Hyperspectral Remote Sensing for Winter Wheat Leaf Area Index Assessment in Precision Agriculture

Dissertation

Zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

des Fachbereichs Mathematik/Informatik

der Universität Osnabrück

Vorgelegt von

Dipl. Geograph Bastian Siegmann

Betreuer:

Dr. Thomas Jarmer

Universität Osnabrück
Institut für Informatik
Arbeitsgruppe Fernerkundung

20.07.2016

This thesis is dedicated to my brother *Manuel*, whom I miss very much.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor *Thomas Jarmer* for his continuous support throughout my Ph.D. study and related research, as well as for his patience, motivation, enthusiasm, and immense knowledge in the field of remote sensing and statistics. Thank you *Thom* for not only being an excellent supervisor, but also for all your help outside work, along with the numerous professional and private conversations and discussions.

My sincere thanks also goes to *Manfred Ehlers* for offering me the opportunity to work as a lecturer at the former Institute for Geoinformatics and Remote Sensing (IGF), as well as his willingness to introduce me to the wonderful world of image fusion.

Furthermore, I am very grateful to *Patrick Hostert* for being my co-supervisor and giving me the chance to present my intermediate Ph.D. results in the LSSC-colloquium in Berlin. Moreover, I want to thank him for his motivating words after my first presentation in front of a large audience during the second EnMAP workshop in 2012.

I would also like to thank my colleagues in the projects HyLand and DryLand – namely, *Holger Lilienthal*, *Bernhard Höfle*, *Thomas Selige*, *Nicole Richter* and *Heike Gerighausen* – for the fruitful discussions we had during project meetings and the help during numerous field campaigns in Köthen, Braunschweig and Demmin in the last few years.

I also extend my gratitude to several colleagues at the University of Osnabrück for assisting me during my doctoral studies and supporting me when it came to scientific and technical issues. I cannot mention everybody, but the following people must be singled out: *Florian Beyer*, *Martin Kanning*, *Jan Bauer, Sascha Klonus*, *Florian Hillen*, *Andreas Wichmann, Sandra Dützer*, *Timo Otte-Vinke* and *Maik Wannink*.

Finally, I must express my very profound gratitude to my family and to my wife *Jasmin* for providing me with unfailing support and continuous encouragement throughout my years of study, as well as in the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Abstract

Remote sensing provides temporal, spectral and spatial information covering a wide area. Therefore, it has great potential in offering a detailed quantitative determination of the leaf area index (LAI) and other crop parameters in precision agriculture. The spatially differentiated assessment of LAI is of utmost importance for enabling an adapted field management, with the aim of increasing yields and reducing costs at the same time.

The scientific focus of this work was the investigation of the potential of hyperspectral remote sensing data of different spectral resolutions, which were acquired at different spatial scales, for a precise assessment of wheat LAI. For this reason, three research experiments were conducted: 1) a comparison of different empirical-statistical regression techniques and their capabilities for a robust LAI prediction; 2) a determination of the required spectral resolution and important spectral regions/bands for precise LAI assessment; and 3) an investigation of the influence of the ground sampling distance of remote sensing images on the quality of spatial LAI predictions.

The first part of this thesis compared three empirical-statistical regression techniques – namely, partial least-squares regression (PLSR), support vector regression (SVR) and random forest regression (RFR) – and their achieved model qualities for the assessment of wheat LAI from field reflectance measurements. In this context, the two different validation techniques – leave-one-out cross-validation (cv) and independent validation (iv) – were applied for verifying the accuracy of the different empirical-statistical regression models. The results clearly showed that model performance markedly depends on the validation technique used to assess model accuracy. In the case of leave-one-out cross-validation, SVR provided the best results, while PLSR proved to be superior to SVR and RFR when independent validation was applied.

In the second part of this thesis, the spectral characteristics of the hyperspectral airborne sensor aisaDUAL (98 spectral bands) and the upcoming hyperspectral satellite mission EnMAP (204 spectral bands) were investigated to show their capability regarding the precise determination of wheat LAI. Moreover, the feature selection algorithm RReliefF, combined with a randomized sampling approach, was applied to identify the spectral bands that were most sensitive to changes in LAI. The results demonstrated that only three spectral bands of aisaDUAL, as well as EnMAP, at specific locations within the investigated spectral range (400–2,500 nm) were necessary for an accurate LAI prediction.

The third part of this thesis dealt with the influence of the spatial resolution of aisaDUAL (3 m) and simulated EnMAP (30 m) image data on the assessment of wheat

LAI. While the ground sampling distance (GSD) of aisaDUAL allowed a robust regression model calibration and validation, LAI predictions based on simulated EnMAP image data led to poor results because of the distinct difference in size between the EnMAP pixels (900 m^2) and the sampled field plots (0.25 m^2) for which the LAI was measured. In order to enable a more precise determination of wheat LAI from EnMAP image data, the two different approaches of image aggregation and image fusion were examined. In this context, the fusion approach has proven to be the more suitable method, which allowed a more accurate LAI prediction compared to the results based on the EnMAP data with a GSD of 30 m.

In summary, the findings of the research reported in this thesis demonstrated that the accuracy of spatial LAI predictions from remote sensing data depends on several factors. Besides the applied empirical-statistical retrieval- and validation method, the spatial and spectral characteristics of the used image data sets played an important role. With the forthcoming hyperspectral satellite missions (e.g., EnMAP, HyspIRI), the area-wide assessment of LAI and other crop parameters (e.g., biomass, chlorophyll content) will be strongly supported. The moderate spatial resolutions of these satellites systems, however, require a combined use with spatial higher resolution multi- or superspectral satellite data (e.g., RapidEye, Sentinel-2). This multisensoral approach offers great potential for the prompt identification of spatial variations in crop conditions on sub-field scale, which is a mandatory prerequisite for precision agricultural applications.

Zusammenfassung

Fernerkundungsdaten haben großes Potential für die quantitative Erfassung von Pflanzenparametern im Bereich der Präzisionslandwirtschaft, da sie spektrale Informationen großer Flächen liefern, die zu verschiedenen Zeitpunkten innerhalb der Wachstumsperiode aufgenommen werden können. Die räumlich differenzierte Erfassung des Blattflächenindex (engl. LAI) ist dabei von besonderer Bedeutung für eine angepasste Bewirtschaftung von Feldern mit dem Ziel die Erträge zu steigern und gleichzeitig die Kosten zu senken.

Die vorliegende Arbeit beschäftigt sich mit der Untersuchung des Potentials hyperspektraler Fernerkundungsdaten mit unterschiedlicher spektraler und räumlicher Auflösung zur genauen Erfassung des LAI von Winterweizen. Zu diesem Zweck wurden drei verschiedene Forschungsexperimente durchgeführt: 1) Vergleich verschiedener empirisch-statistischer Regressionsverfahren und deren Fähigkeiten zur genauen Vorhersage des LAI; 2) Ermittlung der nötigen spektralen Auflösung und wichtiger spektraler Regionen/Kanäle zur möglichst exakten Erfassung des LAI; und 3) Untersuchung des Einflusses der räumlichen Auflösung von Fernerkundungsdaten auf die Genauigkeit von räumlichen LAI Vorhersagen.

Im ersten Teil der Arbeit werden verschiedene Ergebnisse der drei empirisch-statistischen Regressionsverfahren Partial Least Squares Regression (PLSR), Support Vector Regression (SVR) und Random Forest Regression (RFR) verglichen, die im Zuge der Vorhersage des LAI von Winterweizen auf Grundlage von Feldspektrometermessungen erzielt wurden. In diesem Zusammenhang wurden die zwei Validierungsmethoden „leave-one-out“ Kreuzvalidierung (cv) und eine Form der unabhängigen Validierung (iv) vergleichend angewandt, um die Genauigkeit der trainierten Regressionsmodelle zu ermitteln. Die Ergebnisse verdeutlichen, dass die Modellgüte sehr stark von der verwendeten Validierungsmethode abhängt. Während im Fall der „leave-one-out“ Kreuzvalidierung das Regressionsverfahren SVR das beste Ergebnis lieferte, konnte bei der unabhängigen Validierung die höchste Vorhersagegenauigkeit mit dem Regressionsverfahren PLSR erzielt werden.

Der zweite Teil der Arbeit beschäftigt sich mit der Untersuchung der spektralen Eigenschaften des flugzeuggetragenen Sensors aisaDUAL (98 spektrale Kanäle) und des zukünftigen Satelliten EnMAP (204 spektrale Kanäle) zur Erfassung des LAI. Darüber hinaus wurden mit dem Variablenselektionsverfahren RReliefF und zwei daran angeschlossenen randomisierten Experimenten die spektralen Kanäle ermittelt, die am sensitivsten auf Veränderungen des LAI reagieren. Die dabei erzielten Ergebnisse haben gezeigt, dass sowohl bei aisaDUAL- als auch EnMAP jeweils nur drei spektrale Kanäle an bestimmten Positionen im untersuchten Spektralbereich (400–2.500 nm) nötig waren, um genaue LAI Vorhersagen zu ermöglichen.

Im dritten Teil der Arbeit wurde die Vorhersagegenauigkeit des LAI in Abhängigkeit von der räumlichen Auflösung der Sensoren aisaDUAL (3 m) und EnMAP (30 m) analysiert. Während die räumliche Auflösung von aisaDUAL eine stabile Kalibrierung und Validierung eines Regressionsmodells ermöglichte, führten die Vorhersagen des LAI auf Grundlage simulierter EnMAP-Bilddaten zu wesentlich niedrigeren Genauigkeiten. Der Grund dafür bestand im deutlichen Größenunterschied der abgedeckten Flächen eines EnMAP-Pixels (900 m^2) im Vergleich zu den im Feld beprobten Plots (0,25 m^2), für die der LAI erfasst wurde. Aus diesem Grund wurden die beiden Verfahren Bildaggregation und Bildfusion auf den simulierten EnMAP-Daten angewendet um genauere Vorhersagen des LAI zu ermöglichen. Dabei hat sich die Bildfusion als das geeignetere Verfahren herausgestellt.

Zusammenfassend verdeutlichen die Ergebnisse dieser Arbeit, dass die räumliche Vorhersagegenauigkeit des LAI auf Grundlage von Fernerkundungsdaten von verschiedenen Faktoren abhängt. Neben den verwendeten empirisch-statistischen Regressionsverfahren und den verschiedenen Validierungsmethoden haben vor allem die räumliche und spektrale Auflösung der Bilddaten erheblichen Einfluss auf die genaue Ableitung des LAI. Mit den zukünftigen hyperspektralen Satellitenmissionen (z.B. EnMAP, HyspIRI) wird eine noch präzisere Erfassung des LAI und anderer wichtiger Pflanzenparameter (z.B. Biomasse, Chlorophyllgehalt) möglich sein. Allerdings stellt die zu geringe räumliche Auflösung dieser Satelliten für bestimmte Anwendungen innerhalb der Präzisionslandwirtschaft ein Problem dar und erfordert eine Kombination mit räumlich höher aufgelösten Daten multi- oder superspektraler Satelliten (z.B. RapidEye, Sentinel-2). Ein solcher multisensoraler Ansatz bietet großes Potential für die zeitnahe Erfassung von Veränderungen der Wachstumsbedingungen und ermöglicht die Ableitung wichtiger Informationen über den Zustand der Pflanzen, was wiederum eine wichtige Voraussetzung für eine angepasste und räumlich differenzierte Bewirtschaftung von Feldern ist.

Table of Contents

List of Figures

List of Tables

Index of Abbreviations

AIC	Akaike information criterion
BBCH scale	scale of Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie
BRDF	bidirectional reflectance distribution function
BRF	bidirectional reflectance factor
Cab	leaf chlorophyll content
CARI	chlorophyll absorption ratio index
CART	classification and regression trees
CRCWD	continuum removed chlorophyll well depth
cv	leave-one-out cross-validation
DEM	digital elevation model
DGVI	derivative green vegetation index
DM	above-ground biomass dry matter
EeteS	EnMAP end-to-end simulation tool
EnMAP	Environmental Mapping and Analysis Program
EWT	equivalent water thickness
FFT transform	fast Fourier transform
FM	above-ground biomass fresh matter
FOV	field of view
FWHM	full width at half maximum
GIS	geographic information system
GNSS	global navigation satellite system
GSD	ground sampling distance
HCS	Hyperspherical Color Sharpening

HyspIRI	Hyperspectral Infrared Imager
IHS transform	intensity-hue-saturation transform
iv	independent validation
JRC	Joint research Centre of the European Union
LAI	leaf area index
LAI_{eff}	effective leaf area index
LIDAR	light detection and ranging
LV	latent variables
MARS	multivariate adaptive regression splines
MLR	multiple linear regression
MODTRAN	MODerate resolution atmosheric TRANsmission
NDVI	normalized differenced vegetation index
NIR	near infrared
NNR	neural net regression
PAR	photosynthetically active radiation
PCA	principle component analysis
PCR	principle component regression
PLSR	partial least squares regression
PVI	perpendicular vegetation index
PWC	plant water content
R	correlation coefficient
R^2	coefficient of determination
radar	radio detection and ranging
RAMI	Radiation Transfer Model Inter-comparisons initiative
REIP	red edge inflection point
RFR	random forest regression

RMSE	root-mean-square error
RPD	residual prediction deviation
RTM	radiative transfer model
SAVI	soil-adjusted vegetation index
SNR	signal-to-noise ratio
SVM	support vector machine
SVR	support vector regression
SWIR	short wave infrared
TWI	topographic wetness index
VI	vegetation index
VIS	visible light
WI	water index
α_{spec}	spectral angle
ρ_{canopy}	canopy reflectance

1 Introduction

At present, more than 7.2 billon people live on earth [U. S. Census Bureau, 2015], while forecasts indicate that by 2050, the world's population will rise to more than nine billion [DESA, 2015]. To feed the growing population, more than 500 million ha of additional land have been cultivated in the last 50 years [FAO, 2015] and this process will continue well into the future [Alexandratos and Bruinsma, 2012]. In total 37.72 % of the earth's surface is used for agriculture [FAO, 2015]. Hence, agriculturalists may be regarded as the chief managers of terrestrial usable land [Tilman et al., 2002].

In addition to rice and maize, wheat is one of the most important food grain sources for humans, which is grown on more land area than any other commercial crop in the world [CIMMYT, 2015], because it has a high yield potential and a good adaptability to different climatic conditions and soil types [BLV, 2014]. Due to limited natural resources (e.g., suitable soils, water), it is not possible to increase food production based on wheat and other crops indefinitely in the future. For this reason, modern agriculture faces considerable challenges.

To counter the rising global population and the limited availability of arable land, the European Commission declared the further development and rapid diffusion of sustainable agricultural practices and precision agricultural approaches as two of the major objectives in the coming decade [European Commission, 2012]. Precision agriculture is defined as an information-based management system using sustainable methods to increase crop productivity and farm profitability, while reducing negative environmental impacts that are often caused by conventional agriculture (e.g., over-supply of fertilizers, pesticides or herbicides) [Gebbers and Adamchuk, 2010; Mulla, 2013; Zhang et al., 2002].

The first precision agricultural approaches were applied in the 1980s to develop the right treatment of agricultural fields in the right place at the right time [Gebbers and Adamchuk, 2010; Robert, 2002]. Since the 1990s, it has been practised commercially in the USA, Europe, Australia, Canada and Japan [Mulla, 2013]. In accordance with Crookston [2006], it belongs to the top ten revolutions in agriculture today. One of the most important factors in the development and dissemination of precision agriculture is the increasing awareness of variation in soil and crop conditions. In addition the emergence and further development of technologies, such as geographic information systems (GIS), global navigation satellite systems (GNSS), miniaturized computers, mobile computing, advanced information processing, telecommunications, and proximal and remote sensing, have made an important contribution to the dissemination of precision agriculture [Gebbers and Adamchuk, 2010; Zhang et al., 2002].

In particular, remote sensing is a key technology because airborne and spaceborne sensors provide spatial, spectral and temporal information about agricultural fields, which can be used to identify infield variability and support decision-making. This is

especially important for the development and implementation of an adapted field treatment to avoid over- and/or undersupply of agricultural inputs, a frequent cause of ecological problems and economic drawbacks [Gebbers and Adamchuk, 2010; Moran et al., 1997; Seelan et al., 2003; Zarco-Tejada et al., 2005; Zhang et al., 2002].

The fast, cost-effective and non-destructive assessment of relevant biochemical and structural vegetation properties is of the utmost importance when characterizing the crop status at leaf and canopy level. In this regard, parameters, such as chlorophyll content, above-ground biomass dry matter, nitrogen status, canopy water content and leaf area index (LAI), provide important information for describing current growth conditions, which can be converted into yield-driving state variables (e.g., dry mass increase) and used for the re-parameterization of agricultural production models [Cox, 2002; Delécolle et al., 1992; Mulla, 2013; Schueller, 1992].

The LAI is one of the most important plant parameters and serves as an essential variable for assimilating remote sensing data into crop growth models [Machwitz et al., 2014; Moulin et al., 1998]. It is defined as the ratio of the total one-sided leaf surface area per unit soil surface area [Herrmann et al., 2011; Monteith and Unsworth, 2007; Watson, 1947]. As an indicator for the current biotic and abiotic conditions, the LAI provides information about the photosynthetic 'potential' of plants and is an important input parameter for yield modelling, since it is significantly influenced by yield-limiting and -reducing factors, such as plant diseases and mismanagement [Boegh et al., 2002; Carter, 1994; Daughtry et al., 1992].

Various methods for LAI determination exist, differing in the type of measurement methodology and the requirements of the technical equipment [Bréda, 2003; Zheng and Moskal, 2009]. Destructive methods for measuring LAI normally provide more precise results, but the assessment is time-consuming, expensive and, therefore, often limited to small areas [Bréda, 2003]. The increasing use of multispectral remote sensing data in recent decades has significantly improved the spatial and non-destructive assessment of LAI. Additionally, a rising number of hyperspectral airborne and spaceborne sensors are in operation nowadays. While multispectral data often provide an insufficient number and/or a non-optimal location of spectral bands per pixel, hyperspectral data have a large amount of spectral bands that cover a wide spectral range. Each pixel in a hyperspectral data cube contains spectral information about a large number of adjacent bands. These can be displayed as a continuous spectral curve. The additional spectral information can be used to improve the spatial assessment of LAI and other plant parameters [Gray and Song, 2012; Haboudane et al., 2004; Lee et al., 2004].

In recent years, two complementary groups of approaches have been developed for LAI estimation from remote sensing data. On the one hand, the inversion of physically based radiative transfer modelling approaches provides the possibility to retrieve LAI and other vegetation parameters [Bacour et al., 2002; Fang et al., 2003;

Jacquemoud et al., 1995; Jacquemoud et al., 1996; Vohland and Jarmer, 2008; Weiss et al., 2000]. On the other hand, empirical-statistical regression models (e.g. partial least squares regression (PLSR), support vector regression (SVR), random forest regression (RFR), neural net regression (NNR), or multivariate adaptive regression splines (MARS)) can be calibrated with in situ LAI measurements to establish a relation to reflectance spectra or spectral vegetation indices (VIs) [Curran, 1994; Duchemin et al., 2006; Koppe et al., 2010; Thenkabail et al., 2000; Viña et al., 2011].

While physically based radiative transfer models do not need in situ data about a biophysical/-chemical target variable for model calibration, they are very complex (several input variables are needed) and the ill-posed nature of model inversion can result in an incorrect parameter prediction [Atzberger, 2004; Darvishzadeh et al., 2011]. In contrast to radiative transfer models, the calibration of the empirical-statistical models requires field measurements of the target variable as a basic prerequisite. However, the number of biophysical/-chemical parameter samples measured during a field campaign is often too low for allowing an independent validation of an empirical-statistical model [Cohen et al., 2003]. In this case, cross-validation is widely used as an appropriate method for determining the quality of a regression model [e.g., Efron, 1983; Oredein et al., 2011; Snee, 1977; Steyerberg et al., 2001]. Nevertheless, a well-known problem of empirical-statistical models is overfitting. This phenomenon can be defined as an error that occurs in regression modelling when a model is too closely fit to a limited number of samples [Babyak, 2004; Hawkins, 2004]. Consequently, transferring a cross-validated model developed from data of one agricultural field to another field with the same crop yields mostly limited success and often leads to inaccurate results.

In addition to the methodology on which the parameter retrieval is based, the properties of remote sensing data play an important role. There is no uniform definition of hyperspectral data. According to Baltsavias [2002], a hyperspectral sensor must have at least 10 spectral bands (usually more than 100 bands) and a narrow bandwidth of typically a few tens of nanometres or less for visible light (VIS) and near infrared (NIR). Field spectroradiometers (e.g., ASD Field Spec III (ASD Inc., USA), SVC HR-1024 (Spectra Vista Corporation, USA)) and hyperspectral airborne sensors (e.g., HySpex (Norsk Elektro Optikk, Norway), aisaDUAL (Specim Ltd., Finland)) correspond to this definition. EO-1 Hyperion also meets these criteria and demonstrates the potential of hyperspectral imaging technology for earth observation. Additionally, in the near future, new hyperspectral satellite missions, such as the Environmental Mapping and Analysis Program (EnMAP) or the Hyperspectral Infrared Imager (HyspIRI) scheduled for launch in 2018 and 2022, respectively, will continue hyperspectral earth observation and provide data with improved qualities

with respect to signal-to-noise ratio (SNR) and spatial coverage [Guanter et al., 2015; Lee et al., 2015].

Unfortunately, from a statistical point of view, large parts of hyperspectral data are redundant because adjacent spectral bands are highly correlated [Ghosh et al., 2013; Sohaib et al., 2012]. For this reason, different feature selection techniques have been developed in remote sensing or adapted from other scientific fields [Bolón-Canedo et al., 2013; Guyot et al., 1992; Saeys et al., 2007] in order to reduce the high dimensionality of hyperspectral data and simultaneously identify the most important features with the highest explained variance regarding the target variable (e.g., LAI). In the context of supervised classification and regression methods, these feature selection techniques can be distinguished in filter, wrapper and embedded methods [Guyot et al., 1992; Saeys et al., 2007]. In the field of remote sensing, several feature selection techniques have been developed and applied to classification problems, using hyperspectral field and image data [e.g., Bajwa et al., 2004; Benediktsson et al., 1995; Thenkabail et al., 2004]. Additionally, numerous studies have been conducted using feature selection techniques for the optimization of regression models to predict the LAI [e.g., Darvishzadeh et al., 2011; Li et al., 2014].

Furthermore, the spatial resolution (ground sampling distance (GSD)) of hyperspectral sensors, operating from various spatial scales, is completely different. While field spectroradiometers can only measure small footprints of an investigated crop canopy, airborne sensors acquire image data of entire fields, while satellites even provide information of complete agricultural regions. Due to their GSD, hyperspectral data from field spectroradiometers and airborne sensors were often used for precision agricultural applications in the past [e.g., Haboudane et al., 2004; Jarmer, 2013; Tanaka et al., 2015; Zarco-Tejada et al., 2005]. In contrast, EO-1 Hyperion, the only available scientific hyperspectral satellite system, has a GSD of 30 m, which means that a single pixel represents an averaged spectral reflectance signature of a large area [Chen et al., 2014; Mayumi and Iwasaki, 2011]. This is especially a problem in precision agriculture for monitoring the small-scale intra-field variability of crops, which is important to identify areas affected by stress, diseases and physical damage [Johnson, 2014; Moran et al., 1997; Pinter et al., 2003]. Additionally, remote sensing data with high spatial and spectral resolution covering large areas are highly demanded because they serve as a vital information source for applying an adapted field treatment on a sub-field scale and, thus, facilitate a more precise management practice (e.g., fertilization [Baret et al., 2006] and irrigation [Vuolo et al., 2015]). To overcome the limitation of the medium spatial resolution of current and future hyperspectral satellite sensors, pan-sharpening is an adequate method to enhance the GSD of acquired image data. In this context, hyperspectral image data of lower spatial resolution, which contain numerous spectral bands, are fused with a panchromatic image of higher spatial resolution [Ehlers et al., 2010; Pohl and van Genderen, 2015;

Zhang, 2012]. Numerous studies dealing with pan-sharpening, however, were mostly not focused on a specific application, but rather on the development and validation of new algorithms. Until now, only a few studies focusing on precision agriculture have been conducted using fused multispectral satellite data [Amorós-López et al., 2013; Zurita-Milla et al., 2008] or merged multi- and hyperspectral information acquired from different platforms (satellites, aircraft, unmanned aerial vehicles (UAVs)) [Delalieux et al., 2014; Gevaert et al., 2014] to assess crop conditions.

The presented work in this thesis is based on data collected within the framework of the HyLand research project, 'Hyperspectral remote sensing for the assessment of crop and soil parameters in precision farming and yield estimation'. HyLand was a preparation project for the hyperspectral German satellite mission EnMAP and funded by the German Aerospace Center (DLR) with financial resources from the Federal Ministry of Economics and Technology. During the project, wheat crops were mainly investigated. Consequently, the different methods of LAI assessment and spatial predictions of LAI from airborne and spaceborne data are focused on wheat.

The described approaches and associated problems in the process of estimating the LAI from hyperspectral data lead to the following three primary research questions and ensuing tasks:

1. *What potential do empirical-statistical regression models have for an accurate prediction of crop LAI based on hyperspectral data?*

This topic has been examined by Siegmann and Jarmer [2015][1]. The following concrete tasks were defined to answer the question:

- Compare different empirical-statistical regression algorithms in terms of model suitability and performance for the retrieval of LAI.
- Investigate the influence of the measured LAI range and type of applied validation technique on regression model accuracy and transferability to unknown data.

[1] Siegmann, B. and Jarmer, T. (2015): Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *International Journal of Remote Sensing*, 36(18):4519–4534.

2. *What spectral resolution and location of spectral bands are necessary to enable a precise estimation of the parameter LAI?*

This question is part of a research paper that is not yet published. To answer the question, two main tasks were formulated:

- Use a feature selection approach to identify the spectral bands and/or regions that are most important to derive the parameter LAI.
- Compare LAI prediction accuracies of empirical-statistical regression models using full and reduced spectral resolution.

3. *What influence does the ground sampling distance of remote sensing images have on the spatial assessment of the LAI and the associated regression model validation?*

A study conducted by Siegmann et al. [2015][2] investigated this issue in detail. In this regard, the following two tasks were defined to answer the question:

- Examine the spatial LAI prediction accuracy based on the hyperspectral data of different spatial resolutions (aisaDUAL and EnMAP) when using in situ LAI measurements of small plots as the basis for validation.
- Investigate aggregation and pan-sharpening as two different methods to obtain a more valid LAI prediction based on medium-resolution EnMAP data.

[2] Siegmann, B., Jarmer, T., Beyer, F. and Ehlers, M. (2015). The potential of pan-sharpened EnMAP data for the assessment of wheat LAI. *Remote Sensing*, 7(10):12737–12762.

2 Remote Sensing of Vegetation

Remote sensing systems used for vegetation studies are generally divided into active systems, such as LIDAR (light detection and ranging) and radar sensors (radio detection and ranging), and passive systems, such as optically reflective and thermal sensors. While active systems emit a signal and record the amount of radiant flux scattered back towards the sensor, passive systems record the naturally occurring electromagnetic radiation, which is reflected or emitted from the surface [Jensen, 2000]. Since this thesis is focused on optically reflective remote sensing in the wavelength range from 350 to 2,500 nm, the first part of the chapter provides a detailed description of how light interacts with a single leaf, as well as with an entire canopy, in this spectral domain. The second part of the chapter then summarizes the different methods of in situ and remote sensing LAI retrieval, which have been developed in the previous decades.

2.1 Optical Properties of Plants

Plants need solar energy to carry out photosynthesis. For that reason, the interaction of plants with incident radiant energy is of the utmost importance in many fields of science (e.g., botany, ecology, agronomy) [Gates et al., 1965]. The understanding of the radiative properties of plant canopies is essential in order to interpret the reflectance signal acquired by remote sensing sensors. The obtained reflectance signal provides meaningful information on plant chemical ingredients, plant structure and, therefore, information on the current condition of plants [Jones and Vaughan, 2010]. This is especially important in agriculture, where this information is needed in order to apply an adapted treatment of fields [Moran et al., 1997]. The following subchapters first describe the radiative properties of plant leaves, and then the interactions of radiation with a plant canopy is presented in detail.

2.1.1 Properties on Leaf Level

The theory of how light interacts with a leaf was originally described by Willstätter and Stoll [1918]. Based on this work, Allen et al. [1969] and Gates et al. [1965], for example, conducted experiments that proved Willstätter and Stoll's theory [Kneubühler, 2002]. Incident radiation, which impinges on a green leaf, is either reflected, absorbed or transmitted. Figure 2.1 (left) shows the spectral reflectance curve of a green leaf simulated with the radiative transfer model (RTM) PROSPECT [Jacquemoud and Baret, 1990]. The same reflectance curve is illustrated in Figure 2.1 (right), together with the simulated transmittance and absorbance spectra. The three components are closely interrelated and correspond to the total amount of incident radiation. While the transmittance spectrum has approximately the same

shape as the reflectance spectrum, the absorbance spectrum is the opposite of the other two [Atzberger, 1998].

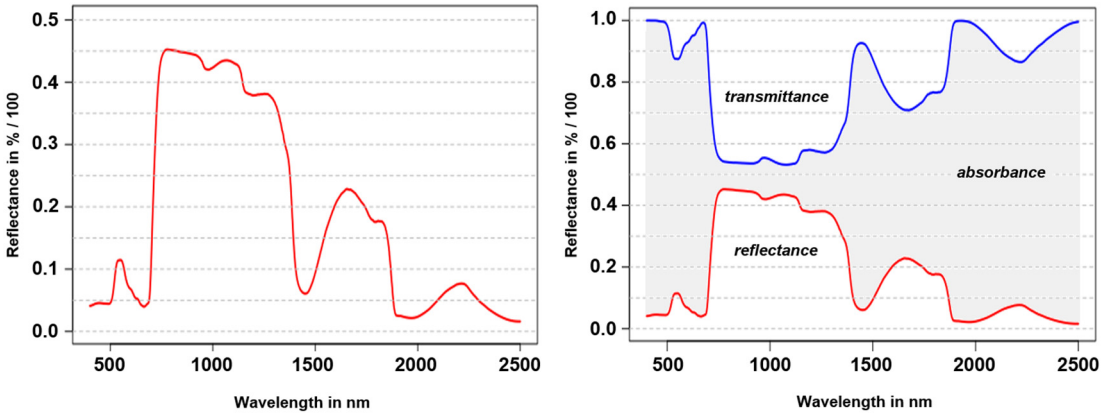


Figure 2.1: *Reflectance spectra of a green leaf simulated with PROSPECT (left); reflectance and corresponding transmittance spectrum of a green leaf simulated with PROSPECT (right).*

Leaf optical properties can be described as a function of leaf structure, concentration of biochemical ingredients and water content [Curran, 1989; Fourty et al., 1996; Gates et al., 1965; Kneubühler, 2002]. Leaf surface reflectance is mainly an interaction of diffuse reflection processes, which are caused by multiple scattering of radiation within the leaf structure [Grant, 1987]. As reported by Jones and Vaughan [2010], only

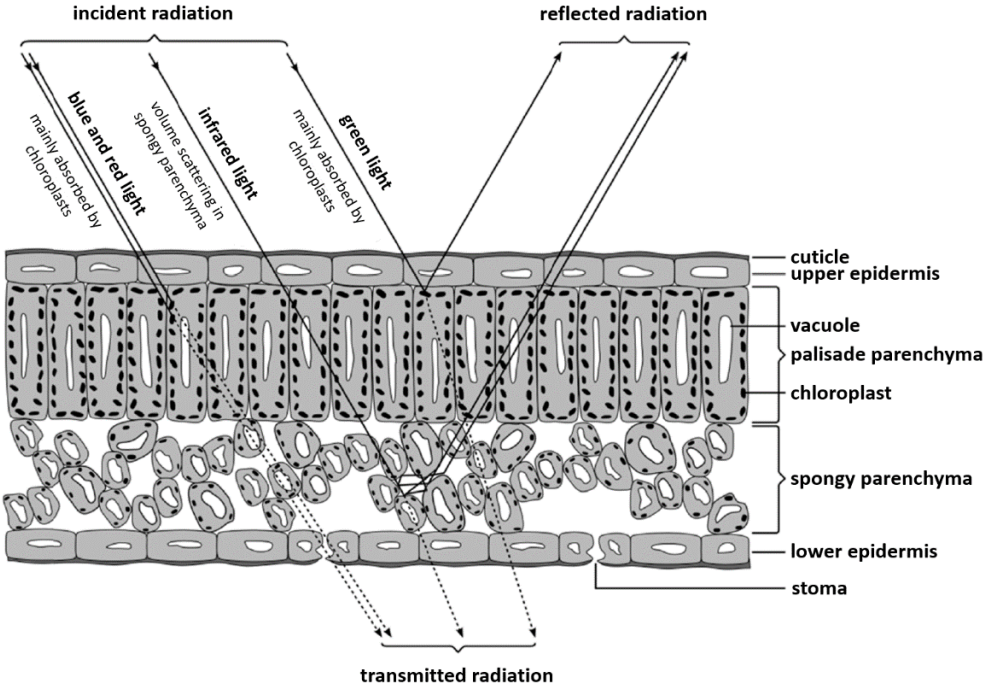


Figure 2.2: *Schematic cross-section of a leaf and the paths of radiation within the leaf for visible (VIS) and near infrared (NIR) light [Guyot,* 1990*].*

a small part of the incident radiation is reflected at the upper cuticle surface. The rest of the radiation, which penetrates the leaf surface, is either transmitted, absorbed by chlorophyll and/or other pigments, or scattered at air-water interfaces at the surface of cells. The scattering at the interfaces is called volume scattering. It is mainly influenced by the different values of refraction indices for the different components within a leaf. Figure 2.2 shows a schematic cross-section of a leaf and the paths of radiation within the leaf.

In the visible part of the spectrum (VIS: 350–700 nm), the reflectance is dominated by the light absorption of the leaf pigments. Since this spectral domain is very important for the process of photosynthesis, it is also known as the range of photosynthetically active radiation (PAR) [Alados et al., 1996]. In contrast, the near infrared (NIR: 700–1,300 nm) and the short wave infrared part of the spectrum (SWIR: 1,300–2,500) are controlled by the internal leaf structure. However, in the SWIR region of the spectrum, the shape of the curve is additionally influenced by the leaf water content, which strongly attenuates the reflectance signal [Grant, 1987].

Visible Light (350–700 nm)

The reflection of a green leaf in the visible domain is mainly dominated by the absorption of the foliar pigments, chlorophyll and carotenoids [Gitelson et al., 2002]. Figure 2.3 illustrates the absorbance of different plant pigments in the wavelength range of 350–750 nm. Chlorophyll a and b have strong absorption peaks in the blue and the red region of the electromagnetic spectrum. The absorptions of chlorophyll a (peak at 430 nm) and b (peak at 440 nm) in the blue region overlap with the absorptions of the carotenoids. In the red region of the spectrum, only chlorophyll a (peak at 660 nm) and b (peak at 640 nm) absorb incident radiation [Curran, 1989; Sims and Gamon, 2002]. Although chlorophyll a and b are the most important plant pigments in order to perform photosynthesis, carotenoids consisting of carotenes and xanthophylls also play an important role. They contribute energy converted from absorbed radiation to photosynthesis and protect the reaction centre in case of an energy oversupply [Blackburn, 2007]. Since the chlorophyll concentration of green leaves is, on average, ten times higher than that of carotenoids, the effect of carotenoids on the reflectance spectrum is masked by that of the chlorophylls [Verdebout et al., 1994].

From 670 to 780 nm, the reflectance distinctly increases because of leaf internal scattering processes and the decrease of chlorophyll absorption [Curran et al., 1991]. This region is called the red edge. According to Collins [1978] and Horler et al. [1983], the shift of the red edge inflection point (REIP), which is the point of maximum slope in the red-infrared transition area, is related to the chlorophyll concentration of leaves from different species (e.g., winter wheat and sorghum).

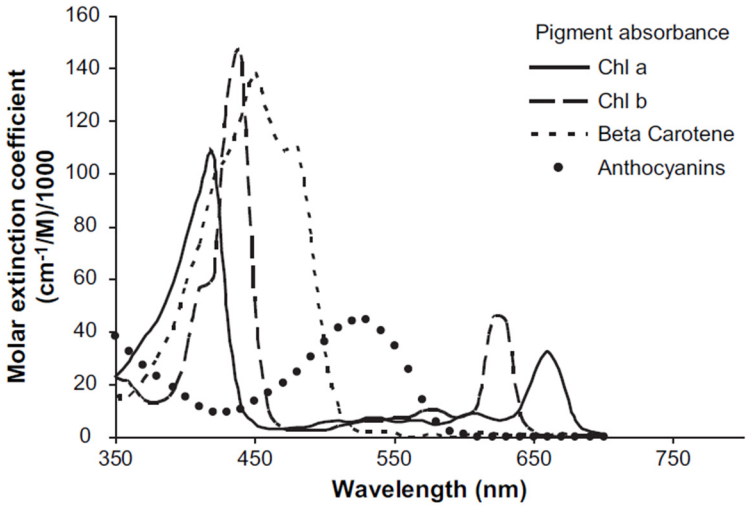


Figure 2.3: *Absorption spectra of the major plant pigments [Blackburn, 2007].*

Near Infrared (700–1,300)

The near infrared portion of a leaf spectrum is also known as the infrared plateau because of the high reflectance and transmittance level. In this spectral region, incident radiation penetrates deeper into the leaf and, therefore, provides more information about the leaf structure [Verdebout et al., 1994]. The reflectance increases with the increase of intercellular air spaces because of scattering effects caused by the transition of light from hydrated cell walls to intercellular air spaces with different refractive indices [Gausman and Allen, 1973]. The NIR region is also characterized by minor water absorption features located near 960 and 1,200 nm, which vary significantly in shape and depth, and may be related to both leaf cellular arrangement and hydration state [Wessman, 1990]. In the phenological development, the NIR reflectance level of a leaf is almost constant until the growth stage of maturation is reached. When leaves begin to senesce, cell water content decreases and cell geometry changes. Consequently, the NIR reflectance greatly increases. The same effect can be observed under drought stress conditions [Sinclair et al., 1971; Verdebout et al., 1994].

Short Wave Infrared (1,300–2,500)

The SWIR domain of leaf reflectance is mainly influenced by the absorption of water and other foliar biochemical components. The dominant water absorption features are located around 1,450, 1,900 and 2,500 nm [Verdebout et al., 1994]. Allen [1969] introduced the term equivalent water thickness (EWT) using the plate model to measure the influence of leaf water content and intercellular air spaces on leaf reflectance and transmittance [Kneubühler, 2002; Seelig et al., 2008]. This

concept has been used as the basis for the most popular leaf radiative transfer model PROSPECT, which was developed by Jacquemoud and Baret [1990].

A decrease in leaf water content leads to a distinct increase in SWIR reflectance. Thus, small absorption features, caused by other leaf components (e.g., cellulose, lignin, starch), become visible, having previously been masked by water. A detection of these absorption features with fresh leaves is difficult because water represents 60–90 % of the fresh weight of leaves [Verdebout et al., 1994]. A detailed overview of absorption features related to leaf chemical ingredients was presented by Curran [1989].

2.1.2 Properties on Canopy Level

To understand the complex process of canopy radiative transfer, it is insufficient simply to apply the reflectance properties of a single leaf on canopy scale. Rather, canopy reflectance is a composition of the optical properties of different plant components (e.g., leaf, stem), distribution of plants within the canopy, soil background, illumination conditions and viewing geometry [Hurcom et al., 1996]. According to Asner [1998] and Kneubühler [2002], the canopy reflectance (ρ_{canopy}) can be described as follows:

$$\rho_{canopy} = f\,(geometry,\ plant\ biochemistry,\ soil\ reflectance,\ structure) \qquad [2.1]$$

The parameter *geometry* describes the relation between the observer viewing direction (observer zenith and azimuth angle) and solar illumination conditions (sun zenith and azimuth angle) (Figure 2.4). The specific reflectance and transmittance signals of different plant components are partly a result of their chemical ingredients and forms

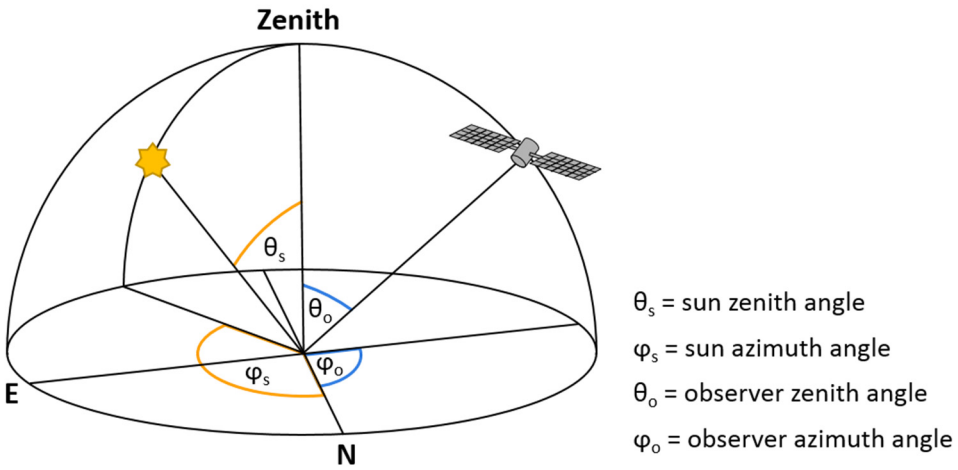


Figure 2.4: *Geometry of solar irradiance and observer viewing direction.*

the second parameter *biochemistry* (cf. Chapter 2.1.1). The third parameter *soil reflectance* is determined by the mineral and moisture properties of the soil underneath the canopy. Especially in canopies with low vegetation cover, soil reflectance has a considerable influence on canopy reflectance. Figure 2.5 shows different reflectance spectra simulated with the canopy RTM PROSAIL [Jacquemoud et al., 1995]. For all simulated spectra, the same input parameter combinations were used; only the LAI was changed stepwise from zero (bare soil) to seven. It can clearly be seen that the soil reflectance has a major impact on the reflectance spectrum of canopies with very low LAI (LAI < 1). The *structure*, as the last parameter, can be described as a composition of different sub-parameters. This includes LAI, horizontal distribution of the plants, leaf angle distribution and plant height.

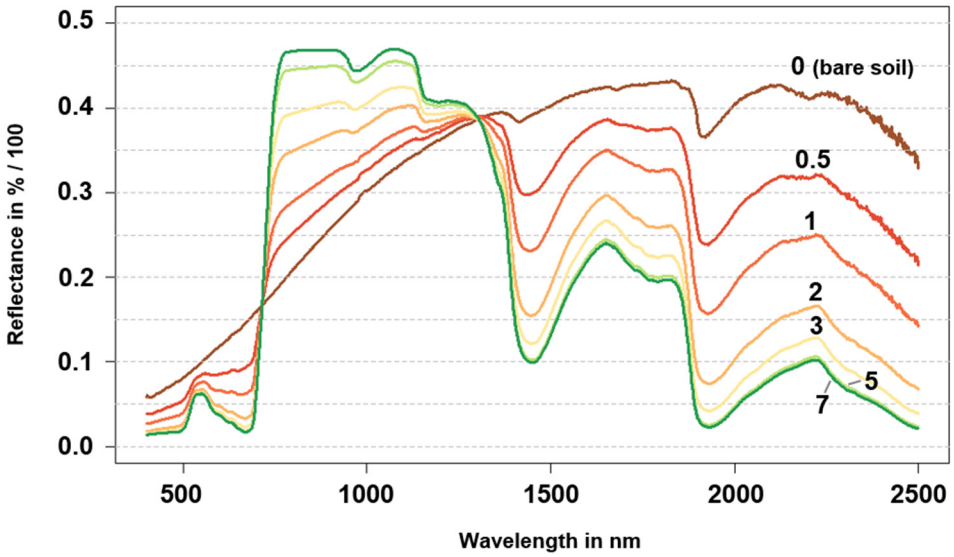


Figure 2.5: *Canopy reflectance spectra simulated with PROSAIL. All the spectral curves were simulated with the same input parameter combinations, except LAI, which was varied stepwise from zero (bare soil) to seven.*

The LAI, which is one of the most important structural parameters, was first defined by Watson [1947] as the ratio of the total one-sided leaf surface area per unit soil surface area (m^2 leaf area/m^2 soil area). Since several horizontal leaf layers can cover a defined soil area, LAI values greater than one are possible. Figure 2.6 schematically illustrates a soil area covered with one fully closed leaf layer (LAI = 1) and three fully closed leaf layers (LAI = 3), respectively. Together with the leaf angle distribution, the LAI strongly controls the relationship between leaf and canopy reflectance [Asner, 1998]. Thus, the LAI acts as a key variable to draw conclusions about leaf biochemical properties from canopy reflectance [Verdebout et al., 1994] and is, therefore, an important parameter in ecological and agricultural modelling [Broge and

Mortensen, 2002; Moran et al., 1995]. As reported by Asner [1998], a changing LAI has the greatest impact on the NIR and the least on the VIS domain of canopy reflectance. This finding is consistent with the simulated canopy reflectance spectra presented in Figure 2.5. While distinct differences in the reflectance signal of canopies with low LAI (LAI < 3) can be detected in the VIS, NIR and SWIR region, only the NIR region is additionally sensitive to small LAI changes in canopies with higher LAI (LAI ≥ 3). For that reason, the NIR range of canopy reflectance is best suited to derive information about leaf biochemical properties. However, it is often very difficult to distinguish between biochemical and structural effects in the NIR domain. One example in this regard is the widening and deepening of the chlorophyll absorption feature in the red region of the spectrum, which is caused by either a higher LAI or the increase in chlorophyll content. Both lead to a shift in the red edge inflection point, which causes a change of the NIR reflectance [Horler et al., 1983; Kneubühler, 2002].

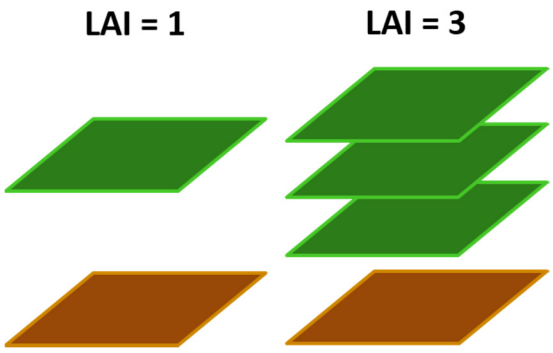


Figure 2.6: *Schematic representation of a defined soil area completely covered with one leaf layer (left) and three leaf layers (right) (based on Stovall [2012].*

In addition to the parameters mentioned in Equation 2.1, surface roughness, shading and gaps must be considered as important factors influencing canopy reflectance [Asam, 2014; Roujean and Lacaze, 2002]. Consequently, a vegetation canopy cannot be regarded as a Lambertian surface, which reflects incident radiation uniformly in all directions. On the contrary, a canopy is an anisotropic surface whose reflectance properties vary strongly with wavelength, illumination angle and viewing angle. The bidirectional reflectance distribution function (BRDF) provides a mathematical description of canopy reflectance for every wavelength when taking into account the current illumination condition and viewing angles [Lillesand et al., 2004]. The BRDF is defined under the assumption of infinitesimal solid angles for the solar incident irradiance from one direction (θ_s, φ_s) and the reflected radiance in one direction (θ_o, φ_o). However, it cannot be determined under realistic conditions in the field because irradiance does not consist of a single direction and remote sensing sensors do

not integrate over infinitesimal but large solid viewing angles [Asam, 2014; Schaepman-Strub et al., 2006]. For that reason, the BRDF can be simplified to the bidirectional reflectance factor (BRF). The BRF is the ratio of the radiance reflected from a surface in a specific direction and the radiance reflected from a lossless Lambertian reference (reflectance standard of known reflectivity) measured under identical illumination geometry [Sandmeier and Itten, 1999].

2.2 Retrieval Methods for LAI Assessment

Since the LAI of wheat is the focus of this thesis, the various methods for measuring this parameter will be described in more detail. These methods differ in the type of measurement methodology and the requirements of the technical equipment [Bréda, 2003; Zheng and Moskal, 2009]. Different methods exist to measure the LAI directly or indirectly in the field. Alternatively, plant leaves can be harvested and the LAI determination is made under laboratory conditions afterwards. Furthermore, remote sensing data acquired from proximal, airborne or spaceborne sensors offer the possibility for a spatial LAI assessment [Gray and Song, 2012; Lee et al., 2004]. The following sections first provide an overview of the various methods of in situ LAI measurements, after which different approaches for deriving the LAI from remote sensing data are presented.

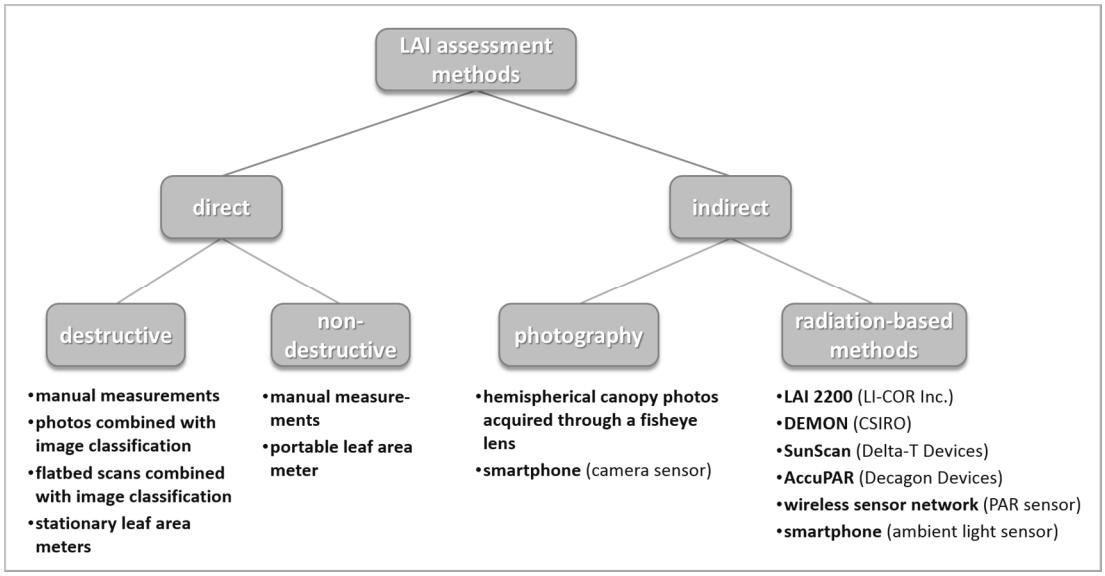


Figure 2.7: *Overview of in situ LAI assessment methods (based on Bréda [2003] and Jonckheere et al. [2004]).*

2.2.1 In Situ Assessment of LAI[3]

In situ methods for measuring the LAI can be distinguished in direct or indirect techniques. The direct methods are further subdivided in destructive (harvesting) and non-destructive (non-harvesting) methods. While direct methods enable LAI determination in a direct way, indirect assessment is based on the derivation of LAI from more easily measurable parameters [Gower et al., 1999; Jonckheere et al., 2004]. The different in situ LAI assessment methods are summarized in Figure 2.7.

Direct LAI Assessment

For the direct assessment of leaf area, the investigated plants need to be physically touched in some manner [Asam, 2014]. On the one hand, direct measurements can be made with portable leaf area meters in the field (e.g., LI-3000C (LI-COR Inc., USA), CI-201 (Delta-T Devices Ltd., UK)). This has the advantage that the leaves do not have to be harvested and thus, a monitoring of LAI changes over longer periods is possible. On the other hand, several direct measuring methods have been developed in recent years, which require a destructive sampling of leaves. In this context, one method is to fix harvested leaves on a flat surface and measure their area manually or take photos of known sizes and derive the leaf area using image classification techniques. Moreover, different stationary leaf area meters are available (e.g., LI-3100 (LI-COR Inc., USA), CI-251 (Delta-T Devices Ltd., UK)), which enable LAI assessment of a large number of harvested leaves in laboratory after a field campaign. Such devices consist of a fluorescent light source and a transparent conveyer belt with constant speed on which the leaves are placed. A well-calibrated scanning camera, mounted below the moving conveyor belt, acquires image data of the leaves and a small processing unit determines their leaf area [Jonckheere et al., 2004]. The LI-3100 stationary leaf area meter, for example, which is depicted in Figure 2.8, falls into this category of devices.

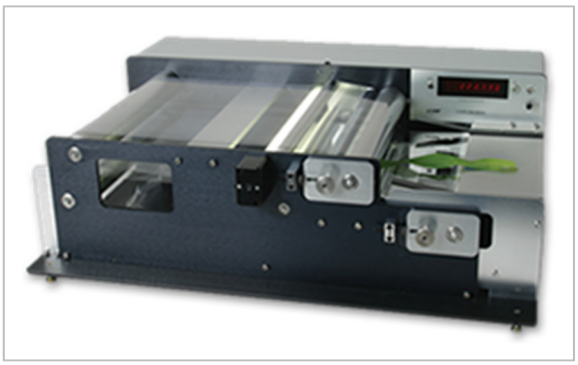


Figure 2.8: *LI-COR LI-3100 stationary leaf area meter [LI-COR Inc., 2016a].*

[3] Parts of this section have previously been published by Bauer et al. [2014].

Since only direct methods are directly related to foliage and therefore allow the most precise LAI determination, they were used as references for the calibration and evaluation of indirect methods in numerous studies [e.g., Wilhelm et al., 2000]. Moreover, direct measurements provide additional information on the shape, size and number of leaves [Bréda, 2003].

Indirect LAI Assessment

Although direct LAI assessment usually provides more precise results, data acquisition is time-consuming, expensive and, therefore, often limited to small areas [Bréda, 2003; Jonckheere et al., 2004]. For that reason, several optical indirect methods have been developed in recent years, which require no physical contact with the leaves. Using these methods, the LAI is derived from measurements of light transmission through the canopy [Jonckheere et al., 2004].

Hemispherical photography belongs to the category of indirect measurement methods (Figure 2.7). A fisheye lens mounted on an upward-pointing digital camera is used to acquire photos from beneath the canopy. The subsequent analysis of the photos with appropriate image processing software enables the assessment of gap size distribution (exploiting the contrast between leaves and sky) and the determination of the LAI [Jonckheere et al., 2004]. Although hemispherical photography can be used for all types of canopies [Rich, 1990], this technique is mainly applied in forestry, where the assessment of incoming radiation above the canopy is difficult to measure [Welles, 1990].

In agricultural studies, which are mainly focused on low and regular canopies, the LAI is often calculated with devices that simply compare measurements of incoming light above and below the canopy [Jonckheere et al., 2004]. Four commercial devices are available, which measure the fraction of transmitted radiation that passes through a plant canopy. They all work on the same principle. The smaller the amount of radiation transmitted by the leaves, the higher is the LAI, and vice versa. In this context, Figure 2.9 illustrates the effect of multiple leaf layers on canopy reflectance and transmittance. Since the optical properties of each leaf are different, the portion of reflected and transmitted radiation are described by the factors a to z in the figure.

The LAI-2000 and its successor, the LAI-2200 (LI-COR Inc., USA), as well as the DEMON (CSIRO, Australia), determine the LAI from gap fraction measurements made for different incident angles in the blue domain of VIS. Although measurements collected with the SunScan (Delta-T Devices Ltd., UK) and the AccuPAR (Decagon Devices, USA) are based on gap fraction analysis as well, these devices use the entire PAR range for LAI determination. According to their measurement principle, the four devices also belong to the group of indirect measurement methods (Figure 2.7). Additionally, several studies in recent years have been conducted using low-cost

sensor nodes [e.g., Bauer et al., 2016a; Bauer et al., 2014; Qu and Sun, 2010] or smartphones [Bauer et al., 2016b] equipped with ambient light sensors, which also facilitate robust measurements of gap fraction and LAI. Given that, in this thesis, the in situ LAI measurements of wheat were collected with an LAI-2000 and a SunScan device, the function and the design of both instruments will be explained in the subsequent section.

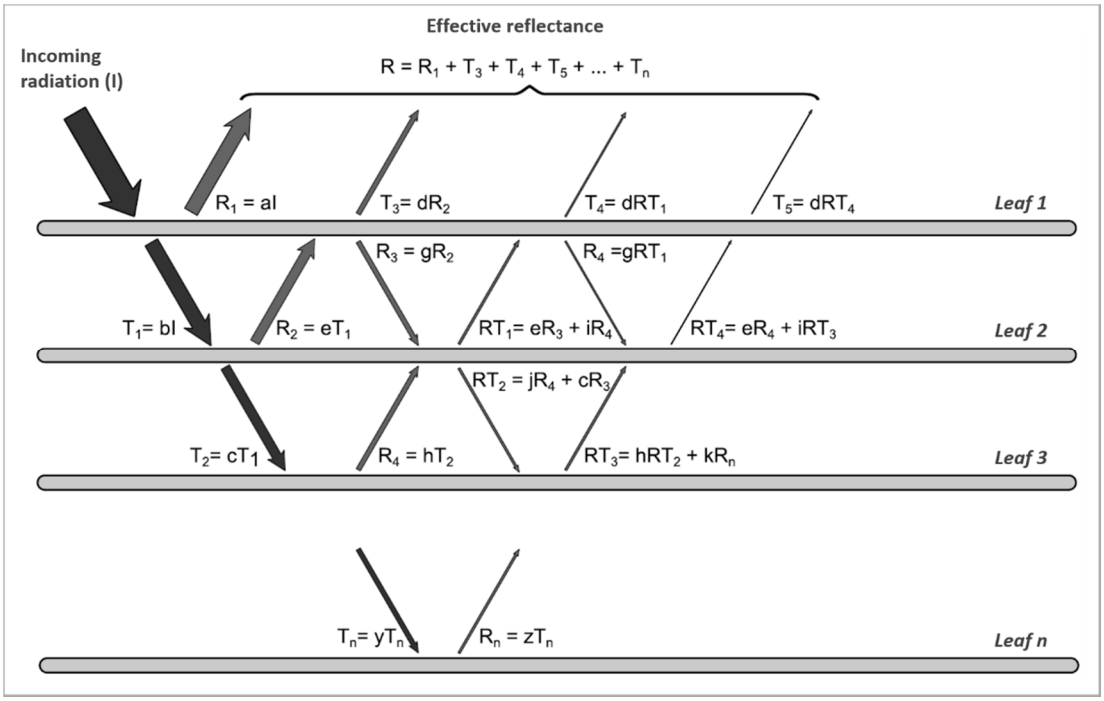


Figure 2.9: *Effect of multiple-leaf layers on canopy reflectance and transmittance (modified according to Spengler [2013] and Hoffer [1978]). (I=incoming radiation, R=reflected radiation, T=transmitted radiation, a…z=portion of reflected and transmitted radiation)*

The typical behaviour of light transmittance of plants in the visible spectrum is leveraged by both devices for indirect LAI determination. The Beer-Lambert law establishes the theoretical basis for deriving LAI from the quantitative interaction between solar radiation and the plant canopy [Jonckheere et al., 2004]. According to this law, there is a logarithmic dependency between the LAI and the solar radiance that is partially absorbed by the canopy, which can be described by the following equation:

$$B = Ae^{-CLAI}. \qquad [2.2]$$

Here, A is the light intensity observed above the canopy and B is the corresponding sensor reading below the canopy. Furthermore, the constant term C is the light extinction coefficient, which is given by the quantity of the specific light absorption

property of leaves from different plant species. In addition, the leaf orientation angle and solar altitude have a distinct impact on this coefficient. Thus, it is both cultivar- and site-specific [Jonckheere et al., 2004; Qu and Sun, 2010]. By exponential transformation of Equation 2.2, the Monsi-Saeki model [Monsi and Saeki, 2005] applies the Beer-Lambert law for LAI estimation. This results in:

$$LAI = -\frac{1}{C}\, log + \left(\frac{B}{A}\right), \qquad [2.3]$$

where $\frac{B}{A}$ is the transmittance used by the indirect methodology and gathered by the LAI-2000 and the SunScan device. Since these devices often do not take the scattered radiation from leaf surfaces into account, it is recommended that the light intensity is measured under fully diffused sky conditions [Kobayashi et al., 2013]. An inherent weakness in LAI estimation, based on this approach, is that it does not distinguish photosynthetically active leaves from other plant elements such as branches and stems [Jonckheere et al., 2004]. Therefore, the term effective LAI (LAI_{eff}) is occasionally used to describe the measured values [Chen and Black, 1992; Garrigues et al., 2008]. Since this effect is less pronounced in homogenous plant canopies, such as wheat [Richter et al., 2011], only the term LAI is used in the rest of this thesis.

The LAI-2000 consists of a measurement wand, which is attached to a control unit. At the end of the measurement wand, there is a fisheye optical sensor (Figure 2.10 (left)). The sensor's field of view is divided into five rings with different centre angles. To ensure accurate LAI measurements, an optical filter is installed, which rejects any incoming radiation above 490 nm. Applying the filter results in a distinct improvement in the contrast between sky and leaf material because, in the blue range of VIS green leaves have a very low reflectance, such that they are detected as black bodies, which results in a more accurate LAI assessment [Garrigues et al., 2008; LI-COR Inc., 2016b].

The SunScan device enables the measurement of incident (above the canopy) and transmitted PAR (below the canopy). It consists of a probe for below canopy readings equipped with 64 equidistant calibrated photodiodes (Figure 2.10 (right)). The measurements of the photodiodes can be stored as single values or as an averaged reading. It is also possible to record the incident PAR above the canopy with the probe. In this measuring mode, the SunScan works in a similar way to the LAI-2000. Alternatively, above canopy readings can be stored with a beam fraction sensor, which is connected to the probe. This allows the simultaneous assessment of below canopy readings with the probe, as well as above canopy measurements of direct and diffuse incident radiation [Bréda, 2003].

Figure 2.10: *LAI-2000 (left) [LI-COR Inc., 2016c] and SunScan (right) [Delta-T Devices Ltd., 2016].*

2.2.2 Remote Sensing Assessment of LAI

The estimation of biochemical and -physical (e.g., LAI) parameters from remote sensing data is a very important topic in modern agriculture. For that reason, in the last 45 years, numerous methods have been developed or adapted from other fields of science, in order to derive information about the current condition of plants from the spectral reflectance acquired by proximal and remote sensing sensors. These methods can be divided into two main groups: empirical-statistical regression and physically based methods [Dorigo et al., 2007].

Empirical-Statistical Regression Methods

Empirical-statistical regression methods are used to relate in situ measured leaf or canopy parameters to remote sensing data [Cohen et al., 2003; Dorigo et al., 2007]. They can be distinguished in parametric and non-parametric approaches. Figure 2.13 provides an overview about the different categories of empirical-statistical regression methods used in the field of remote sensing.

Parametric regression methods describe the relationship between a plant parameter and a single spectral observation [Verrelst et al., 2015]. Most of the parametric approaches are based on broadband VIs, which can be further subdivided into ratio and orthogonal indices [Broge and Mortensen, 2002]. The most famous ratio index is the normalized difference vegetation index (NDVI) [Rouse et al., 1974], which has been used in several studies to predict the LAI [e.g., Broge and Mortensen, 2002; Elvidge and Chen, 1995]. In contrast to ratio indices, orthogonal indices, such as the perpendicular vegetation index (PVI) [Richardson et al., 1982] or the soil-adjusted vegetation index (SAVI) [Huete, 1988], also consider soil background effects. As such, they have also been widely applied in the past to derive LAI and other plant para-

meters from remote sensing data [e.g., Darvishzadeh et al., 2008a; Wu et al., 2007]. With the increasing availability of hyperspectral remote sensing data in recent years, several narrow band vegetation indices have been explored using the information contained in narrow absorption features [Dorigo et al., 2007]. Along with ratio indices, such as the chlorophyll absorption ratio index (CARI) and the water index (WI), which were mainly developed to retrieve chlorophyll [Haboudane et al., 2002] and water content [Peñuelas et al., 1997], new categories of indices based on spectral shape (e.g., derivative green vegetation index (DGVI)), REIP and spectral continuum measures (e.g., continuum removed chlorophyll well depth (CRCWD)) have been introduced for the assessment of LAI [e.g., Broge and Leblanc, 2001; Clevers et al., 2002; Elvidge and Chen, 1995].

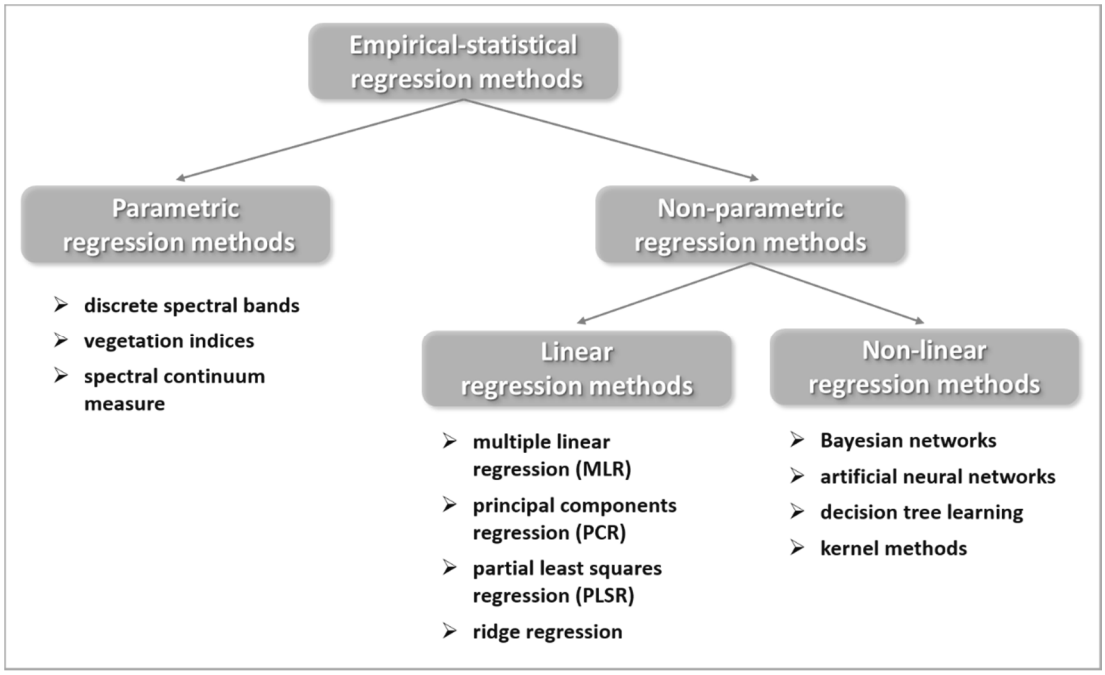


Figure 2.11: *Overview of the different empirical-statistical regression methods (modified according to Verreslt [2015]).*

VIs are often sensitive to more than one plant parameter, meaning that it is difficult to identify which parameter change causes a change in the VI [Dorigo et al., 2007; Govaerts et al., 1999]. Furthermore, the relationship between numerous VIs and the LAI is non-linear, and especially for wheat crops, the problem of saturation from moderate to high values of LAI (LAI > 3) has been observed in different studies [Nguy-Robertson et al., 2014; Serrano et al., 2000]. This effect is illustrated in Figure 2.12, which shows the relationship between the LAI of wheat crops and the two indices, NDVI and SAVI.

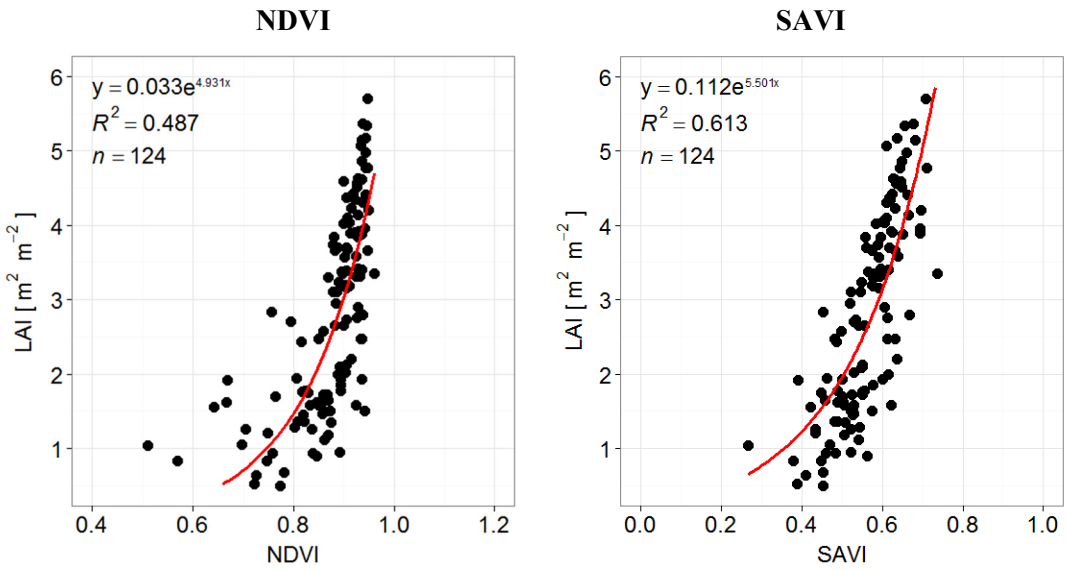


Figure 2.12: *Scatter plots of measured LAI vs. calculated NDVI (left) and SAVI (right) from field spectral measurements of 124 wheat plots (unpublished data).*

In the case of using VIs to predict plant parameters, linear regression analysis is normally applied to model the relationship between the dependent variable (e.g., LAI) and the independent variable (VI). In general, however, VIs are based on only two or three wavelengths, while the rest of the spectral data space provided by multi- or hyperspectral sensors remains unconsidered. For that reason, non-parametric regression methods can be used because they allow for the inclusion of numerous spectral bands to predict a plant parameter. Approaches belonging to this group make use of a calibration and validation phase for model optimization [Dorigo et al., 2007; Verrelst et al., 2015]. Figure 2.13 schematically depicts the basic concept of this type of regression methods. In Figure 2.11 it is also shown that non-parametric regression approaches are further subdivided into linear and non-linear methods.

Multiple linear regression (MLR) can be regarded as the simplest non-parametric linear regression method. Unfortunately, in MLR, the independent variables (spectral bands) are assumed to be uncorrelated. This is normally not the case and can result in the problem of regression model overfitting [Dorigo et al., 2007; Yoder and Pettigrew-Crosby, 1995].

In this context, principle components analysis (PCA) is an appropriate linear method reducing the high dimensional data space into a small set of uncorrelated components, which contain most of the spectral information [Cohen et al., 2003]. The uncorrelated principle components can be used to substitute the often highly correlated independent variables X (spectral bands) in MLR analysis. This method is called principle components regression (PCR) and is able to handle the problem of multicollinearity [Geladi and Kowalski, 1986]. However, PCR has two major

disadvantages. Firstly, as the generated components are dependent on the dataset, they are not transferable to other image data. Secondly, the coefficients (scores) of the PCA components are generated without regard for the relationship between the components and the dependent variable [Cohen et al., 2003]. Although MLR and PCR have some weaknesses, they have been used in several studies to retrieve plant parameters from remote sensing data [e.g., Atzberger et al., 2010; Darvishzadeh et al., 2008b].

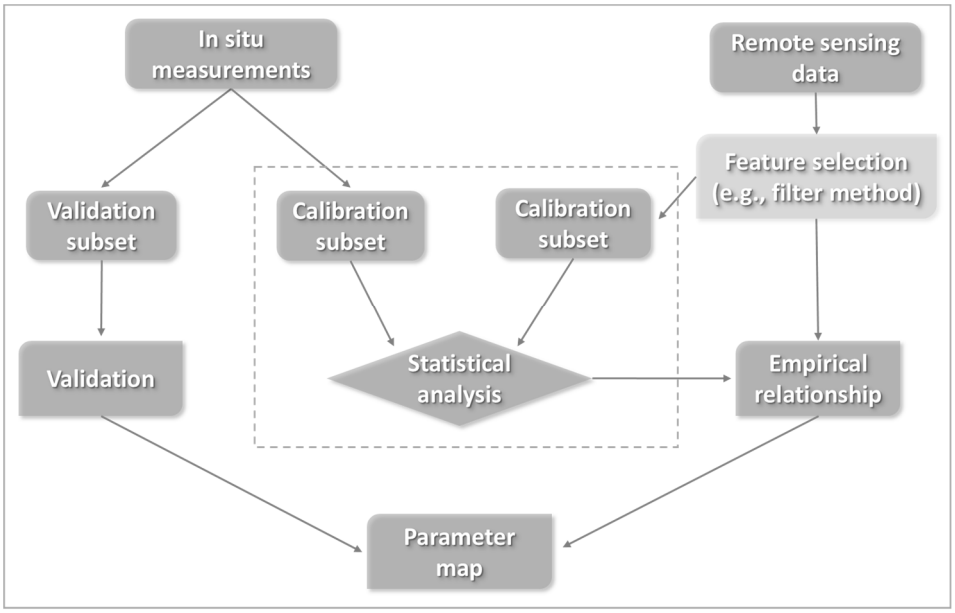


Figure 2.13: *Schematic concept of the empirical-statistical approach. Feature selection is an optional step (modified according to Asam [2014] and Dorigo [2007]).*

Taking into account these weaknesses, PLSR was developed, which can be regarded as an extension of MLR. In contrast to MLR, however, PLSR is less restrictive because model building is possible with fewer dependent (e.g., observations of plant parameters) than independent variables (spectral bands) [Dorigo et al., 2007]. PLSR and PCR are based on a similar function principle. However, while PCR only uses the PCA scores of the independent variables (spectral bands), PLSR takes into account the scores of both the independent (spectral bands) and the dependent variable(s) (e.g., plant parameter(s)) by relating them with each other [Verrelst et al., 2015]. Since PLSR is one of the non-parametric linear empirical-statistical regression approaches applied in this thesis, a detailed overview of this method is provided in Chapter 6.1.1.

Along with non-parametric linear regression models, several non-parametric non-linear regression techniques have been developed and adapted to remote sensing images in order to retrieve plant parameters. These methods are based on non-linear transformation procedures and have the advantage that they can handle non-linear relationships of image features without knowledge about the underlying data

distribution [Verrelst et al., 2015]. Non-linear regression methods are also known as machine learning techniques. According to Verrelst et al. [2015], they can be subdivided into four categories: Bayesian networks, artificial neural networks, decision tree learning and kernel methods.

Regression algorithms from all four groups have been used to predict the LAI from remote sensing images in recent years. Kalacska et al. [2005] and Mustafa et al. [2012], for example, successfully applied Bayesian networks to estimate the LAI from Landsat 7 and MODIS data. Moreover, artificial neural networks, and their suitability for the retrieval of LAI from multi- and hyperspectral data, have been investigated by Verrelst et al. [2012] and Jensen et al. [2012]. RFR is the main representative of decision tree learning methods for regression applications. This technique is also applied in this thesis in order to estimate the LAI of winter wheat. Therefore, a detailed description of the functionality of RFR is presented in Chapter 6.1.3. Kernel-based methods form the last group of machine learning regression techniques. This group includes algorithms using kernel functions to quantify similarities between input samples of a data set [Shawe-Taylor and Cristianini, 2004; Verrelst et al., 2015]. As SVR is the most widely known method in this group, it has been chosen to predict the LAI from spectral reflectance data in this thesis. Chapter 6.1.2 provides a detailed description of SVR.

In general, empirical-statistical regression methods are widely used in remote sensing to retrieve plant parameters because these techniques are simple to apply and computationally inexpensive in most cases [Verrelst et al., 2015]. However, the estimation of LAI based on VIs, mainly using the red and NIR spectral range, has often led to poor results (cf. Figure 2.12) because of saturation effects in canopies with higher vegetation density [Hansen and Schjoerring, 2003]. In contrast, non-parametric empirical-statistical regression techniques explore the complete spectral information, such that they are also referred to as full-spectrum methods [Verrelst et al., 2015].

Although regression algorithms can deal with the high dimensionality of the hyperspectral data, many studies have reported that a preselection of features (spectral bands) can improve the regression model quality and prediction capability [Renzullo et al., 2006; Viscarra-Rossel and Lark, 2009]. The training procedure of some non-parametric empirical-statistical regression algorithms is already based on feature selection techniques (e.g., RFR [Breiman, 2001], MARS [Friedman, 1991]). These algorithms are known as embedded methods [Hilario and Kalousis, 2008]. Additionally, wrapper methods can be applied to identify the most important features with the highest explained variance regarding a target variable (e.g., LAI) [Guyot et al., 1992; Saeys et al., 2007]. Wrapper methods, such as recursive feature elimination [Guyon et al., 2002] or genetic algorithm [Holland, 1992], conduct the search for the most appropriate variables by a repeated calibration and validation of a previously selected regression algorithm (e.g., SVR) with varying subsets of all

features. In the end, the subset that is chosen is the one that provided the highest regression model quality [Kuhn and Johnson, 2013]. In contrast to embedded and wrapper methods, feature selection based on filter methods is completely decoupled from the process of regression model building. As shown in Figure 2.13, a filter method can be optionally applied prior to regression model calibration to select important spectral bands. In this thesis, the RReliefF filter method [Robnik-Sikonja and Kononenko, 1997, 2003] was used prior to regression analysis in order to identify the most important spectral bands for the estimation of LAI. The theoretical background of this method is given in Chapter 6.2.1.

Alongside the numerous benefits that empirical-statistical regression methods provide for the retrieval of LAI and other plant parameters, there are also some drawbacks. To establish an empirical-statistical relationship between remote sensing data and a plant parameter of interest, it is mandatory to collect field measurements of the investigated parameter for the purpose of model calibration and validation. Furthermore, the parameter of interest should be measured for different stages of plant development in order to ensure the coverage of a wide range of possible values. This is important for the regression model quality and prediction capability [Sims and Gamon, 2002]. However, field measurement campaigns are very time-consuming and expensive, such that in situ data sampling is often limited to small areas [Bréda, 2003]. This often leads to the calibration of regression models, which are only representative for the investigated area at a certain time. Therefore, the transferability of locally developed regression models was often criticized for their poor results when applied to data acquired under varying canopy (e.g., phenological stage, canopy geometry) and external (e.g., local soil properties, atmospheric conditions, sun and view geometry) conditions [Baret and Guyot, 1991; Dorigo et al., 2007]. To minimize these effects, several VIs and regression algorithms use transformation techniques. Nevertheless, a robust evaluation of an empirical-statistical relationship is mandatory to test the predictive performance of developed models over independent data sets [Dorigo et al., 2007]. In this context, a recent study of Verrelst et al. [2013] demonstrated the reliable transfer of a locally calibrated non-linear regression model to other regions (independent data sets), which led to promising LAI prediction results. Nonetheless, the authors of the study also pointed out that this field of science needs to be further investigated in the future.

Physically Based Methods

Beside empirical-statistical regression methods, physically based RTMs form the second group of techniques used in the retrieval of plant parameters from remote sensing data. They enable to calculate top-of-canopy reflectance based on physical laws by simulating the interactions between solar radiation and the elements forming

the canopy [Dorigo et al., 2007]. In general, RTMs can be distinguished in models simulating leaf optical properties and models simulating canopy reflectance.

Alongside plate models, such as PROSPECT [Jacquemoud and Baret, 1990], LEAFMOD [Ganapol et al., 1998] and DLM [Stuckens et al., 2009], which enable the determination of foliage leaf reflectance and transmittance, several other models have been introduced to explain the radiative transfer at leaf level. These include LIBERTY [Dawson et al., 1998], a compact spherical particle model especially developed to simulate the reflectance of conifer needles, N-flux models, such as the 4-flux model presented by Richter and Fukshansky [1996], stochastic models, such as SLOP [Maier et al., 1999], and ray-tracing models, such as RAYTRAN [Govaerts and Verstraete, 1998].

In addition to leaf optical models, numerous approaches have been developed, which describe the radiative transfer of entire canopies. According to Goel [1988], these models can be divided into four categories: turbid medium models (e.g., SAIL [Verhoef, 1984]), geometrical models (e.g., 4-scales [Chen and Leblanc, 1997]), hybrid models (e.g., TRIM [Goel and Grier, 1988]) and computer simulation models (e.g., drat [Lewis, 1999]). A comprehensive comparison and evaluation of a large number of canopy RTMs was conducted by the RAdiation transfer Model Intercomparison (RAMI) initiative at the Joint Research Centre (JRC) of the European Union [Widlowski et al., 2013].

While geometric and hybrid models were mainly used in forest canopies to calculate the radiative transfer, turbid medium models were often applied to simulate the top-of-canopy reflectance of agricultural crops [e.g., Bacour et al., 2002]. The PROSAIL model [Jacquemoud et al., 1995], which is a combination of the PROSPECT leaf optical model and the SAIL canopy reflectance model, is the most popular model in this category and a valuable tool for the characterization of horizontally homogeneous and dense leafy canopies, such as crops stands [Asam, 2014; Schlerf and Atzberger, 2006]. Furthermore, ray-tracing methods are increasingly applied to calculate the canopy reflectance of cereals [e.g., Kuester et al., 2014]. These techniques, however, are very computationally intensive, due to the complex description of how light photons interact with simulated canopies consisting of detailed three-dimensional models of single plants.

For the retrieval of plant parameters, canopy RTMs have to be inverted. Inversion techniques can be divided into numerical optimization, lookup table (LUT) and hybrid methods [Dorigo et al., 2007; Verrelst et al., 2015]. Numerical optimization can be regarded as the classical inversion approach and has been applied in many studies [e.g., Jacquemoud et al., 1995; Zarco-Tejada et al., 2001]. In this context, a cost function is used to find the best fit between the simulated and the measured reflectance by iteratively running the canopy RTM with different sets of input variables [Dorigo et al., 2007]. The LUT approach is based on using an RTM in forward mode

to generate thousands, or even millions, of spectra for a large number of input parameter combinations. Subsequently, measured spectra are compared to those in the LUT using a cost function. The input parameters of the LUT spectrum with the closest match to the measured spectrum can be regarded as the ultimate solution in the inversion process [Dorigo et al., 2007; Verrelst et al., 2015]. The LUT approach is the most widely used inversion technique and has been applied in different studies to estimate the LAI of wheat [e.g., Richter et al., 2011; Rivera et al., 2013]. A schematic concept of the LUT approach to estimate plant parameters from remote sensing data is illustrated in Figure 2.14. Along with numerical optimization and the LUT approach, hybrid methods are increasingly used. These techniques combine RTM modelling with non-linear regression algorithms. At first, similar to the LUT approach, a RTM is used in the forward mode to build up a LUT. Subsequently, a non-linear regression model is trained with all available data from the LUT. As the last step, the trained model can be applied to measured spectral data in order to predict vegetation parameters [Verrelst et al., 2015]. Several studies have been carried out successfully using NNR [Atzberger, 2010; Vohland and Jarmer, 2008] as an inversion technique to predict the LAI. Moreover, initial results also show the potential of SVR [Durbha et al., 2007] and RFR [Doktor et al., 2014] as suitable inversion techniques.

Physically based models and their inversion enable a sound retrieval of plant parameters from remote sensing data using the entire spectrum. The functioning of the forward mode of RTMs is based on the relationships between input variables and the spectral reflectance. This physical relationship is well understood. In contrast, the process of model inversion is less transparent because the performance of the inversion for different plant parameters strongly depends on the applied cost function [Rivera et al., 2013; Verrelst et al., 2015]. Moreover, some input parameters of RTMs are very difficult to measure during field campaigns. One of the input parameters of the PROSAIL model, for example, is the average leaf inclination angle. This parameter can vary greatly within an investigated canopy, even if it consists of a single variety. However, the leaf inclination angle is important to describe the leaf angle distribution, which on the other hand strongly controls the relationship between leaf and canopy reflectance [Asner, 1998]. In general, prior knowledge of the range and distribution of the different RTM input parameters can be very useful to limit the number of possible parameter combinations in a LUT. Thus, the problem of ill-posedness, which describes that different parameter combinations can result in identical simulated reflectance spectra, can be minimized in the process of model inversion [Dorigo et al., 2007]. In summary, reflectance modelling is a suitable tool for the prediction of LAI from reflectance data. The user, however, needs a great deal of expert knowledge for RTM parameterization and optimization in order to achieve accurate results.

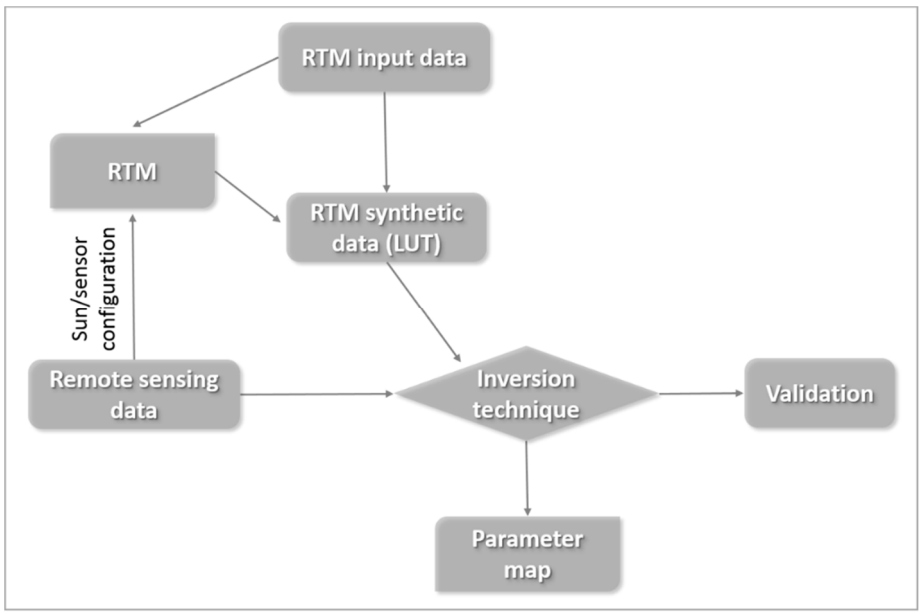


Figure 2.14: *Schematic concept of the RTM inversion using the LUT approach (adapted from Verrelst [2015]).*

3 Remote Sensing in Precision Agriculture

According to Moran [1997], precision agriculture is an agricultural management system designed to target crop and soil inputs, according to within-field requirements to optimize profitability and protect the environment. The first precision agricultural practices were developed in the USA in the 1980s [Gebbers and Adamchuk, 2010; Robert, 2002]. Since the 1990s these practices have also been adopted and commercially used in Europe, Australia, Canada, and Japan [Mulla, 2013].

The main objective of precision agriculture is to apply the right treatment to agricultural fields in the right place at the right time. For that reason, different geospatial technologies, including GIS, GNSS and remote sensing, are nowadays extensively utilized as information sources and bases for decision-making [Ge et al., 2011; Gebbers and Adamchuk, 2010]. In particular, proximal and remote sensing technologies are increasingly important spatio-temporal information sources in precision agriculture for identifying variation in soil and crop conditions on a sub-field scale. In recent years, various remote sensing sensors have been developed and applied in agriculture, which use different wavelength regions ranging from the ultraviolet to microwave portions of the electromagnetic spectrum [Mulla, 2013].

The first section of this chapter provides a short overview of the history of optically reflective remote sensing in precision agriculture. In the second section, the necessary requirements of precision agriculture on remote sensing sensors and the current limitations of earth-observing technologies will be addressed.

3.1 Contribution of Remote Sensing to Precision Agriculture

Since the early 1970s, remote sensing data were used for a wide range of applications in agriculture, such as crop yield assessment [e.g., Serrano et al., 2000], crop nutrient and water stress detection [e.g., Bastiaanssen et al., 2000], infestations of weeds [e.g., Thorp and Tian, 2004], detection of insects and plant diseases [e.g., Seelan et al., 2003], and estimation of soil properties [e.g., Stevens et al., 2010]. The first study using satellite data acquired by Landsat 1, with the aim of classifying agricultural fields, was conducted by Bauer and Cipra in 1973. Since then, numerous optical satellites have been launched, which had/have different characteristics regarding their spatial, spectral and temporal resolution [Mulla, 2013]. An overview of former and current commercial and scientific optically reflective satellite systems, which have been used for agricultural applications, is given in Table 3.1. Apart from the study of Bauer and Cipra [1973], as well as other investigations focused on classification issues, the first precision agricultural application was published by Bhatti et al. [1991], who used Landsat 5 data to identify spatial patterns in soil organic matter content. In subsequent years, multispectral Landsat 5, SPOT 1 and IRS 1A

Table 3.1: *Selection of former, current and future optically reflective remote sensing satellites used for conventional and precision agriculture (modified according to Mulla [2013] and Whelan and Taylor [2013]). (PA=precision agriculture, L=low, M=medium, H=high)*

Satellite (year)	Spatial resolution[4]	Spectral resolution[5]	Return frequency[6]	Suitability for PA
Landsat 1/2/3 (1972/75/78)	79 m	4 bands	18 d	L
Landsat 5 (1984)	30 m	6 bands	16 d	M
SPOT 1/2/3 (1986/90/93)	20 m	3 bands	2-6 d	M
IRS 1A (1988)	72 m	4 bands	22 d	M
SPOT 4 (1998)	20 m	4 bands	2–6 d	M
IKONOS (1999)	3.2 m	5 bands	3 d	H
Landsat 7(1999)	30 m	7 bands	16 d	M
EO-1 Hyperion (2000)	30 m	220 bands	16 d	H
EO-1 ALI (2000)	30 m	10 bands	16 d	M
QuickBird (2001)	2.62 m	5 bands	1–4 d	H
PROBA/CHRIS (2001)	18–36 m	19 bands	7 d	H
SPOT 5 (2002)	10–20 m	5 bands	1-4 d	M
RapidEye (2008)	6.5 m	5 bands	1 d	H
GeoEye-1 (2008)	1.6 m	6 bands	2–8 d	H
WorldView-2 (2009)	1.84 m	8 bands	1.1 d	H
Pleiadis-1A/B (2011/12)	2 m	5 bands	1 d	H
SPOT 6/7 (2012/14)	6 m	5 bands	1 d	H
Landsat 8 (2013)	30 m	9 bands	16 d	M
SkySat-1/2 (2013/14)	2 m	5 bands	2 d	H
WorldView-3 (2014)	1.24–3.7 m	17 bands	1 d	H
Sentinel-2A (2015)	10–60 m	13 bands	10 d	M
Sentinel-2B (2017)[7]	10–60 m	13 bands	10 d	M
EnMAP (2018)[7]	30 m	242 bands	27 d	H
HyspIRI (2022)[7]	30 m	214 bands	16 d	H

[4] spatial resolution (GSD) of multi-/super-/hyperspectral sensors
[5] number of bands including panchromatic band when available
[6] off-nadir revisit when possible
[7] planned launch

data were used in several studies to develop and support precision agricultural applications. Although data from these satellite systems covered large areas with spectral bands in the VIS, NIR and SWIR domain of the spectrum that enabled vegetation studies, the return frequency (except SPOT 1) and the spatial resolution (20–30 m) were only suitable for use in precision agriculture to a limited extent [Mulla, 2013].

As a consequence, satellites, such as IKONOS and QuickBird, were developed and launched in 1999 and 2001, respectively, which better met the demands of precision agriculture in terms of spatial (GSD < 5 m) and temporal resolution (< 5 days). Image data of both satellites were extensively used, for example, to map crop yield variability on a sub-field scale [Enclona et al., 2004; Yang et al., 2006]. The launch of the five-satellite constellation known as RapidEye in 2008 was another milestone in satellite-based precision agriculture. RapidEye contains five spectral bands (including a novel red edge band) and has a spatial resolution of 6.5 m. Given the five identical satellites involved, it is possible to record data of any location on the globe on a daily basis. Therefore, RapidEye images are very well suited to investigations that require high spatio-temporal resolution, such as the monitoring of important plant parameters in agriculture with the aim of optimizing crop production [Kross et al., 2015].

The latest generation of spatial very high-resolution multispectral satellites has GSDs of 2 m (Pleiadis-1A/B, SkySat-1/2) or even higher (GeoEye-1, WorldView-2, WorldView-3) and revisit cycles of one to two days. These systems provide a high potential for precision agricultural applications, such as site-specific irrigation or fertilization [e.g., de la Fuente et al., 2013]. However, these satellites are all operated by private companies, meaning that the image data acquisition can be very expensive, which is especially the case for the monitoring of large areas or multitemporal analysis in precision agriculture, where several scenes are required. For that reason, moderate resolution satellite images (e.g., Landsat 8) are still frequently used for agricultural applications because they are less expensive or even freely accessible [Johnson, 2014]. However, due to its spatial resolution of 30 m, Landsat 8 has only a limited value to precision agriculture. In contrast, the recently launched Sentinel-2 system provides images with a finer resolution (four spectral bands with 10 m and six bands with 20 m GSD). Moreover, Sentinel-2A data are free of charge and the recorded images cover large areas; meanwhile, in several preliminary studies based on simulated image data, the potential of Sentinel-2 for precision agricultural applications was demonstrated [e.g., Richter et al., 2011; Verrelst et al., 2012]. With the planned launch of Sentinel-2B (identical to Sentinel-2A) in 2017, image data of the same location on the globe will be available every five days. Hence, Sentinel-2 data will also have great potential for multitemporal analysis in the future.

In addition to multispectral satellite remote sensing, hyperspectral image data have been used for a wide range of applications in precision agriculture. The majority of

previous studies are mostly based on hyperspectral image data acquired by airborne sensors. The first sensor of this category was AVIRIS in 1987, which was built by the Jet Propulsion Laboratory of NASA. In subsequent years, the sensor was further developed and AVIRIS image data were applied in different studies that were focussed on the estimation of plant parameters from crops [e.g., Jacquemoud et al., 1995; Lee et al., 2004]. Until today, several comparable sensors, such as HyMap (HyVista Corp., Australia), CASI (ITRES Research Ltd., Canada), aisaDUAL (Specim Ltd., Finland) or HySpex (Norsk Elektro Optikk, Norway), have also been developed and extensively used, for example, for yield prediction [e.g., Uno et al., 2005], canopy nitrogen assessment [e.g., Siegmann et al., 2013] or monitoring of several crop parameters [e.g., Jarmer, 2013]. Furthermore, an increasing number of low weight multi- and hyperspectral sensors is now available, which can be mounted on UAVs. Although these systems can only cover small areas, UAVs are flexibly usable and their operation is relatively inexpensive. Initial studies have already proved their suitability for precision agriculture [e.g.,

Aasen et al., 2015; Zarco-Tejada et al., 2012].

Beside the large number of existing airborne hyperspectral systems, EO-1 Hyperion is currently the only satellite system in operation that provides spatial medium-resolution hyperspectral data free of charge. A study on the evaluation of EO-1 Hyperion data by Miglani et al. [2009] concluded that the data have a lot of potential for investigating agricultural crops. This potential was previously demonstrated, for example, through the retrieval of wheat chlorophyll content [Bannari et al., 2008] and the estimation of rice yield [Datt et al., 2003]. In the near future, new hyperspectral satellite missions, such as EnMAP (planned launch in 2018) and HyspIRI (planned launch in 2022), will continue hyperspectral earth observation. Both systems provide great potential for agricultural applications. In comparison to EO-1 Hyperion, however, EnMAP and HyspIRI data will have improved qualities with respect to SNR and spatial coverage [Guanter et al., 2015; Lee et al., 2015].

3.2 Requirements of Precision Agriculture on Remote Sensing

From the precision agricultural perspective, different information is required in order to apply the right treatment of fields in the right place at the right time. In this context, remote sensing data serve as a vital source of information, which are still used for many agricultural applications. According to Whelan and Taylor [2013], the benefit of using remote sensing data, however, depends on the following three essential properties of sensor systems, which are vital to their use in precision agriculture:

- temporal resolution
- spectral resolution coverage

- spatial resolution

The temporal resolution is very important in order to enable the monitoring of crops over the growing season. Temporal frequency has been improved dramatically from Landsat 1 (18 d) to modern satellite systems (e.g., RapidEye, Worldview-3), which have revisit cycles of one day. Moreover, the increasing availability of remote sensing image data from airborne platforms (UAVs and aircraft) additionally allow for the collection of remote sensing data at multiple times, which is highly in demand in precision agriculture in order to conduct near real-time soil, crop and pest management [Mulla, 2013]. Alongside the higher flexibility of airborne data acquisition to obtain image data at a given time, it is also possible to collect data under cloudy conditions, which is one of the major drawbacks of optically reflective satellite remote sensing, especially in Central Europe. Together with temporal resolution, the turnaround time of acquired image data is very important for time-critical applications in precision agriculture. Regarding turnaround time, Moran [1997] refers to the time that the user has to wait for the desired, processed satellite or airborne data. This includes both the delivery time from acquisition to user and the processing time for conversion of raw data to information. However, modern processing chains and improved computational power enable a fast availability of processed satellite data shortly after data acquisition. ESA, for example, provides access to processed Sentinel-2 data (radiometrically and geometrically corrected) within a few hours after data recording.

The spectral resolution of a remote sensing sensor is another important issue. Most optical sensors, which have been used for agricultural applications, are multispectral systems with three to ten spectral bands and a spectral coverage from 400 to 1,000 nm [Whelan and Taylor, 2013] (cf. Chapter 3.1). Some satellites have additional bands in the SWIR region (e.g., Landsat 8) that are better suited to characterize soils or retrieve the water content of plants. Multispectral images are also called broadband images because each band covers a certain spectral range (e.g., blue, green, red, NIR) with a bandwidth of several hundred nanometres [Yao et al., 2012]. Some optically reflective satellites of the latest generation are classified into the new group of superspectral sensors (e.g., Sentinel-2, Worldview-3). These are systems having more than ten and less than 50 spectral bands with narrower bandwidths compared to multispectral sensors [Herrmann et al., 2011; Malin, 2007]. According to Stoorvogel et al. [2015], satellites belonging to this group will provide the opportunity for estimating biophysical properties of crops with improved accuracy. Moreover, the increasing number of the already mentioned hyperspectral airborne sensors, and the availability of hyperspectral satellites with improved data quality in the near future, offer/will offer the capability of sensing a wide variety of soil and crop characteristics simultaneously [Mulla, 2013]. Compared to superspectral sensors, hyperspectral sensors often have more than 100 bands and an even narrower bandwidth of typically a few tens of nanometres or less [Baltsavias, 2002]. Thus, they provide the potential for a much

more detailed extraction of plant and soil information, including parameters such as soil moisture status, soil organic matter, nutrients, chlorophyll, carotenoids, cellulose, LAI and crop biomass [Mulla, 2013; Yao et al., 2012].

Although hyperspectral data contain information in a large number of spectral bands, EO-1 Hyperion and near-future hyperspectral satellite sensors have/will have a relatively small spatial coverage and a long revisit cycle compared to current multi- and superspectral systems. The spatial resolution, which Whelan and Taylor [2013] identified as another important characteristic, can also be regarded as a drawback of hyperspectral satellite sensors for applications in precision agriculture. Unfortunately, the combination of a good SNR for a spatial high-resolution instrument is extremely difficult from a technical point of view, especially when dealing with hyperspectrality. Therefore, the spatial resolution of hyperspectral satellites is currently limited to 30 m GSD, which allows every narrow spectral band to gather enough light and in turn ensure a high SNR [Villafranca et al., 2012].

For most applications in precision agriculture, however, a higher spatial resolution than 30 m is required to enable an adapted treatment on a sub-field scale. According to Mulla [2013], the necessary spatial resolution depends on the type of application. While variable rate application of fertiliser and variable rate irrigation are possible using data with 5–10 m spatial resolution, other applications, such as the estimation of spatial patterns in crop yield and variable rate spraying of herbicides for spot weed control, require distinctly higher resolution image data with GSDs of 3 m or even higher. Newer satellites, such as Worldview-3, Pleiadis or RapidEye, for example, offer this high-resolution images but data acquisition is expensive and, therefore, often only profitable for large commercial farms [Mulla, 2013]. For this reason, freely available medium-resolution satellite images (e.g., Landsat 8, EO-1 ALI, Sentinel-2, EO-1 Hyperion) are still used for agricultural applications. Beside the insufficient spatial resolution of these sensors for most precision agricultural applications, it is also very difficult to validate derived parameter maps based on these images. Satellites such as Landsat 8 or EO-1 Hyperion have a pixel size of 900 m². This makes it very complicated or even impossible to collect in situ parameter values (e.g., LAI, biomass, chlorophyll content) that are representative for such large areas. To overcome the limitation of insufficient spatial resolution, pan-sharpening can be used as an adequate method to enhance the GSD of these data sets [Whelan and Taylor, 2013]. In this context, many multispectral satellites also have a single panchromatic band, which provides a finer spatial resolution in comparison to their multispectral bands. SPOT 6/7, EO-1 ALI and Landsat 8, for example, have panchromatic bands covering the VIS range with spatial resolutions of 1.5, 10 and 15 m, respectively. These higher resolution data can be used to increase the spatial resolution of lower resolution multi- or hyperspectral image data, while preserving their spectral characteristics at the same time [Ehlers et al., 2010; Pohl and van Genderen, 2015]. On the one hand, the

improved spatial resolution as the result of pan-sharpening increases the image data suitability for precision agricultural applications; on the other hand, the validation of derived parameter maps will be facilitated because only smaller areas must be sampled during field campaigns, which are spatially representative for the smaller pixel size after pan-sharpening. In this context, the Ehlers Fusion, a pan-sharpening technique combining colour space transformation with filtering in the frequency domain, was applied in this thesis in order to investigate the potential of pan-sharpened hyperspectral EnMAP data for the precise spatial retrieval of LAI. An overview about the methodological background of this pan-sharpening technique is presented in Chapter 6.3.1.

In general, the selection of suitable optically reflective image data largely depends on the requirements of their specific use in agriculture. For that reason, Whelan and Taylor [2013] provide a checklist of sensor specifications, which should be considered when choosing image data:

- temporal resolution
- spectral resolution and coverage
- spatial resolution
- area captured in the image
- availability and timing of delivery
- price of image data and required processing

In this regard, the overview of available satellites and their corresponding characteristics, which are presented in Table 3.1, can help to select data with suitable spatial, spectral and temporal resolution, which meet the actual requirements of a precision agricultural application.

4 Study Area[8]

The study area (11°54′E, 51°47′N) is located in the eastern part of Germany in the federal state of Saxony-Anhalt and belongs to the Magdeburgian-Anhaltinian Börde (Figure 4.1). Within the study area, four wheat fields (A and B (2011) and C and D (2012)) were investigated. The wheat fields are located on a plain between the city of Köthen in the east and the city of Bernburg in the west.

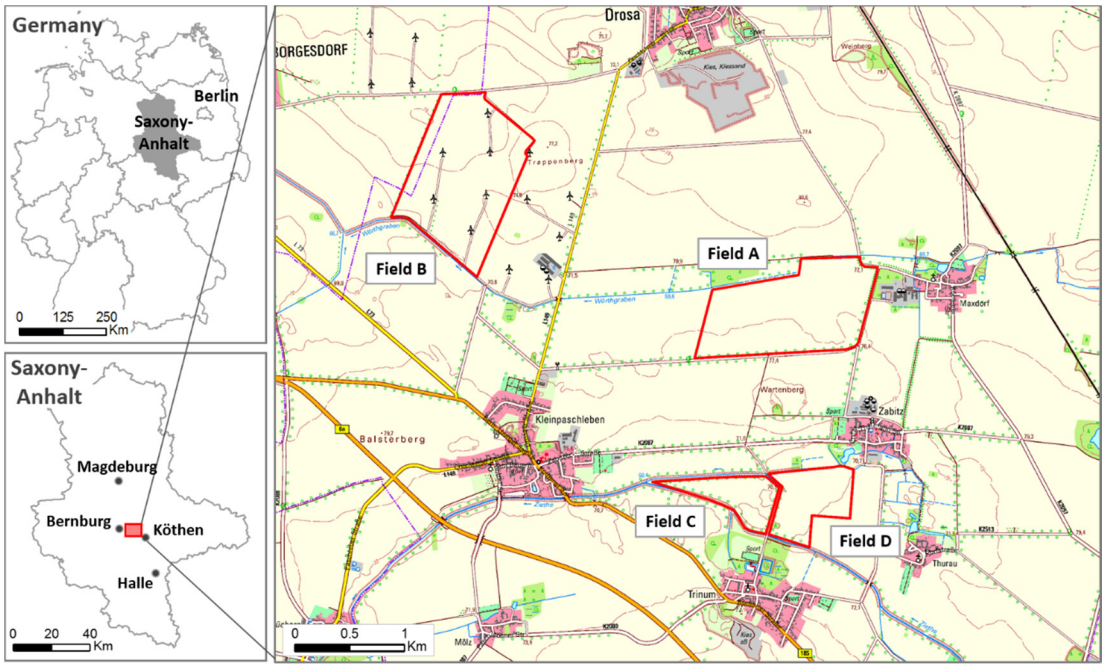


Figure 4.1: *Location of the study site in Germany (top left) and in Saxony-Anhalt (bottom left); the four investigated wheat fields (A–D) within the study site (right), background: digital topographic map on the scale of 1: 25,000 (TK 25), © GeoBasis-DE [LVermGeo LSA, 2015].*

4.1 Geographical Aspects

The entire area was formed during the last ice age from 10,000 years ago. At this time, the region was the preferred run-off area of the main local rivers (i.e., Elbe, Saale, Mulde, Weiße Elster), which accumulated Pleistocene sediments in the region. Later, during the Weichselian period, the area was covered by loess, such that the existing relief structures were nearly compensated [Schröder, 2000]. Today, the region is characterized by a slightly undulated plain. The altitude of the study site varies between 65 and 88 m above sea level. While Field C and Field D are relatively flat with an average altitude of 70 m, Field A and Field B show differences in altitude up to 4 m and 7 m, respectively (Figure 4.2 a)).

[8] Parts of this chapter have previously been published by Siegmann and Jarmer [2015] and Siegmann et al. [2015].

The area is covered by a loess layer up to 1.2 m deep. Chernozems, in conjunction with Cambisols and Luvisols (Figure 4.2 b)), represent the predominant soil type characterized by highly diverse soil properties, resulting in a fine-scale pattern of soil texture and organic matter. The soils from the study site have an average quality (Ackerzahl (in German)) of about 85 and, therefore, belong to the most fertile soils in Germany [Schröder, 2000].

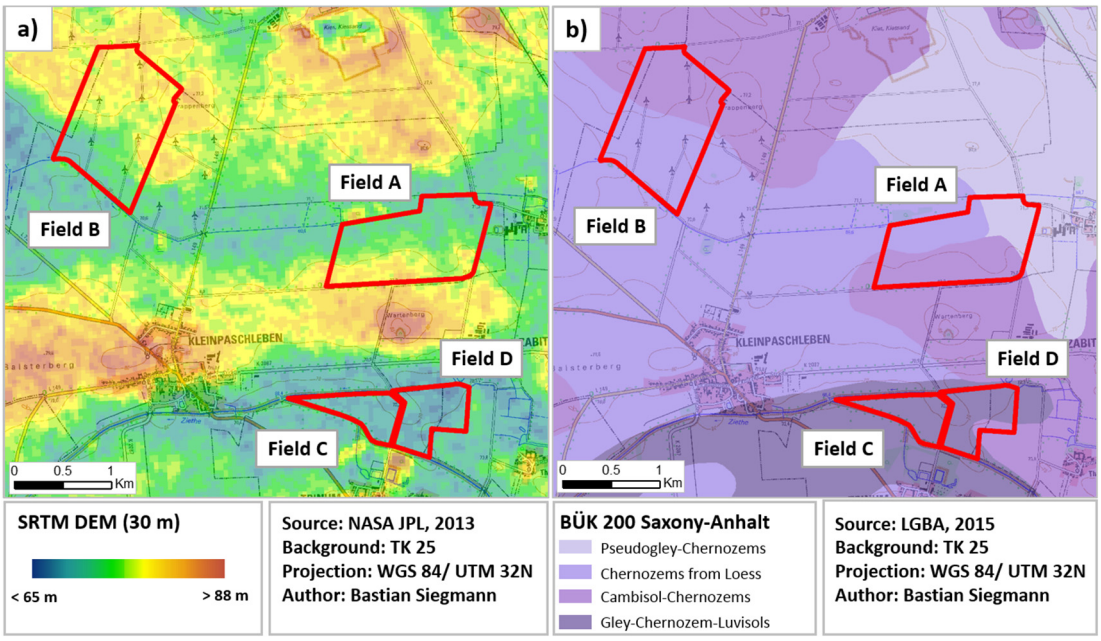


Figure 4.2: *a) Digital elevation model of the study site (based on SRTM) with a spatial resolution of 30 m [NASA JPL, 2013]. b) Soil map of the study site (based on BÜK 200) on the scale of 1:200,000 [LGBA, 2015]. Both maps are underlaid by the digital topographic map on a scale of 1:25,000 (TK 25), © GeoBasis-DE [LVermGeo LSA, 2015].*

Since the area is situated in the rain shadow of the Harz Mountains, it is distinctly dry with 500 mm of mean annual precipitation. The mean annual temperature is 9°C (Figure 4.3 a)). Due to less pronounced differences in the relief structure, the entire region has similar climatic conditions. In comparison to the long-term average, the winter of the 2010/11 growing season was relatively cold and wet (Figure 4.3 b)). This was followed by a period of less precipitation in spring 2011, leading to drought stress symptoms within the observed wheat fields. In contrast, the summer was characterized by a lot of rain, which is less advantageous for the harvest that is normally conducted in late July or early August (Figure 4.3 c)). In general, the weather conditions of the 2011/12 growing season matched more closely to the long-term temperature and precipitation profile. Although late winter and early spring were again a little dry, the months of May, June and July brought much more rain and, thus, guaranteed decent growth conditions (Figure 4.3 d)).

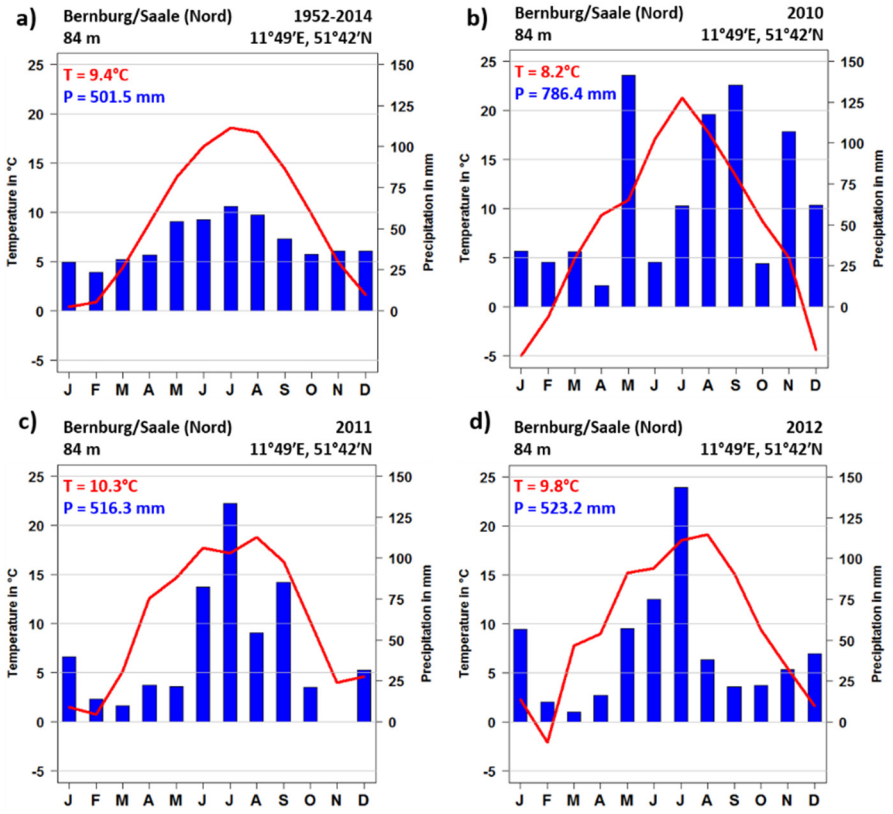


Figure 4.3: *Climate graphs of the weather station Bernburg/Saale (Nord): a) 1952-2014, b) 2010, c) 2011 and d) 2012 (based on data of DWD [2015]).*

4.2 Phenological Aspects of Winter Wheat

Due to its high yield rates and its adaptability to different soil conditions, winter wheat is one of the most important cultivated plants [BLV, 2014]. For that reason, it is grown on more land area than any other commercial crop in the world [CIMMYT, 2015]. In Germany, wheat is also the most important agricultural plant (Figure 4.4). In particular, wheat in Saxony-Anhalt was cultivated on 330,000 ha in both 2011 and 2012. This corresponds to 60 % of the total cropping cultivation area of the entire federal state [SLS-A, 2015]. Thus, the proportion was a little higher than the national average (cf. Figure 4.4).

The phenological development of winter wheat is typically described by different growth stages. In the past, several scales have been developed explaining the growth stages in detail (e.g., Feekes scale [Large, 1954], Zadoks scale [Zadoks et al., 1974] and the scale of Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie (BBCH scale) [Meier, 2001]). Since it is the most frequently used scale in Europe, the BBCH scale was applied in this thesis to describe the growth stages of winter wheat. The BBCH scale consists of ten principle growth stages (Table 4.1). Each of the ten principle stages is further divided into ten secondary stages. During the

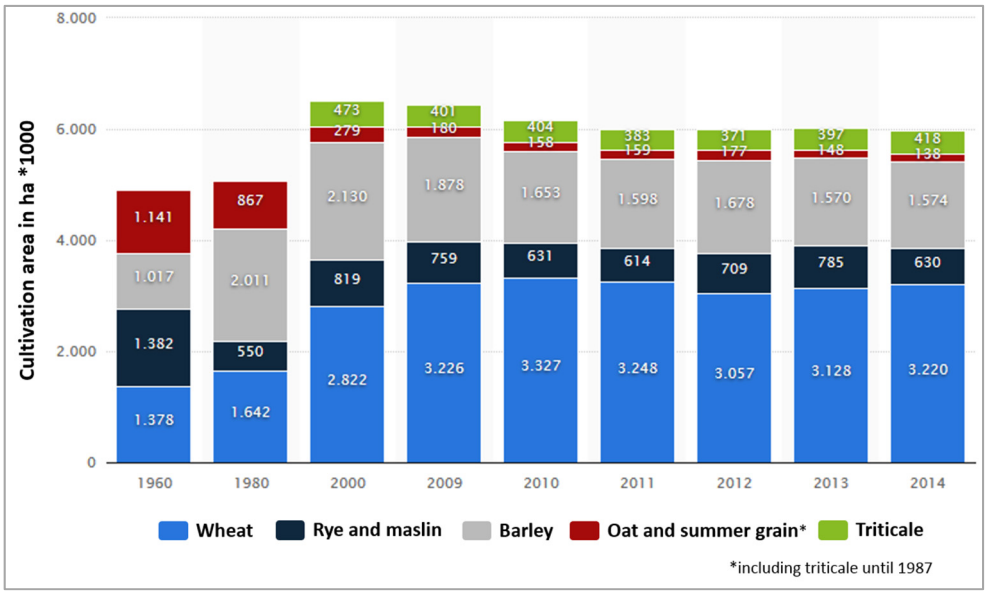


Figure 4.4: *Cultivation area of relevant crops in Germany in 1965–2014 (modified according to Statista [2015] based on data of Statistisches Bundesamt [2015]).*

plant development, the different secondary stages normally overlap because of slightly different site conditions within a field. Therefore, only the principle stages are used to describe the winter wheat development in this thesis. In addition, Figure 4.5 illustrates computer-graphical models of a wheat plant in the different growth stages simulated by Spengler [2013].

Table 4.1: *BBCH scale, principle growth stages of winter wheat [Meier,* 2001*].*

Stage	Principle stage
0	Germination / sprouting / bud development
1	Leaf development (main shoot)
2	Formation of side shoots / tillering
3	Stem elongation or rosette growth / shoot development (main shoot)
4	Development of harvestable vegetative plant parts or vegetative propagated organs / booting (main shoot)
5	Inflorescence emergence (main shoot) / heading
6	Flowering (main shoot)
7	Development of fruit
8	Ripening or maturity of fruit and seed
9	Senescence, beginning of dormancy

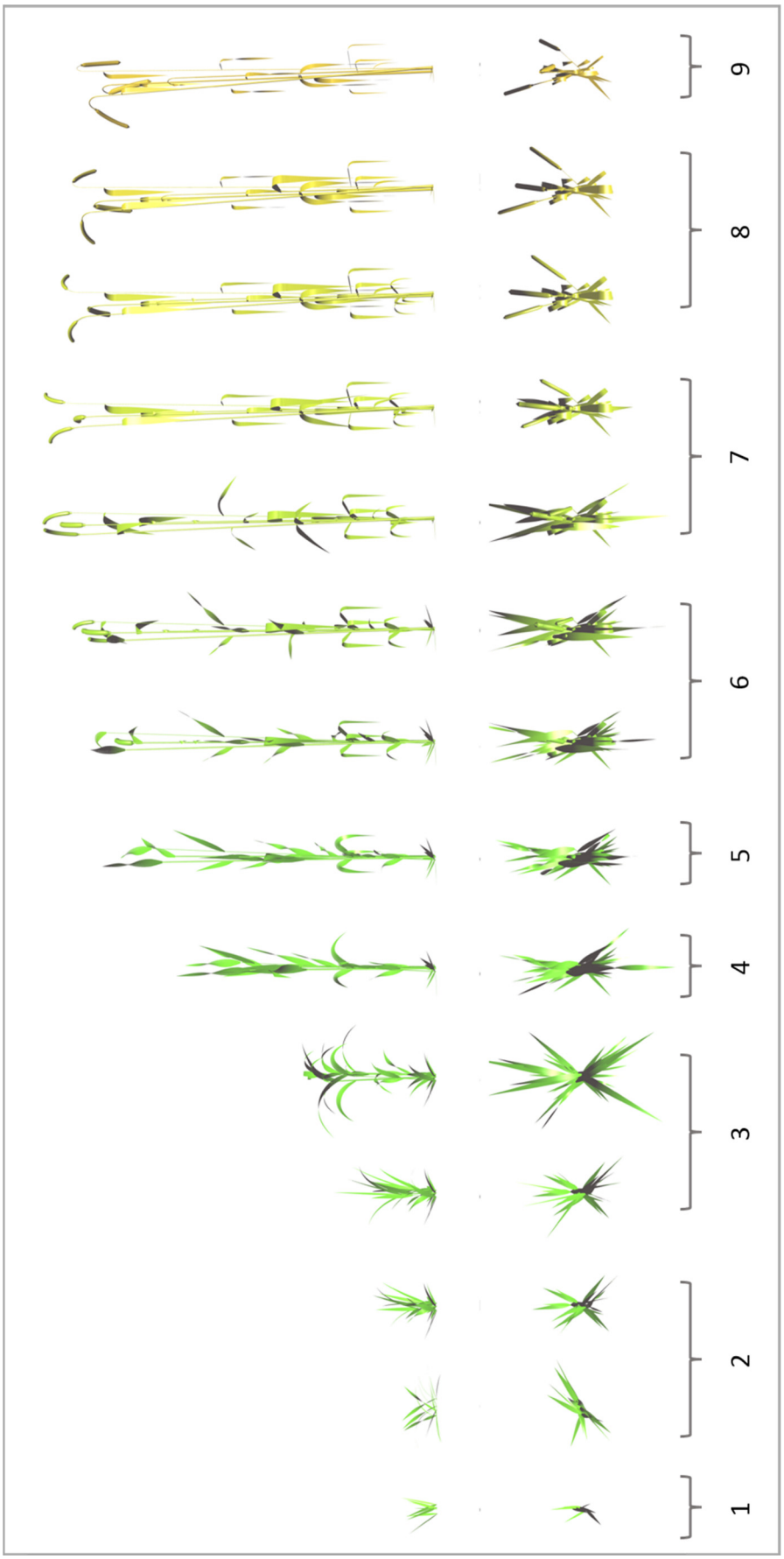


Figure 4.5: *Modelled wheat plants with corresponding BBCH stages (side view (top), top view (down)), (modified according to Spengler [2013]).*

The sowing date of the investigated fields in 2011 (Field A and Field B) and 2012 (Field C and Field D) was in late October in each of the previous years. Before winter, the wheat plants passed the first three development stages (i.e., germination, leaf development, tillering). Afterwards, winter wheat needs a period of cold temperatures between 0° and 5° C to initiate the flowering. This process is called vernalization, which is an important requirement for stem elongation in spring [BLV, 2014; Bowden et al., 2008]. Until summer, the plants passed growth stages four to eight (i.e., booting, heading, flowering, fruit development, ripening). Depending on the weather conditions, the harvest normally takes place in late July or at the beginning of August. Field A and Field B were harvested on 11 August 2011, and Field C and Field D on 10 August 2012.

5 Data and Pre-processing[9]

5.1 Field Data

During two field campaigns in May 2011(7th–10th) and May 2012 (24th–25th), the four wheat fields (A–D) were investigated (Figure 5.1). The fields varied in size between 21 and 94 ha. A digital elevation model (DEM) and Landsat 5 data from previous years were used prior to the field campaigns to develop an adjusted strategy for allowing a representative sampling of the four investigated fields. In this regard, the topographic wetness index (TWI) [Beven and Kirkby, 1979] was calculated using the DEM to identify regions with different soil conditions. Different soil properties in turn led to spatially inhomogeneous growth conditions for plants. In total 124, winter wheat plots (each of size 50 × 50 cm), which were distributed across the investigated fields, were sampled. Measurements from 57 plots were collected in May 2011. One year later, in May 2012, 67 plots were sampled (see Table 5.1). For each plot, several plant parameters were determined and spectral reflectance measurements were collected. A detail description of field data acquisition is given in the next sections.

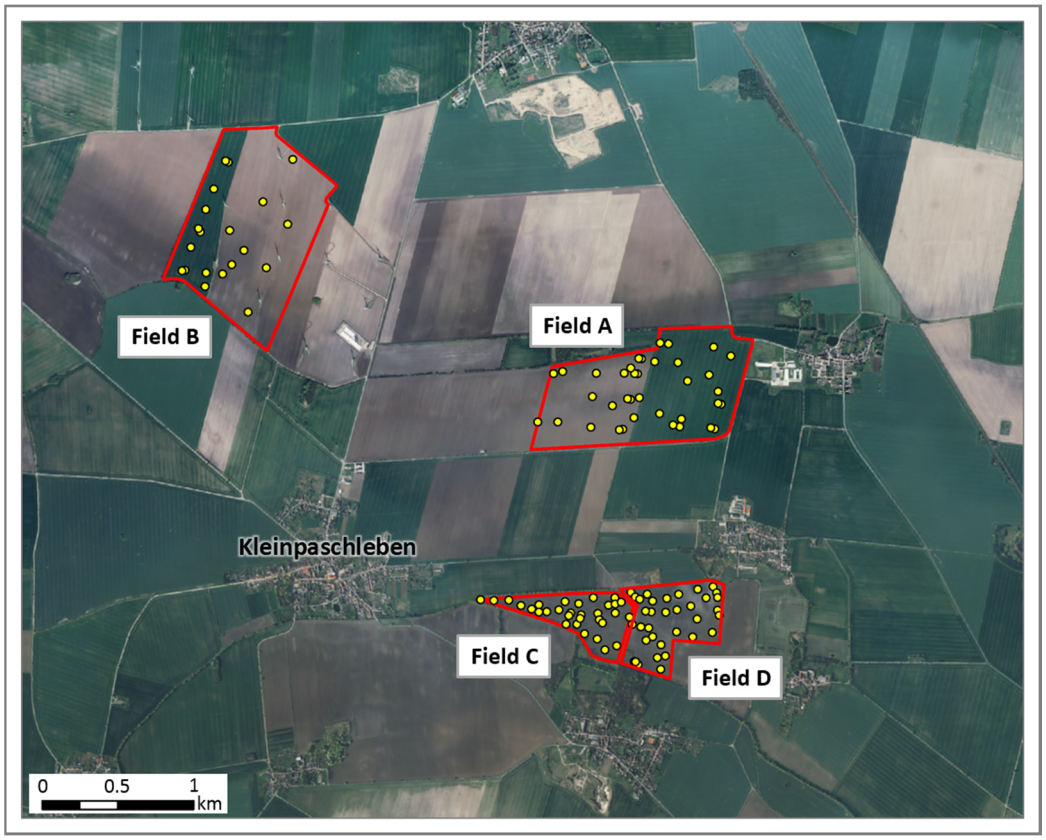


Figure 5.1: *The four investigated wheat fields (A–D) (red polygons) and the sampled plots within the fields (yellow dots), background: digital aerial photos of 2009 [MLU, 2009].*

[9] Parts of this chapter have previously been published by Siegmann and Jarmer [2015] and Siegmann et al. [2015].

Table 5.1: *Investigated fields in 2011 and 2012, with information on field size and number of sampled plots.*

Field	Size in ha	Number of plots	Year
A	82	37	2011
B	94	20	2011
C	21	31	2012
D	29	36	2012

5.1.1 Wheat Parameter Measurements

During the field campaigns, a standardized measurement procedure was applied to every plot. Firstly, the chlorophyll content (Cab) of ten different top leaves was measured with a SPAD 502 Chlorophyll Meter (Konica Minolta Holdings Inc., Japan) and plant height as well as the phenological stage of plant development (BBCH scale) were determined. Additionally, LAI measurements were acquired for every plot (see details below), the above-ground biomass fresh matter (FM) was harvested completely and soil samples of every plot were collected (only in 2012). After the field campaigns, the plant material was dried in an oven at 60° C for at least 24 hours, after which the carbon and nitrogen content of the above-ground biomass dry matter (DM) of each plot was determined in the laboratory. Furthermore, the weight difference of FM and DM enabled the determination of the plant water content (PWC) of every plot.

Since LAI assessment is the focus of this thesis, its determination in the field should be described more in detail. The LAI of all plots was measured non-destructively with a SunScan device in 2011 and an LAI-2000 in 2012. For covering a wide range of values, the LAI was acquired for the two different phenological stages of stem elongation (BBCH 3) in 2011 and inflorescence emergence (BBCH 5) in 2012. In both years, the LAI was measured six times for each plot. Mean LAI values of each plot were calculated from all measurements of a plot within the simple standard deviation to exclude outliers from further investigations.

The LAI of all measured wheat plots covered a range between a minimum of 0.50 m^2 m^{-2} and a maximum of 5.70 m^2 m^{-2}. The mean LAI of 2.80 m^2 m^{-2} indicated that lower LAI values predominated. Due to the earlier stage of plant development (BBCH 3), the LAI values of the wheat plots in 2011 were relatively low, varying between 0.50 and 3.40 m^2 m^{-2}, with an average of 1.64 m^2 m^{-2}. In 2012, the measured plots had a higher LAI because plant development was already more advanced (BBCH 5) than in 2011. The range of measured LAI, with a minimum of 1.04 m^2 m^{-2} and a maximum of 5.70 m^2 m^{-2}, was relatively high, suggesting a more pronounced variability of the wheat stands in comparison to those measured in 2011. However, only three plots in 2012 had values less than 2.00 m^2 m^{-2}. Most of the LAI measurements varied between 2.50 and 5.70 m^2 m^{-2}. Thus, the values measured in both

years covered a different value range (Figure 5.2). Descriptive statistics of the winter wheat samples for the individual fields, the individual years and when taken all together are summarized in Table 5.2.

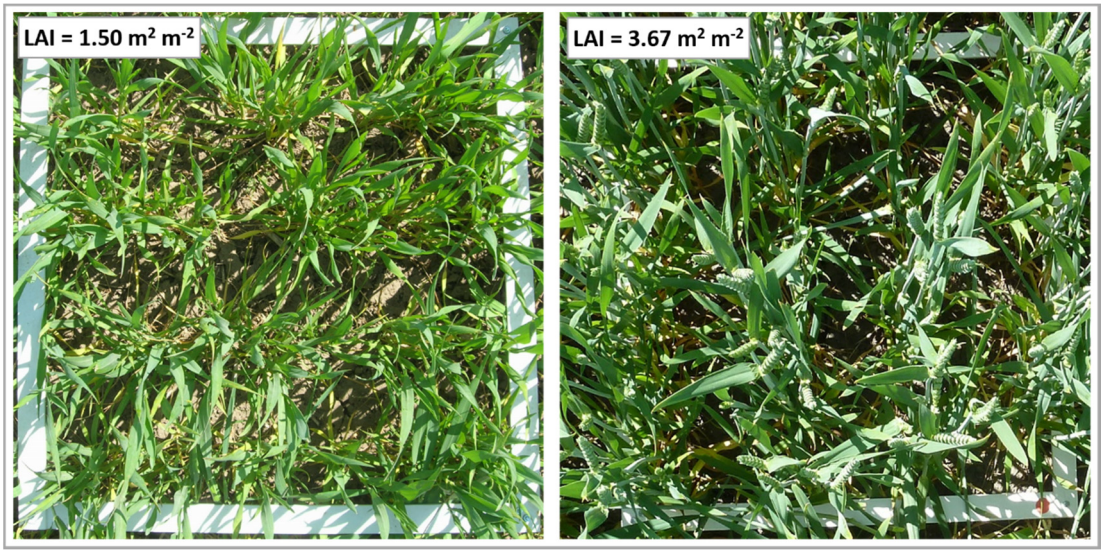


Figure 5.2: *Examples of sampled field plots and corresponding LAI values measured in 2011 (left) and 2012 (right).*

Table 5.2: *Descriptive statistics of winter wheat LAI values (in m^2 m^{-2}).*

Year	Field	n	Min	Max	Mean	SD	Median
2011	Field A	37	0.50	3.40	1.54	0.71	1.50
	Field B	20	0.94	2.95	1.84	0.50	1.75
	Field A & B	57	0.50	3.40	1.64	0.66	1.58
2012	Field C	31	1.04	5.34	3.60	0.96	3.58
	Field D	36	1.91	5.70	3.94	0.82	3.97
	Field C & D	67	1.04	5.70	3.78	0.90	3.84
2011 & 2012	Field A–D	124	0.50	5.70	2.80	1.33	2.78

As one of the structural canopy parameters, LAI is correlated with other structural and biochemical canopy properties [Verrelst et al., 2015]. In this thesis, the calculated correlation coefficients between LAI and DM ($r = 0.95$, $p < 0.01$), as well as between LAI and PWC ($r = 0.98$, $p < 0.01$), illustrated a high statistical relationship between the measured variables. Although the correlation between LAI and Cab was less pronounced, the calculated coefficient of 0.73 ($p < 0.01$) also indicated a medium-high relationship. These correlations between LAI and other plant parameters have been reported in several scientific studies [Filella and Peñuelas, 1994; Stagakis et al., 2010].

5.1.2 Spectral Reflectance Measurements

In parallel to the in situ plant parameter determination, reflectance measurements of each plot were collected with two ASD FieldSpec III spectroradiometers (Analytical Spectral Device Inc., USA) in 2011 and one ASD FieldSpec III, as well as one SVC HR-1024 spectroradiometer (Spectra Vista Corporation, USA) in 2012 (Figure 5.3). All instruments covered the spectral range 350–2,500 nm. The measurements were taken 1.4 m above the ground in the nadir view using 25° field of view (FOV) optics in the case of the ASD instruments, and a 14° FOV optic with the SVC instrument. All measurements were converted to absolute reflectance values using a reflectance standard of known reflectivity (Spectralon, Labsphere Inc., USA). Five spectral measurements were acquired per plot and averaged to one single measurement for further analysis. The exact position of each plot was located by a GPS device and photos were taken for documentation. Given the lower spectral resolution of the SVC HR-1024 spectral data (1,024 spectral bands), the measured ASD FieldSpec III spectral signatures (2,151 spectral bands) were resampled to the spectral resolution of the SVC HR-1024 data to make both data sets comparable. The bands in the region of atmospheric water absorption (1,350–1,480 nm and 1,780–1,990 nm) were deleted, leaving 867 spectral bands between 350 and 2,500 nm.

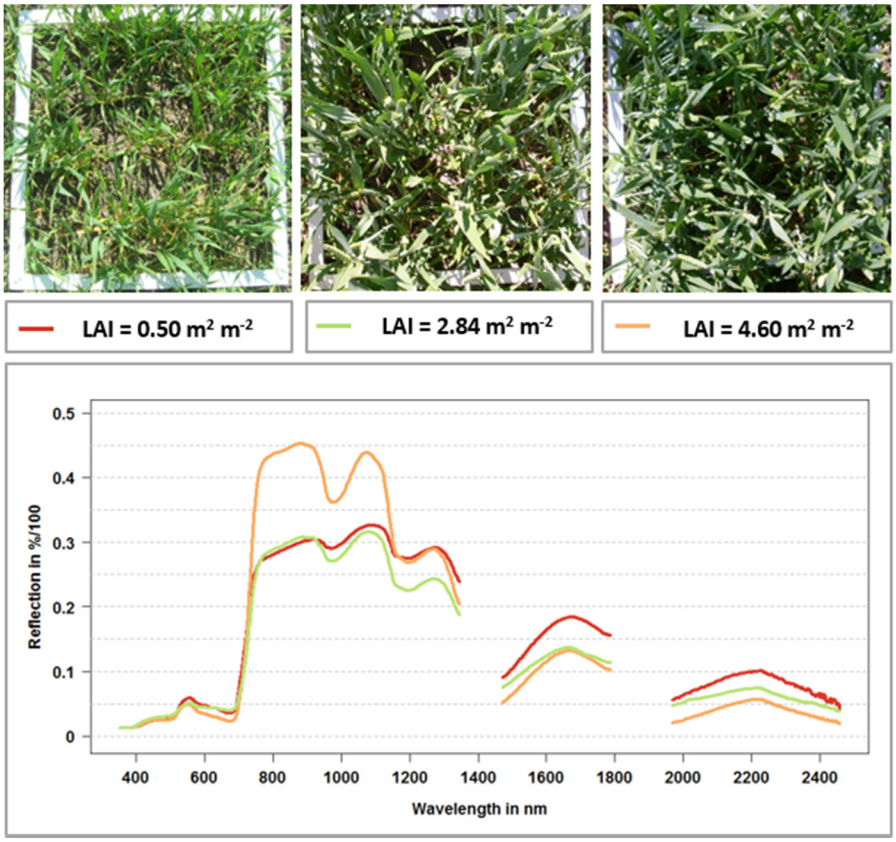


Figure 5.3: *In situ spectral reflectance measurements of plots with different LAI values.*

5.2 Multi- and Hyperspectral Image Data

5.2.1 Airborne Data

In parallel to the field campaigns, image data of the test site were acquired by the airborne hyperspectral scanner aisaDUAL (Specim Ltd., Finland) for allowing spatial predictions of LAI of the observed fields. The aisaDUAL system is a hyperspectral push broom scanner consisting of the two separate sensors, namely, aisaEAGLE (VIS/NIR, 400–1,000 nm) and aisaHAWK (SWIR, 1,000–2,500 nm) (Figure 5.4). The recorded image data have a geometric resolution of 3 m with 367 spectral bands in the wavelength range 400–2,500 nm.

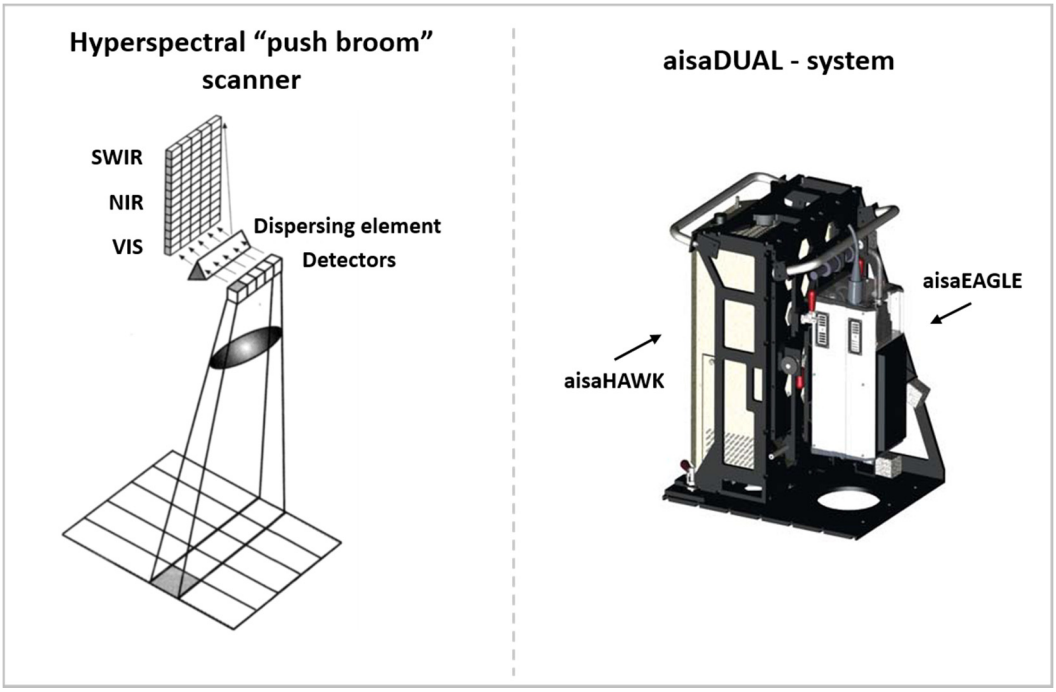


Figure 5.4: *Principle of a hyperspectral push broom scanner (left) (modified according to Jensen [2004]), and an aisaDUAL airborne system consisting of the two sensors, aisaEAGLE and aisaHawk (right) [Specim Ltd., 2015].*

For data correction, the ROME destriping algorithm [Rogaß et al., 2011] was used to reduce miscalibration effects, which were present as deficient lines along the track in the images. Afterwards, an atmospheric correction was conducted using MODTRAN4 (MODerate resolution atmospheric TRANsmission) for transferring the radiance values of each pixel to reflectance data [Berk et al., 1998]. Additionally, an empirical line correction with spectral ground measurements of different dark and bright targets, which were collected at the study site during the time of aisaDUAL overpass, was necessary for removing spectral artefacts that were still in the data after atmospheric correction.

The geometric correction of the aisaDUAL data was realized with the software CaliGeo (Specim Ltd., Finland), while orthorectification was performed with the software ENVI (Exelis Inc., USA). Due to noise in the processed aisaDUAL spectra, an additional spectral binning was carried out on the data. Spectral binning is a commonly used method to reduce noise in hyperspectral data. In this context, adjacent spectral bands were summed up to one new single binned spectral band to enhance the SNR of the data [Dell'Endice et al., 2009]. Therefore, three adjacent spectral bands of the aisaDUAL data were averaged to generate one new spectral band. Thus, the number of spectral bands was reduced from 367 to 122 with improved SNR. Furthermore, 22 spectral bands in the range of the water vapour absorption bands (1,354–1,432 nm, 1,807–1,996 nm), as well as selected bands at the beginning and at the end of the aisaDUAL spectral range (400–430 nm, 2,390–2,500 nm) were deleted because of the high noise of the aisaDUAL system in these spectral regions, thereby leaving 98 spectral bands for further analysis. The different steps of aisaDUAL pre-processing are summarized in Figure 5.5.

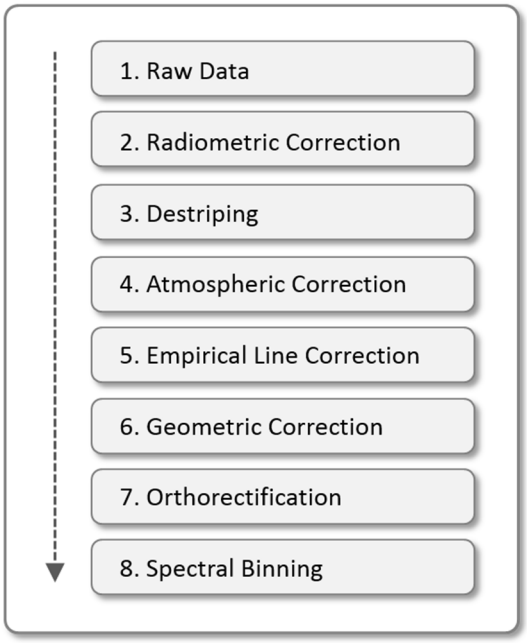


Figure 5.5: *Overview of aisaDUAL pre-processing.*

Although image data of Field B were acquired in 2011, a geometric correction of the aisaDUAL flight stripe covering the field was not possible because of an outage of the GPS device connected to the airborne scanner. Figure 5.6 a) illustrates the uncorrected image of Field B, which cannot be subsequently used in the rest of this thesis. Additionally, aisaDUAL image data of Field C have not been considered for the spatial prediction of LAI because of numerous small-scale inhomogeneous growth conditions within the field. Consequently, the measured plant parameters (including

LAI) of many field plots were only representative for areas, which were smaller than an aisaDUAL pixel with a size of 9 m^2. Moreover, two large areas within the field were not cultivated because of waterlogging at the time of sowing. Figure 5.6 b) shows one of these areas.

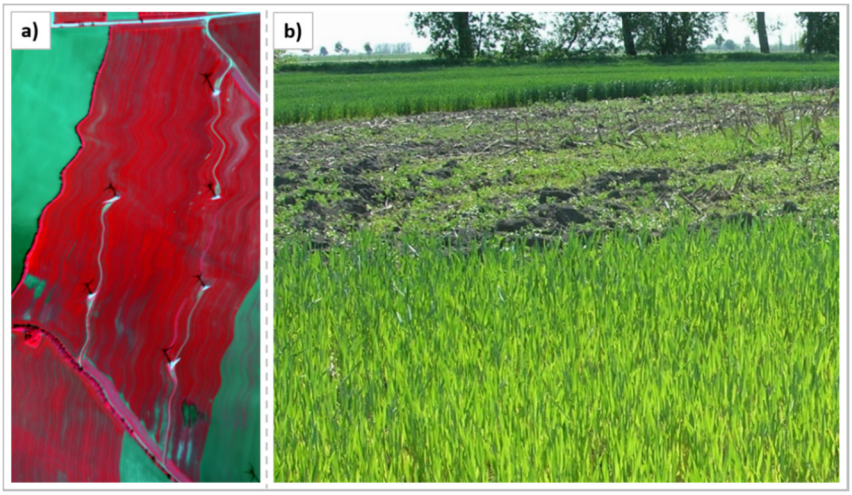


Figure 5.6: *a) Geometrically uncorrected aisaDUAL image data of Field B. b) Uncultivated area within Field C due to waterlogging during sowing.*

For that reason, Figure 5.7 a) only shows the processed aisaDUAL data subsets covering Field A (2011) and Field D (2012). In addition, two drainless hollows in the northern and south-eastern parts of Field A, which were investigated in 2011, were masked by building a decision tree based on NDVI because there was no vegetation cover as a result of waterlogging in early spring in that particular year.

Furthermore, the 33 spectral bands covering the spectral range from 450 to 900 nm were averaged to generate panchromatic images (aisaDUAL pan) out of the aisaDUAL data (Figure 5.7 b)). This spectral range was chosen because it approximately corresponds to the spectral range of the panchromatic bands of current multispectral satellite systems. The generated panchromatic images were a requirement for the pan-sharpening of the EnMAP data to be presented later in this thesis.

5.2.2 Satellite Data

Image data from the German hyperspectral satellite EnMAP (scheduled for launch in 2019) and the ESA superspectral satellite Sentinel-2 (launched on 23 June 2015) were simulated using the aisaDUAL data sets of the study site acquired in 2011 and 2012. Table 5.3 provides an overview of their specifications.

The simulations were conducted using the EnMAP end-to-end simulation tool (EeteS). The tool consists of a complete forward (EnMAP Scene Simulator) and

Table 5.3: *Specifications of EnMAP [DLR, 2015] and Sentinel-2 [ESA, 2015].*

	EnMAP	Sentinel-2
Spectral range	420–2,450 nm	443–2,190 nm
Spectral bands	204	13
Spectral sampling	6.5 nm (420–1,000 nm) 10 nm (900–2,450 nm)	15–180 nm band dependent
Signal-to-noise ratio	500 (at 495 nm) 150 (at 2,200 nm)	154 (at 490 nm) 100 (at 2,190 nm)
Radiometric resolution	14 bit	12 bit
Geometric resolution	30 m	10 m (4 bands) 20 m (6 bands) 60 m (3 bands)
Swath width	30 km	290 km
Orbit altitude	652 km	795 km
Return frequency	27 days	10 days

as a detailed physical sensor model of EnMAP, which enables the simulation of realistic EnMAP data. A detailed description of EeteS is given in Segl et al. [2012] and Guanter et al. [2009]. Although the software was especially developed to simulate EnMAP data, the flexible modular structure of EeteS can easily be adapted to simulate image data from other sensors (e.g., Sentinel-2). Thus, EeteS offers the possibility to investigate the potential of future satellite missions for numerous applications and supports the development of new sensor designs [Segl et al., 2012].

As a result of the simulations, two EnMAP (Figure 5.7 c)) and two Sentinel-2 (Figure 5.7 d)) data sets (one each in 2011 (Field A) and 2012 (Field D)) were generated, which have the specific sensor characteristics of EnMAP and Sentinel-2, respectively. Due to the different GSDs of the single spectral bands, which will be provided by Sentinel-2, only the four spectral bands with the highest spatial resolution of 10 m (band 2: 458–522 nm, band 3: 543–577 nm, band 4: 650–680 nm, band 8: 785–900 nm [ESA, 2015]) were further used for pan-sharpening the EnMAP data. Comparable to the aisaDUAL data, panchromatic images (Sentinel-2 pan) for both years were created out of the Sentinel-2 scenes by averaging the four Sentinel-2 spectral bands (Figure 5.7 e)). Afterwards, the simulated EnMAP and Sentinel-2 data sets were resized to match the spatial dimensions of the aisaDUAL data sets covering the investigated fields. Dependent on the GSD of the panchromatic image used for pan-sharpening, the EnMAP data had to be resampled (with cubic convolution) to the same spatial resolution (aisaDUAL pan: 3 m, Sentinel-2 pan: 10 m).

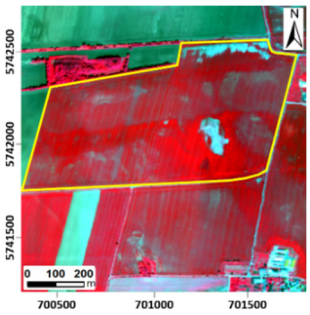
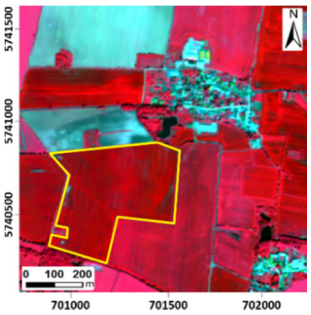
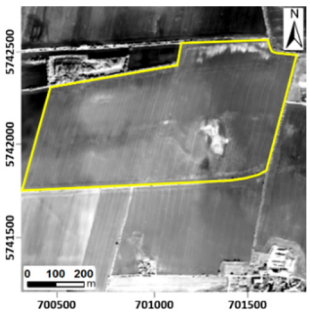
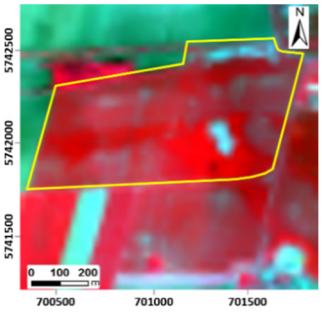
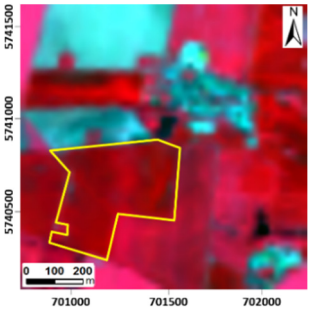
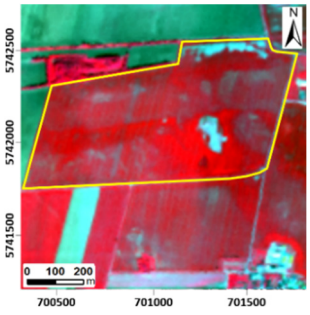
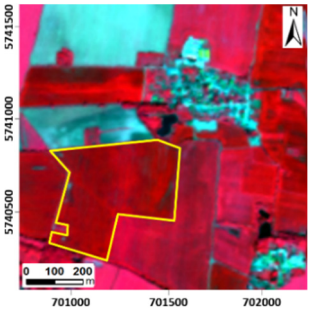
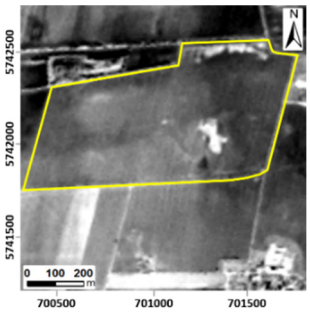
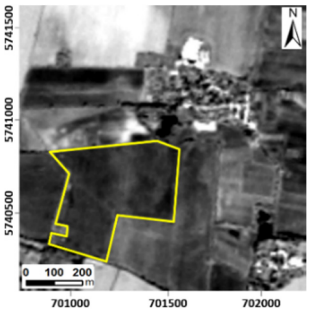
	Field A - 2011	Field D - 2012	Sensor properties
a) aisaDUAL			➢ GSD: 3 m ➢ 98 spectral bands ➢ range: 440–2,370 nm ➢ displayed images in RGB (863/652/548 nm)
b) aisaDUAL pan			➢ GSD: 3 m ➢ 1 spectral band ➢ range: 440–900 nm
c) EnMAP (simulated)			➢ GSD: 30 m ➢ 204 spectral bands ➢ range: 440–2,370 nm ➢ displayed images in RGB (863/652/548 nm)
d) Sentinel-2 (simulated)			➢ GSD: 10 m ➢ 4 spectral bands ➢ range: 458–900 nm ➢ displayed images in RGB (842/665/559 nm)
e) Sentinel-2 pan			➢ GSD: 10 m ➢ 1 spectral band ➢ range: 458–900 nm

Figure 5.5: *Image data of the investigated fields (Field A (2011) and Field D (2012); yellow polygons) with specific sensor characteristics. a) aisaDUAL data; b) panchromatic data sets based on aisaDUAL; c) simulated EnMAP data; d) simulated Sentinel-2 data and e) panchromatic data sets based on Sentinel-2 simulations.*

6 Methodology

6.1 Empirical-Statistical Regression Models[10]

Empirical-statistical regression models can be broadly divided into parametric methods and non-parametric methods (cf. Figure 2.11). While parametric regression approaches only use a limited number of spectral bands to calculate indices, for example, which are related to soil or plant parameters, non-parametric methods use the full spectrum. The latter group of regression methods consist of a learning phase based on training data to optimize the prediction model. During this phase, weights (coefficients) must be adjusted to minimize the estimation error of the target variable (e.g., LAI) [Verrelst et al., 2015].

The following sections provide a detailed description of the three non-parametric regression methods that are applied in this thesis. At first, the theoretical background of PLSR is presented, which belongs to the linear non-parametric group of models. Afterwards, an overview is given of the functional principals of the two non-linear non-parametric models, SVR and RFR.

6.1.1 Partial Least Squares Regression

PLSR was originally developed for data analysis in econometrics [Wold, 1985]. Later, the algorithm was transferred to chemical applications [Wold et al., 1987] and has also been used in remote sensing for many years to predict biophysical and biochemical parameters [e.g., Darvishzadeh et al., 2008b; Jarmer, 2013; Pu, 2012]. In this thesis, PLSR was performed with the R package ‘pls’ [Mevik and Wehrens, 2007]. The following explanation of the algorithm is based on a study of Geladi and Kowalski [1986], who provide a detailed description of PLSR.

The starting point of PLSR involves two data matrices called the X and Y blocks. The X block has the dimension $n \times m$, where n represents the number of samples (e.g., number of spectral reflectance measurements), while m is the number of independent variables (e.g., reflection values of a spectral curve at certain wavelengths). The Y block represents the dependent variable o (e.g., LAI) whose n values belong to the corresponding sample of the X block. Y has the dimension $n \times$ o. A prerequisite in PLSR modelling is that the variables of both blocks have to be mean-centred before model calibration.

In general, a PLSR model consists of an outer relation and an inner relation. While the outer relation describes the individual consideration of the X block and Y block, respectively, the inner relation defines the linkage of both blocks. The outer relation describes the X block in the form of its principle components:

[10] Parts of this section have previously been published by Siegmann and Jarmer [2015].

$$X = TP' + E = \sum t_h p_h' + E \qquad [6.1]$$

where h can be regarded as a dummy index counting the used components, hereinafter referred to as latent variables (LV). T represents the scores matrix (t_h), p' the loadings matrix (p_h') and E is the residual of X. Figure 6.1 depicts how the scores and loadings are obtained by projecting X into vectors.

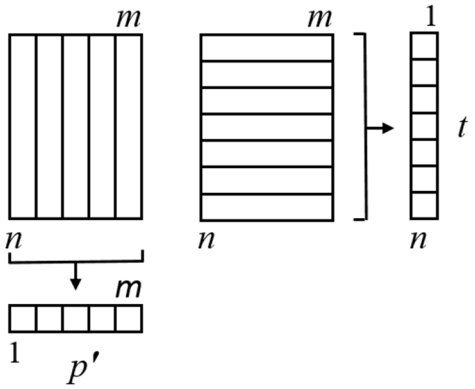
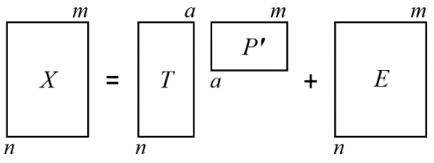


Figure 6.1: *Scores (t) and loadings (p') obtained by projecting X into vectors (modified according to Geladi and Kowalski [1986])*

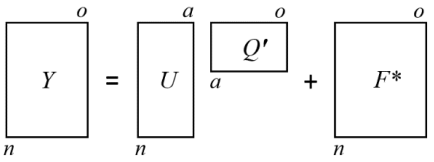
Replacing X by its principal components solves the problem of multi-collinearity. This is because the high number of independent variables can be reduced to a distinctly lower number of uncorrelated LVs. These represent an orthogonal linear combination of the original independent variables [Atzberger et al., 2010; Yu et al., 2015]. The outer relation of the Y block can be built in the same way:

$$Y = UQ' + F^* = \sum u_h q_h' + F^* \qquad [6.2]$$

Graphically, both outer relations can be represented as follows.



[6.3]



[6.4]

The goal is to describe Y as well as possible, keep $\|F^*\|$ as low as possible and establish a suitable relation between X and Y at the same time. Therefore, the scores of both blocks are related to each other for each LV (h). This can be done by the inner relation:

$$\hat{u}_h = b_h\, t_h \qquad [6.5]$$

where b_h can be regarded as the regression coefficient. Given that the LVs of both blocks were calculated separately, the best possible model cannot be found. As a result, both blocks have a weak relation to each other. For this reason, it is advantageous to give both blocks information about each other by slightly rotating their LVs to establish a better fit to the regression line. This is done by an iterative process. A problem in this context is that the scores (t_h) of the X block are no longer orthogonal. Therefore, the loadings (p_h') in Equation 6.1 need to be replaced by weights (w') to again generate orthogonal score values (t_h), after which the residuals of both blocks can be calculated as follows:

$$E_h = E_{h-1} - t_h\, p_h' \;;\; X = E_0 \qquad [6.6]$$

$$F_h^* = F_{h-1}^* - u_h\, q_h' \;;\; Y = F_0 \qquad [6.7]$$

For the outer relation of the Y block, u_h is substituted by $\hat{u}_h$ from Equation 6.5 to finally establish a mixed relation of both blocks:

$$Y = TBQ' + F^* = \sum b_h\, t_h\, q_h' + F^* \qquad [6.8]$$

with:

$$F_h = F_{h-1} - b_h\, t_h\, q_h' \qquad [6.9]$$

The relation of both blocks enables the possibility predicting the target variable from unknown data. Figure 6.2 additionally illustrates the relation of the X and Y blocks in PLSR analysis.

The maximum number of allowed LVs was set to ten, except less than ten spectral bands were involved in PLSR model building. In this case, the number was adapted in line with the number of used spectral bands. Furthermore, the Akaike information criterion (AIC) was calculated for the different PLSR models to determine the model with the optimum number of LVs [Akaike, 1974]:

$$AIC = n \log (RMSE) + 2\,h \qquad [6.10]$$

where n represents the number of samples and h is the number of LVs. Additionally, the root-mean-square error (RMSE) of cross-validation has to be determined [Viscarra-Rossel, 2008]. Finally, the number of LVs (and hence the corresponding PLSR model) was chosen, which provides the lowest AIC [Gerighausen, 2013; Li et al., 2002].

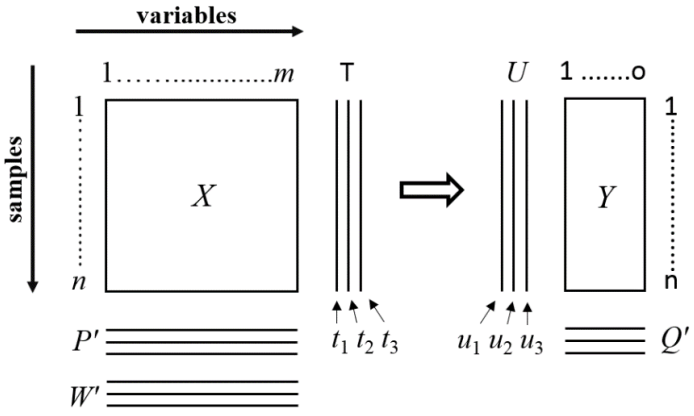


Figure 6.2: *Relation of X block and Y block in PLSR analysis (modified according to Wold et al. [2001]).*

6.1.2 Support Vector Machine Regression

Support vector machines (SVMs) can be traced back to statistical learning theory, which was mainly developed by Vapnik and Chervonenkis [1974] and Vapnik [1995]. In general, learning theory describes the process of characterizing properties by learning machines, which enables them to classify or predict unknown data. Originally, SVM classifiers were used in optical character recognition and later, they were applied to regression problems [Smola and Schölkopf, 2004]. For some years, SVMs have also been adapted to solve classification and regression problems in the field of remote sensing. Karimi et al. [2008], Siegmann et al. [2013] and Tuia et al. [2011], for example, showed the potential of this regression technique for the estimation of plant parameters in agriculture.

In this thesis, standard ε-SVR models with a Gaussian radial basis function (RBF) kernel were built using the R package 'kernlab' [Karatzoglou et al., 2004]. In accordance with the notation used by Camps-Valls et al. [2006] and Heinert [2010], SVR is based on the following theoretical background. An input data set consists of a certain number of samples, each of which is characterized by the independent variable $x_i \in \mathbb{R}^d$ (e.g., a spectral curve with a certain number of spectral bands) and the

corresponding dependent variable, $y_i \in \mathbb{R}$ (e.g., a LAI value). Firstly, the independent variable x_i is mapped onto a higher dimensional space $\mathcal{H}$ (kernel feature space) using the kernel trick:

$$x_i = \Phi(x_i). \qquad [6.11]$$

A detail description of the kernel trick is given, for example, in Smola and Schölkopf [2004]. In this way, it is possible to transfer a non-linear regression problem in the input feature space to a linear regression problem in the kernel feature space, which is easier to solve (Figure 6.3 (left and middle)).

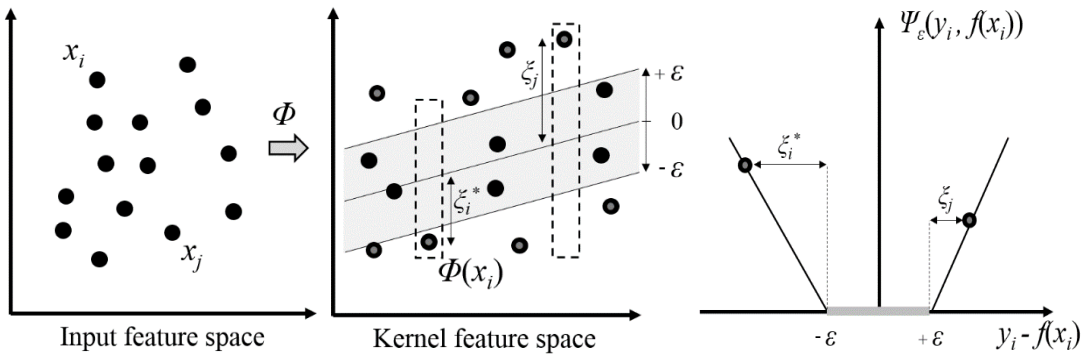


Figure 6.3: *Scheme of support vector regression. Data set in the input feature space (left), data mapped onto the kernel feature space (middle) and ε-insensitive cost function (right) (modified according to Camps-Valls et al. [2006] and Heinert [2010]).*

Then y_i can be determined as follows:

$$\hat{y}_i = f(x_i, w) = \Phi^T(x_i)w + b \qquad [6.12]$$

where w is a weight factor, b is a bias term and $\hat{y}_i$ is an estimate of y_i. Additionally, an error term e_i is introduced, which corresponds to the residuals of the model ($\hat{y}_i = y_i + e_i$). In standard ε-SVR, the goal is to minimize residuals e_i by using the square norm of the model coefficients:

$$F(w, e_i) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \Psi(e_i) \ . \qquad [6.13]$$

In SVR, the hyperplane, which is used in support vector classification for separating two classes, must be modified. All samples located on both sides of the hyperplane in the range $-\varepsilon...+\varepsilon$ are not involved in the optimization process. For all samples outside of the range, non-negative slack variables ξ_i and ξ_i^* are determined (Figure 6.3 (middle and right)). This results in an ε-insensitive function Ψ_ε, such that the loss function has to be redefined as follows:

$$\Phi(w,\xi_i,\xi_i^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad [6.14]$$

with respect to w, b, ξ_i, and ξ_i^* constrained to

$$y_i - \Phi^T(x_i)w - b \leq \varepsilon + \xi_i \quad \forall\, i = 1...n \qquad [6.15]$$

$$\Phi^T(x_i)w - b - y_i \leq \varepsilon + \xi_i^* \quad \forall\, i = 1...n \qquad [6.16]$$

$$\xi_i, \xi_i^* \geq 0 \quad \forall\, i = 1...n\,. \qquad [6.17]$$

The introduced constant C ($C > 0$) can be regarded as a regularization parameter, which determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated [Smola and Schölkopf, 2004]. The minimization of the loss function (Equation 6.14) can equally be solved by the maximization of the SVR function:

$$Q(\alpha_i,\alpha_i^*) = -\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{(i,j)=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\,K(x_i,x_j) \qquad [6.18]$$

under the conditions

$$\sum_{i=1}^{n}\alpha_i = \sum_{i=1}^{n}\alpha_i^* \quad \text{and} \quad 0 \leq \alpha_i,\alpha_i^* \leq C \quad \forall\, i = 1...n\,. \qquad [6.19]$$

The solution of the regression problem for any x is then:

$$f(x) = \sum_{i=1}^{n} \left(\alpha_i - \alpha_i^*\right) K\left(x, x_i\right) + b \qquad [6.20]$$

where K is the kernel, and α_i and α_i^* are the Lagrange multipliers, which can be interpreted as the contribution of the i-th example of the input data to the ultimate solution of w. In the process of model building, the parameters ε, ξ_i and ξ_i^* must be chosen in accordance with C, in order to reduce the number of support vectors and, therefore, the model complexity. In order to find the optimal parameter values for C and g (kernel parameter), the SVR implementation in the R package 'kernlab' uses a grid search technique based on an implementation in 'libsvm' [Chang and Lin, 2011].

6.1.3 Random Forest Regression

RFR was realized with the R package 'randomForest', which contains a direct implementation of Breiman's random forest algorithm for classification and regression problems [Breiman, 2001]. To date, only a few studies have demonstrated the potential of this regression algorithm to derive plant parameters from remote sensing data [Mutanga et al.2012; Powell et al., 2010; Vuolo et al., 2013].

The random forest method was originally developed to solve classification problems. It belongs to the ensemble learning methods, where many 'weak' classifiers (in this case single trees) are aggregated into a combined classifier (in this case a forest) with improved predictive power [Liaw and Wiener, 2002]. Along with the classification of categorical variables, random forest can also be used for regression applications to predict continuous variables [Breiman, 2001].

Each regression tree of a random forest is built with two thirds of the entire data set (bootstrap samples), which are randomly selected with replacement. For growing a single regression tree the CART (classification and regression tree) method, developed by Breiman et al. [1984], is used. According to Hastie et al. [2009], the basic functioning of CART can be described as follows. In a binary case, the continuous dependent variable Y (e.g., LAI) is described by two independent variables, X_1 and X_2 (e.g., two spectral bands). In Figure 6.4 (left), the feature space is divided into rectangles, with a simple model fitted into each rectangle. Thus, it is possible to describe Y in each rectangle by using different constants. Firstly, the feature space is partitioned into two regions, while the output is modelled by the mean of Y in each region. Afterwards, both regions are each split into two more regions. This process is repeated until a stopping criterion is reached. The split values t_1 to t_5 are chosen in such a way that an optimal separation is achieved. Figure 6.4 (right) shows the same model (as in Figure 6.4 (left)) in the form of a tree, which divides the feature space by successive binary decisions. In this context, observations that meet the conditions of a

node are assigned to the left branch, while observations that do not meet the conditions are assigned to the right branch. The terminal nodes (or leaves of the tree) represent the regions R_1 to R_5 (rectangles) in Figure 6.4 (left).

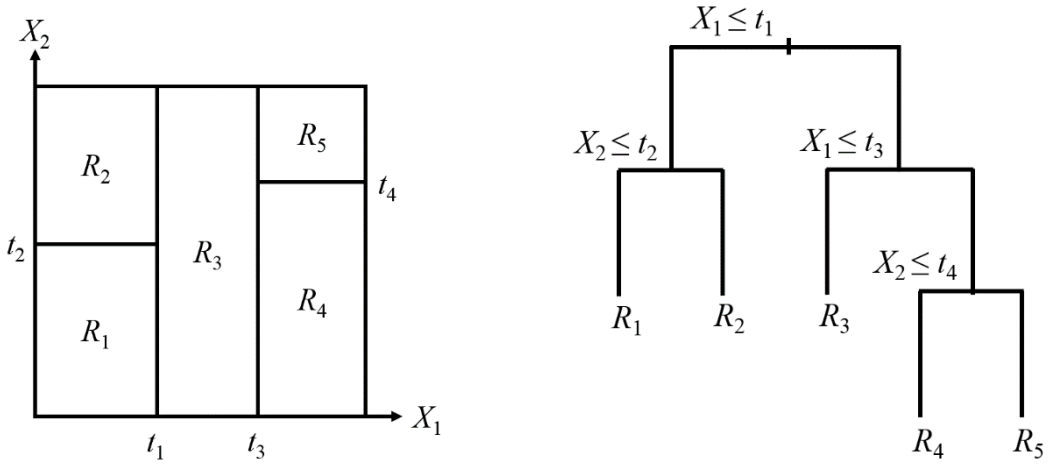


Figure 6.4: *Partitions in CART. Recursive binary splitting of a two-dimensional feature space (left) and tree-based recursive binary splitting (right) (Hastie et al. [2009]).*

One of the most important steps in growing a regression tree is making a decision at each interior node, which divides the data set into two groups. For this reason, a splitting rule must be defined. The aim is to partition the data set into m regions (R_1, R_2,…, R_m), which correspond to the terminal nodes of the tree (LAI values of the training samples). Therefore, a constant term c_m (Equation 6.21) has to be modelled in each region. If the minimization of the sum of squares is used as criterion for binary splitting, then c_m corresponds to the mean response value (e.g., LAI) of a considered region R_m.

$$c_m = \frac{\sum_i y_i I\,(x \in R_m)}{\sum_i I\,(x \in R_m)} \qquad [6.21]$$

For example, at the first node, the mean of all dependent variables (LAI values) of the training data set is used as c_m. Subsequently, the optimal splitting variable j (spectral band) and split point s (reflection in spectral band j) have to be defined. For each splitting variable j, the optimal value s can be determined by:

$$min_{j,s}\left[min_{c_1}\sum_{x_i \in R_1(j,s)}\left(y_i - c_1\right)^2 + min_{c_2}\sum_{x_i \in R_2(j,s)}\left(y_i - c_2\right)^2\right] . \qquad [6.22]$$

This procedure is repeated at every node until the tree is fully grown. At the end, every terminal node represents one of the target regions R_m.

For the assessment of LAI in this thesis, each RFR model was made up of 500 individual trees. At each node of a tree, only a few variables (square root of the number of all the spectral bands) were randomly selected with replacement for decision-making. As already mentioned above, each tree was built with two thirds of the training data, which were randomly selected (bootstrap samples). The remaining third (the out-of-bag samples) were used for the internal validation of each tree. Hence, the error rates of the single trees enabled the determination of an error rate of the entire random forest. After building a random forest, it is possible to predict the target variable (e.g., LAI) from unknown data samples (unknown spectral reflectance measurements). In this context, the predicted values from each tree are averaged to generate the result of the entire forest.

6.1.4 Evaluation of Empirical-Statistical Regression Models

As criteria for regression model accuracy, the coefficient of determination (R^2) and RMSE were calculated. Moreover, the residual prediction deviation (RPD) was computed as an additional indicator of the robustness of the regression models. The RPD represents the ratio of the standard deviation of the observed LAI and the RMSE of the predictions [Malley et al., 2004]. Williams [2001] developed a RPD scale based on high-resolution NIR laboratory spectra with different value classes for applications in food chemistry. Due to the lower spectral resolution of the data used in this study, a RPD scale with the following value limits, which was suggested by Dunn et al. [2002], was instead applied to evaluate the robustness of the regression models: RPD < 1.6 is a poor model; RPD = 1.6−2.0 is an acceptable model; RPD > 2.0 is an excellent model.

6.2 Feature Selection

The high spectral dimensionality of hyperspectral data is one of the major problems in the process of deriving soil or plant parameters from remote sensing images. This phenomenon is known as the curse of dimensionality [Bellman, 1962]. It describes the exponential growth in the number of examples required to maintain a given sampling strategy on the one hand, and the exponential growth in the complexity of the target function with increasing dimensionality on the other hand. For that reason, the selection of the most important features (spectral bands) is a crucial processing step when using regression algorithms to estimate biophysical and -chemical parameters from hyperspectral data [Camps-Valls et al., 2011].

The subsequent subsections first give a detailed overview of the RReliefF algorithm, which belongs to the group of filter methods within the family of feature selection techniques. Following this, a randomized sampling approach is explained, which enables the validation of the results provided by RReliefF.

6.2.1 RReliefF

The Relief algorithm belongs to the filter methods used for feature selection in different fields of science. The basic Relief idea was invented by Kira and Rendell [1992] and was solely designed to solve classification problems with two classes. Later, the algorithm was extended to ReliefF by Kononenko [1994] in order to deal with multi-class classification problems, as well as incomplete and noisy data sets. Due to its characteristics, ReliefF is well suited to classification problems in remote sensing, such that it has been used in some remote sensing studies to detect the optimal spectral bands to separate multiple classes [Ghosh et al., 2013; Wu et al., 2013]. In contrast to classification problems, where discrete classes are available, the estimated values are continuous in regression scenarios. Therefore, Robnik-Šikonja and Kononenko [1997, 2003] further developed the algorithm to RReliefF (Regressional ReliefF), which is able to estimate the quality of features in regression problems.

1. set all weights $W[B] := 0.0$;
2. **for** i := 1 **to** m **do begin**
3. randomly select an instance T_i;
4. find n nearest hits H_j;
5. **for** each class $C \neq class(T_i)$ **do**
6. from class C find n nearest misses $M_j(C)$;
7. **for** $B := 1$ **to** b **do**
8. $W[B] := W[B] - \sum_{j=1}^{n} diff(B,T_i,H_j)/(m \cdot n) +$
9. $\sum_{C \neq class(T_i)} \left[\frac{P(C)}{1-(class(T_i))} \sum_{j=1}^{n} diff(A,T_i,M_j(C)) \right] / (m \cdot n)$;
10. **end**;

Figure 6.5: *Pseudocode of the ReliefF algorithm [Robnik-Šikonja and Kononenko, 2003].*

The ReliefF algorithm for multi-class problems randomly selects an instance (T_i), which corresponds to a spectral curve of a certain class in a remote sensing data set. Afterwards, the algorithm searches the n nearest neighbours of the same class (nearest

hits – H_j) and the n nearest neighbours of different classes (nearest misses – M_j), which it does by calculating the summarized Manhattan distance over all features (spectral bands - B) between the selected instance (T_i) and all other instances of the data set. The calculated distances of the n hits and the n misses of an instance are then averaged, with the n misses additionally weighted with a prior estimated probability of the corresponding class ($P(C)$). The already calculated distances for the single features (B) are then used to determine the importance of every feature regarding its suitability for class separability. The whole procedure is repeated m times (user defined parameter) to update the quality estimation $W[B]$ (depending on T_i, H_j and M_j) and finally ascertain the weight factors of all features. The pseudocode of ReliefF is shown in Figure 6.5.

Due to the lack of classes and thus, not clear definable hits and misses, ReliefF cannot be transferred to regression problems without adjustments. Therefore, in RReliefF, the following probabilities have to be determined to distinguish whether the values of two instances (T_i) are different. First, the probability of a feature B of an instance T_i differs from the same feature B of its nearest neighbour instance (I_j) (Equation 6.23). Second, the probability of a predicted value τ (e.g., LAI) of an instance T_i differs from its nearest neighbour instance (I_j) (Equation 6.24). Third, the probability of a predicted value τ, as well as feature B, differs from its nearest neighbour instance (I_j) (Equation 6.25).

$$P_{diffB} = P(\text{different value of } B \mid \text{nearest instances } I_j) \quad [6.23]$$

$$P_{diffC} = P(\text{different prediction } \tau \mid \text{nearest instances } I_j) \quad [6.24]$$

$$P_{diffC|diffB} = P(\text{different prediction } \tau \mid \text{different value of } B \text{ and nearest instances } I_j) \quad [6.25]$$

Finally, the different probabilities are merged in Equation 6.26, with Bayes' rule used for estimating the weight factor of a feature $W[B]$:

$$W[B] = \frac{P_{diffC|diffB} \; P_{diffB}}{P_{diffC}} - \frac{P_{diffC|diffB} \; P_{diffB}}{1 - P_{diffC}} \quad [6.26]$$

Analogous to ReliefF, the instances (T_i) were chosen randomly, whereas the n nearest neighbour instances (I_j), as well as the number of repetitions (m), were defined by the user. Figure 6.6 illustrates a summary of RReliefF in the form of a pseudocode

of the algorithm. The weight factors of the different prediction values τ (line 6), the different features B (line 8), and the different predictions values and features (line 9) are collected in N_{dC}, $N_{dB}[B]$ and $N_{dC\&dB}[B]$, respectively. Additionally, a weighting term $d(i,j)$ was introduced, taking into account the influence of the n nearest neighbour instances (I_j) regarding their different distances to T_i. Finally, the estimation of the weight of each feature $W[B]$ can be calculated using Equation 6.26 (line 14).

In this thesis, RReliefF was applied to field spectral measurements and the corresponding measured LAI in order to identify the spectral bands that provide the most important information to predict the LAI. For this purpose, the R software environment was used again with the additional 'FSelector' package, which holds an implementation of the RReliefF algorithm [Romanski and Kotthoff, 2014]. The number of the n nearest neighbours was set to ten and the entire procedure was repeated 100 times (m). The determined weighting factors for every spectral band from each run were stored and averaged in order to ascertain the final importance of every spectral band.

1. set all N_{dC}, $N_{dB}[B]$, $N_{dC\&dB}[B]$, $W[B] := 0.0$;
2. for i := 1 to m do begin
3. randomly select an instance T_i;
4. select n instances Ij nearest to Ti;
5. for $j := 1$ to n do begin
6. $N_{dC} := N_{dC} + \text{diff}(\tau,T_i,I_j) \cdot d(i,j)$;
7. for $B := 1$ to b do begin
8. $N_{dB}[B] := N_{dB}[B] + \text{diff}(B,T_i,I_j) \cdot d(i,j)$;
9. $N_{dC\&dB}[B] := N_{dC\&dB}[B] + \text{diff}(\tau,T_i,I_j) \cdot \text{diff}(B,T_i,I_j) \cdot d(i,j)$;
10. end;
11. end;
12. end;
13. for $B := 1$ to b do
14. $W[B] := N_{dC\&dB}[B] / N_{dC} - (N_{dB}[B] - N_{dC\&dB}[B]) / (m - N_{dC})$;

Figure 6.6: *Pseudo code of the RReliefF algorithm [Robnik-Šikonja and Kononenko, 2003].*

6.2.2 Evaluation of Feature Selection

In order to verify the spectral bands selected as the most important by RReliefF for LAI prediction, randomized experiments were conducted. In this context, numerous regression models were calibrated with spectral bands selected by simple random sampling with replacement. This method enabled the random selection of subsets of individuals (subsets of spectral bands) from the entire population (all available spectral bands) [Ardilly and Tillé, 2006]. In the end, the subset of spectral bands that provided the model with the highest R^2 was chosen and the spectral bands involved in regression model building were compared to those selected using RReliefF.

6.3 Pan-sharpening[11]

In remote sensing, image fusion techniques are frequently used to combine two or more different images acquired by the same or different sensors to form a new image with enhanced characteristics by using a certain algorithm [Jiang et al., 2011]. One of the most popular image fusion methods is known as pan-sharpening [Pohl and van Genderen, 2015]. This technique is commonly applied to merge the spatial details of a high-resolution panchromatic image and the spectral information of a low-resolution multi- or hyperspectral image, with the aim of generating a spatial high-resolution multi-or hyperspectral image (hybrid product) [Fonseca et al., 2011].

In this thesis, the Ehlers Fusion is applied to merge image data with high spatial and high spectral resolution to create a composite image with new qualities, which provides more information for the spatially accurate prediction of LAI than the original images. The theoretical background of this pan-sharpening technique is presented in the next section. Following this, the validation techniques are presented, which are used to evaluate the fusion results.

6.3.1 Ehlers Fusion

The Ehlers Fusion was developed specifically for a spectral characteristics preserving image merging [Klonus and Ehlers, 2007]. As it is based on an IHS (intensity, hue, saturation) transform, coupled with a Fourier domain filtering, it is representative of the hybrid image fusion methods [Pohl and van Genderen, 2015].

The principal idea behind a spectral characteristics preserving image fusion is that the high-resolution image has to sharpen the multi-/hyperspectral image without adding new grey level information to its spectral components. An ideal fusion algorithm would enhance high-frequency changes, such as edges and grey level discontinuities in an image, without altering the spectral components in homogeneous regions [Ehlers et al., 2010]. To facilitate these demands, two prerequisites have to be

[11] Parts of this section have previously been published by Siegmann et al. [2015].

addressed. First, spectral and spatial information need to be separated. Second, the spatial information content has to be manipulated in a way that allows an adaptive enhancement of the images. This is achieved by a combination of colour and Fourier transforms.

For optimal separation of spatial and spectral information from an RGB (red, green, blue) image, an IHS transform is used. In this regard, intensity *I* refers to the total brightness of the colour, hue *H* refers to the dominant or average wavelength of the light contributing to the colour, and saturation *S* refers to the purity of colour. This is comparable to human perception [Klonus and Ehlers, 2007]. According to Gonzalez and Woods [2007], the *H* component of each pixel of a given RGB image can be determined using the following equation:

$$H = \begin{cases} \theta & \text{if } B \leq G \\ \theta - 360 & \text{if } B > G \end{cases} \quad [6.27]$$

with:

$$\theta = \cos^{-1}\left\{ \frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(R - B)]^{1/2}} \right\}. \quad [6.28]$$

The *S* and the *I* components are calculated as follows:

$$S = 1 - \frac{3}{(R + B + G)} [\min (R, G, B)] \quad [6.29]$$

$$I = \frac{1}{3} (R + B + G) \quad [6.30]$$

This technique is extended to include more than three bands by using multiple IHS transforms until the number of bands of the multi-/hyperspectral image is exhausted. If the assumption of spectral characteristics preservation holds true, there is no dependency on the selection or order of bands for the IHS transform.

Subsequently, Fourier transforms of the intensity component, which contains the spatial information of the multi-/hyperspectral image, and the panchromatic image are conducted to transfer the images into the frequency domain. For that purpose, the two-dimensional fast Fourier transform (FFT) is used, which is given by the equation:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f(x,y)\, e^{-j2\pi(ux/M+vy/N)}\right] \quad [6.31]$$

where M is the number of horizontal image pixels, N is the number of vertical image pixels, x and y are the spatial variables, u and v are the frequency variables, and j represents the imaginary component of a complex number [Klonus, 2011]. Afterwards, the generated intensity power spectrum in the frequency domain is filtered with a Gaussian low-pass filter (Equation 6.32), whereas the power spectrum of the high-resolution panchromatic image is filtered with the inverse Gaussian high-pass filter (Equation 6.33).

$$L(u,v) = e^{-D^2(u,v)/2\sigma^2} \quad [6.32]$$

$$H(u,v) = 1 - e^{-D^2(u,v)/2\sigma^2} \quad [6.33]$$

In both filter functions, D represents the distance from the origin of the Fourier power spectrum and σ is a measure of the spread of the Gaussian curve [Gonzalez and Woods, 2007]. After filtering, the images are transformed back into the spatial domain with an inverse FFT using the following equation:

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\left[F(u,v)\, e^{j2\pi(ux/M+vy/N)}\right]. \quad [6.34]$$

As a next step, both images are added together to form a fused intensity component with the low-frequency information from the spatial low-resolution multi-/hyper-spectral image and the high-frequency information from the spatial high-resolution panchromatic image. This new intensity component has to be histogram matched to the original intensity component for mapping it into the value range of the original image. Afterwards, the fused and histogram-matched intensity component and the original hue and saturation components of the multi-/hyperspectral image form a new IHS image. Finally, an inverse IHS transform produces a fused RGB image, which contains the spatial resolution of the panchromatic image and the spectral characteristics of the multi-/hyperspectral image. Gonzalez and Woods [2007] provide a comprehensive description of the inverse IHS transform. At the beginning, the H

component is multiplied by 360 to return the hue to its original range (0°−360°). Subsequently, the RGB values are determined as follows:
If $0° \leq H < 120°$:

$$B = I\,(1 - S) \quad [6.35]$$

$$R = I\left[1 + \frac{S \cos H}{\cos(60° - H)}\right] \quad [6.36]$$

$$G = 3I - (R + B) \quad [6.37]$$

If $120° \leq H < 240°$:

$$H = H - 120° \quad [6.38]$$

Then the RGB components are:

$$R = I\,(1 - S) \quad [6.39]$$

$$G = I\left[1 + \frac{S \cos H}{\cos(60° - H)}\right] \quad [6.40]$$

$$B = 3I - (R + G). \quad [6.41]$$

If $240° \leq H < 360°$:

$$H = H - 240° \quad [6.42]$$

Then the RGB components are:

$$G = I\,(1 - S) \quad [6.43]$$

$$B = I\left[1 + \frac{S \cos H}{\cos(60° - H)}\right] \quad [6.44]$$

$$R = 3I - (G + B). \quad [6.45]$$

The different steps of the Ehlers Fusion can be repeated with successive three band selections until all bands of the multi-/hyperspectral image are fused with the panchromatic image. The order of bands and the inclusion of spectral bands for more than one IHS transform are not critical because of the colour preservation of the procedure [Ehlers et al., 2010]. The entire fusion process is presented in Figure 6.7.

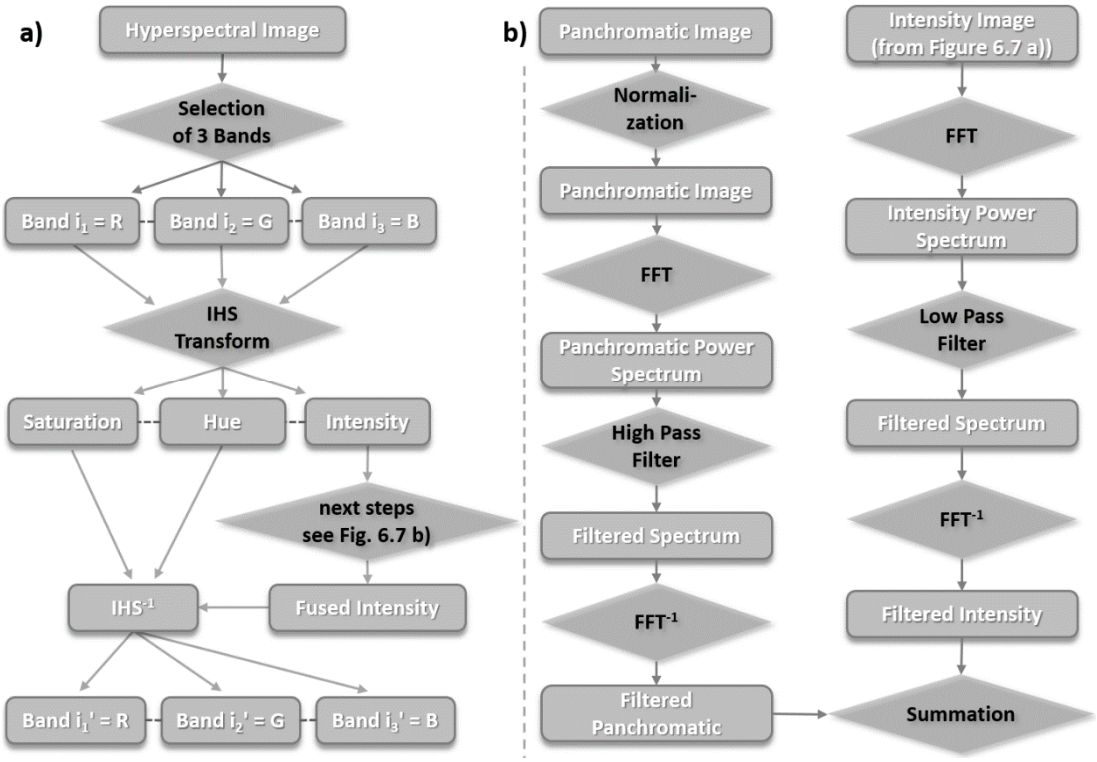


Figure 6.7: *Scheme of the Ehlers Fusion with i_1, i_2, $i_3 \in \{1, 2, \ldots, n\}$ (modified according to Klonus and Ehlers [2007]).*

6.3.2 Evaluation of Pan-sharpened Images

Since the visual interpretation of the fusion results can be considered as very subjective and always depends on the experiences of the human interpreter, two statistical evaluation criteria were calculated to measure the spectral preservation of the results. These methods are objective, quantitative and reproducible. First, the spectral angles (α_{spec}) between corresponding pixels of the original image data and the fusion results were determined [Kruse et al., 1993]. α_{spec} can be calculated in hyper-dimensional space for all bands of two pixels (two spectral curves) at once. An angle of zero indicates an ideal fusion result, while larger angles imply a poorer spectral preservation [Alparone et al., 2007]. α_{spec} was calculated (per pixel) to determine the spectral performance achieved by the Ehlers Fusion in order to identify spatial differences of the spectral preservation in the fusion results.

$$\alpha_{spec} = cos^{-1}\left(\frac{\sum_{i=1}^{n} t_i\, r_i}{\sqrt{\sum_{i=1}^{n} t_i^2}\ \sqrt{\sum_{i=1}^{n} r_i^2}}\right) \qquad [6.46]$$

where n is the number of spectral bands, t_i is the fused spectral curve and r_i is the original spectral curve.

As a second evaluation criterion, the correlation coefficient (R) was calculated for corresponding bands of the original data and the fusion results. In contrast to αspec, R was determined to evaluate the goodness of spectral preservation achieved for every single band. R has a value range from -1 to 1, where 1 indicates a perfect match for two compared spectral bands.

$$R = \left(\frac{\sum_i^n (t_i - \mu_t)\ (r_i - \mu_r)}{\sqrt{\sum_i^n t_i^2}\ \sqrt{\sum_i^n r_i^2}} \right) \qquad [6.47]$$

where n is the number of spectral bands, t_i is the fused spectral curve, r_i is the original spectral curve and $\mu_{t/r}$ is the mean of t or r.

7 Results and Discussion

In this chapter, the results obtained in the course of this thesis are presented and discussed. To this extent, different research experiments were conducted to answer the three research question developed in Chapter 1. The different tasks formulated with regard to each research question form the thread for the following subchapters. First, in Chapter 7.1, different empirical-statistical regression models are evaluated regarding their suitability for an accurate LAI prediction from field reflectance measurements. Then, in Chapter 7.2 and Chapter 7.3, the spectral and spatial properties of hyperspectral image data acquired from different platforms and scales are investigated in detail to identify the potential and the limitations of these types of data for the spatial prediction of LAI in precision agriculture.

7.1 Quality Assessment of Empirical-Statistical Regression Techniques[12]

Empirical-statistical regression models are frequently applied techniques to estimate plant or soil parameters from remote sensing data. Although numerous studies have used these kinds of models to derive the LAI and other parameters, model validation aimed at verifying the prediction accuracy, as well as model transferability to unknown data, has been investigated insufficiently to date. For that reason, in the next sections, the research question, provided below, ought to be answered:

What potential do empirical-statistical regression models have for an accurate prediction of crop LAI based on hyperspectral data?

7.1.1 Influence of Validation Procedure on Regression Model Performance

As a first step, PLSR, SVR and RFR models were separately generated for the data from the two individual years (2011 and 2012). In all cases, model validation was performed by leave-one-out cross-validation (cv), which means that each sample was estimated by a regression model that was calibrated using the remaining (n – 1) samples [Otto, 2007]. SVR provided the best results for the separate analysis of 2011 and 2012 data, with R^2_{cv} values of 0.739 and 0.850, respectively; meanwhile, $RMSE_{cv}$ was relatively low at 0.242 and 0.376. A reason for the higher $RMSE_{cv}$ value in 2012 was the wider range of measured LAI in the field, which is typical for subsequent stages of canopy development. PLSR yielded lower R^2_{cv} values (R^2_{cv} = 0.696 (LV = 6) in 2011, R^2_{cv} = 0.790 (LV = 5) in 2012) than SVR. However, the $RMSE_{cv}$ of PLSR ($RMSE_{cv}$ = 0.244 in 2011, $RMSE_{cv}$ = 0.408 in 2012) was comparable to that of SVR.

[12] This section has previously been published by Siegmann and Jarmer [2015].

In contrast, RFR showed significantly lower model accuracy compared to SVR and PLSR. Particularly, in 2011, it was not possible to build a robust model with RFR (R^2_{cv} = 0.420, $RMSE_{cv}$ = 0.337). In 2012, however, RFR provided much better results (R^2_{cv} = 0.715, $RMSE_{cv}$ = 0.476), but still showed lower model performance than SVR and PLSR. RPD_{cv} values calculated for SVR (2.709 for 2011 and 2.383 for 2012) and PLSR (2.668 and 2.196, respectively) were all higher than 2.0. According to Dunn et al. [2002], regression models with RPD values higher than 2.0 are well suited, while values between 1.6 and 2.0, like those of RFR for both years, can only be regarded as acceptable. Table 7.1 summarizes the results of the cross-validated models using different regression algorithms.

Table 7.1: *Cross-validated results of LAI prediction with PLSR, SVR and RFR, for 2011 and 2012.*

n = 57 (2011) *n* = 67 (2012)	R^2_{cv}		$RMSE_{cv}$		RPD_{cv}	
	2011	2012	2011	2012	2011	2012
PLSR	0.696	0.790	0.244	0.408	2.668	2.196
SVR	0.739	0.850	0.242	0.376	2.709	2.383
RFR	0.420	0.715	0.337	0.476	1.943	1.884

The scatter plots in Figure 7.1 underline the results of the regression models based on their statistical values. More robust models for predicting the LAI could be built with PLSR and SVR. Although PLSR models had lower R^2_{cv} values in both years, observed offsets were smaller compared to SVR models and the regression lines were closer to the 1:1 line. This fact indicates a more reliable model performance of PLSR in comparison to SVR. RFR scatter plots suggested lower model accuracies in relation to the other algorithms. The scatter plot for 2011 particularly illustrated a poor relationship between measured and predicted LAI values. In general, the model performance of all regression algorithms was much better in 2012. The later phenological stage of wheat development results in a wider range of measured LAI, which is one reason for higher model accuracy. Furthermore, the spectral field measurements were less influenced by soil reflectance because of the denser vegetation coverage in 2012. This fact may be another reason for better model performance. Additionally, the overestimation of lower and the underestimation of higher values by the algorithms are problematic and occurred in the scatter plots of all models.

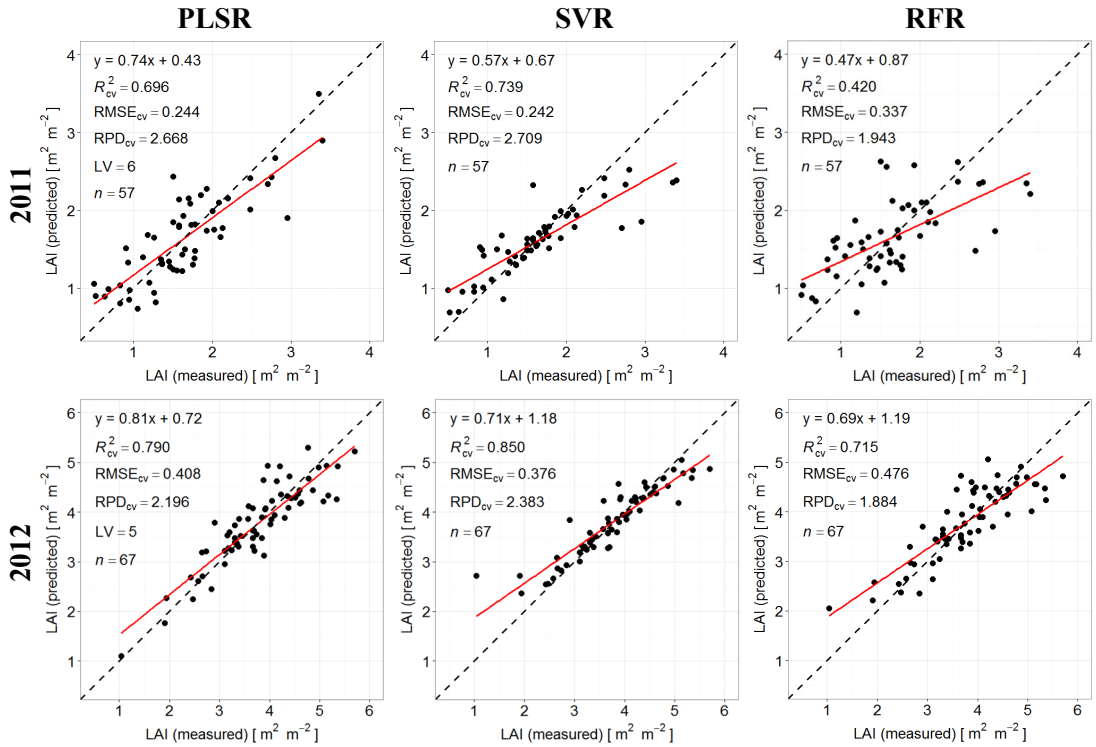


Figure 7.1: *Scatter plots of predicted (cross-validated) and measured LAI values for winter wheat, based on PLSR, SVR and RFR, for 2011 and 2012. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

As a second step, leave-one-out cross-validation was conducted to determine model quality for the entire data set (for 2011 and 2012). Table 7.2 summarizes the results for the models obtained by the different regression algorithms. PLSR and SVR provided results with high R^2_{cv} values of 0.901 (LV = 5) and 0.944, respectively. The relatively low $RMSE_{cv}$ of 0.417 (PLSR) and 0.322 (SVR), together with RPD_{cv} values well over 3.0 for PLSR, and even over 4.0 for SVR, indicated the robustness and successful calibration of both models [Dunn et al., 2002]. Although RFR also provided high model accuracy, R^2_{cv} was distinctly lower (R^2_{cv} = 0.799) when compared to PLSR and SVR. In addition, an $RMSE_{cv}$ of 0.597 and an RPD_{cv} of 2.222 showed distinctly higher and lower values, respectively, in contrast to the other models.

Figure 7.2 illustrates the scatter plots of measured and corresponding estimated LAI. For PLSR and SVR, observed offsets in LAI estimation were negligible, while RFR clearly showed an overestimation of smaller and an underestimation of higher LAI. RFR also had a higher scattering of sample points compared to the other models. The differences between PLSR and SVR accuracy were also obvious in the scatter plots. While offset and slope of the regression line had slightly better values for PLSR concerning the fit to the 1:1 line, the scattering was slightly higher compared to the SVR model, especially for higher values.

Table 7.2: *Cross-validated results of LAI prediction with PLSR, SVR and RFR (2011 and 2012).*

$n = 124$	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}
PLSR	0.901	0.417	3.190
SVR	0.944	0.322	4.134
RFR	0.799	0.597	2.222

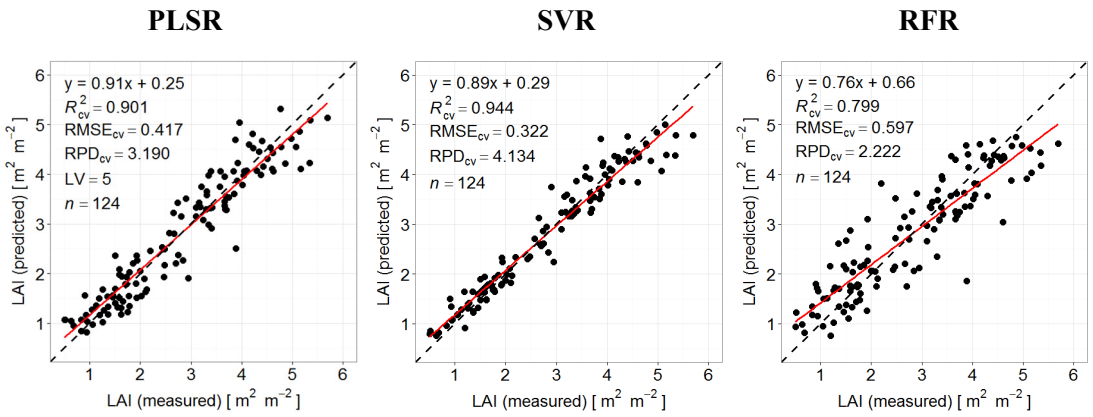


Figure 7.2: *Scatter plots of predicted (cross-validated) and measured LAI for winter wheat, based on PLSR, SVR and RFR, for 2011 and 2012. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

In general, similar to the results from the separate data sets already highlighted, all models more or less overestimated lower values (LAI ≤ 2), while higher values (LAI ≥ 4) were underestimated. Furthermore, the scatter plots indicated an under-representation of LAI at around 3, which can be ascribed to the fact that LAI was only measured in the phenological stages of wheat plant development (BBCH 3 in 2011; BBCH 5 in 2012), where lower or higher LAI dominated.

Independent validation (iv) can be considered the most robust method for assessing regression model accuracy, given that validation data are not involved in the process of model development. Nevertheless, it is common to carry out cross-validation when data division is not possible because of the inadequate availability of samples, as in the case of the separate years in the context of this thesis. Although the entire data set had a relatively high number of samples, leave-one-out cross validation was performed to compare the results with those from independent validation presented below.

To prepare the data for independent validation, the 124 LAI values were arranged in ascending order according to their size. Finally, the data set was split into three subsets by choosing every third value for each subset. Thus, three approximately equal-sized data subsets were generated (subset A with 42 samples, subsets B and C both with 41 samples), each of which almost covered the range of values measured in

2011 and 2012. Subsequently, the corresponding field spectral measurements were arranged into the corresponding three subsets. For independent validation, each subset of the data set was used once for validation, while the models were built using the samples from the other two subsets.

Table 7.3 shows the independently validated results for the different regression algorithms. For the three models based on the three different subsets, PLSR provided very reliable results, with R^2_{iv} values around 0.9 (mean R^2_{iv} = 0.912, LV for subsets A/B/C = 5/5/5) and $RMSE_{iv}$ of about 0.4 (mean $RMSE_{iv}$ = 0.402). The high R^2_{iv} and the RPD_{iv} of well over 3.0 (mean RPD = 3.336) were indicators of successful model calibrations [Dunn et al., 2002]. SVR (mean R^2_{iv} = 0.769, mean $RMSE_{iv}$ = 0.645) and RFR (mean R^2_{iv} = 0.770, mean $RMSE_{iv}$ = 0.643) prediction accuracies were comparable to each other, but much lower than PLSR prediction accuracy. Additionally, the mean RPD_{iv} for both regression algorithms was just over 2.0, suggesting the lower robustness of these models.

Table 7.3: *Independent validated results of LAI prediction, with PLSR, SVR and RFR for the data subsets A, B, and C (ncal = 83/82, nval = 42/41).*

	R^2_{iv}				$RMSE_{iv}$				RPD_{iv}			
	A	B	C	Mean	A	B	C	Mean	A	B	C	Mean
PLSR	0.898	0.924	0.913	0.912	0.450	0.367	0.396	0.402	3.067	3.595	3.346	3.336
SVR	0.708	0.796	0.802	0.769	0.757	0.592	0.585	0.645	1.821	2.231	2.268	2.107
RFR	0.635	0.839	0.835	0.770	0.842	0.546	0.540	0.643	1.638	2.418	2.457	2.171

A general trend can be seen in the independent validation of all regression models. While using data subsets B and C as independent validation data led to nearly identical results in model internal comparison, there seemed to be a deviation in data subset A. The validation accuracy for PLSR with data subset A (R^2_{iv} = 0.898) was almost negligibly lower compared to data subsets B (R^2_{iv} = 0.924) and C (R^2_{iv} = 0.913). Clearer differences in model performance could be detected by comparing SVR and RFR validation results of data set A (SVR: R^2_{iv} = 0.708, RFR: R^2_{iv} = 0.635) to data subsets B (SVR: R^2_{iv} = 0.796, RFR: R^2_{iv} = 0.839) and C (SVR: R^2_{iv} = 0.802, RFR: R^2_{iv} = 0.835), respectively.

The scatter plots for the independent validated data subsets A, B and C, shown in Figure 7.3, illustrate the same trend of overestimating lower and underestimating higher values already observed in the scatter plots of the cross-validated regression models (see Figure 7.1 and Figure 7.2). PLSR provided nearly identical results to those achieved with the cross-validated model using the entire data set and the regression lines fit close to the 1:1 line, which is an indicator of a robust and well-calibrated model. In contrast, the scatter plots for independently validated SVR and RFR showed

a larger offset and a more pronounced scattering. In particular, the aforementioned problem using subset A for validation was highlighted by the scatter plots of SVR and RFR, where the regression lines substantially deviated from the 1:1 line.

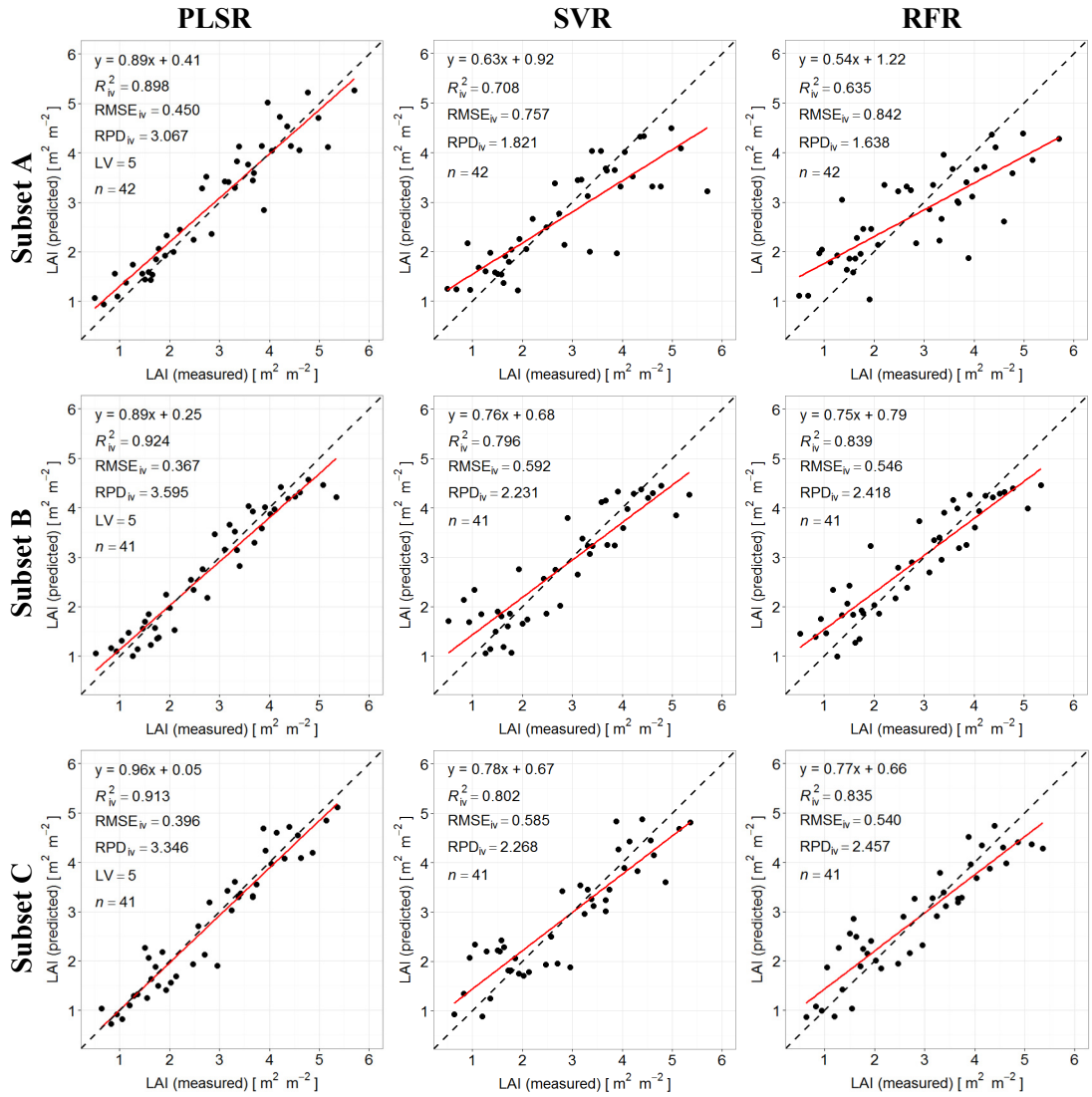


Figure 7.3: *Scatter plots of estimated (independently validated) and measured LAI for winter wheat, based on PLSR, SVR and RFR, for the different calibration and validation data subsets (A, B and C). The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

7.1.2 Comparison of Regression Model Results

The individual data sets of the separate years consisted of a relatively low number of samples (57 samples in 2011, 67 samples in 2012). Consequently, leave-one-out cross-validation was performed to determine model quality. In comparison to the cross-validated regression models built with the data of both years, model accuracies

for the single-year data sets were distinctly lower for all considered regression algorithms (Table 7.1 and Table 7.2). This was mainly due to the fact that using the entire data set (2011 and 2012) increased the number of samples and enlarged the LAI range by including LAI measurements of different phenological stages. This is in accordance with studies by Haboudane et al. [2004], Herrmann et al. [2011], and Nguy-Robertson et al. [2014], all of whom also achieved better correlations between VIs and LAI measurements using data of different phenological stages during the growing season. In addition, the scatter plots for the regression models, when incorporating the data of both years, illustrated a substantial increase in model robustness, less scattering and a better fit of the regression line to the 1:1 line in relation to the scatter plots of the single-year regression models (cf. Figure 7.1 and Figure 7.2). Furthermore, the order of regression algorithm performance was equal in both the single-year and complete data sets. In each case, SVR provided the best results, closely followed by PLSR and then RFR, albeit with distinctly lower model performance.

The subsequent independent validation of the complete data set led to a very different result. For all validation data subsets, PLSR produced the most robust models, with R^2_{iv} and RPD_{iv} values comparable to those of the cross-validation. In contrast, independently validated SVR models provided much lower accuracies, which were on the same level as the RFR results (cf. Table 7.3). While the model performances of PLSR and RFR were almost identical for cross- and independent validation of the combined data set, SVR showed a significant decline in R^2_{iv} and RPD_{iv}, as well as an extensive increase in $RMSE_{iv}$ at the same time. For the data used in this thesis, it seemed that SVR models were overfitted with respect to the training data. In the case of cross-validation, this led to very high model accuracy but, at the same time, an unknown sample could not be predicted with the same high accuracy, which was shown in the case of independent validation. The same problem of overfitting was not detected for PLSR and RFR. Both regression algorithms produced almost stable accuracy for both validation strategies, with PLSR providing results at a substantially higher level. The better and more stable results of PLSR, compared to SVR, differed from recent remote-sensing studies, where machine learning approaches, such as SVR, have outperformed PLSR, or have at least provided comparable results for the retrieval of biophysical and -chemical parameters from agricultural fields [e.g., Arenas-Garcia and Camps-Valls, 2008; Wang et al., 2011].

Obviously, the validation technique for verifying the performance of a regression model is very important. The results showed that PLSR and RFR performance was less sensitive to the validation technique used. In contrast, SVR model quality depended more on the type of validation. Independent validation was the more robust approach for verifying regression models because training and validation data are statistically independent from each other. Thus, model overfitting during calibration can be prevented.

In recent years, several different validation techniques have been developed. For example, Wang et al. [2011] randomly divided their data into two thirds for calibration and one third for validating regression models. Another method for independent validation was presented by Tuia et al. [2011], who conducted two experiments by randomly splitting the data into 80 and 50 % for calibrating the model and validation, respectively. Each experiment was repeated 40 times, with the results averaged to avoid skewed data distribution. In this thesis, a random division of calibration and validation data was not planned. Instead, the LAI values were arranged in ascending order according to their size, while every third value was chosen for one of the three resulting data subsets. Therefore, the data range and distribution of LAI values in each subset were very similar, which is a key requirement because statistically robust model validation can only be conducted on independent data that cover the same data range as used for model calibration. In this context, it was not possible to build regression models with the data set acquired in 2011 and predict the data of 2012, and vice versa, because the value range covered for the two years was different. The LAI range covered by the LAI data from both years was more appropriate for model building to predict LAI from unknown spectral data. However, in order to build a 'universal model', the LAI measurements of earlier (BBCH 1–2), later (BBCH 6–7) and phenological stages in between (BBCH 4) also had to be integrated into model calibration to cover the entire range of possible LAI during the growing season.

Additionally, the spatial domain plays an important role in calibrating and validating regression models. The data used in this thesis were acquired from four different agricultural fields within a region and mixed to build independent data subsets for model calibration and validation. Consequently, the samples were not spatially independent. Therefore, further research is required to investigate an approach that is spatially more independent. In this context, regression models have to be calibrated with data collected from one field. Subsequently, the model validation can be conducted on data from a neighbouring field with the same crop. This approach can help in the development of spatially independent regression models that are transferable within a region.

Besides the validation procedure, which was applied for assessing regression model quality, the use of different spectroradiometers (ASD FieldSpec III and SVC 1024) and LAI measuring devices (LAI-2000 and SunScan) can have an influence on the regression results. In order to ensure comparable spectral measurements, all instruments were calibrated by the manufacturers less than one year before the field campaigns and the spectral data were converted to absolute reflectance values using reflectance standards of known reflectivity. Additionally, the spectral measurements were acquired at the same height above ground in both years. Nevertheless, the divergent FOV of the spectroradiometers used, as well as the different plant heights in 2011 and 2012, may have had some influence on the spectral measurements. The

different LAI devices used in 2011 (SunScan) and 2012 (LAI-2000) were also calibrated by the manufacturers. Moreover, according to Wilhelm et al. [2000], comparable measurements are possible, as long as the operator follows the procedures described in the user manuals. Thus, the use of different measuring devices as an influencing factor on the regression results can be reduced to a minimum.

Furthermore, multicollinearity is a known problem using regression analysis applied to hyperspectral data sets. This can also have an impact on model calibration, validation and, in turn, on model prediction. The best results in the case of independent validation, however, were achieved for PLSR, which is less influenced by this problem because the high number of spectral bands is reduced to a distinctly lower number of uncorrelated LVs, representing an orthogonal linear combination of the original spectral bands [Atzberger et al., 2010; Yu et al., 2015].

The next section addresses the problem of multicollinearity when using hyperspectral data. In this regard, the influence of the spectral resolution on regression model performance to predict the LAI was investigated in greater detail.

7.2 Influence of Spectral Resolution on LAI Determination

Hyperspectral sensors often have much more than 100 spectral bands and a typical bandwidth of a few tens of nanometres or even less [Baltsavias, 2002]. The high spectral resolution in combination with a wide spectral range make these sensors ideally suited to numerous applications. The hyperspectral EnMAP mission, for example, has the goal to provide valuable information for different research fields, including agriculture, forestry, natural ecosystems, geology, soil science, coastal and inland waters, and urban areas [Guanter et al., 2015]. For a specific application, however, such as the assessment of LAI, not all the spectral information is necessary to enable a robust regression model calibration [Thenkabail et al., 2004]. This fact leads to the following research question, which should be answered:

What spectral resolution and location of spectral bands are necessary to enable a precise estimation of the parameter LAI?

7.2.1 LAI Determination Based on aisaDUAL and EnMAP Spectral Resolution

Since useful image data from aisaDUAL, EnMAP (simulated) and Semtinel-2 (simulated) were only available for Field A (2011) and Field D (2012) (cf. Chapter 5.2.1), the total number of field spectral reflectance and corresponding LAI measurements (n = 124) was reduced to the number of samples collected from those fields (n = 73). This was a necessary step to make the in situ data set comparable to the image data sets for the ensuing course of the investigations. Moreover, one sampled

plot of Field A and two sampled plots of Field D were also excluded, due to their location at the edge of the fields and the resulting problem of mixed pixels in the EnMAP image data. Thus, 70 samples were used in further analysis. Since both fields were investigated during different stages of plant development (Field A: BBCH 3, Field D: BBCH 5), the covered range of LAI measurements had a comparable width to that of the entire data set (cf. Table 5.2). Hence, an important requirement has been met, which has already been identified as a decisive factor for a robust calibration of regression models in Chapter 7.1.2.

The field reflectance measurements (867 spectral bands) were first spectrally resampled to the resolution of aisaDUAL (98 spectral bands) and EnMAP (204 spectral bands), respectively. Spectral resampling was conducted with the software ENVI using a Gaussian model with a full width at half maximum (FWHM) equal to the band spacings of both sensors. Hence, it was possible to investigate whether the reduced spectral resolution has any influence on the achieved model quality. For that reason, regression models were built based on the three data sets with different spectral resolutions. Due to the number of samples (n = 70), leave-one-out cross-validation was applied to evaluate the model quality. PLSR was selected as regression technique, since the accuracy obtained using this method has proven to be less affected by the type of validation technique (cf. Chapter 7.1.1).

At first, a PLSR model was built based on the full-resolution field spectral measurements of the 70 samples. The high R^2_{cv} of 0.96 (LV = 7), in combination with a relatively low $RMSE_{cv}$ of 0.279, which was achieved by the leave-one-out cross-validated model, indicated that the LAI could be predicted with high accuracy. Furthermore, the fairly high RPD_{cv} of 5.073 served as additional evidence for a robust model calibration [Dunn et al., 2002]. In comparison to the PLSR model, which was built with the entire data set (cf. Table 7.2), a distinct improvement in model performance was again possible. This is also obvious when comparing the scatter plots of both models. While the PLSR model, built with the 70 samples from Field A and Field D, only depicts distinctly less scattering and a nearly perfect match between the regression and the 1:1 line (Figure 7.4 (left)), the scatter plot of the model based on the data of all fields is characterized by more pronounced scattering and a slight deviation between the regression and the 1:1 line (Figure 7.2 (left)).

The two other models using the 70 field spectral measurements, which were resampled to aisaDUAL and EnMAP resolution, provided model qualities comparable to those of the model built with full-resolution spectra. While the EnMAP model (LV = 7) had almost identical R^2_{cv} (0.960), $RMSE_{cv}$ (0.281) and RPD_{cv} values (5.046), the aisaDUAL model showed an almost negligible poorer model performance (R^2_{cv} = 0.958, $RMSE_{cv}$ = 0.288, RPD_{cv} = 4.923, LV = 6). All three data sets enable regression model calibration and validation on a very high level of accuracy. Table 7.4 summarizes the results of the cross-validated PLSR models based on data sets with

different spectral resolutions. The scatter plots of the three models presented in Figure 7.4 also look almost identical. The models had the same regression functions, while the regression lines were very close to the 1:1 line. In all three cases, higher LAI values were marginally underestimated whereas lower values were marginally overestimated. The same effect, albeit more pronounced, has already been observed in the regression models based on the entire data set.

Table 7.4: *Cross-validated results of LAI prediction with PLSR for full-resolution field spectra, as well as field spectra resampled to aisaDUAL and EnMAP spectral resolution.*

$n = 70$	No. of bands	spectral sampling	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	LV
Field spectra	867	≤ 3.8 nm	0.960	0.279	5.073	7
aisaDUAL resampling	98	13–19 nm	0.958	0.288	4.923	6
EnMAP resampling	204	6.5–10 nm	0.960	0.281	5.046	7

The achieved results confirm that the spectral sampling rate and the resulting spectral resolution of the different data sets had no noticeable effect on the LAI assessment of wheat. Even the data set with the lowest spectral resolution (aisaDUAL – 98 spectral bands) enabled a LAI prediction on a very high level of accuracy. These findings are consistent with a study conducted by Atzberger et al. [2010], in which the authors also applied PLSR and obtained accurate LAI predictions of wheat LAI from field reflectance measurements, which were resampled to the spectral resolution of the HyMap hyperspectral airborne scanner (127 spectral bands), which had a spectral

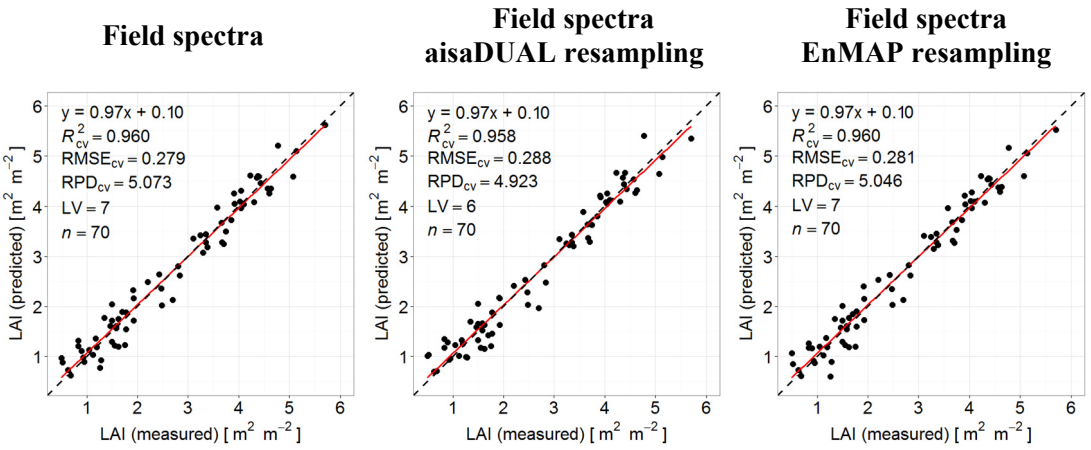


Figure 7.4: *Scatter plots of estimated (cross-validated) and measured LAI for winter wheat, based on PLSR for full-resolution field spectra and field spectra, resampled to aisaDUAL and EnMAP spectral resolution. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

configuration comparable to that of aisaDUAL used in this thesis. Moreover, numerous studies proved that the LAI may also be derived from lower-resolution spectral data with high accuracy because the LAI is sensitive to several different regions along the electromagnetic spectrum [e.g., Asner, 1998; Lee et al., 2004; Thenkabail et al., 2004]. As such, the location of spectral bands, which are sensitive to changes in LAI, can be regarded as a more important factor than their pure number. To this extent, the next section is aimed at identifying the spectral bands with the highest amount of explained variance regarding the target variable LAI.

7.2.2 Detection of the Most Important Bands for LAI Determination

In the previous section, it was demonstrated that the higher spectral resolution of the full-resolution spectral field measurements did not result in higher LAI prediction accuracy. For this reason, and since precision agriculture requires detailed spatial information on the distribution of plant parameters, only the field spectral measurements, which were resampled to the spectral resolution of the aisaDUAL and EnMAP imaging sensors, were further investigated to locate the most important spectral bands for the assessment of wheat LAI.

The red bars in Figure 7.5 and Figure 7.6 illustrate the weight factors for each spectral band, calculated by the RReliefF feature selection algorithm after 100 runs, which were determined for the field spectral measurements resampled to aisaDUAL as well as EnMAP spectral resolution. Positive values implied that a band is relevant for LAI prediction, whereas bands with negative values were irrelevant. Higher positive weight factors indicate a higher importance of the band in order to predict the LAI [Robnik-Sikonja and Kononenko, 1997]. Underlying the bar plot is a typical wheat reflectance spectrum to enable a better understanding of the information content of the spectral bands, which were weighted by RReliefF.

RRelief allowed the identification of eight spectral regions, each of which consisted of more than one band. The eight regions were manually selected by the user based on negative ranked bands or local maxima and minima. Thus, two different regions were separated either by negative ranked bands (e.g., Region 2 and Region 3) or by local minima (bands with distinctly lower positive RReliefF ratings) between two local maxima (e.g., Region 3 and Region 4). Table 7.5 gives an overview of the spectral regions identified with the help of RReliefF, based on the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution. Apart from minor differences, it became clear that the same spectral regions were selected in both data sets. Furthermore, all regions (except one) included spectral wavelengths/ bands, which have also been identified as important in other research studies [Asner, 1998; Darvishzadeh et al., 2008b; Lee et al., 2004; Thenkabail et al., 2004].

Due to the fact that adjacent spectral bands within the eight selected regions were highly correlated to each other [Ghosh et al., 2013; Sohaib et al., 2012], only the bands with the highest RReliefF rating in each group were selected for further analysis. In order to verify that the highest-ranked band per region provided the most important

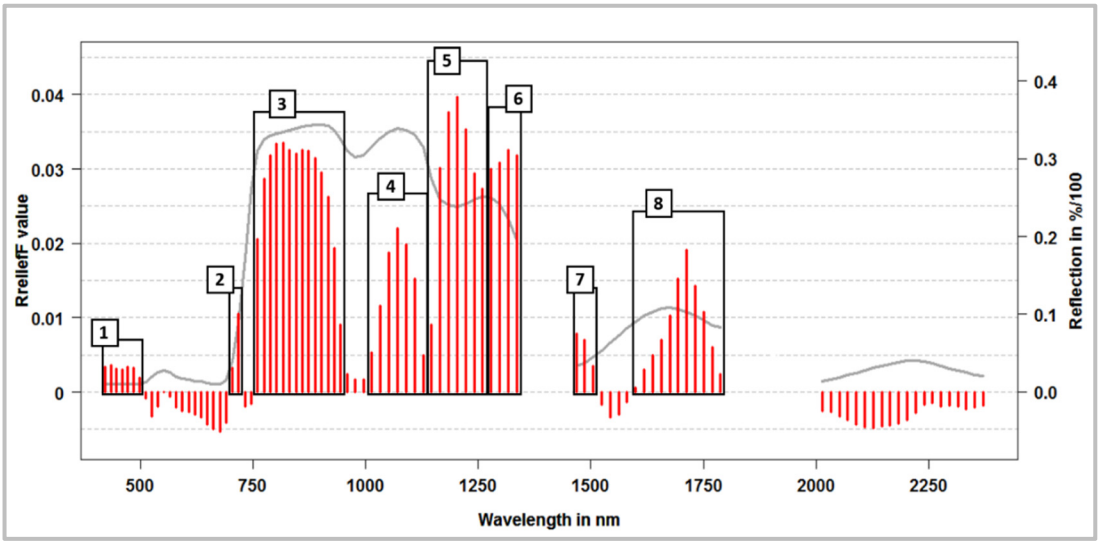


Figure 7.5: *Weight factors determined by RReliefF after 100 runs for each spectral band of the field spectral measurements resampled to aisaDUAL spectral resolution (red bars). Underlying the bar plot is a typical wheat reflectance spectrum (grey line). The eight frames represent the selected spectral ranges (Regions 1-8) with the most important information for the prediction of LAI.*

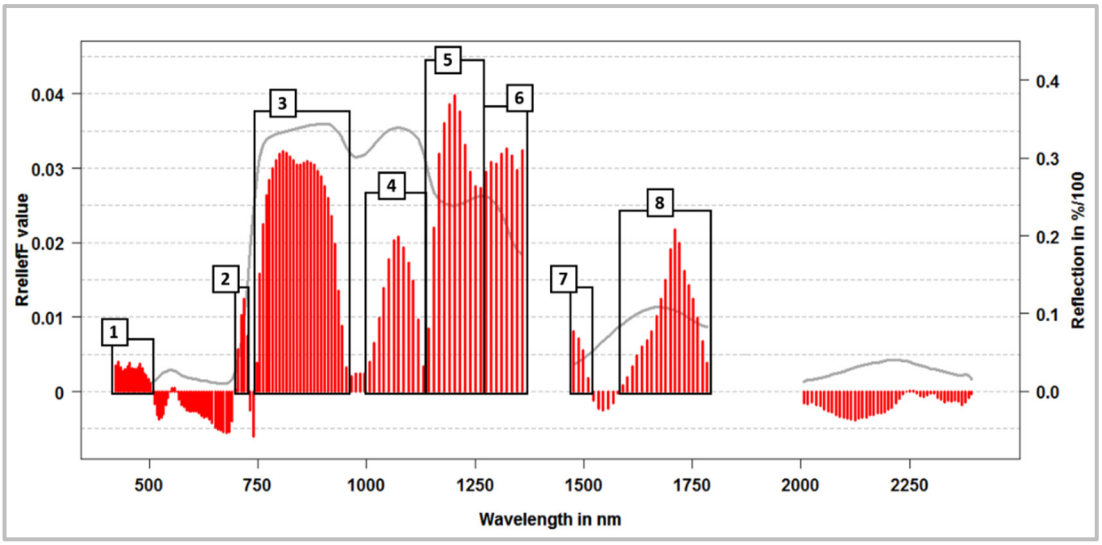


Figure 7.6: *Weight factors determined by RReliefF after 100 runs for each spectral band of the field spectral measurements resampled to EnMAP spectral resolution (red bars). Underlying the bar plot is a typical wheat reflectance spectrum (grey line). The eight frames represent the selected spectral ranges (Regions 1-8) with the most important information for the prediction of LAI.*

Table 7.5: *Spectral regions, manually selected with the help of RReliefF, from the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

RReliefF regions	Sensors	No. of bands per region	Spectral range of the regions	Importance of the regions in other studies
1	aisaDUAL	5	438–497 nm	-
	EnMAP	15	443–503 nm	
2	aisaDUAL	2	703–717 nm	x[13, 14]
	EnMAP	4	705–726 nm	
3	aisaDUAL	13	759–930 nm	x[15, 16]
	EnMAP	26	755–944 nm	
4	aisaDUAL	7	1,041–1,127 nm	x[15]
	EnMAP	10	1,018–1,120 nm	
5	aisaDUAL	7	1,146–1,260 nm	x[13]
	EnMAP	8	1,155–1,239 nm	
6	aisaDUAL	4	1,278–1,345 nm	x[13,14,15]
	EnMAP	10	1,251-1,348 nm	
7	aisaDUAL	3	1,450–1,505 nm	x[13, 14]
	EnMAP	4	1,457–1,511 nm	
8	aisaDUAL	10	1,618–1,788 nm	x[13, 14, 16]
	EnMAP	18	1,601–1,783 nm	

information regarding the target variable, a randomized experiment was conducted. In this context, 500 leave-one-out cross-validated PLSR models were built, each with one spectral band per region randomly selected to predict the LAI. Afterwards, the models were arranged in ascending order according to their R^2_{cv} value. The 50 models with the highest R^2_{cv} were subsequently chosen and a majority voting was conducted for each region to ascertain which spectral band was most frequently involved in model building. The eight spectral bands that were determined during the randomized experiment may be regarded as the most important spectral bands for deriving the LAI. Figure 7.7 and Figure 7.8 illustrate the location of the most important band per region, which was identified with the help of RReliefF (blue bars) and randomized sampling (green bars) from the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution. In addition, Table 7.6 provides an overview of the exact

[13] Asner [1998]
[14] Lee et al. [2004]
[15] Thenkabail et al. [2004]
[16] Darvishzadeh et al. [2008b]

positions of the most important spectral bands per region ascertained by both approaches. In most cases, the spectral bands of a region in each data set, determined by RReliefF and randomized sampling, were very close to each other, or even had exactly the same position. Only in two regions the identified spectral bands had

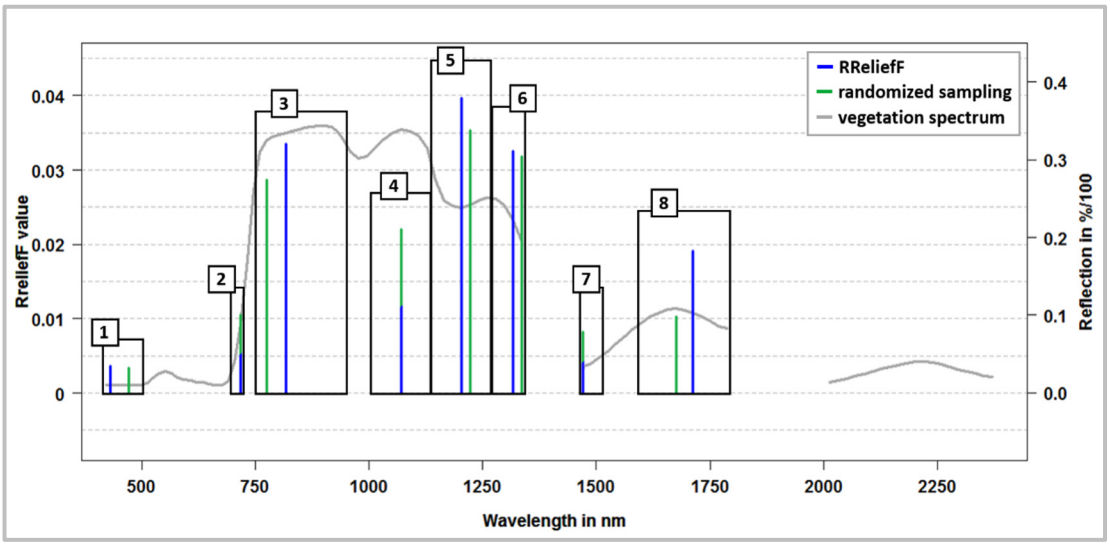


Figure 7.7: *Comparison of the most important spectral bands determined by RReliefF (blue bars) and randomized sampling (green bars) for LAI prediction, based on the field spectral measurements resampled to aisaDUAL spectral resolution. The height of a bar corresponds to the RReliefF weight factor. Underlying the bar plot is a typical wheat reflectance spectrum (grey line). The frames represent the eight previously selected spectral ranges (Regions 1-8).*

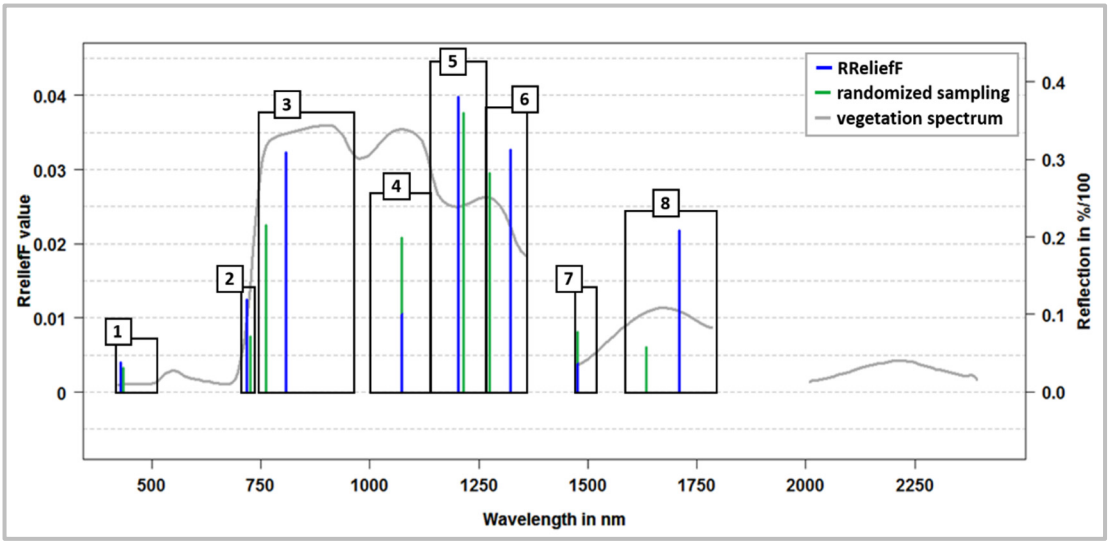


Figure 7.8: *Comparison of the most important spectral bands determined by RReliefF (blue bars) and randomized sampling (green bars) for LAI prediction, based on the field spectral measurements resampled to EnMAP spectral resolution. The height of a bar corresponds to the RReliefF weight factor. Underlying the bar plot is a typical wheat reflectance spectrum (grey line). The frames represent the eight previously selected spectral ranges (Regions 1-8).*

distances of more than 50 nm. In the case of the aisaDUAL data set, this was Region 3 (distance: 57 nm), while in the case of the EnMAP data set, this was Region 8 (distance: 78 nm). Moreover, when separately comparing the selected bands from both datasets for each region, it was apparent that the highest-ranked RReliefF bands lie very close together. Only in Region 1 were spectral bands determined with a slightly larger distance (17 nm), whereas the smaller distances in some other regions resulted from the different spectral sampling intervals of both sensors and, therefore, the location of the spectral bands in both data sets. In contrast, the randomized sampling approach led to larger distances between the selected aisaDUAL and EnMAP bands in several regions (Region 1 (54 nm), Region 6 (60 nm), Region 8 (41 nm)).

Table 7.6: *Comparison of spectral bands ascertained by RReliefF and randomized sampling for the assessment of wheat LAI based on the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

RReliefF regions	Sensors	Highest ranked band per region	
		RReliefF	Randomized sampling
1	aisaDUAL	457 nm	497 nm
	EnMAP	440 nm	443 nm
2	aisaDUAL	717 nm	717 nm
	EnMAP	719 nm	726 nm
3	aisaDUAL	816 nm	759 nm
	EnMAP	808 nm	762 nm
4	aisaDUAL	1,070 nm	1,070 nm
	EnMAP	1,074 nm	1,074 nm
5	aisaDUAL	1,203 nm	1,222 nm
	EnMAP	1,203 nm	1,215 nm
6	aisaDUAL	1,316 nm	1,335 nm
	EnMAP	1,323 nm	1,275 nm
7	aisaDUAL	1,450 nm	1,450 nm
	EnMAP	1,457 nm	1,457 nm
8	aisaDUAL	1,713 nm	1,675 nm
	EnMAP	1,710 nm	1,634 nm

As the next step, leave-one-out cross-validated PLSR regression models were again built, but only with the eight spectral bands identified with the help of RReliefF and randomized sampling for both data sets. Since only eight bands were involved in regression model building, the allowed number of latent variables was set to eight. Table 7.7 summarizes the results of the cross-validated models. The achieved results were almost identical to those obtained for the models built with all the spectral bands

of aisaDUAL (98 bands) and EnMAP (204 bands), respectively (cf. Table 7.4). Both the aisaDUAL and the EnMAP models, which were based on the eight bands selected by randomized sampling, provided slightly higher R^2_{cv} and RPD_{cv}, as well as lower $RMSE_{cv}$ values, in comparison to the two RReliefF models. Furthermore, in the case of randomized sampling, only six LVs were used for model building, whereas the RReliefF models are based on seven and eight LVs, respectively. In this context, a PLSR model with a lower number of LVs can be regarded as more robust than a model with a higher number of LVs because less noise is introduced into the model by excluding lower order LVs [Li et al., 2002]. Therefore, the randomized sampling models can be regarded as the models with higher statistical quality.

***Table 7.7**: Cross-validated results of wheat LAI prediction with PLSR for the eight spectral bands determined by RReliefF and randomized sampling, based on the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

$n = 70$	No. of bands	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	LV
aisaDUAL resampling (RReliefF)	8	0.956	0.294	4.814	7
aisaDUAL resampling (randomized s.)	8	0.960	0.281	5.045	6
EnMAP resampling (RReliefF)	8	0.958	0.289	4.904	8
EnMAP resampling (randomized s.)	8	0.960	0.280	5.056	6

Figure 7.9 illustrates the scatter plots of the RReliefF and randomized sampling PLSR models, which also exhibited strong similarities with the scatter plots of the models built with all the spectral bands of aisaDUAL and EnMAP (cf. Figure 7.4). All four models in Figure 7.9 show less scattering and a close match of the regression to the 1:1 line. Moreover, as the models have almost the same regression function, they can be regarded as equivalent.

Although the randomized sampling models provided slightly better model performances, the differences were negligibly small and both approaches identified almost the same wavelength in each region as important. For that reason, the randomized sampling can be regarded as an appropriate method to evaluate whether the highest-ranked RReliefF spectral bands were most suited for an accurate LAI assessment.

The next question was concerned with examining, whether the spectral information of each identified region was truly necessary for an accurate LAI prediction. Therefore, a second randomized experiment was conducted. In this context, the number of the highest-ranked spectral bands selected with the help of RReliefF (one spectral band from each region) was gradually reduced from eight to two spectral bands. In each

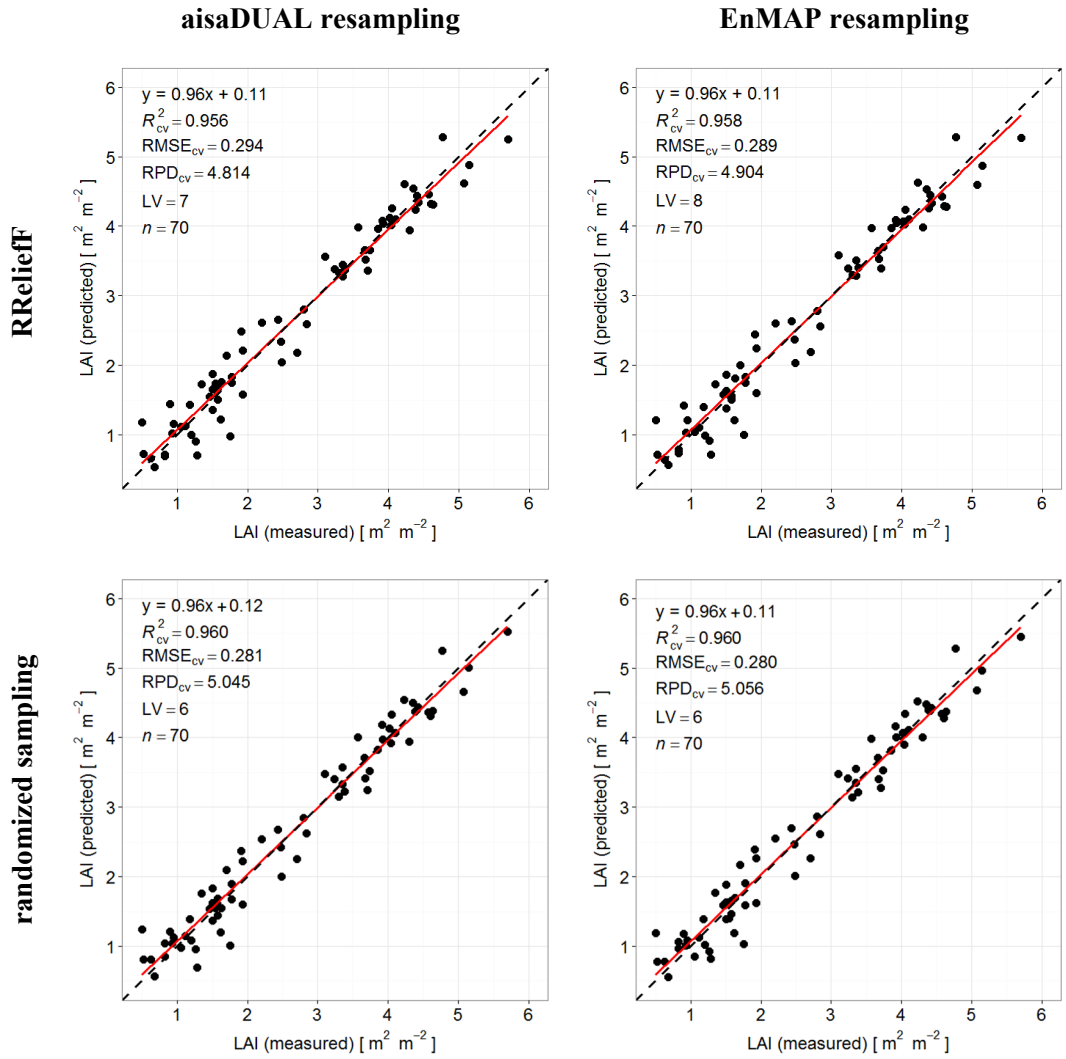


Figure 7.9: *Scatter plots of estimated (cross-validated) and measured LAI for winter wheat, based on PLSR for the eight spectral bands, which were determined by RReliefF and randomized sampling from the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

step, 100 leave-one-out cross-validated PLSR models were generated to predict the LAI. Each of the 100 models was built with n – 1 randomly selected spectral bands, while the number of allowed LVs was adapted to the number of used spectral bands. Afterwards, the model with the highest R^2_{cv} was selected and the band that was not involved in model building was determined. This spectral band was subsequently excluded from further analysis in the next step. Figure 7.10 illustrates the changes in R^2_{cv} and RMSE$_{cv}$ for the aisaDUAL and EnMAP data set when the number of spectral bands was gradually reduced. The figure clearly shows that the R^2_{cv} and RMSE$_{cv}$ of the aisaDUAL and EnMAP models had similar or even identical values. Therefore, they are described together below. From eight (Table 7.7) to four spectral bands (R^2_{cv} = 0.952, RMSE$_{cv}$ = 0.307), the decrease in R^2_{cv} and the increase in RMSE$_{cv}$ were

very small. Although the further reduction to three spectral bands resulted in a slightly more pronounced decline in R^2_{cv} (0.946) and rise in $RMSE_{cv}$ (0.327), the PLSR models based on three aisaDUAL and three EnMAP spectral bands still provided high model qualities. In contrast, the subsequent PLSR modelling with only two spectral bands led to a distinct decrease in model accuracy. The model quality, however, was still rather high with, an R^2_{cv} of 0.912 and an $RMSE_{cv}$ of 0.418.

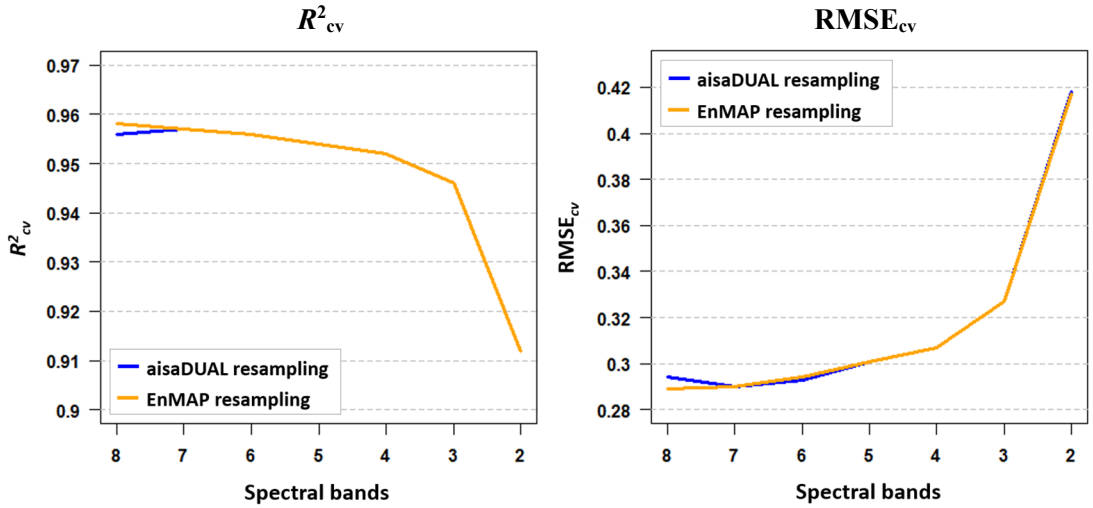


Figure 7.10: *Changes in R^2_{cv} and $RMSE_{cv}$ regarding the prediction of winter wheat, based on PLSR, when gradually reducing the number of spectral bands from the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

The models generated with three spectral bands can be regarded as the most robust models because they provided distinctly higher model qualities than the models based on two spectral bands, as well as possessed almost the same accuracies compared to the models built with four to eight spectral bands. This indicates that more than three spectral bands are incapable of further increasing the model qualities to a significant degree. For both the aisaDUAL and the EnMAP data set, the highest ranked bands from Regions 4, 5, and 8 were identified as sufficient to predict the LAI with high accuracy. Table 7.8 summarizes the most important spectral bands determined from the spectral field measurements resampled to aisaDUAL and EnMAP spectral resolution.

In Region 4, the aisaDUAL spectral band located at 1,070 nm and the EnMAP spectral band located at 1,074 nm were identified as important for the assessment of LAI. This is well in line with a study by Thenkabail et al. [2004], in which a similar wavelength position (1,085 nm) was found to be sensitive to changes in the LAI of shrubs, grasses, weeds and crops. Furthermore, the spectral bands at 1,203 nm in Region 5 were highly rated in both data sets. The identified position was close to the small water vapour absorption feature on the NIR plateau, which is at around

1,150 nm [Rollin and Milton, 1998]. This wavelength position was also determined to be relevant when predicting the LAI by Asner [1998], who observed a deepening of this absorption feature with increasing LAI. Furthermore, important spectral bands for the assessment of wheat LAI were found in Region 8 at 1,713 nm and 1,710 nm in the aisaDUAL and EnMAP data set, respectively. This is consistent with the findings of previous studies by Darvishzadeh et al. [2008b] and Lee et al. [2004], all of whom ascertained strong relationships between spectral bands in the SWIR I region (1,500–1,900 nm) and the LAI of grasslands and crops.

Table 7.8: *Comparison of the most important spectral bands ascertained for the assessment of wheat LAI, based on the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

RReliefF Region	Sensor	Highest ranked RReliefF band
4	aisaDUAL	1070 nm
	EnMAP	1074 nm
5	aisaDUAL	1203 nm
	EnMAP	1203 nm
8	aisaDUAL	1713 nm
	EnMAP	1710 nm

The results of the leave-one-out cross-validated PLSR models built with three aisaDUAL and three EnMAP spectral bands are presented in Table 7.9. Additionally, the results of the models based on eight as well as all the spectral bands are shown again to allow for a better comparison. The achieved results prove that highly accurate LAI predictions based on only three spectral bands were possible. The three-band model qualities for aisaDUAL and EnMAP were similar (R^2_{cv} = 0.946, $RMSE_{cv}$ = 0.327), as well as only slightly poorer compared to the models based on eight as well as all the spectral bands. RPD_{cv} values well over 4.0 additionally indicated

Table 7.9: *Cross-validated results of wheat LAI prediction with PLSR for models, based on three (RReliefF), eight (RReliefF) and all the spectral bands of the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution.*

n = 70	No. of bands	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	LV
aisaDUAL resampling (RReliefF)	3	0.946	0.327	4.333	3
EnMAP resampling (RReliefF)	3	0.946	0.327	4.333	3
aisaDUAL resampling (RReliefF)	8	0.956	0.294	4.814	7
EnMAP resampling (RReliefF)	8	0.958	0.289	4.904	8
aisaDUAL resampling	98	0.958	0.288	4.923	6
EnMAP resampling	204	0.960	0.281	5.046	7

the robustness and successful calibration of both models [Dunn et al., 2002]. The scatter plots in Figure 7.11 confirmed the results of the three-band regression models, based on their statistical values. Both models had the same regression function, exhibited less scattering and provided a close match between the regression and the 1:1 line.

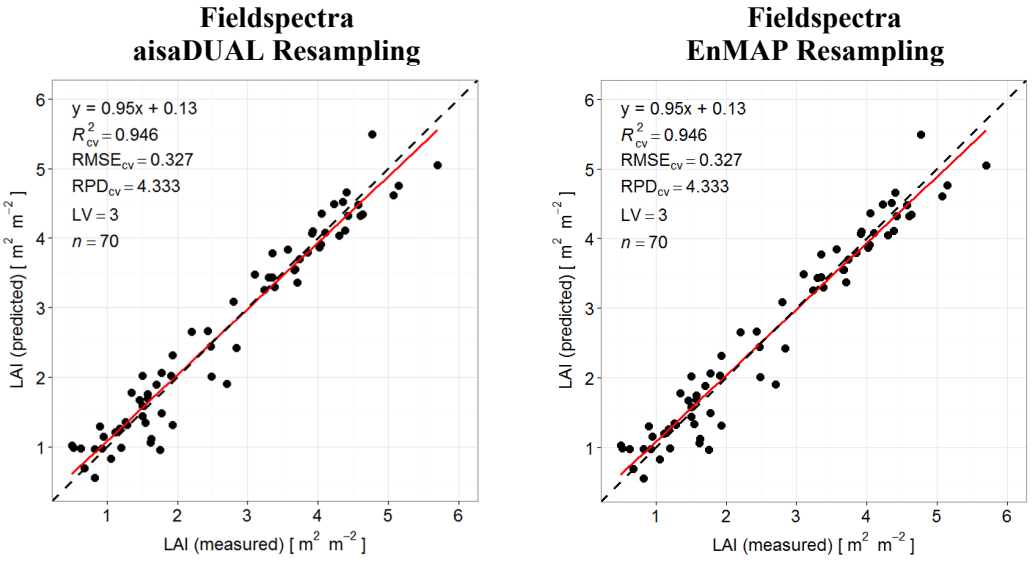


Figure 7.11: *Scatter plots of estimated (cross-validated) and measured LAI for winter wheat, based on PLSR for the three most important spectral bands of the field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

In general, the weight factors determined by RReliefF were only partially helpful in identifying the most important spectral bands for the assessment of wheat LAI. While the first randomized experiment proved that the highest-ranked RReliefF spectral band in each region mostly had the highest information content in order to predict the LAI, the second randomized experiment clearly showed that only the spectral bands with the highest RReliefF rating from Regions 4, 5, and 8 were necessary for highly accurate LAI predictions. As already described above, the three most important bands were also found to be sensitive to changes in LAI in other studies. While the importance of the selection of spectral bands in Regions 4 and 5 can probably be explained by the fact that changes in canopy LAI strongly influence the level of reflectance height in the NIR domain [Asner, 1998], the high rating of spectral bands in Region 8 (SWIR), which is mainly influenced by the water content of plants [Verdebout et al., 1994], may result from the high correlation of LAI and PWC (r = 0.98) determined in this thesis (cf. Chapter 5.1.1). In this context, Verrelst et al. [2015] suggested that the estimation of a structural plant parameter

(e.g., LAI) can be supported by a biochemical plant parameter (e.g., PWC), if both are correlated to each other. Spectral information from VIS and the transition area between VIS and NIR (700-900 nm), however, seem to be less important in the assessment of wheat LAI in this thesis. This stands in contrast to studies conducted by Lee et al. [2004] and Thenkabail et al. [2004], all of whom determined strong relationships between these spectral regions and the parameter LAI.

The results showed that only three spectral bands at specific locations were necessary to predict the LAI of wheat with an accuracy comparable to those achieved for the models based on all the spectral bands. This fact proved that PLSR can deal efficiently with a low number of specific spectral bands without any substantial loss in model accuracy. This is in accordance with a study conducted by An et al. [2015], in which the authors also achieved comparable prediction results for the LAI of turfgrass using PLSR models calibrated with continuous spectra and models, which were only generated with a small number of selected spectral bands. Identifying the most important spectral regions from hyperspectral data for the assessment of a certain plant parameter can be very helpful in predicting the same parameter for other parts of the same study site, for which only multispectral data are available that possibly contain spectral bands in the same spectral ranges. Unfortunately, none of the three regions identified in this thesis as important in the retrieval of LAI is covered by spectral bands of current multi-/superspectral satellite systems (e.g., Landsat 8, Sentinel-2).

Due to the limited number of sampled wheat plots used in this thesis, which were collected from only one study site during only two stages of plant development, the generated PLSR models cannot simply be transferred to other regions. For that reason, LAI and corresponding reflectance measurements of numerous regions, which cover several growth stages, must be examined in future studies in order to develop more robust and regionally transferable regression models, based on the most relevant spectral bands. In this regard, first promising results were achieved by Gerighausen et al. [2016], who successfully transferred PLSR models, which were calibrated with data collected from one study site to a different region, in order to predict the LAI of crops.

The derivation of LAI from remote sensing data for the assessment of the current growth conditions of crops in agriculture requires high accuracy. In this section, it was shown that both aisaDUAL and EnMAP provide the required spectral characteristics to enable an accurate prediction of wheat LAI. Moreover, the location of the most important spectral bands of both sensor systems were determined in order to demonstrate that, by far, not all the spectral bands of aisaDUAL and EnMAP hyperspectral data are necessary for estimating the LAI. Together with the spectral specifications of imaging remote sensing sensors, the spatial resolution plays a decisive role in many applications in agriculture. For that reason, the next section

investigates the GSD of aisaDUAL and EnMAP to ascertain their potential for the spatial prediction of LAI in precision agriculture.

7.3 Influence of Spatial Resolution on LAI Determination

For precision agricultural applications, the spatial resolution is one of the most decisive factors to enable an adapted treatment of fields [Whelan and Taylor, 2013]. Most of these applications (e.g., irrigation, fertilization) require GSDs of less than ten metres [Mulla, 2013]. Hyperspectral data acquired from UAVs or aircraft usually meet this requirement. Current and near future hyperspectral satellites, however, provide/will provide spatial resolutions of only 30 m. Nevertheless, from a spectral point of view, hyperspectral data, typically with more than 100 spectral bands, offer the potential to derive the LAI on a level of accuracy that often cannot be achieved using multispectral data [Lee et al., 2004]. Therefore, this section examines the impact of the spatial resolution of hyperspectral image data on the retrieval of the parameter LAI. In this regard, the following research question was formulated, which should be answered in the ensuing sections:

What influence does the ground sampling distance of remote sensing images have on the spatial assessment of LAI and the associated regression model validation?

7.3.1 LAI Prediction Based on aisaDUAL and EnMAP Image Data

In Chapter 7.2, it was already demonstrated that the spectral characteristics of aisaDUAL and EnMAP were well suited to the precise assessment of LAI. The previous investigations in this thesis, however, were exclusively conducted using field spectral measurements resampled to aisaDUAL and EnMAP spectral resolution. To ascertain the potential of real aisaDUAL and simulated EnMAP image data for the spatial prediction of LAI, the reflectance spectra of the image pixels, which corresponded to the geographic location of the different sampled wheat plots of Field A and Field D, were extracted from both data sets. Afterwards, PLSR models were generated based on the extracted spectra. First, all bands of aisaDUAL and EnMAP were used for model calibration and then models were built with only the three most important bands identified for each sensor, as given in Chapter 7.2.2. Table 7.10 summarizes the results of the generated PLSR models.

The high R^2_{cv} of both aisaDUAL models (98 bands: 0.890, 3 bands: 0.892), in combination with a relatively low $RMSE_{cv}$ (98 bands: 0.466, 3 bands: 0.461), achieved by leave-one-out cross-validation, indicated that the LAI could be predicted with high accuracy. Although the model accuracies were lower when compared to the models calibrated with the field spectral measurements resampled to aisaDUAL spectral

resolution (cf. Table 7.9), both models built with real aisaDUAL data provided an RPD_{cv} higher than 3.0, which indicated a robust model calibration [Dunn et al., 2002]. In contrast, both EnMAP models showed distinctly lower model accuracies with respect to R^2_{cv} (204 bands: 0.666, 3 bands: 0.678) and $RMSE_{cv}$ (204 bands: 0.831, 3 bands: 0.814) when compared to the models, which were based on the field spectral measurements resampled to EnMAP spectral resolution (cf. Table 7.9). In addition, RPD_{cv} values of 1.703 for the EnMAP model, built with all the spectral bands and 1.740 for the three-band EnMAP model only suggested moderate model performances [Dunn et al., 2002].

Table 7.10: *Cross-validated results of wheat LAI prediction with PLSR for models, based on all as well as only three spectral bands of aisaDUAL and EnMAP.*

n = 70	No. of bands	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	LV
aisaDUAL	98	0.890	0.466	3.037	4
aisaDUAL (RReliefF)	3	0.892	0.461	3.069	3
EnMAP	204	0.666	0.831	1.703	3
EnMAP (RReliefF)	3	0.678	0.814	1.740	3

The scatter plots of the aisaDUAL and EnMAP models are presented in Figure 7.12. Both aisaDUAL plots showed more pronounced scattering compared to the scatter plots of the models, which were built with the field spectral measurements resampled to aisaDUAL spectral resolution (cf. Figure 7.9 and Figure 7.11). In general, however, the scattering was on a relatively low level. The offsets of both models were small, such that, the regression lines fitted close to the 1:1 lines. In contrast, the EnMAP models provided distinctly more scattering and the regression lines considerably deviated from the 1:1 lines. Furthermore, both aisaDUAL and EnMAP scatter plots illustrated the same trend of overestimating lower and underestimating higher values, which were already observed in the scatter plots of the regression models in the previous sections.

For both data sets, the three-band models provided comparable model accuracies to those achieved for the full-resolution models. This proved that the three identified spectral bands of both sensors in Chapter 7.2.2 also carried the most important information for the assessment of wheat LAI in the aisaDUAL and EnMAP image data. The quality of the three-band aisaDUAL model was only slightly lower compared to the field spectral measurements resampled to aisaDUAL spectral resolution, given that the locations of the small field plots (size: 0.25 m^2) were chosen on the basis that each plot was representative of a larger surrounding area of approximately 5 x 5 m with comparable growth conditions. Therefore, the LAI measurements of the field plots appeared to be suitable for calibrating and validating robust PLSR models, based

on aisaDUAL image pixels, each of which covered an area of 9 m^2. An EnMAP pixel, however, has a size of 900 m^2; therefore, the sampled field plots, which were 225 times smaller, were less representative for model calibration and validation. Consequently, poor model qualities were achieved. Figure 7.13 illustrates the size of an aisaDUAL and an EnMAP pixel in comparison to a sampled field plot in order to emphasize the problem.

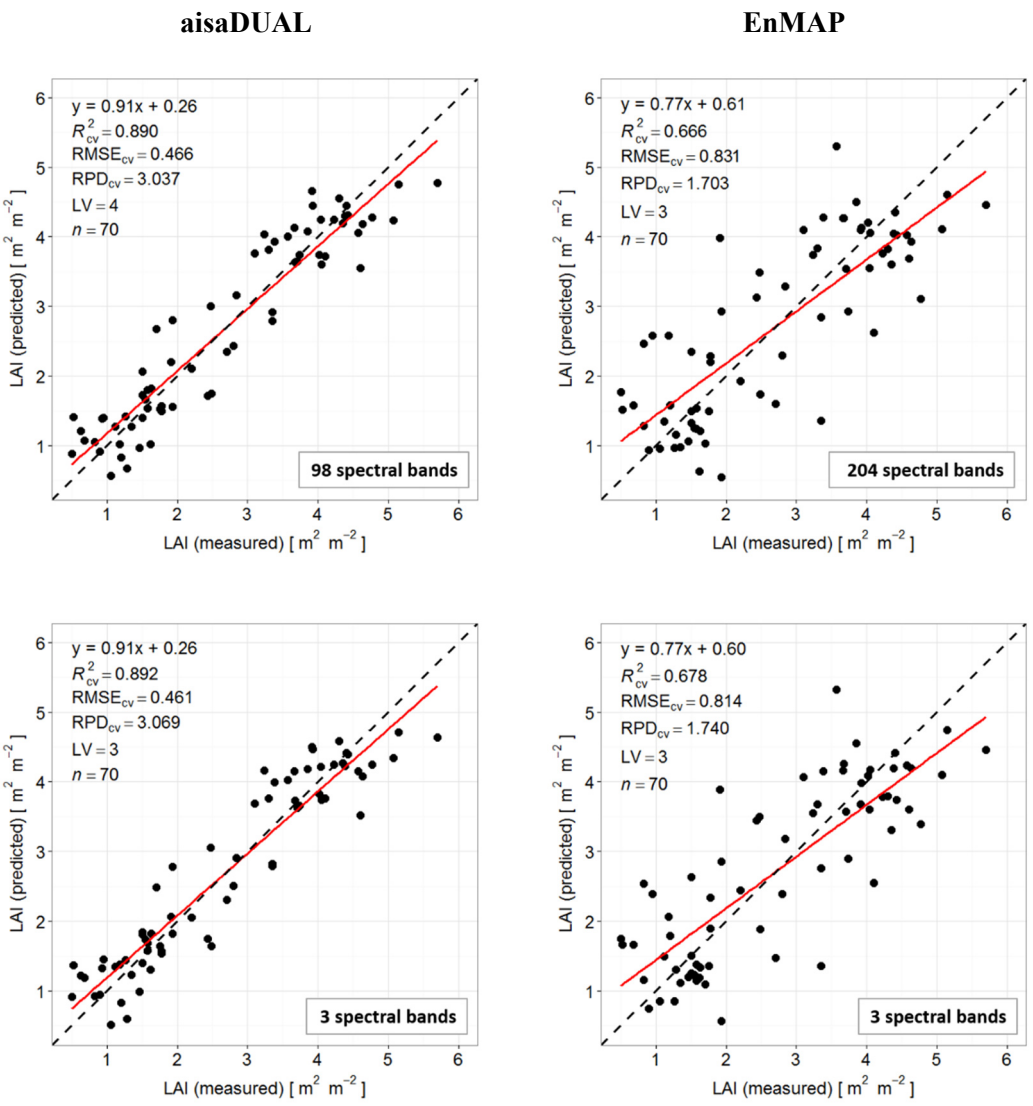


Figure 7.12: *Scatter plots of estimated (cross-validated) and measured LAI for winter wheat, based on PLSR for all as well as for only the three most important spectral bands of aisaDUAL and EnMAP. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

Although it was only the aisaDUAL three-band model that provided high accuracy, both three-band models were transferred to the image data of Field A and Field D to predict the spatial distribution of LAI. Figure 7.14 illustrates the LAI maps based on

the two three-band prediction models. In the rest of this section, the three-band models are only referred to as either the aisaDUAL or the EnMAP model, respectively. Due to the higher quality of the aisaDUAL model, the predicted aisaDUAL LAI maps served as reference; as such, they are described in more detail below.

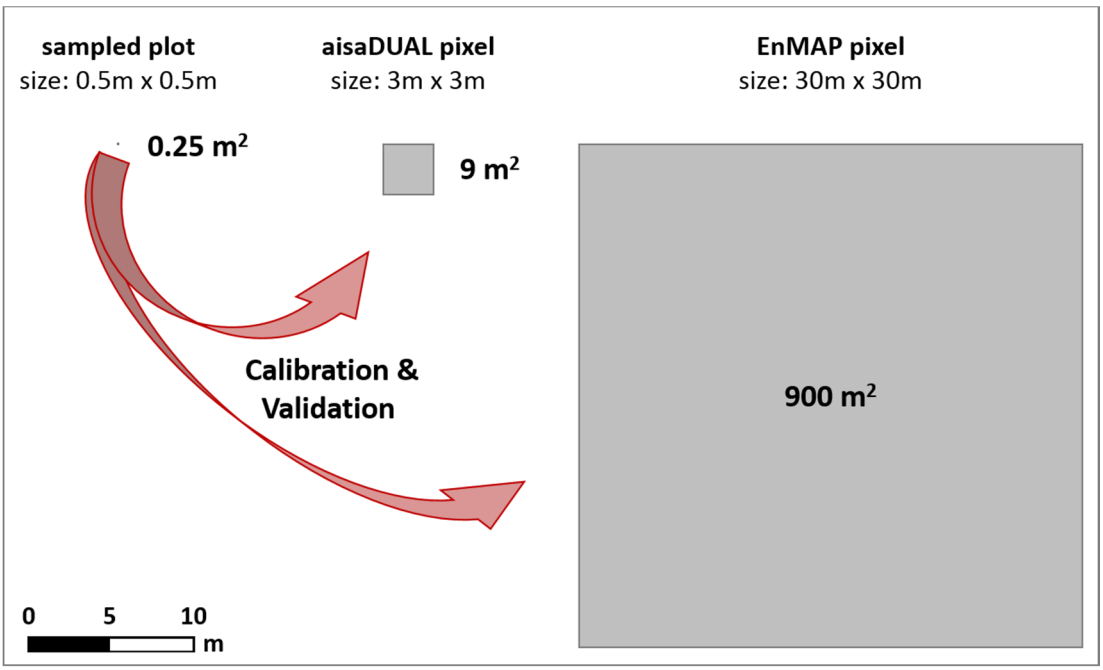


Figure 7.13: *Size of an aisaDUAL and an EnMAP image pixel in comparison to the size of a sampled field plot.*

The aisaDUAL maps for both years showed a very detailed spatial LAI distribution across the fields and the tramlines caused by agricultural machinery were clearly visible. Field A had a very heterogeneous spatial LAI distribution, while, in some parts of the field, the LAI was already above two (middle southern part); other parts, however, were characterized by an LAI of distinctly lower than one (eastern and north-western parts). Between the parts with low and high LAI, there were small transitional zones with an LAI of around one. This unequal distribution of LAI is typical for the early stages of plant development and represents intra-field small-scale differences in water supply, different soil conditions and/or differences in sowing density. These factors have a strong influence on plant growth and, thus, on the LAI. In contrast to Field A, Field D had a relative homogeneous spatial LAI distribution. The main parts of the field were characterized by an LAI of between three and four. Only three small areas on the western edge and two areas on the eastern edge were characterized by an LAI of lower than two, which differed from the rest of the field. During the more advanced stages of plant development (as reported for 2012 in this thesis), the intra-field spatial differences of LAI distribution are not quite as pronounced as they are in

earlier development stages. Additionally, Field D was much smaller than Field A and, apparently, had more comparable growth conditions in spatial terms. The maps based on the EnMAP model also represented the general LAI structure of the fields. In comparison to the aisaDUAL maps, small variations could not be detected. The reason for this was that the lower spatial resolution of the EnMAP image data only allowed the calibration of a regression model with moderate prediction capabilities.

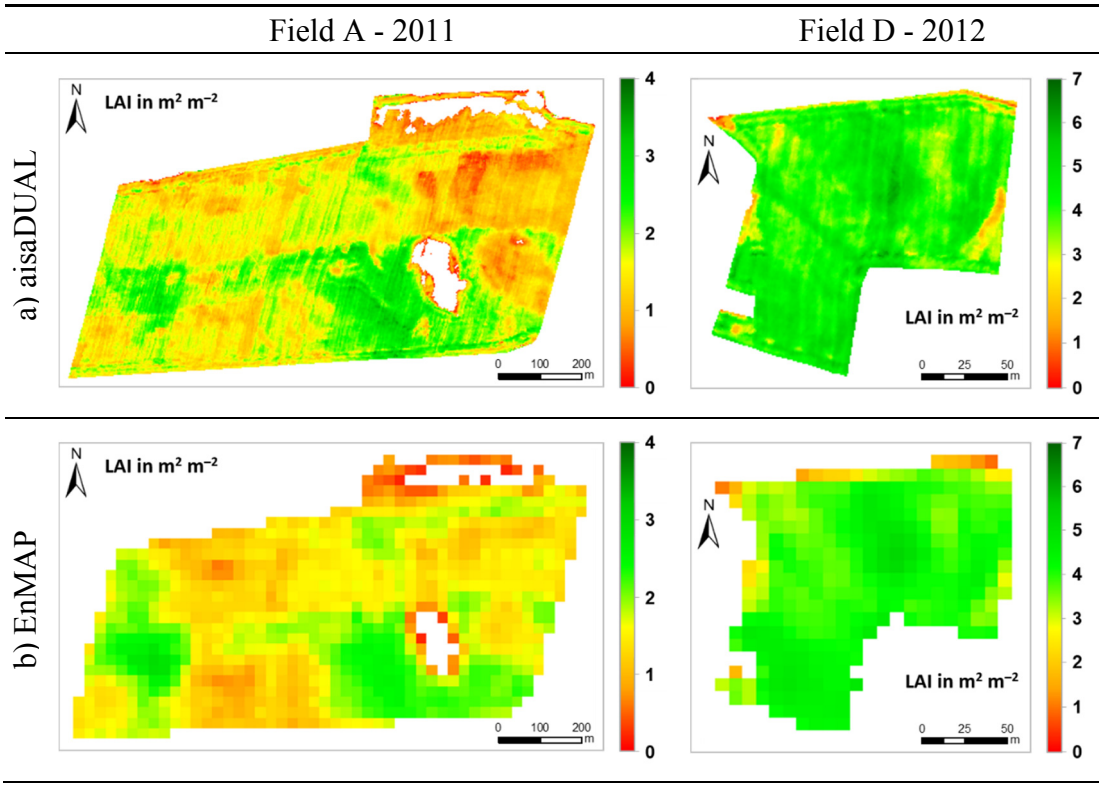


Figure 7.14: *Spatial LAI prediction of Field A and Field D, based on aisaDUAL and EnMAP.*

Due to the medium spatial resolution of EnMAP, robust calibration and validation of regression models, using reference data collected from small field plots, to predict the LAI or other plant parameters was only possible to a limited extent. For that reason, the next section deals with two different approaches, which enable a more precise validation of spatial LAI predictions based on EnMAP, in which image data of higher spatial resolution is also available.

7.3.2 Validation of Spatial LAI Prediction Based on EnMAP Image Data[17]

According to Wu and Li [2009], the retrieval of LAI from remote sensing data acquired at different spatial scales is feasible using the same method. Thus, it was possible to use PLSR models for the spatial prediction of LAI based on the aisaDUAL and EnMAP data sets of Field A and Field D. For a robust calibration and validation of the EnMAP model, however, in situ LAI measurements are necessary, which are representative of an EnMAP pixel with a size of 900 m^2. Since the LAI measurements collected during the field campaigns were only representative of areas that were equal to or less than 25 m^2, a spatial comparability between the reflectance values of the EnMAP pixels (independent variables) and the in situ LAI measurements (dependent variables) was not given, such that a valid model calibration was not possible. Nevertheless, in order to enable a reasonable spatial prediction of LAI based on EnMAP image data, two different approaches were applied in this thesis. The first one was to upscale the aisaDUAL LAI result maps of both fields to the spatial resolution of EnMAP, and then comparing them with the LAI result maps obtained from the EnMAP images. This approach was referred to as aggregation [Morisette et al., 2006]. Another possibility was to improve the spatial resolution of the EnMAP images by merging them with higher resolution image data, and then generating a new PLSR prediction model based on the merged data set, which allows for a more robust spatial LAI assessment of both fields. This method was called fusion in the further course of this section.

Aggregation

The aim of the aggregation approach, also known as the bridging method, was to make the LAI values, which were derived from coarser resolution EnMAP data, equal to the average of LAI values derived independently from finer resolution aisaDUAL data [Shi et al., 2015; Tian et al., 2003]. In order to achieve this, the aisaDUAL LAI result maps presented in Figure 7.14 a) were spatially scaled up (aggregated) to 30 m spatial resolution using an EnMAP pixel grid (Figure 7.15 a)). Hence, the aggregated aisaDUAL and the EnMAP LAI maps of both fields (Figure 7.15 b)) had the same spatial resolution, which allowed for a comparison of the predicted values at the pixel level. Figure 7.15 c) shows the difference maps calculated for both fields. In this context, the EnMAP LAI maps were subtracted from the aisaDUAL LAI maps, which were scaled up to a GSD 30 m, since the spatial predictions based on aisaDUAL served as references.

[17] Parts of this section have previously been published by Siegmann et al. [2015].

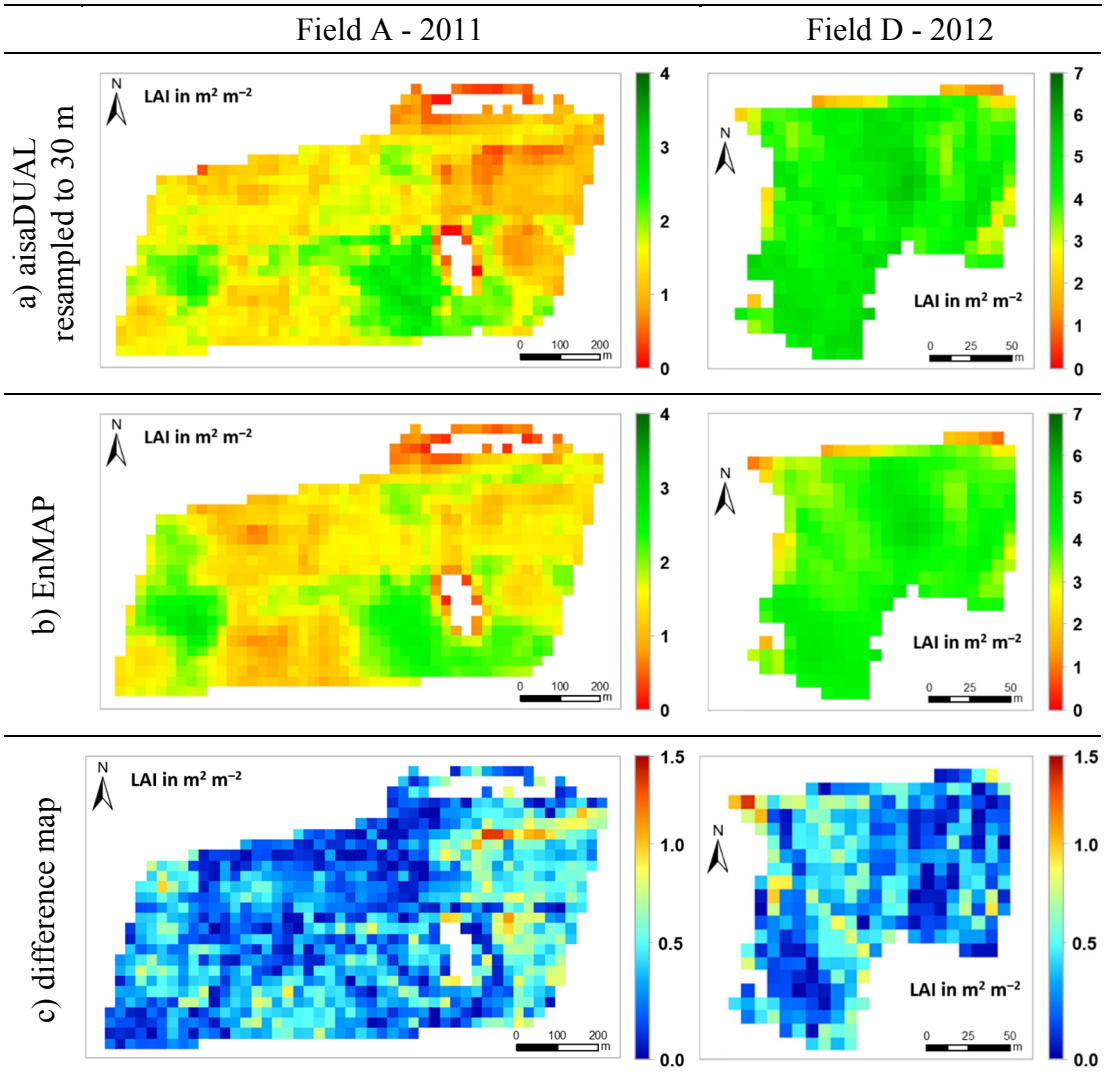


Figure 7.15: *Spatial LAI prediction of Field A and Field D. a) aisaDUAL aggregated to a spatial resolution of 30 m; b) EnMAP; and c) resulting difference maps for both fields (difference of aisaDUAL (aggregated) and EnMAP predictions).*

Bluish pixels in the difference maps represented areas with small LAI differences, which were achieved for both data sets, while yellow, orange and red pixels were indicators of higher deviations. The differences in the spatial LAI predictions were on the same level for both fields. In the western and north-western parts of Field D, two areas with larger deviations were identified. During the field campaign in 2012, both areas exhibited low vegetation cover, meaning that they had low LAI values. In comparison to the aggregated aisaDUAL LAI map, however, the LAI of both areas was distinctly underestimated in the EnMAP LAI map. The difference map of Field A also provided mainly bluish pixels. Only in the eastern and north-eastern part of the field were a few pixels with higher deviations observed. In these areas, the LAI was slightly overestimated in the LAI prediction map, which was based on EnMAP in comparison to the aggregated aisaDUAL LAI map. The descriptive statistics of the

LAI difference maps are presented in Table 7.11. The mean LAI deviation of 0.409, which was determined for Field D, was slightly higher than that for Field A, which was only 0.265. Furthermore, a combined scatter plot was generated to illustrate the goodness of fit between the aggregated aisaDUAL and the EnMAP LAI maps (Figure 7.16). The scatter plot also clarified the higher conformities of lower LAI values (LAI < 3), which were mainly determined for Field A, and the slightly more pronounced deviations for higher LAI values (LAI > 4), which were exclusively derived for Field D. The higher scattering of LAI values greater than four was possibly predicated on the fact that the more complex geometry of the further developed wheat plants, which were investigated in 2012 (Field D), had more influence on the prediction of LAI from different spatial scales. The deviation of the regression line from the 1:1 line in the scatter plot further confirmed the overestimation of lower and the underestimation of higher LAI values in the EnMAP prediction maps, which were already observed in the difference maps computed for both fields (Figure 7.15 c)).

Table 7.11: *Descriptive statistics of the LAI differences maps achieved for Field A and Field D*

Year	Field	*n*	Min	Max	Mean	SD	Median
2011	Field A	901	0.000	1.323	0.265	0.170	0.255
2012	Field D	310	0.000	1.478	0.409	0.279	0.384
2011 & 2012	Field A &D	1211	0.000	1.478	0.318	0.238	0.282

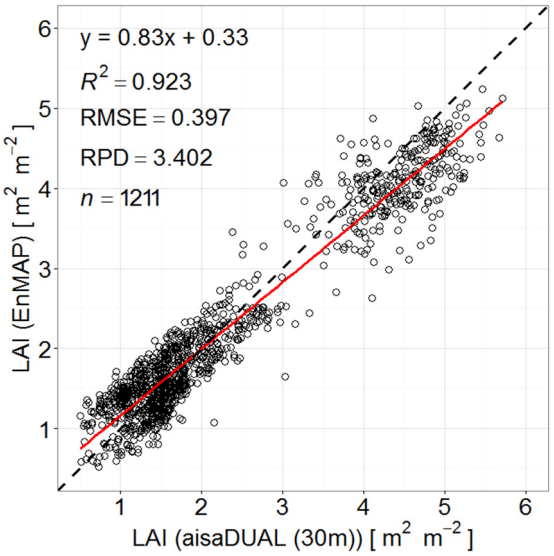


Figure 7.16: *Scatter plot of spatial LAI predictions achieved for Field A and Field D based on aisaDUAL (resampled to a GSD of 30 m) and EnMAP LAI result maps.*

Fusion

The objective of the fusion approach was to sharpen the medium spatial resolution hyperspectral EnMAP images of both fields (30 m GSD) using panchromatic aisaDUAL (3 m GSD) and Sentinel-2 image data (10 m GSD) of higher spatial resolution without altering the spectral characteristics of EnMAP. As described in Chapter 5.2, the different data sets were prepared and the Ehlers Fusion algorithm was applied to merge the images.

Figure 7.17 illustrates the fusion results for both fields in comparison to the simulated EnMAP data sets (Figure 7.17 a)). A predefined filter design for rural regions was used in the fusion process, which took into account the predominant land cover of the investigated area. As a result of merging simulated EnMAP data with aisaDUAL or Sentinel-2 panchromatic images, fused data sets were created with the spectral characteristics of EnMAP (204 spectral bands) and the spatial resolution of aisaDUAL pan (Figure 7.17 b)) or Sentinel-2 pan (Figure 7.17 c)), respectively.

As a first step, a visual analysis of the fusion results was conducted. In this context, the EnMAP–aisaDUAL fusion result of Field A showed a plausible spatial enhancement compared to the EnMAP data and a good agreement with the original aisaDUAL image. In contrast, the EnMAP–aisaDUAL fusion of Field D appeared more blurred. This was particularly evident for the villages with their small-scale structures in the northern and south-eastern parts of the image. Both EnMAP–aisaDUAL fusion results (Field A and Field D) showed some artefacts. Due to the frequency content of the image covering Field D, these artefacts were less notable. The results of the EnMAP–Sentinel-2 fusion showed a similar trend. The result for Field A seemed to be sharper compared to the result achieved for the images of Field D. Furthermore, the lower spatial resolution of Sentinal-2 provoked fewer artefacts in the result images in comparison to the EnMAP–aisaDUAL fusion results. In terms of visual evaluation, the colour infrared composites (RGB 863/652/548 nm) of all fusion results provided reasonable colour preservation. This kind of interpretation, however, can be regarded as very subjective and only allowed the analysis of a single three-band combination. For a more objective spectral evaluation, α_{spec} was calculated using corresponding pixels from the original aisaDUAL data sets and the fused images. In this context, the original aisaDUAL datasets, which consisted of 367 spectral bands, first had to be resampled to match the spectral resolution of the fusion results (204 spectral bands). Second, due to the different spatial resolution of the original aisaDUAL data (3 m GSD) and the EnMAP–Sentinel-2 fusion results (10 m GSD), a spatial resampling of the aisaDUAL GSD to 10 m was necessary. Thus, the data sets were prepared to apply a pixel-based α_{spec} determination.

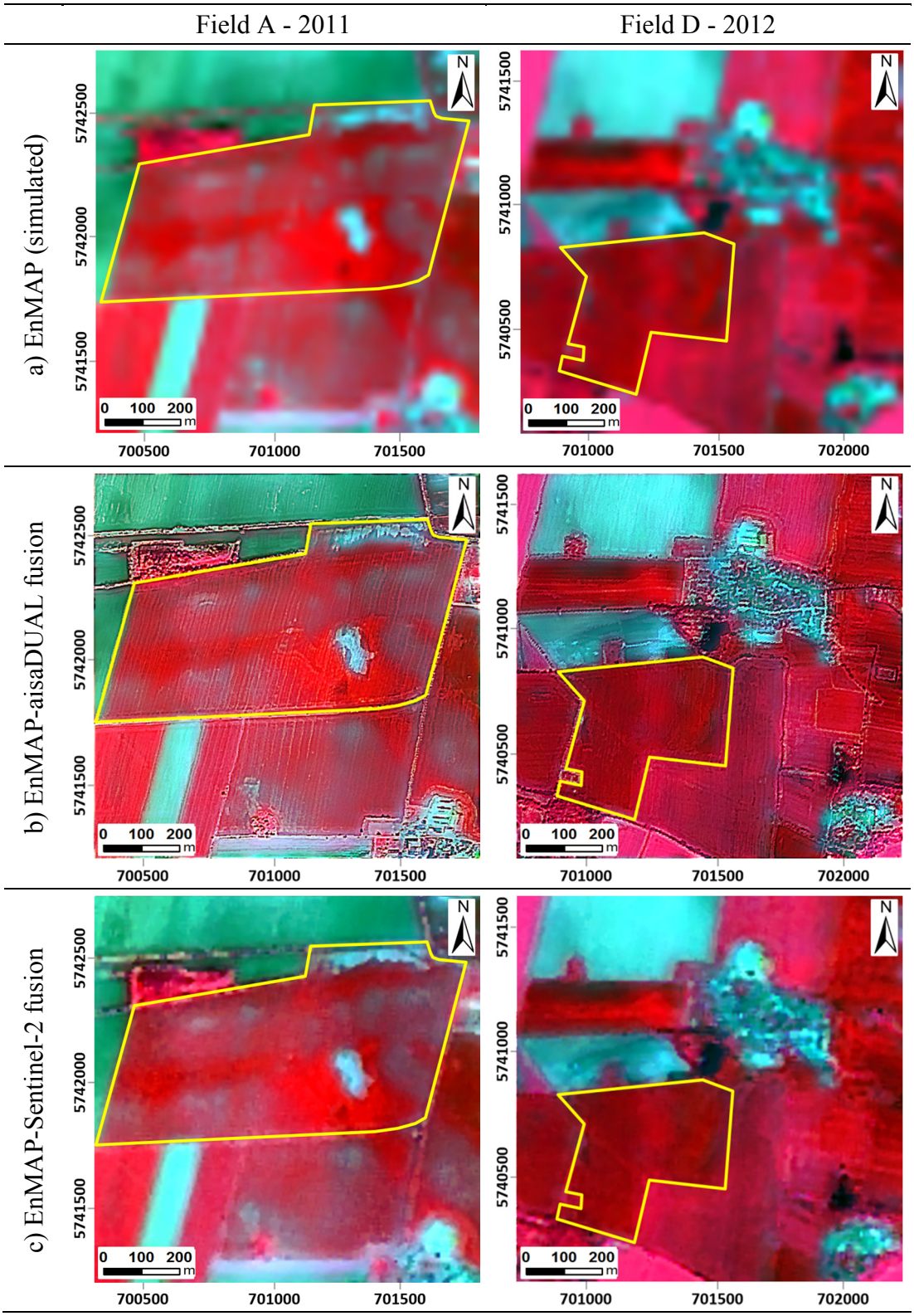


Figure 7.17: *Comparison of the EnMAP data sets with the fusion results. Displayed images in RGB (863/652/548 nm): a) simulated EnMAP data sets; b) EnMAP–aisaDUAL fusion; and c) EnMAP–Sentinel-2 fusion. The two investigated fields are indicated by the yellow polygons.*

Based on the computed α_{spec} map of each fusion result, descriptive statistics were calculated for the pixel values covering the investigated wheat fields (Table 7.12). The minimum α_{spec} values determined for the EnMAP–aisaDUAL fusion and the EnMAP–Sentinel-2 fusion were nearly the same. In contrast, the maximum α_{spec} values showed distinct differences. The maximum angles of the EnMAP–Sentinel-2 fusion results for both fields were much higher, when compared to those achieved for the EnMAP–aisaDUAL fusion results. In contrast, the mean α_{spec} values of both EnMAP–Sentinel-2 fusion results were somewhat lower. This finding indicates a slightly better spectral preservation of the EnMAP–Sentinel-2 fusion results achieved for Field A and Field D.

Table 7.12: *Descriptive statistics of α_{spec} maps (in degrees) for both years.*

	Min		Max		Mean		SD	
Field	A	D	A	D	A	D	A	D
EnMAP–aisaDUAL fusion	1.765	1.738	35.614	50.608	5.798	4.486	3.651	2.456
EnMAP–Sentinel-2 fusion	1.989	1.859	88.605	89.147	5.109	4.048	5.618	3.102

Figure 7.18 illustrates the pixel-based calculated spectral angles of the investigated fields. As already mentioned in Chapter 5.2, due to waterlogging in early spring, which allowed no plant growth, the two drainless hollows in the northern and south-eastern parts of Field A, which was investigated in 2011, were masked and not considered for α_{spec} determination.

In general, the fusion results of Field D showed a better spectral preservation capability. One reason might be the more advanced plant development stage in 2012 compared to 2011, resulting in a spectrally more homogeneous surface, which caused less problems in the fusion process. Only an area with less vegetation in the eastern part of the field and three smaller areas at the western, north-western and south-western edges of the field provided higher spectral angles and, therefore, a poorer spectral preservation compared to the rest of the field. Furthermore, some fusion artefacts present as horizontal edges crossing the middle part of the field were recognized in the EnMAP–aisaDUAL fusion result. That was most likely the result of the insufficient correction of defective image lines during aisaDUAL pre-processing (cf. Chapter 5.2.1). In comparison to Field D, the α_{spec}, maps based on the fusion results of Field A, had a distinctly higher number of cyan pixels, indicating only a medium quality spectral preservation. The yellow, orange and red areas at the edges of the drainless hollows in both α_{spec} maps of Field A represent regions with large spectral angles and, thus, a poor spectral preservation. In these transition areas, the spectral reflection was characterized by a small-scaled and fast changing mixture of vegetation

and soil, resulting in a spectral inhomogeneous surface structure, which apparently led to problems during the fusion. Additionally, the same problem was detected at the western edge of Field A in the EnMAP–aisaDUAL fusion result. In this area, the sharp spectral transition of the wheat field to the neighbouring field, which was covered with bare soil, could not be properly reproduced in the fusion result. In the EnMAP–Sentinel-2 fusion result, the problem was less pronounced. One possible reason for this was the lower spatial resolution of Sentinel-2 and, therefore, the lower spatial fusion ratio of 1:3, in comparison to the EnMAP–aisaDUAL fusion with a ratio of 1:10.

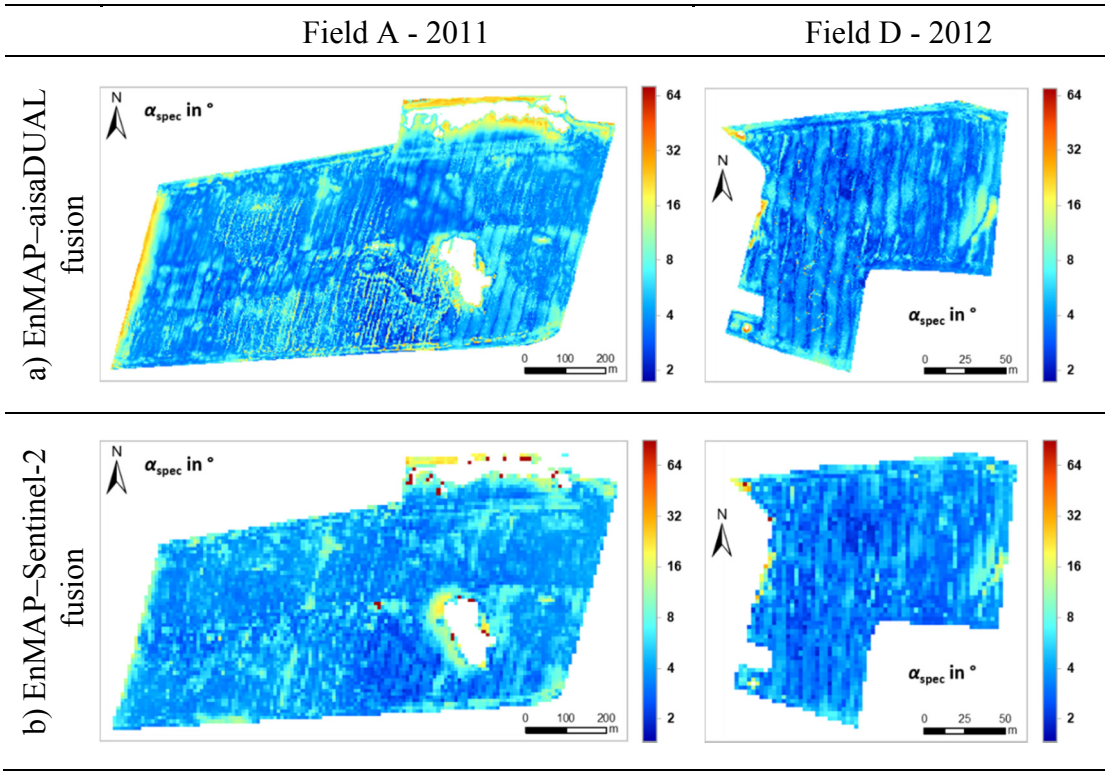


Figure 7.18: *α_{spec} maps based on a) EnMAP–aisaDUAL fusion and b) EnMAP–Sentinel-2 fusion for Field A and Field D.*

Besides the pixel-based performance of the fusion procedure, which provided information on the spatial distribution of the spectral preservation capability, it was equally important to determine the spectral quality of the single bands. Therefore, the correlation coefficient *R* was calculated for every spectral band of the fusion results. Similar to the α_{spec} calculations, *R* was computed solely for the pixels covering the investigated fields. Table 7.13 shows the general results for the EnMAP–aisaDUAL and the EnMAP–Sentinel-2 fusions. In this context, the mean value provided the most important information. Consistent with the results of the α_{spec} analysis, the EnMAP–Sentinel-2 fusion produced a better spectral preservation. This was confirmed by a higher correlation average across all the spectral bands compared to the mean *R,* which

was determined for the EnMAP–aisaDUAL fusion. With respect to the single band analysis, Figure 7.19 illustrates the correlation of every spectral band of the EnMAP–aisaDUAL fusion as well as the EnMAP–Sentinel-2 fusion with the original aisaDUAL bands. It can be seen that, on average, the EnMAP–Sentinel-2 fusion showed a higher correlation than the EnMAP–aisaDUAL fusion.

***Table 7.13**: Image statistics based on calculated correlation coefficients (R), taking into account all the spectral bands of EnMAP (204 spectral bands) for the pixels covering the two investigated fields.*

	Min	Max	Mean	SD
EnMAP–aisaDUAL fusion	0.508	0.872	0.707	0.103
EnMAP–Sentinel-2 fusion	0.520	0.889	0.773	0.066

In general, the shape of the graphs representing the *R* of the EnMAP–aisaDUAL and the EnMAP–Sentinel-2 fusions were characterized by a similar behaviour; however, in some spectral regions, the *R* values were on different levels. In the blue range of the VIS, the correlation coefficients were very low. A possible reason for this can be the very low reflectance of green vegetation in the blue part of the spectrum, which has an impact on the histogram match applied to the fused intensity component matching it to the original intensity component during the Ehlers Fusion. The same problem appeared in the red domain of VIS. However, it was less pronounced because the reflectance in this spectral range was on average a little higher. Furthermore, a sharp decline of correlation in both fusion cases was detected for the spectral bands in the red edge range. This can also be explained by one of the operations during the fusion process. To conduct the IHS transforms three successive spectral bands were used each time. In the red edge range, these bands had a strongly deviating level of reflectance height caused by the steep increase in green vegetation reflectance in this spectral region. This probably had a negative impact on the generated intensity component, which was necessary for the further steps in the fusion procedure. The same effect was also detected for the spectral bands around 1,150 nm and less pronounced around 940 nm. This fact can also be explained by an abrupt change in the reflectance level of green vegetation because of the small water vapour absorption bands in these spectral ranges. In general, the spectral preservation was on a higher level in the NIR domain compared to the VIS. Maybe this is due to the fact that the reflectance of green vegetation is distinctly higher in this spectral range, which obviously had a positive impact on the fusion results achieved with the Ehlers method. This assumption was confirmed by the lower *R* values determined for both fusion cases in the SWIR domain, which is also characterized by lower spectral reflectance due to strong water absorption in this spectral range. Moreover, it is important to note that the

fusion procedure led to a distinctly better spectral preservation of the SWIR spectral bands in the case of merging EnMAP with Sentinel-2, compared to the results obtained in case of the EnMAP–aisaDUAL fusion.

Figure 7.19: *Correlation coefficients (R) calculated between the original aisaDUAL data sets and the fusion results for the single spectral bands: the grey line represent a typical green vegetation reflectance spectra added for better interpretation and understanding purposes. Furthermore, the red bars mark the location of the most important spectral bands for LAI prediction, identified with the help of RReliefF.*

Since only three EnMAP spectral bands were necessary in order to accurately predict the LAI, the spectral preservation of these bands was of particular interest. The location of the three spectral bands is marked by red bars in Figure 7.19 and Table 7.14 shows the corresponding *R* values. While the correlation coefficients of the spectral band located at 1,074 nm were on a high level and almost similar for both fusion cases (EnMAP–aisaDUAL: $R = 0.812$, EnMAP–Sentinel-2: $R = 0.808$), the achieved *R* values for the other two spectral bands showed some differences. At 1,203 nm, the spectral preservation was better for the EnMAP–aisaDUAL fusion ($R = 0.834$) than for the EnMAP–Sentinel-2 fusion ($R = 0.764$). In contrast, the *R* value for the band located at 1,710 nm was distinctly higher in the case of the EnMAP–Sentinel-2 fusion ($R = 0.801$), compared to the EnMAP–aisaDUAL fusion ($R = 0.747$).

Based on these results, it was difficult to assess whether the fused EnMAP–aisaDUAL images or the EnMAP–Sentinel-2 images were better suited to the assessment of LAI. For that reason, comparable to Chapter 7.3.1, PLSR models were again built with the spectral information of the image data sets. Therefore, the reflectance spectra of the fused image pixels, which corresponded to the geographic location of the different sampled wheat plots of Field A and Field D were extracted and afterwards the spectral information were reduced to the three spectral bands located at 1,074, 1,203, and 1,710 nm.

***Table 7.14**: Correlation coefficients (R) of the fusion results for the three most important EnMAP spectral bands, which were ascertained for the assessment of wheat LAI.*

RReliefF regions	Highest ranked RReliefF band	Fused data sets	*R*
4	1074 nm	EnMAP–aisaDUAL	0.812
		EnMAP–Sentinel-2	0.808
5	1203 nm	EnMAP–aisaDUAL	0.834
		EnMAP–Sentinel-2	0.764
8	1710 nm	EnMAP–aisaDUAL	0.747
		EnMAP–Sentinel-2	0.801

The model built for the EnMAP–aisaDUAL fusion resulted in a slightly lower model performance (R^2_{cv} = 0.742, $RMSE_{cv}$ = 0.713), compared to the EnMAP–Sentinel-2 model (R^2_{cv} = 0.760, $RMSE_{cv}$ = 0.689). The slightly higher model quality achieved for the EnMAP–Sentinel-2 fusion was also confirmed by an RPD_{cv} value higher than 2.0, while the RPD_{cv} of the EnMAP–aisaDUAL fusion model was slightly below 2.0. The results are summarized in Table 7.15. In order to achieve better comparability, the results of the three-band aisaDUAL and the three-band EnMAP model are again presented in the table.

***Table 7.15:** Cross-validated results of wheat LAI prediction, based on PLSR for the three-band models of aisaDUAL, EnMAP, EnMAP–aisaDUAL fusion and EnMAP–Sentinel-2 fusion.*

n = 70	No. of bands	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	LV
aisaDUAL (RReliefF)	3	0.892	0.461	3.069	3
EnMAP (RReliefF)	3	0.678	0.814	1.740	3
EnMAP–aisaDUAL fusion	3	0.742	0.713	1.985	3
EnMAP–Sentinel-2 fusion	3	0.760	0.689	2.055	3

Due to the almost equal regression functions and calculated model evaluation criteria, the three-band models built for both fusion results have almost identical scatter plots (Figure 7.20). One reason for the equal performance of the fusion models can be the spatial fusion ratio. The higher the ratio, the more difficult the filtering of the intensity component becomes in the Fourier space. For this reason, the fusion ratio of 1:3 in the case of the EnMAP–Sentinel-2 fusion was obviously well suited for a robust LAI estimation of the investigated fields. Using a higher ratio, such as 1:10 in the EnMAP–aisaDUAL fusion, allowed for no additional improvement in the LAI prediction.

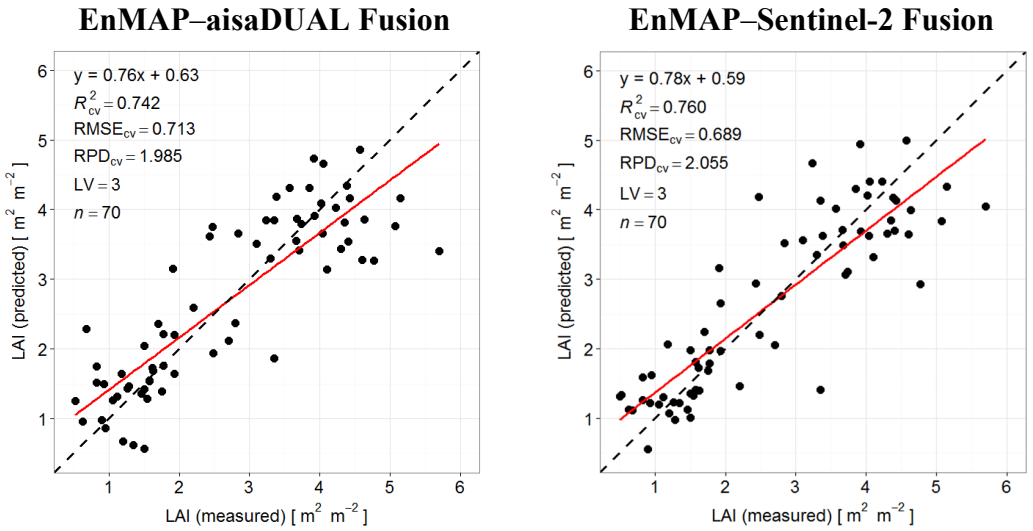


Figure 7.20: *Scatter plots of estimated (cross-validated) and measured LAI for winter wheat, based on PLSR models for three spectral bands of the fused EnMAP–aisaDUAL images and the EnMAP–Sentinel-2 images. The solid red line represents the regression line, while the dashed black line represents the 1:1 line.*

Although the quality of both three-band fusion models were considerably lower than that for the aisaDUAL three-band model, an improved LAI prediction was possible in comparison to the EnMAP three-band model. The general trend of overestimating lower and underestimating higher LAI values was also observed in both fusion cases. This suggested a successful pan-sharpening because the fused datasets obviously contain spectral information, which is comparable to that of the simulated EnMAP data, as well as the original aisaDUAL hyperspectral data.

Finally, both three-band fusion models were transferred to the corresponding images of Field A and Field D to enable a spatial LAI prediction. Figure 7.21 shows the LAI maps obtained for the fused EnMAP–aisaDUAL images and the fused EnMAP–Sentinel-2 images in comparison to the maps already generated for aisaDUAL and EnMAP in the previous chapter. The spatial LAI prediction from the regression models, based on the aisaDUAL (Figure 7.21 a)) and the EnMAP data sets (Figure 7.21 b)), served as reference for the LAI estimates, which were achieved by the three-band regression models of the fused data sets. The maps derived from the EnMAP–aisaDUAL fusion results (Figure 7.21 c)) enabled a more detailed spatial LAI assessment compared to the EnMAP results, as well as showed more precise correspondence with the aisaDUAL maps. However, areas of Field A with low LAI in the eastern and south-western parts of the field, along with areas with high LAI around the drainless hollow in the south, were underestimated. For Field D, the largest differences occurred at the eastern and southern edges of the field, where the LAI was again underestimated. In general, the spatial improvement by a factor of ten seemed to be problematic. In this context, the LAI map of Field D showed some unrealistic

structures. These structures reflected artefacts caused by the fusion process using images with a distinctly different GSD and, thus, a high fusion ratio (1:10). Although the LAI map of Field A had no visible artefacts, the spatial improvement appeared

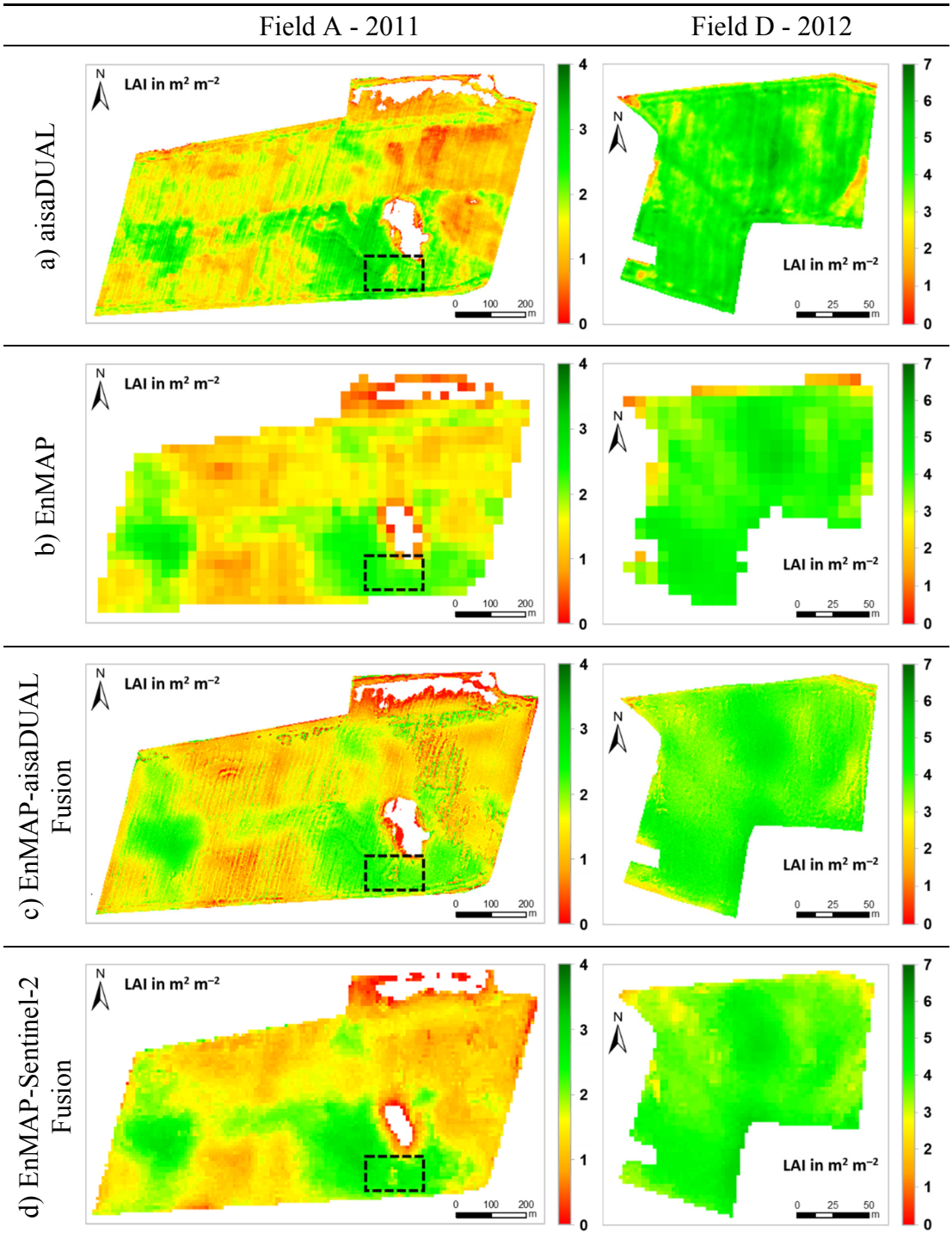


Figure 7.21: *Spatial LAI prediction based on a) aisaDUAL, b) EnMAP, c) EnMAP–aisaDUAL fusion and d) EnMAP–Sentinel-2 fusion. The dashed rectangle indicates the area presented in Figure 7.22.*

exaggerated. This fact indicated that the filter settings for the high pass pan filter were set to a frequency that was probably too low and, thus, included too much spatial information from the pan images during the fusion process. The LAI maps of both fields, which were derived from the EnMAP–Sentinel-2 fusion results (Figure 7.21 d)), provided a more realistic LAI distribution within the fields. Due to the lower spatial resolution, very small-scaled structures within the field, which can be detected in the aisaDUAL LAI maps, cannot be recognized. In contrast to the EnMAP LAI maps, however, more detailed LAI predictions were possible. The EnMAP–Sentinel-2 LAI map of Field D showed a high agreement with the aisaDUAL LAI map of Field D. The map of Field A also reflected the spatial pattern of LAI very well. Only two areas with low LAI in the northern part of the field were underestimated in comparison to the corresponding areas in the aisaDUAL LAI map.

As an additional criterion for evaluating the fusion results, a profile covering an area with highly variable LAI over a short distance was extracted from all LAI maps of Field A. In Figure 7.21, the area of interest was bordered with a dashed black line in all LAI maps. Figure 7.22 illustrates the enlarged areas with the profiles on the right side and a line graph with the extracted LAI of the pixels on the left side.

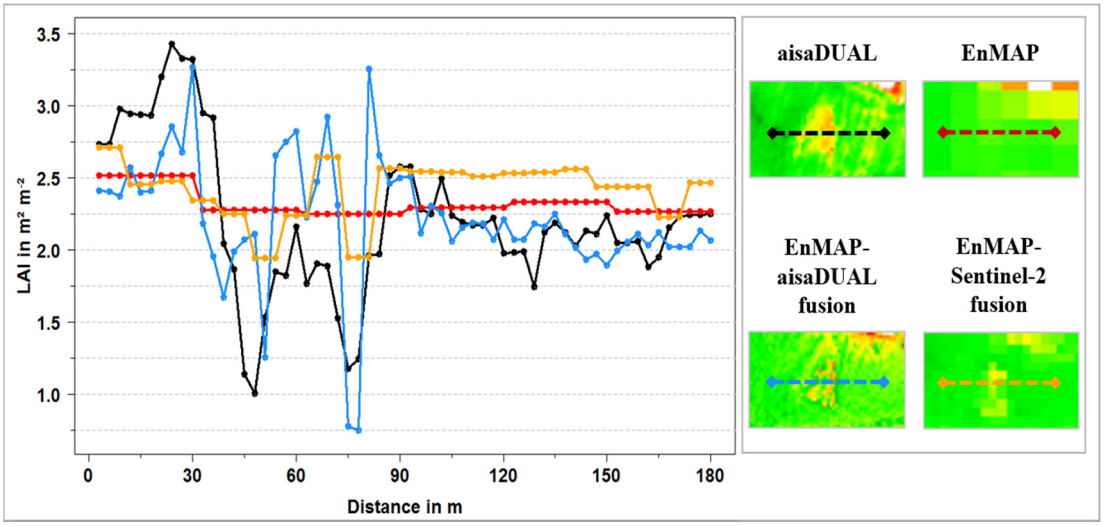


Figure 7.22: *LAI profile for a specific area of the field investigated in 2011 based on a) aisaDUAL (black), b) EnMAP (red), c) EnMAP–aisaDUAL fusion (blue), and d) EnMAP–Sentinel-2 fusion (orange).*

The black graph represents the reference profile of the aisaDUAL LAI map. The profile has a length of 180 m, which corresponds to 60 pixels and, thus, 60 LAI values. At the beginning, the LAI increased from 2.75 to 3.50, after which it gradually decreased to a value of approximately 1.00. In the range between 45 and 75 m, the LAI went up and down again. After 75 m, the LAI increased again sharply, then levelled out at a value of about 2.25 with slight fluctuations. The red graph showed the profile

extracted from the EnMAP LAI map. Due to the spatial resolution, one EnMAP pixel corresponded to ten aisaDUAL pixels. Therefore, the EnMAP profile could not reflect the curve of the aisaDUAL profile and had almost the same value over the entire distance, which corresponded approximately to the average of the aisaDUAL profile. In general, the shape of the EnMAP–aisaDUAL fusion LAI profile was well in accordance with the aisaDUAL profile. Only in areas characterized by sharp changes along the profile, the LAI was slightly over- or underestimated. The same problem, albeit more pronounced, was also observed in the EnMAP–Sentinel-2 fusion profile. While the LAI values were underestimated at the beginning of the profile, an overestimation could be observed from a distance of 30 m onwards. In contrast to the EnMAP profile, however, the EnMAP–Sentinel-2 graph showed a better match with the aisaDUAL graph in most areas. In both fusion cases, it seems the Ehlers Fusion had some difficulties in dealing with the abrupt changes in LAI from high to low values, and back, over short distances. Nevertheless, in comparison to the EnMAP graph, the LAI profiles, which were based on both fusion results, better reflected the profile of aisaDUAL on average.

Comparison of Aggregation and Fusion

Both aggregation and fusion can be used to evaluate the quality of spatial LAI predictions based on EnMAP image data. In each case, however, image data of higher spatial resolution are a necessary requirement to enable a robust validation.

During the aggregation process applied in this thesis, the higher spatial resolution aisaDUAL LAI maps of both fields were reduced to the spatial resolution of EnMAP. As a result, each aggregated aisaDUAL pixel represented an averaged value of several original aisaDUAL pixels. Due to the averaging, the small-scale LAI variability within the investigated fields was lost. Although the aggregation approach has been recognized as an effective way to validate satellite products with medium spatial resolution in different studies [e.g., Cohen et al., 2003], its feasibility is influenced by the limited number of high-resolution images [Shi et al., 2015]. Moreover, the acquisition of airborne hyperspectral data, such as the aisaDUAL data used in this thesis, is very expensive, time-consuming and usually only covers small areas. In contrast, the fusion method only requires a spatial high-resolution panchromatic data set to sharpen the EnMAP images. Such data sets are more often available compared to high-resolution multi-/hyperspectral data. Several medium-resolution multispectral satellites with relatively short revisit cycles also have a panchromatic band with higher spatial resolution (e.g., Landast 8 Pan (GSD: 15 m) [Roy et al., 2014], ALI Pan (GSD: 10 m) [Chander et al., 2009], SPOT-5 Pan (GSD: 5 m) [Gleyzes et al., 2003]), which is well suited to the pan-sharpening of EnMAP image data. Moreover, due to the fact that the fusion method was applied directly to the EnMAP image data, rather

than to the LAI maps based on EnMAP, the small-scale LAI variability within the fields could be largely preserved. This is an advantage of the fusion over the aggregation method for applications in precision agriculture. However, some problems remain, even though the first EnMAP–aisaDUAL and EnMAP–Sentinel-2 fusion results were promising. In this thesis, all used data sets were based on aisaDUAL airborne data. Consequently, all data sets exhibit the same acquisition conditions (e.g., same day and time, same atmospheric conditions, same sensor view angles) and are optimally co-registered with each other. Given that, in future work, the pan-sharpening of real EnMAP images with real Sentinel-2 data may cause problems, which could not be considered in this thesis. In this context, different acquisition dates and atmospheric conditions can lead to major problems during the pan-sharpening process. Nevertheless, using simulated data in this thesis, acquired under the exact same conditions, can also be regarded as an advantage. Thus, it was possible to identify problems caused by the applied pan-sharpening algorithm itself.

8 Conclusions

Remote sensing can be regarded as one of the key tools in precision agriculture because it allows for the spatial assessment of important crop parameters and, thus, supports decision-making for adapted intra-field treatment [Mulla, 2013]. Precise spatial information on the current growth conditions of wheat is of particular interest because wheat is one of the most vital food grain sources for humans [CIMMYT, 2015]. In order to monitor the actual status of wheat development, the area-wide determination of LAI is critical because it serves as an indicator of the photosynthetic performance of plants and, therefore, is a decisive factor in agricultural modelling [Moran et al., 1995].

For that reason, the scientific focus of this thesis was to investigate the potential of hyperspectral remote sensing data acquired at different spatial scales for the precise assessment of wheat LAI. Three research experiments were conducted for the following reasons: 1) to identify the empirical-statistical regression technique that enables the most robust LAI predictions; 2) to determine which spectral resolution and which spectral regions or single bands are essential for a precise LAI assessment; and 3) to investigate the influence of the ground sampling distance of remote sensing images on the quality of spatial LAI predictions. The results and scientific findings of the three research experiments are concluded in the following paragraphs.

Research question 1: What potential do empirical-statistical regression models have for an accurate prediction of crop LAI based on hyperspectral data?

In Chapter 7.1, three regression algorithms (PLSR, SVR, RFR) were investigated in order to predict wheat LAI from field reflectance spectra, which were collected from different fields and at different plant development stages. The results clearly show that model performance markedly depends on the validation technique used to assess model accuracy. In the case of leave-one-out cross-validation, SVR provided the best results for individual years and the entire data set, but revealed significant weaknesses when independent model validation was performed. This suggested a problem in data overfitting during model calibration, which cannot be proven by leave-one-out cross-validation, as model calibration and validation are performed using almost identical data. PLSR and RFR yielded a lower performance for leave-one-out cross-validated models of individual years. Although model accuracy for the entire data set, achieved with both regression algorithms, showed a clear improvement, it was still lower compared to SVR. However, in contrast to SVR model accuracy, independently validated PLSR and RFR models provided nearly identical model performance, similar to that obtained with leave-one-out cross-validation. This indicated that PLSR and RFR are less influenced by data overfitting, meaning that they are less sensitive to the

validation techniques used. However, PLSR provided distinctly better results than RFR and was, therefore, better suited for predicting the parameter LAI in this thesis.

Furthermore, it was shown that an extended data range through the inclusion of the LAI data at different stages of plant development was essential for improving model performance for all regression algorithms considered. Therefore, empirical-statistical regression models, calibrated with hyperspectral reflection, and corresponding LAI measurements of numerous growth stages, which cover a wide range of values, have the capability to provide accurate LAI predictions from unknown wheat spectral measurements acquired within the same region.

The portability of locally developed regression models has often been criticized for their poor results when applied to data acquired under varying canopy (e.g., phenological stage, canopy geometry) and external conditions (local soil properties, atmospheric conditions, sun and view geometry) [Baret and Guyot, 1991; Dorigo et al., 2007]. The results achieved in this thesis, however, show that PLSR has the capability to accurately predict the LAI of wheat from spectral data, which are not involved in model calibration. Therefore, it seems at least possible to transfer PLSR models within a region with comparable soil conditions. The influence of further external factors on the transferability of a regression model, such as atmospheric conditions and sun and view geometry, may be reduced to a minimum by collecting the reflectance spectra, as well as the LAI measurements, under comparable irradiance conditions always using the same measurement set-up. Moreover, in order to take into account varying canopy conditions, data from different plant development stages were used in model calibration. This was another important step to enable the calibration of PLSR models with high predictive power when applied to unknown data.

The robust evaluation of an empirical-statistical relationship is mandatory [Dorigo et al., 2007], as well as being an important research topic in many remote sensing studies, especially in relation to agricultural applications. The conducted research experiment demonstrated that independent validation is necessary to verify the ‘real’ robustness of a regression model. PLSR and RFR, however, proved to be less influenced by the applied type of validation. This finding needs to be further investigated in future work, in which the calibration and validation of different empirical-statistical regression models should be conducted by using spatially independent data sets collected from different regions.

Based on the comparison of different empirical-statistical regression models and validation techniques, the following main conclusions can be drawn:

- *The data range of LAI used in model building strongly affected the quality of the prediction model.* Including data from different phenological stages during

wheat development distinctly improved the accuracy of all applied regression techniques (PLSR, SVR, RFR).

- *The validation technique had a major influence on the predictive power of the generated empirical-statistical regression models.* PLSR proved to be superior to SVR and RFR when independent validation was applied, while SVR outperformed the other regression techniques in the case of leave-one-out cross-validation.

Research question 2: What spectral resolution and location of spectral bands are necessary to enable a precise estimation of the parameter LAI?

In Chapter 7.2, it was first examined whether the spectral characteristics of aisaDUAL and EnMAP were suitable for the precise assessment of wheat LAI. For that reason, the field spectral measurements (867 spectral bands) were resampled to the spectral resolution of aisaDUAL (98 spectral bands) and EnMAP (204 spectral bands). Subsequently, PLSR models were calibrated to determine the predictive power of both data sets. The results clearly showed that the same high model qualities were achieved for the models, based on the field spectral measurements resampled to the spectral resolution of aisaDUAL and EnMAP, respectively, as for the model calibrated with full-resolution field spectra. The fact that the data set with the lowest spectral resolution (aisaDUAL) still provided sufficient spectral information for a precise LAI retrieval is consistent with results achieved by Atzberger et al. [2010]. The authors also used PLSR modelling for the precise assessment of wheat LAI, based on a hyperspectral data set with comparable spectral characteristics to those of aisaDUAL.

Since several studies have proved that the LAI can be derived from lower-resolution spectral data [e.g., Asner, 1998], the second part of Chapter 7.2 was aimed at identifying the most important spectral bands/regions for an accurate LAI assessment. For this purpose, the RReliefF feature selection algorithm was applied because this method enables the selection of important spectral bands prior to the regression model calibration and, therefore, can be combined with any empirical-statistical regression technique. RReliefF allowed for the selection of eight spectral regions, each of which consisted of more than one band. This is in accordance with numerous studies in which several spectral regions had been identified as sensitive in order to predict the LAI [e.g., Lee et al., 2004; Thenkabail et al., 2004]. Due to the high correlation of adjacent spectral bands, a randomized experiment was conducted to prove whether the highest-ranked RReliefF band of each of the eight regions contributed the most important information for the prediction of LAI. In most regions, the spectral band selected by randomized sampling was very close to the highest rated RReliefF band, or even had exact the same position. Consequently, it could be demonstrated that the spectral bands

with the highest RReliefF rating in the different regions had high predictive power to estimate the LAI.

Additionally, a seconded randomized experiment was conducted to test whether it was necessary to use the highest-ranked RReliefF band of each region to obtain precise LAI predictions. The achieved results proved that the highest-ranked bands of Region 4 (aisaDUAL: 1,070 nm, EnMAP: 1,074 nm), Region 5 (aisaDUAL and EnMAP: 1,203 nm), and Region 8 (aisaDUAL: 1,713 nm, EnMAP: 1,710 nm) were sufficient to predict the LAI with high accuracy. These findings were consistent with studies of Asner [1998] (Regions 5 and 8), Darvishzadeh et al. [2008b] (Region 8), Lee et al. [2004] (Region 8), and Thenkabail et al. [2004] (Region 4), in which similar wavelength positions were found to be most sensitive to changes in LAI. For that reason, it can be concluded that applying RReliefF enabled the selection of the three most important spectral bands. At the same time, however, additional spectral bands were also identified as important, which were unable to contribute to a significant improvement in LAI prediction. The selection of these spectral bands was possibly the result of the high correlation of LAI with other plant parameters (e.g., Cab, DM, PWC), which influence the spectral reflectance in this range of the electromagnetic spectrum [Verrelst et al., 2012].

It can be summarized that RReliefF was a supportive algorithm in determining the relevant spectral bands for the assessment of LAI. The interpretation of the RReliefF weightings obtained for each spectral band, however, required expert knowledge to identify important spectral regions/bands. Additionally, a repeated randomized sampling procedure was necessary to finally ascertain the three most important spectral bands. For that reason, the entire feature selection applied in this thesis was a semi-automatic procedure.

Since the three-band aisaDUAL and EnMAP PLSR models provided almost identical high prediction qualities as the models built with all the spectral bands, it could be proved that PLSR is able to provide precise parameter predictions from numerous highly correlated spectral bands, as well as from a few spectral bands located in specific wavelengths regions [An et al., 2015]. Although a feature selection was not necessary to predict the LAI with high accuracy from hyperspectral data, the identification of the most important spectral bands was advantageous in proving whether multispectral satellites, which normally have shorter revisit cycles and larger spatial coverage than hyperspectral systems, provide information in the same spectral regions. Unfortunately, frequently used multi-/superspectral satellites, such as Landsat 8 and Sentinel-2, do not have spectral bands at the locations identified as relevant by this thesis.

Based on the experiment aimed at determining the required spectral resolution and location of the most relevant spectral bands for a precise LAI prediction, the following main conclusions can be drawn:

- *The spectral resolution of aisaDUAL (98 spectral bands) and EnMAP (204 spectral bands) was suitable for the precise assessment of wheat LAI.* In comparison to the PLSR model calibrated with full-resolution field spectra, the same high model qualities could be achieved for the PLSR models, based on the field reflectance measurements resampled to aisaDUAL and EnMAP spectral resolution.

- *The RReliefF feature selection algorithm, in combination with a randomized sampling procedure, enabled the identification of the most important spectral bands for the assessment of wheat LAI.* Two spectral bands on the NIR plateau and one spectral band in the SWIR I region were selected as most sensitive to changes in LAI from both the aisaDUAL and the EnMAP data sets.

Research question 3: What influence does the ground sampling distance of remote sensing images have on the spatial assessment of LAI and the associated regression model validation?

In Chapter 7.3, the findings of the previous chapters were initially used to build PLSR models, based on aisaDUAL and EnMAP image data, in order to predict the spatial LAI distribution of two wheat fields. Therefore, PLSR models were generated with all the spectral bands of each sensor, and then with only the three most relevant spectral bands of aisaDUAL and EnMAP, which were identified during the feature selection procedure in Chapter 7.2.2. The results based on the image spectra of both sensors showed that the three-band models had the same predictive power as the models generated with all spectral bands.

The models built with the aisaDUAL image spectra exhibited only slightly lower prediction accuracies compared to the models based on the field spectral measurements resampled to aisaDUAL spectral resolution. In contrast, distinctly poorer model qualities were achieved for the models based on EnMAP image spectra in comparison to the models generated with the field spectral measurements resampled to EnMAP spectral resolution. This is explained by the different spatial resolutions of the aisaDUAL (3 m) and EnMAP images (30 m). In the case of aisaDUAL, the collected LAI values of the field plots (each with a size of 0.25 m^2) were sufficiently representative of pixels of 9 m^2 in size. In contrast, the small field plots were definitely unsuitable for robust regression model calibration and validation, based on EnMAP pixels, each of which has a size of 900 m^2.

In order to enable a precise assessment of wheat LAI from EnMAP image data, two different approaches were examined. First, the aisaDUAL LAI result maps were scaled up (aggregated) to the spatial resolution of EnMAP. Thus, a direct comparison of the EnMAP LAI maps and the aggregated aisaDUAL LAI maps was possible. The results showed that lower LAI values had a high level of agreement in both prediction maps, while higher LAI values exhibited deviations that were slightly more pronounced. A possible reason for this was the more complex geometry of the further developed wheat plants investigated in 2012, which may have affected LAI prediction from different spatial scales.

As a second step, the EnMAP data sets of both fields were pan-sharpened with artificially generated aisaDUAL and Sentinel-2 panchromatic images using the spectral preserving Ehlers Fusion. Image spectra of the generated EnMAP–aisaDUAL (3 m GSD) and EnMAP–Sentinel-2 fusion results (10 m GSD) of both fields were used to build PLSR models for the spatial assessment of wheat LAI. Although the qualities of the EnMAP–aisaDUAL and the EnMAP–Sentinel-2 fusion models were distinctly lower compared to the quality of the aisaDUAL model, both fusion models provided higher accuracies in comparison to the model based on EnMAP image data. Therefore, the fused data sets were deemed to be better suited for the prediction of wheat LAI compared to the EnMAP images. However, the fusion procedure also showed some problems, which can be explained by the methodology on which the Ehlers Fusion is based. In this context, the fusion ratio of the fused data sets had a substantial impact on the fusion results, but not on the LAI predictions. The fused EnMAP–aisaDUAL data sets (fusion ratio: 1:10) did not lead to a more precise LAI estimate in comparison to the prediction based on the fused EnMAP–Sentinel-2 data sets (fusion ratio: 1:3). A fusion ratio in between the two (e.g., 1:5 or 1:6) possibly would allow for more accurate LAI estimates, since artefacts occurring in the EnMAP–aisaDUAL fusion results could be avoided. Furthermore, in comparison to the fused EnMAP–Sentinel-2 images, more spatial details of the panchromatic image could be transferred into the merged data sets.

In general, merging EnMAP images with higher resolution panchromatic data provides more potential for the precise assessment of wheat LAI in precision agriculture than aggregating LAI result maps, which are based on hyperspectral airborne data, to the spatial resolution of EnMAP. The acquisition of hyperspectral airborne images is very expensive and normally only covers small areas. Thus, the aggregation method can only be used to a limited extent for the validation of parameter maps based on EnMAP images. In contrast, spatially higher resolution panchromatic satellite images or artificial panchromatic images, derived from high-resolution multispectral satellite data, are already available for large areas with high temporal resolution. These images can be used to sharpen entire EnMAP scenes. Consequently, fused EnMAP data with improved spatial resolution would provide high potential for

applications in precision agriculture. In this context, the successful fusion of simulated EnMAP and Sentinel-2 data presented in this thesis has already demonstrated the advantage of combining the spectral and spatial characteristics of both satellite systems. In future work, however, the true benefit of merging 'real' EnMAP and 'real' Sentinel-2 data for the assessment of LAI and other plant parameters needs to be examined.

Based on the experiment determining the influence of the spatial resolution on a precise LAI prediction, the following main conclusions can be drawn:

- *The spatial resolution of aisaDUAL and EnMAP had a major impact on the validation of the regression models and, therefore, on the achieved LAI prediction accuracies.* While the size of the sampled field plots enabled robust model calibration and validation based on aisaDUAL airborne data, regression modelling based on EnMAP image data led to poor results because of the distinct difference in size between the EnMAP pixels and the sampled field plots.

- *The pan-sharpening of EnMAP image data offers great potential for the assessment of LAI in precision agriculture.* The EnMAP–aisaDUAL and the EnMAP–Sentinel-2 fusion results of the wheat fields allowed for a distinctly more precise LAI prediction, when compared to that achieved from the EnMAP data with the original GSD of 30 m.

9 Outlook[18]

Given the increasingly growing world population and the limited availability of arable land, the further dissemination of precision agriculture is urgently needed to increase crop productivity and farm profitability, while reducing negative environmental impacts [European Commission, 2012; Mulla, 2013]. Since the concept of precision agriculture is based on up-to-date spatial information of variations in soil and crop conditions, remote sensing can be regarded as one of the key technologies in this field, given that airborne and spaceborne sensors provide this information for large areas [Gebbers and Adamchuk, 2010; Robert, 2002].

Although remote sensing image data contain information about the actual crop status, the temporal, spectral and spatial resolution of the data sets, as well as the applied retrieval methods, have a major influence on the prediction accuracy of plant parameters, such as the LAI. For this reason, different aspects regarding the assessment of wheat LAI from hyperspectral remote sensing data were investigated in this thesis, in order to show the benefits as well as the limitations of this information source in relation to applications in precision agriculture.

The combination of the RReliefF spectral feature selection technique, which was expanded by a randomized sampling procedure, with PLSR modelling proved to be reliable for the precise assessment of wheat LAI from field reflectance measurements resampled to the spectral characteristics of aisaDUAL airborne and EnMAP spaceborne data. In future research, the identification of important spectral bands using the developed feature selection procedure can be helpful in designing new VIs. These VIs, would possibly be less affected by the main problems of numerous established VIs used to predict the LAI: namely, the influence of several plant parameters on the index value and the problem of saturation from moderate to high LAI values [Govaerts et al., 1999; Nguy-Robertson et al., 2014].

This thesis, however, has only focussed on the parameter LAI using data from one study site. Since additional information about crops is required to monitor the growth conditions, follow-up studies should be aimed at the prediction of other plant parameters (e.g., Cab, DM) using data sets, which are collected from several regions at different stages of plant development. This would help to evaluate the robustness of the combined use of RReliefF and PLSR modelling. Furthermore, data from different study sites would facilitate a spatially more independent validation, which would help to improve the spatial transferability of empirical-statistical regression models. A study investigating this topic was recently conducted by Gerighausen et al. [2016]. The authors demonstrated that PLSR models, which were developed with data collected in one region, could be successfully transferred to another region in order to predict the

[18] Parts of this section have previously been published by Siegmann and Jarmer [2015] and Siegmann et al. [2015].

LAI, as long as the data range and the distribution of measured LAI values are comparable in both regions. Furthermore, the spatial validation of a predicted parameter map (e.g., LAI map) strongly depends on a meaningful interpretation method, which provides quantifiable prediction accuracies on pixel basis. Normally validation data are only available for a small number of pixels within an investigated field (location of sampled plots). New non-parametric non-linear regression techniques, such as Gaussian processes regression, also provide uncertainty measures for every predicted pixel in a parameter map [Fang et al., 2012; Verrelst et al., 2015]. This is very new research topic, which needs to be further explored in the future in order to ascertain the area-wide prediction capability of empirical-statistical regression models at pixel level.

Beside the spatial transferability of regression models and the spectral characteristics of remote sensing images, their spatial resolution is of the utmost importance. Precision agriculture requires image data of equal to or less than 10 m GSD in order to enable an adapted treatment of fields on sub-field scale. Although the aisaDUAL data met this requirement, airborne campaigns are cost-intensive and normally only cover small areas. In contrast, future spaceborne imaging spectrometers, such as EnMAP and HyspIRI, provide data with larger spatial dimesnsions. These data have the potential to serve as an information base for the fast, cost-effective and non-destructive assessment of relevant biochemical and structural crop properties. The medium spatial resolution of EnMAP and HyspIRI image data, however, will be unsuitable for precision agricultural applications. To overcome this problem simulated EnMAP images were sharpened with panchromatic data sets of higher spatial resolution in this thesis. The results showed that the fused EnMAP images allowed for more accurate LAI predictions, when compared to the original EnMAP data with 30 m spatial resolution. Therefore, pan-sharpening can be regarded as an adequate method to make EnMAP data more usable for the retrieval of LAI and other crop parameters in precisions agriculture.

Since the entire pan-sharpening procedure and the subsequent LAI retrieval were based on simulated data, panchromatic bands of operational satellites (e.g., SPOT 5 (GSD: 5 m), EO-1 ALI (GSD: 10 m)) or artificial panchromatic bands, which were derived from multi-/superspectral satellite information (e.g., RapidEye (GSD: 5 m), Sentinel-2 (GSD: 10)), need to be integrated in the fusion process as a next step in investigating the capability of fused EnMAP data for the assessment of LAI and other important crop parameters. Moreover, in future research, more attention must be paid to an inter-comparison of the Ehlers method with other techniques enabling the fusion of hyperspectral data sets, such as the Hyperspherical Color Sharpening (HCS), introduced by Padwick et al. [2010], or the wavelet-based Bayesian fusion developed by Zhang [2012] and Zhang et al. [2009]. Besides an inter-comparison, uniform fusion evaluation criteria have to be defined to ensure an objective validation of the fusion

results. Although numerous criteria measuring the spectral preservation already exist, there is still a great demand for meaningful spatial evaluation criteria.

With the forthcoming hyperspectral satellite missions, the further dissemination of precision agricultural practices will be strongly supported. Hyperspectral, in combination with temporal and spatial higher resolution multi- and superspectral satellite image data, will enable the area-wide monitoring of crops at several stages of plant development over a growing season. In particular, the spatially differentiated assessment of LAI, as shown in this thesis, in combination with remote sensing information on crop DM [e.g., Marshall and Thenkabail, 2015] offer great potential for improving yield estimates [e.g., Dente et al., 2008]. This is because both parameters serve as essential variables for assimilating remote sensing image data into crop growth models [Fang et al., 2008; Moulin et al., 1998]. Thus, using temporal, spatial and spectral information from different satellites in an integrated manner provides the possibility for optimizing yield quantity and quality [Whelan and Taylor, 2013], which can help to secure sufficient food supplies for the world's growing population. The further spread of remote sensing image data as a valuable source of information in agriculture, however, strongly depends on the successful integration of the image data and derived products into agricultural decision support systems [Kneubühler, 2002], which will make it easier for agriculturalists to benefit from the large amount of information.

References

Aasen, H., Burkart, A., Bolten, A., and Bareth, G. (2015). Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:245–259.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Alados, I., Foyo-Moreno, I., and Alados-Arboledas, L. (1996). Photosynthetically active radiation: measurements and modelling. *Agricultural and Forest Meteorology*, 78(1–2):121–131.

Alexandratos, N. and Bruinsma, J. (2012). World agriculture towards 2030/2050: The 2012 revision. ESA Working Paper No. 12-03. Available online: http://www.fao.org/docrep/016/ap106e/ap106e.pdf (accessed on 29. September 2015).

Allen, W. A., Gausman, H. W., Richardson, A. J., and Thomas, J. R. (1969). Interaction of isotropic light with a compact plant leaf. *Journal of the Optical Society of America*, 59(10):1376–1379.

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., and Bruce, L. (2007). Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021.

Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., and Camps-Valls, G. (2013). Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 23:132–141.

An, N., Goldsby, A. L., Price, K. P., and Bremer, D. J. (2015). Using hyperspectral radiometry to predict the green leaf area index of turfgrass. *International Journal of Remote Sensing*, 36(5):1470–1483.

Ardilly, P. and Tillé, Y. (2006). *Sampling Methods*. Springer New York.

Arenas-Garcia, J. and Camps-Valls, G. (2008). Efficient kernel orthonormalized PLS for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, 46(10):2872–2881.

Asam, S. (2014). *Potential of high resolution remote sensing data for Leaf Area Index derivation using statistical and physical models*. PhD thesis, Julius-Maximilians-Universität Würzburg.

Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64(3):234–253.

Atzberger, C. (1998). *Estimates of winter wheat production through remote sensing and crop growth modelling.* PhD thesis, Trier University.

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment*, 93(1–2):53–67.

Atzberger, C. (2010). Inverting the PROSAIL canopy reflectance model using neural nets trained on streamlined databases. *Journal of Spectral Imaging*, 1(1):1–13.

Atzberger, C., Guérif, M., Baret, F., and Wernert, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture*, 73(2):165–173.

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3):411–421.

Bacour, C., Jacquemoud, S., Leroy, M., Hautecoeur, O., Weiss, M., Prévot, L., Bruguier, N., and Chauki, H. (2002). Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne polder data. *Agronomie*, 22(6):555–565.

Bajwa, S. G., Bajcsy, P., Groves, P., and Tian, L. F. (2004). Hyperspectral image data mining for band selection in agricultural applications. *Transactions of the American Society of Agricultural Engineers*, pages 895–907.

Baltsavias, E. P. (2002). Special section on image spectroscopy and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(3):169–170.

Bannari, A., Khurshid, K. S., Staenz, K., and Schwarz, J. (2008). Potential of Hyperion EO-1 hyperspectral data for wheat crop chlorophyll content estimation. *Canadian Journal of Remote Sensing*, 34(1):139–157.

Baret, F. and Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 35(2–3):161–173.

Baret, F., Houlès, V., and Guérif, M. (2006). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany*, 58(4):869–880.

Bastiaanssen, W. G., Molden, D. J., and Makin, I. W. (2000). Remote sensing for irrigated agriculture: examples from research and possible applications. *Agricultural Water Management*, 46(2):137–155.

Bauer, J., Siegmann, B., Jarmer, T., and Aschenbruck, N. (2016a). On the potential of wireless sensor networks for the in-situ assessment of crop leaf area index. *Computers and Electronics in Agriculture*, 128:149–159.

Bauer, J., Siegmann, B., Jarmer, T., and Aschenbruck, N. (2016b). Smart fLAIr: A smartphone application for fast LAI retrieval using ambient light sensors. *In 11th IEEE Sensors Applications Symposium (SAS)*, pages 1–5.

Bauer, J., Siegmann, B., Jarmer, T., and Aschenbruck, N. (2014). On the potential of wireless sensor networks for the in-field assessment of bio-physical crop parameters. In *2014 IEEE 39th Conference on Local Computer Networks Workshops (LCN Workshops)*, pages 523–530.

Bauer, M. E. and Cipra, J. E. (1973). Identification of agricultural crops by computer processing of ERlS MSS data. LARS Technical Reports 20, The Laboratory of Applications of Remote Sensing, Purdue University, India.

Bellman, R. E. (1962). *Adaptive Control Processes: A Guided Tour (Princeton Legacy Library)*. Princeton University Press.

Benediktsson, J., Sveinsson, J., and Amason, K. (1995). Classification and feature extraction of AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 33(5):1194–1205.

Berk, A., Bernstein, L., Anderson, G. P., Acharya, P., Robertson, D. C., Chetwynd, J. H., and Adler-Golden, S. M. (1998). MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Remote Sensing of Environment*, 65(3): 367–375.

Beven, K. J. and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1):43–69.

Bhatti, A., Mulla, D., and Frazier, B. (1991). Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images. *Remote Sensing of Environment*, 37(3):181–191.

Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany*, 58(4):855–867.

BLV (2014). *Landwirtschaftlicher Pflanzenbau*. BLV Buchverlag GmbH & Co. München., 13 edition.

Boegh, E., Soegaard, H., Broge, N., Hasager, C., Jensen, N., Schelde, K., and Thomsen, A. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment*, 81(2–3):179–193.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems*, 34(3):483–519.

Bowden, P., Edwards, J., McNee, T., Roberts, K., Schipp, A., Schulze, K., and Wilkins, J. (2008). *Wheat growth & development*. NSW Department of Primary Industries, 1 edition.

Bréda, N. J. J. (2003). Ground based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, 54(392):2403–2417.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall, New York.

Broge, N. and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2):156–172.

Broge, N. and Mortensen, J. (2002). Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote Sensing of Environment*, 81(1):45–57.

Camps-Valls, G., Bruzzone, L., Rojo-Alvarez, J., and Melgani, F. (2006). Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 3(3):339–343.

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., and Malo, J. (2011). *Remote Sensing Image Processing*. Synthesis Lectures on Image, Video & Multimedia Processing. Morgan & Claypool Publishers.

Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, 15(3):697–703.

Chander, G., Markham, B. L., and Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5):893–903.

Chang, C. C. and Lin, C. J. (2011). Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27.

Chen, J. and Leblanc, S. (1997). A four-scale bidirectional reflectance model based on canopy architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1316–1337.

Chen, J. M. and Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, 15(4):421–429.

Chen, Z., Pu, H., Wang, B., and Jiang, G.-M. (2014). Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods. *IEEE Geoscience and Remote Sensing Letters*, 11(8):1418–1422.

CIMMYT - International Maize and Wheat Improvement Center (2015). Available online: http://wheat.org/wheat-in-the-world/ (accessed on 07 October 2015).

Clevers, J. G. P. W., Jong, S. M. D., Epema, G. F., Meer, F. D. V. D., Bakker, W. H., Skidmore, A. K., and Scholte, K. H. (2002). Derivation of the red edge index using the MERIS standard band setting. *International Journal of Remote Sensing*, 23(16):3169–3184.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., and Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84(4):561–571.

Collins, W. (1978). Remote sensing of crop type and maturity. *Photogrammetric Engineering and Remote Sensing*, 44(1):43–55.

Cox, S. (2002). Information technology: the global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture*, 36(2–3):93–111.

Crookston, R. K. (2006). A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. *Crop Science*, 46(5):2253–2262.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3):271–278.

Curran, P. J. (1994). Imaging spectrometry. *Progress in Physical Geography*, 18(2):247–266.

Curran, P. J., Dungan, J. L., Macler, B. A., and Plummer, S. E. (1991). The effect of a red leaf pigment on the relationship between red edge and chlorophyll concentration. *Remote Sensing of Environment*, 35(1):69–76.

Darvishzadeh, R., Atzberger, C., Skidmore, A., and Schlerf, M. (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6):894–906.

Darvishzadeh, R., Skidmore, A., Atzberger, C., and van Wieren, S. (2008a). Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture. *International Journal of Applied Earth Observation and Geoinformation*, 10(3):358–373.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., and Cho, M. (2008b). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(4):409–426.

Datt, B., Jupp, D., and McVicar, T. (2003). Time series analysis of EO-1 Hyperion data for yield estimation at an agricultural site. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 564–566.

Daughtry, C., Gallo, K., Goward, S., Prince, S., and Kustas, W. (1992). Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies. *Remote Sensing of Environment*, 39(2):141–152.

Dawson, T. P., Curran, P. J., and Plummer, S. E. (1998). LIBERTY–modeling the effects of leaf biochemical concentration on reflectance spectra. *Remote Sensing of Environment*, 65(1):50–60.

de la Fuente, D., Suarez, J., Yague, J., and Pedrazzani, D. (2013). Potentiality of World-View 2 data for precision agriculture. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2825–2828.

Delalieux, S., Zarco-Tejada, P., Tits, L., Jimenez Bello, M., Intrigliolo, D., and Somers, B. (2014). Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2571–2582.

Delécolle, R., Maas, S., Guérif, M., and Baret, F. (1992). Remote sensing and crop production models: present trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 47(2):145–161.

Dell'Endice, F., Nieke, J., Koetz, B., Schaepman, M. E., and Itten, K. (2009). Improving radiometry of imaging spectrometers by using programmable spectral regions of interest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6):632–639.

Delta-T Devices Ltd. (2016). SS1 - SunScan. Available online: http://www.delta-t.co.uk/product-display.asp?id=SS1%20Product&div=Plant%20Science (accessed on 28 January 2016).

Dente, L., Satalino, G., Mattia, F., and Rinaldi, M. (2008). Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. *Remote Sensing of Environment*, 112(4):1395–1407.

DESA (2015). World population prospects: The 2015 revision, key findings and advance tables. Technical Report Working Paper No. ESA/P/WP.241, United Nations, Department of Economic and Social Affairs, Population Division.

DLR - German Aerospace Center (2015). Environmental Mapping and Analysis Program (EnMAP). Available online: http://www.enmap.org/ (accessed on 28 October 2015).

Doktor, D., Lausch, A., Spengler, D., and Thurner, M. (2014). Extraction of plant physiological status from hyperspectral signatures using machine learning methods. *Remote Sensing*, 6(12):12247–12274.

Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., and Schaepman, M. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2):165–193.

Duchemin, B., Hadria, R., Erraki, S., Boulet, G., Maisongrande, P., Chehbouni, A., Escadafal, R., Ezzahar, J., Hoedjes, J., Kharrou, M., Khabba, S., Mougenot, B., Olioso, A., Rodriguez, J.-C., and Simonneaux, V. (2006). Monitoring wheat phenology and irrigation in central morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices. *Agricultural Water Management*, 79(1):1–27.

Dunn, B., Beecher, H., Batten, G., and Ciavarella, S. (2002). The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the riverine plain of south- eastern australia. *Australian Journal of Experimental Agriculture*, 42(5):607–614.

Durbha, S. S., King, R. L., and Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, 107(1–2):348–361.

DWD - Deutscher Wetterdienst (2015). Climate Data Center (Zugang zu den Klimadaten des Deutschen Wetterdienstes). Available online: ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/monthly/kl/historical/ (accessed on 25 November 2015).

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331.

Ehlers, M., Klonus, S., Åstrand, P. J., and Rosso, P. (2010). Multi-sensor image fusion for pansharpening in remote sensing. *International Journal of Image and Data Fusion*, 1(1):25–45.

Elvidge, C. D. and Chen, Z. (1995). Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment*, 54(1): 38–48.

Enclona, E. A., Thenkabail, P. S., Celis, D., and Diekmann, J. (2004). Within-field wheat yield prediction from IKONOS data: a new matrix approach. *International Journal of Remote Sensing*, 25(2):377–388.

ESA - European Space Agency (2015). Sentinel online. Available online: https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed on 28 October 2015).

European Commission (2012). Sustainable agriculture for the future we want. Available online: http://ec.europa.eu/agriculture/events/2012/rio-side-event/brochure_en.pdf (accessed on 29 September 2015).

Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., and Cavigelli, M. (2008). Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *International Journal of Remote Sensing*, 29(10): 3011–3032.

Fang, H., Liang, S., and Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. *Remote Sensing of Environment*, 85(3):257–270.

Fang, H., Wei, S., Jiang, C., and Scipal, K. (2012). Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sensing of Environment*, 124:610–621.

FAO - Food and Agriculture Organization of the United Nations (2015). Change in world's agricultural area 1961-2011 in %. Available online: http://faostat3.fao.org/download/E/EL/E (accessed on 29 September 2015).

Filella, I. and Peñuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, 15(7):1459–1470.

Fonseca, L., Namikawa, L., Castejon, E., Carvalho, L., Pinho, C., and Pagamisse, A. (2011). Image fusion for remote sensing applications. In Zheng, Y., editor, *Image Fusion and Its Applications*, chapter 9, pages 153–178. InTech.

Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., and Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sensing of Environment*, 56:104–117.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–67.

Ganapol, B. D., Johnson, L. F., Hammer, P. D., Hlavka, C. A., and Peterson, D. L. (1998). LEAFMOD: A new within-leaf radiative transfer model. *Remote Sensing of Environment*, 63(2):182–193.

Garrigues, S., Shabanov, N., Swanson, K., Morisette, J., Baret, F., and Myneni, R. (2008). Intercomparison and sensitivity analysis of leaf area index retrievals from LAI-2000, AccuPAR, and digital hemispherical photography over croplands. *Agricultural and Forest Meteorology*, 148(8–9):1193–1209.

Gates, D. M., Keegan, H. J., Schleter, J. C., and Weidner, V. R. (1965). Spectral properties of plants. *Applied Optics*, 4(1):11–20.

Gausman, H. W. and Allen, W. A. (1973). Optical parameters of leaves of 30 plant species. *Plant Physiology*, 52(1):57–62.

Ge, Y., Thomasson, J. A., and Sui, R. (2011). Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, 5(3):229–238.

Gebbers, R. and Adamchuk, V. I. (2010). Precision agriculture and food security. *Science*, 327(5967):828–831.

Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185(0):1–17.

Gerighausen, H. (2013). *Abbildende Spektroskopie erosionsrelevanter Boden-parameter in Agrarökosystemen*. PhD thesis, Bonn University.

Gerighausen, H., Lilienthal, H., Jarmer, T., and Siegmann, B. (2016). Evaluation of leaf area index and dry matter predictions for crop growth modeling and yield estimation based on field reflectance measurements. *EARSeL eProceedings*, 14(S2):71–90.

Gevaert, C. M., Tang, J., Garcia-Haro, F. J., Suomalainen, J., and Kooistra, L. (2014). Combining hyperspectral UAV and mulitspectral FORMOSAT-2 imagery for precision agriculture applications. In *Proceedings of the 6th WHISPERS Workshop on Hyperspectral Image and Signal Processing*, Lausanne, Switzerland.

Ghosh, A., Datta, A., and Ghosh, S. (2013). Self-adaptive differential evolution for feature selection in hyperspectral image data. *Applied Soft Computing*, 13(4):1969–1977.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., and Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, 75(3):272–281.

Gleyzes, J. P., Meygret, A., Fratter, C., Panem, C., Baillarin, S., and Valorge, C. (2003). SPOT5: system overview and image ground segment. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pages 300–302, Toulouse, France.

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): 1–212.

Goel, N. S. and Grier, T. (1988). Estimation of canopy parameters for inhomogeneous vegetation canopies from reflectance data: III. Trim: A model for radiative transfer in heterogeneous three-dimensional canopies. *Remote Sensing of Environment*, 25(3):255–293.

Gonzalez, R. C. and Woods, R. E. (2007). *Digital Image Processing*. Prentice Hall, Upper saddle River, third edition.

Govaerts, Y. and Verstraete, 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(2):493–505.

Govaerts, Y. M., Verstraete, M. M., Pinty, B., and Gobron, N. (1999). Designing optimal spectral indices: A feasibility and proof of concept study. *International Journal of Remote Sensing*, 20(9):1853–1873.

Gower, S. T., Kucharik, C. J., and 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(1):29–51.

Grant, L. (1987). Diffuse and specular characteristics of leaf reflectance. *Remote Sensing of Environment*, 22(2):309–322.

Gray, J. and Song, C. (2012). Mapping leaf area index using spatial, spectral, and temporal information from multiple sensors. *Remote Sensing of Environment*, 119:173–183.

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. J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., and Sang, B. (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, 7(7):8830–8857.

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

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422.

Guyot, G. (1990). *Applications of Remote Sensing in Agriculture*, chapter Optical properties of vegetation canopies, pages 19–43. Butterworths, London.

Guyot, G., Baret, F., and Jaquemoud, S. (1992). *Imaging Spectroscopy: Fundamentals and Prospective Applications*, chapter Imaging spectroscopy for vegetation studies, pages 145–165. Kluwer Academic Publisher, Norwell, Dordrecht.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90(3):337–352.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81(2–3):416–426.

Hansen, P. and Schjoerring, J. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4): 542–553.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer, New York.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1):1–12.

Heinert, M. (2010). Support vector machines – Teil 1: Ein theoretischer Überblick. [in German]. *ZFV - Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 135(3):179–189.

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., and Bonfil, D. (2011). LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands. *Remote Sensing of Environment*, 115(8):2141–2151.

Hilario, M. and Kalousis, A. (2008). Approaches to dimensionality reduction in proteomic biomarker studies. *Briefings in Bioinformatics*, 9(2):102–118.

Hoffer, R. M. (1987). *Remote Sensing: the quantitative approach*, chapter Biological and Physical Considerations in Applying Computer-Aided Analysis Techniques to Remote Sensor Data, pages 227–289. McGraw-Hill, New York.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. A Bradford Book.

Horler, D. N. H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2):273–288.

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3):295–309.

Hurcom, S. J., Harrison, A. R., and Taberner, M. (1996). Assessment of biophysical vegetation properties through spectral decomposition techniques. *Remote Sensing of Environment*, 56(3):203–214.

Jacquemoud, S. and Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34(2):75–91.

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F., and Jaggard, K. (1995). Extraction of vegetation biophysical parameters by inversion of the PROSPECT + SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote Sensing of Environment*, 52(3):163–172.

Jacquemoud, S., Ustin, S., Verdebout, J., Schmuck, G., Andreoli, G., and Hosgood, B. (1996). Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, 56(3):194–202.

Jarmer, T. (2013). Spectroscopy and hyperspectral imagery for monitoring summer barley. *International Journal of Remote Sensing*, 34(17):6067–6078.

Jensen, J. R. (2000). *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice Hall, Upper Saddle River, second edition.

Jensen, J. R. (2004). *Introductory Digital Image Processing*. Prentice Hall, third edition.

Jensen, R. R., Hardin, P. J., and Hardin, A. J. (2012). Estimating urban leaf area index (LAI) of individual trees with hyperspectral data. *Photogrammetric Engineering & Remote Sensing*, 78(5):495–504.

Jiang, D., Zhuang, D., Huang, Y., and Fu, J. (2011). Survey of multispectral image fusion techniques in remote sensing applications. In Zheng, Y., editor, *Image Fusion and Its Applications*, chapter 1, pages 1–22. InTech.

Johnson, B. (2014). Effects of pansharpening on vegetation indices. *ISPRS International Journal of Geo-Information*, 3(2):507–522.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., and Baret, F. (2004). Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121(1–2):19–35.

Jones, H. G. and Vaughan, R. A. (2010). *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford University Press.

Kalacska, M., Sanchez-Azofeifa, G., Caelli, T., Rivard, B., and Boerlage, B. (2005). Estimating leaf area index from satellite imagery using Bayesian networks. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8):1866–1873.

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab - An S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9):1–20.

Karimi, Y., Prasher, S., Madani, A., and Kim, S. (2008). Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. *Canadian Biosystems Engineering*, 50:713–720.

Kira, K. and Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the Ninth International Workshop on Machine Learning*, ML92, pages 249–256, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Klonus, S. (2011). *Optimierung und Auswirkungen von ikonischen Bildfusions-verfahren zur Verbesserung von fernerkundlichen Auswerteverfahren*. PhD thesis, University of Osnabrück.

Klonus, S. and Ehlers, M. (2007). Image fusion using the Ehlers spectral characteristics preservation algorithm. *GIScience & Remote Sensing*, 44(2): 93–116.

Kneubühler, M. (2002). *Spectral Assessment of Crop Phenology Based on Spring Wheat and Winter Barley*. PhD thesis, University of Zurich.

Kobayashi, H., Ryu, Y., Baldocchi, D. D., Welles, J. M., and Norman, J. M. (2013). On the correct estimation of gap fraction: How to remove scattered radiation in gap fraction measurements? *Agricultural and Forest Meteorology*, 174–175:170–183.

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In Bergadano, F. and De Raedt, L., editors, *Machine Learning: ECML-94*, volume 784 of *Lecture Notes in Computer Science*, pages 171–182. Springer Berlin Heidelberg.

Koppe, W., Li, F., Gnyp, M. L., Miao, Y., Jia, L., Chen, X., Zhang, F., and Bareth, G. (2010). Evaluating multispectral and hyperspectral satellite remote sensing data for estimating winter wheat growth parameters at regional scale in the north China plain. *PFG Photogrammetrie, Fernerkundung, Geoinformation*, 2010(3):167–178.

Kross, A., McNairn, H., Lapen, D., Sunohara, M., and Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation*, 34:235–248.

Kruse, F., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. (1993). The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2-3):145–163.

Kuester, T., Spengler, D., Barczi, J.-F., Segl, K., Hostert, P., and 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(4):2096–2108.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer, New York.

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

Lee, K.-S., Cohen, W. B., Kennedy, R. E., Maiersperger, T. K., and Gower, S. T. (2004). Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment*, 91(3–4):508–520.

Lewis, P. (1999). Three-dimensional plant modelling for remote sensing simulation studies using the Botanical Plant Modelling System. *Agronomie*, 10(3–4): 185–210.

LGBA LSA - Landesamt für Geologie und Bergwesen Sachsen-Anhalt (2015). Bodenübersicht Sachsen-Anhalt (BÜK200) (WMS-Darstellungsdienst). Available online: http://www.geofachdatenserver.de/de/geodienste.html (accessed on 26 November 2015).

Li, B., Morris, J., and Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 64(1):79–89.

Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., Song, X., and Yang, G. (2014). Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sensing*, 6(7):6221–6241.

LI-COR Inc. (2016a). LI-COR LAI-2000 Plant Canopy Analyzer. Available online: https://www.licor.com/env/products/leaf_area/LAI-2000/ (accessed on 28 January 2016).

LI-COR Inc. (2016b). LI-COR LAI-2200 Plant Canopy Analyzer - Instruction Manual. Available online: https://app.boxenterprise.net/s/fqjn5mlu8c1a7zir5qel/ (accessed on 26 January 2016).

LI-COR Inc. (2016c). LI-COR LAI-3100 leaf area meter. Available online: https://www.licor.com/env/products/leaf_area/LI-3100C/ (accessed on 21 January 2016).

Liaw, A. and Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3):18–22.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2004). *Remote Sensing and Image Interpretation*. Wiley, fifth edition.

LVermGeo LSA - Landesamt für Vermessung und Geoinformation Sachsen-Anhalt (2015). Topographische Karte 1:25 000 (TK 25). © GeoBasis-DE (52c).

Machwitz, M., Giustarini, L., Bossung, C., Frantz, D., Schlerf, M., Lilienthal, H., Wandera, L., Matgen, P., Hoffmann, L., and Udelhoven, T. (2014). Enhanced biomass prediction by assimilating satellite data into a crop growth model. *Environmental Modelling & Software*, 62:437–453.

Maier, S., Lüdeker, W., and Günther, K. (1999). SLOP: A revised version of the stochastic model for leaf optical properties. *Remote Sensing of Environment*, 68(3):273–280.

Malin, D. (2007). *The Focal Encyclopedia of Photography*, chapter Scientific Photography: Expanded Vision, pages 497–626. Focal Press, 4 edition.

Malley, D. F., Martin, P. D., and Ben-Dor, E. (2004). *Near-Infrared Spectroscopy in Agriculture*, volume 44 of *Agronomy Monograph*, chapter Application in Analysis of Soils, pages 729–784. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Marshall, M. and Thenkabail, P. (2015). Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:205–218.

Mayumi, N. and Iwasaki, A. (2011). Image sharpening using hyperspectral and multispectral data. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 519–522.

Meier, U. (2001). *Growth stages of mono-and dicotyledonous plants. BBCH Monograph*. Federal Biological Research Centre for Agriculture and Forestry, 2 edition.

Mevik, B.-H. and Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18(2):1–24.

Miglani, A., Ray, S. S., Pandey, R., and Parihar, J. S. (2009). Evaluation of EO-1 Hyperion data for agricultural applications. *Journal of the Indian Society of Remote Sensing*, 36(3):255–266.

MLU - Ministerium für Landwirtschaft und Umwelt Sachsen-Anhalt (2009). InVeKoS -Digitale Orthophotos (DOP) 2009.

Monsi, M. and Saeki, T. (2005). On the factor light in plant communities and its importance for matter production. *Annals of Botany*, 95(3):549–567.

Monteith, J. and Unsworth, M. (2007). *Principles of Environmental Physics, Third Edition*. Academic Press, San Diego, CA, USA.

Moran, M., Inoue, Y., and Barnes, E. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61(3):319–346.

Moran, M. S., Maas, S. J., and Jr., P. J. P. (1995). Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sensing Reviews*, 12(3–4):335–353.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G., and Cook, R. (2006). Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7): 1804–1817.

Moulin, S., Bondeau, A., and Delecolle, R. (1998). Combining agricultural crop models and satellite observations: From field to regional scales. *International Journal of Remote Sensing*, 19(6):1021–1036.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4): 358–371.

Mustafa, Y. T., Stein, A., Tolpekin, V. A., and Van Laake, P. E. (2012). Improving forest growth estimates using a Bayesian Network approach. *Photogrammetric Engineering & Remote Sensing*, 78(1):45–51.

Mutanga, O., Adam, E., and Cho, M. A. (2012). High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18(0):399–406.

NASA Jet Propulsion Laboratory (2013). NASA Shuttle Radar Topography Mission Global 1 arc second. Available online: https://lpdaac.usgs.gov/dataset_discovery/measures/measures_products_table/srtmgl1_v003/ (accessed on 02 March 2016).

Nguy-Robertson, A. L., Peng, Y., Gitelson, A. A., Arkebauer, T. J., Pimstein, A., Herrmann, I., Karnieli, A., Rundquist, D. C., and Bonfil, D. J. (2014). Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. *Agricultural and Forest Meteorology*, 192–193:140–148.

Nguyen, H. T. and Lee, B.-W. (2006). Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. *European Journal of Agronomy*, 24(4):349–356.

Oredein, A. I., Olatayo, T. O., and Loyinmi, A. C. (2011). On validating regression models with bootstraps and data splitting techniques. *Global Journal of Science Frontier Research*, 11(6):1–6.

Otto, M. (2007). *Chemometrics*. Wiley-VCH, second edition.

Padwick, C., Deskevich, M., and Pacifici, F.and Smallwood, S. (2010). Worldview-2 pan-sharpening. In *Proceedings of the 2010 ASPRS Annual Conference*, San Diego, USA.

Peñuelas, J., Pinol, J., Ogaya, R., and Filella, I. (1997). Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *International Journal of Remote Sensing*, 18(13):2869–2875.

Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S., and Upchurch, D. R. (2003). Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, 69(6):647–664.

Pohl, C. and van Genderen, J. (2015). Structuring contemporary remote sensing image fusion. *International Journal of Image and Data Fusion*, 6(1):3–21.

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., and Ohmann, J. L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114(5):1053–1068.

Pu, R. (2012). Comparing canonical correlation analysis with partial least squares regression in estimating forest leaf area index with multitemporal Landsat TM imagery. *GIScience & Remote Sensing*, 49(1):92–116.

Qu, Y. and Sun, G. (2010). Research on wireless sensor node for measurement of vegetation structure parameters. In *Proceedings of the World Automation Congress (WAC)*, pages 437–442.

Renzullo, L., Blanchfield, A., and Powell, K. (2006). A method of wavelength selection and spectral discrimination of hyperspectral reflectance spectrometry. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1986–1994.

Rich, P. M. (1990). Characterizing plant canopies with hemispherical photographs. *Remote Sensing Reviews*, 5(1):13–29.

Richardson, A., Wiegand, C., Arkin, G., Nixon, P., and Gerbermann, A. (1982). Remotely-sensed spectral indicators of sorghum development and their use in growth modeling. *Agricultural Meteorology*, 26(1):11–23.

Richter, K., Atzberger, C., Vuolo, F., and D'Urso, G. (2011). Evaluation of Sentinel-2 spectral sampling for radiative transfer model based LAI estimation of wheat, sugar beet, and maize. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2):458–464.

Richter, T. and Fukshansky, L. (1996). Optics of a bifacial leaf: 1. A novel combined procedure for deriving the optical parameters. *Photochemistry and Photobiology*, 63(4):507–516.

Rivera, J. P., Verrelst, J., Leonenko, G., and Moreno, J. (2013). Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. *Remote Sensing*, 5(7):3280–3304.

Robert, P. (2002). Precision agriculture: A challenge for crop nutrition management. *Plant and Soil*, 247(1):143–149.

Robnik-Šikonja, M. and Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53(1-2):23–69.

Robnik-Sikonja, M. and Kononenko, I. (1997). An adaptation of Relief for attribute estimation in regression. In *Proceedings of the Fourteenth International Conference on Machine Learning*, ICML '97, pages 296–304, San Francisco, USA. Morgan Kaufmann Publishers Inc.

Rogaß, C., Spengler, D., Bochow, M., Segl, K., Lausch, A., Doktor, D., Roessner, S., Behling, R., Wetzel, H.-U., and Kaufmann, H. (2011). Reduction of radiometric miscalibration-applications to pushbroom sensors. *Sensors*, 11(6):6370–6395.

Rollin, E. and Milton, E. (1998). Processing of high spectral resolution reflectance data for the retrieval of canopy water content information. *Remote Sensing of Environment*, 65(1):86–92.

Romanski, P. and Kotthoff, L. (2014). *FSelector: Selecting attributes*. R package version 0.2, Available online: https://cran.r-project.org/web/packages/FSelector/index.html (accessed on 02 November 2015).

Roujean, J.-L. and Lacaze, R. (2002). Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: A pragmatic method and its validation. *Journal of Geophysical Research: Atmospheres*, 107(D12):ACL 6–1–ACL 6–14.

Rouse, J. W., Jr, Deering, D. W., Schell, J. A., and Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Technical report, NASA.

Roy, D., Wulder, M., Loveland, T., C.E., W., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R., Scambos, T., Schaaf, C., Schott, J., Sheng, Y., Vermote, E., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Hipple, J., Hostert, P., Huntington, J., Justice, C., Kilic, A., Kovalskyy, V., Lee, Z., Lymburner, L., Masek, J., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R., and Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145:154–172.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517.

Sandmeier, S. and Itten, K. (1999). A field goniometer system (FIGOS) for acquisition of hyperspectral BRDF data. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):978–986.

Schaepman-Strub, G., Schaepman, M., Painter, T., Dangel, S., and Martonchik, J. (2006). Reflectance quantities in optical remote sensing–definitions and case studies. *Remote Sensing of Environment*, 103(1):27–42.

Schlerf, M. and Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, 100(3):281–294.

Schröder, H. (2000). *Abriß der physischen Geographie und Aspekte des Natur- und Umweltschutzes Sachsen-Anhalts*, volume 247 of *Forschungen zur deutschen Landeskunde*. Deutsche Akademie für Landeskunde.

Schueller, J. (1992). A review and integrating analysis of spatially-variable control of crop production. *Fertilizer research*, 33(1):1–34.

Seelan, S. K., Laguette, S., Casady, G. M., and Seielstad, G. A. (2003). Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing of Environment*, 88(1–2):157–169.

Seelig, H.-D., Hoehn, A., Stodieck, L., Klaus, D., III, W. A., and Emery, W. (2008). Relations of remote sensing leaf water indices to leaf water thickness in cowpea, bean, and sugarbeet plants. *Remote Sensing of Environment*, 112(2):445–455.

Segl, K., Guanter, L., Kaufmann, H., Schubert, J., Kaiser, S., Sang, B., and Hofer, S. (2010). Simulation of spatial sensor characteristics in the context of the EnMAP hyperspectral mission. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):3046–3054.

Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., Sang, B., Mogulsky, V., and Hofer, S. (2012). EeteS–The EnMAP End-to-End Simulation Tool. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):522–530.

Serrano, L., Filella, I., and Peñuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Science*, 40(3):723–731.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York.

Shi, Y., Wang, J., Qin, J., and Qu, Y. (2015). An upscaling algorithm to obtain the representative ground truth of LAI time series in heterogeneous land surface. *Remote Sensing*, 7(10):12887–12908.

Siegmann, B. and Jarmer, T. (2015). Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *International Journal of Remote Sensing*, 36(18):4519–4534.

Siegmann, B., Jarmer, T., Beyer, F., and Ehlers, M. (2015). The potential of pan-sharpened EnMAP data for the assessment of wheat LAI. *Remote Sensing*, 7(10):12737–12762.

Siegmann, B., Jarmer, T., Selige, T., Lilienthal, H., Richter, N., and Höfle, B. (2013). Comparison of narrow band vegetation indices and empirical models from hyperspectral remote sensing data for the assessment of wheat nitrogen concentration. In *Proceedings of the EARSeL 8th SIG-Imaging Spectroscopy Workshop 2013*.

Sims, D. A. and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2–3):337–354.

Sinclair, T. R., Hoffer, R. M., and Schreiber, M. M. (1971). Reflectance and internal structure of leaves from several crops during a growing season. *Agronomy Journal*, 63(6):864–868.

SLS-A - Statistisches Landesamt Sachsen-Anhalt (2015). Anbauflächen ausgewählter Fruchtarten in Sachsen-Anhalt nach Jahren. Available online: http://www.statistik.sachsen-anhalt.de/Internet/Home/Daten_und_Fakten/4/41/412/41241/Anbauflaechen_ausgewaehlter_Fruchtarten_nach_Jahren.html (accessed on 27 November 2015).

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

Snee, R. D. (1977). Validation of regression models: Methods and examples. *Technometrics*, 19(4):415–428.

Sohaib, M., Haq, I., and Mushtaq, Q. (2012). Dimensional reduction of hyperspectral image data using band clustering and selection through k-means based on statistical characteristics of band images. *International Journal of Computer Science*, 2(4):146–151.

Specim Ltd. (2015). aisaDUAL Hyperspectral Sensor. Available online: http://www.channelsystems.ca/documents/AISADualver1-07.pdf (accessed on 28 October 2015).

Spengler, D. (2013). *Charakterisierung von Getreidearten aus hyperspektralen Fernerkundungsdaten auf der Basis von 4D-Bestandsmodellen*. PhD thesis, Technical University Berlin.

Stagakis, S., Markos, N., Sykioti, O., and Kyparissis, A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations. *Remote Sensing of Environment*, 114(5):977–994.

Stat-Bund - Statistisches Bundesamt. Land- & Forstwirtschaft, Fischerei (2015). Available online: https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/LandForstwirtschaftFischerei/LandForstwirtschaft.html (accessed on 27 November 2015).

Statista GmbH (2015). Anbaufläche von Getreide in Deutschland nach Art in den Jahren 1960 bis 2014 (in 1.000 Hektar). Available online: http://de.statista.com/statistik/daten/studie/28887/umfrage/anbauflaeche-von-getreide-in-deutschland-seit-1960/ (accessed on 27 November 2015).

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., and van Wesemael, B. (2010). Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, 158(1–2):32–45.

Steyerberg, E. W., Jr, F. E. H., Borsboom, G. J., Eijkemans, M., Vergouwe, Y., and Habbema, J. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54(8):774–781.

Stoorvogel, J. J., Kooistra, L. K., and Bouma, J. (2015). Managing soil variability at different spatial scale as a basis for precision agriculture. In Lal, R. and Stewart, B. A., editors, *Soil-Specific Farming: Precision Agriculture*, number 22 in Advances in Soil Science, chapter 2, pages 37–72. CRC Press.

Stovall, J. (2012): Stand Dynamics: Stand Function. Available online: http://forestry.sfasu.edu/faculty/stovall/silviculture/images/textbook/lai.png (accessed on 12 May 2015).

Stuckens, J., Verstraeten, W. W., Delalieux, S., Swennen, R., and Coppin, P. (2009). A dorsiventral leaf radiative transfer model: Development, validation and improved model inversion techniques. *Remote Sensing of Environment*, 113(12):2560–2573.

Tanaka, S., Kawamura, K., Maki, M., Muramoto, Y., Yoshida, K., and Akiyama, T. (2015). Spectral index for quantifying leaf area index of winter wheat by field hyperspectral measurements: A case study in Gifu prefecture, central Japan. *Remote Sensing*, 7(5):5329–5346.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., and Meer, B. V. D. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91(3–4):354–376.

Thenkabail, P. S., Smith, R. B., and Pauw, E. D. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71(2):158–182.

Thorp, K. and Tian, L. (2004). A review on remote sensing of weeds in agriculture. *Precision Agriculture*, 5(5):477–508.

Tian, Y., Wang, Y., Zhang, Y., Knyazikhin, Y., Bogaert, J., and Myneni, R. B. (2003). Radiative transfer based scaling of LAI retrievals from reflectance data of different resolutions. *Remote Sensing of Environment*, 84(1):143–159.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, 418: 671–677.

Tuia, D., Verrelst, J., Alonso-Chorda, L., Pérez-Cruz, F., and Camps-Valls, G. (2011). Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geoscience and Remote Sensing Letters*, 8(4):804–808.

U. S. Census Bureau (2015). U.S. and world population clock. Available online: http://www.census.gov/popclock/ (accessed on 28 September 2015).

Uno, Y., Prasher, S., Lacroix, R., Goel, P., Karimi, Y., Viau, A., and Patel, R. (2005). Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Computers and Electronics in Agriculture*, 47(2):149–161.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York, USA.

Vapnik, V. N. and Chervonenkis, A. Y. (1974). *Theory of Pattern Recognition*. Nauka, Moscow, Russia.

Verdebout, J., Jacquemoud, S., and Schmuck, G. (1994). *Imaging Spectrometry – a Tool for Environmental Observations*, volume 4 of *Remote Sensing*, chapter Optical Properties of Leaves: Modelling and Experimental Studies, pages 169–191. Kluwer Academic Publishers, Dordrecht.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16(2):125–141.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., and Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108:273–290.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., and 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. P., Moreno, J., and Camps-Valls, G. (2013). Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86:157–167.

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., and Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 115(12):3468–3478.

Villafranca, A. G., Corbera, J., Martín, F., and Marchán, J. F. (2012). Limitations of hyperspectral earth observation on small satellites. *Journal of Small Satellites*, 1(1):19–29.

Viscarra-Rossel, R. A. (2008). ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems*, 90(1): 72–83.

Viscarra-Rossel, R. A. and Lark, R. M. (2009). Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *European Journal of Soil Science*, 60(3):453–464.

Vohland, M. and Jarmer, T. (2008). Estimating structural and biochemical parameters for grassland from spectroradiometer data by radiative transfer modelling (PROSPECT+SAIL). *International Journal of Remote Sensing*, 29(1):191–209.

Vuolo, F., Essl, L., and Atzberger, C. (2015). Costs and benefits of satellite-based tools for irrigation management. *Frontiers in Environmental Science*, 3(52):1–12.

Vuolo, F., Neugebauer, N., Bolognesi, S. F., Atzberger, C., and D'Urso, G. (2013). Estimation of leaf area index using DEIMOS-1 data: Application and transferability of a semi-empirical relationship between two agricultural areas. *Remote Sensing*, 5(3):1274–1291.

Wang, F.-M., Huang, J.-F., and Lou, Z.-H. (2011). A comparison of three methods for estimating leaf area index of paddy rice from optimal hyperspectral bands. *Precision Agriculture*, 12(3):439–447.

Watson, D. J. (1947). Comparative physiological studies on the growth of field crops I. Variation in net assimilation rate and leaf area between species and varieties, and within and between years. *Annals of Botany*, 11(1):41–76.

Weiss, M., Baret, F., Myneni, R. B., Pragnère, A., and Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20(1):3–22.

Welles, J. M. (1990). Some indirect methods of estimating canopy structure. *Remote Sensing Reviews*, 5(1):31–43.

Wessman, C. A. (1990). *Remote Sensing of Biosphere Functioning*, chapter Evaluation of Canopy Biochemestry, pages 135–156. Number 79 in Ecological Studies. Springer, New York.

Whelan, B. and Taylor, J. (2013). *Precision Agriculture for Grain Production Systems*. CSIRO Publishing.

Widlowski, J.-L., Pinty, B., Lopatka, M., Atzberger, C., Buzica, D., Chelle, M., Disney, M., Gastellu-Etchegorry, J.-P., Gerboles, M., Gobron, N., Grau, E., Huang, H., Kallel, A., Kobayashi, H., Lewis, P. E., Qin, W., Schlerf, M., Stuckens, J., and Xie, D. (2013). The fourth radiation transfer model intercomparison (RAMI-IV): Proficiency testing of canopy reflectance models with ISO-13528. *Journal of Geophysical Research: Atmospheres*, 118(13):6869–6890.

Wilhelm, W., Ruwe, K., and Schlemmer, M. R. (2000). Comparison of three leaf area index meters in a corn canopy. *Crop Science*, 40(4):1179–1183.

Williams, P. C. and Norris, K. H. (2001). *Near-Infrared Technology in the Agricultural and Food Industries*. American Association of Cereal Chemistry, second edition.

Willstätter, R. and Stoll, A. (1918). *Untersuchungen über die Assimilation der Kohlensäure*. Springer, Berlin.

Wold, H. (1985). *Encyclopedia of the Statistical Sciences*, volume 6, chapter Partial Least Squares, page 581–591. John Wiley & Sons, Inc., New York.

Wold, S., Geladi, P., Esbensen, K., and Öhman, J. (1987). Multi-way principal components-and PLS-analysis. *Journal of Chemometrics*, 1(1):41–56.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

Wu, B., Chen, C., Kechadi, T. M., and Sun, L. (2013). A comparative evaluation of filter-based feature selection methods for hyper-spectral band selection. *International Journal of Remote Sensing*, 34(22):7974–7990.

Wu, H. and Li, Z.-L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors*, 9(3):1768–1793.

Wu, J., Wang, D., and Bauer, M. E. (2007). Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field Crops Research*, 102(1):33–42.

Yang, C., Everitt, J. H., and Bradford, J. M. (2006). Evaluating high-resolution QuickBird satellite imagery for estimating cotton yield. *Transactions of the American Society of Agricultural and Biological Engineers*, 49(5):1599–1606.

Yao, H., Tang, L., Tian, L., Brown, R. L., Bhatnagar, D., and Cleveland, T. E. (2012). Using hyperspectral data in precision farming applications. In Thenkabail, P. S., Lyon, J. G., and Huete, A., editors, *Hyperspectral Remote Sensing of Vegetation*, chapter 25, pages 591–607. CRC Press.

Yoder, B. J. and Pettigrew-Crosby, R. E. (1995). Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. *Remote Sensing of Environment*, 53(3):199–211.

Yu, K., Gnyp, M. L., Gao, L., Yuxin, M., Xinping, C., and Bareth, G. (2015). Estimate leaf chlorophyll of rice using reflectance indices and partial least squares. *PFG Photogrammetrie, Fernerkundung, Geoinformation*, 2015(1):45–54.

Zarco-Tejada, P., González-Dugo, V., and Berni, J. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sensing of Environment*, 117:322–337.

Zarco-Tejada, P., Miller, J., Noland, T., Mohammed, G., and Sampson, P. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1491–1507.

Zarco-Tejada, P. J., Ustin, S. L., and Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agronomy Journal*, 97(3):641–653.

Zhang, N., Wang, M., and Wang, N. (2002). Precision agriculture–a worldwide overview. *Computers and Electronics in Agriculture*, 36(2–3):113–132.

Zhang, Y. (2012). Wavelet-based Bayesian fusion of multispectral and hyperspectral images using Gaussian scale mixture model. *International Journal of Image and Data Fusion*, 3(1):23–37.

Zhang, Y., Backer, S. D., and Scheunders, P. (2009). Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3834–3843.

Zheng, G. and Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. *Sensors*, 9(4):2719–2745.

Zurita-Milla, R., Clevers, J., and Schaepman, M. (2008). Unmixing-based Landsat TM and MERIS FR data fusion. *IEEE Geoscience and Remote Sensing Letters*, 5(3):453–457.

Publications

Peer Reviewed Journal Paper

Siegmann, B. and Jarmer, T. (2015): Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *International Journal of Remote Sensing*, 36(18):4519–4534.

Siegmann, B., Jarmer, T., Beyer, F. and Ehlers, M. (2015). The potential of pan-sharpened EnMAP data for the assessment of wheat LAI. *Remote Sensing*, 7(10):12737–12762.

Bauer, J., **Siegmann, B.**, Jarmer, T., and Aschenbruck, N. (2016). On the potential of wireless sensor networks for the in-situ assessment of crop leaf area index. *Computers and Electronics in Agriculture*, 128:149–159.

Koenig, K., Höfle, B., Hämmerle, M., Jarmer, T., **Siegmann, B.** and Lilienthal, H. (2015). Comparative classification analysis of post-harvest growth detection from terrestrial LiDAR point clouds in precision agriculture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104: 112–125.

Beyer, F., Jarmer, T. and **Siegmann, B.** (2015). Identification of agricultural crop types in Northern Israel using multitemporal RapidEye data. *Journal for Photogrammetry, Remote Sensing and Geoinformation Science*, 2015(1): 21-32.

Conferences Paper

Siegmann, B., Jarmer, T., Selige, T., Lilienthal, H., Richter, N. and Höfle, B. (2013). Comparison of narrow band vegetation indices and empirical models from hyperspectral remote sensing data for the assessment of wheat nitrogen concentration. In *Proceedings of the EARSeL 8th SIG-Imaging Spectroscopy Workshop 2013*.

Siegmann, B., Jarmer, T., Selige, T., Lilienthal, H., Richter, N. and Höfle, B. (2012). Using hyperspectral remote sensing data for the assessment of topsoil organic carbon from agricultural soils. In *SPIE Proceedings - Remote Sensing for Agriculture, Ecosystems, and Hydrology XIV 2012*.

Gerighausen, H., Lilienthal, H., Jarmer, T., and **Siegmann, B.** (2016). Evaluation of leaf area index and dry matter predictions for crop growth modeling and yield estimation based on field reflectance measurements. In *EARSeL eProceedings*, 14(S2), pages 71–90.

Bauer, J., **Siegmann, B.**, Jarmer, T., and Aschenbruck, N. (2016). Smart fLAIr: A smartphone application for fast LAI retrieval using Ambient Light Sensors. In *2016 IEEE Sensors Applications Symposium (SAS)*, pages 1–6.

Bauer, J., **Siegmann, B.**, Jarmer, T., and Aschenbruck, N. (2014). On the potential of wireless sensor networks for the in-field assessment of bio-physical crop parameters. In *2014 IEEE 39th Conference on Local Computer Networks Workshops (LCN Workshops)*, pages 523–530.

Jarmer, T., **Siegmann, B.**, Lilienthal, H., Höfle, B., Selige, T. and Richter, N. (2013). LAI assessing of wheat stands from AISA-DUAL imagery. In *Proceedings of the EARSeL 8th SIG-Imaging Spectroscopy Workshop 2013*.

Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich / unentgeltlich geholfen.

1. .....

   .....

2. .....

   .....

3. .....

   .....

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

....

(Ort, Datum) (Unterschrift)