optical depth, precipitable water and O_3 content [110]. For the manual fitting process, it is assumed that LAI has been correctly measured. In reality, the method of inverting the gap fraction from LAI2200C data introduces uncertainties as well, particularly for non-homogeneous canopies [111]. The impact of a ±10% error range of LAI on model results shown in Figure 2.6.4 confirms, however, that these uncertainties have only a minor influence on the observed deviations between measured and modelled spectra. Spectral output based on optimized ALIA conversely suggests that leaf inclination is

the most important factor for an accurate parameterization of SAIL and that more research is needed to improve its representation in real-life applications. Previous studies have shown that algorithms trained with BOA spectra are more valid for spaceborne sensors with higher spatial resolution due to a reduced averaging of canopy geometrical effects [112,113]. A re-analysis of the modelled and measured spectral observations from space will eventually reveal if the deviations linked to phenological development are consistent and robust towards scaling effects.

5. Conclusion

Various authors have successfully carried out inversions for LAI and leaf chlorophyll content from a variety of crop types using the widely known and applied PROSAIL. However, many studies obtained rather large errors in the retrieval of other important biophysical and biochemical variables, indicating a disagreement between model input and actual in situ canopy properties. We showed that the manual fitting of PROSAIL parameters to match model output with spectra of winter wheat and maize fields led to an adjusted set of ALIA, EWT and C_{brown}. These adapted values show only marginal correlation with the observed in situ values and reveal a distinct crop- and phenology-specific behaviour. It is concluded that disturbing effects from phyto-elements like ears or stalks can either be counterbalanced by an altered ALIA or simulated in more geometrical detail by 3D-models or adapted PROSAIL versions like SLC. Even then, the optimized ALIA cannot be compared to in situ measured leaf angles but rather represents a SAIL-internal parameterization. Setting it to measured values would disallow the retrieval of other PROSAIL parameters, so it is suggested to leave it as a calibration parameter of plant structural traits.

EWT, as retrieved from PROSAIL inversions from bands in the near infrared, will still carry unwanted canopy information. It may thus be useful to investigate the influence of stalks and fruits on the inversion and to decouple leaf variables from residual signals of the canopy structure. Similarly, C_{brown} has a strong impact on the red edge region of vegetation spectra. It cannot be linked to visible leaf browning nor can it be measured directly. Fixing it to a single value will conceal dynamics in the biochemistry of leaves and corrupt the retrieval of other pigments. Instead, it can be well approximated by matching reflectances in the convex NIR shoulder.

For winter wheat, optimized leaf inclinations were found to be flatter than measured in situ to cope with generally high NIR reflectances observed in the field during the peak vegetative stages. After finishing length growth, a converse trend is found that overestimates leaf inclinations in the reproductive stages. In contrast to that, optimization of the maize variables did not follow seasonal patterns but increased towards senescence. Highest accuracies in the parameter retrieval from both winter wheat and maize spectra can be expected at the earlier stages of canopy development. In the turbid medium approach of PROSAIL, ALIA should be considered as a concept to model scattering processes within the canopy. Simply transferring it to real-life situations and ascribing a strict physical meaning to it has shown to be problematic. Thus, it is recommended to allow free assignment of ALIA within constrained ranges even if true leaf inclinations are known. The retrieval of structural parameters from PROSAIL should conversely be performed in one global step instead of decoupling them by fixating single parameters.

Supplementary Materials: Figure S1: Deviations between PROSAIL model and Field Spectral Data; Winter Wheat 2014, 2015, 2017, 2018—As Measured. Figure S2: Deviations between PROSAIL model and Field Spectral Data; Maize 2014, 2017, 2018—As Measured. Figure S3: Deviations between PROSAIL model and Field Spectral Data; Winter Wheat 2014, 2015, 2017, 2018—Optimized ALIA & EWT. Figure S4: Deviations between PROSAIL model and Field Spectral Data; Maize 2014, 2017, 2018—Optimized ALIA & EWT. Figure S5: Deviations between PROSAIL model and Field Spectral Data; Winter Wheat 2014, 2015, 2017, 2018 – Optimized ALIA, EWT & Cbrown. Figure S6: Deviations between PROSAIL model and Field Spectral Data; Maize 2014, 2017, 2018—Optimized ALIA, EWT & Cbrown. Table S7: Manual fitting of ALIA, EWT and C_{brown} for four winter wheat growing seasons. Table S8: Manual fitting of ALIA, EWT and C_{brown} for three silage maize growing seasons.

Author Contributions: Conceptualization, M.D., K.B. and T.H.; Data curation, M.D. and M.W.; Funding acquisition, W.M. and T.H.; Investigation, M.D. and K.B.; Methodology, M.D., K.B., M.W. and T.H.; Project

administration, W.M. and T.H.; Software, M.D. and M.W.; Supervision, W.M. and T.H.; Validation, M.W.; Visualization, M.D.; Writing—original draft, M.D.; Writing—review & editing, M.D., K.B., M.W. and T.H..

Funding: The research presented in this article was conducted at the Chair of Geography and Remote Sensing, Department of Geography, Ludwig-Maximilians-Universität Munich. The financial support through the Space Administration of the German Aerospace Center (DLR) in the frame of the project "EnMAP Scientific Advisory Group Phase III—Developing the EnMAP Managed Vegetation Scientific Processor" through funding by the German Ministry of Economics and Technology under the grant code 50EE1623 is gratefully acknowledged.

Acknowledgments: We want to thank the communal farms of Munich for granting access to the sampling plots and for providing auxiliary data.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript and in the decision to publish the results.

References

1. Hanes, J. Biophysical applications of satellite remote sensing. Springer-Verlag Berlin: Heidelberg, Germany, 2013; p XIV, 230.
2. Webber, H.; Ewert, F.; Kimball, B.; Siebert, S.; White, J.; Wall, G.; Ottman, M.; Trawally, D.; Gaiser, T. Simulating canopy temperature for modelling heat stress in cereals. *Environ. Model. Softw.* **2016**, *77*, 143–155.
3. Chalker-Scott, L. Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochem. Photobiol.* **1999**, *70*, 1–9.
4. Daughtry, C. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote. Sens. Environ.* **2000**, *74*, 229–239.
5. Gitelson, A.A.; Gamon, J.A.; Solovchenko, A. Multiple drivers of seasonal change in PRI: Implications for photosynthesis 1. Leaf level. *Remote. Sens. Environ.* **2017**, *191*, 110–116.
6. Schweiger, A.K.; Schütz, M.; Risch, A.C.; Kneubühler, M.; Haller, R.; Schaepman, M.E. How to predict plant functional types using imaging spectroscopy: Linking vegetation community traits, plant functional types and spectral response. *Methods Ecol. Evolut.* **2017**, *8*, 86–95.
7. Van Der Tol, C.; Verhoef, W.; Timmermans, J.; Verhoef, A.; Su, Z. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences* **2009**, *6*, 3109–3129.
8. Shangguan, Z.; Shao, M.; Dyckmans, J. Effects of Nitrogen Nutrition and Water Deficit on Net Photosynthetic Rate and Chlorophyll Fluorescence in Winter Wheat. *J. N.a. Physiol.* **2000**, *156*, 46–51.
9. Zhang, Y.; Guanter, L.; Berry, J.A.; Joiner, J.; Van Der Tol, C.; Huete, A.; Gitelson, A.; Voigt, M.; Kohler, P. Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models. *Chang. Boil.* **2014**, *20*, 3727–3742.
10. Pearce, R.B.; Brown, R.H.; Blaser, R.E. Relationships between Leaf Area Index, Light Interception and Net Photosynthesis in Orchardgrass1. *Crop. Sci.* **1965**, *5*, 553.
11. Richards, R. Selectable traits to increase crop photosynthesis and yield of grain crops. *J. Exp. Bot.* **2000**, *51*, 447–458.
12. Hank, T.B.; Bach, H.; Mauser, W. Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote. Sens.* **2015**, *7*, 3934–3965.
13. Sid'Ko, A.; Botvich, I.; Pisman, T.; Shevyrnogov, A. Estimation of chlorophyll content and yield of wheat crops from reflectance spectra obtained by ground-based remote measurements. *N.a. Crop.* **2017**, *207*, 24–29.
14. Thorp, K.; Wang, G.; Bronson, K.; Badaruddin, M.; Mon, J.; Thorp, K. Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield. *Comput. Electron. Agric.* **2017**, *136*, 1–12.
15. Pantazi, X.; Moshou, D.; Alexandridis, T.; Whetton, R.; Mouazen, A. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **2016**, *121*, 57–65.
16. Ustin, S.L.; Gitelson, A.A.; Jacquemoud, S.; Schaepman, M.; Asner, G.P.; Gamon, J.A.; Zarco-Tejada, P. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote. Sens. Environ.* **2009**, 113, 67-77.

17. Ceccato, P.; Flasse, S.; Tarantola, S.; Jacquemoud, S.; Grégoire, J.-M. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote. Sens. Environ.* **2001**, *77*, 22–33.
18. Cornelissen, J.H.C.; Lavorel, S.; Garnier, E.; Díaz, S.; Buchmann, N.; Gurvich, D.E.; Reich, P.B.; Ter Steege, H.; Morgan, H.D.; Van Der Heijden, M.G.A.; et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust. J. Bot.* **2003**, *51*, 335–380.
19. Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Mohammed, I.A. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 697–723.
20. Sonobe, R.; Wang, Q. Nondestructive assessments of carotenoids content of broadleaved plant species using hyperspectral indices. *Comput. Electron. Agric.* **2018**, *145*, 18–26.
21. Haboudane, D. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote. Sens. Environ.* **2004**, *90*, 337–352.
22. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. N.a. Physiol.* **2004**, *161*, 165–173.
23. Thorp, K.; Gore, M.; Andrade-Sanchez, P.; Carmo-Silva, A.; Welch, S.; White, J.; French, A.; Thorp, K. Proximal hyperspectral sensing and data analysis approaches for field-based plant phenomics. *Comput. Electron. Agric.* **2015**, *118*, 225–236.
24. Gilbertson, J.K.; Van Niekerk, A. Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Comput. Electron. Agric.* **2017**, *142*, 50–58.
25. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Sens.* **2011**, *66*, 247–259.
26. Verger, A.; Baret, F.; Camacho, F. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote. Sens. Environ.* **2011**, *115*, 415–426.
27. Atzberger, C.; Guérif, M.; Baret, F.; Werner, W. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Comput. Electron. Agric.* **2010**, *73*, 165–173.
28. Verrelst, J.; Malenovský, Z.; Van Der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.-P.; Lewis, P.; North, P.; Moreno, J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surv. Geophys.* **2018**, 1–41.
29. Baret, F.; Buis, S. Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems. In *Advances in Land Remote Sensing*; Springer Nature, 2008; pp. 173–201.
30. Kuester, T.; Spengler, D. Structural and Spectral Analysis of Cereal Canopy Reflectance and Reflectance Anisotropy. *Remote. Sens.* **2018**, *10*, 1767.
31. Disney, M.; Lewis, P.; North, P. Monte Carlo ray tracing in optical canopy reflectance modelling. *Sens. Rev.* **2000**, *18*, 163–196.
32. Gastellu-Etchegorry, J.; Demarez, V.; Pinel, V.; Zagolski, F. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote. Sens. Environ.* **1996**, *58*, 131–156.
33. Govaerts, Y.; Verstraete, M. Raytran: a Monte Carlo ray-tracing model to compute light scattering in three-dimensional heterogeneous media. *IEEE Trans. Geosci. Sens.* **1998**, *36*, 493–505.
34. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. Prospect + sail models: A review of use for vegetation characterization. Remote Sensing of Environment 2009, 113, Supplement 1, S56–S66.
35. Féret, J.-B.; François, C.; Asner, G.P.; Gitelson, A.A.; Martin, R.E.; Bidel, L.P.; Ustin, S.L.; Le Maire, G.; Jacquemoud, S. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote. Sens. Environ.* **2008**, *112*, 3030–3043.
36. Féret, J.-B.; Gitelson, A.; Noble, S.; Jacquemoud, S. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote. Sens. Environ.* **2017**, *193*, 204–215.
37. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote. Sens. Environ.* **1990**, *34*, 75–91.
38. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote. Sens. Environ.* **1984**, *16*, 125–141.
39. Verhoef, W.; Jia, L.; Xiao, Q.; Su, Z. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. *IEEE Trans. Geosci. Sens.* **2007**, *45*, 1808–1822.

40. Pu, R. Hyperspectral remote sensing: Fundamentals and practices. CRC Press: Boca Raton, USA, 2017.
41. Hank, T.B.; Berger, K.; Bach, H.; Clevers, J.G.P.W.; Gitelson, A.; Zarco-Tejada, P.; Mauser, W. Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. *Surv. Geophys.* **2018**, 1–37.
42. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote. Sens.* **2015**, *7*, 8830–8857.
43. Candela, L.; Formaro, R.; Guarini, R.; Loizzo, R.; Longo, F.; Varacalli, G. In The prisma mission, Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, Beijing, China, 2016; IEEE: Beijing, China, pp 253–256.
44. Feingersh, T.; Ben Dor, E. *SHALOM - A Commercial Hyperspectral Space Mission*; Wiley, 2015; pp. 247–263.
45. Lee, C.M.; Cable, M.L.; Hook, S.J.; Green, R.O.; Ustin, S.L.; Mandl, D.J.; Middleton, E.M. An introduction to the nasa hyperspectral infrared imager (hyspiri) mission and preparatory activities. Remote Sens. Environ. 2015, 167, 6–19.
46. Nieke, J.; Rast, M. In Towards the copernicus hyperspectral imaging mission for the environment (chime), IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018; IEEE: Valencia, Spain, pp 157–159.
47. Berger, K.; Atzberger, C.; Danner, M.; D'Urso, G.; Mauser, W.; Vuolo, F.; Hank, T. Evaluation of the prosail model capabilities for the future enmap model environment: A review study. *Remote Sens.* **2017**, under review.
48. Richter, K.; Hank, T.B.; Vuolo, F.; Mauser, W.; D'Urso, G. Optimal Exploitation of the Sentinel-2 Spectral Capabilities for Crop Leaf Area Index Mapping. *Remote. Sens.* **2012**, *4*, 561–582.
49. Danner, M.; Berger, K.; Wocher, M.; Mauser, W.; Hank, T. Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy. *Remote. Sens.* **2017**, *9*, 726.
50. Verrelst, J.; Rivera, J.P.; Leonenko, G.; Alonso, L.; Moreno, J. Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. *IEEE Trans. Geosci. Sens.* **2014**, *52*, 257–269.
51. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote. Sens. Environ.* **2008**, *112*, 2592–2604.
52. Locherer, M.; Hank, T.; Danner, M.; Mauser, W. Retrieval of Seasonal Leaf Area Index from Simulated EnMAP Data through Optimized LUT-Based Inversion of the PROSAIL Model. *Remote. Sens.* **2015**, *7*, 10321–10346.
53. Kimes, D.; Knyazikhin, Y.; Privette, J.; Abuelgasim, A.; Gao, F. Inversion methods for physically-based models. *Sens. Rev.* **2000**, *18*, 381–439.
54. Rivera, J.P.; Verrelst, J.; Leonenko, G.; Moreno, J. Multiple Cost Functions and Regularization Options for Improved Retrieval of Leaf Chlorophyll Content and LAI through Inversion of the PROSAIL Model. *Remote. Sens.* **2013**, *5*, 3280–3304.
55. Lauvernet, C.; Baret, F.; Hascoët, L.; Buis, S.; Le Dimet, F.-X. Multitemporal-patch ensemble inversion of coupled surface–atmosphere radiative transfer models for land surface characterization. *Remote. Sens. Environ.* **2008**, *112*, 851–861.
56. Atzberger, C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote. Sens. Environ.* **2004**, *93*, 53–67.
57. Broge, N.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote. Sens. Environ.* **2001**, *76*, 156–172.
58. Weiss, M.; Baret, F.; Myneni, R.B.; Pragnère, A.; Knyazikhin, Y. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* **2000**, *20*, 3–22.
59. Berger, K.; Atzberger, C.; Danner, M.; Wocher, M.; Mauser, W.; Hank, T. Model-Based Optimization of Spectral Sampling for the Retrieval of Crop Variables with the PROSAIL Model. *Remote. Sens.* **2018**, *10*, 2063.
60. Atzberger, C.; Darvishzadeh, R.; Schlerf, M.; Le Maire, G. Suitability and adaptation of PROSAIL radiative transfer model for hyperspectral grassland studies. *Sens. Lett.* **2013**, *4*, 55–64.
61. ASDInc. Fieldspec 3 user manual (http://www.Geo-informatie.Nl/courses/grs60312/material2017/manuals/600540-jfieldspec3usermanual.Pdf). 2010.
62. Danner, M.; Locherer, M.; Hank, T.; Richter, K. Enmap field guides technical report - spectral sampling with the asd fieldspec 4. 2015.

63. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
64. Suunto. Suunto precision instruments user guide (https://ns.Suunto.Com/manuals/pm-5/userguides/suunto_precisioninstruments_qg_de.Pdf?_ga=2.98826141.267561439.1552297146-115087547.1552297146). 2017.
65. Campbell, G. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. *Agric. Meteorol.* **1986**, *36*, 317–321.
66. Danner, M.; Locherer, M.; Hank, T.; Richter, K. Enmap field guides technical report - measuring leaf area index (lai) with the li-cor lai 2200c or lai-2200 (+ 2200clear kit). 2015.
67. LICOR-Biosciences. Lai-2200c plant canopy analyzer instruction manual. https://licor.app.boxenterprise.net/s/fqjn5mlu8c1a7zir5qel (2019/5/11).
68. Jay, S.; Bendoula, R.; Hadoux, X.; Féret, J.-B.; Gorretta, N. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. *Remote. Sens. Environ.* **2016**, *177*, 220–236.
69. Baret, F.; Hagolle, O.; Geiger, B.; Bicheron, P.; Miras, B.; Huc, M.; Berthelot, B.; Niño, F.; Weiss, M.; Samain, O., et al. Lai, fapar and fcover cyclopes global products derived from vegetation: Part 1: Principles of the algorithm. *Remote Sens. Environ.* **2007**, *110*, 275–286.
70. Jiang, J.; Comar, A.; Burger, P.; Bancal, P.; Weiss, M.; Baret, F. Estimation of leaf traits from reflectance measurements: comparison between methods based on vegetation indices and several versions of the PROSPECT model. *N.a. Methods* **2018**, *14*, 23.
71. Suess, A.; Danner, M.; Obster, C.; Locherer, M.; Hank, T.; Richter, K. Enmap field guides technical report - measuring leaf chlorophyll content with the konica minolta spad-502plus. 2015.
72. Lichtenthaler, H.K. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods in Enzymology* **1987**, *148*, 350–382.
73. Baret, F.; Andrieu, B.; Guyot, G. A Simple Model for Leaf Optical Properties in Visible and Near-Infrared: Application to the Analysis of Spectral Shifts Determinism. In *Applications of Chlorophyll Fluorescence in Photosynthesis Research, Stress Physiology, Hydrobiology and Remote Sensing*; Springer Nature, 1988; pp. 345–351.
74. Bleiholder, H.; Weber, E.; Lancashire, P.; Feller, C.; Buhr, L.; Hess, M.; Wicke, H.; Hack, H.; Meier, U.; Klose, R. Growth stages of mono-and dicotyledonous plants, bbch monograph. In Federal biological research centre for agriculture and forestry, berlin/braunschweig, germany, Meier, U., Ed. Vienna, Austria, 2001; p 158.
75. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421.
76. Van Der Walt, S.; Colbert, S.C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
77. Danner, M.; Wocher, M.; Berger, K.; Mauser, W.; Hank, T. Developing a Sandbox Environment for Prosail, Suitable for Education and Research. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* **2018**, 783–786.
78. Rabe, A.; Jakimow, B.; Thiel, F.; Hostert, P.; van der Linden, S. In Enmap-box 3 a free and open source python plug-in for qgis, IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018; IEEE: Valencia, Spain, pp 7764–7766.
79. François, C.; Ottle, C.; Olioso, A.; Prévot, L.; Bruguier, N.; Ducros, Y. Conversion of 400-1100 nm vegetation albedo measurements into total shortwave broadband albedo using a canopy radiative transfer model. *Agronomie* **2002**, *22*, 611–618.
80. Cannavò, F. Sensitivity analysis for volcanic source modeling quality assessment and model selection. *Comput. Geosci.* **2012**, *44*, 52–59.
81. Lillesaeter, O. Spectral reflectance of partly transmitting leaves: Laboratory measurements and mathematical modeling. *Remote. Sens. Environ.* **1982**, *12*, 247–254.
82. A Sims, D.; A Gamon, J. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. *Remote. Sens. Environ.* **2003**, *84*, 526–537.
83. Bull, C. Wavelength selection for near-infrared reflectance moisture meters. *J. Agric. Eng.* **1991**, *49*, 113–125.
84. Van Der Walt, S.; Schonberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T.; Gomez, S. scikit-image: image processing in Python. *PeerJ* **2014**, *2*, 453.

85. Richter, K.; Atzberger, C.; Hank, T.B.; Mauser, W. Derivation of biophysical variables from Earth observation data: validation and statistical measures. *J. Appl. Sens.* **2012**, *6*, 63557.
86. Datt, B. Remote Sensing of Water Content in Eucalyptus Leaves. *Aust. J. Bot.* **1999**, *47*, 909.
87. Suits, G. The calculation of the directional reflectance of a vegetative canopy. *Remote. Sens. Environ.* **1971**, *2*, 117–125.
88. Knipling, E.B. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote. Sens. Environ.* **1970**, *1*, 155–159.
89. Carter, G.A. Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves. *Am. J. Bot.* **1991**, *78*, 916.
90. Verhoef, W.; Bach, H. Coupled soil–leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote. Sens. Environ.* **2007**, *109*, 166–182.
91. Casa, R.; Baret, F.; Buis, S.; López-Lozano, R.; Pascucci, S.; Palombo, A.; Jones, H.G.; Jones, H. Estimation of maize canopy properties from remote sensing by inversion of 1-D and 4-D models. *Precis. Agric.* **2010**, *11*, 319–334.
92. Botha, E.J.; LeBlon, B.; Zebarth, B.J.; Watmough, J. Non-destructive estimation of wheat leaf chlorophyll content from hyperspectral measurements through analytical model inversion. *Int. J. Sens.* **2010**, *31*, 1679–1697.
93. Tripathi, R.; Sahoo, R.N.; Sehgal, V.K.; Tomar, R.K.; Chakraborty, D.; Nagarajan, S. Inversion of prosail model for retrieval of plant biophysical parameters. *J. Indian Soc. Remote Sens.* **2012**, *40*, 19–28.
94. Zou, X.; Mõttus, M. Retrieving crop leaf tilt angle from imaging spectroscopy data. *Agric. Meteorol.* **2015**, *205*, 73–82.
95. Zou, X.; Mõttus, M. Sensitivity of Common Vegetation Indices to the Canopy Structure of Field Crops. *Remote. Sens.* **2017**, *9*, 994.
96. Atzberger, C.; Richter, K. Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote. Sens. Environ.* **2012**, *120*, 208–218.
97. Laurent, V.C.; Verhoef, W.; Clevers, J.G.; Schaepman, M.E. Inversion of a coupled canopy–atmosphere model using multi-angular top-of-atmosphere radiance data: A forest case study. *Remote. Sens. Environ.* **2011**, *115*, 2603–2612.
98. Clevers, J.; Kooistra, L.; Schaepman, M.; Clevers, J.; Schaepman, M. Estimating canopy water content using hyperspectral remote sensing data. *Int. J. Appl. Earth Obs. Geoinformation* **2010**, *12*, 119–125.
99. Jacquemoud, S. Comparison of Four Radiative Transfer Models to Simulate Plant Canopies Reflectance Direct and Inverse Mode. *Remote. Sens. Environ.* **2000**, *74*, 471–481.
100. Newnham, G.; Burt, T. In Validation of a leaf reflectance and transmittance model for three agricultural crop species, Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International, Sydney, Australia, 2001; IEEE: Sydney, Australia, pp 2976–2978.
101. Baret, F.; Fourty, T. Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie* **1997**, *17*, 455–464.
102. Jacquemoud, S.; Ustin, S. In Application of radiative transfer models to moisture content estimation and burned land mapping, 4th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management, Ghent, Belgium, 2003; Ghent, Belgium.
103. Wocher, M.; Berger, K.; Danner, M.; Mauser, W.; Hank, T. Physically-Based Retrieval of Canopy Equivalent Water Thickness Using Hyperspectral Data. *Remote. Sens.* **2018**, *10*, 1924.
104. Huber, K.; Dorigo, W.; Bauer, T.; Eitzinger, S.; Haumann, J.; Kaiser, G.; Linke, R.; Postl, W.; Rischbeck, P.; Schneider, W. In Changes in spectral reflectance of crop canopies due to drought stress, Remote Sensing for Agriculture, Ecosystems, and Hydrology VII, 2005; International Society for Optics and Photonics: p 59761I.
105. Landi, M.; Tattini, M.; Gould, K.S. Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* **2015**, *119*, 4–17.
106. Wang, W.-M.; Li, Z.-L.; Su, H.-B. Comparison of leaf angle distribution functions: Effects on extinction coefficient and fraction of sunlit foliage. *Agric. Meteorol.* **2007**, *143*, 106–122.
107. Ali, A.M.; Darvishzadeh, R.; Skidmore, A.K.; Van Duren, I.; Heiden, U.; Heurich, M. Estimating leaf functional traits by inversion of PROSPECT: Assessing leaf dry matter content and specific leaf area in mixed mountainous forest. *Int. J. Appl. Earth Obs. Geoinformation* **2016**, *45*, 66–76.
108. Wang, L.; Qu, J.J.; Hao, X.; Hunt Jr, E.R. Estimating dry matter content from spectral reflectance for green leaves of different species. *Int. J. Remote Sens.* **2011**, *32*, 7097–7109.

109. Casas, A.; Riano, D.; Ustin, S.; Dennison, P.; Salas, J. Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote. Sens. Environ.* **2014**, *148*, 28–41.
110. Dong, T.; Wu, B.; Meng, J.; Du, X.; Shang, J. Sensitivity analysis of retrieving fraction of absorbed photosynthetically active radiation (FPAR) using remote sensing data. *N.a. Ecol. Sin.* **2016**, *36*, 1–7.
111. Jonckheere, I.; Fleck, S.; Nackaerts, K.; Muys, B.; Coppin, P.; Weiss, M.; Baret, F. Review of methods for in situ leaf area index determination: Part i. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* **2004**, *121*, 19–35.
112. Si, Y.; Schlerf, M.; Zurita-Milla, R.; Skidmore, A.; Wang, T. Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote. Sens. Environ.* **2012**, *121*, 415–425.
113. Claverie, M.; Vermote, E.F.; Weiss, M.; Baret, F.; Hagolle, O.; Demarez, V. Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. *Remote. Sens. Environ.* **2013**, *139*, 216–230.

2.7. Preparation of Scientific Publication III

In the third publication of this thesis, the initial objective of creating an automated scientific processor for the retrieval of biochemical and biophysical variables from hyperspectral data was finalized. With the experience of publication II it was decided to abandon the idea of a step-wise inversion of different classes of crop variables. This hierarchical approach was intended to allow the retrieval of each parameter with a highly customized setting of the inversion routine. But even though e.g. LAI and C_w may be optimally inverted with individual settings, the fixation of known parameters turned out to prune the performance by forcing the inversion outside the ranges in which correct results can be obtained. For this reason, the retrieval of LAI, ALIA, C_{cab} and C_m was aspired in a global inversion approach.

A main driver for the experimental work of publication III were practical challenges in making the processor available to the community of users of future EnMAP data. LUT inversions have proven to yield satisfying results for the given task, but even when storing them in raw binary format, they exceed by far an acceptable size for distribution along with the EnMAP-Box. A solution was found in the use of machine learning regression algorithms (MLRAs), which were able to learn the link between spectral signatures and underlying vegetation properties from massive sets of synthetic training data.

In the first part of the paper, four MLRAs were set up: Artificial Neural Network (ANN), Random Forest Regression (RFR), Support Vector Machine Regression (SVR) and Gaussian Process Regression (GPR). Each algorithm was optimized in an individual process by maximizing accuracies at learning and predicting PROSAIL synthetic data. Relevant hyperparameters of the algorithms were first identified and then optimized in a multidimensional grid-search approach.

In the second part, the optimized algorithms were used to predict structural canopy parameters from real-life data. However, none of the models succeeded in this task due to the overfitting that happened at training with synthetic data. The issue was solved by adding artificial noise to PROSAIL spectra in the training process and by re-calibrating outputs with in situ data from the field campaigns.

2.8. Scientific Publication III

Training Machine Learning Regression Algorithms to predict biophysical & biochemical variables from resampled hyperspectral EnMAP data using PROSAIL

Martin Danner[a] *, Matthias Wocher[a], Katja Berger[a], Wolfram Mauser[a] & Tobias Hank[a]

[a] Department of Geography, Ludwig-Maximilians-Universität München, Luisenstraße 37, D-80333 Munich, Germany; m.wocher@iggf.geo.uni-muenchen.de; katja.berger@iggf.geo.uni-muenchen.de; w.mauser@lmu.de; tobias.hank@lmu.de

* Correspondence: m.danner@lmu.de; Tel.: +49-89-2180-6673

Abstract: With access to an unprecedented stream of remote sensing images, there arises also the need for tools and applications to extract relevant information from all this data on an operational basis. Algorithms to automatically retrieve biophysical and biochemical plant traits have been published for sensors like MODIS or Sentinel-2. In this study, we investigate the potential of a scientific processor of managed vegetation from future hyperspectral EnMAP data. Said processor will execute pre-trained Machine Learning Regression Algorithms (MLRA) and deliver information about Leaf Area Index, Average Leaf Inclination Angle, Leaf Chlorophyll Content and Leaf Mass per Area. Training data comprises a synthetic Look-Up-Table (LUT) of PROSAIL modelled vegetation spectra and their associated parameterization. Four MLRAs, namely Artificial Neural Networks (ANN), Random Forest Regression, Support Vector Machine Regression and Gaussian Process Regression were found to predict target biophysical variables with high accuracy of relative error scores between 0.13 and 0.25. Applying the models to real-world data turned out to deliver poor results. Predictive power was restored by applying additive Gaussian noise of σ = 4 % and re-calibrating results with in situ data from winter wheat and silage maize. ANNs excelled in terms of accuracy and – from a practical point of view – also in model size and execution time when spectral bands were transformed into 15 components and their signal scaled by a z-transformation. The optimized ANN algorithms will be provided via the EnMAP-Box to feature a Managed Vegetation Scientific Processor.

Keywords: machine learning; reflectance modelling; hyperspectral remote sensing; EnMAP-Box; radiative transfer model; PROSAIL

1. Introduction

Biophysical and biochemical variables express the state of a plant or canopy. They serve as quantified representations necessary to monitor and model biological, chemical and structural characteristics of vegetation over space and time (Asner 1998; Hanes 2013). On leaf scale, biochemical traits like pigment concentrations or water content affect the very elementary processes such as photosynthesis and biomass production (Gitelson 2018; Thenkabail 2017). The architecture of photosynthetically active plant parts is described on canopy level, mainly by the Leaf Area Index (LAI; Darvishzadeh et al. 2008) and the Leaf Inclination Distribution Function (LIDF; Norman and Campbell 1989; Wang et al. 2007). Complex 3D canopy models like DART (Gastellu-Etchegorry et al. 1996) or HySimCaR (Kuester et al. 2014) need further information about the shape and size of those compartments to be able to trace electromagnetic waves as they pass through the canopy and interact with biomass. A more feasible way to simulate reflectance and transmittance on an operational basis

with fewer input parameters is the use of radiative transfer models based on turbid medium assumptions. In the past three decades, the leaf optical properties model PROSPECT (Féret et al. 2017) and the canopy architecture model SAIL (Verhoef 1984) have emerged in their respective fields of application. There exist several coupled versions of PROSPECT and SAIL which are summarized under the term PROSAIL (Jacquemoud et al. 2009a).

The mentioned biophysical and biochemical variables can be obtained from reflectance signatures by inversion of PROSAIL. This means that canopy reflectance models could potentially be used on an operational basis to deliver input for e.g. land surface process simulations from remote sensing data. Variables like LAI, chlorophyll$_{a+b}$ content (C_{ab}) and Leaf Mass per Area (LMA, i.e. dry matter weight) are of particular interest for driving crop growth models and improve predictions on harvest yield estimations (Hank et al. 2015; Ziliani et al. 2018). In applied ecologies, simulated biophysical and biochemical parameters indicate the state of ecosystems (Feilhauer et al. 2018) and their associated Essential Biodiversity Variables to which remote sensing makes a major contribution (Kattenborn et al. 2018; Paganini et al. 2016).

Approximately 60 % of the earth's land surface is biologically productive. If we subtract the share of forests, a remaining 50,000,000 km² is used for agriculture or provides other ecosystem services (Zabel et al. 2014). Operational retrieval of biophysical parameters is sometimes made available alongside with the spectral product of spaceborne missions (e.g., Level-4 MODIS Leaf Area Index; Myneni et al. 2000). Other data sources rely on software solutions, as is the case with the SNAP toolbox and its Biophysical Processor, which retrieves LAI, fraction of Absorbed Photosynthetically Active Radiation (fAPAR), fraction of Vegetation Cover (fCover), C_{ab} and canopy water content from Sentinel-2 reflectance data (Weiss and Baret 2016).

In a similar manner, the forthcoming mission EnMAP (Environmental Mapping and Analysis Program) shall provide applications to obtain biophysical parameters from hyperspectral signatures (Guanter et al. 2015). The algorithms that link reflectance information with ground estimations need to be trained beforehand to be distributed with future hyperspectral sensor data. Machine Learning Regression Algorithms (MLRA) are able to learn a mathematical relationship between pairs of predictors (X) and targets (Y). After a successful training phase, they estimate new targets from unseen predictors autonomously and are thus very appealing for applications of operational processing of larger data sets. Combining radiative transfer with machine learning opened up a field of hybrid regression methods that solve the inverse problem with varying accuracy and computational demand (Rivera-Caicedo et al. 2017; Verrelst et al. 2018).

In this study, we clarify the following research questions in prospect of a future EnMAP Managed Vegetation Scientific Processor: (1) Which machine learning algorithm does the best job in learning and predicting biophysical and biochemical parameters from PROSAIL synthetic datasets? (2) How do these best-performing settings predict target variables from real world data? (3) How do we overcome practical challenges in making these algorithms available as an automatic processor?

2. Materials and Methods

2.1. Test set data

Data for the test set are taken from the study area Munich-North Isar (MNI), Bavaria, southern Germany (48°16′04″ N, 11°42′45″ E). Full range spectral information was gathered with an ASD FieldSpec 3 Jr. (Boulder, CO, USA) in the vegetative phases of winter wheat in 2014, 2015, 2017 and 2018 as well as of silage maize in 2014, 2017 and 2018. After post-processing of the raw spectral data, they were converted into pseudo-EnMAP reflectances by applying band-wise spectral response functions (Segl et al. 2012). Spectral bands that coincide with regions of atmospheric water absorption (1359 nm – 1465 nm and 1731 nm – 1998 nm) as well as noisy far-SWIR bands (2401 nm – 2439 nm) were excluded in the subsequent analysis.

In situ biophysical variables were measured at the same locations right after spectral data was recorded. For determination of the LAI, a LI-COR Biosciences LAI-2200C was used with particular

attention to the row structure of wheat and maize (LICOR-Biosciences 2015). Average Leaf Inclinations (ALIA) were read from a Suunto PM-5/360 inclinometer placed along the leaf petiole axis and averaged over multiple observations from different leaves of the same plant and multiple plants altogether. The Leaf Inclination Distribution Function was calculated from the ALIA via Campbell's ellipsoidal distribution (Campbell 1986). A Konica-Minolta SPAD-502 was used to sample leaf chlorophyll$_{a+b}$ content (C_{ab}). Leaf Mass per Area (LMA) was sampled in the nearby laboratory by measuring leaf sizes and weighing the samples after dehydration.

In total, the data archive comprises 40 full sets of canopy reflectances and their related in situ variables over four seasons of winter wheat and 26 sets over three seasons of silage maize. A comprehensive overview over the layout of the MNI sampling campaigns and the processing of spectral data and in situ variables is given in (Danner et al. 2019).

All considered MLRAs are able to link spectral input to a set of several target variables at once. This opens the opportunity to either learn biophysical and biochemical parameter one after the other, or conversely several – or all – at once. The global approach neglects the fact that different classes of parameters are better trained with different parameterizations of the MLRA. For this study, the canopy structural parameters (LAI & ALIA) were combined into one mutual group and trained in one step, whereas C_{ab} and LMA were estimated with individually trained models.

2.2. Look-up-Table creation

A data set of synthetic reflectance signatures and biophysical parameters was built with a combination of PROSPECT-D and 4SAIL (called PROSAIL hereinafter). Parameters were drawn at random from Gaussian normal and uniform distributions (Danner et al. 2019). Assuming that physical memory is not a limiting factor, most authors argue in favor of relatively large LUTs of size 50,000 and above. On the other hand, it has been observed that also small LUTs allow a proper inversion, as long as parameters are well distributed within their ranges. Bearing in mind that algorithms will be iteratively optimized to learn the relationship between synthetic spectra and PROSAIL parameters, smaller LUTs allow quicker learning and require less memory. For this study, the basic LUT size is first set to 2000 members. Six of these LUTs were created, each with a different Sun Zenith Angle (30° - 55° in steps of 5°) to cover the range of sun geometries observed in the test data set. The observing sensor is considered in nadir position. Two more synthetic data sets with 5,000 and 10,000 members were built to quantify the effects of LUT size.

2.3. Machine Learning Algorithms

2.3.1. Data pre-processing

The interlink between spectral and biophysical information is complex and non-linear. MLRAs are capable of 'learning' from pairs of training data to predict values which they have not seen at that point in the process. We used python's *scikit-learn* package (Pedregosa et al. 2011) to do the conjoint pre-processing, fitting, prediction and evaluation of target variables. The models were saved to hard drive using binary protocols of the *pickle* module.

The diversity of machine learning requires a flexible chain in preprocessing data. All algorithms were tested for sensitivity on the following transformations:

Scaling of the predictors: the predictors in this case are reflectance outputs of PROSAIL. Different techniques involve the Standard Scaler as a regular z-transformation (Eq. 1), the Min-Max Scaler transforming spectra within a given range (here: Eq. 2 shows the special case of min=0 and max=1), the Max-Abs-Scaler which is similar to the Min-Max-Scaler in setting 1.0 as the upper boundary, but maintaining sparsity by keeping the center of the data in place (Eq. 3), as well as the Robust Scaler which uses the interquartile range to scale the data, making it robust towards outliers (Eq. 4).

$$X_{scaled} = \frac{X - \mu}{\sigma} \quad \text{Equation 1}$$

$$X_{scaled} = \omega \cdot (X - \min(X))$$
$$\omega = \frac{1}{\max(X) - \min(X)} \quad \text{Equation 2}$$

$$X_{scaled} = \frac{X}{\max(|X|)} \quad \text{Equation 3}$$

$$X_{scaled} = \frac{X - Q_{0.25}(X)}{Q_{0.75}(X) - Q_{0.25}(X)} \quad \text{Equation 4}$$

Dimensionality reduction: the magnitude of spectral bands in hyperspectral studies hampers the training process of MLRAs (Rivera-Caicedo et al. 2017). A subset of bands is found by minimizing the covariance of the predictors, namely by a Principal Component Analysis (PCA). The effect of the number of components used for calibration is tested on the best performing parameterization of each algorithm.

2.3.2. Artificial Neural Network

Artificial Neural Networks (ANN) is the umbrella term for all algorithms that learn in layers of interconnected neurons. In this study, a MultiLayer Perceptron network Regression (MLPR) was trained. This widely used approach makes use of error backpropagation through the network of perceptrons without relying on training data of any specific statistical distribution (White 1992). On the downside, ANNs depend stronger on the quality of the input and struggle with unseen values outside the range in which they were trained (Waske et al. 2009). In search of an optimal ANN environment, the following settings were changed:

Activation Function: The activation function is used to transfer the information from each neuron into output signals for the consecutive layer. Activation functions used: Identity function, logistic function, hyperbolic tangent (tanh) function, Rectified Linear Unit (ReLU) function (see Karlik and Olgac 2011 for overview).

Solver: The actual mathematical core process of the MLPR is the optimization of the activation functions until convergence. Two different solvers were tested, namely the ADAM solver (Kingma and Ba 2014) and the L-BFGS solver (Sohl-Dickstein et al. 2014).

Alpha: The regularization term, also known as L2 penalty. Regularization here means that less important features are penalized towards zero, so they gain less weight in the outcome of the regression. Alpha is varied to hold control of (over-)fitting the data in the calibration process. Alpha ∈ [0.001, 0.01, 0.1, 1.0, 10.0, 100.0].

Max_iter: Allowing the solver to cycle through a larger number of iterations increases its chance to converge but takes longer to calculate. Max_iter was increased in the steps of [100, 300, 500, 1000, 2000, 5000, 10000].

2.3.3. Random Forest Regression

The Random Forest Regression (RFR) is an ensemble algorithm based on multiple decision trees. The mixture of bootstrapping and random feature selection makes it less sensitive to outliers and noise (Waske et al. 2009). RFRs are out-of-bag predictors, meaning that they can only predict values they had learned in the training process (Breiman 2001). Not being able to extrapolate, this restricts their field of application, but may also serve as an advantage when estimations are supposed to lie within valid ranges. The following parameters were identified to hold most of the control over accuracy in learning and predicting:

N_Estimators: The number of decision trees grown in the random forest. Too few trees lead to a narrow set of values to pick from; too many trees make the algorithm bulky and slow. In this approach, n_estimators was raised in the range of [20, 50, 100, 200, 300, 400, 500].

Min_Samples_Leaf: A decision tree will only be split if there are at least n samples in each of the branches. This parameter is known to have smoothing effects for regression tasks (Breiman 2001). The parameter was increased in the range of [1, 2, 3, 4, 5].

Max_Features: Defines the number of best features to analyze and decide about each new split. Since absolute values depend on the number of features selected for training beforehand – be it bands or components – fractions of one are used to select a relative number of features: [1.0, 0.75, 0.5, 0.25].

2.3.4. Support Vector Regression

Support Vector Machine Regression (SVR) do not rely on statistical assumptions of the training data, which makes them appealing to statistical learning tasks especially in geosciences (Camps-Valls et al. 2009; Mountrakis et al. 2011). Under the right calibration, they were found to produce equally good results as neural networks (Pal and Mather 2005). Their biggest strength lies in their contentment in terms of size of the trainings data set, since only the values closest to the hyperplane fitted in a higher-dimensional space are relevant to define the final model (Waske et al. 2009; Wu et al. 2008). For this study, we used a radial basis function (RBF) kernel and evaluated the best performing hyperparameters by a grid-search with cross-validation:

γ: The kernel coefficient 'gamma' parameter affects the outreach of training samples in influencing the model as support vectors of the RBF. Higher γ allow constructing more complex relationships, but will eventually lead to an overfitted model, so γ is tried to be kept as low as possible. It was varied in decadic logarithmic steps from $1.0 \cdot 10^{-5}$ to $1.0 \cdot 10^{-1}$.

C: The C-parameter regulates the SVR's attempt to a) maximize the distance between support vectors and hyperplane (lower values of C) and b) correctly estimate as many data points as possible (higher values of C). The steps of C were set to the inverse multiplicative of γ, i.e. $1.0 \cdot 10^{0}$ to $1.0 \cdot 10^{5}$.

2.3.5. Gaussian Process Regression

Gaussian Process Regression (GPR) is another example for kernel-based learning. As a probabilistic approach, it trains a model that finds individual functions to all training data by fitting a mean and a covariance function (Gehler et al. 2009). For this, the marginal likelihood is maximized by changing the input function (prior) until an optimum is reached (posterior). Different kernels are frequently used to perform this task and while the balance between penalization and data fitting happens automatically, each kernel still requires some different parameters to hold control of the processes (Rasmussen 2003). All of them share the parameter length-scale, which defines the distance between two points of the training data necessary to significantly influence each other in the output dimension. It was varied in the range of [1, 2, 3, 4, 5, 10]. The alpha-parameter (α_{GPR}) additionally scales the noise level of the target

variables if the GPR is run without a White-Kernel. α_{GPR} was varied in the range of [0.001, 0.01, 0.1, 1.0, 10.0] for all kernels.

Radial Basis Function Kernel: The RBF-Kernel, also known as squared exponential kernel, produces smooth covariance functions. There are no additional parameters to adjust.

Matérn Kernel: The Matérn kernel corresponds to the RBF-kernel in a generalized form. The ν-parameter ("nu") controls the smoothness of the kernel-function and becomes more like the RBF-kernel the higher ν. It was varied in the range of [0.1, 0.2, 0.3, 0.5, 1.5, 2.5].

Rational Quadratic Kernel: Just like the Matérn kernel, the RQ Kernel is a modification of the basic RBF kernel. It is controlled by the scale mixture, which is represented by the alpha-parameter. To avoid confusion with α_{GPR}, we will call the scale mixture α_{RQ} and vary it in the range of [0.01, 0.1, 1.0, 10.0].

When the GPR is run to predict target variables from input data, the result comes together with an estimation of the confidence interval. This option makes the GPR unique among the field of machine learning, rendering it very useful for remote sensing tasks. On the downside, GPRs tend to be very demanding in terms of memory size, since all pairs of X and Y data need to be stored together with the actual mathematical model.

2.4. Artificial Noise

Adding an artificial signal to the spectral output of PROSAIL raises the error in the training process, but arguably increases the accuracy when dealing with real data that is subjected to a certain degree of uncertainty. In this study, we test two different approaches of injecting randomness to the training data: additive (Equation 5) and multiplicative Gaussian noise (Equation 6).

$$R_{addnoise}(\lambda) = R_{Prosail}(\lambda) + \chi(0,\sigma) \quad \text{Equation 5}$$

$$R_{multnoise}(\lambda) = R_{Prosail}(\lambda) \cdot (1 + \chi(0,\sigma)) \quad \text{Equation 6}$$

With $\chi(0,\sigma)$ as the random noise term at wavelength λ, drawn from a Gaussian Normal Distribution with $\mu = 0$ and varying values for standard deviation σ, also referred to as *noise level*.

3. Results

3.1. Training and testing with synthetic data

3.1.1. Effects of data preprocessing

The number of PCA components needed for an accurate training was very similar for all MLRAs (see Figure 2.8.1). A saturation in prediction accuracy for LAI & ALIA is observable for all algorithms. RFR and GPR need fewer components for a stable performance than ANN and SVR. A higher number of components increased the prediction time for SVR (5.2 % per added component) and GPR (2.5 % per added component) but not for ANN and RFR. Basic model size increased with each added component, except for RFR. The smallest increase is observed for ANN and the largest for the SVR.

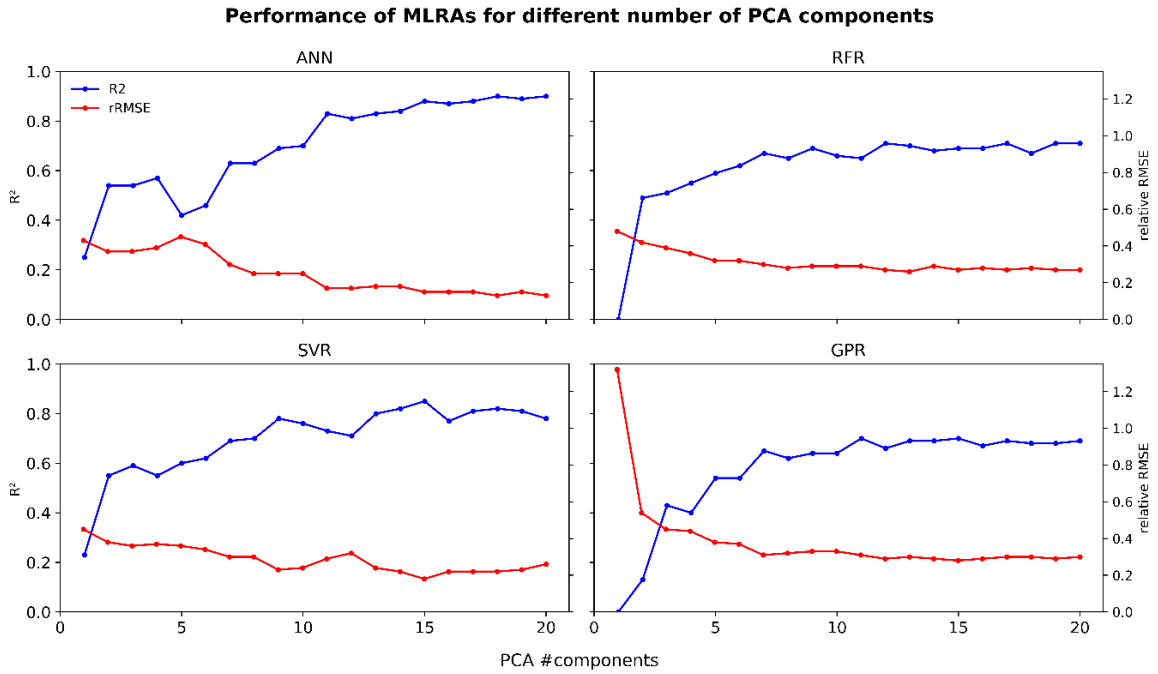


Figure 2.8.1. Performance of four MLRAs in predicting PROSAIL LAI and ALIA from a LUT (n=5,000) in dependence of the number of PCA components. Each MLRA was initialized with its optimized set of hyperparameters.

The choice of the scaler function for the predictors had little impact on the inversion results (Figure 2.8.2). The Standard Scaler mostly performed slightly better than the Robust Scaler which in turn had little advantage to the Min-Max Scaler. It is evident, though, that there is a general need for scaling the input when using ANNs, SVRs or GPRs, whereas the RFR does not rely on scaled data. Scaling the targets, i.e. the biophysical parameters, conversely did not have any effect on the quality of the inversion.

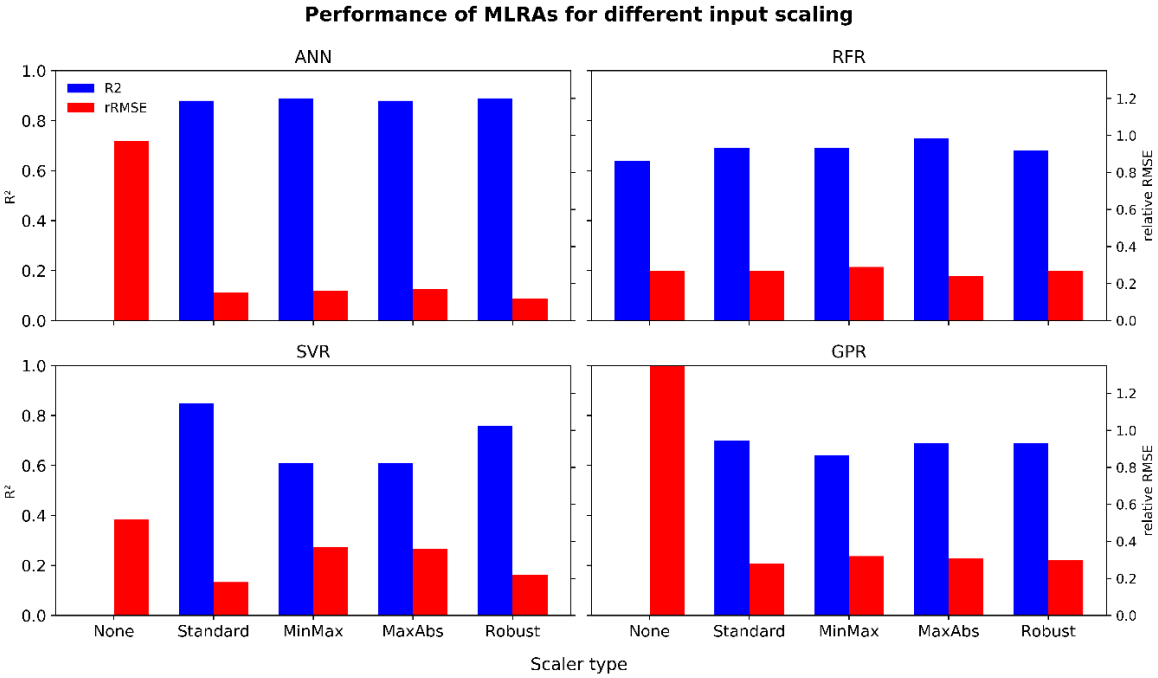


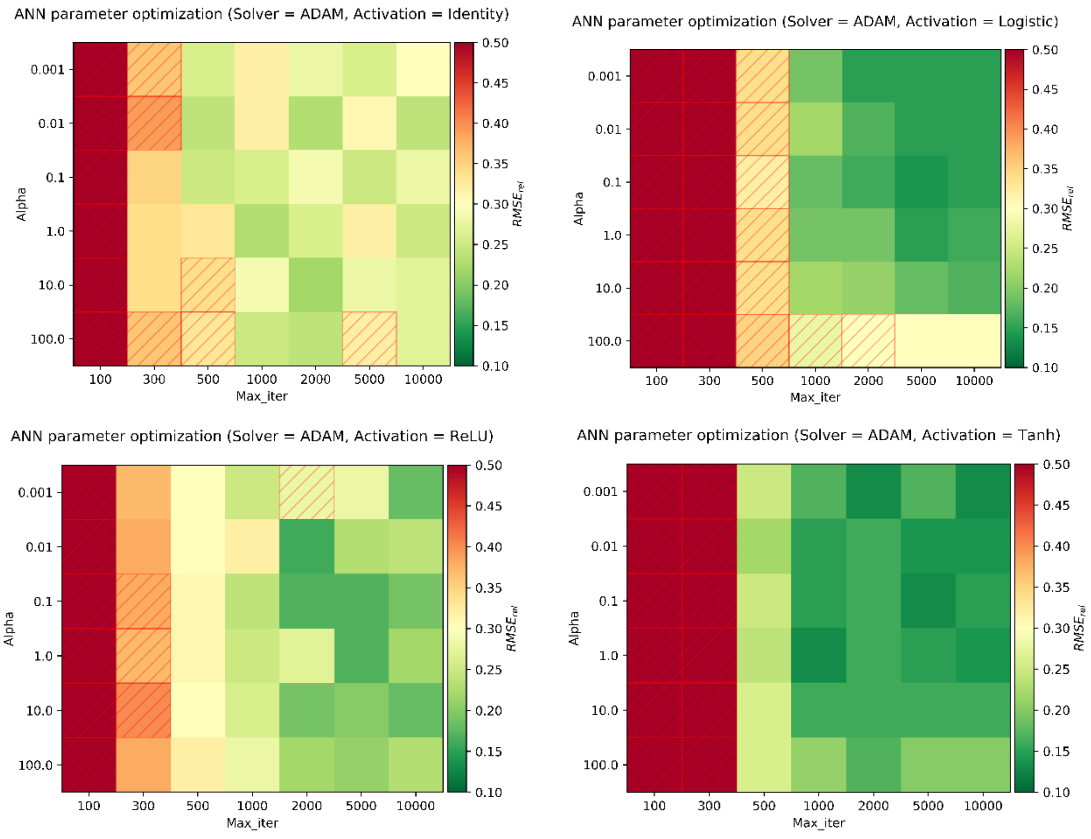
Figure 2.8.2. Performance of four MLRAs in predicting PROSAIL LAI and ALIA from a LUT (n=5,000) for varying scaler types. Each MLRA was initialized with its optimized set of hyperparameters.

3.1.2. Performance of ANN

The choice of solver turned out to be of minor importance for prediction accuracy of the ANNs if both model variants are trained well (Figure 2.8.3). ADAM is strongly dependent on the maximum number of iterations performed by the algorithm, whereas L-BFGS is more sensitive to α_{ANN}. All four activation functions were able to produce outputs with relative errors below 20 % and R^2 above 0.8. The overall highest accuracy ($rRMSE_{training}$: 0.11, $rRMSE_{test}$: 0.13, R^2_{test}: 0.87) was obtained with the following setting:

- Activation: tanh
- Solver: ADAM
- α_{ANN}: 0.001
- max_iter: 10,000

It shall be noted that all other activation functions with both solvers performed nearly equally well, except for the identity activation function. The differences in performance are within ranges of slight random deviations at initialization.



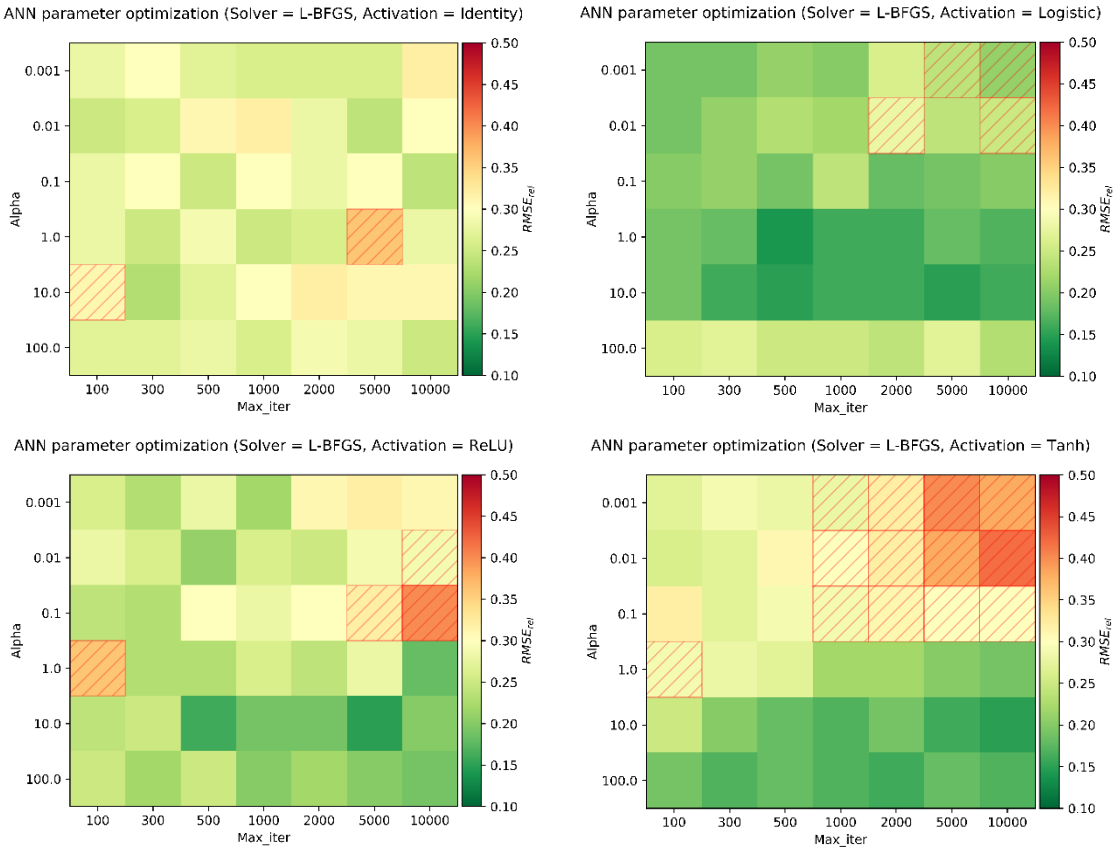


Figure 2.8.3. Optimization results of ANN with different hyperparameters to predict LAI & ALIA together from PROSAIL. Each image shows the relative RMSEs for different alpha and max_iter values. Simulation runs with a $R^2 < 0.6$ are hatched in red.

The average model size was 76 KB for all runs with the ADAM solver and 18 KB for all runs with the L-BFG solver. Relative training time is mostly depending on the maximum number of iterations which scales with $\mathcal{O}$(max_iter) for the L-BFG solver but saturates for ADAM, which was faster by factor 3.5 for max_iter = 5000 in the training process. Absolute training time for LAI & ALIA with a LUT size of 2000 and max_iter of 2000 took 5.2 sec using ADAM as solver method.

Increasing the size of the LUT to 10,000 members further improved the performance of the algorithm up to $rRMSE_{Test} = 0.11$ while the size of the model stayed unaffected. Training time increased only by 0.9 sec and prediction time by 0.0007 sec per 1000 additional LUT members in the synthetic dataset.

3.1.3. Performance of RFR

The variation of Max_features did not have a significant impact on the model performance. Best results for estimation of PROSAIL LAI and ALIA were obtained when at least half of the features were used for training (max_feature = 0.5). A higher number of estimators in the Random Forest turned out to be slightly beneficial for the training success. The minimum number of samples per leaf for each decision tree shows a non-linear influence on model accuracy. The best result ($rRMSE_{training}$: 0.23, $rRMSE_{test}$: 0.25, R^2_{test}: 0.72) was obtained with the following parametrization:

- Max_features: 0.75
- min_samples_leaf: 1
- n_Estimators: 500

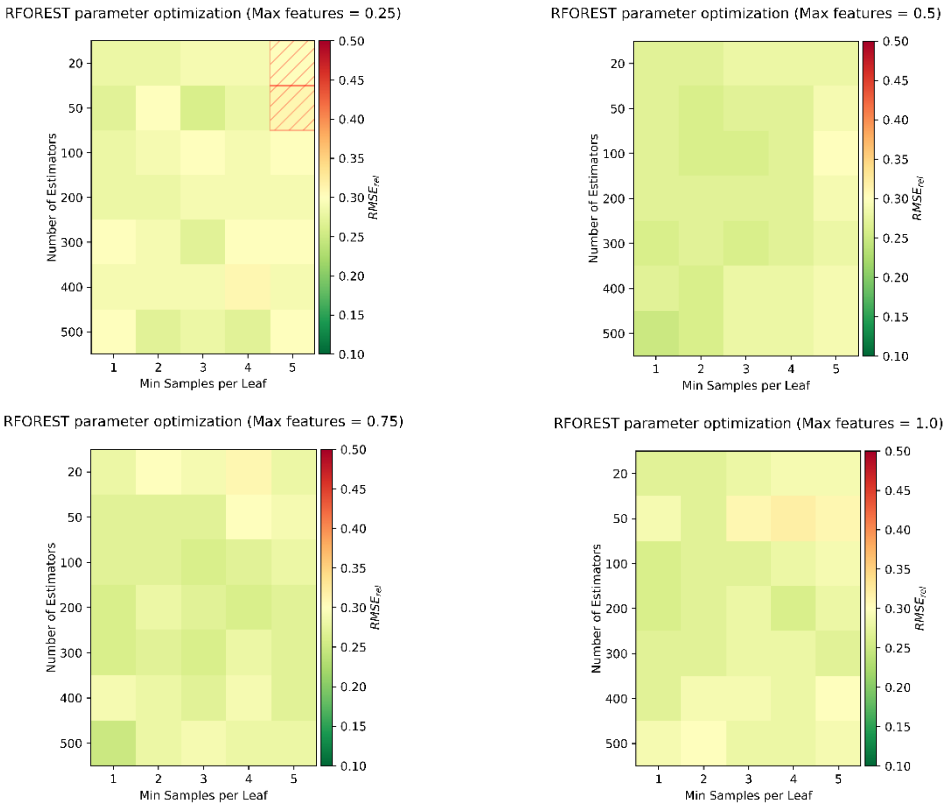


Figure 2.8.4. Optimization results of RFR with different hyperparameters to predict LAI & ALIA together from PROSAIL. Each image shows the relative RMSEs for different minimum of samples per leaf and number of estimators. Simulation runs with a $R^2 < 0.6$ are hatched in red. Max_features was increased from 0.25 to 1.0 in each image.

Max_features does not have any impact on model size, but the minimum number of samples per leaf scales with inverse potency. Increasing the number of estimators has a linear effect on the final size of the model. A multivariate non-linear regression yielded an exemplary empirical relationship (Equation 7) to calculate the resulting model size for Random Forest when trained on the LUT with 2000 members ($R^2 > 0.99$) with scikit-learn and pickle:

$$Model\ Size\ [KB] = 134.13 \cdot x_{\text{min\_leaf\_samples}}{}^{-1.31} \cdot x_{n\_Estimators} \qquad \text{Equation 7}$$

Training time scales linearly with $\mathcal{O}$(min_leaf_samples) and $\mathcal{O}$(n_Estimators). Absolute training time for the optimal hyperparameter setting and 2000 LUT-members was 3.9 sec, prediction time was 0.05 sec.

Larger synthetic datasets led to a marginal improvement in accuracy (rRMSE for LUT-size 10,000 = 0.22) but also to very large models. For the best performing parameter set, the physical memory demand of a single model increased by 34 MB, training time by 25 sec and prediction time by 0.2 sec per 1000 additional LUT-members.

3.1.4. Performance of SVR

The grid search for optimized parametrization of the SVR indicates that lower radial outreach (γ) and higher degree of regularization (C) yield best results to estimate LAI and ALIA from PROSAIL synthetic data ($rRMSE_{training}$: 0.22, $rRMSE_{test}$: 0.22, R^2_{test}: 0.81). Best results were achieved with:

- γ: 0.001
- C: 1000

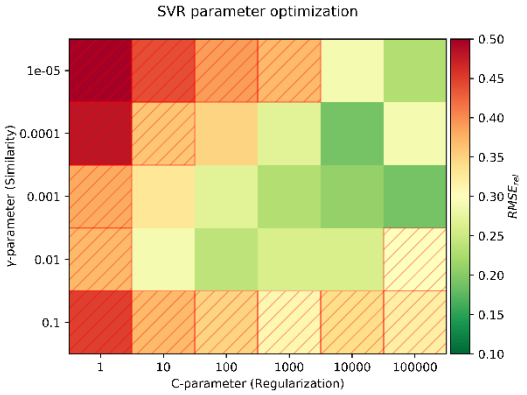


Figure 2.8.5. Optimization results of SVR with different hyperparameters to predict LAI & ALIA together from PROSAIL. The relative RMSEs in the γ-C-Feature Space is shown for an RBF kernel. Simulation runs with $R^2 < 0.6$ are hatched in red.

Model size was independent of γ and C with a mean size of 399 KB for LUT size of 2000. Calculation time scaled linearly with C for very low values of g ($R^2 > 0.98$ for g in [0.0001, 0.001, 0.01]), but this correlation broke down when g was increased further. Consequently, training time for the SVR grows at speed $\mathcal{O}$(C) when g is parameterized such that the model performs well.

Larger sets of training data did not improve the accuracy of the SVR, but resulted in an increased model size (186 KB), training time (5.4 sec) and prediction time (0.1 sec, each per 1000 additional LUT-members).

3.1.5. Performance of GPR

Best results for the training of GPRs was obtained using Matérn or RBF kernels. Training and testing results relied stronger on the exact parameterization than ANN, SVR and RFR did, but offered the potential to generate robust results with high accuracy. When using the Matérn kernel, good results were achieved for all considered variations of ν and length scales. The noise level of the data, i.e. α_{GPR}, showed a notable sensitivity in the training process. Low and moderate ranges for α_{GPR} between 0.01 and 0.1 turned out to yield best results. This effect is even stronger for the RBF-Kernel, for which only $\alpha_{GPR} = 0.1$ led to posteriors with an estimation accuracy of $R^2 > 0.7$.

The best performing parametrization of the GPR was:

- Kernel: Matérn kernel (ν = 1.5)
- α_{GPR}: 0.5
- length_scale: 3

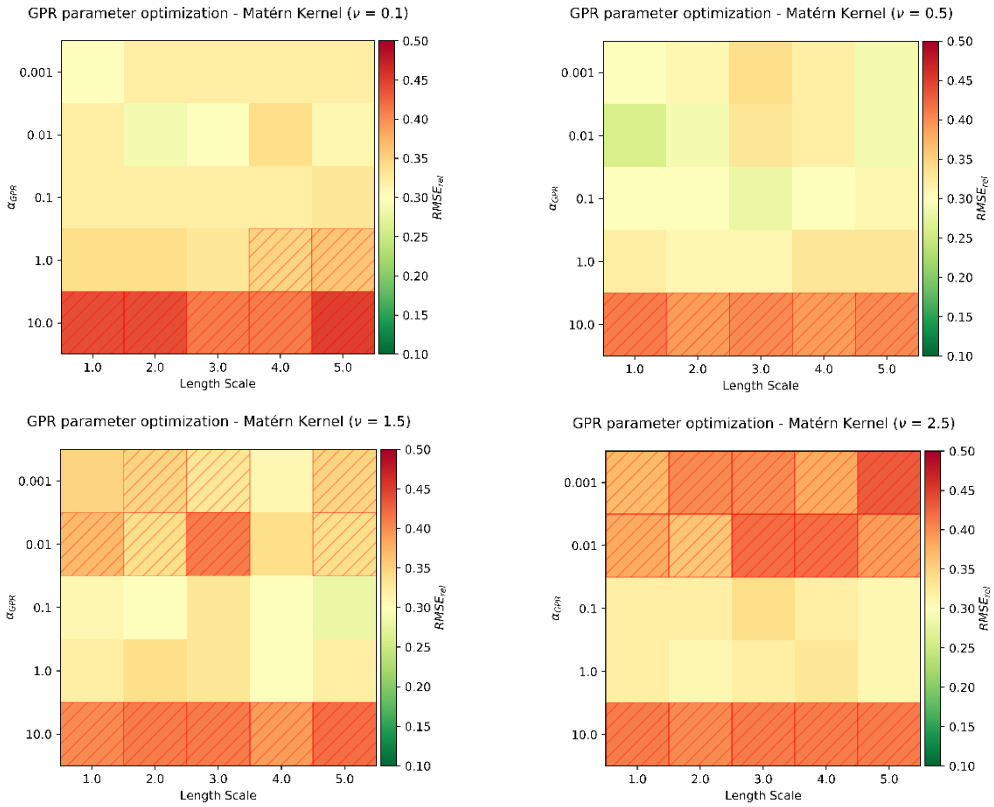


Figure 2.8.6. Results of the optimization of hyperparameters of the GPR with Matérn Kernel to learn LAI and ALIA from PROSAIL synthetic data. Each image shows the relative RMSEs for different α_{GPR} and length scale. Smoothness (ν) was increased from 0.1 to 2.5 in each image. Simulation runs with a $R^2 < 0.6$ are hatched in red.

Optimal parametrization of all GPR-kernels and their accuracies are listed in Table 2.8.1:

Table 2.8.1. Hyperparameters of the different GPR kernels and their according errors in conjointly predicting the PROSAIL parameters LAI and ALIA.

Kernel	Kernel-parameters	GPR-parameters	rRMSE	R^2
Matérn	ν = 1.5	α_{GPR} = 0.5; length_scale = 3	0.22	0.75
RBF	---	α_{GPR} = 0.1; length_scale = 2	0.22	0.75
RQ	α_{rq} = 0.5	α_{GPR} = 0.5; length_scale = 4	0.23	0.70

A bigger issue when dealing with GPR was the size for storing the algorithms on the computer drive. Its model size only depends on the size of the training data set and not the parameterization. For a LUT-size of 2000, model size was 19,200 KB, which was larger than the actual LUT.

Training speed is kernel dependent. For Matérn kernels, calculation time is strongly reduced (factor 10+) when using intermediate values for ν, i.e. 0.5, 1.5 and 2.5. α_{GPR} and the length scale do not influence training speed. Average calculation time for intermediate ν-values was 2.97 sec for LUT size of 2000. GPRs using the RBF-kernel were slightly quicker in training for lower values of α_{GPR} and length scales of 1 or 10. Average training time was 3.3 sec. The RQ-Kernel training speed was sensitive to the scale mixture parameter with longer training times for higher values of α_{RQ} and 7.00 sec on average for the tested setting.

The GPRs were most sensitive to LUT-size. Training the best performing parameterization of that algorithm with 10,000 instead of 2,000 members led to an improvement from $rRMSE_{Test}$ = 0.31 to 0.23. On the downside, this came to the expense of a massive increase of computational demand. Model size

grew with $\mathcal{O}^2$(LUT-size), e.g. 480 MB for 10,000 members and 3 GB for 25,000 members. Training time increased even stronger with $\mathcal{O}^{2.6}$(LUT-size) so that it took more than 30 minutes for one single model to be trained and 4.2 sec for one spectrum, i.e. one pixel in the sensor image, to be inverted.

3.2. Validation and re-calibration on real Spectral Data

The best performing version of each algorithm was validated using field spectral and in situ data of the MNI test site. The results, however, were poor. None of the algorithms were able to reproduce in situ LAI observations with acceptable deviations, revealing the discrepancy between intrinsic model optimization and the real-world situation (see Figure 2.8.7).

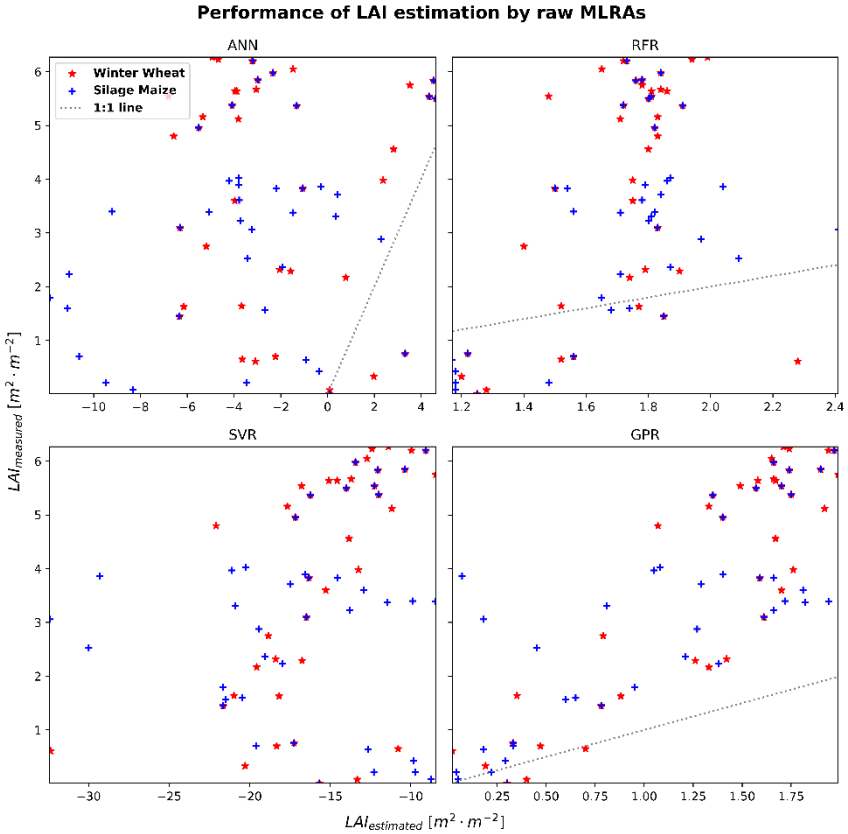


Figure 2.8.7. In situ measured LAI versus LAI estimations of the four MLRAs with optimized hyperparameters.

The GPR showed promising correlations, but the overall estimation error was weak for all algorithms shown. Accordingly, they were trained again, but this time a filter of artificial noise was applied to the PROSAIL spectra to allow training of more robust models. As shown in Table 2.8.2, additive noise is preferred over multiplicative noise in terms of prediction accuracy of real-life data.

Table 2.8.2. Impact of artificial noise on biophysical variable estimations by MLRAs. The examples represent the R^2 between output of ANNs fed with PROSAIL synthetic spectra of different levels of noise and in situ measured LAI and ALIA

	LAI		ALIA	
Noise σ [%]	Additive	Multiplicative	Additive	Multiplicative
0	0.04	0.04	0.00	0.00
1	0.45	0.02	0.02	0.00
2	0.71	0.00	0.00	0.00
3	0.66	0.16	0.11	0.01
4	**0.81**	0.21	0.21	0.00
5	0.77	0.23	0.18	0.01
10	0.65	0.43	0.20	0.00
15	0.41	0.63	0.00	0.07

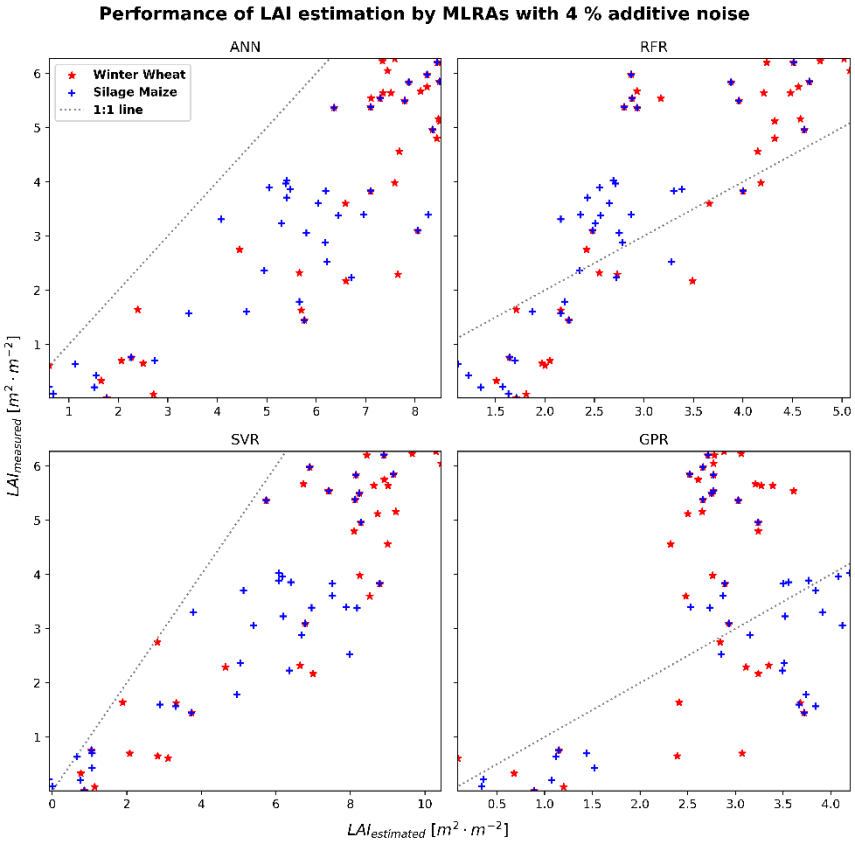


Figure 2.8.8. In situ measured LAI versus LAI estimations of the four MLRAs with optimized hyperparameters trained on PROSAIL synthetic spectra with 4 % additive noise.

Additive noise with σ = 4 % drastically improved the correlation for all algorithms except for GPR. Relative errors of LAI retrieval, however, were still too high for operational practice, e.g. ANN: $rRMSE_{LAI}$ = 0.71. For this reason, linear models were fit to a training set of in situ measured variables and validated on the remaining data, which in turn served as the test set. The final success was evaluated by means of rRMSE between the transformed output of the algorithms and in situ representations.

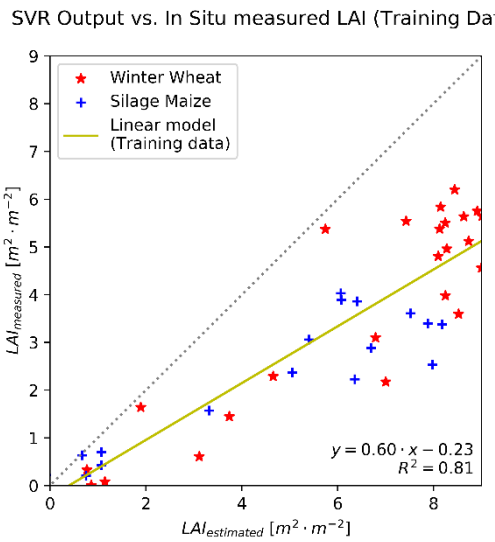
The results of that linear regression are summarized in Table 2.8.3. At the end of the process, LAI observations could be estimated by calibrated ANNs with an accuracy of rRMSE = 0.24. The linear regression model for ALIA had less predictive power (R^2 = 0.30) but approximated leaf inclinations with a relative error of 0.18. The rRMSE for C_{ab} was 0.19 and for LMA it was 0.22. The calibrated SVR outputs

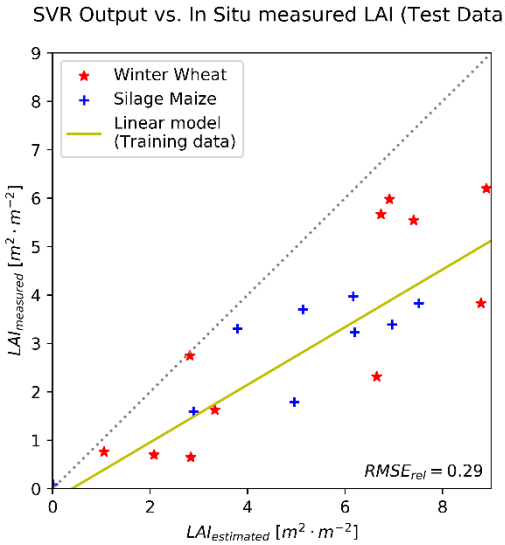
show similar ranges of errors for all parameters, but the linear correction model had a slightly higher correlation for LAI and ALIA than the ANN could achieve. An illustration of the linear regression is shown in Figure 2.8.9, taking the example of SVR. With Random Forest Regression the obtained results were poorer except for the estimation of LMA in which a weak linear signal of $R^2 = 0.17$ was found in the training data. The GPR with Matérn kernel was outperformed by all three algorithms, as it did not improve neither with the added noise nor after linear transformation.

Table 2.8.3. Summary of the final accuracy measures of the four MLRAs after linear correction. The R^2 is calculated between algorithm output and in situ measured variables of the training set. The relative error is calculated after correction with a linear model for the remaining test set of the field data.

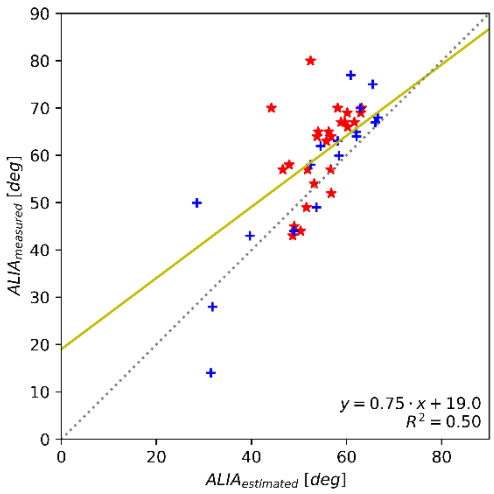
Algorithm	Parameter	R^2 (Trainig set)	Regression slope (Training set)	Regression intercept (Training set)	rRMSE (Test set)
ANN	LAI	0.75	0.69	-0.60	0.24
	ALIA	0.30	0.78	20.16	0.18
	C_{ab}	0.51	0.78	36.52	0.19
	LMA	0.09	0.32	0.00	0.22
RFR	LAI	0.71	1.50	-1.04	0.32
	ALIA	0.16	0.55	25.96	0.19
	C_{ab}	0.37	0.59	27.75	0.20
	LMA	0.17	0.88	0.00	0.27
SVR	LAI	0.81	0.60	-0.23	0.29
	ALIA	0.50	0.75	19.02	0.16
	C_{ab}	0.48	0.75	32.88	0.16
	LMA	0.01	0.06	0.00	0.21
GPR	LAI	0.23	1.01	0.68	0.58
	ALIA	0.07	0.21	52.97	0.26
	C_{ab}	0.24	0.86	33.63	0.23
	LMA	0.08	-0.22	0.01	0.18

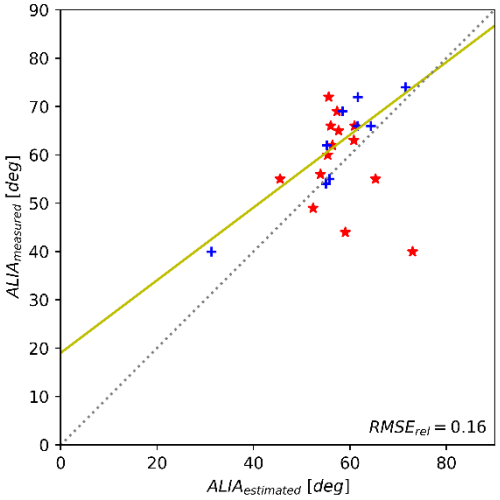
a)


SVR Output vs. In Situ measured LAI (Training Data)
Winter Wheat
Silage Maize
Linear model (Training data)
$y = 0.60 \cdot x - 0.23$
$R^2 = 0.81$
$LAI_{measured}$ $[m^2 \cdot m^{-2}]$
$LAI_{estimated}$ $[m^2 \cdot m^{-2}]$

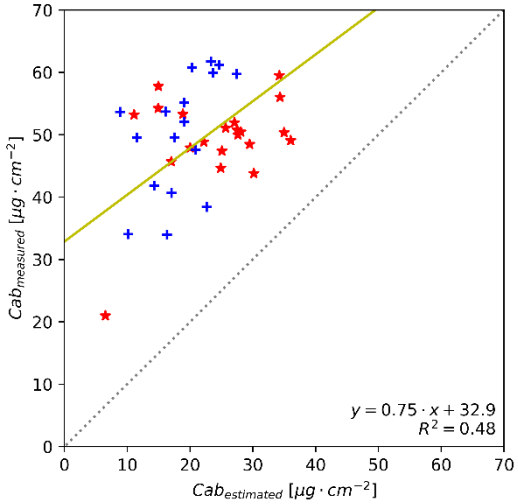

SVR Output vs. In Situ measured LAI (Test Data)
Winter Wheat
Silage Maize
Linear model (Training data)
$RMSE_{rel} = 0.29$
$LAI_{measured}$ $[m^2 \cdot m^{-2}]$
$LAI_{estimated}$ $[m^2 \cdot m^{-2}]$

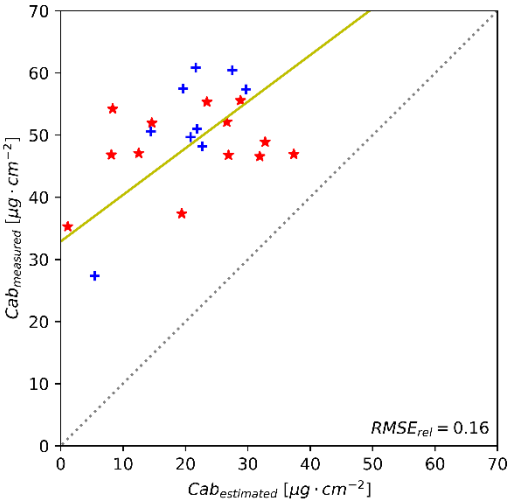
b)


SVR Output vs. In Situ measured ALIA (Training Data)
$y = 0.75 \cdot x + 19.0$
$R^2 = 0.50$
$ALIA_{measured}$ [deg]
$ALIA_{estimated}$ [deg]


SVR Output vs. In Situ measured ALIA (Test Data)
$RMSE_{rel} = 0.16$
$ALIA_{measured}$ [deg]
$ALIA_{estimated}$ [deg]

c)


SVR Output vs. In Situ measured Cab (Training Data)
$y = 0.75 \cdot x + 32.9$
$R^2 = 0.48$
$Cab_{measured}$ $[\mu g \cdot cm^{-2}]$
$Cab_{estimated}$ $[\mu g \cdot cm^{-2}]$


SVR Output vs. In Situ measured Cab (Test Data)
$RMSE_{rel} = 0.16$
$Cab_{measured}$ $[\mu g \cdot cm^{-2}]$
$Cab_{estimated}$ $[\mu g \cdot cm^{-2}]$

d)

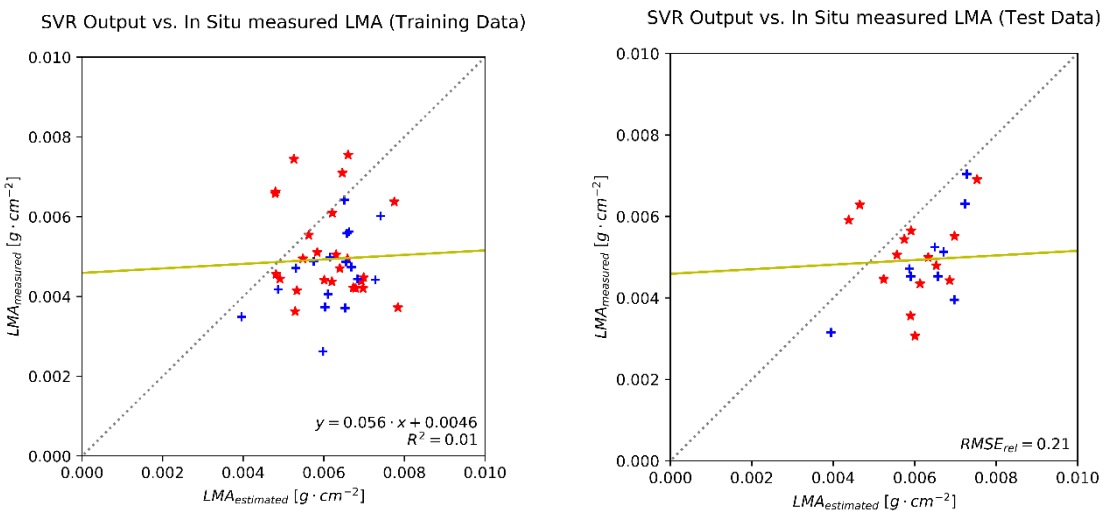


Figure 2.8.9. Illustration of the calibration of the SVR output on in situ measured variables and the validation of the resulting linear model on the test set.

A separate calculation of the linear models for each crop type did not change the further outcome of the inversion. This was tested by a tenfold shuffling of the training subset. Four times maize data was fit with higher accuracies than wheat, six times it was the other way around. In four of the ten variations, best results were achieved when the model was trained on both crops together.

4. Discussion

4.1. Performance of the MLRAs on synthetic data

All four machine learning regression algorithms tested – Artificial Neural Network, Random Forest Regression, Support Vector Regression and Gaussian Process Regression – were successful in learning the inverse mode of the PROSAIL model. They were trained to predict biophysical and biochemical parameters from corresponding reflectance signatures of a synthetic PROSAIL Look-Up-Table. High to very high accuracies could be achieved since training and validation data represented the same statistical population, i.e. they were generated from the same model. An indispensable preparation of the predictors includes dimensionality reduction and scaling.

ANNs did particularly well in learning PROSAIL output and predicting the associated set of parameters. They benefit from large LUTs and were also the fastest in training and prediction. Estimation accuracy for the test set was only marginally lower than for the training set, indicating that overfitting could be avoided during training. The ADAM solver was supposed to be computationally more efficient (Kingma and Ba 2014), which is confirmed by our analysis. The L-BFGS solver converged after fewer iterations, but ADAM performed the minimization task with overall higher accuracies and was faster by factor 2.5 on average. The tanh activation function is often used for binary classification problems, but has also gained popularity in a variety of regression applications (Karlik and Olgac 2011) and worked better than the other activation functions considered in this study.

RFR is supposed to deliver high accuracy while being unaffected from outliers and noise (Breiman 2001). Our findings support these assumptions and show an astounding insensitivity towards parameterization of the hyperparameters in the considered ranges. A total of 140 variations of the minimum samples per leaf, maximum numbers of features and number of estimators was tested, and all resulted in rRMSE between 0.25 and 0.32. Neither a poor nor a great result could be achieved with RFR when a sufficient number of estimators was considered.

As expected, the combination of C and γ had a crucial impact on retrieval accuracy of the SVR. High regularization combined with a low outreach of the training samples led to second highest

accuracies – only outperformed by ANN – but only for the small LUT of size 2000. The algorithm is recommended when the size of the training set is limited, as larger data sets lead to very long training times but not to better predictions. This is perfect in line with (Waske et al. 2009) who found that in many studies SVR performed better or equally well in comparison to other algorithms, but also highlighted their advantage especially for small training sets. It is even argued that SVR could potentially achieve results of the same quality as an ANN if the right settings were used (Pal and Mather 2005). This is confirmed for the training of PROSAIL data with an RBF Kernel and properly tuned hyperparameters.

Training of a GPR algorithm mostly means finding the right kernel that suits the given task. In the study case, most kernels could be trained to achieve accuracies of rRMSE = 0.30, but it was difficult to reach results beyond that line. Interestingly, noise played an important role in the training of all GPRs. For the RBF Kernel, the corridor of valid predictions of the algorithm lies between $\alpha_{GPR} = 0.1$ and 1.0, but undercuts the R^2 threshold of 0.6 for all other instances. The Matérn kernel performed better for higher values of ν, i.e. less sharp ridges in the covariance function or smoother realizations of the Gaussian process. In doing so, lower noise levels could be applied and the accuracy was raised to $rRMSE_{Test} = 0.26$. It remains problematic that there is an infinite amount of possible combinations of kernels and that also their combination would have to be optimized to obtain ideal results. Using only one kernel means that there is no guarantee an optimal parameterization is found. Confidence intervals can be calculated for each GPR prediction to indicate a certainty of the estimation of the target variables. However, for a proper evaluation of the final model accuracy, the comparison with validation data is still inevitable. Another issue is that GPRs tend to become extremely large and slow with growing size of the training data. Even if training success was higher than that of the ANN, the model could hardly be distributed to other users and the inversion of image scenes would take disproportionately long. Extrapolating the necessary 0.58 seconds per single inversion on the tested system, it would take up to 160 hours for yet a small image of 1,000 by 1,000 pixels.

4.2. Performance of the MLRAs on real data

The predictive power of the analyzed MLRAs was raised close to optimum by exhaustive optimization of the hyperparameters. During this process, the algorithm is improved in its capability to link PROSAIL spectral output (X) to PROSAIL vegetation input parameters (Y). Even if the radiative transfer model is complex, there is a mathematical relationship between X and Y that can be learned and reproduced. It turned out that the resulting model became over-determined in the optimization process so that the algorithms were highly specialized on the model setting. When ANN are run on real vegetation signatures, they extrapolate if spectral information is out of their training range. This leads to the shifted and distorted model outputs shown in Figure 2.8.7. Even though the same happened for SVR and GPR, ANNs in particular are well known to react unpredictably upon that issue (Waske et al. 2009). Adding noise to the synthetic data before training is a popular choice to make the algorithms more robust towards unseen data from a different distribution or source (Rivera-Caicedo et al. 2017). An additive noise term from a Gaussian Normal Distribution with standard deviation of 4 % reflectance best accounted for the uncertainties of PROSAIL simulations as well as the errors of in situ measurements of reflectances and crop variables. The same noise level had already been proposed when comparing PROSAIL output with real sensor data (Bacour et al. 2006). Comparing the predictions of LAI, ALIA, C_{ab} and LMA with the variables as measured in the field, striking deviations are still observed. A re-calibration of the MLRA output, as suggested by (Fang and Liang 2003) was the logical last step to make the algorithms perform well also on real-life data. Simple linear models were chosen over higher degree polynomial or exponential curve fitting to keep control over the ranges of prediction in case of extrapolating beyond the observations of the MNI field campaign. Adding noise to spectral data did not improve predictions for the GPR, which already uses α_{GPR} to account for noise in the data. Winter wheat and maize did not demand individual regressions to reach best accuracies, assuring that in this context the inversion of PROSAIL works independently of the crop type.

The final performance of the retrieval of LAI is lower compared to other studies. Some approaches may produce lower relative errors below, but they are hard to compare to our findings for different

reasons. Sticking to synthetic data only (Fang and Liang 2003; Koetz et al. 2005; Weiss et al. 2000) relates to the first step of this study in which the performance of within-model-predictability was tested. Algorithms trained on real data also produce better outputs (Verrelst et al. 2016; Yang et al. 2011), but are sensitive to slight variations in illumination, background signal or crop type. In most cases, parameter retrievals from real data are realized directly with LUT inversions (Danner et al. 2017; Locherer et al. 2015; Verger et al. 2014; Verrelst et al. 2014) with usually higher performance, because additional uncertainties of the MLRAs are avoided. The decreased performance of machine learning when transferring the algorithm from synthetic to field data is confirmed for LAI e.g. in (Doktor et al. 2014).

The retrieval of ALIA was accomplished with relative errors between 0.16 and 0.26. However, this means that predicted leaf inclinations are on average 21 % away from the data mean, which is mainly attributed to the dense scatter cloud and the clusters of in situ measured ALIA around 50 – 60°. The linear models explain between 30 and 50 % of the variations of the in situ measurements. Correcting the inversion results of each MLRA accordingly, posed a large improvement in predictive power especially for ANN. Chlorophyll contents were underestimated by all algorithms, but a linear model helped to correct this offset and reduced relative errors down to 0.16 (SVR) and 0.19 (ANN). The inversion of LMA, however, at the moment is still subjected to quite large deviations. In the case of SVR and ANN, $LMA_{Estimated}$ and $LMA_{Measured}$ were weakly correlated, but scattering around their data means, so high RMSEs prevailed even after optimization of the training and post-processing of the algorithms.

4.3. Limitations of the study

Significance of the performance of each MLRA is restricted to the use of python's scikit-learn module. It should be kept in mind that software solutions of other packages or programming languages running on other machines may be faster or produce different output than in this study. Furthermore, there is a growing field of research dealing with speed optimization of MLRAs, which is exhaustively reviewed in (Camps-Valls et al. 2016).

For a comprehensive interpretation of the results, there are limitations that must be considered. In situ variables from the MNI campaigns were measured in the field, but each of them introduces its own error sources. LiCor LAI-Meters invert leaf area from the measured gap fraction. This method is well established in the theoretical and practical research of canopy architecture, but is known to saturate for values above 5 m^2 m^{-2} (Gower et al. 1999). In fact, the estimations for LAI seem to fit better below that value (see Figure 2.8.9a), so the actual accuracy may as well be better than indicated by the scores. Average Leaf Inclinations, besides being potentially subjected to a bias at reading of the analogue display, are only one factor in the description of crop geometry. For instance, a canopy constructed from fully erected and flat leaves only would have the same ALIA as a canopy with perfect medium inclinations (both ALIA = 45°) but their spectral signal would be quite different. The deficits in the inversion of canopy geometry have been addressed by using a training and test set of ALIA that was adjusted by manual fitting (Danner et al. 2019).

Given that in situ variables were measured with sufficient accuracy, there remains the challenging transfer from model findings to applications of the real world. In the global inversion approach, PROSAIL parameters not analyzed in this study were inverted together with LAI, ALIA, C_{ab} and LMA at the same time. A mismatch for e.g. the leaf structure parameter (N) could not be quantified due to a lack of field methods to measure this parameter. It would subsequently affect the inversion of parameters with similar spectral ranges of sensitivity, with a suspected influence on e.g. the accuracy of LMA estimations.

The validation of the machine learning approach was done for data from field instruments. An operational use of the Biophysical Vegetation Processor will introduce scaling effects from leaf to canopy and from proximal to spaceborne spectral observations (Malenovský et al. 2019). Real sensor data of hyperspectral satellites will be crucial for a re-calibration of the linear models. Such data should also include a larger variety of crops – especially soybean, rice and sorghum – and other types of vegetation like grassland and forests to extend validity to the major part of vegetated surfaces around the globe.

5. Conclusion

We demonstrated that Artificial Neural Networks (ANN), Random Forest Regression (RFR), Support Vector Machine Regression (SVR) and Gaussian Process Regression (GPR) were all able to learn the inverse case of PROSAIL in a way that they predict biophysical and biochemical parameters from spectral data with accuracies of up to rRMSE = 0.13. Transforming the input data by applying a PCA into at least 15 components and a subsequent z-transformation were a prerequisite for good retrieval results just like an optimal tuning of hyperparameters of the algorithms. ANNs were best suited for this task, but the model could not simply be transferred to real-life observations. Instead, artificial noise and re-calibration with a training set of in situ measured variables was needed to account for uncertainties occurring outside of the ideal model world. Estimation errors for this approach are higher than for most other techniques using synthetic data. These deviations need to be accepted for building a vegetation processor that retrieves biophysical and biochemical variables operationally from hyperspectral imaging data on its own. Such an application needs to be fast in processing, to take up little memory and to be easily distributed to a broad user community. It is concluded that ANNs excel in all these scopes. They were outperformed by SVRs in terms of final accuracy of the parameter retrieval, but are several orders of magnitude smaller and faster and are thus considered the better choice for the future EnMAP Managed Vegetation Scientific Processor. Further development is encouraged to improve the performance of the algorithms to estimate also the less sensitive parameters with higher accuracy and less prediction time.

Acknowledgement: The research presented in this article was conducted at the Chair of Geography and Remote Sensing, Department of Geography, Ludwig-Maximilians-Universität Munich. The financial support through the Space Administration of the German Aerospace Center (DLR) in the frame of the project "EnMAP Scientific Advisory Group Phase III—Developing the EnMAP Managed Vegetation Scientific Processor" through funding by the German Ministry of Economics and Technology under the grant code 50EE1623 is gratefully acknowledged. We want to thank the communal farms of Munich for granting access to the sampling plots and for providing auxiliary data. Special thanks also to the Geomatics team of Humboldt-Universität zu Berlin for their devoted support throughout the development process of the EnMAP Managed Vegetation Scientific Processor.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

Asner, G.P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. Remote Sensing of Environment, 64, 234-253

Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. Remote Sensing of Environment, 105, 313-325

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32

Campbell, G.S. (1986). Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. Agricultural and Forest Meteorology, 36, 317-321

Camps-Valls, G., Muñoz-Marí, J., Gómez-Chova, L., Richter, K., & Calpe-Maravilla, J. (2009). Biophysical parameter estimation with a semisupervised support vector machine. Ieee Geoscience and Remote Sensing Letters, 6, 248-252

Camps-Valls, G., Verrelst, J., Munoz-Mari, J., Laparra, V., Mateo-Jimenez, F., & Gomez-Dans, J. (2016). A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation. IEEE Geoscience and Remote Sensing Magazine, 4, 58-78

Danner, M., Berger, K., Wocher, M., Mauser, W., & Hank, T. (2017). Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy. Remote Sensing, 9, 21

Danner, M., Berger, K., Wocher, M., Mauser, W., & Hank, T. (2019). Fitted PROSAIL Parameterization of Leaf Inclinations, Water Content and Brown Pigment Content for Winter Wheat and Maize Canopies. Remote Sensing, 11, 1150

Darvishzadeh, R., Skidmore, A., Atzberger, C., & van Wieren, S. (2008). Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture. International Journal of Applied Earth Observation and Geoinformation, 10, 358-373

Doktor, D., Lausch, A., Spengler, D., & Thurner, M. (2014). Extraction of Plant Physiological Status from Hyperspectral Signatures Using Machine Learning Methods. Remote Sensing, 6, 12247-12274

Fang, H., & Liang, S. (2003). Retrieving leaf area index with a neural network method: Simulation and validation. IEEE Transactions on Geoscience and Remote Sensing, 41, 2052-2062

Feilhauer, H., Schmid, T., Faude, U., Sánchez-Carrillo, S., & Cirujano, S. (2018). Are remotely sensed traits suitable for ecological analysis? A case study of long-term drought effects on leaf mass per area of wetland vegetation. Ecological Indicators, 88, 232-240

Féret, J.B., Gitelson, A.A., Noble, S.D., & Jacquemoud, S. (2017). PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sensing of Environment, 193, 204-215

Gastellu-Etchegorry, J.-P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative transfer in heterogeneous 3-D vegetation canopies. Remote Sensing of Environment, 58, 131-156

Gehler, P.V., Schölkopf, B., Camps-Valls, G., & Bruzzone, L. (2009). An introduction to kernel learning algorithms. Kernel methods for remote sensing data analysis, 25-45

Gitelson, A.A. (2018). 1 Recent Developments in Remote Estimation of Crop Biophysical and Biochemical Properties at Various Scales. Biophysical and Biochemical Characterization and Plant Species Studies, 1

Gower, S.T., Kucharik, C.J., & Norman, J.M. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. Remote Sensing of Environment, 70, 29-51

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., & Sang, B. (2015). The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sensing, 7, 8830-8857

Hanes, J. (2013). Biophysical Applications of Satellite Remote Sensing. Heidelberg, Germany: Springer-Verlag Berlin

Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. Remote Sensing, 7, 3934-3965

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., Francois, C., & Ustin, S.L. (2009). PROSPECT plus SAIL models: A review of use for vegetation characterization. Remote Sensing of Environment, 113, S56-S66

Karlik, B., & Olgac, A.V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems, 1, 111-122

Kattenborn, T., Fassnacht, F.E., & Schmidtlein, S. (2018). Differentiating plant functional types using reflectance: which traits make the difference? Remote Sensing in Ecology and Conservation

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. Remote Sensing of Environment, 95, 115-124

Kuester, T., Spengler, D., Barczi, J.-F., Segl, K., Hostert, P., & Kaufmann, H. (2014). Simulation of multitemporal and hyperspectral vegetation canopy bidirectional reflectance using detailed virtual 3-D canopy models. IEEE Transactions on Geoscience and Remote Sensing, 52, 2096-2108

LICOR-Biosciences (2015). LAI-2200C Plant Canopy Analyzer Instruction Manual. In. Lincoln, USA

Locherer, M., Hank, T., Danner, M., & Mauser, W. (2015). Retrieval of Seasonal Leaf Area Index from Simulated EnMAP Data through Optimized LUT-Based Inversion of the PROSAIL Model. Remote Sensing, 7, 10321-10346

Malenovský, Z., Homolová, L., Lukeš, P., Buddenbaum, H., Verrelst, J., Alonso, L., Schaepman, M.E., Lauret, N., & Gastellu-Etchegorry, J.-P. (2019). Variability and Uncertainty Challenges in Scaling Imaging Spectroscopy Retrievals and Validations from Leaves Up to Vegetation Canopies. Surveys in Geophysics, 1-26

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. Isprs Journal of Photogrammetry and Remote Sensing, 66, 247-259

Myneni, R., Running, S., Glassy, J., & Votova, P. (2000). User's guide: fPAR, LAI (ESDT: MOD15A2) 8-day composite. NASA MODIS Land Algorithm

Norman, J.M., & Campbell, G.S. (1989). Canopy structure. Plant physiological ecology (pp. 301-325): Springer

Paganini, M., Leidner, A.K., Geller, G., Turner, W., & Wegmann, M. (2016). The role of space agencies in remotely sensed essential biodiversity variables. Remote Sensing in Ecology and Conservation, 2, 132-140

Pal, M., & Mather, P. (2005). Support vector machines for classification in remote sensing. International Journal of Remote Sensing, 26, 1007-1011

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12, 2825-2830

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

Rivera-Caicedo, J.P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. Isprs Journal of Photogrammetry and Remote Sensing, 132, 88-101

Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., & Hofer, S. (2012). EeteS - The EnMAP End-to-End Simulation Tool. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 5, 522-530

Sohl-Dickstein, J., Poole, B., & Ganguli, S. (2014). Fast large-scale optimization by unifying stochastic gradient and quasi-Newton methods. In, International Conference on Machine Learning (pp. 604-612)

Thenkabail, P. (2017). Hyperspectral remote sensing of vegetation and agricultural crops

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., & Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. Remote Sensing of Environment, 152, 654-664

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. Remote Sensing of Environment, 16, 125-141

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., & Moreno, J. (2018). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. Surveys in Geophysics, 1-41

Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J., & Camps-Valls, G. (2016). Spectral band selection for vegetation properties retrieval using Gaussian processes regression. International Journal of Applied Earth Observation and Geoinformation, 52, 554-567

Verrelst, J., Rivera, J.P., Leonenko, G., Alonso, L., & Moreno, J. (2014). Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. Geoscience and Remote Sensing, IEEE Transactions on, 52, 257-269

Wang, W.-M., Li, Z.-L., & Su, H.-B. (2007). Comparison of leaf angle distribution functions: effects on extinction coefficient and fraction of sunlit foliage. Agricultural and Forest Meteorology, 143, 106-122

Waske, B., Fauvel, M., Benediktsson, J.A., & Chanussot, J. (2009). Machine learning techniques in remote sensing data analysis. Kernel methods for remote sensing data analysis, 3-24

Weiss, M., & Baret, F. (2016). S2ToolBox level 2 products: LAI, FAPAR, FCOVER. Institut National de la Recherche Agronomique (INRA), Avignon

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. Agronomie, 20, 3-22

White, H. (1992). Artificial neural networks: approximation and learning theory. Cambridge, USA: Blackwell Publishers, Inc.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., & Philip, S.Y. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14, 1-37

Yang, X., Huang, J., Wu, Y., Wang, J., Wang, P., Wang, X., & Huete, A.R. (2011). Estimating biophysical parameters of rice with remote sensing data using support vector machines. Science China Life Sciences, 54, 272-281

Zabel, F., Putzenlechner, B., & Mauser, W. (2014). Global agricultural land resources–a high resolution suitability evaluation and its perspectives until 2100 under climate change conditions. PLoS ONE, 9, e107522

Ziliani, M., Ershadi, A., Zheng, B., Ait-El-Fquih, B., Altaf, M., Aragon, B., Hoteit, I., Franz, T., Chapman, S., & McCabe, M. (2018). Corn-Yield Estimation through Assimilation of Remotely Sensed LAI Data into APSIM. In, AGU Fall Meeting Abstracts

3. Synthesis & Outlook

The initial scientific objectives of this thesis were to test and improve approaches to retrieve biochemical and biophysical variables from hyperspectral data like that of the future EnMAP-HSI. After the experimental work with different LUT inversions it could be concluded that parameter estimations with high accuracy are possible not only from nadir but also from images recorded in across-track pointing mode. A sought hierarchical approach failed mainly because of the separation of the structural canopy variables LAI and Average Leaf Inclination Angle. A global retrieval of LAI, ALIA and C_{cab} was instead achieved with the use of an Artificial Neural Networks and a Support Vector Machine Regression when the synthetic training data was treated with artificial gaussian noise. Accuracies with this method are lower compared to retrieval schemes devoted to all-synthetic or all-real-world data. These studies may serve as road signs for the general capability of novel methods, but they are hardly realizable in actual practical scopes.

The integration of a priori information could improve model outputs as well as multi-temporal or multi-angular observations. These advancements will gain more relevance when a stream of spaceborne hyperspectral images becomes available with repetition rates of few days. EnMAP, however, is planned as a scientific mission with an expected strong competition between institutions and study areas. The EnMAP managed vegetation scientific processor is considered to be a practical tool for quick estimation of relevant crop properties, but the development of that application will need to persist beyond the completion of this thesis.

Consecutive field campaigns should include additional crops like sugar beet, rice or potatoes to test whether the proposed solutions are indeed independent of crop type. Further MLRAs will have to be trained to represent a variety of illumination conditions and sensor tilts. The retrieval of C_w from hyperspectral data has been successfully tested by (Wocher et al. 2018) by using the physical-based approach of the 970 nm domain of liquid water absorption. This method will be added to the existing scheme and run in advance of the MLRA to further check whether a pixel contains the signal of a vital plant or not. With these refinements ahead, an application is evolving that has the potential to be used by a broad user community and ultimately contribute another piece to the complex puzzle of modern farming practice.

References

Abburu, S., & Golla, S.B. (2015). Satellite image classification methods and techniques: A review. *International journal of computer applications, 119*

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., & Sousa, J. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing, 9*, 1110

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing, 7*, 16398-16421

Allen, W.A., Gausman, H.W., Richardson, A.J., & Thomas, J.R. (1969). Interaction of Isotropic Light with a Compact Plant Leaf*. *Journal of the Optical Society of America, 59*, 1376-1379

Asner, G.P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment, 64*, 234-253

Atzberger, C., Jarmer, T., Schlerf, M., Kötz, B., & Werner, W. (2003). Retrieval of wheat bio-physical attributes from hyperspectral data and SAILH+ PROSPECT radiative transfer model. In, *Proceedings of the 3rd EARSeL Workshop on imaging spectroscopy* (pp. 473-482): Citeseer

Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment, 105*, 313-325

Baret, F., & Fourty, T. (1997). Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie, 17*, 455-464

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., & Hank, T. (2018). Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sensing, 10*, 85

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32

Campbell, G.S. (1986). Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. *Agricultural and Forest Meteorology, 36*, 317-321

Campbell, J.B., & Wynne, R.H. (2011). *Introduction to remote sensing*. Guilford Press

Camps-Valls, G., Muñoz-Marí, J., Gómez-Chova, L., Richter, K., & Calpe-Maravilla, J. (2009). Biophysical parameter estimation with a semisupervised support vector machine. *Ieee Geoscience and Remote Sensing Letters, 6*, 248-252

Camps-Valls, G., Verrelst, J., Munoz-Mari, J., Laparra, V., Mateo-Jimenez, F., & Gomez-Dans, J. (2016). A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine, 4*, 58-78

Candela, L., Formaro, R., Guarini, R., Loizzo, R., Longo, F., & Varacalli, G. (2016). The PRISMA mission. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International* (pp. 253-256): IEEE

Conway, J., Brown, L., Veck, N., Wielogorski, A., & Borgeaud, M. (1991). A model-based system for crop classification from radar imagery. In, *1991 Seventh International Conference on Antennas and Propagation, ICAP 91 (IEE)* (pp. 616-619): IET

Danner, M., Berger, K., Wocher, M., Mauser, W., & Hank, T. (2017). Retrieval of Biophysical Crop Variables from Multi-Angular Canopy Spectroscopy. *Remote Sensing, 9*, 21

Danner, M., Berger, K., Wocher, M., Mauser, W., & Hank, T. (2019). Fitted PROSAIL Parameterization of Leaf Inclinations, Water Content and Brown Pigment Content for Winter Wheat and Maize Canopies. *Remote Sensing, 11*, 1150

Danner, M., Hank, T., Locherer, M., Mauser, W., & Ieee (2015). ANALYZING UNCERTAINTIES IN SIMULATED CANOPY REFLECTANCE THROUGH EXHAUSTIVE COMPARISON WITH IN-SITU MEASURED OPTICAL PROPERTIES. *2015 Ieee International Geoscience and Remote Sensing Symposium* (pp. 4009-4012)

Danner, M., Wocher, M., Berger, K., Mauser, W., & Hank, T. (2018). Developing a Sandbox Environment for Prosail, Suitable for Education and Research. In, *IEEE International Geoscience and Remote Sensing Symposium* (pp. 783-786). Valencia, Spain

Darvishzadeh, R., Skidmore, A., Atzberger, C., & van Wieren, S. (2008). Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture. *International Journal of Applied Earth Observation and Geoinformation, 10*, 358-373

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys, 27*, 326-327

Doktor, D., Lausch, A., Spengler, D., & Thurner, M. (2014). Extraction of Plant Physiological Status from Hyperspectral Signatures Using Machine Learning Methods. *Remote Sensing, 6*, 12247-12274

EOC (2018a). EnMAP Mission. http://www.enmap.org/mission.html *(2019/06/03)*

EOC (2018b). EnMAP Organization. http://www.enmap.org/organization.html *(2019/06/03)*

Fang, H., & Liang, S. (2003). Retrieving leaf area index with a neural network method: Simulation and validation. *IEEE Transactions on Geoscience and Remote Sensing, 41*, 2052-2062

FAO (2016). The State of Food and Agriculture 2016. Climate change, agriculture and food security. In: Food and Agriculture Organization of the United Nations, FAO

Feilhauer, H., Schmid, T., Faude, U., Sánchez-Carrillo, S., & Cirujano, S. (2018). Are remotely sensed traits suitable for ecological analysis? A case study of long-term drought effects on leaf mass per area of wetland vegetation. *Ecological Indicators, 88*, 232-240

Feingersh, T., & Dor, E.B. (2015). SHALOM–A Commercial Hyperspectral Space Mission. *Optical Payloads for Space Missions*, 247

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

Féret, J.B., Gitelson, A.A., Noble, S.D., & Jacquemoud, S. (2017). PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment, 193*, 204-215

Folkman, M.A., Pearlman, J., Liao, L.B., & Jarecke, P.J. (2001). EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration. In, *Hyperspectral Remote Sensing of the Land and Atmosphere* (pp. 40-52): International Society for Optics and Photonics

Gastellu-Etchegorry, J.-P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sensing of Environment, 58*, 131-156

Gehler, P.V., Schölkopf, B., Camps-Valls, G., & Bruzzone, L. (2009). An introduction to kernel learning algorithms. *Kernel methods for remote sensing data analysis*, 25-45

Gitelson, A.A. (2018). 1 Recent Developments in Remote Estimation of Crop Biophysical and Biochemical Properties at Various Scales. *Biophysical and Biochemical Characterization and Plant Species Studies*, 1

Goel, N.S. (1988). Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sensing Reviews, 4*, 1-212

Govaerts, Y.M., & Verstraete, M.M. (1998). Raytran: A Monte Carlo ray-tracing model to compute light scattering in three-dimensional heterogeneous media. *IEEE Transactions on Geoscience and Remote Sensing, 36*, 493-505

Gower, S.T., Kucharik, C.J., & Norman, J.M. (1999). Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sensing of Environment, 70*, 29-51

Greenspan, H., Goodman, R., Chellappa, R., & Anderson, C.H. (1994). Learning texture discrimination rules in a multiresolution system. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 894-901

Guanter, L., Kaufmann, H., Förster, S., Brosinsky, A., Wulf, H., Bochow, M., Boesche, N., Brell, M., Buddenbaum, H., & Chabrillat, S. (2016). Enmap Science Plan

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., & Sang, B. (2015). The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing, 7*, 8830-8857

Hanes, J. (2013). *Biophysical Applications of Satellite Remote Sensing*. Heidelberg, Germany: Springer-Verlag Berlin

Hank, T., Bach, H., & Mauser, W. (2015). Using a Remote Sensing-Supported Hydro-Agroecological Model for Field-Scale Simulation of Heterogeneous Crop Growth and Yield: Application for Wheat in Central Europe. *Remote Sensing, 7*, 3934-3965

Hank, T., Danner, M., Berger, K., & Mauser, W. (2017). Erfolgskontrollbericht EnMAP Scientific Advisory Group Phase II - Monitoring the Phenological Development of Agricultural Crops. In. Munich, Germany

Hank, T.B., Berger, K., Bach, H., Clevers, J.G.P.W., Gitelson, A., Zarco-Tejada, P., & Mauser, W. (2018). Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. *Surveys in Geophysics*, 1-37

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment, 34*, 75-91

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., Francois, C., & Ustin, S.L. (2009a). PROSPECT plus SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment, 113*, S56-S66

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., & Ustin, S.L. (2009b). PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment, 113, Supplement 1*, S56-S66

Jones, H.G., & Vaughan, R.A. (2010). *Remote sensing of vegetation: principles, techniques and applications.* . New York: Oxford University Press

Karlik, B., & Olgac, A.V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems, 1*, 111-122

Kattenborn, T., Fassnacht, F.E., & Schmidtlein, S. (2018). Differentiating plant functional types using reflectance: which traits make the difference? *Remote Sensing in Ecology and Conservation*

Khan, M.J., Khan, H.S., Yousaf, A., Khurshid, K., & Abbas, A. (2018). Modern trends in hyperspectral image analysis: a review. *IEEE Access, 6*, 14118-14129

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment, 95*, 115-124

Kubat, M., Holte, R.C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine Learning, 30*, 195-215

Kubelka, P., & Munk, F. (1931). Ein Beitrag zur Optik der Farbanstriche, Zeitshrift tech. In: Physik

Kuester, T., Spengler, D., Barczi, J.-F., Segl, K., Hostert, P., & Kaufmann, H. (2014). Simulation of multitemporal and hyperspectral vegetation canopy bidirectional reflectance using detailed virtual 3-D canopy models. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 2096-2108

Lary, D.J. (2010). Artificial intelligence in geoscience and remote sensing. *Geoscience and Remote Sensing New Achievements*: IntechOpen

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., & Middleton, E.M. (2015). An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment, 167*, 6-19

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System. *Agronomie, 19*, 185-210

Li, X., & Strahler, A.H. (1985). Geometric-optical modeling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing*, 705-721

LICOR-Biosciences (2015). LAI-2200C Plant Canopy Analyzer Instruction Manual. *https://licor.app.boxenterprise.net/s/fqjn5mlu8c1a7zir5qel (2019/5/11)*

Liu, D., Zeng, X.-A., & Sun, D.-W. (2015). Recent developments and applications of hyperspectral imaging for quality evaluation of agricultural products: a review. *Critical reviews in food science and nutrition, 55*, 1744-1757

Locherer, M., Hank, T., Danner, M., & Mauser, W. (2015). Retrieval of Seasonal Leaf Area Index from Simulated EnMAP Data through Optimized LUT-Based Inversion of the PROSAIL Model. *Remote Sensing, 7*, 10321-10346

Lopatin, J., Fassnacht, F.E., Kattenborn, T., & Schmidtlein, S. (2017). Hyperspectral classification of grassland species: towards a UAS application for semi-automatic field surveys. In, *EGU General Assembly Conference Abstracts* (p. 8703)

Malenovský, Z., Homolová, L., Lukeš, P., Buddenbaum, H., Verrelst, J., Alonso, L., Schaepman, M.E., Lauret, N., & Gastellu-Etchegorry, J.-P. (2019). Variability and Uncertainty Challenges in Scaling Imaging Spectroscopy Retrievals and Validations from Leaves Up to Vegetation Canopies. *Surveys in Geophysics*, 1-26

Martin, E., Gastellu-Etchegorry, J.P., Dhalluin, R., Ieee, Ieee, & Ieee (2003). *Model Intercomparison for validating the 2003 DART model*.

Mauser, W., Bach, H., Hank, T., Zabel, F., & Putzenlechner, B. (2012). How spectroscopy from space will support world agriculture. In, *2012 IEEE international geoscience and remote sensing symposium* (pp. 7321-7324): IEEE

Mauser, W., & Hank, T. (2012). EnMAP Core Science Team Phase II - Monitoring the Phenological Development of Agricultural Crops. In. Munich, Germany

Mitchell, T., Buchanan, B., DeJong, G., Dietterich, T., Rosenbloom, P., & Waibel, A. (1990). Machine learning. *Annual review of computer science, 4*, 417-433

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *Isprs Journal of Photogrammetry and Remote Sensing, 66*, 247-259

Mulla, D.J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering, 114*, 358-371

Murphy, K.P. (2012). *Machine learning: a probabilistic perspective*. MIT press

Myneni, R., Running, S., Glassy, J., & Votova, P. (2000). User's guide: fPAR, LAI (ESDT: MOD15A2) 8-day composite. *NASA MODIS Land Algorithm*

Nieke, J., & Rast, M. (2018). Towards the Copernicus Hyperspectral Imaging Mission for the Environment (CHIME). In, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 157-159): IEEE

Norman, J.M., & Campbell, G.S. (1989). Canopy structure. *Plant physiological ecology* (pp. 301-325): Springer

Ortenberg, F., Thenkabail, P., Lyon, J., & Huete, A. (2011). Hyperspectral sensor characteristics: airborne, spaceborne, hand-held, and truck-mounted; Integration of hyperspectral data with Lidar. In (pp. 39-68): Boca Raton, FL, USA: CRC Press

Paganini, M., Leidner, A.K., Geller, G., Turner, W., & Wegmann, M. (2016). The role of space agencies in remotely sensed essential biodiversity variables. *Remote Sensing in Ecology and Conservation, 2*, 132-140

Pal, M., & Mather, P. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26*, 1007-1011

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*, 2825-2830

Pinter Jr, P.J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S., & Upchurch, D.R. (2003). Remote sensing for crop management. *Photogrammetric Engineering & Remote Sensing, 69*, 647-664

Pu, R. (2017). *Hyperspectral Remote Sensing: Fundamentals and Practices*. CRC Press

Rabe, A., Jakimow, B., Thiel, F., Hostert, P., & van der Linden, S. (2018). EnMAP-Box 3 a free and open source Python plug-in for QGIS. In, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 7764-7766): IEEE

Ramakrishnan, D., & Bharti, R. (2015). Hyperspectral remote sensing and geological applications. *Current science*, 879-891

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

Rees, W.G. (2013). *Physical principles of remote sensing*. Cambridge University Press

Rivera-Caicedo, J.P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *Isprs Journal of Photogrammetry and Remote Sensing, 132*, 88-101

Sabins, F.F. (2007). *Remote sensing: principles and applications*. Waveland Press

Schaepman-Strub, G., Schaepman, M.E., Martonchik, J., Painter, T., & Dangel, S. (2009). Radiometry and reflectance: From terminology concepts to measured quantities. *The SAGE Handbook of Remote Sensing* (pp. 215-228): Sage

Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., & Hofer, S. (2012). EeteS - The EnMAP End-to-End Simulation Tool. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 5*, 522-530

Smola, A., & Vishwanathan, S. (2008). Introduction to machine learning. *Cambridge University, UK, 32*, 34

Sohl-Dickstein, J., Poole, B., & Ganguli, S. (2014). Fast large-scale optimization by unifying stochastic gradient and quasi-Newton methods. In, *International Conference on Machine Learning* (pp. 604-612)

Storch, T., Bachmann, M., Honold, H.-P., Kaufmann, H., Krawczyk, H., Müller, R., Sang, B., Schneider, M., Segl, K., & Chlebek, C. (2014). EnMAP data product standards. In, *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 2586-2589): IEEE

Suits, G.H. (1971). The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment, 2*, 117-125

Summerfield, M. (2007). *Rapid GUI programming with Python and Qt: the definitive guide to PyQt programming*. Pearson Education

Thenkabail, P. (2017). Hyperspectral remote sensing of vegetation and agricultural crops

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment, 91*, 354-376

van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitão, P., Okujeni, A., Schwieder, M., Suess, S., & Hostert, P. (2015). The EnMAP-Box—A Toolbox and Application Programming Interface for EnMAP Data Processing. *Remote Sensing, 7*, 11249-11266

van der Meer, F., De Jong, S.M., & Bakker, W.H. (2001). Imaging Spectrometry: Basic Analytical Techniques. In F. Van der Meer, & S.M. De Jong (Eds.), *Imaging Spectrometry* (pp. 17-61). The Netherlands: Kluwer Academic

van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., & Su, Z. (2009). An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences, 6*, 3109-3129

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., & Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sensing of Environment, 152*, 654-664

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment, 16*, 125-141

Verhoef, W., & Bach, H. (2003). Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of Environment, 87*, 23-41

Verhoef, W., & Bach, H. (2007). Coupled soil–leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sensing of Environment, 109*, 166-182

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., & Moreno, J. (2018). Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surveys in Geophysics*, 1-41

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment, 118*, 127-139

Verrelst, J., Rivera, J., & Moreno, J. (2015). ARTMO's Global Sensitivity Analysis (GSA) toolbox to quantify driving variables of leaf and canopy radiative transfer models. *EARSeL eProc, 14*, 1-11

Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J., & Camps-Valls, G. (2016). Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation, 52*, 554-567

Verrelst, J., Rivera, J.P., Leonenko, G., Alonso, L., & Moreno, J. (2014). Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. *Geoscience and Remote Sensing, IEEE Transactions on, 52*, 257-269

Wang, L.a., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal, 4*, 212-219

Wang, W.-M., Li, Z.-L., & Su, H.-B. (2007). Comparison of leaf angle distribution functions: effects on extinction coefficient and fraction of sunlit foliage. *Agricultural and Forest Meteorology, 143*, 106-122

Wang, Z., Skidmore, A.K., Darvishzadeh, R., Heiden, U., Heurich, M., & Wang, T. (2015). Leaf Nitrogen Content Indirectly Estimated by Leaf Traits Derived From the PROSPECT Model. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*, 3172 - 3182

Warmerdam, F. (2008). The geospatial data abstraction library. *Open source approaches in spatial data handling* (pp. 87-104): Springer

Waske, B., Fauvel, M., Benediktsson, J.A., & Chanussot, J. (2009). Machine learning techniques in remote sensing data analysis. *Kernel methods for remote sensing data analysis*, 3-24

Weiss, M., & Baret, F. (2016). S2ToolBox level 2 products: LAI, FAPAR, FCOVER. *Institut National de la Recherche Agronomique (INRA), Avignon*

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie, 20*, 3-22

Weiss, M., Troufleau, D., Baret, F., Chauki, H., Prévot, L., Olioso, A., Bruguier, N., & Brisson, N. (2001). Coupling canopy functioning and radiative transfer models for remote sensing data assimilation. *Agricultural and Forest Meteorology, 108*, 113-128

White, H. (1992). *Artificial neural networks: approximation and learning theory*. Cambridge, USA: Blackwell Publishers, Inc.

Wocher, M., Berger, K., Danner, M., Mauser, W., & Hank, T. (2018). Physically-Based Retrieval of Canopy Equivalent Water Thickness Using Hyperspectral Data. *Remote Sensing, 10*, 17

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., & Philip, S.Y. (2008). Top 10 algorithms in data mining. *Knowledge and information systems, 14*, 1-37

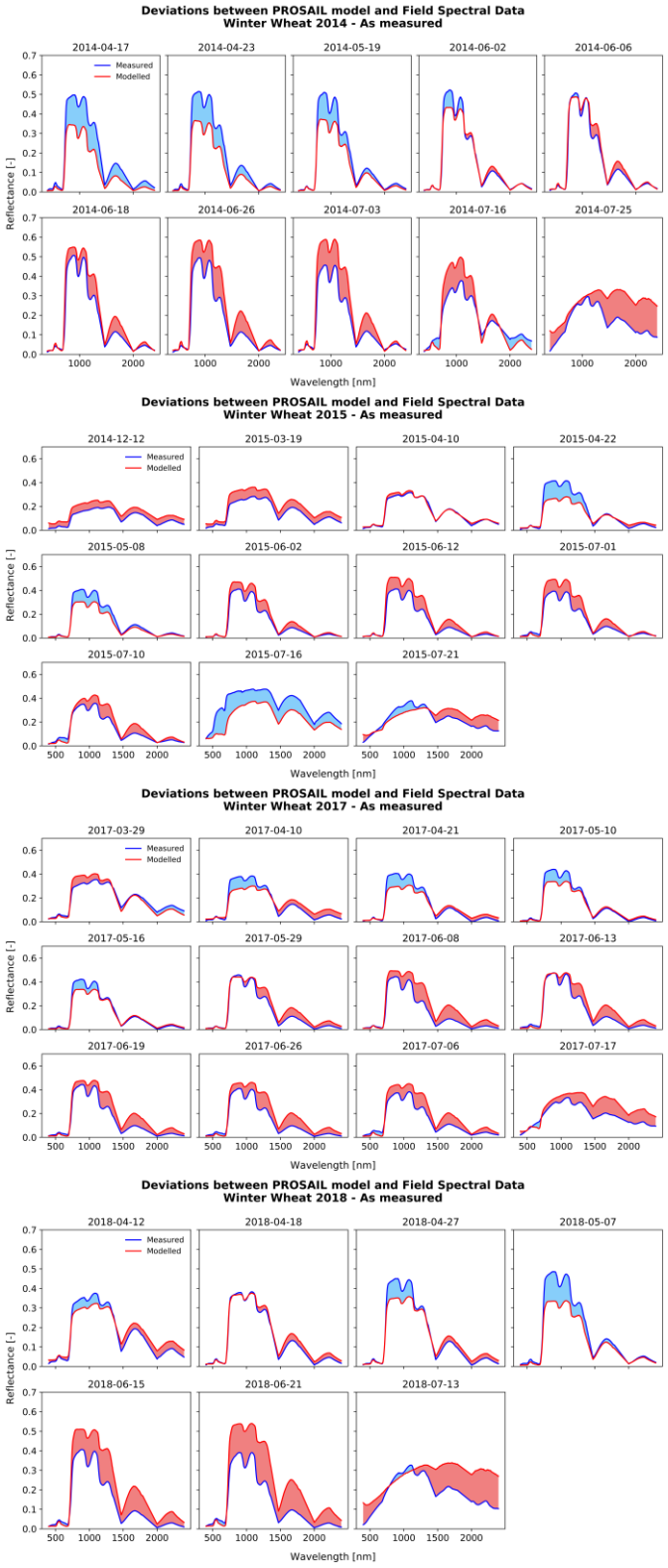
Yang, X., Huang, J., Wu, Y., Wang, J., Wang, P., Wang, X., & Huete, A.R. (2011). Estimating biophysical parameters of rice with remote sensing data using support vector machines. *Science China Life Sciences, 54*, 272-281

Zabel, F., Putzenlechner, B., & Mauser, W. (2014). Global agricultural land resources–a high resolution suitability evaluation and its perspectives until 2100 under climate change conditions. *PLoS ONE, 9*, e107522

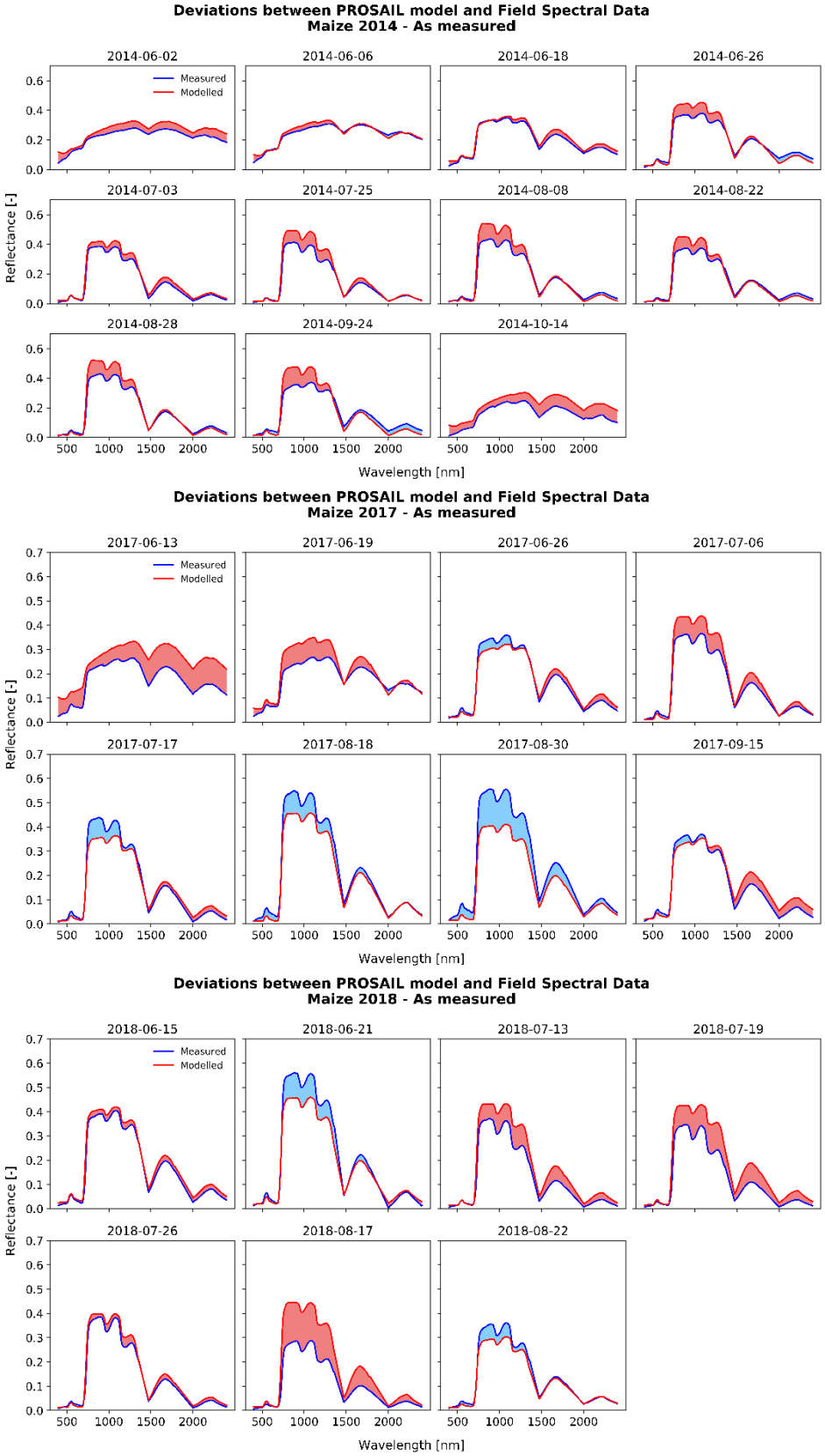
Ziliani, M., Ershadi, A., Zheng, B., Ait-El-Fquih, B., Altaf, M., Aragon, B., Hoteit, I., Franz, T., Chapman, S., & McCabe, M. (2018). Corn-Yield Estimation through Assimilation of Remotely Sensed LAI Data into APSIM. In, *AGU Fall Meeting Abstracts*

Supplementary Material

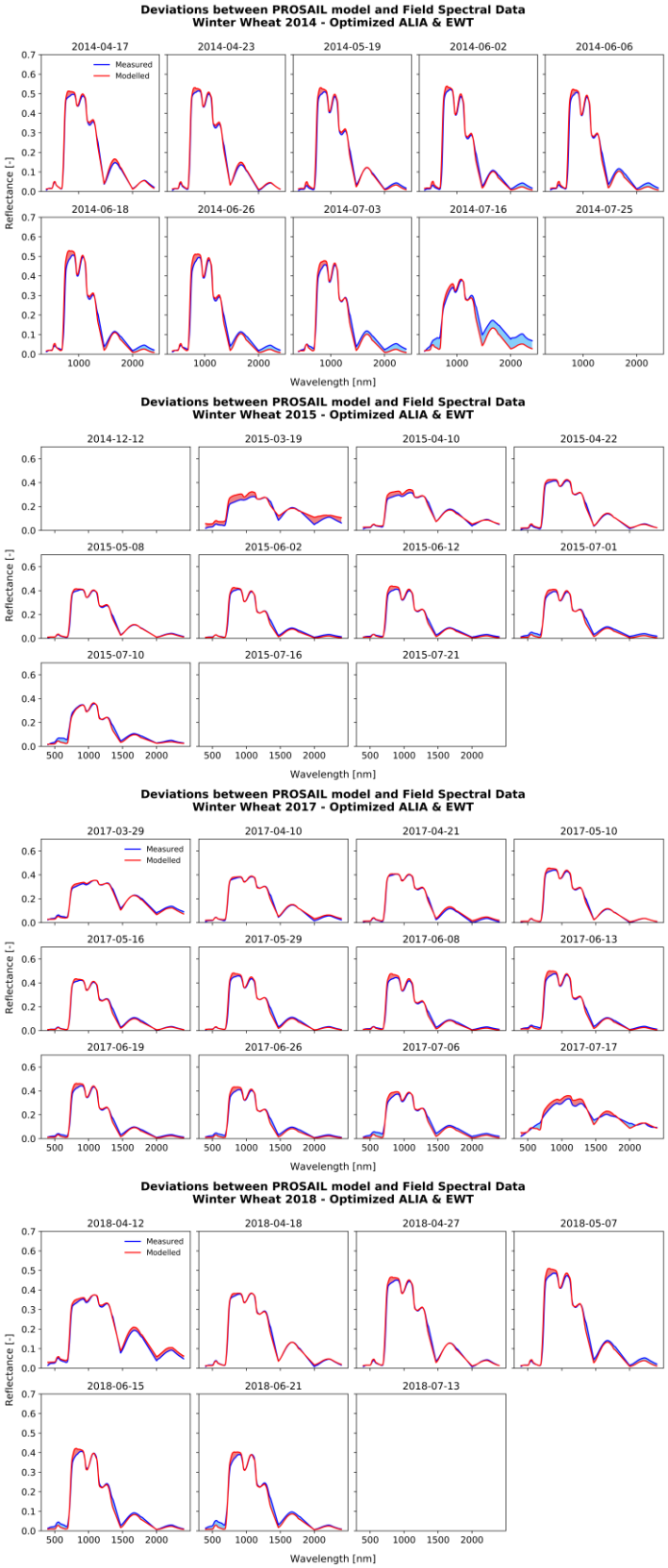
S1. Deviations between the PROSAIL model and field spectral data for winter wheat seasons of 2014, 2015, 2017, 2018 – as measured.



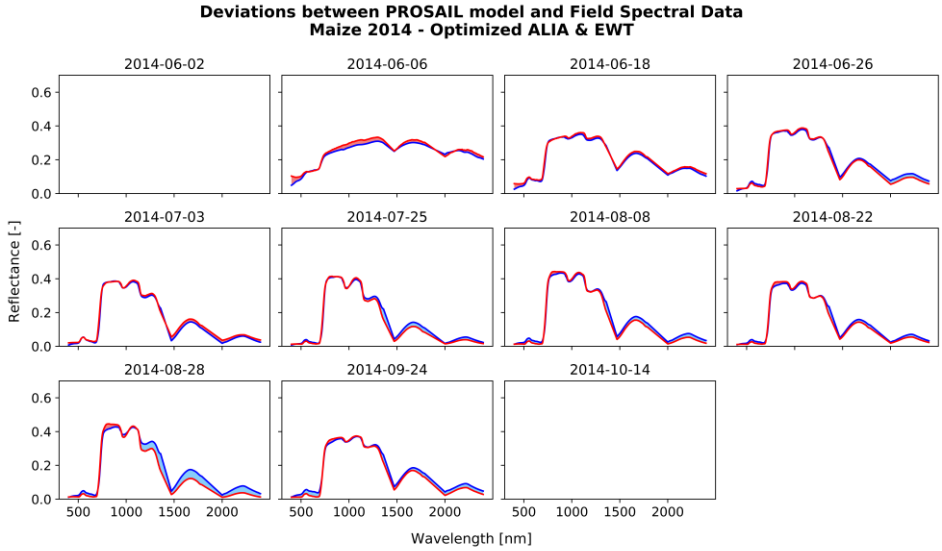
S2. Deviations between the PROSAIL model and field spectral data for maize seasons of 2014, 2017, 2018 – as measured.

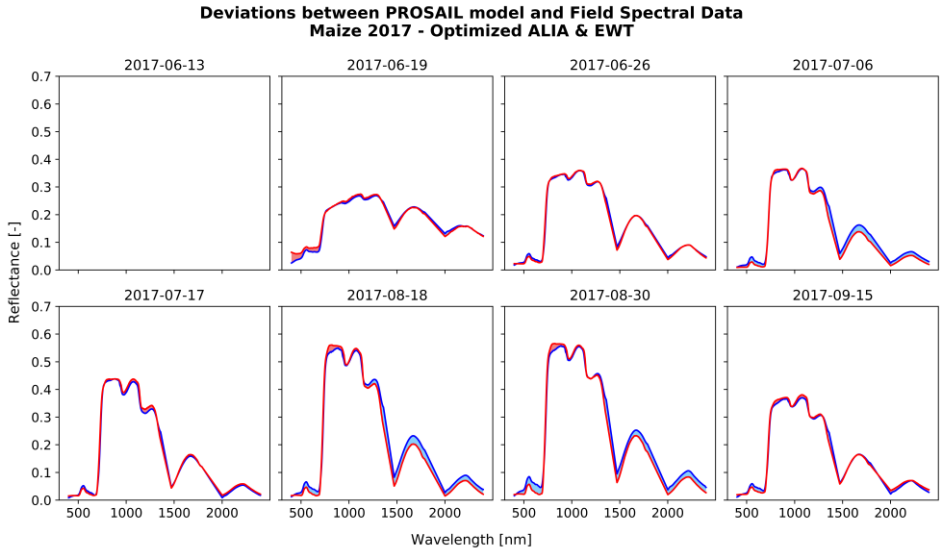


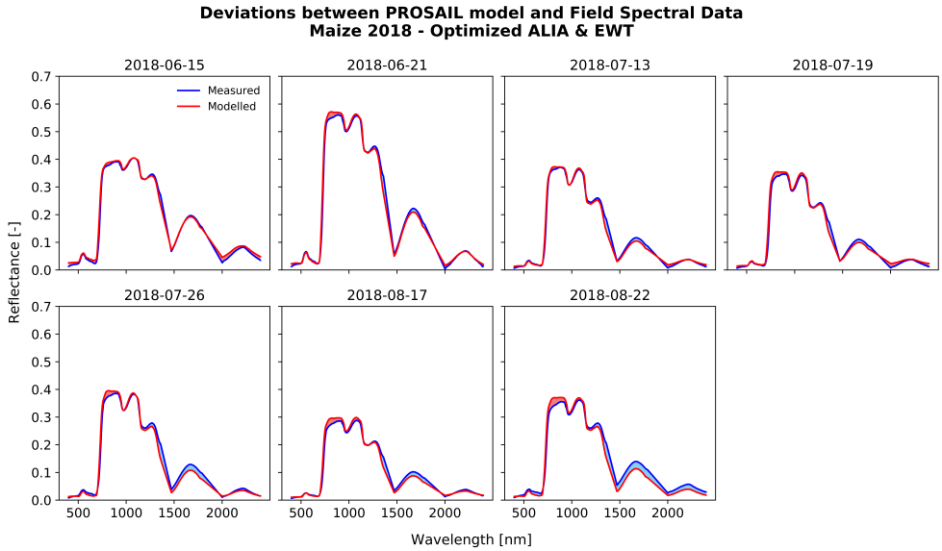
S3. Deviations between the PROSAIL model and field spectral data for winter wheat seasons of 2014, 2015, 2017, 2018 – optimized ALIA & EWT.



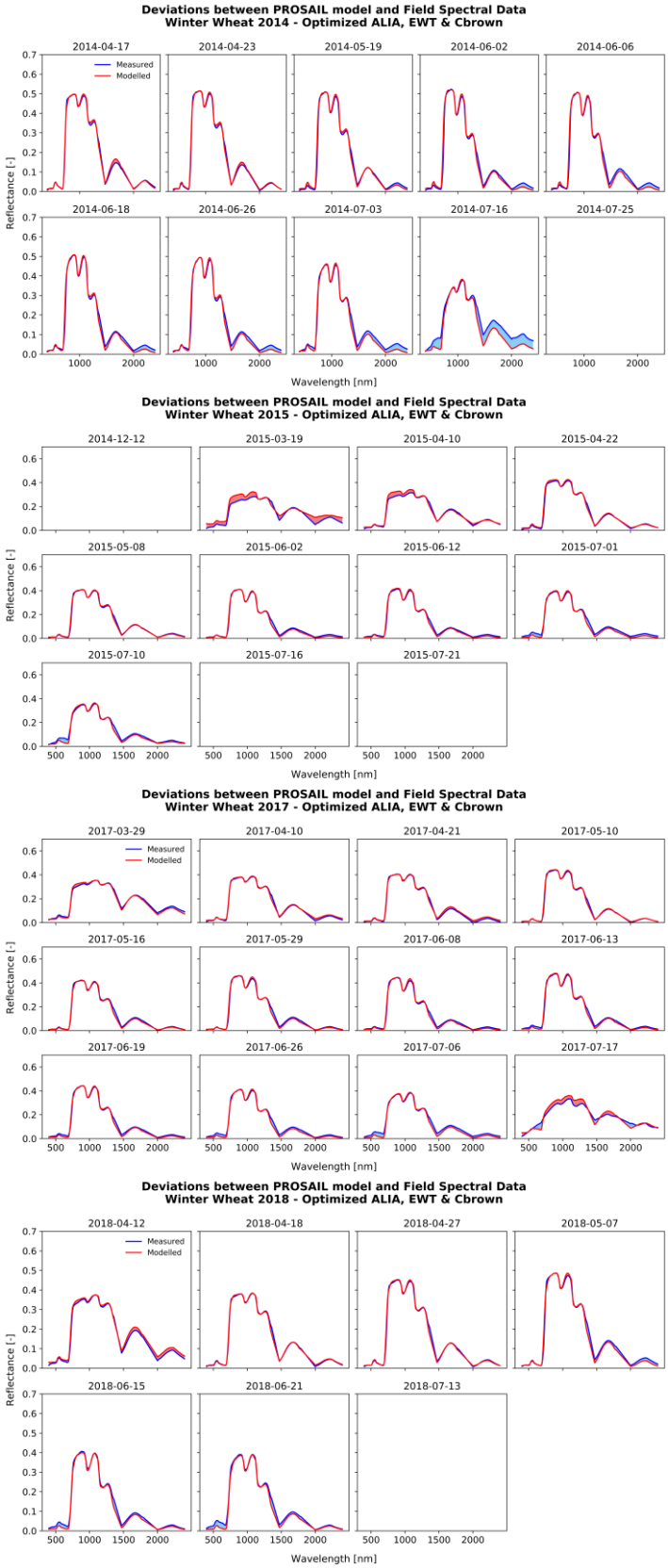
S4. Deviations between the PROSAIL model and field spectral data for maize seasons of 2014, 2017, 2018 – optimized ALIA & EWT.



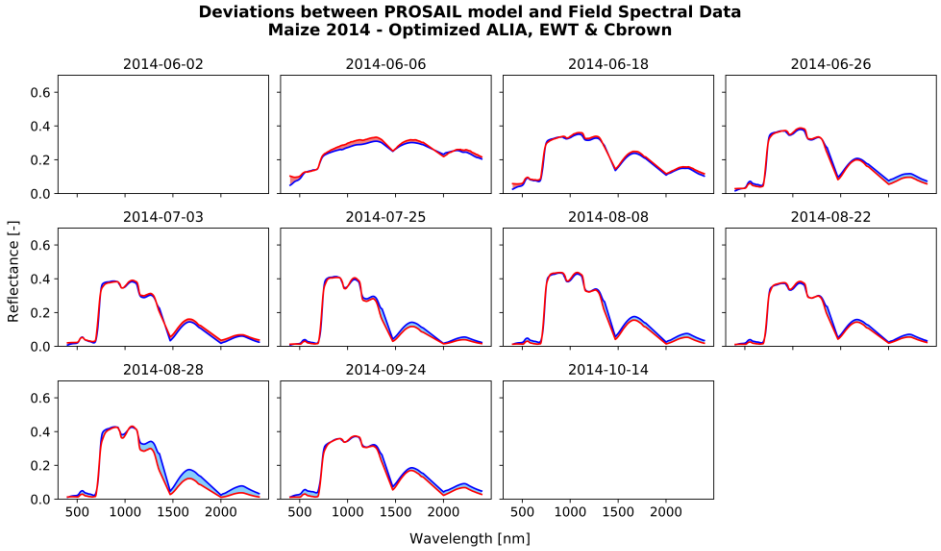


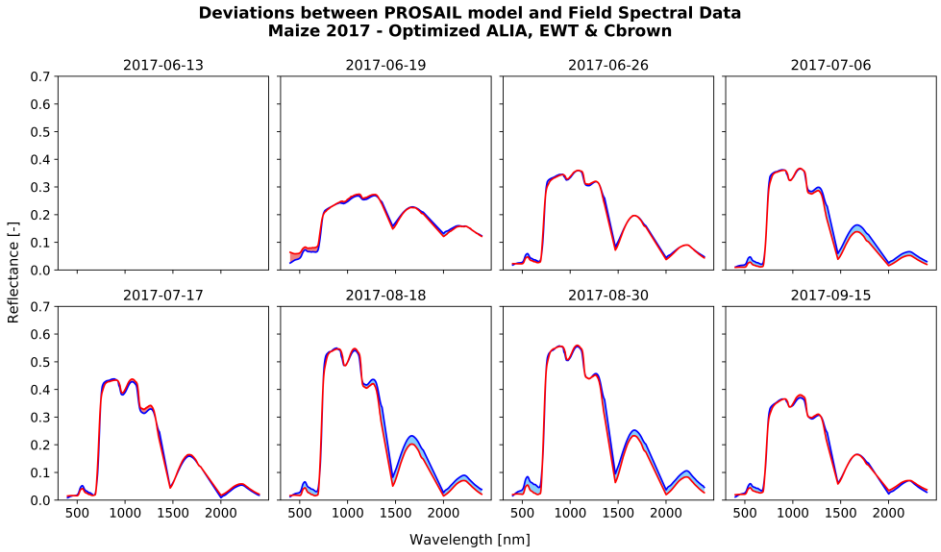


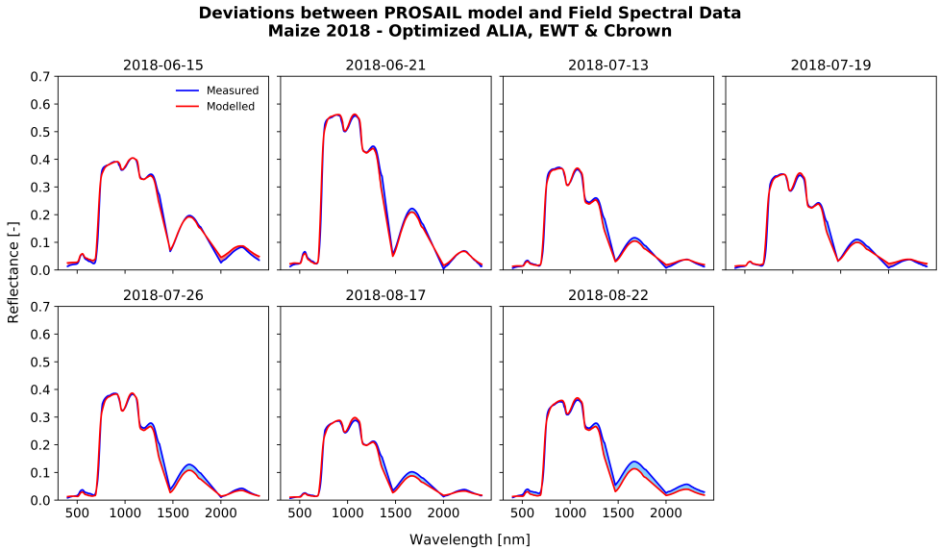
S5. Deviations between the PROSAIL model and field spectral data for winter wheat seasons of 2014, 2015, 2017, 2018 – optimized ALIA, EWT & C_{brown}.



S6. Deviations between the PROSAIL model and field spectral data for maize seasons of 2014, 2017, 2018 – optimized ALIA, EWT and C_{brown}.







S7. Manual fitting of ALIA, EWT and C_{brown} for four winter wheat growing seasons.

Date	Phenology [BBCH]	ALIA [deg]		EWT [cm]		C_{brown} [-]	
		in situ	*optimized*	in situ	*optimized*	in situ	*optimized*
2014-04-17	3	75	*57*	0.035	*0.026*	0.0	*0.10*
2014-04-23	3	74	*55*	0.032	*0.032*	0.0	*0.08*
2014-05-19	4	74	*56*	0.028	*0.042*	0.0	*0.1*
2014-06-02	5	67	*49*	0.025	*0.060*	0.0	*0.1*
2014-06-06	6	58	*49*	0.026	*0.058*	0.0	*0.15*
2014-06-18	7	40	*43*	0.027	*0.058*	0.1	*0.25*
2014-06-26	7	25	*44*	0.026	*0.060*	0.1	*0.28*
2014-07-03	8	28	*54*	0.030	*0.056*	0.1	*0.33*
2014-07-16	8	39	*62*	0.026	*0.040*	0.5	*1.00*
2014-07-25	8	40	**	0.012	**	1.0	**
2014-12-12	1	80	**	0.030	**	0.0	**
2015-03-19	2	40	*55*	0.026	*0.080*	0.0	*0.00*
2015-04-10	2	68	*66*	0.032	*0.035*	0.0	*0.00*
2015-04-22	3	77	*52*	0.034	*0.043*	0.0	*0.05*
2015-05-08	3	77	*67*	0.031	*0.034*	0.0	*0.07*
2015-06-02	5	66	*70*	0.026	*0.048*	0.0	*0.10*
2015-06-12	6	56	*66*	0.028	*0.049*	0.0	*0.15*
2015-07-01	7	60	*69*	0.025	*0.044*	0.1	*0.20*
2015-07-10	8	55	*63*	0.026	*0.066*	0.5	*0.30*
2015-07-16	8	35	**	0.018	**	0.9	**
2015-07-21	8	70	**	0.008	**	0.9	**
2017-03-29	2	54	*65*	0.023	*0.018*	0.0	*0.00*
2017-04-10	2	77	*63*	0.018	*0.037*	0.0	*0.05*
2017-04-21	2	78	*67*	0.020	*0.031*	0.0	*0.05*
2017-05-10	3	76	*65*	0.018	*0.035*	0.0	*0.08*
2017-05-16	3	77	*69*	0.018	*0.037*	0.0	*0.08*
2017-05-29	4	69	*64*	0.013	*0.043*	0.0	*0.10*
2017-06-08	6	64	*65*	0.014	*0.052*	0.0	*0.15*
2017-06-13	6	64	*58*	0.014	*0.050*	0.1	*0.20*
2017-06-19	6	64	*64*	0.014	*0.048*	0.1	*0.20*
2017-06-26	7	65	*66*	0.013	*0.050*	0.1	*0.25*
2017-07-06	8	65	*69*	0.013	*0.037*	0.1	*0.30*
2017-07-17	8	45	*44*	0.003	*0.032*	0.8	*1.00*
2018-04-12	2	69	*57*	0.019	*0.027*	0.0	*0.05*
2018-04-18	3	69	*67*	0.019	*0.031*	0.0	*0.05*
2018-04-27	3	73	*60*	0.018	*0.038*	0.0	*0.10*
2018-05-07	3	76	*57*	0.018	*0.036*	0.0	*0.15*
2018-06-15	6	61	*70*	0.014	*0.045*	0.0	*0.15*
2018-06-21	6	58	*72*	0.012	*0.038*	0.0	*0.20*
2018-07-13	8	45	**	0.001	**	1.0	**

** The low amount of biomass did not allow a proper fitting for this variable and date.

S8. Manual fitting of ALIA, EWT and C_{brown} for three maize growing seasons.

Date	Phenology [BBCH]	ALIA [deg]		EWT [cm]		C_{brown} [-]	
		in situ	*optimized*	in situ	*optimized*	in situ	*optimized*
2014-06-02	1	40	**	0.031	**	0.00	**
2014-06-06	3	42	*50*	0.031	*0.022*	0.00	*0.15*
2014-06-18	3	45	*44*	0.026	*0.036*	0.00	*0.05*
2014-06-26	3	36	*55*	0.027	*0.031*	0.00	*0.10*
2014-07-03	3	54	*60*	0.031	*0.035*	0.00	*0.05*
2014-07-25	6	56	*66*	0.026	*0.037*	0.00	*0.05*
2014-08-08	7	42	*62*	0.028	*0.025*	0.00	*0.05*
2014-08-22	7	62	*70*	0.027	*0.023*	0.00	*0.08*
2014-08-28	7	49	*62*	0.025	*0.037*	0.00	*0.20*
2014-09-24	8	51	*69*	0.028	*0.017*	0.07	*0.20*
2014-10-14	9	75	**	0.011	**	0.81	**
2017-06-13	3	49	**	0.016	**	0.00	**
2017-06-19	3	51	*77*	0.021	*0.028*	0.00	*0.05*
2017-06-26	3	71	*64*	0.012	*0.022*	0.00	*0.10*
2017-07-06	5	65	*72*	0.014	*0.022*	0.00	*0.06*
2017-07-17	6	70	*58*	0.018	*0.028*	0.00	*0.06*
2017-08-18	8	64	*43*	0.014	*0.025*	0.00	*0.10*
2017-08-30	8	64	*28*	0.017	*0.024*	0.00	*0.10*
2017-09-15	8	68	*63*	0.014	*0.030*	0.05	*0.18*
2018-06-15	3	51	*54*	0.025	*0.032*	0.00	*0.10*
2018-06-21	3	49	*14*	0.025	*0.033*	0.00	*0.10*
2018-07-13	6	60	*67*	0.023	*0.045*	0.00	*0.08*
2018-07-19	7	60	*68*	0.020	*0.050*	0.00	*0.15*
2018-07-26	7	65	*65*	0.025	*0.042*	0.00	*0.15*
2018-08-17	8	55	*74*	0.022	*0.042*	0.00	*0.20*
2018-08-22	8	75	*66*	0.022	*0.039*	0.01	*0.20*

** The low amount of biomass did not allow a proper fitting for this variable and date.