

Development of An Earth Observation Processing Chain for Crop Biophysical Parameters at Local and Global Scale

Autor: Manuel Campos Taberner

Directors: Francisco Javier García Haro i Gustau Camps Valls



VNIVERSITAT
DE VALÈNCIA

Doctorat en Teledetecció

Maig de 2017

PHD THESIS IN REMOTE SENSING

BY

MANUEL CAMPOS TABERNER

DEPARTAMENT DE FÍSICA DE LA TERRA I TERMODINÀMICA. FACULTAT DE FÍSICA



VNIVERSITAT
ID VALÈNCIA

DEVELOPMENT OF AN EARTH OBSERVATION PROCESSING CHAIN FOR CROP BIOPHYSICAL PARAMETERS AT LOCAL AND GLOBAL SCALE

Advisors:

FRANCISCO JAVIER GARCÍA HARO

GUSTAU CAMPS VALLS

Copyright © 2017 MANUEL CAMPOS TABERNER
Book cover right by MARIA CARDA SANZ
First printing, May, 2, 2017



Teledetección
doctorado



[Física] Facultat de Física

DR. FRANCISCO JAVIER GARCÍA HARO, Doctor en Física, Professor Titular de Universitat del Departament de Física de la Terra i Termodinàmica de la Facultat de Física de la Universitat de València i

DR. GUSTAU CAMPS VALLS, Doctor en Física, Professor Titular de Universitat del Departament d'Enginyeria Electrònica de l'Escola Tècnica Superior D'Enginyeria de la Universitat de València

FAN CONSTAR QUÈ:

el Llicenciat en Física MANUEL CAMPOS TABERNER ha realitzat sota la seva direcció el treball titulat *Development of an earth observation processing chain for crop biophysical parameters at local and global scale*, que es presenta en aquesta memòria per optar al grau de Doctor per la Universitat de València.

I per tal que així conste a efectes oportuns, i donant el vistiplau per a la presentació d'aquest treball davant el Tribunal de tesi que corresponga, signem el present certificat a València el 2 de Maig 2017.

Francisco Javier García Haro

Gustau Camps Valls

TESI DOCTORAL:

Development of an earth observation processing chain for crop biophysical parameters at local and global scale

AUTOR:

Manuel Campos Taberner

DIRECTORS:

Dr. Francisco Javier García Haro

Dr. Gustau Camps Valls

El tribunal nombrat per jutjar la Tesis Doctoral citada anteriorment, compost per:

President: _____

Vocal: _____

Secretari: _____

Acorda otorgar-li la qualificació de _____

I per a què així conste a efectes oportuns, signem el present certificat.

A Burjassot el de de 2017

NOTE TO THE READER

According to the University of Valencia Doctorate Regulation¹ this Thesis Dissertation is presented as a compendium of at least three publications in international journals containing the results of the conducted work. The Thesis achievements including the development of the processing chain have been reported in four international peer-reviewed publications. Nonetheless, this Thesis also provides a global overview of the developed processing chain in order to elaborate in detail its main components and obtained results. In addition, following the aforementioned regulation in the framework of Valencian teaching and with the aim to foster the University of Valencia own language in research and educational activity, the Dissertation includes an extended abstract in Valencian.

¹Reglament sobre depòsit, avaluació i defensa de la tesi doctoral aprovat pel Consell de Govern de 28 de Juny de 2016. ACGUV 172/2016.

Pla d'increment de la docència en valencià (ACGUV 129/2012) aprovat i modificat pel Consell de Govern de 22 de desembre de 2016. ACGUV 308/2016.

AGRAÏMENTS

Quan s'arriba a una fita com és la culminació d'aquest treball, hom sempre recorda aquelles persones que han format part d'una manera o altra del dia a dia durant la realització de la tesi. En primer lloc, no puc més que donar les gràcies als meus directors de tesi, Javi i Gus, dels quals només tinc paraules d'agraïment per tot el suport rebut i per haver depositat en mi la seva confiança des del primer moment. Amb ells he crescut científica i professionalment, i una bona part del que sóc, sens dubte, és gràcies a ells. Per a mi ha estat un privilegi treballar amb vosaltres i espere seguir fent-ho durant molt de temps.

De la mateixa manera, a tots els membres de l'*Environmental Remote Sensing Group of the University of Valencia* (UV-ERS) per tractar-me com u més des del primer dia i per la seva ajuda en tot el que els he demanat. A Amparo i Bea, pels vostres consells i experiència de la qual he après molt durant aquests anys. A Álvaro, amb qui he après i “frikejat” de software, hardware i de manteniment de servidors. A Sergio, per eixes xarrades sobre el món de les publicacions, farem front a l'*establishment*! A Gonçal, amb qui he compartit hores i hores de mesures enfangats dins els arrossars sota un sol de justícia en ple estiu sobrevivint als atacs de les camallongues.

Donar les gràcies també a la resta de persones que formen el Departament de Física de la Terra i Termodinàmica, a la gent d'EOLAB i sobretot a l'*Institute for Electromagnetic Sensing of the Environment* del *National Research Council of Italy*. En especial a Mirco Boschetti i Lorenzo Busetto, amb els que he tingut el plaer de treballar més intensament en el projecte ERMES, i per la seva hospitalitat durant la setmana que vaig passar a Itàlia.

A la meua família i amics, pel seu suport i continua motivació. Ben segur que, com es diu al meu poble, “estaran pagats de mi”. En especial a Esther, per la seva confiança, ànims i per tolerar els meus “vaig a enviar un mail”, “mire una coqueta i acabe” i “vaig a escriure un poquet”, que esdevenen hores davant l'ordinador.

Per últim però no menys important, donar les gràcies als projectes ERMES i Land-SAF, sense el finançament dels quals, la realització d'aquesta tesi no haguera estat possible.

Contents

Preface	xxi
Abstract	xxiii
1 INTRODUCTION	1
1.1 Introduction to remote sensing for Earth observation	2
1.1.1 Remote sensing at global scale	3
1.1.2 Remote Sensing at local scale	3
1.2 Biophysical parameters	4
1.2.1 Leaf Area Index (LAI)	4
1.2.2 Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)	5
1.2.3 Fraction of green Vegetation Cover (FVC)	5
1.3 Biophysical parameter retrieval methods	5
1.3.1 Statistical methods	5
1.3.2 Physical methods	6
1.3.3 Hybrid methods	6
1.4 Calibration and validation of remote sensing products	7
1.4.1 <i>In situ</i> measurements	7
1.4.2 Intercomparison of products	7
1.5 Summary	9
2 Inverting radiative transfer models	11
2.1 Introduction to forward and inverse modelling	12
2.2 PROSAIL	13
2.2.1 PROSPECT	13
2.2.2 SAIL	14
2.2.3 Biochemical and structural parameters	17
2.3 Statistical inverse modelling	18
2.3.1 Introduction to statistical inference	18
2.3.2 Machine learning methods	19
2.3.3 Implementation of the algorithms	24

2.4	Developed processing chain at local scale	24
2.4.1	Remote sensing data and sites	24
2.4.2	Theoretical performances	26
2.4.3	High-resolution LAI maps	28
2.4.4	Influence of the background spectra and utility of GPR uncertainty	30
2.5	Concluding remarks	34
3	<i>In situ</i> measurements	35
3.1	Introduction to field measurements	36
3.2	Classical instruments	37
3.2.1	LICOR-LAI2000	37
3.2.2	Digital hemispherical photography (DHP)	37
3.3	New technologies in geosciences: PocketLAI	39
3.4	Validation of the processing chain at local scale	41
3.4.1	Field campaigns	41
3.4.2	Accuracy assessment over ground LAI	45
3.4.3	Temporal evolution of LAI estimates	47
3.5	Concluding remarks	49
4	Biophysical parameters at global scale	51
4.1	Introduction	52
4.2	Remote sensing existing products at global scale	52
4.2.1	MODIS (MOD15A2)	52
4.2.2	Copernicus (GEOV1)	53
4.2.3	EUMETSAT Land-SAF (SEVIRI/MSG)	54
4.3	Developed processing chain at global scale	55
4.3.1	Inputs: EUMETSAT Polar System BRDF	55
4.3.2	Retrieval chain	56
4.3.3	Global LAI, FVC and FAPAR maps	58
4.4	Global validation	58
4.4.1	Comparison with global products	58
4.4.2	Comparison with local estimates	61
4.5	Concluding remarks	63
5	Conclusion and Discussion	65
6	Summary in Valencian	73
6.1	Estructura	73
6.2	Motivació i objectius	73

6.3	Metodologia	74
6.3.1	Paràmetres biofísics: LAI, FAPAR i FVC	74
6.3.2	Models de transferència radiativa: PROSAIL	75
6.3.3	Algorismes d'inversió	75
6.3.4	Dades de teledetecció	77
6.4	Resultats i validació	77
6.4.1	Mesures <i>in situ</i>	78
6.4.2	Validació	79
6.5	Conclusions	79
 Annex:		
	Scientific Publications	81
	 Acronyms	 143

List of Figures

1.1	Remote sensing from satellite active and passive sensors.	2
1.2	Left: The concept of imaging spectroscopy. Righth: Evolution of the spatial-spectral resolution through the years. Credits: Camps-Valls et al. (2011)	3
1.3	Validation approach (“bottom-up”) at global scale.	8
2.1	Forward (solid lines) and inverse (dashed lines) problems in remote sensing.	12
2.2	Radiation fluxes involved in the four-stream interaction concept applied in SAIL radiative transfer model.	15
2.3	PROSAIL coupling scheme. The parameters are highlighted either in light green (PROSPECT) or dark green (SAIL) color.	18
2.4	False RGB (NIR,Red,Green) images over a rice field as seen by (a) WorldView2 (2.5 m, acquired 01/07/2015) and (b) Landsat 8 OLI (30 m, acquired 06/07/2015). Bottom panels show photos taken in (e) May and (f) July of the small field channels used to drain water and stripes due tractor pass.	19
2.5	Neural network architecture.	20
2.6	Developed processing chain at local scale.	25
2.7	Study areas. Sentinel-2A surface reflectance RGB composite of the Spanish (30 July 2016), Italian (21 July 2016) and Greek (23 July 2016) rice local sites.	26
2.8	GPR (top), NN (middle) and KRR (bottom) theoretical performances using Landsat 8 OLI (left) and SPOT5 (right) simulated data. Plots are obtained by applying the trained models on test data.	27
2.9	Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Spanish study area in mid-May (left), early-June (middle) and mid-August (right). Grey mask covers non-rice areas.	29
2.10	Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Italian study area in mid-May (left), early-July (middle) and mid-August (right). Grey mask covers non-rice areas.	30

2.11	High-resolution LAI map obtained with Sentinel-2A data acquired in the Spanish study area on August 9, 2016. Right panel provides a zoom over a monitored field in the rice season.	31
2.12	LAI difference maps (up) and the corresponding scatter plots of the closest Sentinel-2A and Landsat-8 LAI estimates (bottom) over Spain (left), Italy (middle) and Greece (right). 32	
2.13	Temporal evolution of LAI estimates from SPOT5 data within a rice pixel using different soil background in (top) Spain and (bottom) Italy.	33
2.14	Temporal evolution of LAI and uncertainty (σ) over a rice field in Spain during the 2016 rice season.	33
3.1	Left: LAI-2000, right: Diagram of rings in LAI-2000.	37
3.2	NIKON Coolpix5000 camera and FC-E8 Nikon fisheye used in the digital hemispherical photography.	38
3.3	PocketLAI screenshot (up) and measurement scheme (bottom).	40
3.4	Comparison of effective plant area index measurements collected in the 2014 Spanish ERMES field campaign between: PocketLAI and LAI-2000 (up-left), PocketLAI and DHP (up-right), and DHP and LAI-2000 (bottom).	43
3.5	(Top) Leaf area index prediction maps, (middle) standard deviation maps and (bottom), coefficient of variation, obtained with GPR using PocketLAI, DHP and LAI-2000 ground data on the Landsat 8 image (DoY=212).	44
3.6	Scatter plots of estimated LAI values using Landsat (left) and SPOT5 (right) data versus <i>in situ</i> LAI measurements acquired with PocketLAI over Spain (up) and Italy (bottom). Standard deviation of measurements is drawn as horizontal error bars as well as bisector line (black line). For the sake of visualization, GPR prediction uncertainty (around 1) are not shown.	46
3.7	Scatter plots of Sentinel-2A estimated LAI values versus <i>in situ</i> LAI measurements during the 2016 rice season. For the sake of visualization, standard deviation of measurements and Sentinel-2A prediction uncertainty (around 0.8) are not shown.	47
3.8	Estimated temporal dynamics and the field LAI measurements within four representative ESUs in (top) the Spanish and (bottom) the Italian study area in 2015. The prediction uncertainty provided by the GPRs is shown in shaded blue (GPR-Landsat) and orange (GPR-SPOT5) around the mean prediction; in purple the overlap. The standard deviations of the <i>in situ</i> measurements are displayed as error bars.	48

3.9	Example of merged series of LAI on 3 sites. Dotted vertical bars indicate the sowing date. Numbers indicate fields ID. For the sake of visualization, standard deviations are not shown.	49
4.1	Global 8-biome map used in the retrieval chain of MOD15A2 products. Credits: Yan et al. (2016)	53
4.2	Spectral responses of reflectance inputs used in MOD15A2, GEOV1 and Land-SAF EPS. 56	
4.3	LAI (top), FAPAR (middle) and FVC (bottom) maps derived with the global processing chain in September 15, 2015.	59
4.4	Sites location of the BELMANIP-2.1 network used for intercomparison of EPS estimates and VEGA, GEOV1 and MOD15A2 products.	60
4.5	Scatter plots of estimated LAI, FAPAR and FVC values in 2015 and 2016 years over the BELMANIP-2.1 sites.	60
4.6	Box plots of estimated LAI, FAPAR and FVC values in 2015 and 2016 years over the BELMANIP-2.1 sites.	61
4.7	Scatter plots of estimated LAI values in the 2015 (top) and 2016 (middle and bottom) rice season over the Spanish, Italian and Greece rice sites.	62
4.8	Temporal evolution of representative rice pixels of MOD15A2 and GEOV1.4 products and the resampled Landsat-7/8 and Sentinel-2A LAI estimates over Spain (top), Italy (middle) and Greece (bottom).	63

List of Tables

2.1	Distribution of the canopy, leaf and soil parameters within the PROSAIL RTM.	26
2.2	Statistical scores between the Landsat estimated LAI taking into account different background conditions and ground measurements over the Valencia rice district.	31
3.1	Mean values of the statistical indicators (RMSE, MAE, absolute value of the ME, and R^2) between estimated and measured leaf area index in LAI_{APP} , LAI_{DHP} and LAI_{LIC} validation subsets. Standard deviations are shown in parenthesis.	44
3.2	Statistical indicators (coefficient of determination (R^2), Root Mean Squared Error (RMSE), Mean Error (ME), Mean Absolute Error (MAE), relative Root Mean Squared Error to mean ($rRMSE_m = RMSE/\hat{y}_i$) and relative Root Mean Squared Error to range ($rRMSE_r = RMSE/(max(y_i) - min(y_i))$) between the ground measurements and LAI estimates during the 2016 rice season over the three study areas.	46
4.1	Statistical indicators over the unseen test set (best results in bold). In the case of the single output algorithms, one has to train a model per parameter to retrieve, while for multioutput versions only a single model is learned.	57

PREFACE

Earth vegetation is the basis of our living on Earth. Vegetation plays an essential role in the study of global climate change influencing terrestrial CO₂ flux exchange and variability through plant respiration and photosynthesis. Vegetation monitoring shows valuable information for understanding natural and man-made environments through quantifying vegetation cover and use. Therefore, the assessment of current vegetation status is critical in order to foster Earth vegetation protection and restoration initiatives at both local and global scales. Many different programmes have been developed for monitoring vegetation status. This task has been carried out through meeting legislative responsibilities and international obligations (e.g., United Nations Convention to Combat Desertification, Convention on Biological Diversity) involving organizations worldwide, including national and international governments and authorities at regional and state levels. Furthermore, environmental markets also require vegetation monitoring to allow informed decision making.

Classical vegetation monitoring methods (e.g., field surveys, map interpretation and ancillary data analysis) are not effective to acquire vegetation dynamics because they are time consuming, outdated and often too expensive. In this context, during the last few years, different remote sensing methods are being employed for monitoring remotely detectable properties of vegetation and its dynamics. This interdisciplinary technology offers an increasingly practical and economical means to assess vegetation changes, over large areas at scales ranging from kilometers to decametric spatial resolutions.

In this context, we hypothesize that it is possible to accurately monitor vegetation spatial and time-resolved status from remote sensing by retrieving biophysical parameters at both local and global scales. However, monitoring vegetation by biophysical parameters retrieval from Earth observation (EO) data is a challenging problem. One can find in the literature that biophysical parameters estimation has been conducted using many different approaches: through the empirical relationships between the biophysical parameter of interest and vegetation indices (VIs); using pure statistical regression methods from existing remote sensing observations, products and *in situ* measurements; and inverting physically based canopy reflectance models using either look-up tables (LUTs) or machine learning techniques.

These methods have been applied by the remote sensing scientific community using different spectral and spatial-temporal remote sensing data without setting up a common framework for global and local estimation. In answer to this situation, and to test our hypothesis, we propose and describe the development of a generic processing chain able to retrieve key biophysical parameters such as Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) and Fractional Vegetation Cover (FVC). The biophysical parameters are derived from satellite data with different spatial and spectral characteristics useful for appraising Earth vegetation status at both local and global

scales within a common procedure. The estimation of these vegetation essential climate variables (ECV) is conducted implementing a hybrid method through the inversion of physically-based radiative transfer models with state-of-the-art machine learning regression algorithms. In particular, we focused on Gaussian Processes (GPs) for regression which offer interesting capabilities: not only state-of-the-art approximation results, but also the possibility to obtain confidence intervals for the predictions and automatic input bands ranking from the models.

The proposed retrieval chain was successfully applied to retrieve LAI at local scale, in the framework of ERMES (An Earth obseRvation Model based RicE information Service) project, using time series of Landsat-7 ETM+ (Enhanced Thematic Mapper), Landsat-8 OLI (Operational Land Imager) remote sensing data at 30 m spatial resolution. Similarly, the retrieval chain was applied using time series of Sentinel-2A data at 10 m spatial resolution. On the other hand, the proposed chain was also successfully applied to jointly estimate LAI, FAPAR and FVC at 1 km spatial resolution in the framework the Land Surface Analysis Satellite Applications Facility (Land-SAF) project, for an optimal exploitation of the AVHRR/MetOP data at global scale and in which the Universitat de València is responsible of developing the products of vegetation parameters. Therefore, we conclude that our Thesis is supported by the results shown in this memory.

ABSTRACT

This thesis' topics embrace remote sensing for Earth observation, specifically in Earth vegetation monitoring. The Thesis' main objective is to develop and implement an operational processing chain for crop biophysical parameters estimation at both local and global scales from remote sensing data. Conceptually, the components of the chain are the same at both scales: First, a radiative transfer model is run in forward mode to build a database composed by simulations of vegetation surface reflectance and concomitant biophysical parameters associated to those spectrum. Secondly, the simulated database is used for training and testing non-linear and non-parametric machine learning regression algorithms. The best model in terms of accuracy, bias and goodness-of-fit is then selected to be used in the operational retrieval chain. Once the model is trained, remote sensing surface reflectance data is fed into the trained model as input in the inversion process to retrieve the biophysical parameters of interest at both local and global scales depending on the inputs spatial resolution and coverage. Eventually, the validation of the leaf area index estimates is performed at local scale by a set of ground measurements conducted during coordinated field campaigns in three countries during 2015 and 2016 European rice seasons. At global scale, the validation is performed through intercomparison with the most relevant and widely validated reference biophysical products.

The work elaborated in this Thesis is structured in six chapters including an introduction of remote sensing for Earth observation, the developed processing chain at local scale, the ground LAI measurements acquired with smartphones, the developed chain at global scale, a chapter discussing the conclusions of the work, and a chapter which includes an extended abstract in Valencian. The Thesis is completed by an annex which include a compendium of peer-reviewed publications in remote sensing international journals. The outline of each chapter is summarized as follows:

Chapter 1 introduces the reader into the framework of remote sensing for Earth observation and reviews the main definitions, used methodologies and approaches for biophysical parameter estimation from remote sensing data.

Chapter 2 reviews the fundamentals of radiative transfer model inversion detailing the PROSAIL formulation basis and the main features of the machine learning regression techniques used for the inversion. The last part of the chapter describes the developed processing chain at local scale in the framework of ERMES project.

Chapter 3 reviews the *in situ* measurements acquisition procedure using classical instrumentation such as LAI-2000 and DHP techniques. This chapter introduces and describes the *in situ* LAI measurements acquired with new technologies such the use of smartphones through a dedicated application called PocketLAI. The chapter addresses the validation of

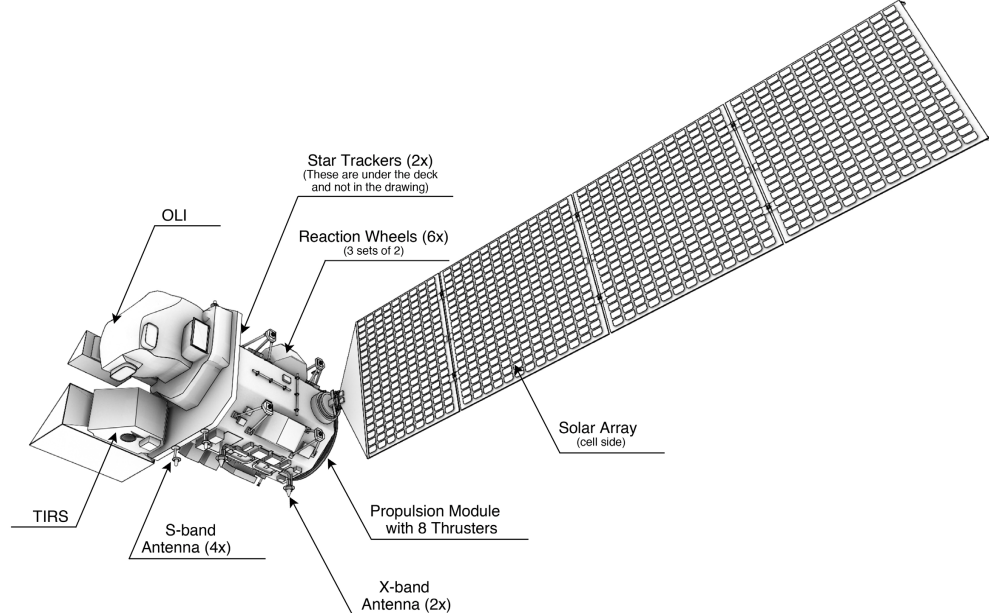
the LAI retrieval chain at local scale with ground data.

Chapter 4 describes the developed processing chain at global scale in the framework of Land-SAF project (exploitation of EUMETSAT satellites) and the indirect validation of the estimates by intercomparison with reference biophysical products such as MOD15A2, GEOV1 and VEGA products.

Chapter 5 concludes up the Thesis' achievements and discusses the main conclusions.

Chapter 6 provides a complete overview of the Thesis in Valencian.

Annex attaches the peer-reviewed scientific publications directly related with the work conducted in this Dissertation.



1. INTRODUCTION

Contents

1.1 Introduction to remote sensing for Earth observation

- 1.1.1 Regional Scale
- 1.1.2 Local scale

1.2 Biophysical parameters

- 1.2.1 Leaf Area Index (LAI)
- 1.2.2 Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)
- 1.2.3 Fractional Vegetation Cover (FVC)

1.3 Biophysical parameter retrieval methods

- 1.3.1 Statistical methods
- 1.3.2 Physical methods
- 1.3.3 Hybrid methods

1.4 Calibration and validation of remote sensing products

- 1.4.1 *In situ* measurements
- 1.4.2 Intercomparison of products

1.5 Summary

1.1 Introduction to remote sensing for Earth observation

Remote sensing can be defined as the science of acquiring and interpreting information about an object from distance. This is a very common day-to-day process: eyes detect electromagnetic energy in the visible light range which is interpreted by the brain as a real world image. On the scientific research and technologic application domains, this process is done by multi and hyperspectral cameras (or even by smartphone cameras) in industrial environments. In the field of Earth observation, remote sensing is performed by sensors typically onboard either satellite or airborne platforms thus recording reflected or emitted electromagnetic energy from Earth's surface. On the other hand, EO remote sensing can be also achieved with “near-surface” approaches using RGB (red, green, blue) and NIR (near infrared) cameras placed on selected sites for monitoring vegetation phenology (i.e. PhenoCam network). These kind of networks provide information needed to link what is actually happening “near-surface” and what is observed by airborne and satellite sensors.

Remote sensing sensors can be classified either as passive or active: passive sensors record reflected Earth's surface energy that was emitted from the sun, whereas active sensors record reflected Earth's surface electromagnetic energy which was previously emitted by themselves. Figure 1.1 represents an EO remote sensing process from active and passive remote sensors.



Figure 1.1: Remote sensing from satellite active and passive sensors.

The information conveyed by remotely sensed platforms is tied to the sensor characteristics and satellite capabilities (Lillesand et al., 2014). Spatial resolution gives the image pixel size (ranging from centimeters to kilometers), spectral resolution provides data at different spectral wavelengths (ranging from solar to thermal spectrums) and also information about spectral width, temporal resolution relates both the acquisition date and the frequency of acquisitions (ranging from a day to decades), and eventually spatial extent covers the ground area detected by the sensor's field of view (up to entire Earth). Consequently, the most suitable sensor/platform must be selected depending on the needs of the application (Benz et al., 2004). For example, Fig. 1.2 (left) shows the process of imaging spectroscopy which illustrates the simultaneous acquisition of spatially co-registered images in many spectral bands.

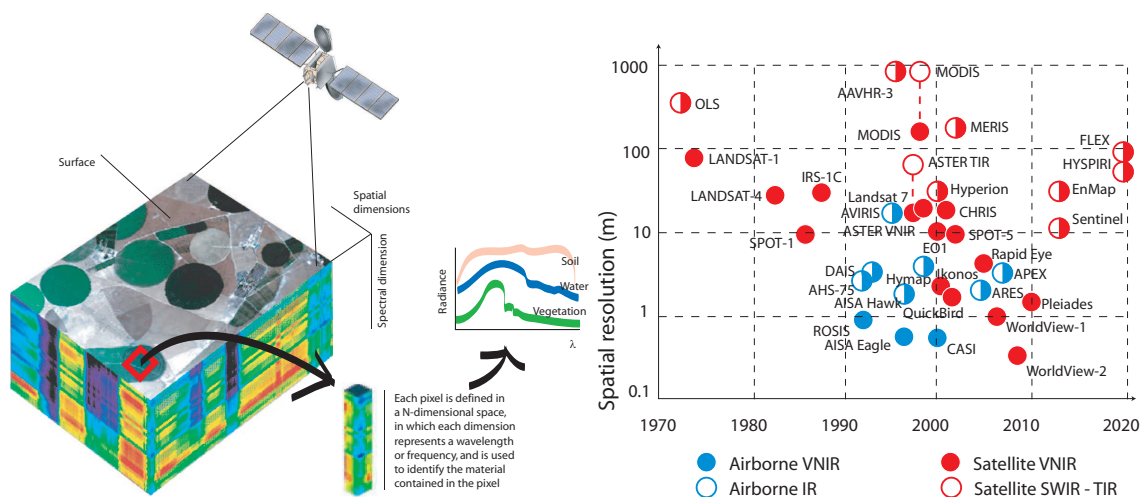


Figure 1.2: Left: The concept of imaging spectroscopy. Right: Evolution of the spatial-spectral resolution through the years. Credits: [Camps-Valls et al. \(2011\)](#)

1.1.1 Remote sensing at global scale

One of the main goals in remote sensing is to study the role of Earth's vegetation in large-scale global processes to improve our knowledge about how Earth functions as a system. This has been partly achieved employing different methods to determine and evaluate vegetation phenology using time series of vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) derived from a variety of operational medium resolution (i.e. kilometric spatial resolution) sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS, 1999 up to date) ([Justice et al., 1998](#)), the *Système Pour l'Observation de la Terre* (SPOT-V, 1999 to May 2014) ([Pasquier & Verheyden, 1998](#)) and the *Project for On-Board Autonomy* (PROBA-V, June 2014 up to date) ([Sterckx et al., 2014](#)). Similarly, the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) provides information for vegetation monitoring from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG), and the National Oceanographic and Atmospheric Administration/Advanced Very High Resolution Radiometer (NOAA/AVHRR) sensor was used to monitor vegetation ([Townshend, 1994](#)).

The availability of these remotely sensed data allowed the production of medium resolution operational products of biophysical parameters at global scales ([Baret et al., 2007](#); [Gao et al., 2008](#); [Ganguly et al., 2008](#); [Baret et al., 2013](#)). Nevertheless, higher spatial resolutions are needed when dealing with vegetation monitoring at a parcel level. Figure 1.2 (right) illustrates the evolution of remote sensing in terms of spatial-spectral resolution.

1.1.2 Remote Sensing at local scale

The recent advances in remote sensing platforms, sensors, statistical models and computational power as well as the huge amount of data availability, have provided new challenges and possibilities in remote sensing image processing ([Campos-Taberner et al., 2016c](#)). Especially, in the context of biophysical parameter estimation they have entailed in many studies of crop management at a parcel level ([Mulla, 2013](#)).

Accurate and timely information at high spatial resolution of crop condition and status is critical for crop management. For example, the Landsat Data Continuity Mission (LDCM) (Roy et al., 2014) and the recently European Sentinel-2 Mission (Drusch et al., 2012) provide free EO high-resolution (HR) information for a wide variety of land applications (Malenovsky et al., 2012) including crop monitoring. Several HR data are also available from different initiatives and platforms such as Worldview or RapidEYE. However, they are not free of charge which limits its usability in continuous and long term applications.

1.2 Biophysical parameters

Biophysical parameters are key variables which provide valuable information in order to assess changes in vegetation and climate processes affecting agriculture, natural resources and ecosystems at all levels. In the following sections, the most important biophysical parameters are presented, which will be subject of our study.

1.2.1 Leaf Area Index (LAI)

Leaf Area Index (LAI) is a key biophysical parameter which represents half the total intercepting leaf area per unit ground surface area (Chen & Black, 1992) being an important structural property of vegetation. LAI plays an important role in vegetation processes such as photosynthesis and transpiration, and is connected to meteorological/climate and ecological land processes. LAI has been widely used in many agricultural and remote sensing studies (Chen & Cihlar, 1996; Myneni et al., 1997; Carlson & Ripley, 1997; Fang & Liang, 2005). Concerning biomass and crop yield estimation, LAI estimates can be assimilated in crop models (Confalonieri et al., 2009) by means of forcing and/or recalibration techniques (Dorigo et al., 2007; Busetto et al., 2017).

Definition of LAI needs some additional remarks though. If no distinction is made between leaves (or needles in the case of conifers) and the other plant elements, the proper term to use is Plant Area Index (PAI) rather than LAI. PAI can be defined as half the total surface area of all above ground vegetation matter. Note that the remote sensing retrieved LAI is mainly corresponding to the green elements of vegetation, and thus the proper term to be used would be GAI (Green Area Index). Nevertheless, this Thesis uses the term LAI for the sake of simplicity and because it has been widely adopted in the remote sensing community.

Remote sensing LAI estimates refer to all green elements of the observed scene such as the understory under forests canopies. However, except when using directional observations, “*actual*” LAI is not directly accessible from remote sensing observations due to the possible heterogeneity in leaf distribution within the canopy volume. Therefore, remote sensing observations are rather sensitive to the “*effective*” LAI which is the value that provides the same diffuse gap fraction while assuming a random distribution of leaves. The difference between the *actual* LAI and the *effective* LAI may be quantified by the Clumping Index (CI) (Chen et al., 2005) that ranges from 0.5 in the case of very clumped canopies to 1 in the case of canopies with randomly distributed leaves.

1.2.2 Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)

The Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) is a biophysical parameter that accounts for the solar radiation absorbed by green vegetation in the 400 nm to 700 nm wavelength region and refers to the energy absorption power of vegetation. Thus, FAPAR can be considered as the fraction of the Photosynthetically Active Radiation (PAR) absorbed by a vegetation canopy. FAPAR depends on illumination conditions and canopy structure being a non-dimensional quantity varying from 0 (bare soil) to almost 1 (very dense canopies). FAPAR can be used to estimate net primary productivity (NPP) and is a key parameter in many land-surface models, climate-ecological and carbon cycle models (McCallum et al., 2009).

Since FAPAR depends on the angular position of the Sun and the relative contributions of the direct and diffuse radiation (illumination conditions) two FAPAR values must be considered: “*black-sky*” FAPAR (assuming only direct radiation) and “*white sky*” FAPAR (assuming the incoming radiation, both isotropic and diffuse). Most of the remote sensing FAPAR products are mainly defined as the “*black-sky*” FAPAR value for the same Sun position as that observed at the satellite overpass.

On the other hand, since the vast majority of the primary productivity models using FAPAR at the daily time step, the FAPAR parameter should correspond to the daily integrated FAPAR value taking into account both direct and diffuse radiation. The definition of FAPAR refers only to vegetation. FAPAR does not include the vegetation reflectance or the solar radiation absorbed by the background, except the background reflected fraction absorbed by the vegetation. FAPAR is approximately linearly related to reflectance.

1.2.3 Fraction of green Vegetation Cover (FVC)

Fraction of green Vegetation Cover (FVC) is defined as the green fraction as seen from the nadir direction (i.e., one minus the gap fraction). FVC is a key parameter used to measure green surface vegetation cover used to distinguish between vegetation and soil in soil-vegetation-atmosphere processes, including temperature and evapotranspiration. Unlike FAPAR, FVC does not depend on variables such as the geometry of illumination being thus a good alternative to vegetation indices for monitoring Earth’s green vegetation. Similarly to FAPAR, FVC is linearly related to reflectance.

1.3 Biophysical parameter retrieval methods

From a methodological point of view, biophysical parameters estimation from remote sensing (optical) data can be faced following either statistical, physical, or hybrid methods (Camps-Valls et al., 2011; Verrelst et al., 2015). In any case, all methods have in common that they rely on a model which aims to relate spectral data with the biophysical parameter of interest and can be seen as a regression problem. Regression methods for biophysical parameter estimation can be categorized in the following subclasses.

1.3.1 Statistical methods

Pure statistical methods extract patterns and trends from a data set, and try to understand the underlying physical laws ruling the relationships between them. That is what we call “learning from data” (Bishop, 2006; Hastie et al., 2009). Statistical methods can be divided

either into parametric or non-parametric methods depending if they assume explicit models of data distributions or not, respectively.

Parametric models rely on physical knowledge of the problem and build explicit parametrized expressions that relate some spectral bands (typically two) with the parameter of interest. They have been developed through empirical relations between *in situ* measurements of the biophysical parameter and vegetation indices derived from airborne and satellite spectra (Baret & Guyot, 1991). Therefore, this kind of empirical approaches rely on selected band ratios suited to retrieve useful information of vegetation status. For example, they include the red and near infrared bands when using the NDVI for LAI retrieval. It is worth noting that VIs are sensor-specific because of their particular spectral configurations.

Non-parametric methods do not assume an explicit (i.e., parametric) relation between the biophysical parameter of interest and the spectral reflectance. These methods estimate the variable of interest using a training database of input-output (i.e spectral reflectance-biophysical parameter) data pairs. In statistical non-parametric approaches, the biophysical parameter of interest to adjust the models come from concomitant *in situ* measurements and associated spectral data from remote sensing platforms. This is typically achieved in field campaigns conducted at the same time the satellite overpasses the study area where biophysical measurements have been acquired.

1.3.2 Physical methods

Physically-based methods are based on the physical knowledge describing the interactions between incoming radiation and vegetation. In this context radiative transfer models (RTMs) can be run in forward (direct) to create a database of spectral signatures covering a wide range of vegetation scenarios (i.e., from sparse to very dense vegetation, water stress, etc.) and allow the simultaneous analysis of the full spectrum as well as sensitivity studies of canopy parameters. In physical methods, the estimation problem reduces to minimize a proper cost or loss functional which implies to, for a given new spectrum, searching for similar spectra in the simulated database and assigning the most likely parameter.

1.3.3 Hybrid methods

Hybrid methods couple statistical non-parametric with physically-based methods. Hybrid methods rely on inverting a database generated by a radiative transfer model exploiting the generality of RTMs and the flexibility and computational efficiency of non-parametric non-linear regression methods. The retrieval is made by learning an inverse mapping of a function $g(\cdot)$ parametrized by weights ω , that approximates a matrix of state vectors, \mathbf{X} , that describe the system (e.g., LAI) using a set of observations as inputs, \mathbf{y} (e.g., reflectance): $\hat{\mathbf{X}} = g(\mathbf{y}, \omega)$.

The advantage of hybrid approaches is that a broad range of land cover situations can be simulated (e.g., up to hundred thousands), leading to a data set much bigger than what can be collected during a field campaign and you can invert the RTM data in a flexible and accurate manner with machine learning methods. A key issue when using hybrid methods is the representability of the simulated database because the inversion process will produce inaccurate estimates if the parameter space is underrepresented.

This Thesis is focused on the generation of a retrieval processing chain using the hybrid approach. Section 2.1 reviews the PROSAIL formulation in detail while Section 2.2 describes the machine learning inversion techniques evaluated in this Thesis.

1.4 Calibration and validation of remote sensing products

In remote sensing, validation can be defined as the process of assessing the quality of the biophysical data products by independent means (Justice et al., 2000). Validation is a key issue for remote sensing users and scientists in order to determine the most appropriate product, or combination of products, to use for their applications.

The international scientific community has discussed and addressed the validation of biophysical products through dedicated meetings and groups to support the validation of those products. The Committee Earth Observing Satellites' Working Group on Calibration and Validation (CEOS WGCV) was created to coordinate this calibration and validation activities worldwide. In this framework, the subgroup called Land Product Validation (LPV), defined the standards of validation protocols and information exchange in order to follow a common validation of biophysical products (Fernandes et al., 2014). Two main validation procedures can be undertaken: (i) *direct validation* using ground measurements and (ii) *indirect validation* through intercomparison of different products, which provided estimates of the same biophysical parameter. In the following sections, we introduce the concept of *in situ* measurements useful for validation as well as the main techniques used for intercomparison of products.

1.4.1 *In situ* measurements

Direct validation of biophysical estimates from remote sensing can be achieved through direct comparison with ground measurements of the biophysical parameter of interest. For this purpose, the ground data must be georeferenced using a GPS (Global Positioning System) in order to match the *in situ* value with the one retrieved from the satellite or airborne platform.

During the biophysical parameter retrieval, skewing effects such as topographic and atmospheric correction must be taken into account in order to derive the most similar parameter and value that is measured in the field. Moreover, when matching the values, a single ground value must be used to compare with the remote sensing product, therefore an *in situ* representative averaged value should be obtained from a set of measurements trying to cover as much as possible the extension of the pixel. This is intended to account for the pixel heterogeneity since a single *in situ* measurement could be biased, and down-weight the influence of erroneous acquisitions.

Chapter 3 reviews the methods and instruments used during this Thesis' which were used for *direct validation* of the biophysical parameters retrieved with the processing chain proposed in this work.

1.4.2 Intercomparison of products

Indirect validation includes the intercomparison of different products. The general approach relies on the comparison of the derived product with a reference product. Usually the products to be compared differ in spatial and temporal resolutions, thus the intercompar-

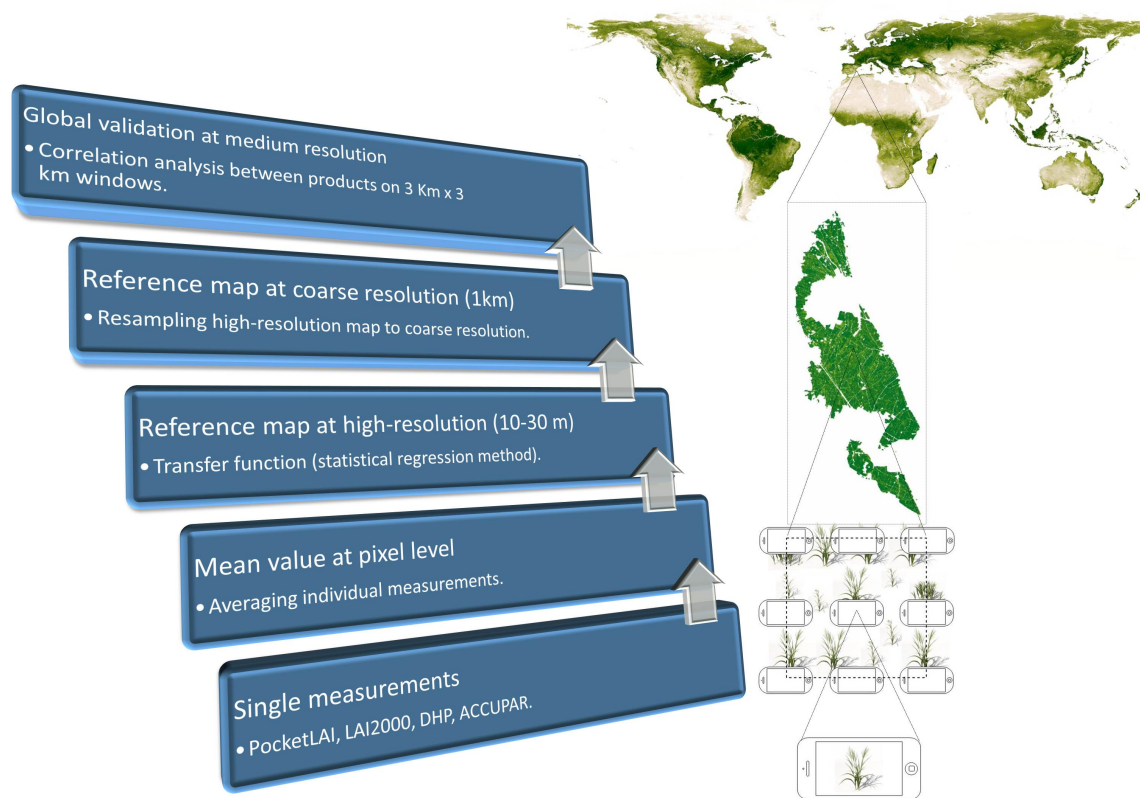


Figure 1.3: Validation approach (“bottom-up”) at global scale.

ison should be conducted for the closest dates of each product at a lowest spatial resolution to minimize differences due to spatial-temporal mismatch (Garrigues et al., 2008a).

After resampling to the coarser resolution, intercomparison can be achieved averaging the biophysical value of the estimates computed over the 3×3 pixels if more than 5 out of the 9 pixels are valid (Morissette et al., 2006) in order to reduce misregistration errors between images and inconsistencies associated to differences in the point spread functions (PSF). Global estimates are typically compared over co-located regions through scatter plots across global datasets or biomes.

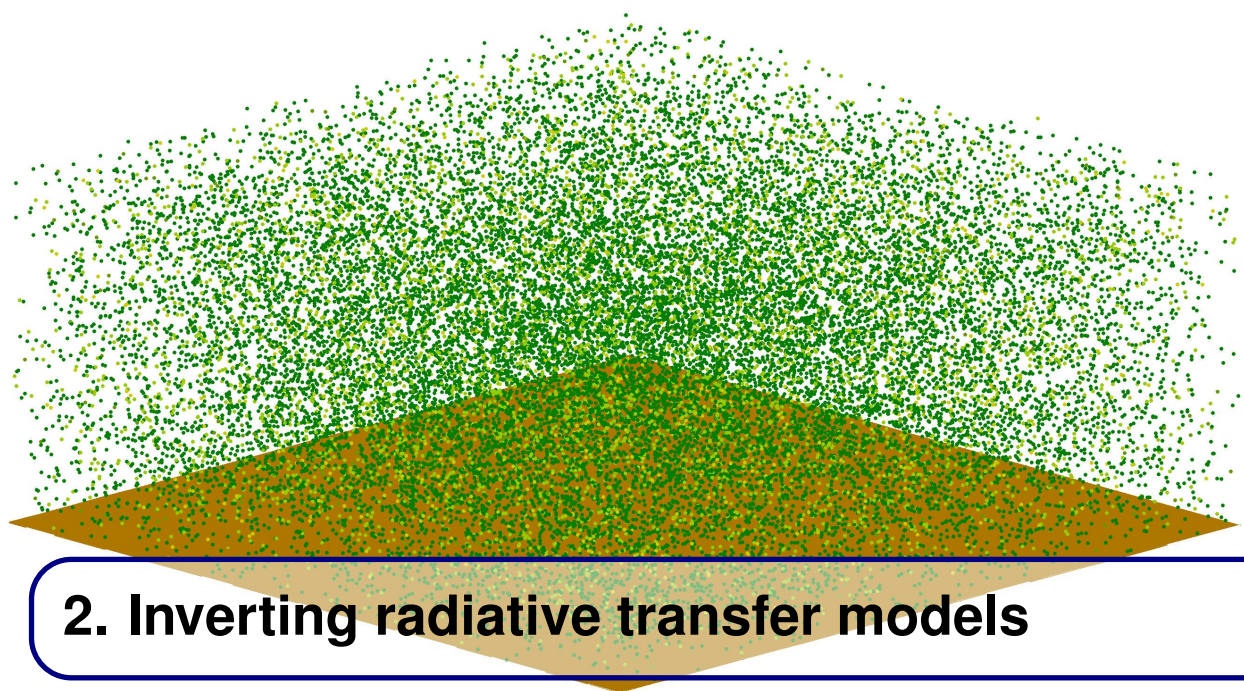
On the other hand, the *in situ* measurements allow to make indirect validation through a “bottom-up” approach from local measurements at field level to global comparison with satellite-derived biophysical products (see Fig. 1.3). This approach is based on an up-scaling process which establishes a relationship, mainly based on the calibration of empirical transfer functions, between the average *in situ* biophysical value from each pixel and the associated spectral (typically multispectral) values from high-resolution imagery. Consequently, this process can be classified as a pure statistical regression method as described in Section 1.3.1. The transfer function is site-specific (Cohen et al., 2003) and the empirical relationships selected are usually based on linear regressions with VIs (Fernandes et al., 2003) or multiple linear regressions either with radiances or reflectances (Baret et al., 2005).

1.5 Summary

This Chapter introduced remote sensing for Earth observation at both local and global scales. The biophysical parameters used in this Thesis have been described also highlighting their relevance in remote sensing. In addition, we have also introduced the biophysical parameter retrieval methods: statistical, physical and hybrid methods.

The last part of the chapter summarized the validation and calibration processes of biophysical parameters remote sensing products. The biophysical parameter *in situ* measurements have been introduced in order to perform direct validation of the retrievals, and the concept of indirect validation has been introduced for intercomparison of remote sensing biophysical products.

This chapter has introduced the reader the main components of the processing chain developed in this Thesis. In the following chapters, we describe in detail the developed processing chain at both local and global scales emphasizing the description of the radiative transfer model inversion, the regression techniques and the *in situ* measurements.



2. Inverting radiative transfer models

Contents

2.1 Introduction to forward and inverse modelling

2.2 PROSAIL

2.2.1 PROSPECT

2.2.2 SAIL

2.2.3 Biochemical and structural parameters

2.3 Statistical modelling

2.3.1 Introduction to statistical inference

2.3.2 Machine learning methods

2.3.3 Implementation of the algorithms

2.4 Developed processing chain at local scale

2.4.1 Remote sensing data and sites

2.4.2 Theoretical performances

2.4.3 High-resolution LAI maps

2.4.4 Influence of the background spectra and utility of GPR uncertainty

2.5 Concluding remarks

This Chapter is partly based on the following publications:

- **Campos-Taberner, M.**, García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A. & Boschetti, M. (2017). “Multitemporal and Multiresolution Leaf Area Index Retrieval for Operational Local Rice Crop Monitoring”, 2016. *Remote Sensing of Environment*, 187, 102 - 118, 2016.
- **Campos-Taberner, M.**, García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Busetto, L., Katsantonis, D., Stavrakoudis, D., Minakou, C., Gatti, L., Barbieri, M., Holecz, F., Stroppiana D. & Boschetti, M., “Exploitation of SAR and optical Sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index”, *Remote Sensing*, 9(3), 248, 2017.

2.1 Introduction to forward and inverse modelling

The interaction between radiation and canopy is modeled with radiative transfer theory. Radiative transfer is the process of transmission of electromagnetic energy through a medium. Radiation energy is lost either by absorption or scattering on canopy objects or increased through emission of heat sources, and the intensity of these processes depends on the wavelength.

Radiative transfer theory uses mathematical models to describe light interception by plant canopies and the relationship between vegetation reflectance and biophysical parameters. A wide range of radiative transfer models have been proposed in the literature since last decades (Suits, 1971–1973; Verhoef, 1984, 1985; Verhoef & Bach, 2007). RTMs describe vegetation properties at leaf and canopy levels. PROSPECT radiative transfer model (Jacquemoud & Baret, 1990) simulates transmittance and directional–hemispherical reflectance of leaves between 400 nm and 2500 nm, while the SAIL (Scattering by Arbitrary Inclined Leaves) model (Verhoef, 1984) accounts for the interactions at canopy level.

This kind of mechanistic models implement the laws of Physics and allow us to compute the data values given a model (Snieder & Trampert, 1999). This is known as the *forward mode* in radiative transfer modeling (Jacquemoud et al., 2000). On the other hand, in the *inverse mode*, the aim is to reconstruct the model from a set of measurements, in our case, remote sensing observations either from satellites or airborne platforms. Figure 2.1 shows the scheme of forward-inverse procedure in remote sensing. Notationally, a forward model describing the system is expressed as $\mathbf{x} = g(\mathbf{y}, \omega)$, where \mathbf{x} is a measurement obtained by the satellite (e.g. reflectance); the vector \mathbf{y} represents the state of the biophysical variables on Earth (which we desire to infer or predict and is often referred to as *outputs* in the inverse modeling approach); ω contains a set of controllable conditions (e.g. wavelengths, viewing direction, time, Sun position, and polarization); and $g(\cdot)$ is a function which relates \mathbf{y} with \mathbf{x} . Such a function g is typically considered to be nonlinear, smooth and continuous. The aim in these problems is to obtain an inverse model, $f(\cdot) \approx g^{-1}(\cdot)$, parametrized by θ , which approximates the biophysical variables \mathbf{y} given the data \mathbf{x} received by the satellite, i.e. $\hat{\mathbf{y}} = f(\mathbf{x}, \theta)$.

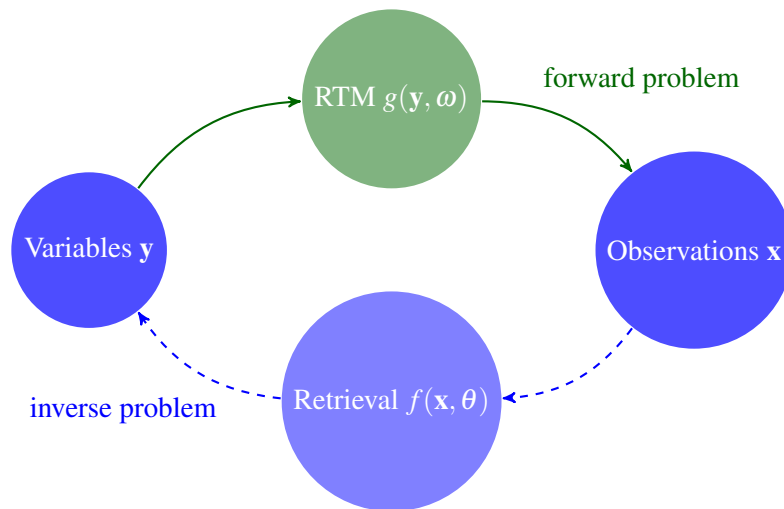


Figure 2.1: Forward (solid lines) and inverse (dashed lines) problems in remote sensing.

In the following sections we describe the PROSAIL RTM as well as the machine learning techniques used in this Thesis for the inversion process (see Sections 2.2 and 2.3). Finally, Section 2.4 presents the developed processing chain at local scale.

2.2 PROSAIL

PROSAIL (Baret et al., 1992) results from the PROSPECT and SAIL coupling. This coupling can be achieved introducing the output leaf reflectance and transmittance of the PROSPECT into the SAIL model to simulate the canopy reflectance in all directions and wavelengths. In the ensuing sections both PROSPECT and SAIL models are described.

2.2.1 PROSPECT

PROSPECT (Jacquemoud & Baret, 1990) is a radiative transfer model based on an improved version of the generalized *plate model* (Allen et al., 1969) which represents the optical properties of plant leaves from 400 nm to 2500 nm. This *plate model* considers a compact plant leaf as a transparent plate with rough plane parallel surfaces. The incoming beam penetrates the leaf with incident directions within a solid angle Ω defined by a maximum incident angle α relative to the normal of the leaf plane. After its penetration inside the leaf, the light flux is assumed to be diffuse and isotropic. In order to take into account the incoming beam within the solid angle, the expressions of the *plate model* (reflectance ρ_α and transmittance τ_α) can be written as

$$\rho_\alpha = [1 - t_{av}(\alpha, n)] + \frac{t_{av}(90, n)t_{av}(\alpha, n)\theta^2[n^2 - t_{av}(90, n)]}{n^4 - \theta^2[n^2 - t_{av}(90, n)]^2}, \quad \text{and} \quad (2.1)$$

$$\tau_\alpha = \frac{t_{av}(90, n)t_{av}(\alpha, n)\theta^2 n^2}{n^4 - \theta^2[n^2 - t_{av}(90, n)]^2}, \quad (2.2)$$

where n is the refractive index, θ is the transmission coefficient of the plate and $t_{av}(\alpha, n)$ is the transmissivity of a dielectric plane surface, averaged over all directions of incidence and over all polarizations.

The generalization of the *plate model* consists of stacking elementary layers. A leaf is then assumed to be composed of a pile of N homogeneous layers separated by $N - 1$ air spaces. The solution of this problem has been given long time ago by Stokes (Stokes, 1860). As the non-diffuse character of the incident beam concerns only the top of the pile, the Stokes system has been modified by separating the first layer from the $(N - 1)$ other ones. The total reflectance and transmittance for N layers are given by

$$R_{N,\alpha} = \rho_\alpha + \frac{\tau_\alpha \tau_{90} R_{N-1,90}}{1 - \rho_{90} R_{N-1,90}}, \quad \text{and} \quad (2.3)$$

$$T_{N,\alpha} = \frac{\tau_\alpha T_{N-1,90}}{1 - \rho_{90} R_{N-1,90}}, \quad (2.4)$$

where ρ_{90} and τ_{90} are the reflectance and transmittance of an internal elementary layer.

On the specific absorbtion spectra of pigments and water

The *plate model* provides a transmission coefficient θ , which is related to the absorption coefficient k through the following equation (Allen et al., 1969):

$$\theta = (1 - k) \exp(-k) + k^2 \int_k^\infty \frac{\exp(-x)}{x} dx. \quad (2.5)$$

The spectral absorption coefficient $k(\lambda)$ can be written in the form

$$k(\lambda) = \sum_i^M K_i(\lambda) C_i + k_e(\lambda), \quad (2.6)$$

where λ is the wavelength, $K_i(\lambda)$ is the spectral specific absorption coefficient relative to the leaf component i , and C_i is the leaf component i content per unit leaf area. k_e accounts for the case of dry flat leaves where the $C_i = 0$ and the absorptance is different from zero.

PROSPECT-4 accounts for Leaf optical properties taking into account the mesophyll structural parameter (N) and the leaf chlorophyll (C_{ab}), dry matter (C_m), and water (C_w) contents. Therefore, the spectral absorbtion coefficient is

$$k(\lambda) = K_{ab} C_{ab} + K_w C_w + \frac{K_m C_m}{N} + k_e(\lambda). \quad (2.7)$$

PROSPECT-5 version also accounts for brown pigments (C_{bp}) treating total chlorophylls and carotenoids (C_{ar}) separately. The parameters related to N , K_{ab} , and K_w , K_m , as well as K_{bp} and K_{ar} in the case of PROSPECT-5, are fitted using experimental data corresponding to a wide range of plant types and status.

2.2.2 SAIL

SAIL radiative transfer model is an extension of the Suits model (Suits, 1971–1973), to simulate the bidirectional reflectance factor of a turbid medium. The model describes a four-stream flux interaction concept taking into account four upward/downward radiative fluxes (see Fig. 2.2): the incoming solar radiation (E_s), the diffuse downward radiation (E^-), the diffuse upward radiation (E^+) and the radiance in the observer's direction (E_o). SAIL solves the radiance in the observer's direction (bidirectional reflectance), taking into account the energy losses due to scattering and absorption in the canopy making assumptions on the morphology of the canopy layer. The model assumes that the layer is horizontal, infinite and homogeneous, and the canopy consists of small flat leaves. SAIL accounts for leaves with any orientation. The orientation of the leaves is described by the leaf orientation density function (LODF) which can be replaced by the Leaf Inclination Density Function (LIDF) assuming that the leaves azimuth is distributed randomly. The LIDF can be represented by an Average Leaf Angle (ALA) with a distribution around it and the number of leaves in the canopy is quantified by the Leaf Area Index (LAI). The hot-spot effect can also be introduced as a function of the ratio of leaf size to canopy height (Kuusk, 1991) which accounts when the zenith and azimuth angle of observation are equal to those of the sun.

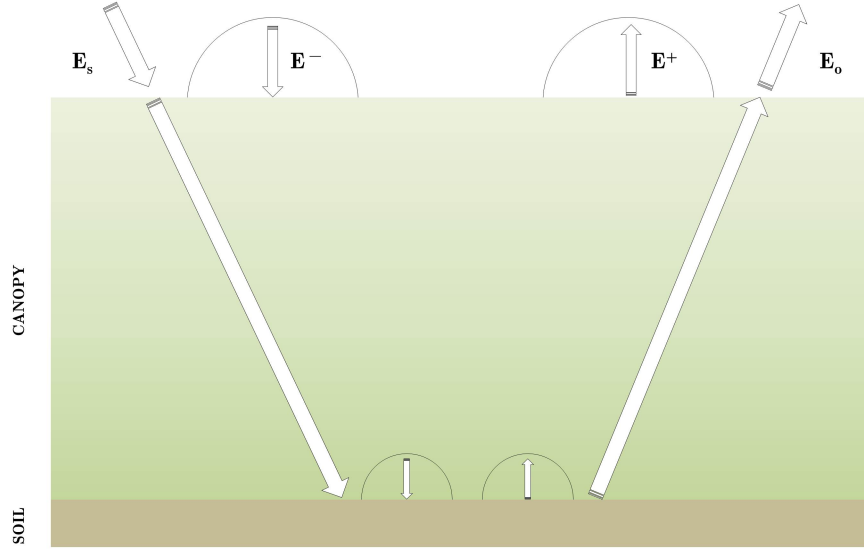


Figure 2.2: Radiation fluxes involved in the four-stream interaction concept applied in SAIL radiative transfer model.

Four-stream radiative transfer theory (Verhoef, 1985) describes the interactions among the four fluxes involved in the process. Assuming infinitesimal leaf layers, this leads to a four linear differential equations system given by:

$$\begin{aligned}
 \frac{d}{Ldx} E_s &= kE_s, \\
 \frac{d}{Ldx} E_- &= -s'E_s + aE^- - \sigma E^+, \\
 \frac{d}{Ldx} E_+ &= s'E_s + \sigma E^+ - aE^-, \quad \text{and} \\
 \frac{d}{Ldx} E_o &= wE_s + vE^- - KE_o
 \end{aligned} \tag{2.8}$$

where x is the relative optical height and L refers to LAI. The optical height runs to -1 (bottom of canopy) to zero (top of canopy). The coefficients describe the interactions among the fluxes and are given in the Bunnik notation (Bunnik, 1978). The formal solution of this differential equations system (Verhoef, 1985) reveals a set of relations between the incident, reflected and transmitted fluxes as follows:

$$\begin{aligned}
 E_s(-1) &= \tau_{ss}E_s(0), \\
 E_-(-1) &= \tau_{sd}E_s(0) + \tau_{dd}E^-(0) + \rho_{dd}E^+(-1), \\
 E_+(0) &= \rho_{sd}E_s(0) + \rho_{dd}E^-(0) + \tau_{dd}E^+(-1), \\
 E_o(0) &= \rho_{so}E_s(0) + \rho_{do}E^-(0) + \tau_{do}E^+(-1) + \tau_{oo}E_o(-1),
 \end{aligned} \tag{2.9}$$

where ρ and τ are either the reflectances and the transmittances of the canopy layer. The double subscripts refer to the type of flux and the interaction, namely: 'so' (bidirectional

reflectance); 'ss' (direct transmittance in the direction of the solar beam); 'sd' (directional-hemispherical solar flux); 'dd' (bihemispherical); 'do' (hemispherical-directional in viewing direction); 'oo' (direct transmittance in the observer's direction).

To account for the soil background interaction, which is placed underneath the vegetation layer, one obtains the equation:

$$E_o(-1) = E^+(-1) = r_s (E_s(-1) + E^-(-1)), \quad (2.10)$$

where r_s is the Lambertian soil reflectance. Combining Eq. 2.10 with the second line of Eq. 2.9, the fluxes at the soil level (bottom of canopy) can be computed and inserted in the last line of Eq. 2.9 to solve the top of canopy flux-equivalent radiance:

$$E_o(0) = \left[\rho_{so} + \frac{(\tau_{sd} + \tau_{ss})r_s(\tau_{do} + \tau_{oo})}{1 - r_s\rho_{dd}} \right] E_s(0) + \left[\rho_{do} + \frac{(\tau_{do} + \tau_{oo})r_s\tau_{dd}}{1 - r_s\rho_{dd}} \right] E^-(0). \quad (2.11)$$

The computation of the bidirectional canopy reflectance ρ_{so} can be modified to account for the hot-spot effect (Verhoef et al., 2007). In addition, to account for the canopy optical properties, the bidirectional gap fraction (τ_{ssoo}) has been added to the output parameters as follows:

$$\begin{aligned} E_o(0) = & \left[\rho_{so} + \tau_{ssoo}r_{so} + \frac{(\tau_{ss}r_{sd} + \tau_{sd}r_{dd})\tau_{do} + (\tau_{sd} + \tau_{ss}r_{sd}\rho_{dd})r_{do}\tau_{oo}}{1 - r_{dd}\rho_{dd}} \right] E_s(0) \\ & + \left[\rho_{do} + \frac{\tau_{dd}(r_{dd}\tau_{do} + r_{do}\tau_{oo})}{1 - r_{dd}\rho_{dd}} \right] E^-(0). \end{aligned} \quad (2.12)$$

Equation 2.12 includes the bidirectional reflectance of the ensemble and the directional reflectance for hemispherical diffuse incident radiation from the sky. The reflection properties of the non-Lambertian soils are expressed by r_{so} , r_{do} , r_{sd} , and r_{dd} . The resulting top-of-canopy reflectance factors are then given by

$$r_{so}^* = \rho_{so} + \tau_{ssoo}r_{so} + \frac{(\tau_{ss}r_{sd} + \tau_{sd}r_{dd})\tau_{do} + (\tau_{sd} + \tau_{ss}r_{sd}\rho_{dd})r_{do}\tau_{oo}}{1 - r_{dd}\rho_{dd}}, \quad \text{and} \quad (2.13)$$

$$r_{do}^* = \rho_{do} + \frac{(\tau_{do} + \tau_{oo})r_s\tau_{dd}}{1 - r_s\rho_{dd}}. \quad (2.14)$$

Since the atmosphere has some influence on measured surface reflectance values, and the irradiance coming from the sky depends on the average surface reflectance of the surroundings, the measured directional reflectance factor at ground level (r_o) is affected by the surroundings and can be computed as:

$$r_o = \frac{r_{so}^*E_{sun} + r_{do}^*E_{sky}}{E_{tot}}, \quad (2.15)$$

where E_{sun} , E_{sky} and E_{tot} are the solar, sky and total irradiance at ground level (Verhoef & Bach, 2003).

In addition, absorptances of the isolated canopy layer for solar and diffuse incident flux, can be expressed as $\alpha_s = 1 - \rho_{sd} - \tau_{sd} - \tau_{ss}$ and $\alpha_d = 1 - \rho_{dd} - \tau_{dd}$. Including the soil background reflectance in these absorptances we obtain

$$\alpha_s^* = \alpha_s + \frac{\tau_{ss}r_{sd} + \tau_{sd}r_{dd}}{1 - r_{dd}\rho_{dd}}\alpha_d, \quad \text{and} \quad (2.16)$$

$$\alpha_d^* = \alpha_d + \frac{\tau_{dd}r_{dd}}{1 - r_{dd}\rho_{dd}}\alpha_d. \quad (2.17)$$

The fraction absorbed photosynthetically active radiation (FAPAR) can be computed by means of the canopy layer absorptances. The absorbed radiation flux density (A) at a certain wavelength can be expressed by

$$A = \alpha_s^* E_{sun} + \alpha_d^* E_{sky}. \quad (2.18)$$

Integrating over the region where the active radiation is photosynthetically absorbed (400 to 700 nm), and dividing by the total incident flux over this range we can calculate the FAPAR:

$$fAPAR = \frac{\alpha_s^* E_{sun} + \alpha_d^* E_{sky}}{E_{tot}} \Big|_{400}^{700}. \quad (2.19)$$

2.2.3 Biochemical and structural parameters

PROSAIL simulates the top of canopy bi-directional reflectance as a function of PROSPECT and SAIL input parameters. Figure 2.3 illustrates the coupling between PROSPECT and SAIL to derive the PROSAIL simulations. Leaf optical properties are expressed taking into account the mesophyll structural parameter (N) and the leaf chlorophyll (C_{ab}), dry matter (C_m), and water (C_w) contents (see Eq. 2.7). The C_w is tied to dry matter content ($C_w = C_m \times C_{wREL} / (1 - C_{wREL})$) assuming that green leaves have a relative water content (C_{wREL}) varying within a relatively small range (Baret et al., 2007). In addition, brown pigments (C_{bp}) and carotenoid content (C_{ar}) can also be taken into account if PROSPECT-5 is used in the coupling.

Canopy structure is characterized by the SAIL parameters ALA, LAI and Hotspot. A multiplicative brightness parameter (β_s) can be introduced and applied to spectral soil signatures to represent different background reflectance types (Baret et al., 2007). The system geometry is described by the solar zenith angle (θ_s), view zenith angle (θ_v), and the relative azimuth angle between both angles ($\Delta\Theta$).

In order to account for pixel mixed conditions, a pixel can be represented by a linear mixture of vegetation (vCover) and bare/flooded soil (1-vCover) spectra. Taking this heterogeneity into account, the pixel reflectance can be expressed as $R = R_{veg} \times vCover + R_{soil} \times (1 - vCover)$, where R_{veg} and R_{soil} account for pure vegetation reflectance and background, respectively. This simple approach was introduced by Baret et al. (2007) to account for clumping at the landscape level.

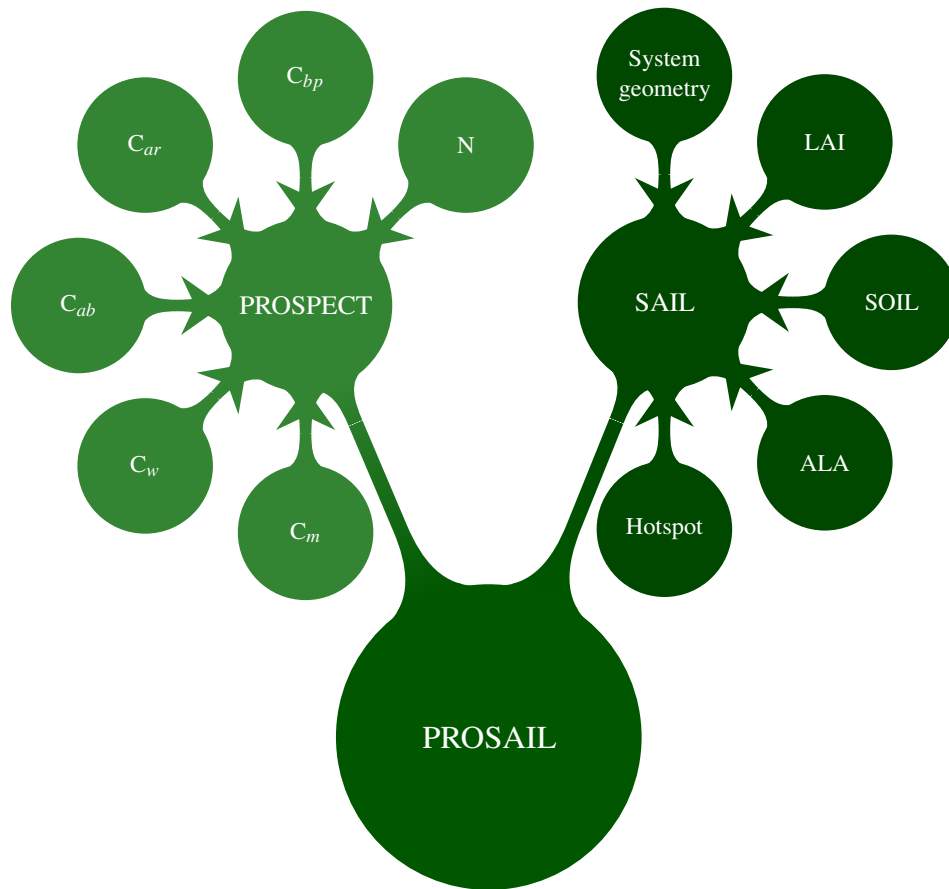


Figure 2.3: PROSAIL coupling scheme. The parameters are highlighted either in light green (PROSPECT) or dark green (SAIL) color.

In addition, vCover can also account for sub-pixel variability at decametric resolution due to several reasons (see Figure 2.4):

1. Non-vegetated areas found at field border due to mix between crop and surrounding land cover.
2. Patches of bare soil, in particular in poor soil fertility area due to germination problems (plant density) related to soil variability.
3. Small stripes due machinery practises.
4. Small channels for water drainage which are always present in some crops.

2.3 Statistical inverse modelling

2.3.1 Introduction to statistical inference

Machine learning is a multidisciplinary field that has become one of the key tools used in a huge range of diverse applications such as signal and image processing, computer science and even in web page ranking. Machine learning relies on the concept that an algorithm can learn from data without being explicitly programmed to perform specific tasks. They learn and adapt from previous computations (iteratively) to produce reliable, repeatable decisions and results.

In the last decades, machine learning has been used in remote sensing applications including target and change detection, image classification and biophysical parameters

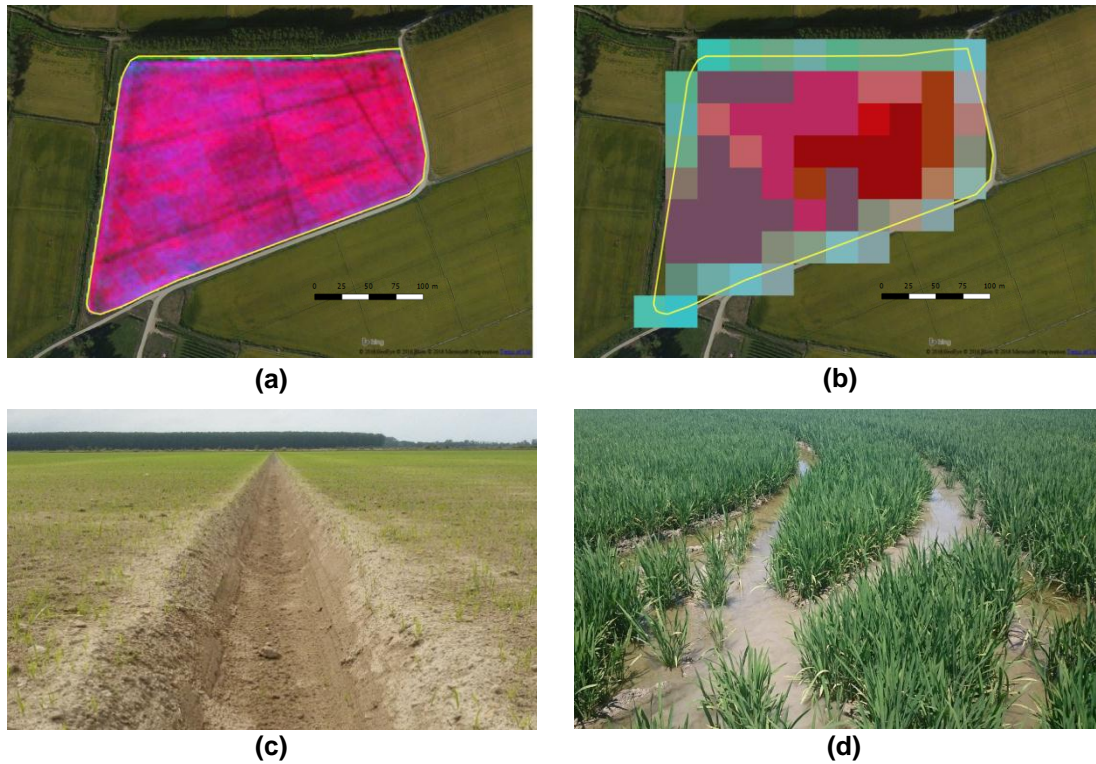


Figure 2.4: False RGB (NIR,Red,Green) images over a rice field as seen by (a) WorldView2 (2.5 m, acquired 01/07/2015) and (b) Landsat 8 OLI (30 m, acquired 06/07/2015). Bottom panels show photos taken in (e) May and (f) July of the small field channels used to drain water and stripes due tractor pass.

retrieval (Camps-Valls et al., 2011). Among them, machine learning regression algorithms (Witten et al., 2016) have been used in many remote sensing studies as powerful tools for the inversion of RTMs (Verrelst et al., 2012; Caicedo et al., 2014). These algorithms are generally robust and in most of the cases they are very fast to apply once trained. Machine learning methods are able to cope with the nonlinear relationship between biophysical parameters and its associated reflectance. However, it is worth noting that RTM inversion poses several methodological problems: it may lead to high computational cost, and being an ill-posed problem, it may give rise to unstable results. Prior information related to the distribution of the canopy variables and representative background spectra can be implemented in the RTM to better address the inversion process (see Section 2.4.2).

2.3.2 Machine learning methods

Neural networks

Neural networks (NN) have been used since 1990s for retrieving vegetation properties (Franci & Panigrahi, 1997; Paruelo & Tomasel, 1997; Kimes et al., 1998; Haykin, 1999) outperforming linear models and becoming the most popular machine learning tool during the last decade in hybrid retrieval methods. NN are essentially a fully connected structure of neurons organized in layers. A neuron basically performs a linear regression followed by a pointwise non-linear function (e.g., a sigmoid). Neurons of different layers are interconnected with weighted links. Therefore, NN formed only by a single neuron,

produces similar (or slightly better because of the nonlinearity) results than a linear regression. A NN is formally defined by the interconnection pattern between the different neuron layers, the learning process which updates the weights of the interconnections, and the activation function that converts the neurons weighted input with its output activation.

The most common NN structure is a feed-forward NN, where information flows from the input nodes (e.g. reflectances), data pass hidden nodes (if any) toward the output nodes (e.g. biophysical parameter). In this standard multi-layer model (Fig. 2.5), the neuron j in layer $l + 1$ yields $x_j^{l+1} = f(\sum_i w_{ij}^l x_i + w_{bj}^l)$, where w_{ij}^l are the weights connecting neuron i in layer l to neuron j in layer $l + 1$, w_{bj}^l are the bias term of neuron j in layer l , and f is a nonlinear activation function. Thus, given a set of inputs x_i , the output is denoted as $y = f(x_i)$.

Training a NN needs selecting a structure (number of hidden layers and nodes per layer), properly initialization of the weights, shape of the nonlinearity, learning rate, and regularization parameters to prevent overfitting. The problem of ANNs is that their performance is typically determined by their design: too few or too many layers and/or neurons may significantly reduce their accuracy. The use of several layers enables a more general application since the performance of networks with a single layer of adaptive weights is limited. However, in remote sensing applications, one hidden layer is most often sufficient. In order to train the network (i.e. fit the weights), one has to select a cost function (i.e least squares loss) and an algorithm to do this (i.e standard back-propagation algorithm). The back-propagation algorithm is one of the most widely used algorithms for training NNs. Basically, after choosing the weights of the network randomly, the back-propagation algorithm fits the weights in the following four steps: i) Feed-forward computation, ii) back-propagation to the output layer, iii) back-propagation to the hidden layer, and iv) weight updates. The algorithm stops when the value of the error function has become sufficiently small.

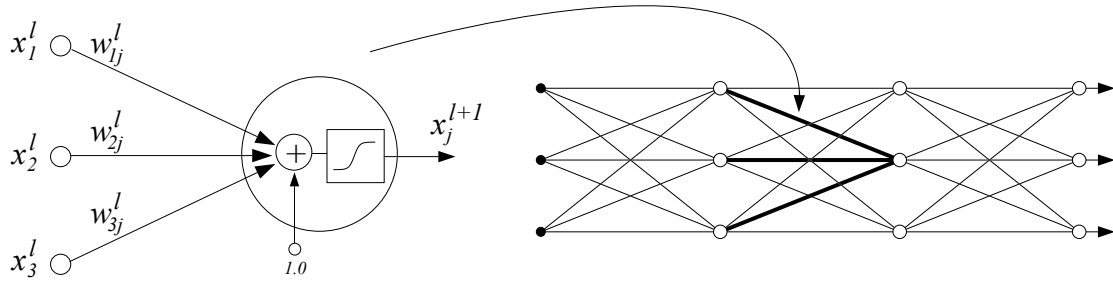


Figure 2.5: Neural network architecture.

Kernel methods

Kernel methods (Shawe-Taylor & Cristianini, 2004) owe their name to the use of kernel functions, which measure similarities between input samples of a dataset. When the input space is not endorsed with a proper notion of distance, kernel methods map the data in to a Hilbert space \mathcal{H} of higher dimensionality, and then compute the (dot product) similarity therein. The advantage in kernel methods is that one does not need to know the feature map to \mathcal{H} , neither to compute the data coordinates in \mathcal{H} to estimate similarities, which can be done implicitly via reproducing kernel functions. Thus, given a dataset of inputs $\mathbf{x} \in \mathcal{X}$,

the feature mapping can be defined as:

$$\phi : \mathcal{X} \rightarrow \mathcal{H}, \mathbf{x} \mapsto \phi(\mathbf{x}).$$

The similarity between the elements in \mathcal{H} can now be measured using its associated dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Thus, we can introduce a function called *kernel* that computes the similarity between inputs $(\mathbf{x}, \mathbf{x}' \in \mathcal{X})$ as:

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}}. \quad (2.20)$$

Kernels that can be written in the form of Eq.2.20 coincide with the class of positive definite kernels and one does not need to know the explicit form of the feature map since it is implicitly defined through the kernel. In addition, Eq. 2.20 is also known in the machine learning literature as the kernel trick (Scholkopf & Smola, 2002) which states that all dot products in \mathcal{H} can be implicitly computed by simply using a kernel function defined on the input data.

Kernel ridge regression

The Kernel ridge regression (KRR) is considered as the nonlinear (kernel) version of the canonical least squares linear regression. In this context, the objective is to compute a canonical least squares linear regression into a Hilbert space \mathcal{H} where samples have been mapped to through a mapping $\phi(\mathbf{x}_i)$. In matrix notation, the model is given by

$$\mathbf{Y} = \Phi \mathbf{W} + \mathbf{b} \quad (2.21)$$

where Φ is the mapped samples matrix $[\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]^\top$, whose dimension is $n \times d_{\mathcal{H}}$. In this case, the matrix \mathbf{W} has very high dimensionality and in principle unknown since the space \mathcal{H} and the vectorial mapping Φ are not defined explicitly. It has been assumed an additive noise model $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{E}$ with Gaussian noise $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, of zero mean and standard deviation σ_n .

Then, similarly to the regularized linear regression setting, we want to minimize the regularized squared loss function:

$$\mathcal{L}_p = \|\mathbf{Y} - \Phi \mathbf{W}\|^2 + \lambda \|\mathbf{W}\|^2 \quad (2.22)$$

with respect to weights \mathbf{W} . Therefore, taking derivatives with respect to \mathbf{W} and equating them to zero, gives

$$\mathbf{W} = (\Phi^\top \Phi + \lambda \mathbf{I})^{-1} \Phi^\top \mathbf{Y} \quad (2.23)$$

This problem is not solvable as the inverse runs on matrix $\Phi^\top \Phi$, which is of size $d_{\mathcal{H}} \times d_{\mathcal{H}}$, and Φ in principle unknown. However, we can express the solution as a linear combination of mapped samples (Representer's theorem) $\mathbf{W} = \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i)$ whose matrix notation is

$$\mathbf{W} = \Phi^\top \boldsymbol{\alpha}.$$

Therefore, the solution can be expressed as a function of the weights α :

$$\alpha = (\Phi\Phi^\top + \lambda\mathbf{I})^{-1}\mathbf{Y} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{Y} \quad (2.24)$$

At this point we only need to compute the inverse of the (regularized) kernel matrix $\Phi\Phi^\top$ (of size $n \times n$) to solve the problem.

The predictions for a new matrix of test examples \mathbf{X}_* is

$$\hat{\mathbf{Y}}_* = \Phi_*\mathbf{W} = \Phi_*\Phi^\top\alpha = \mathbf{K}_*\alpha, \quad (2.25)$$

where \mathbf{K}_* stands for the similarities between all test and training samples.

In conclusion, the KRR can be seen as a linear regularized regression over a features space possibly infinite. The main drawbacks are related to the kernel inversion which can be computationally costly since usually $n \gg d$ and the test retrievals can be slow since every test sample must be compared non-linearly with a train sample.

Gaussian process regression

Gaussian processes (GPs) are Bayesian state-of-the-art tools for discriminative machine learning, i.e., regression (Williams & Rasmussen, 1996), classification (Kuss & Rasmussen, 2005) and dimensionality reduction (Lawrence, 2005). GPs were first proposed in statistics by Tony O'Hagan (O'Hagan & Kingman, 1978) and they are well-known to the geostatistics community as *kriging*. However, due to their high computational complexity they did not become widely applied tools in machine learning until the early XXI century (Rasmussen & Williams, 2006). GPs can be actually interpreted as a family of kernel methods with the additional advantage of providing a full conditional statistical description for the predicted variable, which can be primarily used to establish confidence intervals and to set hyperparameters. In a nutshell, Gaussian processes assume that a Gaussian process prior governs the set of possible latent functions (which are unobserved), and the likelihood (of the latent function) and observations shape this prior to produce posterior probabilistic estimates. Consequently, the joint distribution of training and test data is a multidimensional Gaussian and the predicted distribution is estimated by conditioning on the training data.

Standard regression approximates observations (often referred to as *outputs*) $\mathbf{y} \equiv \{y_n\}_{n=1}^N$ as the sum of some unknown latent function $f(\mathbf{x})$ of the inputs $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$ plus *constant power (homoscedastic)* Gaussian noise, i.e.

$$y_n = f(\mathbf{x}_n) + \varepsilon_n, \quad \varepsilon_n \sim \mathcal{N}(0, \sigma^2). \quad (2.26)$$

Instead of proposing a parametric form for $f(\mathbf{x})$ and learning its parameters in order to fit observed data well, GP regression proceeds in a Bayesian, non-parametric way. A zero mean¹ GP prior is placed on the latent function $f(\mathbf{x})$ and a Gaussian prior is used for each latent noise term ε_n , $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k_\theta(\mathbf{x}, \mathbf{x}'))$, where $k_\theta(\mathbf{x}, \mathbf{x}')$ is a covariance function parametrized by θ and σ^2 is a hyperparameter that specifies the noise power. Essentially, a GP is a stochastic process whose marginals are distributed as a multivariate Gaussian. In particular, given the priors \mathcal{GP} , samples drawn from $f(\mathbf{x})$ at the set of locations $\{\mathbf{x}_n\}_{n=1}^N$ follow a joint multivariate Gaussian with zero mean and covariance matrix \mathbf{K}_{ff} with $[\mathbf{K}_{\text{ff}}]_{ij} = k_\theta(\mathbf{x}_i, \mathbf{x}_j)$.

¹It is customary to subtract the sample mean to data $\{y_n\}_{n=1}^N$, and then to assume a zero mean model.

If we consider a test location \mathbf{x}_* with corresponding output y_* , priors \mathcal{GP} induce a prior distribution between the observations \mathbf{y} and y_* . Collecting available data in $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \dots, N\}$, it is possible to analytically compute the posterior distribution over the unknown output y_* :

$$p(y_* | \mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_* | \mu_{\text{GP}*}, \sigma_{\text{GP}*}^2) \quad (2.27)$$

$$\mu_{\text{GP}*} = \mathbf{k}_{\text{f}*}^\top (\mathbf{K}_{\text{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} = \mathbf{k}_{\text{f}*}^\top \boldsymbol{\alpha} \quad (2.28)$$

$$\sigma_{\text{GP}*}^2 = \sigma^2 + k_{**} - \mathbf{k}_{\text{f}*}^\top (\mathbf{K}_{\text{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_{\text{f}*}. \quad (2.29)$$

which is computable in $\mathcal{O}(n^3)$ time (this cost arises from the inversion of the $n \times n$ matrix $(\mathbf{K}_{\text{ff}} + \sigma^2 \mathbf{I})$, see (Rasmussen & Williams, 2006).

The GPR (Gaussian Process Regression) model not only offers pointwise estimations, $\mu_{\text{GP}*}$ but confidence estimates, $\sigma_{\text{GP}*}^2$, since it finally offers a full posterior probability, $p(y_* | \mathbf{x}_*, \mathcal{D})$. Typically, one is only interested in doing predictions for unseen data. Therefore, in practice, the mean prediction model establishes a relation between the input (e.g., spectral data) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (e.g., biophysical variable) $y \in \mathbb{R}$ of the form:

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^N \alpha_i k_\theta(\mathbf{x}_i, \mathbf{x}) + \alpha_o, \quad (2.30)$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the spectra used in the training phase, α_i is the weight assigned to each one of them, α_o is the bias in the regression function, and k is a function evaluating the similarity between the test spectrum and all N training spectra. Therefore, in order to generate a GP model, one needs to specify a covariance/kernel function k_θ , to infer its hyperparameters θ and model weights α_i .

The core of any kernel method in general, and of GPs in particular, is the appropriate definition of the covariance (or kernel) function. A standard, widely used covariance function is the squared exponential (SE) kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2)),$$

which captures sample similarity well in most of the (unstructured) problems, and only one hyperparameter σ needs to be tuned. Several works (Verrelst et al., 2012; Campos-Taberner et al., 2016b) proposed the use of the so-called automatic relevance determination (ARD) kernel, as an alternative generalization of the isotropic SE prior.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{b=1}^B \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2}\right) + \sigma_n^2 \delta_{\mathbf{x}\mathbf{x}'}, \quad (2.31)$$

where ν is a scaling factor, B is the number of bands, and σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b . Model parameters (ν, σ_b) and model weights α_i can be automatically optimized by maximizing the marginal likelihood in the training set (Rasmussen & Williams, 2006). The obtained weights α_i after optimization gives the relevance of each spectrum x_i , while the inverse of σ_b represents the

relevance of each band b . Hence, low values of σ_b indicate a higher informative content of this certain band b to the training function k .

The corresponding hyperparameters $\{\theta, \sigma_n\}$ are typically selected by Type-II Maximum Likelihood, using the marginal likelihood (also called evidence) of the observations, which is also analytical (explicitly conditioning on θ and σ_n):

$$\log p(\mathbf{y}|\theta, \sigma_n) = \log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{ff}} + \sigma_n^2 \mathbf{I}). \quad (2.32)$$

When the derivatives of Eq. (2.32) are also analytical, which is often the case, conjugated gradient ascend is typically used for optimization. Therefore, the whole procedure for learning a GP model only depends on a very small set of hyper-parameters that combats overfitting efficiently.

Inference of the hyper-parameters and the weights for doing predictions, α , can be performed using this continuous optimization of the evidence.

2.3.3 Implementation of the algorithms

These machine learning algorithms have been integrated in the *SimpleR* MATLAB toolbox which is freely available at the Image Processing Laboratory website (<http://isp.uv.es/>). The toolbox is intended for practitioners with little expertise in machine learning, and that may want to assess advanced methods in their problems easily. The toolbox compares numerically and statistically the algorithms by simply entering the input-output (e.g. reflectances and LAI values) data matrices. The toolbox was used to train the aforementioned inversion models with the PROSAIL databases generated during this Thesis.

2.4 Developed processing chain at local scale

This section is dedicated to describe the developed processing chain at local scale. The processing chain was used in the framework of ERMES (<http://www.ermes-fp7space.eu/>) project where rice monitoring is performed exploiting seasonal remote sensing data (i.e. LAI estimates provided by the developed chain), and crop modeling as a demonstration of potential operation system. The processing chain was initially set up and preliminary tested during 2014 (Campos-Taberner et al., 2015a), and fully applied and validated during the 2015 and 2016 ERMES activities (Campos-Taberner et al., 2016a,b, 2017a).

The processing chain is based on a hybrid method through the generation of a database composed of RTM simulations (e.g. reflectances + LAI) to train machine learning regression methods. Once the methods are trained, the best method in terms of accuracy, bias and goodness-of-fit is selected. Then this method is used for LAI retrieval over rice areas using decametric remote sensing data (Landsat-7/8, SPOT-5 and Sentinel-2A). Figure 2.6 shows the followed workflow of the processing chain at local scale presented in this Thesis. It is worth mentioning that the processing chain is modular and can be adapted to other sensors and sites. As a matter of fact, the chain was adapted with the aim of deriving global scale estimates in Chapter 4.

2.4.1 Remote sensing data and sites

At local scale, the developed processing chain was tested to retrieve LAI during the 2015 and 2016 European rice seasons over local rice areas in Spain, Italy and Greece, which are

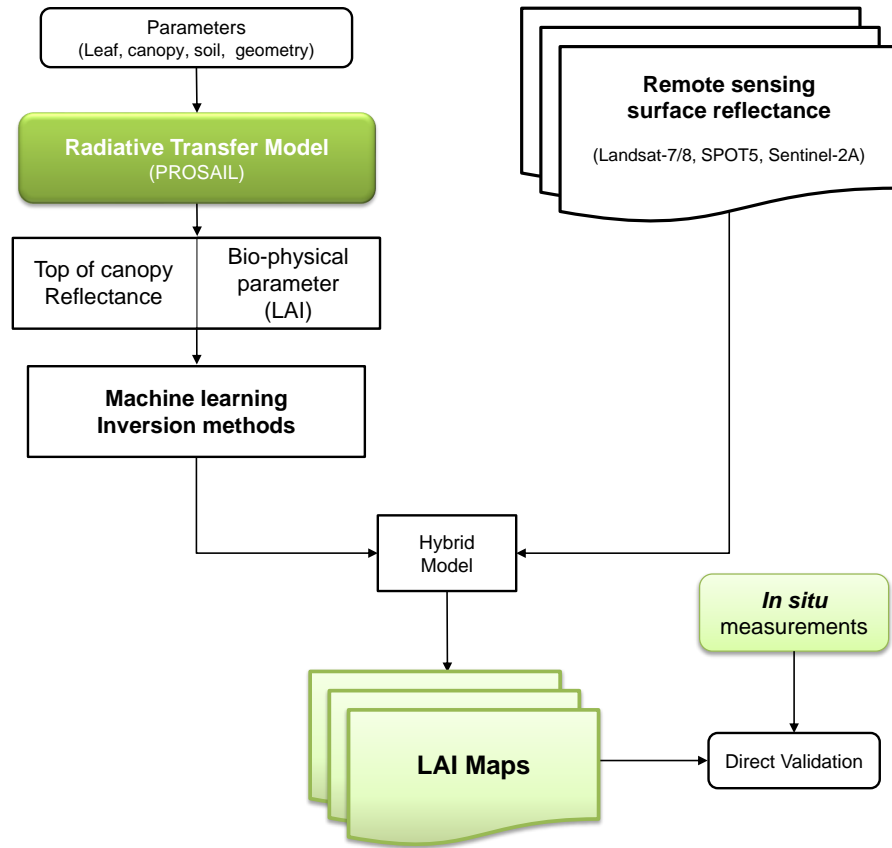


Figure 2.6: Developed processing chain at local scale.

responsible of 85% of total European rice production: Italy (51.9%), Spain (25.4%) and Greece (7.0%) (see Fig. 2.7). The Spanish site is located in the rice district of València (East of Spain) belonging to the *Albufera Natural Park* which is a special protection area thus allowing only rice crop practices. The Italian site belongs to the Lomellina rice district (south-western Lombardy region) where rice is the dominant crop (> 90%). The Greek site is located in the rice district of Thessaloniki which is the main cultivation area for Greece.

In 2015 the remote sensing data inputs of the processing chain were Landsat-7/8 and SPOT-5. The provisional Landsat-8 Surface Reflectance (LaSRC) and the Landsat-7 ETM+ (Enhanced Thematic Mapper) LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing System) products were used as inputs to derive Landsat-7/8 LAI estimates. Images were available every 16 day in Italy and Greece. On the other hand, since the Spanish rice area lies in two Landsat paths within the same row the temporal resolution of the images increased up to seven and nine days. SPOT5 provided 10 m pixel resolution data over the Spanish and Italian local rice sites every 5 days under constant angles from end-April to early-September 2015, thus covering the majority of the vegetation phase for summer crops in Europe. SPOT5 data provided a multispectral top of canopy surface reflectance product (green, red, near infrared and short wave infrared channels).

In 2016 the retrieval chain was run using Landsat-7/8 and Sentinel-2A images. Sentinel-2A Level 1C data (top-of-atmosphere reflectances in cartographic geometry) were used

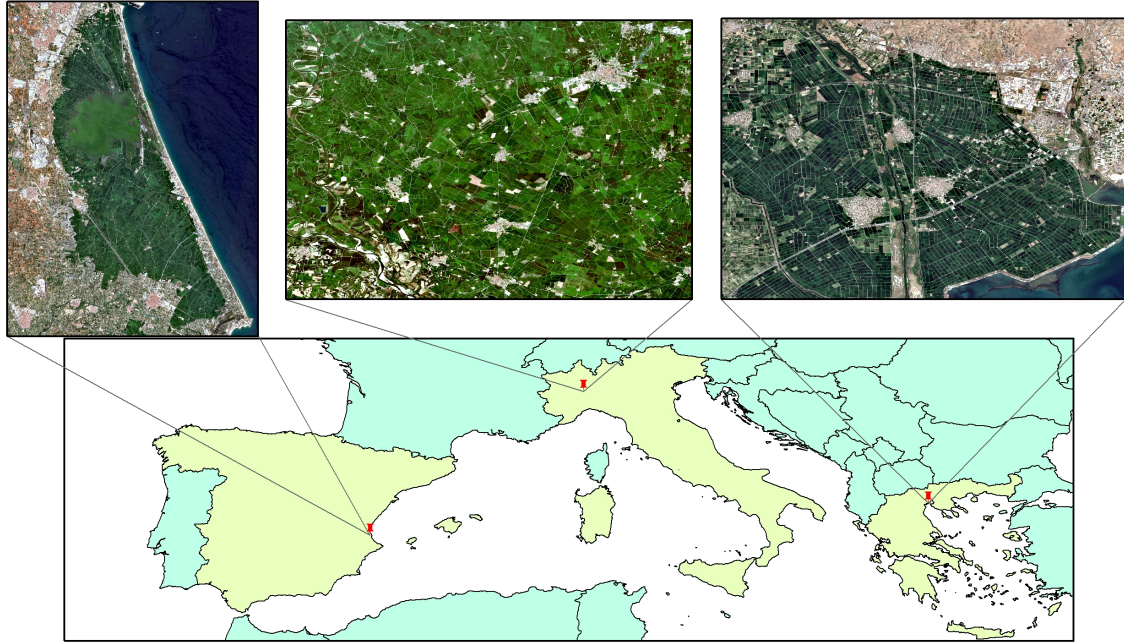


Figure 2.7: Study areas. Sentinel-2A surface reflectance RGB composite of the Spanish (30 July 2016), Italian (21 July 2016) and Greek (23 July 2016) rice local sites.

Table 2.1: Distribution of the canopy, leaf and soil parameters within the PROSAIL RTM.

	Parameter	Min	Max	Mode	Std	Type
Canopy	LAI (m^2/m^2)	0	10	3.5	4.5	Gaussian
	ALA ($^\circ$)	30	80	60	20	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.5	1	1	0.2	Truncated Gaussian
Leaf	N	1.2	2.2	1.5	0.3	Gaussian
	C_{ab} ($\mu\text{g}\cdot\text{cm}^{-2}$)	20	90	45	30	Gaussian
	C_{dm} ($\text{g}\cdot\text{cm}^{-2}$)	0.003	0.011	0.005	0.005	Gaussian
	C_{wREL}	0.6	0.8	-	-	Uniform
Soil	β_s	0.3	1.2	0.9	0.25	Gaussian

providing images over the three selected sites every 10 days in 13 bands in the visible, near infra-red and short wave infra-red spectrum at 10, 20 and 60 m spatial resolution depending on the spectral band.

2.4.2 Theoretical performances

This section is devoted to highlight the theoretical performances of NN, KRR and GPR in a LAI retrieval experiment. Firstly, a database of 2000 data pairs of multispectral reflectances (Landsat-8 and SPOT-5) and corresponding LAI were simulated running the PROSAIL model in forward mode. The leaf and canopy variables as well as the soil brightness parameter were randomly generated following specific distributions (see Table 2.1). A site-specific parameterization of the PROSAIL model based on the available 2014 ERMES field measurements was selected in order to constrain the behavior of the model to Mediterranean rice areas reducing the equifinality of the ill-posed PROSAIL inversion process (Combal

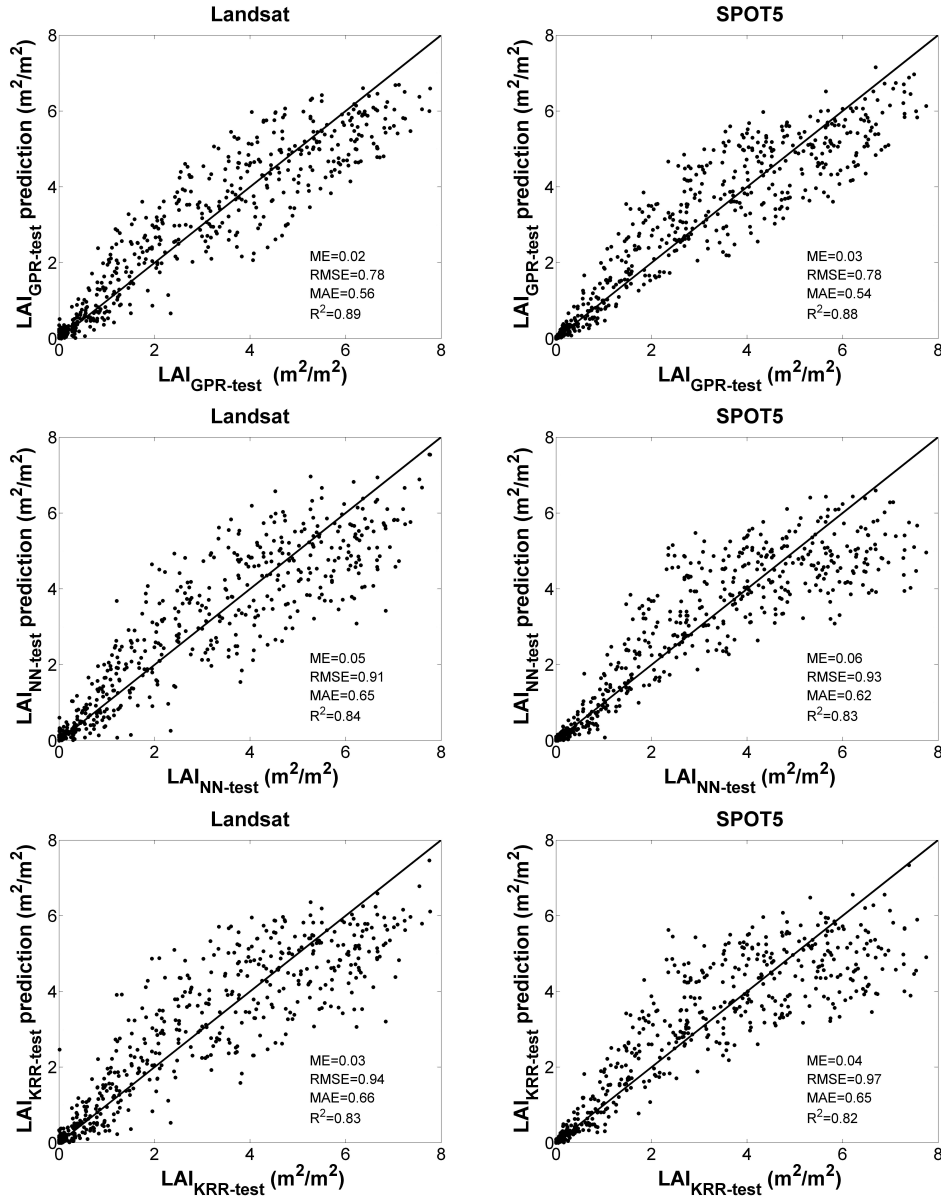


Figure 2.8: GPR (top), NN (middle) and KRR (bottom) theoretical performances using Landsat 8 OLI (left) and SPOT5 (right) simulated data. Plots are obtained by applying the trained models on test data.

et al., 2003). During the 2014 ERMES rice season, leaf chlorophyll content was measured with a SPAD-502 chlorophyll meter (Campos-Taberner et al., 2016a) thus allowing to constrain the C_{ab} range to Mediterranean rice values which typically vary from 35-70 $\mu g \cdot cm^{-2}$ depending on the rice variety. This range was slightly extended in order to represent rice with high leaf chlorophyll concentration due to fertilization practices, as well as low leaf chlorophyll content caused either by possible diseases, blasts or nitrogen deficits. ALA distribution was selected for accounting specific leaf inclination during rice phenological stages (Zhang et al., 2013).

In order to better constrain the retrieval to rice crops, a spectral library of underlying soil background was generated by considering signatures of homogeneous flooded and dry

areas identified within rice fields in the study areas with Landsat-8 and SPOT5 imagery, which was spectrally matched with typical rice background spectra collected by [Boschetti et al. \(2014\)](#). Each background signature was selected randomly from the spectral library and multiplied by the soil brightness parameter, which was assumed to follow a Gaussian distribution (see Table 2.1). For each sensor, a single dataset of PROSAIL simulations was performed, which included all geometrical configurations. The distributions for the system geometry were randomly generated based on information in imagery metadata.

Then, the simulated database was split into a training (70% of the samples) and a testing set (the remaining 30%), whereby 1400 samples were used for training and 600 for testing. This split allowed to mitigate overfitting since every model is tested over unseen data not used during the training phase. Hereafter, we refer to GPR-Landsat and GPR-SPOT5 to the GPR models built using either Landsat or SPOT5 data.

The performance of the all three models is evaluated in Fig. 2.8 over the test dataset. Results revealed good accuracy and low bias in both Landsat-8 and SPOT-5 simulated reflectances. GPR outperformed NN and KRR in all statistical quality measures (see Fig. 2.8). GPR-Landsat and GPR-SPOT5 theoretical performances were robust and very close, revealing biases of 0.02 and 0.03 m^2/m^2 respectively, and a Root Mean Squared Error (RMSE) of 0.78 m^2/m^2 in both GPR-Landsat and GPR-SPOT5 models. High determination coefficients ($R^2 > 0.87$) were obtained in both cases.

2.4.3 High-resolution LAI maps

The application of the chain allowed to obtain multitemporal high-resolution LAI maps over the three rice areas with a spatial resolution of 30 m in the case of using Landsat-7/8 imagery, and 10 m in the case of SPOT5 and Sentinel-2A data. Figure 2.9 shows the GPR-Landsat and GPR-SPOT5 LAI maps derived for the very beginning of season (mid-May), early growing season (early-June), and maximum leaf rice development (mid-August) in the Spanish site. The corresponding HR estimated LAI maps for the Italian study area are shown in Figure 2.10.

First inspection of the maps indicates the occurrence of very low LAI values corresponding to mid-May since the sowing dates were around May, 10-15th in Spain. In Italy, for the same period, some rice fields have higher LAI values (>2) because of the early sowing of some rice varieties. In the Spanish site, it can be seen the expected rice emergence in the early-June LAI maps (see Fig. 2.9 (middle)), while the early-July Italian maps (Fig. 2.10 (middle)) show already higher LAI estimates due to the advanced phenological growing state corresponding to the rice stem elongation phase. Eventually, the mid-August maps (Fig. 2.9 and Fig. 2.10 right panels) show the highest LAI estimates because rice plants reached the heading phase and the LAI seasonal peak.

Similarly to 2015, high-resolution (10 m) LAI retrievals were obtained over the three rice areas during the 2016 rice season applying the trained model to Sentinel-2A data. Figure 2.11 shows the LAI map derived over the Spanish study area using a Sentinel-2A image acquired on August 9, 2016. It can be seen the high spatial detail of estimates which allows to identify significant different values within the same rice field. Those intra-field LAI differences are mainly due to the heterogeneity of the rice field caused by non homogenous seeding and agro-practices ([Ranghetti et al., 2016](#); [Campos-Taberner et al., 2016b](#)).

The spatial consistency between Sentinel-2A and Landsat-8 LAI estimates was assessed

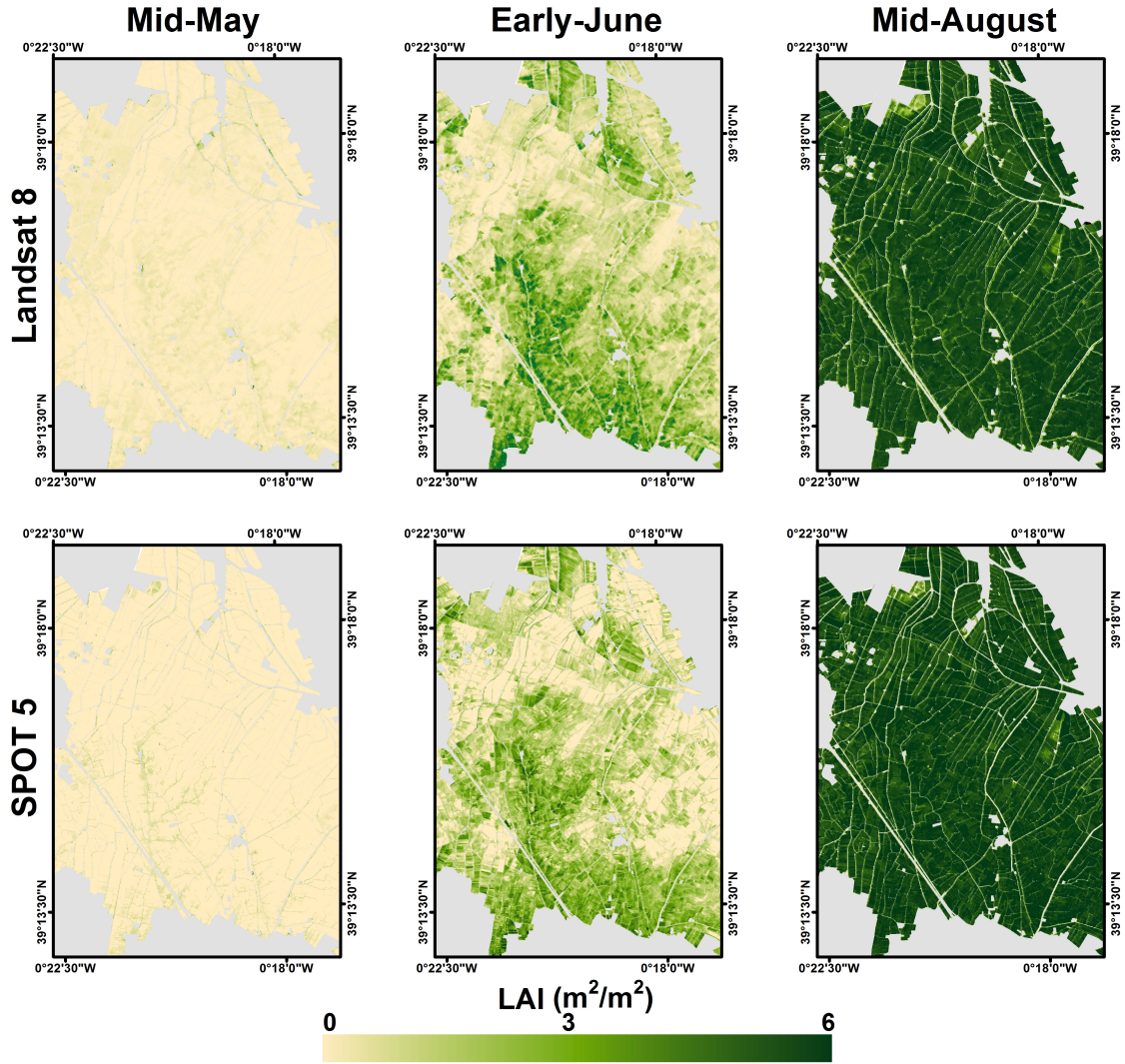


Figure 2.9: Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Spanish study area in mid-May (left), early-June (middle) and mid-August (right). Grey mask covers non-rice areas.

computing the difference map of the pair of the closest acquisition dates in 2016 between Landsat-8 and Sentinel-2A over Spain, Italy and Greece (see Fig. 2.12). For this purpose, the Sentinel-2A images were resampled to Landsat-8 resolution and the comparison was achieved averaging the LAI value computed over a window of 3×3 valid pixels in order to reduce coregistration errors between images (Storey et al., 2016) and inconsistencies associated to differences in the point spread functions (Morissette et al., 2006). Good correlations ($R^2 > 0.82$) were found in all three areas and a notable agreement was found in Italy with practically no bias and low RMSE (see Fig. 2.12). Leaf area index difference maps show the predominance of the light green color which demonstrates the retrievals from the two sensors are coherent exhibiting LAI differences around zero over the majority of the rice areas. Higher differences are mainly due to differences in surface reflectances from Landsat-8 and Sentinel-2A generated by the different atmospheric correction methods. In the case of Spain (Fig. 2.12 up-left), a surrounding halo effect can be discerned. This

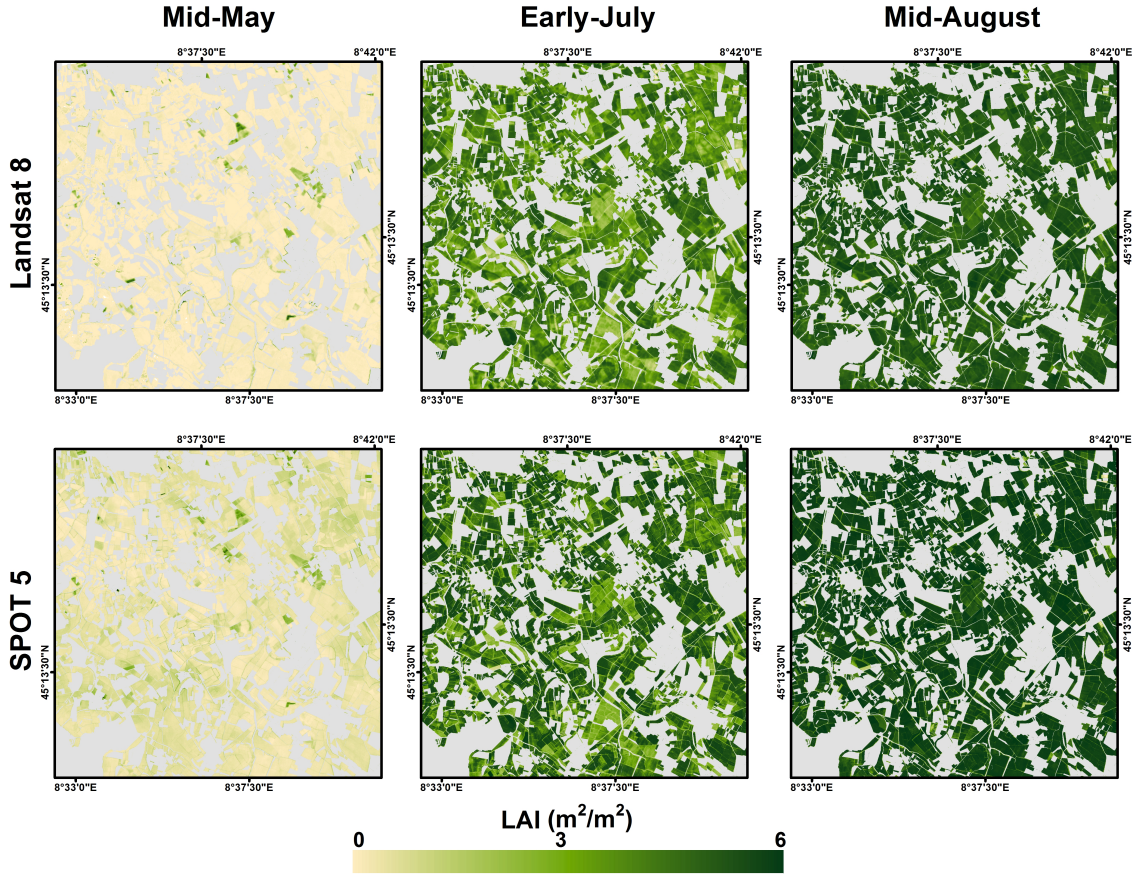


Figure 2.10: Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Italian study area in mid-May (left), early-July (middle) and mid-August (right). Grey mask covers non-rice areas.

effect is due to a suboptimal Landsat-8 atmospheric correction performance over the edge region between land and sea and could explain the wider dispersion observed in the scatter plot (see Fig. 2.12 bottom-left).

All derived high-resolution maps can be investigated by registering and joining as local user into the ERMES web-based geo-portal (<http://ermes.dlsi.uji.es/prototype/geoportal/>) or can be found in the ERMES catalogs (<http://get-it.ermes-fp7space.eu/>).

2.4.4 Influence of the background spectra and utility of GPR uncertainty

In this Thesis, the influence of the background reflectance on LAI estimations was investigated. The influence of background characteristics in LAI estimates was assessed introducing three different types of background spectral signatures in the PROSAIL database as follows:

- Primarily, it was trained and inverted the PROSAIL model using typical spectra of dry bare soils which were present at the very beginning of the rice season.
- Secondly, it as used a set of flooded soil signatures representing typical background conditions during the rice growing (from sowing to harvest).
- Eventually, we used both dry and flooded spectra.

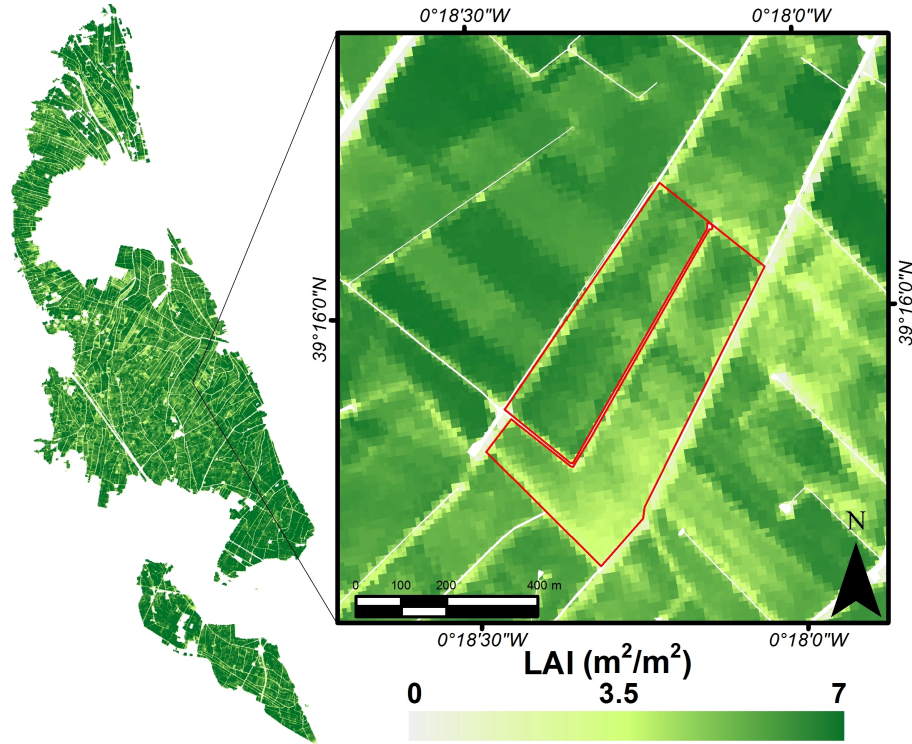


Figure 2.11: High-resolution LAI map obtained with Sentinel-2A data acquired in the Spanish study area on August 9, 2016. Right panel provides a zoom over a monitored field in the rice season.

Table 2.2: Statistical scores between the Landsat estimated LAI taking into account different background conditions and ground measurements over the Valencia rice district.

	ME	RMSE	MAE	R ²
Dry	-0.50	1.09	0.85	0.81
Flooded	0.10	0.62	0.40	0.90
Dry + Flooded	0.08	0.39	0.28	0.95

Note that during these experiments the distributions of PROSAIL parameters for the crop were kept constant (see Table 2.1), and only the set of background spectra was changed (i.e. dry, flooded and both dry + flooded conditions). Table 2.2 exhibits the statistics when estimates from Landsat-8 over the Spanish site are compared with *in situ* LAI measurements. The best results were obtained when a background spectral library composed by flooded and dry soils was used. Conversely, when considering only flooded or dry backgrounds high errors were found during the initial development stages of the plants (during tillering and before panicle formation), as the reflectance of the background constitutes a significant component of the overall spectral signature recorded by the sensor.

In addition, temporal behavior of LAI estimates was also evaluated. Figure 2.13 shows a typical example of a LAI evolution retrieved from SPOT5 data over a rice pixel in Italian and Spanish rice fields. They reflect a common situation in which the soil was dry until day of year (DOY) 132 (Italian area) and DOY 130 (Spanish area), and flooded immediately before sowing and during plant development. The figure compares the retrievals taking into

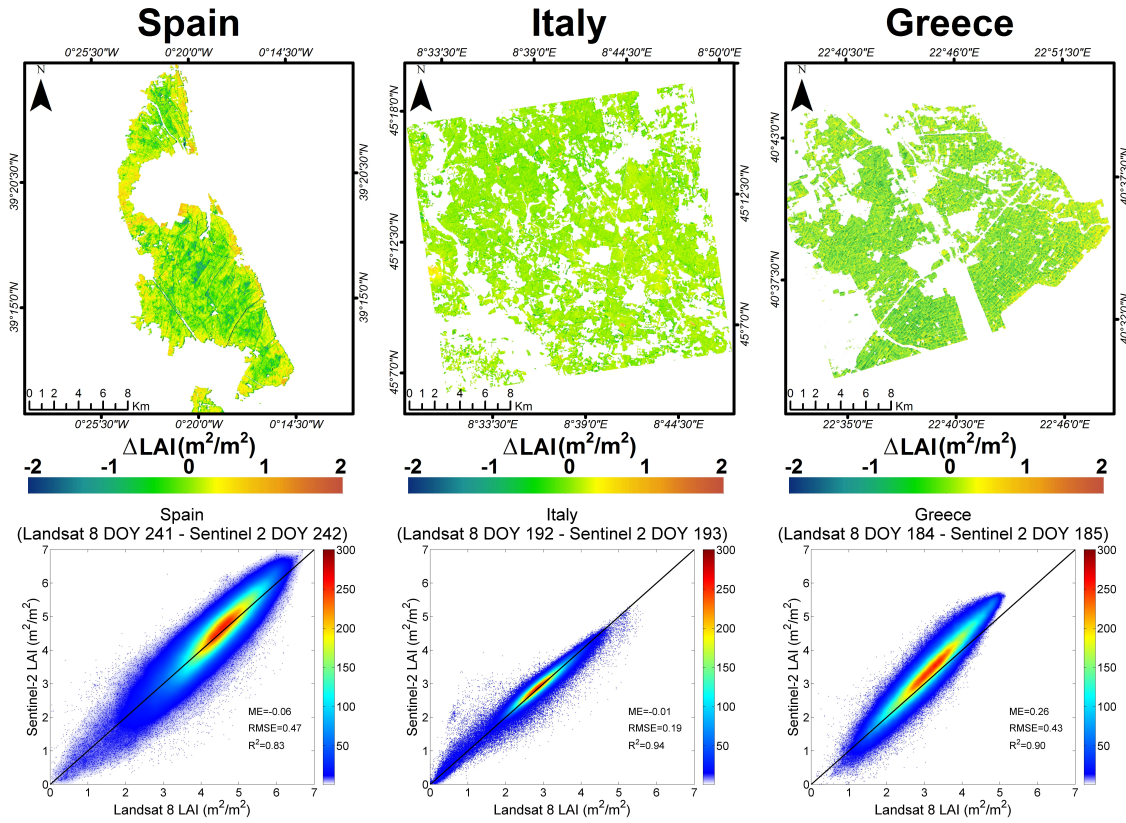


Figure 2.12: LAI difference maps (up) and the corresponding scatter plots of the closest Sentinel-2A and Landsat-8 LAI estimates (bottom) over Spain (left), Italy (middle) and Greece (right).

account both dry and flooded conditions (red line), and the influence of using only dry soils (yellow line) or flooded soils (blue line). It can be observed that the use of a complete and representative spectral database (flooded + dry background) in the PROSAIL simulation provides realistic LAI evolutions (red line) which agree well with field measurements. It should be noted that in this situation LAI remains close to 0 before the rice emergence (from DOY 100 to 150 approx.), irrespectively of the flooded/non-flooded condition of the soil background. This assessment demonstrates the importance of training correctly the RTM to produce a reliable data set for the retrieval process.

The GPR model not only provides a mean estimate for the predictions but also a prediction uncertainty related to the confidence of the estimates (the lower uncertainty the higher the confidence). Although this uncertainty is a numerical value, it must be interpreted as a quality-driven indicator (Verrelst et al., 2013; Campos-Taberner et al., 2017a) related, in the case of rice, to (i) an incorrect simulation of rice background which has a high impact on the reliability of LAI retrievals in rice crops (see Fig. 2.13), and (ii) to spectral data under-represented in the PROSAIL training database such as water bodies and man-made surfaces which its associated LAI estimates presented significant higher prediction uncertainty (lower confidence). An experiment conducted simulating Sentinel-2A reflectances with PROSAIL and retrieving 2016 multitemporal LAI estimates with the GPR over rice areas, revealed unexpected estimates (pronounced drops) during the season. These LAI estimates were related to suboptimal atmospheric correction or undetected

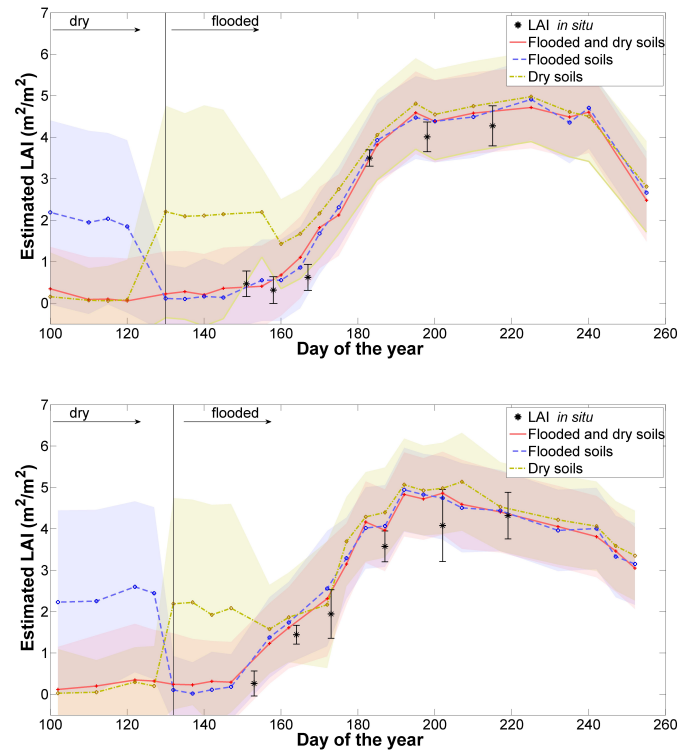


Figure 2.13: Temporal evolution of LAI estimates from SPOT5 data within a rice pixel using different soil background in (top) Spain and (bottom) Italy.

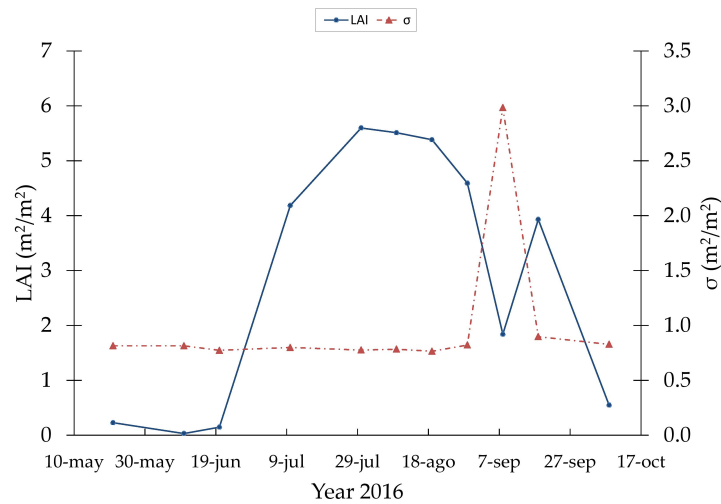


Figure 2.14: Temporal evolution of LAI and uncertainty (σ) over a rice field in Spain during the 2016 rice season.

clouds in Sentinel-2A data which were labeled as poor-quality estimates. Figure 2.14 shows the temporal evolution of Sentinel-2A LAI and uncertainty within a rice field where the unexpected LAI drop reported on 7 September and the associated σ peak were related to a cloudy pixel undetected by the official atmospheric correction tool provided by ESA.

2.5 Concluding remarks

In this chapter we detailed the main components of the developed processing chain at local scale. Firstly, we reviewed the PROSPECT and SAIL RTMs formulation paying attention to the leaf, canopy, soil background and system geometry parameters as well as illustrated the PROSAIL coupling. It is worth mentioning that a specific PROSAIL parametrization was used in order to constrain the LAI retrievals to mediterranean rice cropping systems alleviating thus the RTM ill-posed problem.

In addition, we described the machine learning methods used for the PROSAIL inversion: NN, KRR and GPR. We have assessed the theoretical performance of the inversion methods in an experiment conducted in 2015 where the local processing chain was tested using Landsat-8 and SPOT5 data to retrieve LAI estimates over rice areas in Spain, Italy and Greece. GPR proved to be a highly efficient and accurate method to invert the PROSAIL model, outperforming NN and KRR. However, the flexibility of the processing chain allows its application to any other regression method of choice.

Eventually, we illustrated the derived HR LAI maps which captured the rice LAI spatial distribution of the local areas as well as the rice phenological growing states. In addition, an assessment regarding the impact of simulated background spectra in multitemporal rice LAI retrievals is presented, and the practical usefulness of the GPR provided uncertainty is also highlighted. The validation of the retrievals is addressed in the next chapter, where the *in situ* LAI measurements are described in detail.



3. *In situ* measurements

Contents

3.1 Introduction to field measurements

3.2 Classical instruments

3.2.1 LICOR-LAI2000

3.2.2 Digital hemispherical photography (DHP)

3.3 New technologies in geosciences: PocketLAI

3.4 Validation of the processing chain at local scale

3.4.1 Field campaigns

3.4.2 Accuracy assessment over ground LAI

3.4.3 Temporal evolution of LAI estimates

3.5 Concluding remarks

This Chapter is partly based on the following publications:

- **Campos-Taberner, M.**, Garcia-Haro, F. J., Moreno, A., Gilabert, M. A., Sanchez-Ruiz, S., Martinez, B. & Camps-Valls, G., “Mapping Leaf Area Index With a Smartphone and Gaussian Processes”. *Geoscience and Remote Sensing Letters, IEEE*, 12(12), 2501-2505, 2015.
- **Campos-Taberner, M.**, García-Haro, F. J., Confalonieri, R., Martínez, B., Moreno, Á., Sánchez-Ruiz, S., Gilabert, M.A., Camacho, F., Boschetti, M. & Busetto, L., “Multitemporal Monitoring of Plant Area Index in the Valencia Rice District with PocketLAI”. *Remote Sensing*, 8(3), 202, 2016.
- **Campos-Taberner, M.**, García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A. & Boschetti, M. (2017). “Multitemporal and Multiresolution Leaf Area Index Retrieval for Operational Local Rice Crop Monitoring”, 2016. *Remote Sensing of Environment*, 187, 102 - 118, 2016.
- **Campos-Taberner, M.**, García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Busetto, L., Katsantonis, D., Stavrakoudis, D., Minakou, C., Gatti, L., Barbieri, M., Holecz, F., Stroppiana D. & Boschetti, M., “Exploitation of SAR and optical Sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index”, *Remote Sensing*, 9(3), 248, 2017.

3.1 Introduction to field measurements

Biophysical parameters of the canopy can be estimated from remote sensing observations and can also be measured in the field through different *in situ* methods and techniques. For example, *in situ* LAI measurement methods can be divided in two main categories: direct and indirect (Breda, 2003; Weiss et al., 2004).

Direct methods require an effort in collecting an optimal samples size and estimating plant density which involve destructive harvest techniques (Confalonieri et al., 2006). Direct and indirect methods are complementary as direct LAI measurements can be used a reference or calibration for indirect measurements. Indirect methods allow to infer LAI from observations of another variable. They are generally faster than direct methods and allow larger spatial sample collection.

Indirect methods can be divided into indirect contact LAI measurements and indirect non-contact measurements (Jonckheere et al., 2004). Indirect contact LAI methods are based on the estimation of the contact frequency (Warren-Wilson, 1963) while indirect non-contact methods are based on the estimation of the gap fraction (Ross, 2012). Contact frequency is the probability that a beam (radiation) penetrates inside the canopy and interacts with the vegetation. On the other hand, gap fraction is the integrated value of the gap frequency which accounts for the probability that the beam have no contact with the vegetation until it reaches the ground. Sensors like LAI-2000 or LAI-2200 plant canopy analyzers (LI-COR, Inc., Nebraska, USA) measure the gap fraction from five different angles simultaneously. Digital hemispherical photography (DHP) is another indirect technique for computing the gap fraction through cameras with hemispherical lenses (fish-eye) coupled. This method estimates LAI from measurements of the gap fraction, defined as the fraction of sky seen from below the canopy (upwards photography) or fraction of soil seen from above (downwards photography). Both plant canopy analyzers and DHP are some of the most widely used classical optical instruments for indirect LAI estimation (Jonckheere et al., 2004; Garrigues et al., 2008b). Optical instruments used to measure LAI typically rely on the Beer law which does not addresses the clumping within the canopy, thus these measurements are closer to an *effective* LAI.

Smartphones are nowadays the most popular sensing instrument. The use of smartphone components such as global positioning system (GPS), camera, accelerometer, and core processing power makes them suitable for a number of purposes including remote sensing educational purposes (Campos-Taberner et al., 2017b). Recently, the increasing of the smartphones computational power and the implementation of high-quality sensors jointly with a low-cost of devices, led to the use of smartphones as possible tools for monitoring vegetation properties and geological field studies (Weng et al., 2012; Delgado et al., 2013). A smartphone application called PocketLAI was recently presented and tested for acquiring LAI measurements (Confalonieri et al., 2013).

In the following sections we describe the classical instrumentation used for LAI, FVC and FAPAR non-destructive measurements. In addition, the field campaigns conducted during this Thesis are also described, and eventually, the procedure followed for the validation of the retrieval chain at local scale is detailed.

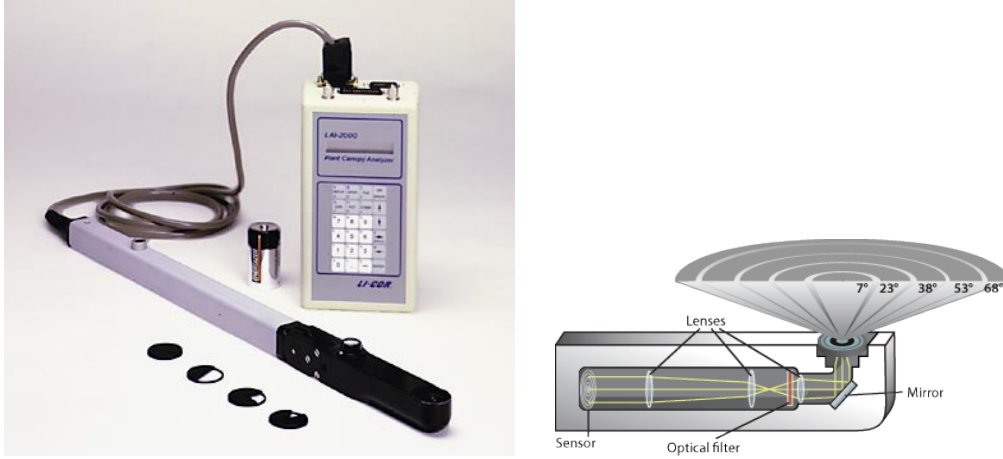


Figure 3.1: Left: LAI-2000, right: Diagram of rings in LAI-2000.

3.2 Classical instruments

3.2.1 LICOR-LAI2000

Li-Cor LAI-2000 estimates effective LAI (LAI_{eff}) using a gap fraction method that determines LAI_{eff} from measurements made above and below the canopy, which are used to determine canopy light interception at five zenith angles (7° , 23° , 38° , 53° and 68°) (see Fig. 3.1). Canopy transmission is measured between 320 and 490 nm. LAI-2000 computes the LAI_{eff} using the Miller's formula (Miller, 1967) as:

$$LAI_{eff} = 2 \int_0^{\pi/2} -\ln P(\theta) \cos(\theta) \sin(\theta) d\theta = 2 \sum_{i=1}^5 \bar{K}_i W_i, \quad (3.1)$$

where K_i and W_i are the contact number and the weighting factor and the subscript i refers to the number of the ring. $\bar{P}(\theta)$ is the average probability of light penetration into the canopy and the gap fraction for a ring i , $G_i(\theta)$, is computed as:

$$G_i(\theta) = \exp(\overline{\ln P(\theta)}) = \exp\left(\frac{1}{N} \sum_{j=1}^N \ln\left(\frac{B_j}{A_j}\right)\right). \quad (3.2)$$

The subscript j is the number of readings ($j = 1 \dots N$) and B_j and A_j are the corresponding readings to below and above the canopy, respectively. The typical LAI_{eff} provided by LAI-2000 is calculated on five rings. Additionally, the FVC can be obtained as the complement of the gap fraction measured by the first zenith angle of the LAI-2000 sensor, $1 - G_1(7^\circ)$ (White et al., 2000).

3.2.2 Digital hemispherical photography (DHP)

Digital hemispherical photography allows the computation of LAI measuring the gap fraction. The digital photographs can be taken in downward-looking or upward-looking. Although the upward photography is easy for gap identification, it may overestimate the gap fractions and consequently underestimate the LAI_{eff} in case of rice (Fang et al., 2014). Thus in the DHP acquisitions conducted during this Thesis, given this underestimation



Figure 3.2: NIKON Coolpix5000 camera and FC-E8 Nikon fisheye used in the digital hemispherical photography.

and the characteristics of the rice plant (low height) and soil conditions (flooded), the downward-looking was selected. In addition, the distance between the camera and the top of rice canopy was set to about 1 m in order to avoid that leaves were too much close to the lens.

The photographs can be processed using the Can-Eye software developed at the Mediterranean environment and agro-hydro system modelisation (EMMAH, *Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes*) laboratory in the French national institute of agricultural research (INRA, *Institut National de la Recherche Agronomique*). Can-Eye computes effective LAI from gap fractions after an interactive thresholding classification process separating rice foliage from the background (downward view). The images should be masked limiting to 60° zenith angle the valid range of fisheye lens to avoid edge distortions and ensuring that the area captured was within the area of interest. Hemispherical photographs are taken maintaining the camera in an approximately horizontal position not-exceeding 10° which is considered the threshold for minimizing source of errors due horizontal camera leveling in estimating LAI (Demarez et al., 2008). Photographs with suboptimal horizontal acquisition should be excluded for further processing.

During the field campaigns conducted in this Thesis, a NIKON Coolpix5000 camera was used at the finest image resolution available (JPEG format at 2560 × 1920 pixels resolution) and a FC-E8 Nikon fisheye lens with a field of view of 183° (see Fig. 3.2). The camera with the fisheye lens was previously calibrated according to the Can-Eye manual (Weiss & Baret, 2010). Similarly to the LAI-2000, LAI_{eff} can be computed using DHP through the Miller's formula as follows:

$$LAI_{eff} = 2 \int_0^{\pi/2} -\ln P_0(\theta) \cos(\theta) \sin(\theta) d\theta. \quad (3.3)$$

Can-Eye software processes simultaneously up to 16 images acquired over the same area of interest. All images belonging to the same area of interest should be acquired with similar illumination conditions to limit the variation of color dynamics between images.

Using DHP, it is not possible to get a value of FVC in the exact nadir direction, and the cover fraction must be integrated over a range of zenith angles. Can-Eye provides FVC estimates with a default value for this range of 0-10°. In addition, Can-Eye provides the

instantaneous “black sky” FAPAR in the solar zenith angle θ_S as

$$FAPAR^{BS}(\theta_S) = 1 - P_0(\theta_S), \quad (3.4)$$

the “white sky” FAPAR as

$$FAPAR^{WS} = 2 \int_0^{\pi/2} (1 - P_0(\theta)) \cos(\theta) \sin(\theta) d\theta, \quad (3.5)$$

and the daily integrated “black sky” FAPAR as

$$FAPAR_{int}^{BS} = \frac{\int_{t_1}^{t_2} \cos(\theta) (1 - P_0(\theta)) d\theta}{\int_{t_1}^{t_2} \cos(\theta) d\theta}, \quad (3.6)$$

where t_1 refers to sunrise and t_2 to sunset.

3.3 New technologies in geosciences: PocketLAI

Classical commercial instruments have proven to be a good alternative to destructive methods in many experimental conditions (Verger et al., 2009) but these instruments are usually quite expensive both to purchase and maintain. Due to their weight and strict requirements concerning acquisition protocol, they can also prove to be quite difficult to use in cases in which access to the canopy or placement of the instrument either below or above the canopy is difficult, such as in the case of flooded rice fields. Specifically, DHP techniques require high effort by the operator during the classification process needed to obtain a LAI estimate. This fact limits the deployment of these measurements in near real time applications. An additional drawback when using these instruments is the time needed for repairing them in case of damage. A delay in the repair or the replacement of the instrument may lead to canceling the field campaign in the worst case.

Smartphone capabilities are growing up day by day, making them a reliable alternative to classical measuring instruments. Specifically, in the context of exploiting the technology implemented in smartphones for studies dealing with natural sciences, a mobile application called PocketLAI was introduced for leaf area index estimation (Confalonieri et al., 2013).

In this Thesis, we propose the use of PocketLAI for *in situ* non-destructive rice LAI_{eff} monitoring. PocketLAI is a smartphone application based on the segmentation of images acquired at 57.5° below the canopy to estimate gap fraction (Confalonieri et al., 2013). It is based on a simplified model of light transmittance under the assumption of a random spatial distribution of infinitely small leaves. In this case, the gap fraction $P_0(\theta_v, \phi_v)$ in the zenith angle θ_v direction and azimuth angle ϕ_v is given by

$$P_0(\theta_v, \phi_v) = \exp \left(-G(\theta_v, \phi_v) \frac{LAI}{\cos \theta_v} \right), \quad (3.7)$$

where $G(\theta_v, \phi_v)$ is the *projection function*, which can be considered as almost independent of *leaf inclination* ($G \cong 0.5$) for a viewing angle of 57.5° ($\cong 1$ radian) (Weiss et al., 2004).



Figure 3.3: PocketLAI screenshot (up) and measurement scheme (bottom).

Inverting Eq. 3.7, LAI can be estimated from the gap fraction at this particular direction as follows:

$$LAI(57.5^\circ) = \exp\left(\frac{-0.5LAI}{\cos(57.5^\circ)}\right) = \frac{-\ln P_0(57.5^\circ)}{0.93}. \quad (3.8)$$

At this directional configuration, the information acquired is independent from leaf angle distribution and minimizes leaf clumping effect (Baret et al., 2010). This property comes from the *projection function* used to compute gap fraction corresponding to the projection of a LAI unit into a given direction.

PocketLAI computes LAI using an inclinometer derived by the device's accelerometer and the camera, to obtain an estimate of the gap fraction at 57.5°. The gap fraction estimate is acquired automatically while rotating the device along its main axis starting with the display orientated downward and concluding with the display in vertical position (or vice versa).

Once the app is set in measuring-mode, the user has five seconds to place the device below the canopy; at that instant the app triggers a vibration and the orientation of the screen surface is continuously recomputed, provided in real time by the accelerometer sensor, by using plain vector algebra. When the angle between the vertical and the normal to the screen reaches 57.5°, a camera frame is captured and transferred to the processing algorithms while the app issues a second vibration event to inform the user of the occurred successful acquisition. Figure 3.3 shows a screenshot of the PocketLAI interface.

Useful Images are automatically processed using an algorithm developed to detect sky

pixels, based on a segmentation strategy in which pixels are classified according to their chromatic values. The algorithm uses two different segmentations for cloudy and clear sky conditions: the first (cloudy sky) is simply based on the pixels intensity and is based on the assumptions that leaves appear darker than the sky in images acquired below the canopy. The second segmentation strategy (clear sky, with high levels of direct radiation) is based on the analysis of pixels chromatic values in an HSB (Hue Saturation Brightness) colour space and is able to detect sky and clouds even from the parts of vegetation that can appear lighter and brighter than the sky. This is possible through the identification of the residual green chromatic component persisting even in the most reflective part of the plants.

The application lets the user to select the total number of LAI measurements to be done that will be used to compute an averaged LAI value. In addition, the user can specify a code for each measurement. Moreover, the measurements can be exported to the most common geographical information systems formats to be loaded, for example, in Google Earth.

This mobile application has shown to perform well in canopies with different structures (Francone et al., 2014) and was proved its reliability in terms of both trueness and precision (Confalonieri et al., 2013). PocketLAI was successfully tested against LAI-2000 and DHP for *in situ* LAI acquisition (Campos-Taberner et al., 2015c) and LAI mapping (Campos-Taberner et al., 2015b) on paddy rice during the 2014 ERMES field activities. In the following sections we describe these tests and also the validation of the retrieval chain at local scale in 2015 and 2016 through LAI measurements acquired with PocketLAI.

3.4 Validation of the processing chain at local scale

As previously explained in Chapter 2.4, the retrieval chain was applied at local scale in the framework of ERMES during the 2015 and 2016 rice seasons. Therefore, the validation of the chain was conducted using the *In situ* measurements acquired during field campaigns in this period. Hereunder, we describe the field campaigns and show the results of the validation of LAI satellite-derived estimates.

3.4.1 Field campaigns

Field campaigns were conducted in Spain and Italy during the rice growing period in 2015 and 2016, while in Greece, field campaigns were conducted in 2016. Ground measurements were acquired within ESUs (Elementary Sampling Units). An ESU is a contiguous spatial region over which the expected value of LAI can be estimated through *in situ* measurement. In 2015, the total number of sampled ESUs in Spain was 40. For some of the ESUs, field acquisitions were made in all field campaigns, whereas in other ESUs the temporal frequency was approximately one every two field campaigns. In the Italian study area, 10 ESUs were fixed and considered during all the field campaigns, increasing the number of ESUs up to 19 for some dates. For its part, in 2016 field campaigns, 32, 16 and 10 ESUs were selected in Spain, Italy and Greece, respectively. The temporal frequency of the campaigns was approximately 7-10 day starting from the very beginning of rice emergence (early-June) up to the maximum green rice LAI development (mid-August). The field campaigns were planned in such a way as LAI measurements within every ESU were either temporally coincident or 1-2 day apart with respect to the satellite images.

Based on land cover distribution in the areas and information provided by farmers at the

very beginning of the rice season, a reliable sampling was achieved selecting ESUs with different rice varieties and sowing dates in order to cover as much as possible the variability of the study areas. The ESUs were selected within homogeneous rice fields. The same sampling scheme was used over each ESU, following the guidelines and recommendations of the Validation of Land European Remote sensing Instruments (VALERI) protocol. For the case of row crops, this protocol suggests to take measurements along small transects between rows incorporating some random acquisitions to prevent possible biases in the characterization of the row effect. We adopted the same schema also for fields with broadcast seeding.

The size of the ESUs was approximately $20\text{ m} \times 20\text{ m}$, and the locations were far from the field borders. In order to characterize the spatial variability within each ESU, a range of 18 to 24 measurements was taken. This number of replicates allows to obtain a statistically significant mean LAI estimate per ESU. The center of the ESU was geo-located using a GPS for later matching and associate the mean LAI estimate with the corresponding satellite spectra. *In situ* LAI estimates were acquired in all the years and rice sites with smartphones using the PocketLAI application. The mobile application was installed on Samsung Galaxy S4 GT-I9505, with a Quad-Core 1.9 GHz processor and 2 GB RAM. The smartphone's camera allowed PocketLAI to take images with a resolution of 4128×3096 pixels.

It is worth mentioning that during this Thesis, a comparison study between PocketLAI and classical instrumentation estimates was conducted in a previous year, in order to assess PocketLAI consistency and performance with LAI-2000 and DHP during an entire rice season. In this case, concomitant PocketLAI, LAI-2000 and DHP measurements were acquired on 26 ESUs from June the 17th to September the 8th in the 2014 rice season following the aforementioned sampling strategy over the Spanish local rice area.

PocketLAI computes the LAI_{eff} of an ESU averaging each LAI_{eff} calculated from each gap fraction reading on the ESU, while DHP and LAI-2000 first calculate the average gap fraction within an ESU and retrieve LAI_{eff} from it. Since the gap fraction- LAI_{eff} relationship is not linear, it is not the same to average first the gap fraction and then estimate the LAI_{eff} than the contrary (Weiss et al., 2004). This fact may be one of the reasons why PocketLAI generally underestimates LAI_{eff} values (see Fig. 3.4). The LAI_{eff} and $\text{LAI}_{eff}(57.5^\circ)$ values obtained with DHP were used to compare with LAI_{eff} estimates from LAI-2000 and PocketLAI, respectively. Different statistics were computed to assess the consistency and performance of the PocketLAI with LAI-2000 and DHP: root-mean-squared error (RMSE) and mean absolute error (MAE) were used to assess the accuracy, Mean error (ME) to evaluate the bias, and coefficient of determination (R^2) to account for the goodness-of-fit and variability between instruments.

Effective LAI values computed with all three instruments are well correlated. Note that since all the elements of the plant were taken into account LAI refers to PAI. Coefficient of determination computed between LAI_{eff} estimates acquired with PocketLAI and classical instruments was $R^2 = 0.95$ and $R^2 = 0.94$ for LAI-2000 and DHP respectively. Comparisons also reveal high accuracy and small bias between instruments ($\text{ME} = -0.38\text{ m}^2/\text{m}^2$, $\text{MAE} = 0.41\text{ m}^2/\text{m}^2$ for LAI-2000 and $\text{ME} = 0.46\text{ m}^2/\text{m}^2$, $\text{MAE} = 0.48\text{ m}^2/\text{m}^2$ for DHP). PocketLAI presents a very small negative bias regarding LAI-2000, although a slight positive bias is found at the beginning of the season ($\text{LAI}_{eff} < 1\text{ m}^2/\text{m}^2$). PocketLAI is also highly consistent with DHP although it tends to produce slightly lower values ($\text{ME} = 0.47$). Com-

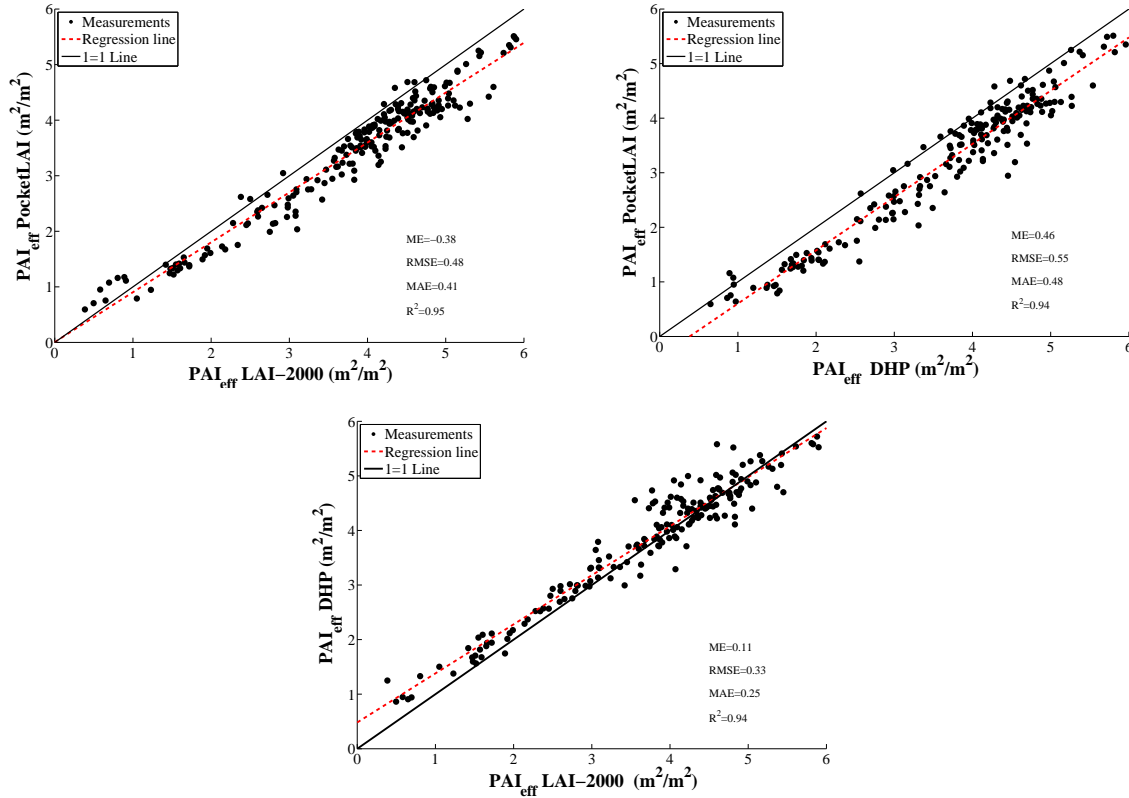


Figure 3.4: Comparison of effective plant area index measurements collected in the 2014 Spanish ERMES field campaign between: PocketLAI and LAI-2000 (up-left), PocketLAI and DHP (up-right), and DHP and LAI-2000 (bottom).

parison between LAI-2000 and DHP instruments, shows also good results in terms of accuracy, bias and variability (RMSE= 0.33 m²/m², ME= 0.11 m²/m², R²= 0.94 m²/m²) (see Fig. 3.4).

Additionally, LAI measurements acquired in 2014 with PocketLAI (LAI_{APP}), DHP (LAI_{DHP}), and LAI-2000 (LAI_{LIC}), allowed to construct a pure statistical regression method for each gathered data set using GPR (see formulation in Section 2.3.2). The three data sets and associated Landsat-8 surface reflectance for every measurement were divided into two different training (80%) and testing subsets (20%) and an independent model was built for each instrument. Each model was constructed by running GPR a hundred of times with different random selections of training and testing subsets. Table 3.1 reports the statistical indicators of the models' performance used for the LAI retrieval. GPR using LAI_{APP} outperforms LAI_{DHP} and LAI_{LIC} models in terms of accuracy and goodness of fit. However, LAI_{APP} bias is slightly higher than the other fits. In general, estimated LAI values with GPR from PocketLAI data were correlated to the classical instruments. Specifically, mean estimates retrieved when using PocketLAI showed similar results to those obtained with LAI-2000. Bias between LAI_{APP}-LAI_{LICOR}, LAI_{APP}-LAI_{DHP}, and LAI_{DHP}-LAI_{LICOR} was 0.06, 0.05 and 0.12 in LAI units, respectively. This bias is slightly smaller than other reported in the literature for crops (Verger et al., 2009).

In addition, Landsat-8 based LAI and associated uncertainty maps were derived (see Figure 3.5). We display generated maps on July 31th (DoY=212). Within-field variations

Table 3.1: Mean values of the statistical indicators (RMSE, MAE, absolute value of the ME, and R^2) between estimated and measured leaf area index in LAI_{APP} , LAI_{DHP} and LAI_{LIC} validation subsets. Standard deviations are shown in parenthesis.

Dataset	RMSE	MAE	ME	R^2
LAI_{APP}	0.51 (0.18)	0.35 (0.19)	0.12 (0.07)	0.94 (0.07)
LAI_{DHP}	0.64 (0.20)	0.43 (0.19)	0.15 (0.10)	0.87 (0.09)
LAI_{LICOR}	0.62 (0.21)	0.48 (0.26)	0.13 (0.08)	0.89 (0.07)

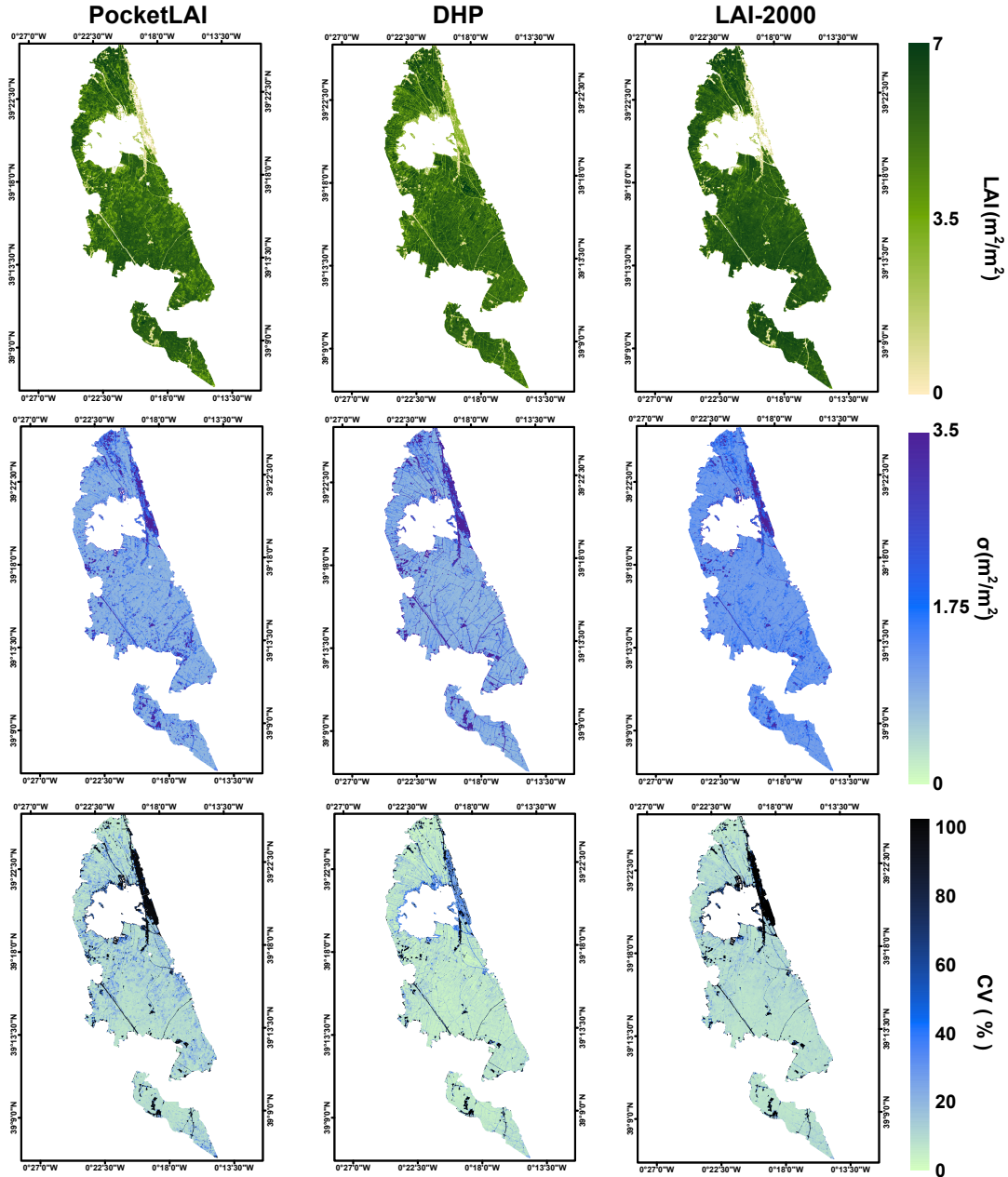


Figure 3.5: (Top) Leaf area index prediction maps, (middle) standard deviation maps and (bottom), coefficient of variation, obtained with GPR using PocketLAI, DHP and LAI-2000 ground data on the Landsat 8 image (DoY=212).

are observed in all three LAI maps due the following reasons: (i) the spatial LAI variability of rice fields corresponding to different varieties, (ii) differences in plant phenological stages and (iii) low LAI values corresponding to non-vegetated areas and boundaries of the rice fields. In the rice fields, LAI estimates fell within the expected range at that phenological state. Secondly, LAI uncertainty (σ) maps show low values within the rice fields. Higher uncertainties values appear over zones corresponding to non-vegetated areas or low LAI estimates. Particularly, on the east side of the maps it can be detected low LAI values and higher uncertainties referred to a different type of vegetation land cover (trees) and several man-made surfaces. This is essentially due to the fact that GPR cannot extrapolate outside the ranges seen in the dataset. Maps based on DHP ground data are able to predict more variability in LAI and uncertainty than the rest, and hence better identify rice field boundaries, roads, farm buildings, etc. Nevertheless, map based on PocketLAI is very similar in terms of LAI variability to LAI-2000. LAI-2000 provides uncertainties slightly higher but also higher values in terms of LAI. Uncertainty map values need to be interpreted as a confidence interval around the mean predictions. For this reason, we computed the coefficient of variation (ratio between LAI uncertainties and mean LAI predictions) $CV = \sigma/\mu * 100$ which provides relative uncertainties (see Fig. 3.5). These maps show that the majority of rice pixels fall below the 20% uncertainty threshold and can be considered well validated as proposed by the Global Climate Observing System (GCOS).

3.4.2 Accuracy assessment over ground LAI

After 2014 testing experiments, the processing chain was applied and assessed in 2015 and 2016 computing RMSE between the estimates and the *in situ* measurements. RMSE values of 0.39, and 0.38 m^2/m^2 were found in Spain and Italy respectively in 2015, showing good accuracy between GPR-Landsat map values and the *in situ* LAI measurements (see Fig. 3.6). The GPR-SPOT5 retrievals revealed good accuracies as well, showing RMSE values of 0.51 and 0.47 m^2/m^2 in Spain and Italy respectively. Other different statistics, such as ME, MAE and the coefficient of determination were also computed to evaluate the bias, accuracy, and the goodness-of-fit between retrievals and *in situ* measurements. A remarkably good correspondence between the 2015 satellite retrievals and *in situ* measurements was found in the Spanish site, with very low bias for both GPR-Landsat and GPR-SPOT5 models. A good agreement and low biases were also observed for the Italian site. In all cases, very high correlations were found with $R^2 > 0.92$. Error bars in Fig. 3.6 refer to the standard deviation of the field measurements, and are thus related with the heterogeneity of the ESU.

In 2016, an overall RMSE of 0.69 m^2/m^2 and a coefficient of determination of 0.95 were found. A slight bias was observed between Sentinel-2A LAI estimates and ground data. More in detail, Table 3.2 shows the statistical indicators between ground data and Sentinel-2A estimates over each study area. In general, very high correlations were found. Low biases were observed in the case of the Spanish study area, while in Italy and Greece small biases were observed with respect the ground data (see Table 3.2). Biases could be explained due the rice ESU heterogeneity and the slight underestimation of PocketLAI regarding other measurements (Campos-Taberner et al., 2015c, 2016a). Relative RMSE to mean ($rRMSE_m$) and relative RMSE to range ($rRMSE_r$) are also reported in Table 3.2 and Fig. 3.7 showing good accuracies and revealing that the processing chain performed well.

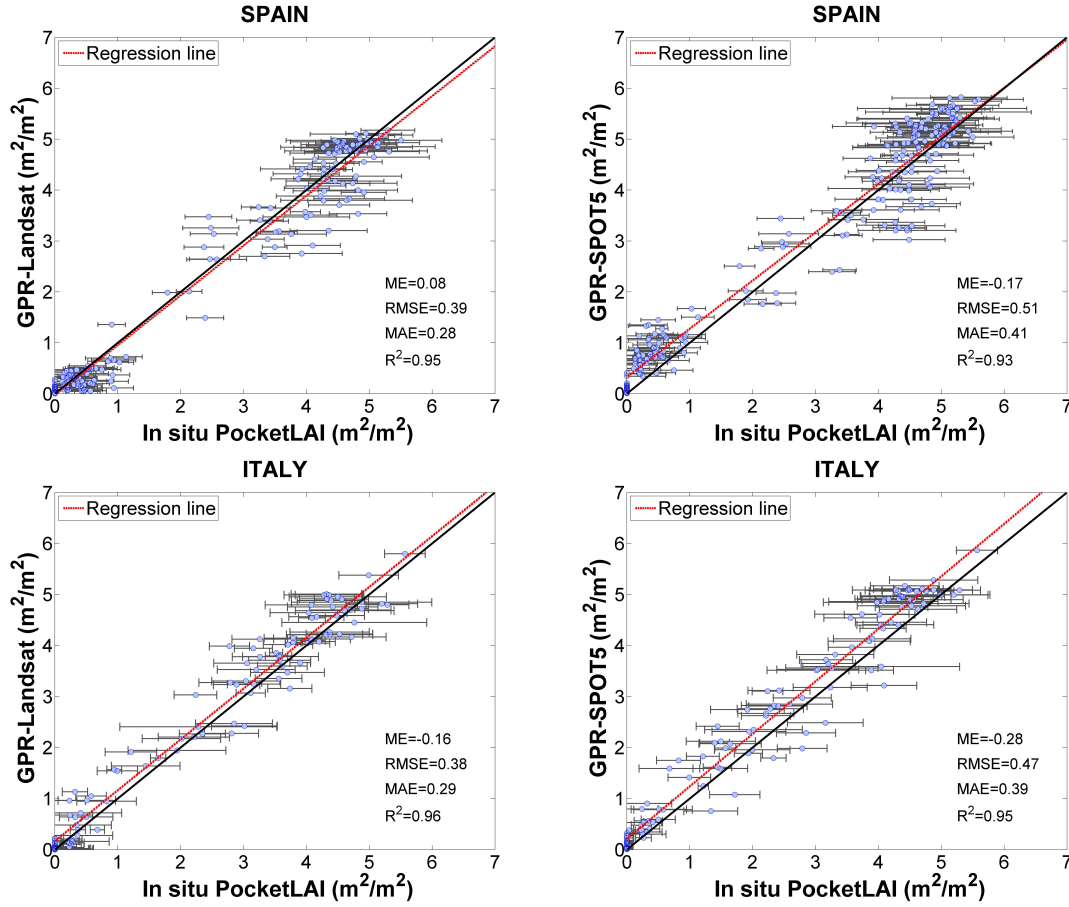


Figure 3.6: Scatter plots of estimated LAI values using Landsat (left) and SPOT5 (right) data versus *in situ* LAI measurements acquired with PocketLAI over Spain (up) and Italy (bottom). Standard deviation of measurements is drawn as horizontal error bars as well as bisector line (black line). For the sake of visualization, GPR prediction uncertainty (around 1) are not shown.

Table 3.2: Statistical indicators (coefficient of determination (R^2), Root Mean Squared Error (RMSE), Mean Error (ME), Mean Absolute Error (MAE), relative Root Mean Squared Error to mean ($rRMSE_m = RMSE/\hat{y}_i$) and relative Root Mean Squared Error to range ($rRMSE_r = RMSE/(max(y_i) - min(y_i))$) between the ground measurements and LAI estimates during the 2016 rice season over the three study areas.

	Sentinel-2A		
	Spain	Italy	Greece
R^2	0.97	0.94	0.94
RMSE (m^2/m^2)	0.56	0.82	0.77
ME (m^2/m^2)	-0.15	-0.67	-0.62
MAE (m^2/m^2)	0.43	0.69	0.60
$rRMSE_m$ (%)	20.1	23.5	22.7
$rRMSE_r$ (%)	9.3	16.5	15.3

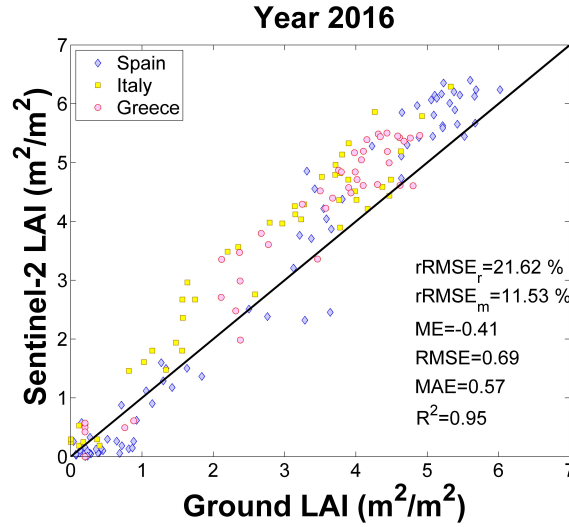


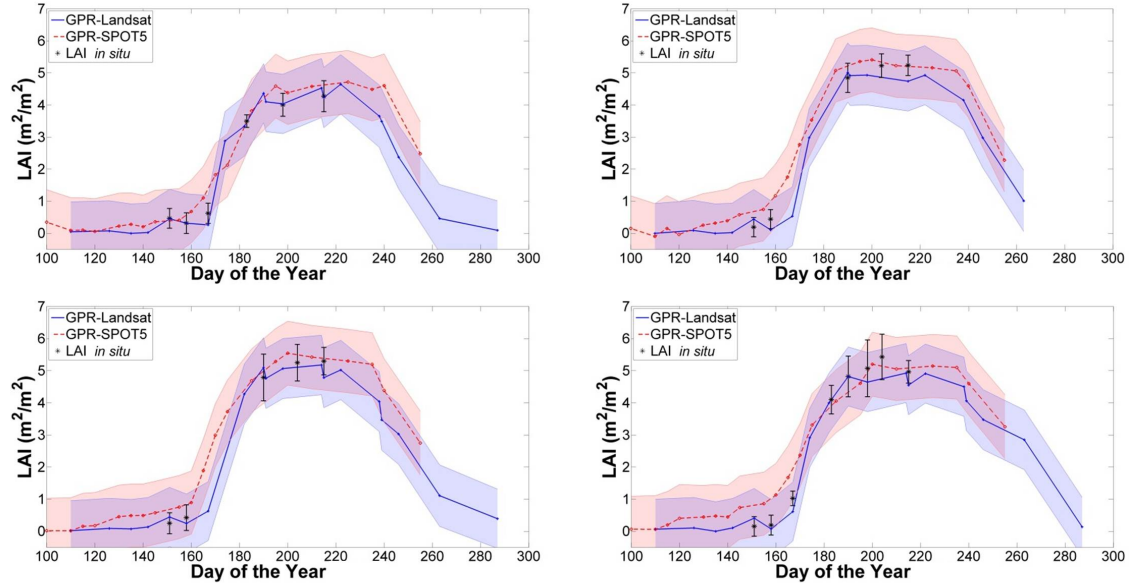
Figure 3.7: Scatter plots of Sentinel-2A estimated LAI values versus *in situ* LAI measurements during the 2016 rice season. For the sake of visualization, standard deviation of measurements and Sentinel-2A prediction uncertainty (around 0.8) are not shown.

3.4.3 Temporal evolution of LAI estimates

The *in situ* LAI data points allowed us to compare the temporal evolution of field measurements over the study areas (see Fig. 3.8). In general, LAI estimates derived from the developed processing chain using GPR-Landsat and GPR-SPOT5 data agree with regard to the 2015 seasonal rice phenological cycle and followed the temporal dynamics of the ground measurements. The different LAI evolutions that can be observed in Fig. 3.8, show coherent temporal behaviors as a consequence of either different rice varieties or sowing dates that determine a shift in the development curve. It is interesting to note that the difference between Landsat-7 and Landsat-8 did not induce large difference in LAI retrieval. For example, LAI estimates of Landsat-8 DOY=190 and Landsat-7 DOY=191 for the Spanish site showed differences lower than $0.25 \text{ m}^2/\text{m}^2$ (see blue profiles in Fig. 3.8).

In addition the 2016 LAI temporal behaviours are shown in Fig. 3.9 which exhibits the temporal consistency of LAI estimates from Sentinel-2A (blue dots) with the ones from Landsat-7/8 (orange/green dots). Each dot is the estimated LAI averaged over a rice field; over each study site. The fields showed in Fig. 3.9 were selected to cover LAI variability as large as possible and, where available, cultivated with local traditional varieties (*Bomba* and *Carnaroli* in top panels of Spain and Italy). All the other panels show modern rice varieties (*Olympiada* and *Mare CL* in Greece, *Sirio CL* and *Selenio* in bottom panels of Spain and Italy respectively). In Italy, there is larger variability of agro-practices, which is highlighted in the example field of Fig. 3.9 by the significant difference in sowing dates (vertical dashed bars). Moreover, according to *in situ* observations, field #68 is characterized by the presence of a cover crop preceding rice sowing, which is also confirmed by LAI profiles showing LAI values before the abrupt drop due to field preparation (ploughing and laser levelling) for rice sowing.

Spain



Italy

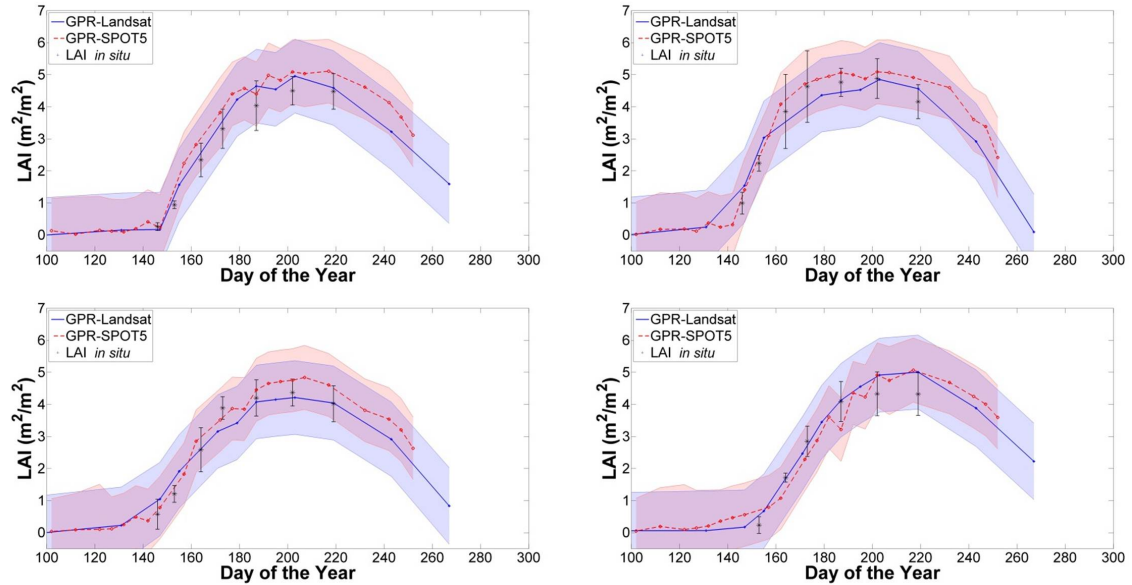


Figure 3.8: Estimated temporal dynamics and the field LAI measurements within four representative ESUs in (top) the Spanish and (bottom) the Italian study area in 2015. The prediction uncertainty provided by the GPRs is shown in shaded blue (GPR-Landsat) and orange (GPR-SPOT5) around the mean prediction; in purple the overlap. The standard deviations of the *in situ* measurements are displayed as error bars.

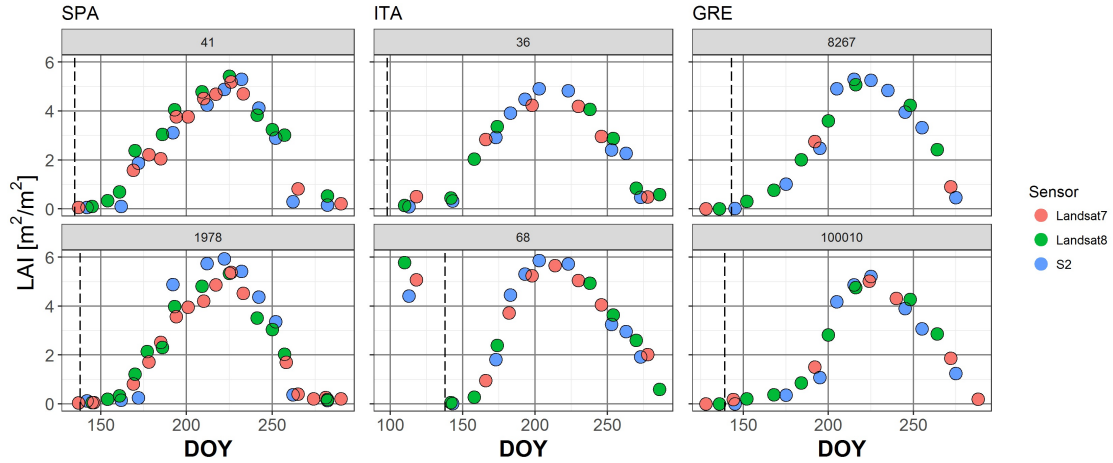


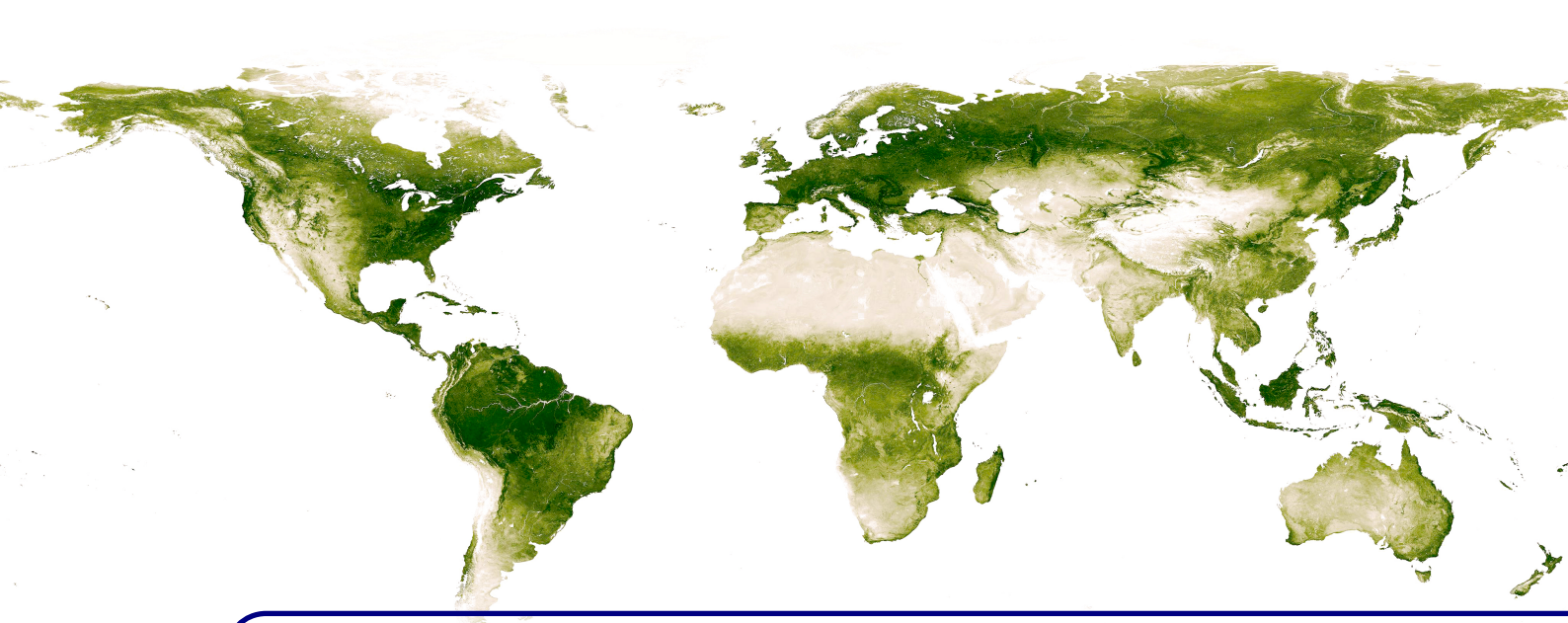
Figure 3.9: Example of merged series of LAI on 3 sites. Dotted vertical bars indicate the sowing date. Numbers indicate fields ID. For the sake of visualization, standard deviations are not shown.

3.5 Concluding remarks

In this chapter we introduced the acquisition of field measurements for biophysical parameter estimation and validation. Two of the most widely used instruments for indirect and non-destructive LAI measurements (e.g. LAI-2000 and DHP) were described as well as the theoretical background in which they rely.

Besides, in this Thesis we propose the use of smartphones for non-destructive LAI measurement through the use of an application called PocketLAI. PocketLAI was tested over rice fields in the 2014 rice season in Spain. The results allow to consider PocketLAI as a very powerful alternative to classical and commercial instruments for LAI monitoring during field campaigns. In particular, its ease of use and cheapness allows to quickly collect a large number of LAI measurements which can be exploited as field data for calibration and validation of remote sensing biophysical land products as well as for LAI mapping through up-scaling methods. Therefore, PocketLAI was used for acquiring *in situ* LAI measurements for the validation of the processing chain at local scale during 2015 and 2016.

In the final part of the chapter, the validation of the developed chain at local scale was performed using the PocketLAI field data. The results showed high accuracy if compared with ground data and the temporal LAI evolution of the estimates followed the rice growing phenological stages.



4. Biophysical parameters at global scale

Contents

4.1 Introduction

4.2 Remote sensing existing products at global scale

4.2.1 MODIS (MOD15A2)

4.2.2 Copernicus (GEOV1)

4.2.3 EUMETSAT Land-SAF (SEVIRI/MSG)

4.3 Developed processing chain at global scale

4.3.1 Inputs: EUMETSAT EPS

4.3.2 Retrieval chain

4.3.3 Global LAI, FVC and FAPAR maps

4.4 Global validation

4.4.1 Comparison with global products

4.4.2 Comparison with local estimates

4.5 Concluding remarks

This Chapter is partly based on the following publications:

- Garcia-Haro, F.J, **Campos-Taberner, M.**, Camacho, F., Muñoz-Marí, J., Laparra, V., Camps-Valls, G. “Multioutput statistical retrieval for global vegetation monitoring in the EUMETSAT Polar System: Principles and Validation”. (submitted)

4.1 Introduction

This Chapter is devoted to give an overview of the main operational remote sensing biophysical products at global scale and describes the development and validation of the processing chain at global scale conducted in this Thesis. The same conceptual scheme that in the case of the local chain was followed. The main difference relies in the spatial-spectral input data and in the PROSAIL database which was simulated covering a worldwide range of vegetation situations. The algorithm has been recently coded and integrated in the Land-SAF system for the near real time (NRT) derivation of vegetation parameters from AVHRR-3/MetOp-B data. In addition, a validation experiment was conducted in order to compare the performance of the global estimates with regard to the most widely used and validated products currently available products. Eventually, we compare the estimates derived by either the local or the global processing chain over the same areas.

4.2 Remote sensing existing products at global scale

4.2.1 MODIS (MOD15A2)

Terra MODIS Collection 5 LAI/FAPAR products (MOD15A2) ([Knyazikhin et al., 1999](#)) have been produced since February 2000 until present at eight-day step over a sinusoidal grid at 1 km spatial resolution. The main algorithm of the products is based on a three-dimensional radiative transfer model ([Knyazikhin et al., 1998](#)) which links surface spectral bi-directional reflectance factors (BRFs) to both canopy and soil spectral and structural parameters ([Myneni & Ross, 2012](#)).

The algorithm estimates LAI and FAPAR by comparing observed and modeled MODIS BRFs that are stored in biome type specific LUTs. An estimate is considered a candidate solution if agrees within specified levels of (model and measurement) uncertainties. A backup algorithm based on NDVI relationships with LAI/FAPAR ([Myneni & Williams, 1994](#)), calibrated over the same radiative transfer model simulations is used if the main algorithm fails. This occurs if the uncertainties of input BRFs are larger than threshold values or due to deficiencies of the RTM that result in incorrect simulated BRFs. For example, the backup algorithm is used for retrievals where residual atmospheric contamination is detected.

The MODIS algorithm inputs are daily surface reflectance at red (648 nm) and near-infrared (858 nm) bands because of high uncertainties in other bands ([Wang et al., 2001](#)). Both main and back-up algorithms use a biome map in which global vegetation is classified into eight biomes: (B1) grasses and cereal crops; (B2) shrublands; (B3) broadleaf crops; (B4) savannas; (B5) evergreen broadleaf forests; (B6) deciduous broadleaf forests; (B7) evergreen needle-leaf forests and (B8) deciduous needle-leaf forests. Figure 4.1 illustrates the biome map. The MODIS algorithm uses this prior information for better addressing the “ill-posed” inverse problem. No LAI/FAPAR values are provided over bare and very sparsely vegetated areas, permanent wetland, ice and snow pixels, urban and water bodies.

The operational retrieval chain uses as input MODIS daily red and near-infrared surface reflectances and a biome map to generate daily LAI/FAPAR estimates without any quality-control masking on inputs. Then, a temporal compositing method is used to select the best retrievals and generate an 8-daily product from daily estimates. Estimates from the main algorithm are selected, and if none are available, back-up algorithm retrievals are selected. The MODIS algorithm accounts for vegetation clumping at the canopy and leaf

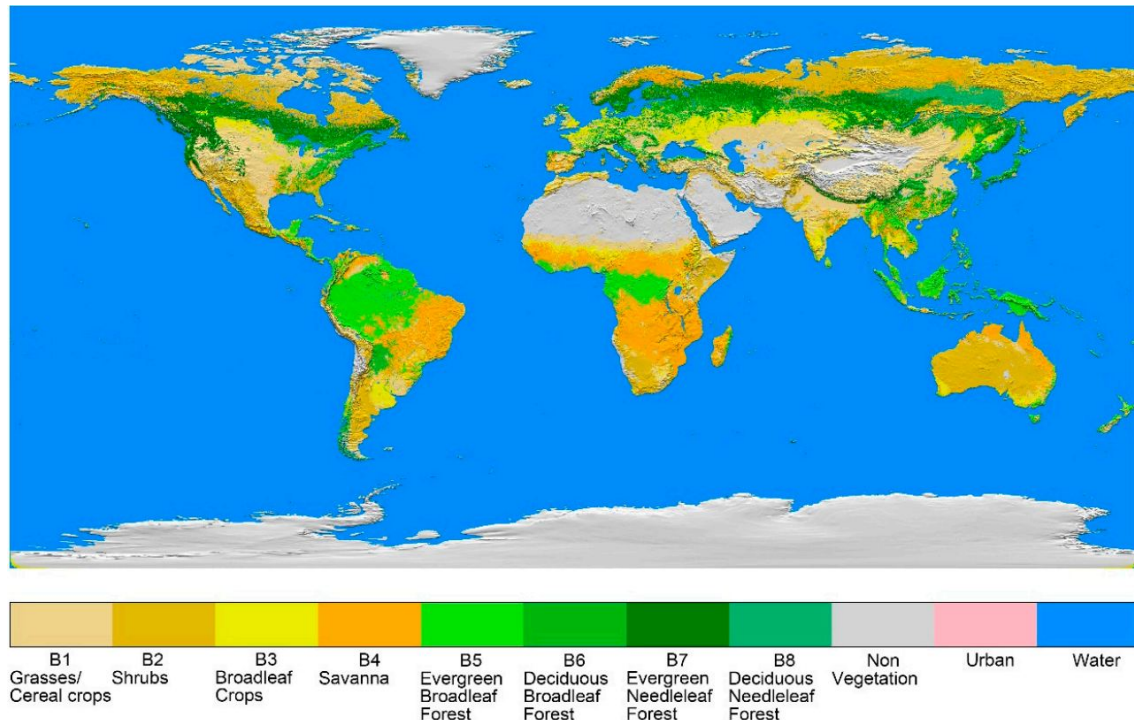


Figure 4.1: Global 8-biome map used in the retrieval chain of MOD15A2 products. Credits: [Yan et al. \(2016\)](#)

scales through 3D radiative transfer formulations. The clumping at the landscape level is partly addressed via mechanism based on radiative transfer theory of canopy spectral invariants ([Huang et al., 2007](#)). In the case of FAPAR, the estimates correspond to the instantaneous black sky value at the MODIS satellite overpass (10:30 a.m.). The retrieved values of the product correspond to when the maximum FAPAR value within the eight days period is observed ([Yang et al., 2006](#)). The compositing window tries to remove contaminated retrievals also reducing the impact of day-to-day artificial variations in surface reflectance due to cloud and residual atmospheric effects.

The MODIS LAI/FAPAR products also include a quality flag indicator. The quality information was used to filter pixels flagged invalid or poor quality (e.g., cloud state not clear or main method failed).

4.2.2 Copernicus (GEOV1)

The Copernicus Global Land Service (<http://land.copernicus.vgt.vito.be/>) provides open access to LAI, FAPAR and FVC family of biophysical products from SPOT/VEGETATION (1999 to May 2014) and PROBA-V (June 2014 up to date) at 1/112° spatial resolution (1 km at the equator) in a *Plate Carrée* projection (regular latitude/longitude grid), namely GEOV1 products.

The LAI retrieval processing chain relies on neural networks trained to generate LAI estimates by fusing and scaling MODIS Collection 5 and CYCLOPES 3.1 products ([Baret et al., 2013](#)). This “fusion” is conducted in order to take advantage of their individual specific performances while limiting the situations where they show deficiencies. The input data of the retrieval chain are top of canopy directional normalized reflectance from either

SPOT/VEGETATION or PROBA-V computed using a kernel-driven BRDF (Bidirectional Reflectance Distribution Function) model (Hagolle et al., 2005). The retrieval methodology uses a temporal weighting function giving the higher weight in the last observation of the composing period. The estimates are built from 30 days composite observations and the products are provided at a 10 days sampling interval.

GEOV1 is highly correlated to CYCLOPES 3.1 for low-medium LAI values and provides systematically larger values for dense canopies due to the contribution of MODIS estimates in the training process. Since the GEOV1 LAI is a weighted combination of CYCLOPES 3.1 and MODIS Collection 5 products and that both original products consider the clumping at different scales on different manners, the resulting GEOV1 LAI is relatively consistent with an *effective* LAI for low values and close to the *actual* LAI for high values. Unlike in MODIS and similarly to CYCLOPES 3.1, no biome classification is required to run the GEOV1 algorithm. GEOV1 provides instantaneous black-sky FAPAR value at around 10:15 a.m. solar time under clear sky conditions, which is a close approximation of the daily integrated black-sky FAPAR value.

4.2.3 EUMETSAT Land-SAF (SEVIRI/MSG)

The Satellite Application Facility (SAF) for Land Surface Analysis (LSA), also known as Land-SAF, is a dedicated processing center serving the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) (<http://landsaf.meteo.pt>). The SAF network complements the product-oriented activities at the EUMETSAT Central Facility in Darmstadt (Germany). The Land-SAF system is located at IPMA (*Instituto Portugues do Mar e Atmosfera*) in Lisbon (Portugal) and VITO (*Vlaamse Instelling Voor Technologisch Onderzoek*) in Mol (Belgium) and has been designed to generate, archive, and disseminate the operational products.

The main purpose of Land-SAF is to develop and implement algorithms that allow an operational use of land surface variables taking full advantage of remotely sensed data from EUMETSAT satellites and sensors to measure land surface variables. Land-SAF has been especially designed to cover the needs of the meteorological community, particularly Numerical Weather Prediction (NWP). However, Land-SAF addresses a much broader community, which includes users from environmental management and land use, natural hazards management and, renewable energy resources assessment, particularly biomass, depending on biophysical parameters. In this framework, biophysical parameters such as LAI, FAPAR and FVC have been retrieved from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG) on a daily and 10-day time step since 2006. These datasets have been recently reprocessed (since 2004 onwards) with the last version of the algorithm to obtain an homogeneous time series of climate data records (CDR).

The retrieval algorithm of SEVIRI/MSG products, namely VEGA, relies on the use of bidirectional reflectance distribution function (BRDF) parameters (Roujean & Lacaze, 2002) which contain specific spectral directional signatures of vegetation reflectances. This kind of parametric BRDF models assume that the surface reflectance can be expressed as a linear combination of angular functions (kernels¹) and three free parameters (k_0 , k_1 , k_2) being k_0 the reflectance at nadir sun-view geometry which constitutes the input of the Land-SAF algorithm. FAPAR is based on simulations of surface reflectances in optimal angular

¹Not to be confused with kernels in machine learning (see Section 2.2.2)

geometries (Roujean & Breon, 1995), while FVC is estimated through the application of a spectral mixture analysis methodology developed taking into account the spectral variability of vegetation in different ecosystems (García-Haro et al., 2005), to visible and near infrared reflectance values. The algorithm relies on a statistical approach, in which soil and vegetation components are represented by a multimodal probability density function. LAI is retrieved from FVC following the methodology developed by (Roujean & Lacaze, 2002), which proved to be more effective than traditional techniques based on spectral vegetation indices. This method relies on a tractable physical model for the interception of solar irradiance by vegetative canopies with an emphasis on shading effects (Roujean, 1996).

4.3 Developed processing chain at global scale

During the current phase of Land-SAF project (2012–2017), the Land-SAF consortium pursues the consolidation of operational activities, upgrading of algorithms and continuously validating products based on the evaluation of user needs (Trigo et al., 2011). Within the consortium, the University of València is the responsible of the development and validation of the vegetation products. In this framework, LAI, FAPAR and FVC will be produced from the Advanced Very High Resolution Radiometer (AVHRR) on board the Meteorological–Operational (MetOp) polar orbiting satellites also known as EUMETSAT Polar System (EPS).

This section describes the strategy adopted for deriving these biophysical parameters, which has followed the hybrid model developed in this Thesis. The rationale of the algorithm is the inversion of the PROSAIL radiative transfer model with a family of non-linear retrieval methods (see Section 2.2). The best method in terms of stability, accuracy and robustness is then implemented into the operational chain to retrieve a family of global EPS ten-day (ET) vegetation products (ETFVC, ETLAI, ETFAPAR) from corresponding AVHRR/MetOp data. The inputs of the chain, the PROSAIL parameterization and the validation methods are described in the following sections.

4.3.1 Inputs: EUMETSAT Polar System BRDF

The EPS is Europe's first polar orbiting operational meteorological satellite. EUMETSAT has the operational responsibility for the MetOp satellites, the first of which (MetOp-A) was successfully launched on October 19, 2006, the second (MetOp-B) in September 17, 2012, whereas the launch of the third (MetOp-C) is foreseen for 2018. MetOp carries on-board a wide range of sensors, and among them, the AVHRR instrument is the main sensor in charge of providing observations useful for most of parameters that Land-SAF supplies.

AVHRR offers capability to observe the whole globe every day at 1 km spatial resolution (at nadir), in the visible and infrared bands of the electromagnetic spectrum. The Land-SAF EPS vegetation products will be delivered at ten-day time step based on the cloud-free BRDF k_0 parameter of the EPS ten-day albedo product (ETAL). Inputs are atmospherically corrected cloud-screened top of canopy k_i parameters in three EPS channels: $0.6\ \mu\text{m}$ (C1), $0.8\ \mu\text{m}$ (C2) and $1.6\ \mu\text{m}$ (C3), namely “HDF5_LSASAF_EPS_ETAL-Channel-K012” (Geiger et al., 2016). They correspond to the directional coefficients of the BRDF model, and are generated through the use of a semiempirical bidirectional reflectance distribution

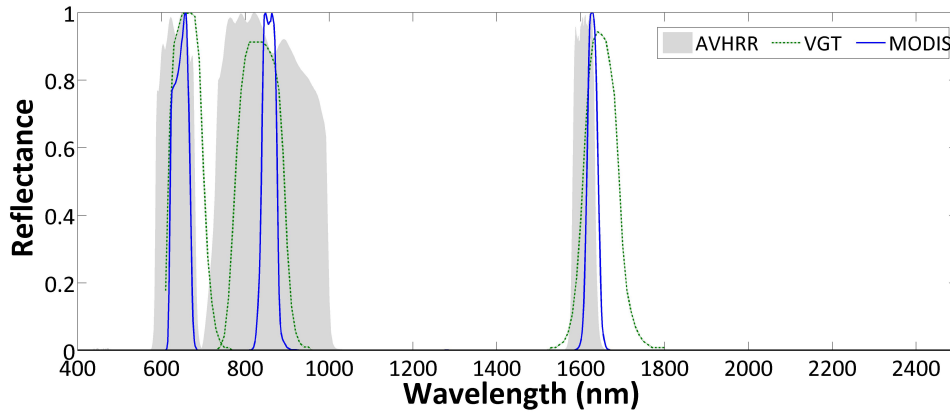


Figure 4.2: Spectral responses of reflectance inputs used in MOD15A2, GEOV1 and Land-SAF EPS.

function (Roujean et al., 1992) which contains specific spectral directional signatures of vegetation reflectances.

4.3.2 Retrieval chain

Similarly to the retrieval chain at local scale (see Section 2.3), the approach relies on the inversion of PROSAIL with machine learning regression techniques. The differences between the local chain are related to the PROSAIL parametrization, surface reflectance input, temporal interval of the estimates and the multioutput retrieval method. The retrieval chain was adapted to cope with this features as detailed in the next subsections.

PROSAIL parametrization at global scale

Conversely to the retrieval chain at local scale, PROSAIL parametrization used at global scale not only includes constrained simulations for a specific crop but also a wide range of simulations covering different vegetation biomes. Thus, these simulations were computed in order to represent as much as possible the Earth's vegetated areas. In addition, the simulations considered a variety of world wide representative soils in order to account for the background spectral component.

The PROSPECT version 5 was used for the coupling with SAIL. PROSPECT-5 explicitly accounts for leaf brown pigments (C_{bp}) and deals with chlorophylls (C_{ab}) and carotenoids (C_{ar}) separately, while PROSPECT-4 uniformly treats all photosynthetic pigments.

Top of canopy reflectance was computed for each wavelength (from 400 to 2500 nm) and then filtered according to the spectral sensitivity of MetOP/AVHRR bands (red, near infrared and short wave infrared channels). Figure 4.2 shows the relative spectral response of the MetOP/AVHRR. This procedure was executed 2000 times accounting for any combination of the input variables thus obtaining a database composed by 2000 pairs of reflectances in the AVHRR channels and the corresponding associated biophysical parameters LAI, FAPAR, and FVC.

It is worth mentioning that brown pigments were intentionally fixed to zero in order to account only for photosynthetic elements of the canopy. Therefore, this assumption implies the simulated LAI, FVC and FAPAR refer only to green elements of the canopy which

match best with requirements of the user community in primary production models (Green et al., 2003). In addition, since FAPAR depends on sun position, FAPAR simulations were run also integrating radiation from sunrise to sunset. This fact allowed us to derive a daily-integrated FAPAR. This kind of daily-integrated green FAPAR is demanded for users since most of the primary productivity models using FAPAR run at a daily time resolution.

Multioutput regression algorithm

One of the main features of the chain at global scale is the capability of deriving jointly the three biophysical parameters which implies the construction of a unique model able to retrieve all three parameters at the same time. This approach was achieved using a multioutput version of the NN, KRR and GPR. As we will see later, improved results and more consistency between predictions is attained with this approach.

Table 4.1: Statistical indicators over the unseen test set (best results in bold). In the case of the single output algorithms, one has to train a model per parameter to retrieve, while for multioutput versions only a single model is learned.

	GPR _{multi}			GPR		
	LAI	FVC	FAPAR	LAI	FVC	FAPAR
R	0.951	0.995	0.984	0.948	0.993	0.982
RMSE	0.949	0.036	0.067	0.978	0.038	0.069
MAE	0.569	0.025	0.043	0.573	0.027	0.046
ME	0.005	0.001	0.001	0.010	0.001	0.001

	NN _{multi}			NN		
	LAI	FVC	FAPAR	LAI	FVC	FAPAR
R	0.950	0.993	0.983	0.951	0.994	0.983
RMSE	0.960	0.038	0.069	0.952	0.037	0.068
MAE	0.570	0.027	0.045	0.569	0.025	0.044
ME	0.009	0.001	0.001	0.009	0.001	0.001

	KRR _{multi}			KRR		
	LAI	FVC	FAPAR	LAI	FVC	FAPAR
R	0.940	0.994	0.982	0.940	0.992	0.979
RMSE	1.052	0.037	0.071	1.052	0.046	0.076
MAE	0.619	0.026	0.046	0.619	0.028	0.048
ME	0.015	0.001	0.002	0.015	0.002	0.002

In the case of NN, the approach is a multioutput algorithm *per se* given their characteristics (i.e. connected layers, weights and biases). In case of kernel methods, the algorithms can be set up to cope with multioutput regression problems simply adapting the kernel hyperparameters for a unique kernel which is able to deal all the outputs. The optimization of the hyperparameters for the unique kernel is done either by cross validation or by maximizing the marginal likelihood in the case of KRR and GPR, respectively.

After the construction of the simulated database, the multioutput algorithms were used to train (i.e. with the 70% of the samples) a global model to jointly retrieve LAI, FAPAR, and FVC. The theoretical performances of the single output and multioutput (denoted by

subscript *multi*) algorithms are shown in table 4.1 where statistics were computed over the unseen test set (i.e. the remaining 30% of the samples).

Table 4.1 shows that GPR and KRR multioutput algorithms improve the results of their single output versions while NN do not. GPR_{multi} outperformed the rest of the regression methods and thus was selected to jointly retrieve the four biophysical parameters at the same time.

4.3.3 Global LAI, FVC and FAPAR maps

The processing chain has been recently coded and integrated in the Land-SAF system for the near real time derivation of vegetation parameters from AVHRR/Metop. Given the spatial-temporal features of the EPS inputs, the application of the trained GPR_{multi} model into the retrieval chain, allowed to obtain multitemporal LAI, FVC and FAPAR maps at global scale at about 1 km pixel size every 10 days. Figure 4.3 shows an example of LAI, FAPAR and FVC maps corresponding to the period (25 June-15 July). The maps show consistent and similar spatial patterns exhibiting the higher values over Earth dense vegetated surfaces whereas lowest values (virtually zero) over sparse and bare areas.

4.4 Global validation

4.4.1 Comparison with global products

The validation of the aforementioned estimates were realized to ensure the consistency with other relevant vegetation products for different biomes and seasons. The estimates have been intercompared over 2 entire years (2015 and 2016) with VEGA (FVC, LAI, FAPAR), GEOV1 (FCOVER, LAI, FAPAR) and MOD15A2 (LAI, FAPAR) products. The intercomparison was conducted over a network of sites, so called BELMANIP-2.1 (Benchmark Land Multisite Analysis and Intercomparison of Products) especially selected for representing the global variability of vegetation making them suitable for global intercomparison of land biophysical products (Baret et al., 2006). BELMANIP-2.1 is an updated version of the original BELMANIP sites which includes 445 sites located in relatively flat and homogeneous areas all over the globe (see Fig. 4.4). The sites are considered to be representative of the different planet biomes over a $10 \times 10 \text{ km}^2$ area, almost flat, and with a minimum proportion of urban area and permanent water bodies.

With the aim of assessing the accuracy of the EPS estimates provided by the retrieval chain at global scale, RMSE, ME, MAE and R were computed over the BEMANIP.2-1 network (see Fig. 4.5). Results show that EPS biophysical estimates are highly correlated to GEOV1 products and present good correlations and accuracies with respect to VEGA estimates as well. EPS LAI is more correlated to GEOV1 and VEGA showing values of $R=0.95$ and $R=0.88$ respectively, while in the case of MOD15A2, EPS LAI estimates seem to saturate at about $5.5 \text{ m}^2/\text{m}^2$ but also good correlations are found ($R=0.83$).

Similar behaviour is observed for EPS FAPAR while EPS FCV estimates present a slight negative bias regarding GEOV1 and VEGA but high correlations $R \geq 0.93$ are obtained. In addition, Fig. 4.6 shows the comparison over biome sites, namely, BEF (Broadleaved Evergreen Forest), BDF (Broadleaved Deciduous Forest), NLF (Needle-leaved Forest), S (Shrublands), H (Herbaceous), CM (Cultivated and Mosaic), SP (Sparse), B (Bare areas). The comparison shows the bias of the majority of biomes fall into the range of $\pm 0.5 \text{ m}^2/\text{m}^2$ in the case of LAI, and ± 0.05 in the case of FAPAR and FVC. The

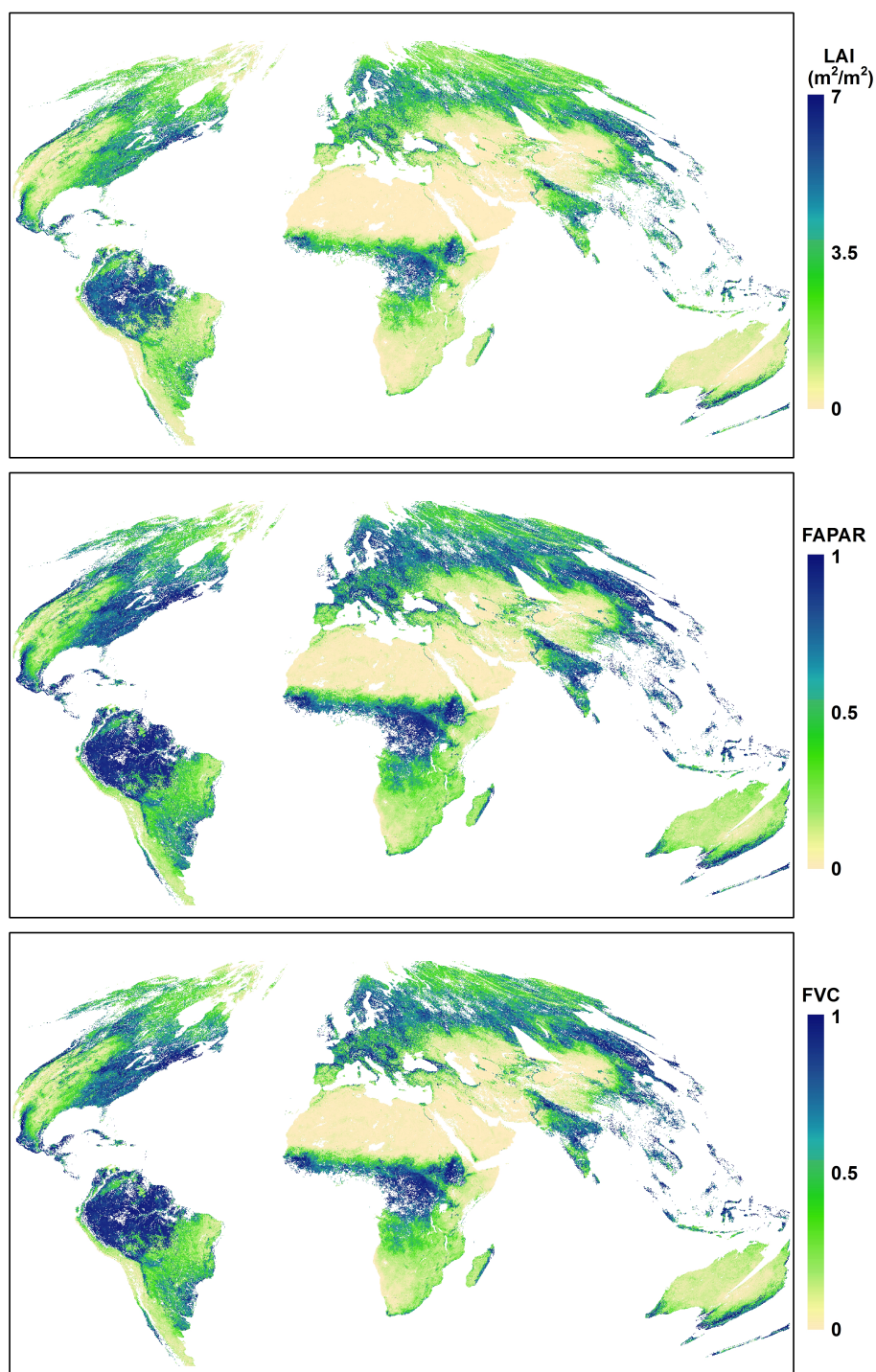


Figure 4.3: LAI (top), FAPAR (middle) and FVC (bottom) maps derived with the global processing chain in September 15, 2015.

higher discrepancies are observed in forested areas mainly when comparing EPS FAPAR with VEGA FAPAR estimates.

In general, results are consistent with the available biophysical products and observed discrepancies in terms of magnitude can be explained either by assumptions made in the retrieval algorithm and product definition, errors in the inputs, sensor spectral responses,

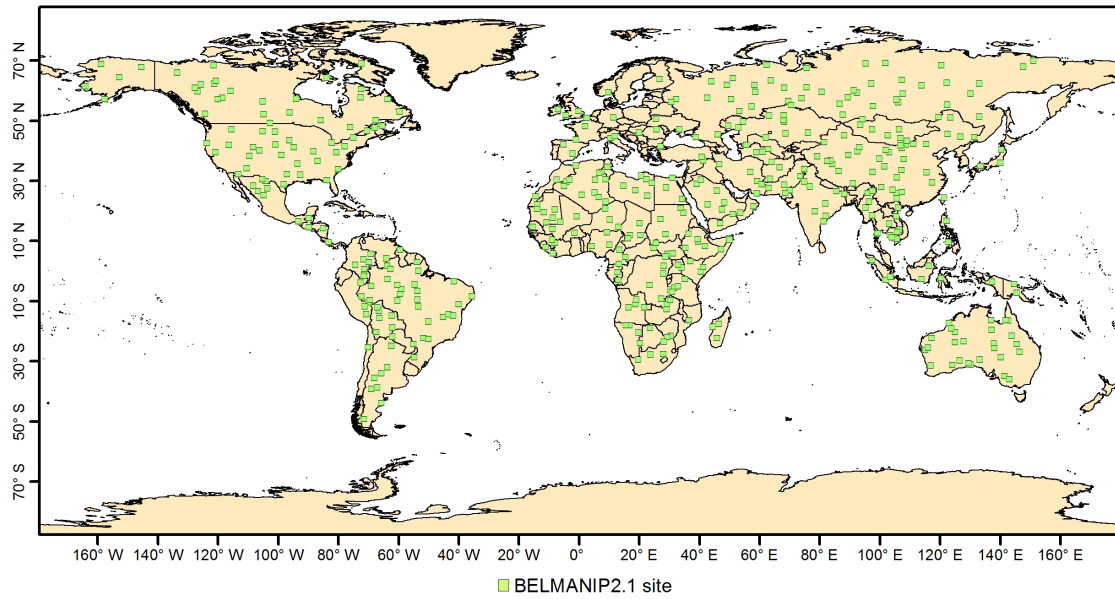


Figure 4.4: Sites location of the BELMANIP-2.1 network used for intercomparison of EPS estimates and VEGA, GEOV1 and MOD15A2 products.

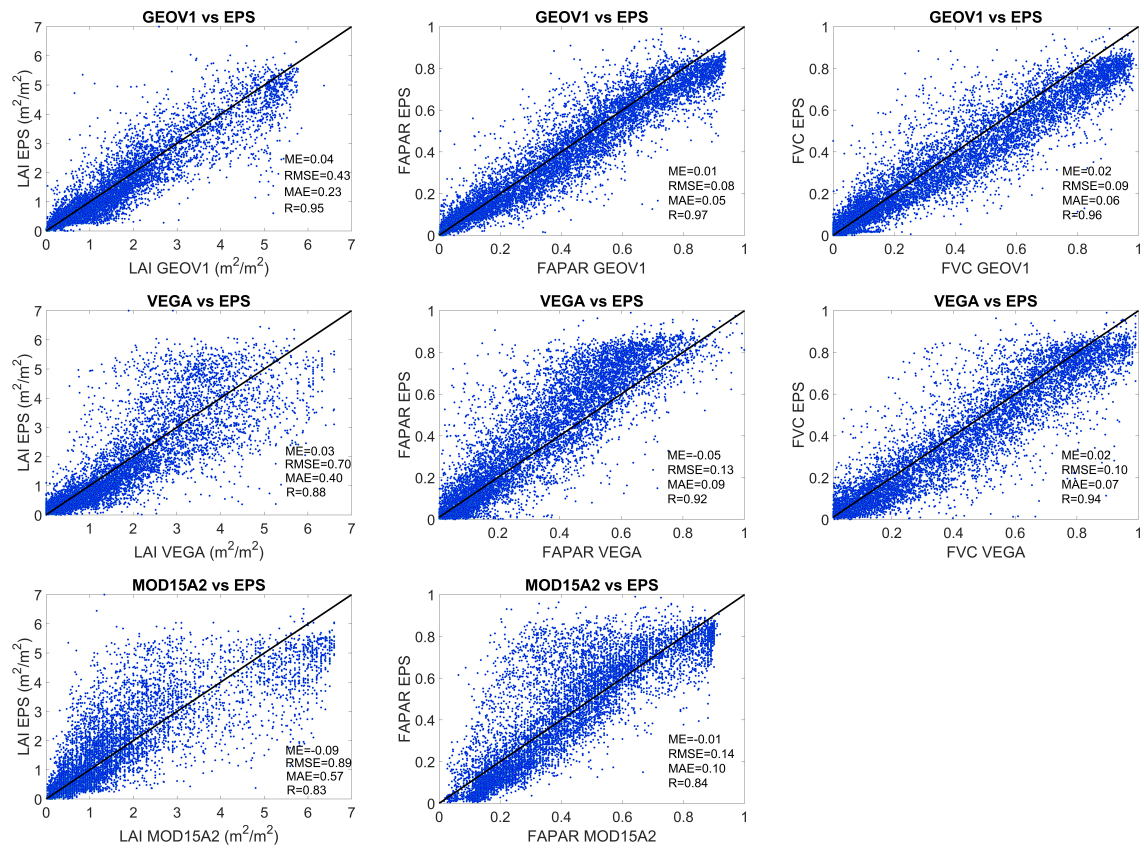


Figure 4.5: Scatter plots of estimated LAI, FAPAR and FVC values in 2015 and 2016 years over the BELMANIP-2.1 sites.

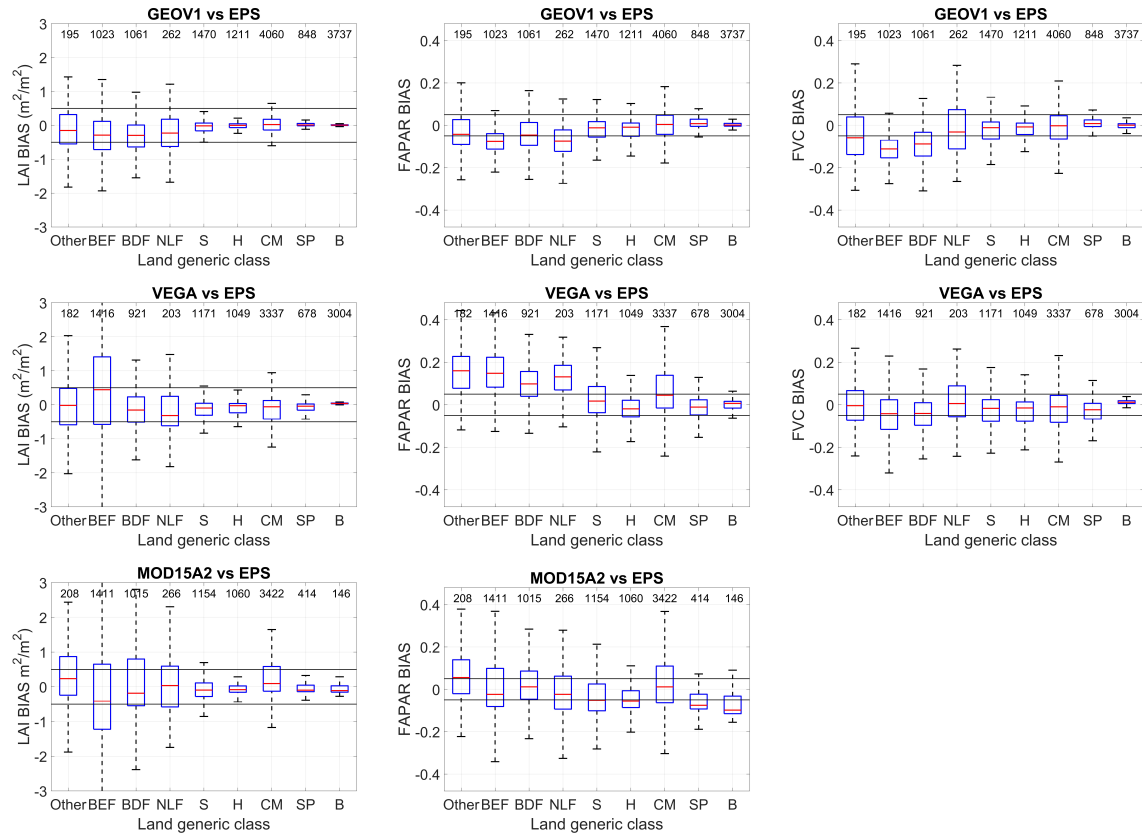


Figure 4.6: Box plots of estimated LAI, FAPAR and FVC values in 2015 and 2016 years over the BELMANIP-2.1 sites.

point spread functions or different atmospheric correction and temporal composite methods. Although this is not a strict validation, it offers a means of assessing the uncertainties (systematic or random) between products. We have used as reference products which have been largely improved and validated in the previous years. Future research includes the assessment of the realism of the time profiles during the 2015-2017 period and complete the validation with accuracy assessment through the comparison between EPS vegetation products and ground data upscaled to a $3 \times 3 \text{ km}^2$ area representative of the main biome types and phenological conditions.

4.4.2 Comparison with local estimates

In this section the comparison between local and global LAI estimates is performed. The local processing chain provided high-resolution (30 and 10 m) 2015 and 2016 LAI maps which were resampled to the spatial resolution of the global maps (1 km). In addition, global maps were clipped to match with the extent of the corresponding high-resolution imagery. The comparison between local and global LAI estimates revealed high correlations and good agreement between the Landsat-7/8 and Sentinel-2A LAI aggregated estimates derived from the local processing chain and the GEOV1 and MOD15A2 estimates. In general, the MODIS and GEOV1 products fit well with the aggregated maps derived from the local processing chain as proved by determination coefficients higher than 0.9 in all cases and years (see Fig. 4.7).

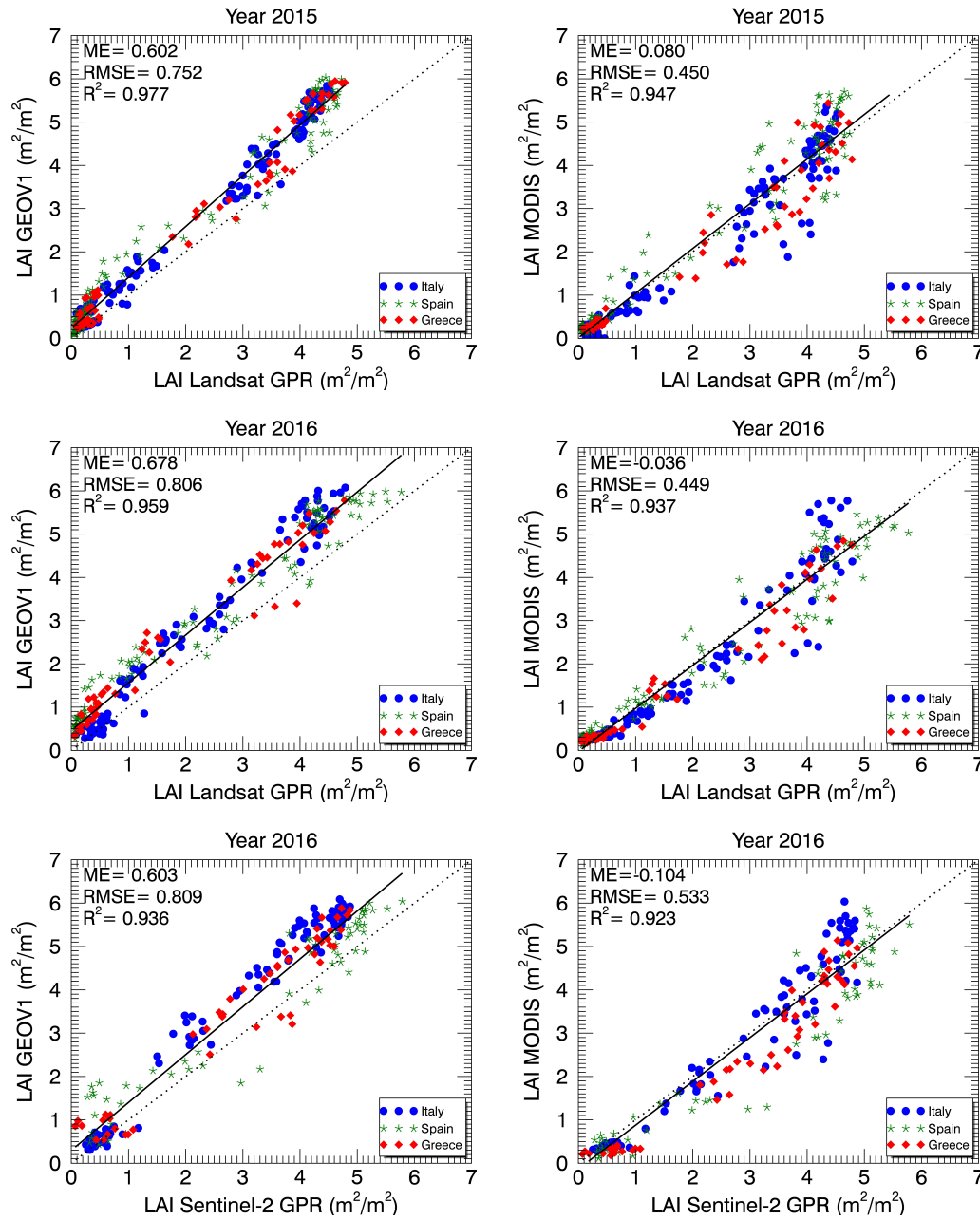


Figure 4.7: Scatter plots of estimated LAI values in the 2015 (top) and 2016 (middle and bottom) rice season over the Spanish, Italian and Greece rice sites.

It is worth mentioning that global products present an overestimation regarding the local ones for the periods of maximum development (high LAI values). These differences are partly due to differences in LAI definition: MOD15A2 and GEOV1.4 are closer to *actual* LAI for high values while the aggregated Sentinel-2A and Landsat-7/8 LAI estimates rely on a turbid radiative transfer model and are thus closer to *effective* LAI. Eventually, Fig. 4.8 shows time series of a representative rice pixel over each study area for the three ERMES rice seasons. Note that in 2014, the Landsat-7/8 estimates provided by the local processing chain were considered for testing purposes and consequently were not delivered

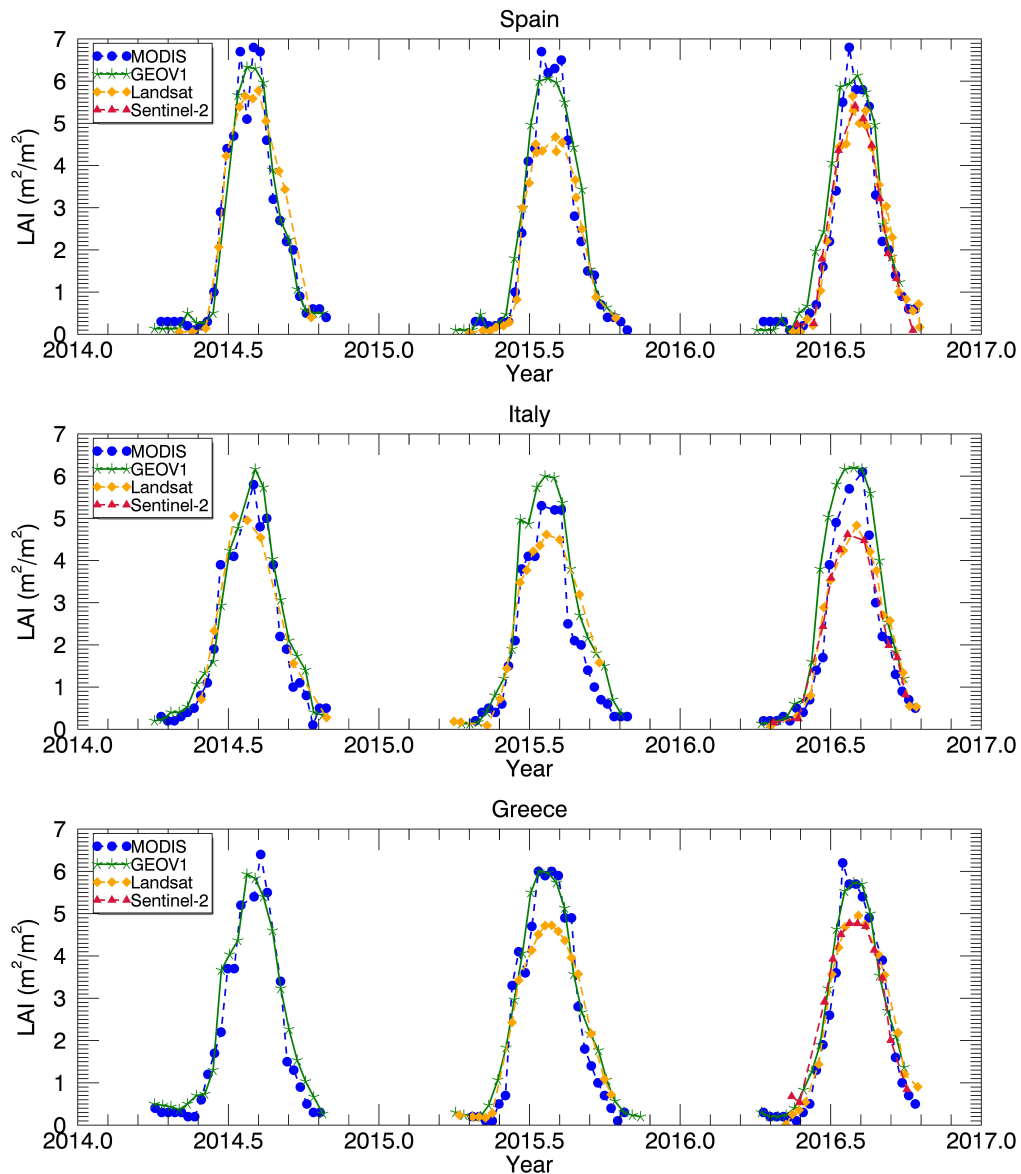


Figure 4.8: Temporal evolution of representative rice pixels of MOD15A2 and GEOV1.4 products and the resampled Landsat-7/8 and Sentinel-2A LAI estimates over Spain (top), Italy (middle) and Greece (bottom).

operationally. Overall evolutions are consistent and differences are mainly regarding the different algorithms for LAI estimation and parameter definition mainly for high values.

4.5 Concluding remarks

This chapter described the developed processing chain at global scale. The first part of the chapter reviewed the main operational remote sensing biophysical products at global scale: MOD15A2, GEOV1 and VEGA. Then, the EPS inputs of the retrieval chain at global scale were described as well as the multioutput regression algorithm used for the PROSAIL inversion. In this case, the PROSAIL was run to simulate a spectral database of world wide

variety of vegetation and soil types. The chain was implemented in the Land-SAF system which has currently initiated the production chain to derive global vegetation parameters at kilometric spatial resolution using AVHRR/MetOP data every 10 days.

In the last part of the chapter, a validation exercise was performed during 2015 and 2016 years in order to compare the estimates with the ones provided by MOD15A2, GEOV1 and VEGA over the BELMANIP-2.1 network. Results showed the EPS retrieval chain provided correlated estimates with the available operational products. In addition, LAI estimates provided by the local processing chain were intercompared with the global products revealing high consistency between estimates over Spain, Italy and Greece.



5. Conclusion and Discussion

This Thesis presented an operational remote sensing processing chain for deriving biophysical parameters at both local and global scale. The developed chain provides biophysical parameter estimates using hybrid methods through the generation of a reflectance and associated biophysical parameters database from the PROSAIL RTM and powerful nonlinear inversion methods, such as neural networks, kernel ridge regression and Gaussian process regression. Despite the differences between the local and global retrieval characteristics, the chain proposed in this Thesis is fully modular and can be adapted to each scale easily. The chain is able not only to retrieve biophysical parameters of a wide range of vegetation types, but also estimates of specific Earth crop by simply adapting the PROSAIL parametrization in order to fit simulations to the crop of interest.

As a matter of fact, decametric spatial resolution data were used at local scale (Landsat-7/8, SPOT5 and Sentinel-2A) for deriving HR LAI estimates over local rice areas in three different countries (Spain, Italy and Greece) during two years (2015 and 2016) in the framework of the European FP7 ERMES project. Similarly, the processing chain at global scale was implemented in EUMETSAT Land-SAF project and successfully integrated in the Land-SAF testing chain which is currently producing near real time vegetation products from AVHRR-3/Metop-B data. After its upgrade to operational, it is foreseen the reprocessing of the EPS products from 2007 onwards to obtain homogeneous time series of Climate Data Records (CDRs), ensuring the continuity of EPS observations from Metop-A until the end of life cycle of the future Metop-C.

Both results at local and global scales revealed the operational nature of the developed processing chain. In particular, direct validation conducted at local scale with *in situ* measurements acquired in three rice areas during different seasons proved the accuracy of the estimates retrieved by the chain. At global level, the estimates were validated over selected sites located around the planet covering different biomes by intercomparison with reference biophysical products such as MOD15A2, GEOV1 and VEGA. The intercomparison showed high consistency between them which highlighted the chain robustness. According to the attained results, we can confirm the hypothesis of the present work raised in the beginning of the Thesis. Nevertheless, we would like to discuss some key aspects:

On the retrieval algorithm

Operational biophysical products provide climate data records (i.e. long time periods) based on the fusion of products through a neural network, in the case of GEOV1, and based on LUT inversion in the case of MODIS. Over the years, these products have been updated, leading to more accurate estimates. However, the core algorithm of each product remains unchanged.

As evidenced in the literature, GPR outperforms other machine learning algorithms such as NN or KRR and also empirical parametric regressions based on vegetation indices. Moreover, GPR provides advantages for biophysical estimation and mapping. For example, Since GPR relies on a Bayesian framework, it provides probabilistic outputs which implies not only to derive a mean estimate for a prediction but also and associated uncertainty. This uncertainty is helpful a) for the scientist as a product to understand where potential errors in the retrieval exist, and b) for expert users, such as crop modelers, in order to weight estimates in the model biophysical assimilation process according to their goodness. The prediction uncertainty can also be used for diagnosing the presence of surfaces not addressed in the simulated database. Therefore, inspection of the uncertainty may provide with insights to (1) refine the selection of inputs, such as certain backgrounds not initially included in the preliminary retrieval, and (2) identify non-vegetated areas and targets for which the model was not calibrated and/or performed well. This feature allows to create a quality flag indicator useful for users and crop modelers.

From a pure operational point of view, NN have been used in biophysical processing chains such is the case of GEOV1. For its part, up to the present kernel-based methods have been considered only in experimental studies and none of them have been implemented into operational applications yet. The processing chain developed in this Thesis have conducted for the first time the implementation of GPR in operational chains at both global and local scale for deriving spatially and time-resolved estimations (not just static estimates) of biophysical parameters. GPR was implemented in ERMES for deriving a single biophysical parameter, while in the framework of Land-SAF three vegetation variables were derived since GPR has the flexibility of retrieving multiple vegetation variables simultaneously with the same model.

On the problems of inverting RTMs

Inverting RTMs is very complex because the number of unknowns is generally larger than the number of independent radiometric information. Also, estimating physical parameters from RTMs is hampered by the presence of high levels of uncertainty and noise, primarily associated to atmospheric conditions and sensor calibration, sun angle, and viewing geometry, as well as the poor sampling of the parameter space in most of the applications. This translates into inverse problems where deemed similar spectra may correspond to very diverse solutions which lead to ill-posed problems.

This kind of ill-posed problems can be alleviated by constraining the RTM simulations to specific type of vegetation and cropping systems. In addition, a correct background spectra of the simulated vegetation type must be introduced in the simulations, otherwise inaccurately estimates may be retrieved. From a remote sensing point of view, vegetation changes due to land cover, land use and agro-practises determine a strong change in the soil background conditions which may vary from dry soil to standing water. This fact generates an uncontrolled reflectance signature which may confound the retrieval of biophysical

parameters if those features are unrepresented by radiative transfer models. In addition, the background reflectance component is most influential for sparse or low vegetation coverage. It is therefore relevant to test methods for this specific situation.

For example, the experiment conducted in Chapter 2 considered a scenario where the training database does not characterize the rice fields flooded conditions, the retrieved LAI was unrealistically high after the flooding during the first crop development stages, but as LAI increases canopy closure minimizes the importance of spectral background in the retrieval. Similarly, when the training database does not characterize the dry conditions, the retrieved LAI was also unrealistically high before the real flooding of the fields. That assessment evidences that a correct characterization of the soil conditions in the training database is mandatory in order to obtain time series of realistic LAI estimates and avoid meaningless values in the first growing stages. These unrealistic effects produced by wrong spectra background can be very critical when time series analysis of rice crop dynamics is conducted. For example, the increase of estimated LAI at the beginning of the crop season would determine an unrealistic simulation of crop growth and identification of a wrong parameters in crop modelling such as sowing date or emergence period.

On the derived maps

The near-real time production of LAI maps is useful in planning the management practices (i.e. fertilization) to minimize the yield pattern variability within each parcel. This information is particularly important for precision farming activities, since farmers are expecting to be supported in prescription maps production for top-dress fertilization. A dense temporal data set of biophysical maps is also fundamental to perform expert crop monitoring and/or improve crop model estimations exploiting assimilation techniques.

It is important to mention that, whenever no external information able to represent the real variability in the field is provided for crop modeling, model simulation will provide the same results. This is the case when the aim is to apply crop models in an operational way at a parcel level. It is in fact not possible to obtain micro-meteorological information able to provide information changing from field to field. Moreover, it is not realistic to have detailed soil maps that usually exist at regional level with a scale ranging from 1:25.000 to 1:100.000. In this case, if the sowing date and crop variety are the same, or slightly different, the only way to capture the real spatial-temporal changes in crop development and production is to assimilate exogenous observation of crop status such as the information provided by EO biophysical maps.

On the other hand, the production of consistent LAI, FAPAR and FVC maps at global scale over a long period of time and for all ecosystems of the terrestrial surface, is very useful to be used in coarse scale models including: regional and global climate modelling, weather forecasting and global change monitoring, environmental management and land use, hydrology and drought assessment.

On the *in situ* measurements

A critical point when deriving estimates is their validation. Validation can be achieved with comparison between estimates and ground measurements. Ground data for validation activities are typically acquired by using non-destructive ground measurements with classical instrumentation such as plant canopy analyzers, digital hemispherical photography techniques or ceptometers. This Dissertation assessed and used for the first time the capabilities of smartphones as ground truth measuring tools for acquiring multitemporal

LAI estimates at field level, thus providing with new insights for LAI mapping and validation making use of new technologies in geosciences. LAI estimates at local scale were validated using ground data acquired with smartphones using a dedicated application (PocketLAI) for LAI estimation during coordinated field campaigns in the framework of ERMES.

We would like to mention that ground data collected in field campaigns is the most valuable information for assessing remote sensing derived information and understand possible biases and discrepancies. Nonetheless, even ground data could lead to erroneous conclusions if associated measurement errors are not taken into account and properly treated. Indirect methods for *in situ* measurements provide LAI_{eff} associated with several sources of measurement error including performance of instruments, illumination conditions, simplification of leaf optical properties, suboptimal spatial sampling within an ESU and saturation of optical signal in dense canopies. Specifically, observed variability in rice ground data when LAI_{eff} values are greater than 4 typically correspond to rice plant in the reproductive phenology stage. At this point, there is a significant change in the rice morphological structure due to the panicle emergence, leading to an increasing variability of the estimates. It is worth mentioning that error, bias and correlation between instruments used in this Thesis are small, and do agree with previous studies in different crops.

Future work

The proposed processing chain was successfully implemented as described in this Thesis. Furthermore, the modularity of the chain opens future research and applications. Future work could consider the application of the chain to ongoing remote sensing missions to cope with new satellite and sensors features at all scales. In addition, the GPR hybrid inversion methodology could be extended by proposing a statistical method blending both *in situ* and simulated data from RTMs by learning how to trade off noise variance in the *in situ* and simulated data (Camps-Valls et al., 2017; Svendsen et al., 2017). In the case of jointly estimate multiple parameters, composite covariances linking spatial-temporal relationships between outputs could be designed and learned for more accurate and robust retrievals (Luengo et al., 2016). In this scenario, we will focus on developing GPR models able to encode multi-sensor and multi-temporal information via the design of cross-correlation kernels.

Achievements and relevance

The Thesis directly related published papers are attached in the Annex. In addition, the work conducted in this Thesis has been also presented on several international conferences and indirectly involved in other scientific publications:

Journal papers

1. F. J. García-Haro, **M. Campos-Taberner**, N. Sabater, F. Belda, A. Moreno, M.A. Gilabert, B. Martínez & J. Meliá. (2014). Vulnerabilidad de la vegetación a la sequía en España. *Revista de Teledetección*, 42, 29–37, [doi:10.4995/raet.2014.2283](https://doi.org/10.4995/raet.2014.2283).
2. **M. Campos-Taberner**, A. Romero-Soriano, C. Gatta, G. Camps-Valls, A. Lagrange, B. Le Saux, A. Beaupère, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Ran-

- drianarivo, M. Ferecatu, M. Shimoni, G. Moser & D. Tuia. (2016). Processing of Extremely High-Resolution LiDAR and RGB Data: Outcome of the 2015 IEEE GRSS Data Fusion Contest–Part A: 2-D Contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5547 - 5559, 2016, doi:10.1109/JSTARS.2016.2569162.
3. J. M. Lopez-Sanchez, F. Vicente-Guijalba, E. Erten, **M. Campos-Taberner** & F. J. García-Haro, (2017). Retrieval of Vegetation Height in Rice Fields using Polarimetric SAR Interferometry with TanDEM-X Data. *Remote Sensing of Environment*, 192, 30–44, 2017, doi:10.1016/j.rse.2017.02.004.
 4. L. Busetto, S. Casteleyn, C. Granell, M. Pepe, M. Barbieri, **M. Campos-Taberner**, R. Casa, F. Collivignarelli, R. Confalonieri, A. Crema, F. J. García-Haro, L. Gatti, I. Z. Gitas, A. González-Pérez, G. Grau-Muedra, T. Guarneri, F. Holecz, D. Katsantonis, C. Minakou, I. Miralles, E. Movedi, F. Nutini, V. Pagani, A. Palombo, F. D. Paola, S. Pascucci, S. Pignatti, A. Rampini, L. Ranghetti, E. Ricciardelli, F. Romano, D. G. Stavrakoudis, D. Stroppiana, M. Viggiano & M. Boschetti. (2017). Downstream services for rice crop monitoring in Europe: from regional to local scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1–19, doi:10.1109/JSTARS.2017.2679159.

Conference papers

1. **M. Campos-Taberner**, F.J. García-Haro, F. Camacho, G. Camps-Valls, M.A. Gilabert, B. Martínez, A. Moreno, S. Sánchez-Ruiz & J. Meliá. (2014). Prototyping of physically based methods to retrieve leaf area index and canopy water content from satellite data. *Proceedings of the Fourth International Symposium Recent Advances in Quantitative Remote Sensing*, 7–10, Ed. José A. Sobrino, ISBN: 978-84-370-9808-1, September 22–26, Torrent (Valencia), Spain.
2. F.J. García-Haro, F. Camacho, B. Martínez, **M. Campos-Taberner** & J. Meliá. (2014). Continuous validation and algorithm improvement in LSA SAF FVC and LAI operational products over Africa. *Proceedings of the Fourth International Symposium Recent Advances in Quantitative Remote Sensing*, 426-430, Ed. José A. Sobrino, ISBN: 978-84-370-9808-1, September 22–26, Torrent (Valencia), Spain.
3. F.J. García-Haro, C. Gevaert & **M. Campos-Taberner**. (2014). Crop cycle monitoring by combining medium and high resolution optical imagery. *Proceedings of the Fourth International Symposium Recent Advances in Quantitative Remote Sensing*, 283–287, Ed. José A. Sobrino, ISBN: 978-84-370-9808-1, September 22–26, Torrent (Valencia), Spain.
4. M.A. Gilabert, A. Moreno, B. Martínez, S. Sánchez-Ruiz, **M. Campos-Taberner** & F.J. García-Haro. (2015). Flujos de carbono en el parque natural de la Albufera a alta resolución. *Proceedings of the 2015 XVI Congreso de la Asociación Española de Teledetección*, ISBN: 978-84-608-1726-0, 444–447, Sevilla, Spain.

5. A. Moreno, M.A. Gilabert, S. Sánchez-Ruiz, B. Martínez, F.J. García-Haro, **M. Campos-Taberner** & A. Carrara. (2015). Repercusión de la eficiencia en el uso de la radiación en la incertidumbre de los productos de GPP diaria. *Proceedings of the 2015 XVI Congreso de la Asociación Española de Teledetección*, ISBN: 978-84-608-1726-0, 476–479, Sevilla, Spain.
6. A. Moreno, M.A. Gilabert, S. Sánchez-Ruiz, B. Martínez, F.J. García-Haro, **M. Campos-Taberner** & A. Carrara. (2015). Un método sencillo para estimar flujos de carbono a partir del NDVI en la Península Ibérica. *Proceedings of the 2015 XVI Congreso de la Asociación Española de Teledetección*, ISBN: 978-84-608-1726-0, 476–479, Sevilla, Spain.
7. **M. Campos-Taberner**, A. Romero, C. Gatta & G. Camps-Valls. (2015). Shared Feature Representations Of LiDAR And Optical Images: Trading Sparsity For Semantic Discrimination. *Proceedings of the 2015 IEEE Geoscience and Remote Sensing Symposium*, 4169–4172, Milan, Italy. [doi:10.1109/IGARSS.2015.7326744](https://doi.org/10.1109/IGARSS.2015.7326744).
8. **M. Campos-Taberner**, F.J. García-Haro, A. Moreno, M.A. Gilabert, B. Martínez, S. Sánchez-Ruiz & G. Camps-Valls. (2015). Development of an Earth Observation processing chain for crop bio-physical parameters at local scale. *Proceedings of the 2015 IEEE Geoscience and Remote Sensing Symposium*, 29–32, Milan, Italy. [doi:0.1109/IGARSS.2015.7325689](https://doi.org/10.1109/IGARSS.2015.7325689).
9. **M. Campos-Taberner**, F. J. García-Haro, R. Confalonieri, B. Martínez, A. Moreno, S. Sánchez-Ruiz, M.A. Gilabert, F. Camacho, M. Boschetti & L. Busetto. (2015). Intercomparison of instruments for measuring Leaf Area Index over rice. *Proceedings of the 2015 IEEE Geoscience and Remote Sensing Symposium*, 3389–3392, Milan, Italy. [doi:10.1109/IGARSS.2015.7326546](https://doi.org/10.1109/IGARSS.2015.7326546).
10. D. Luengo, **M. Campos-Taberner**, & G. Camps-Valls. (2016). Latent force models for earth observation time series prediction. *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*, 1–6. [doi:10.1109/MLSP.2016.7738844](https://doi.org/10.1109/MLSP.2016.7738844).
11. **M. Campos-Taberner**, G. Grau-Muedra, F. J. García-Haro, B. Martínez, S. Sánchez-Ruiz, M.A. Gilabert & R. Confalonieri. (2017). Use of smartphones as measuring tools for educational purposes in remote sensing. *11th International Technology, Education and Development Conference, INTED2017 Proceedings*, 5474–5478, València, Spain. [doi:10.21125/inted.2017.1279](https://doi.org/10.21125/inted.2017.1279).
12. B. Martínez, S. Sánchez-Ruiz, **M. Campos-Taberner**, G. Grau-Muedra, F. J. García-Haro & M.A. Gilabert. (2017). Spectroscopy experiences in post-graduate university education. *11th International Technology, Education and Development Conference, INTED2017 Proceedings*, 5591–5597, València, Spain. [doi:10.21125/inted.2017.1303](https://doi.org/10.21125/inted.2017.1303).

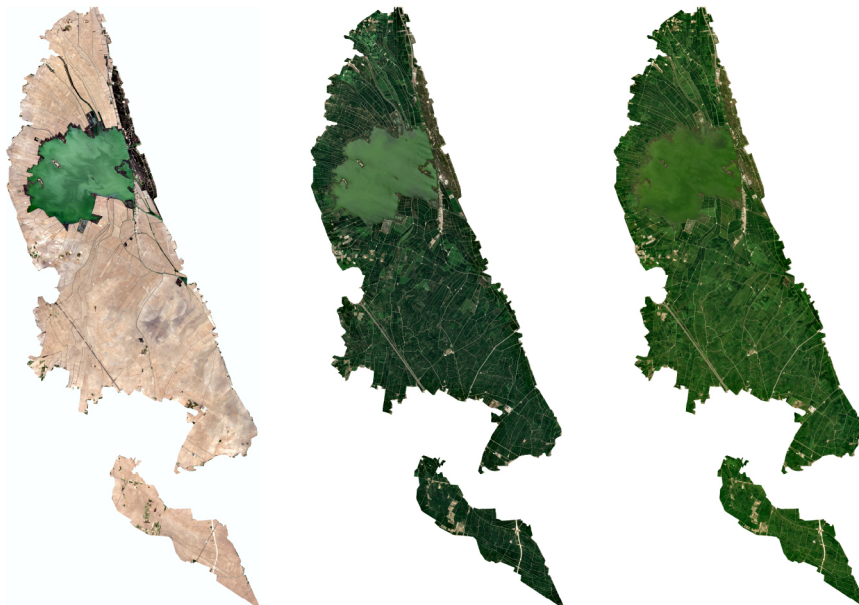
Competitions and awards

The paper *Shared feature representations of LiDAR and optical images: Trading sparsity for semantic discrimination* was awarded as best paper in the [2015 IEEE GRSS Data Fusion contest](#). The paper was presented in a special session at IGARSS 2015 in Milan.

Acknowledgements

The research activities leading to these Thesis results were supported and conducted in the framework of the following international projects:

- ERMES (an Earth obseRvation Model based ricE information Service) FP7 project, which received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement 606983.
- Land-SAF (Land Surface Analysis Satellite Applications Facility) project, which is supported by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT).



6. Summary in Valencian

6.1 Estructura

Aquesta tesi es presenta com a compendi de publicacions científiques, la qual cosa requereix en l'estructura de la tesi la inclusió d'un mínim de tres articles ja publicats o acceptats en revistes indexades internacionals¹. Així doncs, a l'Annex d'aquest treball es presenten quatre articles ja publicats en revistes indexades internacionals que estan directament relacionats amb el treball realitzat durant aquesta tesi. A més, amb la finalitat d'impulsar l'ús de la llengua pròpia de la Universitat de València en l'activitat docent i investigadora, aquesta tesi inclou un ampli resum en valencià dels objectius, metodologia, resultats i conclusions que es descriuen en detall a continuació.

6.2 Motivació i objectius

La vegetació és la base de la nostra vida sobre el planeta Terra. La vegetació juga una funció essencial en l'estudi del canvi climàtic global influïnt en l'intercanvi de CO₂ a través de la respiració de les plantes i la fotosíntesi. La inspecció de la vegetació ens proporciona informació valuosa per millorar el coneixement dels entorns naturals quantificant l'ús i l'estat del dosser vegetal a escala local i global.

Els mètodes clàssics de control i seguiment de la vegetació no resulten eficaços avui en dia degut al fet que requereixen molt de temps d'execució, estan desactualitzats i sovint resulten massa cars. L'avanç tecnològic en el camp de la indústria espacial ha permès desenvolupar mètodes per detectar les propietats de vegetació i la seva dinàmica mitjançant teledetecció des de sensors espectrals a bord d'avions i de satèl·lits. Aquesta tecnologia interdisciplinària ofereix cada cop més solucions pràctiques i econòmiques per avaluar els canvis produïts en la vegetació sobre àrees extenses a resolucions espacials que van des dels kilòmetres fins als decàmetres.

L'estat de la vegetació pot ser avaluat emprant eines de teledetecció a partir de l'estimació de paràmetres biofísics com el LAI (per les seues sigles en anglès, Leaf Area Index), el FAPAR (per les seues sigles en anglès, Fraction of Absorbed Photosyn-

¹Reglament sobre depòsit, avaluació i defensa de la tesi doctoral aprovat pel Consell de Govern de 28 de Juny de 2016. ACGUV 172/2016.

thetically Active Radiation) i l'FVC (per les seues sigles en anglès, Fractional Vegetation Cover). Tanmateix, l'estimació de paràmetres biofísics mitjançant dades de teledetecció és un problema desafiant. Hom pot trobar en la literatura que l'estimació de dits paràmetres ha estat duta a terme mitjançant diverses aproximacions: a través de relacions empíriques entre el paràmetre biofísic d'interès i índexs de vegetació; utilitzant mètodes de regressió purament estadístics, i invertint RTMs (per les seues sigles en anglès, Radiative Transfer Models) mitjançant LUTs (per les seues sigles en anglès, Look-Up Tables) o tècniques d'aprenentatge automàtic (en anglès, machine learning).

Aquests mètodes són aplicats per la comunitat científica utilitzant diferents dades espectrals, espacials i temporals des de plataformes de teledetecció. Malgrat tot, no hi ha un marc comú per a l'estimació de paràmetres biofísics a escala local i global. En aquest context, els objectius d'aquesta tesi són el desenvolupament, la validació i la implementació de manera operacional, d'una cadena de processament de dades d'observació de la Terra per tal d'estimar paràmetres biofísics amb un mètode de regressió híbrid tant a escala local com global.

6.3 Metodologia

De del punt de vista metodològic i conceptual, la cadena de processament desenvolupada en aquest treball conté els mateixos components a escala local i global en un model d'estimació híbrid.

En primer lloc, s'obté una base de dades que conté simulacions de reflectivitat de la vegetació i els seus paràmetres biofísics associats executant un model de transferència radiativa en mode directe. Seguidament, les mostres simulades s'utilitzen per a l'entrenament de models de regressió no lineals i no paramètrics. El millor model en termes de precisió, biaix i bondat de l'ajustament s'escull per tal de ser utilitzat en la cadena de processament. Una vegada entrenat i seleccionat el millor model, les imatges de reflectivitat de la superfície terrestre obtingudes per teledetecció des de satèl·lits s'utilitzen com a dades d'entrada en el procés d'inversió. Aquest procés d'inversió proporciona estimacions dels paràmetres biofísics d'interès tant a escala local com global depenent de la resolució espacial i cobertura de les imatges. Finalment, la validació dels paràmetres estimats a escala local es realitza mitjançant la comparació amb mesures *in situ* adquirides durant campanyes de camp en tres països durant les temporades de conreu d'arròs en 2015 i 2016. Pel que fa a la validació de les estimacions a escala global, aquesta és realitzada mitjançant la comparació amb productes biofísics operacionals de teledetecció.

A continuació descrivim els principals components de la cadena així com les seues característiques més rellevants per a l'estimació de paràmetres biofísics.

6.3.1 Paràmetres biofísics: LAI, FAPAR i FVC

L'estimació de paràmetres biofísics de la vegetació a partir de dades espectrals de teledetecció ha estat objecte de nombrosos estudis a causa de la seua rellevància en aplicacions mediambientals i agronòmiques. Els paràmetres biofísics més utilitzats per aquestes aplicacions són el LAI, el FAPAR i l'FVC.

El LAI és un índex que representa la quantitat d'àrea foliar per unitat d'àrea de superfície que abasta la vegetació que es vol estudiar. El LAI juga un paper rellevant en processos com la fotosíntesi i s'utilitza tant en models climàtics com en estimació de

collita. EL FAPAR és un paràmetre que dona compte de la radiació solar que s'absorbeix per la vegetació (verda) en el rang des dels 400 nm fins als 700 nm. EL FAPAR és una variable adimensional que pren valors des de 0 (sòl nu) fins a 1 (vegetació molt densa) i depèn de les condicions d'il·luminació i de l'estructura de la coberta vegetal. El FAPAR s'utilitza en models d'intercanvi de carboni per a estimar la producció primària neta. Per la seua part, l'FVC es defineix com la fracció de vegetació verda vista des de la direcció de nadir i té el mateix rang de variació que el FAPAR. Contràriament al FAPAR, l'FVC no depèn de la geometria d'il·luminació i s'utilitza en models per diferenciar entre vegetació i sòl.

6.3.2 Models de transferència radiativa: PROSAIL

La interacció entre la radiació i el dosser vegetal es pot modelar mitjançant la teoria de transferència radiativa. Aquesta teoria utilitza models matemàtics per tal de descriure la interceptió de llum per les plantes, i la relació entre la reflectivitat de la vegetació i els paràmetres biofísics que la caracteritzen.

Durant les últimes dècades hom pot trobar a la literatura una àmplia varietat d'RTMs proposats per tal de modelar la interacció de la radiació amb el dosser vegetal. Entre ells, el PROSAIL és el més emprat. El PROSAIL és un RTM que resulta de l'acoblament dels models PROSPECT a nivell de fulla i SAIL (per les seues sigles en anglès, Scattering by Arbitrary Inclined Leaves) a nivell de dosser. Aquest acoblament s'aconsegueix introduint la reflectància i la transmissivitat simulades pel PROSPECT com a dades d'entrada en el SAIL. D'aquesta manera s'aconsegueix simular la reflectància del dosser vegetal en el rang des dels 400 nm fins als 2500 nm tenint en compte tant les propietats bioquímiques com les característiques geomètriques d'il·luminació i observació de la vegetació simulada.

Les propietats òptiques a nivell de fulla descrites pel PROSAIL són tingudes en compte mitjançant un paràmetre estructural mesòfil (N) i els continguts en clorofil·la (C_{ab}), matèria seca (C_m) i aigua (C_w). A més a més, si en l'acoblament s'utilitza la versió 5 del PROSPECT, també es tenen en compte els pigments marrons (C_{bp}) i els carotenoides (C_{ar}). L'estructura del dosser està caracteritzada pels paràmetres del SAIL: ALA (per les seues sigles en anglès, Average Leaf Angle), LAI i Hotspot (paràmetre que té en compte el màxim de radiació difusa en la direcció de retrodispersió). La geometria del sistema ve donada per l'angle solar zenital (θ_s), l'angle d'observació zenital (θ_v) i l'angle relatiu entre ambdós ($\Delta\theta$). Per tal de representar diferents espectres de fons, es pot tindre en compte també un paràmetre multiplicatiu de brillantor (β_s) que s'aplica als espectres de sòl simulats. De manera similar, es pot introduir el paràmetre $\nu Cover$ per tindre en compte condicions en les quals es vol simular una escena formada per una mescla de vegetació i sòl.

La distribució triada per mostrejar cada paràmetre dins el PROSAIL determinarà les signatures espectrals simulades. Per tant, la cadena de processament permet simular reflectància corresponent tant a un tipus de vegetació concret com a una gran varietat de tipus i estats de vegetació alhora.

6.3.3 Algorismes d'inversió

Una vegada construïda la base de dades simulades, és a dir, l'execució del PROSAIL en mode directe, el següent pas en la cadena de processament consisteix en la inversió de les simulacions. La inversió es du a terme mitjançant algorismes no lineals d'aprenentatge

automàtic. L'aprenentatge automàtic és un camp multidisciplinari que ha esdevingut una de les eines claus d'aplicacions tan diverses com el processament de senyals i d'imatges, la medicina i el reconeixement de patrons. La idea principal es basa en el fet que un algorisme pot aprendre a partir d'unes dades com realitzar una tasca sense haver estat explícitament programat per a ajustar-se a aquesta tasca.

En teledetecció l'aprenentatge automàtic ha estat emprat en aplicacions que inclouen la classificació d'imatges, la detecció de canvis i l'estimació de paràmetres biofísics i atmosfèrics. Aquests algorismes són generalment robusts i, en la majoria dels casos, són ràpids d'aplicar una vegada han estat entrenats. A més a més, són capaços de treballar amb possibles relacions no lineals entre els paràmetres biofísics i la reflectància associada sense utilitzar relacions paramètriques.

Entre els mètodes més populars, hom pot trobar les xarxes neuronals, (NN per les seues sigles en anglès, Neural Networks) i els mètodes nucli (coneguts en anglès com kernel methods). Les NN han estat utilitzades des dels anys noranta en l'estimació de paràmetres biofísics. Una NN és essencialment una estructura de neurones plenament connectada que s'organitza en capes. Una neurona bàsicament realitza una regressió lineal seguida per l'aplicació d'una funció d'activació no lineal (p. ex., una sigmoide). Les neurones de capes diferents s'interconnecten amb enllaços ponderats. La xarxa queda formalment definida pel patró d'interconnexió entre diferents capes de neurones, el procés d'aprenentatge que actualitza els pesos dels enllaços, i per la funció d'activació que converteix les entrades pesades per les neurones amb la seua activació de sortida. L'estructura més comuna de les NN és la coneguda com "alimentació cap endavant", on la informació flueix des dels nodes d'entrada (p. ex., reflectàncies), a través dels nodes ocults (si hi ha) cap als nodes de sortida (p. ex., paràmetre biofísic). Pel que fa a l'entrenament, el mètode més utilitzat és el conegut com algorisme de "retro propagació" que, bàsicament, consisteix en l'ajust dels pesos després d'haver-los escollit inicialment a l'atzar, fins que l'error és suficientment petit.

Per altra banda els mètodes nucli com el KRR (per les seues sigles en anglès, Kernel Ridge Regression) i el GPR (per les seues sigles en anglès, Gaussian Process Regression) es basen en la mesura de semblances entre mostres mitjançant funcions nucli. El càlcul de la semblança es realitza mitjançant un mapatge de les dades d'entrada en un espai de característiques de dimensió molt major a través d'un producte interior (escalar). Hom no necessita saber de manera explícita el mapatge, atès que el producte interior es pot substituir per una funció de semblança entre les mostres que es coneix com nucli. El KRR és la versió nucli de la regressió no lineal regularitzada que realitza una regressió no lineal de mínims quadrats en l'espai de Hilbert. L'objectiu és minimitzar una funció quadràtica de pèrdua regularitzada respecte dels pesos.

En el cas del GPR, la seua utilització ens proporciona una aproximació probabilística per a problemes de regressions amb nuclis. En aquest cas, la regressió està basada en la definició d'una distribució (gaussiana) sobre funcions que està completament descrita per una predicció mitjana i una funció de covariància; és a dir, el nucli. La funció de covariància del GPR defineix una distribució sobre els valors de la variable de sortida seguint una distribució gaussiana multivariant, i ens dona la informació de semblança entre mostres o valors d'entrada. El GPR estableix una relació entre les mostres d'entrada i la variable d'eixida a partir d'uns pesos obtesos en l'entrenament i el nucli. En els mètodes nucli la selecció adequada del nucli és clau, ja que el nucli ve definit per uns

hiperparàmetres que han de ser optimitzats per validació creuada en el cas del KRR i maximitzant la distribució marginal de les mostres d'entrenament en el cas del GPR. En aplicacions de teledetecció, la utilització del GPR ens proporciona alguns avantatges com són la rellevància de les bandes espectrals en l'estimació del paràmetre d'interès i un nivell d'incertesa per a cada predicció que es pot interpretar com una confiança de la predicció.

6.3.4 Dades de teledetecció

El següent punt a considerar en la cadena de processament és la utilització d'imatges de teledetecció com a dades d'entrada en els models entrenats a partir de les simulacions realitzades amb el PROSAIL. Durant aquesta tesi s'han utilitzat diverses dades de teledetecció a escala local i global. En primer lloc, i en el marc de treball del projecte ERMES s'han utilitzat imatges de reflectància de superfície de Landsat-7/8, SPOT-5 i Sentinel-2A a escala local. En concret, les dades de Landsat-7/8 proporcionen informació espectral en sis canals (blau, verd, roig, infraroig proper i dos canals en l'infraroig d'ona curta) a 30 m de resolució espacial. SPOT-5 proporciona dades en quatre bandes (verd, roig, infraroig proper i un canal en l'infraroig d'ona curta) a 10 m de resolució espacial, i Sentinel-2A és capaç de proporcionar dades a 10 m, 20 m i 60 m de resolució espacial depenent de les bandes espectrals que abasten des del visible, passant per l'infraroig proper fins a l'infraroig d'ona curta. La resolució temporal dels satèl·lits és de 16, 5 i 10 dies en cas de Landsat-7/8, SPOT-5 i Sentinel-2A respectivament.

A escala global, la cadena de processament ha estat desenvolupada per a treballar amb dades del sensor AVHRR a bord del satèl·lit MetOp en el marc de treball del projecte LandSAF. AVHRR/MetOp proporciona dades globals cada 10 dies a 1 km de resolució espacial en les bandes del roig, infraroig proper i un canal en l'infraroig d'ona curta. La reflectància proporcionada per AVHRR/MetOp s'obté a partir d'una funció de la distribució de la reflectància bidireccional que descriu els valors de reflectància bidireccional tenint en compte totes les combinacions possibles en la geometria de captura de la imatge (posició solar i posició del sensor). Aquests models assumeixen que la reflectància es pot expressar com una combinació lineal de funcions angulars i tres paràmetres lliures (k_0 , k_1 , k_2) on k_0 és la reflectància capturada en geometria nadiral.

6.4 Resultats i validació

L'aplicació dels models utilitzant les dades de teledetecció descrites en l'apartat anterior, ha donat com a resultat l'estimació de paràmetres biofísics tant a escala local com global. A escala local, la cadena de processament ha estat utilitzada de manera operacional dins el projecte ERMES per tal d'estimar LAI sobre les principals zones arrosseres de la conca del Mediterrani durant els anys 2015 i 2016. La cadena de processament va incloure la construcció d'una base de dades de reflectància i LAI mitjançant una parametrització específica del PROSAIL. Aquesta parametrització es va fonamentar en mesures de camp obtingudes durant l'any 2014 sobre plantacions d'arròs. Açò es va dur a terme per tal d'ajustar les simulacions a la reflectància de les plantes d'arròs durant les fases del seu cicle fenològic sota les pràctiques agrícoles comuns en les zones d'estudi. Amb aquestes simulacions s'entrenaren els models KRR, GPR i NN, i resultà el GPR com el millor model en termes estadístics. Conseqüentment, el GPR va ser emprat per estimar LAI utilitzant imatges multitemporals de reflectància adquirides per Landsat-7/8, SPOT-5 i Sentinel-2A,

i donà com a resultat mapes decamètrics de LAI a escala local.

Així mateix, l'aplicació de la cadena de processament dins el projecte LandSAF ha permès obtenir estimacions conjuntes de LAI, FAPAR i FVC a escala global durant els anys 2015 i 2016. Convé esmentar que, en aquest cas, la parametrització del PROSAIL es va ajustar per tal de simular un ampli ventall de signatures espectrals corresponents a molts tipus de vegetació diferents amb l'objectiu de modelar el màxim possible la vegetació que cobreix la Terra. De manera similar al cas local, s'entrenaren els tres models mencionats anteriorment en les seues versions multi sortida; és a dir, capaçs d'estimar al mateix temps LAI, FAPAR i FVC amb un únic entrenament. Les NN són un mètode multi sortida *per se* a causa de la seua estructura de capes i connexions. En el cas dels mètodes nucli, les versions multi sortida s'obtenen de la utilització d'un únic nucli que conté els hiperparàmetres optimitzats per als tres paràmetres alhora. En el cas del KRR s'optimitzen per validació creuada, mentre que en el cas del GPR l'optimització és duu a terme mitjançant el mètode de màxima versemblança. Comparant tots els algorismes (tant uni com multi sortida), el GPR en versió multi sortida va ser el més precís, i així doncs, utilitzat per derivar conjuntament cada deu dies mapes globals dels tres paràmetres biofísics a 1 km de resolució espacial.

6.4.1 Mesures *in situ*

En aquesta tesi, s'han planificat i dut a terme una sèrie de campanyes de camp durant els anys 2014, 2015 i 2016 per tal de mesurar LAI en plantacions d'arròs a Espanya, Itàlia i Grècia. Cal destacar que les mesures *in situ* de LAI es recolliren utilitzant telèfons intel·ligents (en anglès smartphones) mitjançant una aplicació mòbil anomenada PocketLAI. Aquesta aplicació mesura el LAI d'una manera simple i ràpida. Una vegada engegada l'aplicació, l'operador selecciona el mode mesura i disposa de cinc segons per col·locar el dispositiu en la part més baixa de la planta i subjectar-lo en posició vertical. Transcorreguts els cinc segons, l'aplicació activa una vibració per fer saber l'operador que s'està mesurant l'orientació de la pantalla respecte de la vertical, la qual cosa s'aconsegueix en temps real mitjançant l'acceleròmetre de què disposa el telèfon intel·ligent. En aquest instant l'operador rota el dispositiu lentament, i quan l'angle entre la vertical i la normal a la pantalla abasta els 57.5° , l'aplicació captura una imatge que és automàticament processada per un algorisme intern que distingeix entre píxels de vegetació i cel per estimar LAI. A més, el telèfon emet una segona vibració per fer notar l'operador que la mesura s'ha pres correctament. Així doncs, el PocketLAI fa mesures de LAI mitjançant fotografies a 57.5° fonamentant-se en un model simplificat de transmitància de la llum sota la suposició d'una distribució espacial aleatòria de fulles infinitament petites.

Durant les campanyes de camp de 2014 es va realitzar un estudi de comparació entre mesures *in situ* de LAI preses amb el PocketLAI i instrumentació clàssica com el LAI-2000 i càmeres fotogràfiques hemisfèriques DHP (per les seues sigles en anglès, Digital Hemispherical Photography). Aquest estudi de comparació va concloure que les mesures adquirides en les plantacions d'arròs amb el PocketLAI són molt similars a les proporcionades pels instruments clàssics, tot i que el PocketLAI presenta una lleugera infraestimació. A la llum d'aquests resultats, i dels avantatges d'utilitzar telèfons intel·ligents (poc pes, alta manejabilitat i fàcil ús) el PocketLAI va ser escollit com l'instrument de mesura per a les campanyes de 2015 i 2016.

6.4.2 Validació

Les estimacions proporcionades per teledetecció poden ser validades de manera directa mitjançant mesures de camp dels paràmetres estimats. La validació és un aspecte clau a l'hora de proporcionar de manera operacional un producte de teledetecció, a fi que els usuaris puguin determinar quin és el producte o combinació de productes més adients per a les seues aplicacions.

Aquestes mesures *in situ* de LAI han estat utilitzades per validar la cadena de processament a escala local. La validació s'ha fet comparant espacialment (a nivell de píxel) els valors de les estimacions de LAI derivats per la cadena de processament amb les dades *in situ* mesurades sobre els mateixos píxels. Els resultats estadístics de la validació mostraren errors quadràtics mitjans al voltant de mig punt en unitats de LAI i coeficients de determinació propers a la unitat en les tres zones durant els anys 2015 i 2016. A més, l'evolució temporal de les estimacions de LAI seguiren de manera semblant la de les mesures *in situ*. Aquests resultats destaquen la precisió, correlació i consistència entre les estimacions i les mesures de camp.

A escala global la cadena de processament ha estat validada comparant les estimacions de LAI, FAPAR i FVC amb les corresponents estimacions proporcionades pels principals productes operacionals de paràmetres biofísics (MOD15A2, GEOV1 i VEGA). La comparació ha estat duta a terme utilitzant els valors estimats durant dos anys consecutius sobre la versió 2.1 d'una xarxa de llocs distribuïts per tot el planeta anomenada BELMANIP (per les seues sigles en anglès, Benchmark Land Multisite Analysis and Intercomparison of Products). Aquesta xarxa de llocs ha estat emprada per molts estudis de comparació d'estimacions de paràmetres biofísics a escala global, atès que inclou un total de 445 llocs localitzats sobre àrees relativament planes i homogènies representant tot tipus de vegetació al llarg del planeta. La comparació de les estimacions obtingudes per la cadena de processament tragué a la llum l'alta correlació i consistència entre els productes a escala global i per biomes.

Per a finalitzar, el mapes de LAI obtinguts a escala local a 30 i 10 m van ser remostrejats a la resolució espacial dels productes globals per tal de fer una comparació entre les estimacions. En general, les estimacions dels productes locals agregats a 1 km s'ajustaren bé amb els globals. Tanmateix, una petita infraestimació per part dels productes locals es va observar en el període de màxim desenvolupament de la planta d'arròs. Aquestes diferències són esperades i s'expliquen tenint en compte els algorismes i les definicions de LAI que cada producte proporciona: les estimacions locals de LAI proporcionen un LAI *efectiu* mentre que els productes globals proporcionen estimacions de LAI més properes a un LAI *actual*.

6.5 Conclusions

En aquesta tesi hem desenvolupat, descrit i avaluat experimentalment una cadena de processament operacional per estimar paràmetres biofísics del dosser vegetal a escala local i global. Els paràmetres estimats s'obtenen de l'aplicació d'un mètode híbrid que consisteix en la inversió de models de transferència radiativa mitjançant algorismes d'aprenentatge automàtic.

La cadena de processament desenvolupada és completament modular, la qual cosa permet adaptar-la a escala local i global d'una manera senzilla sense comprometre con-

ceptualment i metodològicament els fonaments de la mateixa. Des d'un punt de vista operacional, la cadena ha estat utilitzada a escala local per estimar LAI en tres països durant les temporades de conreu de l'arròs en 2015 i 2016. Així mateix, la cadena ha estat provada per estimar conjuntament LAI, FAPAR i FVC a nivell global durant dos anys sencers. Actualment, la seua implantació i aplicació fa que s'estiguin produint estimacions globals dels tres paràmetres biofísics disponibles per als usuaris a través del portal web de LandSAF (<https://landsaf.ipma.pt/>). Així mateix, aquesta tesi avalua i presenta per primera vegada el GPR com l'algorisme emprat en un procés d'inversió implementat en una cadena de processament operacional de paràmetres biofísics des de teledetecció.

Les estimacions de LAI a escala local van ser validades satisfactòriament amb mesures *in situ* preses per telèfons intel·ligents mitjançant la utilització d'una aplicació especialment dissenyada per a mesurar LAI al camp. Per la seua banda, les estimacions a escala global han sigut validades per comparació amb els principals productes biofísics existents. Cal remarcar que durant els processos de validació hom ha de tindre en compte possibles errors en les mesures de camp per defugir conclusions errònies. Aquests errors estan associats principalment a fallades dels instruments de mesura, condicions d'il·luminació i mostreig no òptimes, i efectes de saturació en vegetació densa. Quan es tracta de comparar amb productes existents s'ha de tenir en compte les possibles diferències a causa de les deficiències i composicions temporals de cada producte així com la projecció espacial i resolució òptima on fer la comparació.

Cal destacar que la inversió de models de transferència radiativa és una tasca molt complexa que pot plantejar problemes atès que espectres similars poden correspondre a diverses solucions, és a dir, a diversos valors del paràmetre a estimar. Malgrat això, una encertada parametrització de la vegetació que es vol simular basada en informació exògena i la utilització d'espectres de sòl característics condueix a una millora en la precisió i estabilitat de les estimacions. Per tal d'alleugerir aquest problema, en aquesta tesi hem emprat unes parametritzacions adients tant a escala local com global basades en dades de camp.

Pel que fa a la utilitat dels mapes biofísics derivats de l'aplicació de la cadena de processament convé destacar que la seua inclusió en models tant locals com globals proporciona una informació clau per millorar les activitats agrícoles, les modelitzacions del clima i els efectes del canvi climàtic, la gestió ambiental i l'ús del sòl, i les avaluacions de períodes de sequera.

A tall de cloenda, el desenvolupament del treball realitzat durant aquesta tesi obre la porta a futures recerques i aplicacions en el camp de les estimacions de paràmetres biofísics des de nous sensors i plataformes de teledetecció. A més, es poden desenvolupar noves tècniques híbrides de regressió combinant dades *in situ* i simulacions, com també dissenyar nuclis capaços de tenir en compte relacions espaciotemporals i relacions creuades entre els paràmetres a estimar.

Annex: Scientific Publications

The achievements and conclusions of this work have been published on high quality international journals as research papers :

Publication I

Campos-Taberner, M., Garcia-Haro, F. J., Moreno, A., Gilabert, M. A., Sanchez-Ruiz, S., Martinez, B. & Camps-Valls, G., “Mapping Leaf Area Index With a Smartphone and Gaussian Processes”, *Geoscience and Remote Sensing Letters, IEEE*, 12(12), 2501-2505 , 2015. [doi:10.1109/LGRS.2015.2488682](https://doi.org/10.1109/LGRS.2015.2488682)

Publication II

Campos-Taberner, M., García-Haro, F. J., Confalonieri, R., Martínez, B., Moreno, Á., Sánchez-Ruiz, S., Gilabert, M.A., Camacho, F., Boschetti, M. & Busetto, L., “Multitemporal Monitoring of Plant Area Index in the Valencia Rice District with PocketLAI”, *Remote Sensing*, 8(3), 202. 2016. [doi:10.3390/rs8030202](https://doi.org/10.3390/rs8030202)

Publication III

Campos-Taberner, M., García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A. & Boschetti, M., “Multitemporal and Multiresolution Leaf Area Index Retrieval for Operational Local Rice Crop Monitoring”, *Remote Sensing of Environment*, 187, 102 - 118, 2016. [doi:10.1016/j.rse.2016.10.009](https://doi.org/10.1016/j.rse.2016.10.009)

Publication IV

Campos-Taberner, M., García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Busetto, L., Katsantonis, D., Stavrakoudis, D., Minakou, C., Gatti, L., Barbieri, M., Holecz, F., Stroppiana D. & Boschetti, M. (2017). “Exploitation of SAR and optical Sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index”, *Remote Sensing*, 9(3), 248, 2017. [doi:10.3390/rs9030248](https://doi.org/10.3390/rs9030248)

Mapping Leaf Area Index With a Smartphone and Gaussian Processes

Manuel Campos-Taberner, Franciso Javier García-Haro, Álvaro Moreno, María Amparo Gilabert, Sergio Sánchez-Ruiz, Beatriz Martínez, and Gustau Camps-Valls, *Senior Member, IEEE*

Abstract—Leaf area index (LAI) is a key biophysical parameter used to determine foliage cover and crop growth in environmental studies. Smartphones are nowadays ubiquitous sensor devices with high computational power, moderate cost, and high-quality sensors. A smartphone app, which is called PocketLAI, was recently presented and tested for acquiring ground LAI estimates. In this letter, we explore the use of state-of-the-art nonlinear Gaussian process regression (GPR) to derive spatially explicit LAI estimates over rice using ground data from PocketLAI and Landsat 8 imagery. GPR has gained popularity in recent years because of its solid Bayesian foundations that offer not only high accuracy but also confidence intervals for the retrievals. We show the first LAI maps obtained with ground data from a smartphone combined with advanced machine learning. This letter compares LAI predictions and confidence intervals of the retrievals obtained with PocketLAI with those obtained with classical instruments, such as digital hemispheric photography (DHP) and LI-COR LAI-2000. This letter shows that all three instruments obtained comparable results, but PocketLAI is far cheaper. The proposed methodology hence opens a wide range of possible applications at moderate cost.

Index Terms—Biophysical parameter retrieval, Gaussian processes (GPs), leaf area index (LAI), smartphone.

I. INTRODUCTION

THE estimation of biophysical parameters from remote sensing data is a key issue for monitoring crop properties. Leaf area index (LAI) has been defined as the total one-sided leaf area in relation to the ground or the total foliage surface area per unit of horizontally projected ground surface area [1]. Two main approaches have been used to quantify LAI from the ground, either *direct* or *indirect*. Direct methods require an effort in collecting an optimal sample size and estimating plant density, which involves destructive harvest techniques [2]. Indirect methods estimate LAI using optical instruments through the computation of the radiation transmitted or the canopy gap fraction [3]. Sensors such as LAI-2000 Plant Canopy Analyzers

(LI-COR, Inc., Nebraska, USA) measure the gap fraction from five different angles simultaneously. Digital hemispherical photography (DHP) is another indirect technique for computing the gap fraction through cameras with coupled hemispherical lenses (fish-eye). This method estimates LAI from measurements of the gap fraction, which is defined as the fraction of sky seen from below the canopy (upward photography) or the fraction of soil seen from above (downward photography). Both plant canopy analyzers and DHP are some of the most widely used classical optical instruments for indirect LAI estimation [2], [3]. Indirect methods actually compute an effective LAI (LAI_{eff}). In this letter, LAI estimates refer to effective LAI values. The difference between the actual LAI and the effective LAI may be quantified by the clumping index Ω through $LAI_{eff} = \Omega \times LAI$ [4]. The clumping index is almost always less than 1, with the exemption of very regularly spaced leaf distributions.

An alternative sensor device to estimate LAI may be currently in our hands. Smartphones are becoming an accessible daily taken instrument for most of the population.¹ The high adoption rate of smartphones in today's society, along with the increase in computational power and sensing capabilities, is being exploited in many fields of science and engineering. The use of smartphone components such as Global Positioning System (GPS), camera, accelerometer, and core processing power makes them suitable for a number of purposes, including methods for indirect LAI estimation. Smartphone capabilities are growing day by day, making them possible future measuring instruments. Recently, a mobile application (PocketLAI) has been successfully introduced for LAI estimation [5], which was further tested on paddy rice against commercial instruments, such as LI-COR LAI-2000 and Decagon AccuPAR ceptometer, and during an entire rice crop season against DHP and LAI-2000 in [6].

Several methods such as physical, statistical, empirical, and hybrid methods have been used to deal with the biophysical parameter estimation [7]. In this letter, we will face the problem of LAI estimation following a modern statistical approximation. Statistical methods based on parametric approaches, such as vegetation indices, use explicit parametric equations and need prior physical knowledge. On the other hand, nonparametric

Manuscript received October 1, 2015; accepted October 6, 2015. Date of publication October 30, 2015; date of current version November 11, 2015. This work was supported in part by the European Union Seventh Framework Program (FP7/2007-2013) under Grant 606983, the European Research Council under ERC-CoG-2014 SEDAL grant, and the Spanish Ministry of Economy and Competitiveness (MINECO) under funding for projects TIN2012-38102-C03-01 and RESET CLIMATE 337 (CGL2012-35831).

M. Campos-Taberner, F. J. García-Haro, A. Moreno, M. A. Gilabert, S. Sánchez-Ruiz, and B. Martínez are with the Department of Earth Physics and Thermodynamics, Faculty of Physics, Universitat de València, 46100 València, Spain (e-mail: manuel.campos@uv.es).

G. Camps-Valls is with the Image Processing Laboratory, Universitat de València, 46980 València, Spain (e-mail: gcamps@uv.es).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2015.2488682

¹In 2014, the number of global users of mobile phones surpassed that of desktop computers. Even in underdeveloped countries, the use of smartphones rises at a much faster rate. Beyond worldwide adoption of this technology, interaction with the smartphones also increases: The average smartphone user downloads three apps per month.

methods do not need, in principle, any prior about the relationship between data, but they infer those relations directly from data analysis. In general, parametric methods for biophysical parameters retrieval become less effective in terms of accuracy, bias, and goodness of fit than nonparametric methods [8].

In the framework of the nonparametric methods, Bayesian approaches have become an alternative proposal to other machine learning methods such as neural networks [9] or support vector machines [10], [11]. GPR [12] has been widely used for biophysical parameters estimation in many remote sensing studies, including chlorophyll content retrieval [13], solar irradiation [14], and vegetation properties [8]. Aside from the robustness and stability, one of the characteristics that make GPR a particular useful tool is the combination of very good prediction accuracy and the ability to provide confidence intervals for the estimates.

The remainder of this letter is organized as follows. Section II briefly describes the theory of the GPR. The methodology followed in this letter is outlined in Section III, describing both the ground data set and the Landsat 8 images. Section IV discusses the results obtained, and finally, Section V concludes this letter with a discussion and outline of the future research.

II. GPR

Standard regression approximates observations (often referred to as *outputs*) $\{y_n\}_{n=1}^N$ as the sum of some unknown latent function $f(\mathbf{x})$ of the inputs $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$ plus *constant power* Gaussian noise, i.e., $y_n = f(\mathbf{x}_n) + \varepsilon_n, \varepsilon_n \sim \mathcal{N}(0, \sigma^2)$. Instead of proposing a parametric form for $f(\mathbf{x})$ and learning its parameters in order to fit observed data well, GPR proceeds in a Bayesian nonparametric way. A zero-mean² Gaussian process (GP) prior is placed on the latent function $f(\mathbf{x})$, and a Gaussian prior is used for each latent noise term ε_n , $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k_\theta(\mathbf{x}, \mathbf{x}'))$, where $k_\theta(\mathbf{x}, \mathbf{x}')$ is a covariance function parameterized by θ , and σ^2 is a hyperparameter that specifies the noise power. Essentially, a GP is a stochastic process whose marginals are distributed as a multivariate Gaussian. In particular, given the priors \mathcal{GP} , samples drawn from $f(\mathbf{x})$ at the set of locations $\{\mathbf{x}_n\}_{n=1}^N$ follow a joint multivariate Gaussian with zero mean and covariance matrix $\mathbf{K}_\mathbf{ff}$, with $[\mathbf{K}_\mathbf{ff}]_{ij} = k_\theta(\mathbf{x}_i, \mathbf{x}_j)$.

If we consider a test location \mathbf{x}_* with corresponding output y_* , \mathcal{GP} defines a joint prior distribution between the observations $\mathbf{y} \equiv \{y_n\}_{n=1}^N$ and y_* . Collecting available data in $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \dots, N\}$, it is possible to analytically compute the posterior distribution over the unknown output y_* , i.e.,

$$\begin{aligned} p(y_* | \mathbf{x}_*, \mathcal{D}) &= \mathcal{N}(y_* | \mu_{\text{GP}*}, \sigma_{\text{GP}*}^2) \\ \mu_{\text{GP}*} &= \mathbf{k}_{\mathbf{f}*}^\top (\mathbf{K}_\mathbf{ff} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} = \mathbf{k}_{\mathbf{f}*}^\top \boldsymbol{\alpha} \\ \sigma_{\text{GP}*}^2 &= \sigma^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^\top (\mathbf{K}_\mathbf{ff} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_{\mathbf{f}*}. \end{aligned}$$

GPs offer some advantages over other regression methods. Since they yield a full posterior predictive distribution over

y_* , it is possible to obtain not only mean predictions for test data, i.e., $\mu_{\text{GP}*}$, but also the so-called “error bars,” i.e., $\sigma_{\text{GP}*}^2$, assessing the uncertainty of the mean prediction. The whole procedure only depends on a very small set of hyperparameters, which combats overfitting efficiently. In addition, the inference of the hyperparameters and the weights $\boldsymbol{\alpha}$ can be performed using continuous optimization of the evidence. Note, however, that the bottleneck of the algorithm is the definition of the covariance (kernel, or Gram) function k : this function should capture the similarity between data instances. A standard widely used covariance function is the *isotropic* squared exponential, i.e., $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2))$, which captures sample similarity in most of the data problems efficiently. In this letter, we use the GPML MATLAB toolbox for the experiments³ [12], which is also available along with many other regression methods in the simpleR MATLAB toolbox.⁴

III. DATA COLLECTION AND METHODOLOGY

This section covers the data collection of both field data and remote sensing images in the study and reviews the adopted methodology for training and validating the GPR models.

A. Field Data

GPR has been trained with the ground data acquired during the 2014 ERMES (An Earth obseRvation Model based information RicE Service) field campaign in Spain.⁵ LAI measurements were taken over selected farms within the rice district of Sueca (39° 16' N, 0° 18' W) situated in the Albufera Natural Park (south of Valencia City, east of Spain). The area has a typical Mediterranean climate, mild, with an average annual humidity of 65%. The average annual temperature is 17 °C. Their mean values range from 11 °C in January and 27 °C in August. The mean annual precipitation is approximately 430 mm, which tends to be intense and concentrated in autumn.

The site is a homogeneous rice planting area of approximately 10 km × 20 km extension. Most of the paddy rice fields are rectangular and flat. The rice cropping practices are common on all the rice districts. The field campaign was carried out on 26 elementary sampling units (ESUs) from June 17 to September 8 during ten days, covering the entire rice season. Measurement dates were selected matching with Landsat 8 overpasses. ESUs were located at least 30 m away from the field borders and were approximately 20 m × 20 m in size. The center of the ESU was geolocated using a GPS. Over each ESU, 16 photographs were acquired with the DHP and subsequently processed using the CAN-EYE software. In this letter, the standard procedure for DHP data processing suggested in [15] was followed. LI-COR LAI-2000 was also used to estimate LAI by making three replications of one reading above and eight below the canopy for each measurement and ESU. In addition, LAI was also acquired with PocketLAI. A representative LAI measurement acquired with the smart app was obtained averaging 18 single measurements over an ESU. The mobile application was installed on

² It is customary to subtract the sample mean to data $\{y_n\}_{n=1}^N$ and then to assume a zero-mean model.

³ <http://www.Gaussianprocess.org/gpml/code/matlab/doc/>

⁴ <http://www.uv.es/gcamps/code/simpleR.html>

⁵ <http://www.ermes-fp7space.eu/>



Fig. 1. (Left) RGB compositions of the two Landsat path/rows (198/33 and 199/33). (Right) Clipped image covering the rice area.

a Samsung Galaxy S4 GT-I9505, with a Quad-Core 1.9-GHz processor and 2-GB random access memory. The PocketLAI uses the smartphone's camera for taking images with a resolution of 4128×3096 pixels.

During the field campaign, five more ESUs per day were identified as nonvegetated land covers such as bare soils, water bodies, and roads. This approach was done in order to represent $\text{LAI} \approx 0$ ESUs in the training set to avoid possible wrong mean estimates and very low confidence values retrieved by the GPR [16].

B. Landsat 8

The Operational Land Imager (OLI) is a multispectral sensor on board the Landsat 8 satellite. The spatial and temporal resolutions of Landsat 8 OLI images provide valuable information for crop monitoring at local scale [17]. Recently, the United States Geological Survey has facilitated free access to Landsat archive data. In this work, Landsat 8 Surface Reflectance product was used for deriving high-resolution LAI maps. Since the rice area lies in two Landsat path/rows (198/33 and 199/33), images were available every seven and nine days rather than the usual interval of 16 days. Six Landsat 8 OLI bands (B, G, R, NIR, SWIR1, and SWIR2) were selected to relate the surface reflectance with LAI measurements. Landsat 8 OLI images were clipped to 1500×800 pixel size covering the entire rice area (see Fig. 1). Pixels covering urban areas, sea, and the lagoon were masked out during the retrieval process in order to avoid meaningless LAI estimates over those surfaces.

C. Training and Testing

During the entire rice growing season, LAI was measured simultaneously using three instruments: PocketLAI, DHP, and LI-COR LAI-2000. Hereafter, we refer to each data set as LAI_{APP} , LAI_{DHP} , and LAI_{LIC} , respectively. The three data sets and the associated Landsat 8 OLI surface reflectance were divided into two different training (80%) and testing subsets (20%). An independent model was built for LAI_{APP} , LAI_{DHP} , and LAI_{LIC} . Each model was constructed by running GPR a hundred of times with different random selections of training and testing subsets. The testing subset was used for validation purposes evaluating the root-mean-squared error (RMSE) and the mean absolute error (MAE) to assess GPR accuracy. Mean error (ME) was used to evaluate the bias and coefficient of

determination (R^2) to account for the goodness of fit between predictions and measurements.

IV. EXPERIMENTAL RESULTS

This section reports the experimental results of the study, both qualitatively through the generated explicit space-resolved high-resolution maps and quantitatively through measures of accuracy, fit, and bias for the three LAI estimation tools.

A. High-Resolution Maps

Landsat 8 OLI-based LAI maps were derived for the rice district of Sueca (València), Spain. Fig. 2 shows the high-resolution maps providing both mean estimate and associated uncertainties maps for LAI_{APP} , LAI_{DHP} , and LAI_{LIC} . We display generated maps on July 31 (DoY = 212). Within-field variations are observed in all three LAI maps due to the following reasons: 1) the spatial LAI variability of rice fields corresponding to different varieties; 2) differences in plant phenological stages; and 3) low LAI values corresponding to nonvegetated areas and boundaries of the rice fields. In the rice fields, LAI estimates fell within the expected range at that phenological state. Second, LAI uncertainty (σ) maps show low values within the rice fields. Higher uncertainties values appear over zones corresponding to nonvegetated areas or low LAI estimates. Particularly, on the east side of the maps, low LAI values and higher uncertainties referred to a different type of vegetation land cover (trees) and several man-made surfaces can be detected. This effect was also observed in [18], which is essentially due to the fact that GPR cannot extrapolate outside the ranges seen in the data set. Maps based on DHP ground data are able to predict more variability in LAI and uncertainty than the rest, and hence better identify rice field boundaries, roads, farm buildings, etc. Nevertheless, map based on PocketLAI is very similar in terms of LAI variability to LAI-2000. LAI-2000 provides uncertainties slightly higher, but also higher values in terms of LAI. Uncertainty map values need to be interpreted as a confidence interval around the mean predictions. For this reason, we computed the coefficient of variation (ratio between LAI uncertainties and mean LAI predictions) $\text{CV} = \sigma/\mu * 100$, which provides relative uncertainties (see Fig. 2). These maps show that the majority of rice pixels fall below the 20% uncertainty threshold and can be considered well validated as proposed by the Global Climate Observing System [19].

B. Statistical Comparison

Table I reports the statistical indicators of the models' performance used for the LAI retrieval. GPR using LAI_{APP} outperforms LAI_{DHP} and LAI_{LIC} models in terms of accuracy and goodness of fit. However, LAI_{APP} bias is slightly higher than the other fits. The predictive mean estimates and associated predictive variances (uncertainties) between models were compared. Fig. 3 shows the comparison of the scatterplots of the mean LAI estimates and associated uncertainties between models. In general, estimated LAI values with GPR from PocketLAI data were correlated to the classical

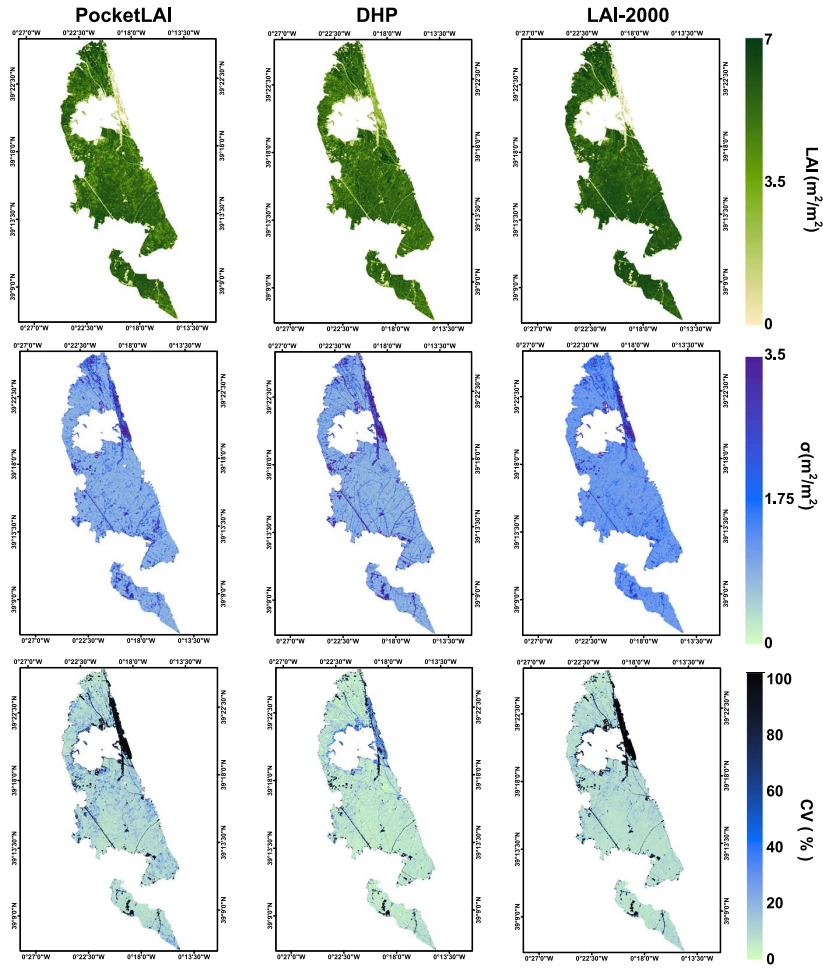


Fig. 2. (Top) LAI prediction maps, (middle) standard deviation maps, and (bottom) coefficient of variation, obtained with GPR using PocketLAI, DHP, and LAI-2000 ground data on the Landsat 8 image (DoY = 212).

TABLE I
MEAN VALUES OF THE STATISTICAL INDICATORS (RMSE, MAE, ABSOLUTE VALUE OF THE ME, AND R^2) BETWEEN ESTIMATED AND MEASURED LAI IN LAI_{APP}, LAI_{DHP}, AND LAI_{LIC} VALIDATION SUBSETS. STANDARD DEVIATIONS ARE SHOWN IN PARENTHESES

Dataset	RMSE	MAE	ME	R^2
LAI _{APP}	0.51(0.18)	0.35(0.19)	0.12(0.07)	0.94(0.07)
LAI _{DHP}	0.64(0.20)	0.43(0.19)	0.15(0.10)	0.87(0.09)
LAI _{LICOR}	0.62(0.21)	0.48(0.26)	0.13(0.08)	0.89(0.07)

instruments. Specifically, mean estimates retrieved when using PocketLAI showed similar results to those obtained with LAI-2000. Bias between LAI_{APP}-LAI_{LICOR}, LAI_{APP}-LAI_{DHP}, and LAI_{DHP}-LAI_{LICOR} was 0.06, 0.05, and 0.12 in LAI units, respectively. This bias is slightly smaller than others reported in the literature for crops [20]. A comparison between associated uncertainties provided by the different models showed that the majority of pixels fall close to the 1:1 line, suggesting good fits. Therefore, using data acquired from PocketLAI does not produce high differences in uncertainties when compared with those produced by DHP and LAI-2000. The statistical

confidence intervals of LAI predictions reliably identify areas that may be inaccurately mapped, such as man-made constructions and roads. Nevertheless, we would like to note that this is a conservative estimate of the LAI uncertainty since the error sources associated to experimental measurements and input reflectances are not well characterized. Modeling such error sources and designing proper GP priors are matters of active research and will be pursued in future studies.

V. DISCUSSION AND FUTURE WORK

This letter has introduced the use of a smartphone app, which is called PocketLAI, to estimate in situ LAI. To assess the usefulness of the estimates, we then combined remote sensing Landsat 8 images with state-of-the-art nonparametric GPR to generate spatially explicit high-resolution LAI maps. We compared LAI predictions and confidence intervals for the estimates using PocketLAI and other classical instruments, such as DHP and LI-COR LAI-2000. The obtained results suggested that the proposed combination of a smartphone app and GPR was an appropriate cost-effective approach, in terms of accuracy, bias, and goodness of fit.

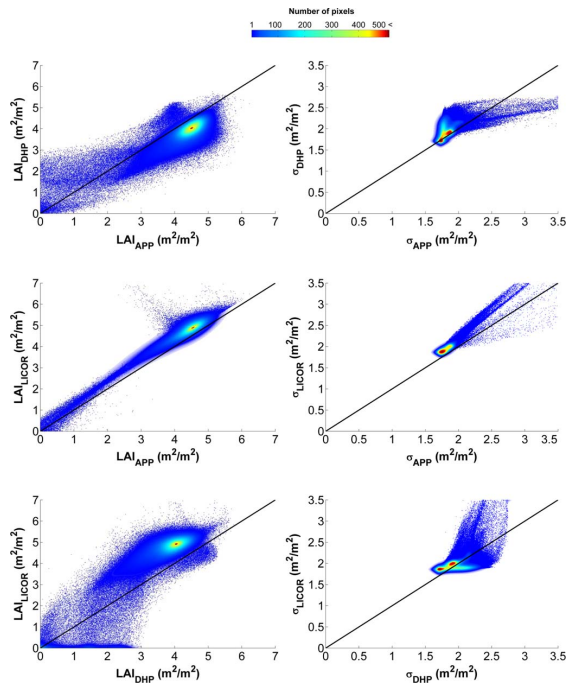


Fig. 3. Scatterplots of LAI (left) mean estimates and (right) uncertainty obtained with all three instruments.

We can conclude that, in general, the performance of the proposed approach based on the PocketLAI smart app has shown to be less expensive in terms of instrument and processing. This avoids errors due to processing from nonexpert operators and makes it suitable for operational LAI monitoring activities. As we have shown, the confidence map obtained with LAI_{APP} remained constant; thus, the upscaling process can be considered stable. These results allow considering PocketLAI as a powerful alternative to other commercial instruments not only for LAI monitoring during field campaigns but also for LAI retrieval. In any case, further studies are required on other crop types and using the app on other smartphones with different cameras, which could affect PocketLAI readings and hence LAI estimates.

PocketLAI may help in collecting a huge amount of data during the phenological cycle of vegetation, particularly over rice, due to the high portability and ease of handling. However, two shortcomings are observed here: the use of large data sets increases the computational effort needed for running GPRs, and the current GPR does not include any temporal information. Two methodological GPRs will be considered in the near future: sparse GPR models to cope with massive data and temporal GPR to deal with nonstationarities, scales, and trends of the acquired time series.

Finally, we would like to mention that, to our knowledge, this is the first work combining smartphones and machine learning for biophysical parameter retrieval. The results are really encouraging and open a wide field for experimentation and biophysical parameter retrieval at affordable cost, both in time and computational and human resources. The forthcoming free-of-charge Sentinel-2 data will be a perfect testbed for the proposed methodology.

© 2015 IEEE. Reprinted, with permission, from Manuel Campos-Taberner, Francisco Javier García-Haro, Álvaro Moreno, María Amparo Gilabert, Sergio Sánchez-Ruiz, Beatriz Martínez, and Gustau Camps-Valls, Mapping Leaf Area Index With a Smartphone and Gaussian Processes, DECEMBER 2015.

ACKNOWLEDGMENT

The authors would like to thank the ERMES and LSA SAF projects, the European Research Council, and the Spanish Ministry of Economy and Competitiveness.

REFERENCES

- [1] J. M. Chen and T. A. Black, "Defining leaf area index for non-flat leaves," *Plant, Cell Environ.*, vol. 15, no. 4, pp. 421–429, 1992.
- [2] N. J. J. Breda, "Ground-based measurements of leaf area index: A review of methods, instruments and current controversies," *J. Exp. Botany*, vol. 54, no. 392, pp. 2403–2417, 2003.
- [3] J. M. Welles and S. Cohen, "Canopy structure measurement by gap fraction analysis using commercial instrumentation," *J. Exp. Botany*, vol. 47, no. 9, pp. 1335–1342, 1996.
- [4] J. Chen and T. Black, "Foliage area and architecture of plant canopies from sunfleck size distributions," *Agricultural Forest Meteorol.*, vol. 60, no. 3/4, pp. 249–266, 1992.
- [5] R. Confalonieri *et al.*, "Development of an app for estimating leaf area index using a smartphone. trueness and precision determination and comparison with other indirect methods," *Comput. Electron. Agriculture*, vol. 96, pp. 67–74, 2013.
- [6] M. Campos-Taberner *et al.*, "Intercomparison of instruments for measuring leaf area index over rice," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Milano, Italy, 2015, pp. 3389–3392.
- [7] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, "Remote sensing image processing," *Synthesis Lectures Image, Video, Multimedia Process.*, vol. 5, no. 1, pp. 1–192, 2011.
- [8] M. Lazaro-Gredilla, M. Titsias, J. Verrelst, and G. Camps-Valls, "Retrieval of biophysical parameters with heteroscedastic Gaussian processes," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 838–842, Apr. 2014.
- [9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [10] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
- [11] G. Camps-Valls, J. Muñoz Mari, L. Gó-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semi-supervised support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 248–252, Feb. 2009.
- [12] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, Jan. 2006.
- [13] Y. Bazi, N. Alajlan, and F. Melgani, "Improved estimation of water chlorophyll concentration with semisupervised Gaussian process regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2733–2743, Jul. 2012.
- [14] S. Salcedo-Sanz, C. Casanova-Mateo, J. Muñoz-Mari, and G. Camps-Valls, "Prediction of daily global solar irradiation using temporal Gaussian processes," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1936–1940, Nov. 2014.
- [15] V. Demarez, S. Duthoit, F. Baret, M. Weiss, and G. Dedieu, "Estimation of leaf area and clumping indexes of crops with hemispherical photographs," *Agricultural Forest Meteorol.*, vol. 148, no. 4, pp. 644–655, 2008.
- [16] J. Verrelst, J. P. Rivera, J. Moreno, and G. Camps-Valls, "Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 157–167, 2013.
- [17] M. Campos-Taberner *et al.*, "Development of an earth observation processing chain for crop bio-physical parameters at local scale," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Milano, Italy, 2015, pp. 29–32.
- [18] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1832–1843, May 2012.
- [19] "Systematic observation requirements for satellite-based products for climate, 2011 update, supplemental details to the satellite-based component of the implementation plan for the global observing system for climate in support of the UNFCCC (2010 update, GCOS-154)," Global Climate Observing System (GCOS), Geneva, Switzerland, Tech. Rep., 2011. [Online]. Available: <http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf>
- [20] A. Verger, B. Martínez, F. C. de Coca, and F. J. García-Haro, "Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area," *Int. J. Remote Sens.*, vol. 30, no. 10, pp. 2685–2704, 2009.



Article

Multitemporal Monitoring of Plant Area Index in the Valencia Rice District with PocketLAI

Manuel Campos-Taberner ^{1,*}, Franciso Javier García-Haro ¹, Roberto Confalonieri ², Beatriz Martínez ¹, Álvaro Moreno ¹, Sergio Sánchez-Ruiz ¹, María Amparo Gilabert ¹, Fernando Camacho ³, Mirco Boschetti ⁴ and Lorenzo Busetto ⁴

¹ Department of Earth Physics and Thermodynamics, Faculty of Physics, Universitat de València, Dr. Moliner, Burjassot 46100, València, Spain; j.garcia.haro@uv.es (F.J.G.-H.); Beatriz.Martinez@uv.es (B.M.); alvaro.moreno@uv.es (A.M.); Sergio.Sanchez@uv.es (S.S.-R.); m.amparo.gilabert@uv.es (M.A.G.)

² Department of Agricultural and Environmental Sciences—Production, Landscape, Agroenergy, Cassandra Lab, Università degli Studi di Milano, Via Celoria 2, 20133 Milan, Italy; roberto.confalonieri@unimi.it

³ EOLAB, Parc Científic Universitat de València, Catedrático A. Escardino, 46980 Paterna, Spain; fernando.camacho@eolab.es

⁴ Institute for Electromagnetic Sensing of the Environment, Italian National Research Council, Via Bassini 15, 20133 Milan, Italy; boschetti.m@irea.cnr.it (M.B.); busetto.l@irea.cnr.it (L.B.)

* Correspondence: manuel.campos@uv.es; Tel: +34-963-543-256; Fax: +34-963-543-385

Academic Editors: Agnes Begue, Clement Atzberger and Prasad S. Thenkabail

Received: 14 December 2015; Accepted: 18 February 2016; Published: 1 March 2016

Abstract: Leaf area index (LAI) is a key biophysical parameter used to determine foliage cover and crop growth in environmental studies in order to assess crop yield. Frequently, plant canopy analyzers (LAI-2000) and digital cameras for hemispherical photography (DHP) are used for indirect effective plant area index (PAI_{eff}) estimates. Nevertheless, these instruments are expensive and have the disadvantages of low portability and maintenance. Recently, a smartphone app called PocketLAI was presented and tested for acquiring PAI_{eff} measurements. It was used during an entire rice season for indirect PAI_{eff} estimations and for deriving reference high-resolution PAI_{eff} maps. Ground PAI_{eff} values acquired with PocketLAI, LAI-2000, and DHP were well correlated ($R^2 = 0.95$, $RMSE = 0.21 \text{ m}^2/\text{m}^2$ for Licor-2000, and $R^2 = 0.94$, $RMSE = 0.6 \text{ m}^2/\text{m}^2$ for DHP). Complementary data such as phenology and leaf chlorophyll content were acquired to complement seasonal rice plant information provided by PAI_{eff} . High-resolution PAI_{eff} maps, which can be used for the validation of remote sensing products, have been derived using a global transfer function (TF) made of several measuring dates and their associated satellite radiances.

Keywords: rice; effective plant area index (PAI_{eff}); PocketLAI; smartphone; high-resolution map

1. Introduction

With the aim of managing plant needs in a more efficient way, precision agriculture has arisen as a rush of technological enhancements to classical farm management tools [1,2]. Detailed geo-spatial information on plant and soil properties is essential knowledge in crop management. In this context, remote sensing has become a very efficient tool for precision farming of large areas through data acquired by sensors on-board satellite platforms [3], airborne imagery [4], and unmanned aerial vehicles (UAVs) [5]. In this framework, rice cultivation is one of the most extended land uses for food production worldwide and has therefore been the main objective of many studies using optical [6,7] and radar [8–10] remote sensing techniques. In this context, leaf area index (LAI) is a key biophysical variable for both crop monitoring and modelling applications, defined as the total one-sided leaf area

in relation to the ground [11]. LAI has been used in agricultural and remote sensing studies [12,13], including precision agriculture [14], and is regarded as a key input in global models of ecosystem, hydrology, climate, ecology, biogeochemistry, and productivity [15].

In situ LAI measurement methods can be divided into two main categories: direct and indirect [16,17]. Direct methods require an effort in collecting an optimal sample size and estimating plant density, which involve destructive harvest techniques [18]. Direct and indirect methods are complementary, as direct LAI measurements can be used as a reference or calibration for indirect measurements. Indirect methods allow the inference of LAI from observations of another variable. They are generally faster than direct methods and allow larger spatial sample collection. Indirect methods can be divided into indirect contact LAI measurements and indirect non-contact measurements [19]. Indirect contact LAI methods are based on the estimation of the contact frequency [20], while indirect non-contact methods are based on the estimation of the gap fraction [21]. Contact frequency is the probability that a beam (radiation) penetrates inside the canopy and interacts with the vegetation. On the other hand, gap fraction is the integrated value of the gap frequency, which accounts for the probability that the beam will have no contact with the vegetation until it reaches the ground. Sensors like LAI-2000 or LAI-2200 Plant Canopy Analyzers (LI-COR, Inc., Nebraska, USA) measure the gap fraction from five different angles simultaneously. Digital hemispherical photography (DHP) is another indirect technique for computing the gap fraction through cameras with hemispherical lenses (fish-eye) coupled.

This method estimates LAI from measurements of the gap fraction, defined as the fraction of sky seen from below the canopy (upwards photography) or fraction of soil seen from above (downwards photography). Both Plant Canopy Analyzers and DHP are some of the most widely used classical optical instruments for indirect LAI estimation [19,22]. Classical commercial instruments have proven to be a good alternative to destructive methods in many experimental conditions [23] but these instruments are usually quite expensive both to purchase and to maintain. Due to their weight and strict requirements concerning acquisition protocol, they can also prove to be quite difficult to use in cases in which access to the canopy or placement of the instrument either below or above the canopy is difficult, such as in the case of flooded rice fields. Specifically, DHP techniques requires high effort by the operator during the classification process needed to obtain a LAI estimate. This fact limits the deployment of these measurements in near real time applications. An additional drawback when using these instruments is the time needed for repairing them in case of damage. A delay in the repair or replacement of the instrument may lead to canceling the field campaign in the worst case.

Recently, in the context of exploiting the technology implemented in smartphones for studies dealing with natural sciences, we introduced a mobile application called PocketLAI for leaf area index estimation [24]. PocketLAI was already successfully tested against Decagon AccuPAR Ceptometer, and it was used to measure LAI both on rice and other crop types also deviating from ideal assumptions of the light transmittance model used [25]. Smartphones are becoming an accessible daily instrument for most of the population. The use of smartphone components such as global position system (GPS), camera, accelerometer, and core processing power makes them suitable for a number of purposes, including methods for indirect LAI estimation. Smartphone capabilities are growing day by day, making them a reliable alternative to classical measuring instruments.

Leaf area index estimates refer only to leaf elements of the plant. Destructive methods only allow for measurement of leaves [26], but when dealing with indirect methods, several important considerations should be taken into account for a proper definition of the measured variable. In particular, for *in situ* LAI acquisitions and remote sensing observations, if no distinction is made between leaves and other plant elements, the proper term to use is PAI (Plant Area Index) rather than LAI [27–29]. Canopies are made of green photosynthetically active elements and other elements which are not green and therefore non-photosynthetically active (senescent leaves, trunks, branches, fruits, and flowers). Hence, to represent the photosynthetic functionality of all elements of the plant, PAI should be corrected to GAI (Green Area Index) [30]. Nevertheless, a proper indirect

determination of GAI requires optical instruments able to distinguish green from non-green elements within the canopy [19]. Since this study deals with measurements and estimates taking into account all elements of the rice plant during all phenological stages, the term PAI will be used throughout the manuscript.

Temporal information on crop status is a requirement for better crop monitoring used to support agronomical management. In this context, other crop parameters, such as leaf chlorophyll content (Chl) or plant phenological stages should be acquired in order to complete temporal information of the crop. The identification of phenological stages can be addressed by visual interpretation of the plant morphological characteristics. This procedure can be carried out following standardized protocols, such as the Biologische Bundesanstalt Bundessortenamt and Chemical industry (BBCH) [31], which provides a description of the major morphological characteristics of each plant development stage assigning a specific numerical code for each one. An example of using the BBCH scale to identify phenological stages over rice fields can be found in [32]. On the other hand, chlorophyll content provides information about the physiological status of plants, nutrient stress, photosynthesis, and growing periods [33,34]. Chlorophyll concentration may change throughout different stages of plant phenology and is affected when crop plants are under stress conditions, mainly due to changes in soil nitrogen content [35]. Thus, leaf chlorophyll content becomes a key issue for agronomists and farmers to make management decisions at critical stages and has been widely studied by the remote sensing community [36–38]. Direct field measurements of chlorophyll content over large areas require a big effort in collecting destructive samples and conducting laboratory chemistry methods. Conversely, the use of handheld devices, such as SPAD-502 (Minolta Osaka Company, Ltd., Japan), are being used for rapid non-destructive sampling of leaf chlorophyll content [36,39]. For these purposes, continuous seasonal field phenology and leaf chlorophyll content were measured to complement seasonal rice plant information provided by PAI acquisitions.

In situ bio-physical parameters acquired during field campaigns are usually used for the validation of coarse satellite-derived products [40–42]. For these purposes, field measurements must follow several good practices and protocols [43]. Validation is commonly addressed through the derivation of a reference high-resolution map of an area covering several moderate resolution pixels [44,45]. The broadest strategy for the validation of satellite products is based on a bottom-up approach: it starts from the scale of the individual measurements that are aggregated over an elementary sampling unit area (ESU) corresponding to a support area consistent with that of the high-resolution imagery used for the up-scaling of ground data [44,46].

The main aim of this study is to monitor the rice plant area index with PocketLAI, assessing its consistency and performance with LAI-2000 and DHP during an entire rice season. The analysis of PocketLAI performance includes inter-comparison of field measurements and derived (upscaled) maps, which can be used for continuous monitoring and validation of LAI products. This study contributes to instrument comparison studies of rice.

The remainder of this study is organized as follows: Section 2 describes the study area and the sampling strategy outlining the *in situ* PAI measurements and describing the instruments used. Section 3 discusses the obtained results and the ancillary bio-physical variables used for rice monitoring, and finally Section 4 concludes the paper with a discussion and outline of the future research.

2. Materials and Methods

2.1. Study Area and Field Campaign

The study area is located in selected farms of the rice district of Sueca (39°16'N, 0°18'W), situated in the south of Valencia, in Eastern Spain (see Figure 1). The area has a typical Mediterranean climate; mild, with an average annual humidity of 65%. The average annual temperature is 17 °C. Their

mean values range from 11 °C in January, and 27 °C in August. The mean annual precipitation is approximately 430 mm, tending to be intense and concentrated in autumn.

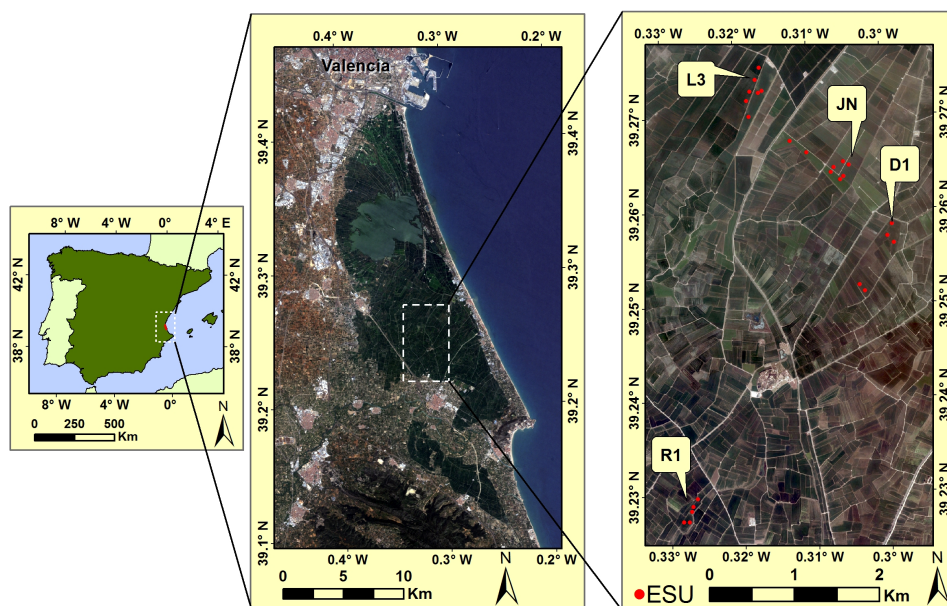


Figure 1. Study area: Location of the study area in Spain (**left**), a Landsat 8 Operational Land Imager (OLI) RGB composite image showing the extent of the rice fields (**middle**), and specific locations of the elementary sampling units (ESUs) within the rice fields (**right**). Four representative ESUs (R1, L3, JN, and D1) were selected for showing phenology stages.

The site is a homogeneous rice planting area of approximately 10 km × 20 km extension. Most of the paddy rice fields are rectangular and flat, approximately 100 m × 200 m. The rice cropping practices are common in the entire rice district. The sowing dates are around early May. The maturity stage is reached in early September, and the rice harvest begins in mid-September. Fields are flooded most of the time during the season. The two main rice varieties are Senia and Bomba, and are under the guarantee granted by the Regulatory Board of Designation of Origin Arròs de València <http://www.arrozdevalencia.org>. These rice varieties have different morphological structures. The Senia variety has more stacked up stems and leaves, while the Bomba variety has a considerably greater height.

ERMES (an Earth observation Model based rice information Service) <http://www.ermes-fp7space.eu/> is an FP7 Project funded by the European Commission with the objective of developing a prototype of downstream service dedicated to the rice sector to support authorities and farmers. In the framework of ERMES, ground measurements of phenology, plant height, chlorophyll content, FAPAR (Fraction of Absorbed Photosynthetically Active Radiation), and PAI were acquired on 26 ESUs from June the 17th to September the 8th in 2014. Measurement dates were selected to cover the entire rice season.

2.2. Spatial Sampling Strategy

The sampling strategy is a critical issue and should include considerations such as the number, dimensions, and spatial distributions of ESUs, driven by the heterogeneity of the study area, and the sampling scheme within each individual ESU. The 26 ESUs were selected within homogeneous rice fields. The fields were selected in order to take into account the main cultivated rice varieties. A reliable sampling covering the maximum bio-physical parameter variability was done. Over each ESU, the same sampling scheme was used as recommended by the VALERI (Validation of Land European Remote sensing Instruments) protocol in the case of row crops. This protocol suggests to

take measurements along small transects between rows and incorporate some random acquisitions to prevent possible biases in the characterization of the row effect. ESUs were located at least 30 m away from the field borders and were approximately 20 m × 20 m in size. In order to characterize the variability within each ESU, a range of 18 to 24 measurements of the bio-physical variables were taken. This number of replicates allows to obtain a statistically significant mean estimate of each bio-physical variable per ESU. The center of the ESU was geo-referenced using GPS for later matching and association of the mean of each bio-physical variable with the reflectance derived from satellite data.

2.3. Effective Plant Area Index (PAI_{eff})

Strictly speaking, the “true LAI” can be measured only using a planimeter [26]. Indirect methods compute an effective leaf area index through the following relation:

$$LAI_{eff} = \Omega \cdot LAI \quad (1)$$

where Ω is the aggregation or dispersion parameter [47,48], also known as the clumping index [11]. It accounts for canopy structure, geometry and foliage clumping. PocketLAI, LAI-2000, and DHP actually provide an estimate of the plant area index, since all parts of the plant contribute to the canopy transmittance. Indirect methods based on gap fraction measurements do not have the ability to determine if some leaves are present behind the stems, branches, flowers, or trunk [29]. In this study, we are using PAI_{eff} , since we are considering the rice fields as a turbid medium (minimum clumping).

2.4. PocketLAI

In this study, we proposed the use of PocketLAI for *in situ* non-destructive rice PAI_{eff} monitoring. PocketLAI is a smartphone application based on the segmentation of images acquired at 57.5° below the canopy to estimate the gap fraction [24]. It is based on a simplified model of light transmittance under the assumption of a random spatial distribution of infinitely small leaves. In this case, the gap fraction $P_0(\theta_v, \phi_v)$ in the zenith angle θ_v direction and azimuth angle ϕ_v is given by:

$$P_0(\theta_v, \phi_v) = \exp \left(-G(\theta_v, \phi_v) \frac{PAI}{\cos \theta_v} \right) \quad (2)$$

where $G(\theta_v, \phi_v)$ is the *projection function*, which can be considered as almost independent of *leaf inclination* ($G \approx 0.5$) for a viewing angle of 57.5° (≈ 1 radian) [17]. Inverting Equation (2), PAI can be estimated from the gap fraction at this particular direction as follows:

$$PAI(57.5^\circ) = \frac{-\ln P_0(57.5^\circ)}{0.93}. \quad (3)$$

At this directional configuration, the information acquired is independent from leaf angle distribution and minimizes leaf clumping effects [30]. This property comes from the *projection function* used to compute gap fraction corresponding to the projection of a PAI unit into a given direction. This mobile application has been shown to perform well in canopies with different structures [25] and has proven its reliability in terms of both trueness and precision [24]. PocketLAI computes the gap fraction using the smartphone’s accelerometer and camera. The operator holds the device vertically below the canopy and rotates the device along its main axis. When the angle between the vertical and the normal to the screen reaches 57.5°, a camera frame is captured and processed using an algorithm based on a segmentation strategy to detect sky pixels. In this study, the smartphone was placed 5 cm above the shallow water. PocketLAI allows the averaging of various numbers of PAI estimates in order to get a representative measure of the ESU based on the visual jackknife method [18]. In this work, the mobile application was installed on a Samsung Galaxy S4 GT-I9505, with a Quad-Core

1.9 GHz processor and 2 GB RAM. The smartphone's camera allowed PocketLAI to take images with a resolution of 4128×3096 pixels. With the aim of obtaining a representative measurement of each ESU, eighteen measurements acquired under similar illumination conditions were computed for each ESU following the strategy described in Section 2.2.

2.5. Digital Hemispherical Photography (DHP)

Digital hemispherical photography allows the computation of PAI measuring the gap fraction. The digital photographs were taken downward-looking and the distance between the camera and the top of the rice canopy was set to about 1 m to avoid the case that leaves were too close to the lens. Although upward photography is easy for gap identification, it may overestimate the gap fractions and consequently underestimate the PAI_{eff} in the case of rice [29]. Thus, in this study, given this underestimation and the characteristics of the rice plant (low height) and soil conditions (flooded), the downward-looking method was selected. During the ERMES field campaign, 16 photographs per ESU were acquired with the DHP and were subsequently processed using the Can-Eye software developed at INRA-CSE Avignon in order to meet the requirements of VALERI. Can-Eye computes effective PAI from gap fractions after an interactive thresholding classification process separating rice foliage from the background (downward view). The images were masked limiting to 60° zenith angle the valid range of fisheye lens to avoid edge distortions and ensuring that the area captured was within the ESU. Hemispherical photographs were taken maintaining the camera in an approximately horizontal position not-exceeding 10° , which is considered the threshold for minimizing errors due horizontal camera levelling in estimating PAI [28]. Photographs with suboptimal horizontal acquisition were excluded. A NIKON Coolpix5000 camera was used at the finest image resolution available (JPEG format at 2560×1920 pixels resolution) and a FC-E8 Nikon fisheye lens with a field of view of 183° . The camera with the fisheye lens was calibrated before the field according to the CAN EYE manual [49]. The PAI_{eff} was computed using the Miller's formula [50] as follows:

$$PAI_{eff} = 2 \int_0^{\pi/2} -\ln P_0(\theta) \cos(\theta) \sin(\theta) d\theta \quad (4)$$

In addition, the Can-Eye software proposes an estimate of the PAI_{eff} derived from the gap fraction measured for a view angle of 57.5° in the same way that is computed in Equation (3). Can-Eye software simultaneously processes up to 16 images acquired over the same ESU. All images belonging to the same ESU were acquired with similar illumination conditions to limit the variation of color dynamics between images.

2.6. LAI-2000 Plant Canopy Analyzer

Li-Cor LAI-2000 was used to estimate the rice PAI using a gap fraction method that determines the PAI from measurements made above and below the canopy, which are used to determine canopy light interception at five zenith angles (7° , 23° , 38° , 53° , and 68°). Canopy transmission is measured between 320 and 490 nm. LAI-2000 computes the PAI_{eff} using the Miller's formula as:

$$PAI_{eff} = 2 \int_0^{\pi/2} -\ln \overline{P(\theta)} \cos(\theta) \sin(\theta) d\theta = 2 \sum_{i=1}^5 \overline{K_i} W_i \quad (5)$$

where K_i and W_i are the contact number and the weighting factor, and the subscript i refers to the number of the ring. $\overline{P(\theta)}$ is the average probability of light penetration into the canopy, and the gap fraction $G_i(\theta)$ is computed as:

$$G_i(\theta) = \exp(\overline{\ln P(\theta)}) = \exp\left(\left(\frac{1}{N}\right) \sum_{j=1}^N \ln\left(\frac{B_j}{A_j}\right)\right) \quad (6)$$

The subscript j is number of readings ($j = 1 \dots N$), and B_j and A_j are the corresponding readings to below and above the canopy, respectively.

In this field campaign, a 270° view cap was used to limit the azimuthal field of view, facing away from the operator. Three replications of one measurement above and eight below the canopy were made for each measurement. Regarding below canopy measurements, the Li-Cor LAI-2000 instrument was placed about 5 cm above the shallow water. Measurements were made under diffuse light conditions in order to avoid incoming radiation from sunlit foliage.

2.7. Complementary Field Data

2.7.1. Phenology

The information regarding the phenology was obtained for all ESUs with *in situ* observations according to the BBCH scale for rice. Representative ESUs of the study area (see Figure 1) were selected in order to show phenology stages of the fields. The BBCH scale provides a continuous numerical range for identifying and describing the plant phenology. Table 1 describes the main rice stages as follows: stages 0 to 49 correspond to the vegetative phase of the rice cycle, stages 50 to 69 correspond to the reproductive phase, and stages 70 to 99 correspond to the maturation phase. The BBCH scale accounts for a single plant. If an operator aims to define the phenology stage of an ESU, at least half of the plants should present the same phenological state.

Table 1. Description of the rice phenological cycle according to the BBCH scale stages.

	Description	Principal Stage	BBCH
Vegetative	Germination	0	0–9
	Leaf development	1	10–19
	Tillering	2	20–29
	Stem elongation	3	30–39
	Booting	4	40–49
Reproductive	Emergence, heading	5	50–59
	Flowering, anthesis	6	60–69
Maturation	Fruit development	7	70–79
	Ripening	8	80–89
	Senescence	9	90–99

2.7.2. Chlorophyll Content

The leaf chlorophyll content has been obtained by means of a SPAD-502, which gives a leaf chlorophyll content estimate taking into account the radiation absorbed by leaves at specific wavelengths. SPAD-502 provides digital counts (DN) which are dimensionless and require an empirical calibration between SPAD-502 DN and extracted chlorophyll values in laboratory. This problem can be addressed using specific calibration curves, which includes linear, exponential, or polynomial calibration functions. In this work, a calibrated SPAD-502 using a power relationship to obtain leaf Chl content physical values in $\mu\text{g} \cdot \text{cm}^{-2}$ was used [36]. Hence, in order to characterize the chlorophyll content of the ESUs, 10–15 readings were made covering each ESU. SPAD-502 readings were made on the last completely unfolded leaf.

2.8. Transfer Function for High-Resolution PAI_{eff} Mapping

In this study, the field measurements were also used for the derivation of a reference high-resolution map within the study area which can be used for remote sensing validation products. The derivation of high-resolution PAI_{eff} maps is a procedure based on an empirical transfer function (TF) that establishes a relationship between the average PAI_{eff} values from each ESU and the multispectral values from sensors onboard either satellite or airborne platforms. In this work, a relationship between the average PAI_{eff} values from each ESU and radiometric values

over concomitant Landsat-8 imagery in four spectral bands, namely green, red, near infrared, and shortwave infrared 1 bands (G, R, NIR, and SWIR1) was used. Following previous works [44] and recommendations [41,43,45], the up-scaling algorithm relies on a robust linear regression that evaluates the band combination that exhibits the lowest error. A multivariate ordinary least squares (OLS) regression is used, which assumes that the prediction Y (*in situ* PAI_{eff} measurement) is related to the independent variable $X_i, i = 1, 2, \dots, q$ (Landsat-8 radiometric values in the four selected bands) through the the following functional relationship:

$$Y_j = \beta_0 + \sum_{j=1}^q \beta_j \cdot X_{ij} + \varepsilon \quad (i = 1, \dots, n) \quad (7)$$

where n is the number of observations, and β_j are the parameters of the multiple linear regression. With the goal of minimizing the influence of outliers, an iteratively re-weighted least squares (IRLS) method is applied. This approach includes a weight factor to adjust the amount of each response value on the estimates provided by the model. ESUs with weights lower than 0.7 are usually linked either to samples located near the field borders or to experimental errors [44]. In order to evaluate the optimal Landsat-8 OLI band combination, we considered the weighted root mean square error (RW) and the cross-validation root mean square error (RC). The RW gives the mean prediction error assumed by the model for all the observations while the RC provides information about the model's performance.

In this study, the response variable represents the three data sets of *in situ* PAI_{eff} measurements (i.e., PAI_{eff}^{APP} , PAI_{eff}^{DHP} , and $PAI_{eff}^{LAI2000}$), whereas the predictor variable is the radiometric information on the four Landsat-8 OLI spectral bands. Landsat 8 OLI images provide valuable information for crop monitoring at the local scale [51] due to the spatial (30 m) and temporal (16 day) resolutions. Images were downloaded as a Level 1T product and atmospherically corrected using the L8SR code, which corrects to surface reflectance from top of atmosphere (TOA) reflectance using ancillary NCEP (US National Centers for Environmental Prediction) water vapor data and TOMS (Total Ozone Mapping Spectrometer) ozone data sets. Landsat 8 OLI images were clipped to 1500×800 pixel size covering the entire rice area.

The standard method of generating a reference map relies on the information provided by spectral bands and vegetation indices from a single date of imagery [44,46]. This general method implies that each map is derived with a different TF made with the corresponding measurements and associated reflectances. One main feature of this work is the combination of field and satellite data from different acquisition dates in order to create a multitemporal data set which was used for building a unique global TF able to derive mutitemporal maps (see Figure 2).

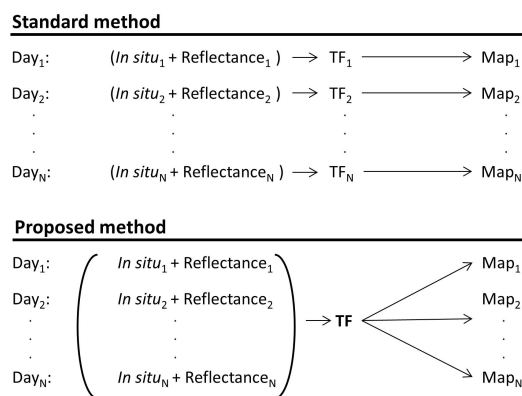


Figure 2. Standard method for derivation of high-resolution reference maps building a transfer function per available measuring date (**up**), and the proposed method made with a multitemporal data set (**bottom**).

3. Results

3.1. On the Temporal Evolution of the PAI_{eff} Field Measurements

A general overview of the PAI_{eff} measurements obtained during the field campaign shows that the range of PAI_{eff} values obtained using all three instruments is according to the values reported in the literature for rice [29,52]. Seasonal variation of rice PAI_{eff} (see Figure 3b) shows the typical behavior throughout the entire rice season. Maximum values of PAI_{eff} were observed on day of year (DoY) 220. Significant differences in PAI_{eff} time profiles were observed in the representative ESUs. Specifically, R1 clearly presented the highest values, about $1 \text{ m}^2/\text{m}^2$ at the beginning of the season and about $5.5 \text{ m}^2/\text{m}^2$ at almost the end of the growing season when the PAI_{eff} values are maximum. R1 and L3 ESUs belong to same rice variety; nevertheless, a different PAI_{eff} evolution was observed (see Figure 3b,c): plant density on R1 (number of rice plants per unit area) was significantly higher than on L3, leading to systematically higher PAI_{eff} values.

3.2. On the Ancillary Data: Phenology and Leaf Chlorophyll Content

As part of the ERMES field campaign, phenology was monitored during 10 days from DoY = 168 to DoY = 251. In general, all representative ESUs have a similar behavior (see Figure 3a), since sowing dates were very similar in the study area. Comparison of rice PAI_{eff} and BBCH seasonal variation (Figure 3c) highlights three most noticeable features: (1) the fast increase in PAI_{eff} during the tillering and stem elongation vegetative stages (BBCH from 20 to 40); (2) the saturation of PAI_{eff} during the stages of flowering, fruit development, and ripening (BBCH from 50 to 90), in which these vegetative and maturation phases the rice plant loses some leaves and becomes drier, PAI_{eff} being practically constant; and (3) D1, JN, and L3 ESUs have a close evolution, while R1 shows higher PAI_{eff} values due to higher plant density.

Rice leaf chlorophyll content measured with SPAD-502 showed a constant behavior during vegetative and reproductive stages, which means that no anomalies due to significant changes in Chl were found. A slight decrease in the rice Chl content was observed in the maturation stage due to the beginning of rice senescence (see Figure 3d). Since SPAD-502 readings were made in the last unfolded leaf, in most cases the estimated Chl content corresponded to leaves that were not senescent; *i.e.*, on leaves where the hydrolytic processes that recycle nitrogen-rich compounds (including chlorophylls and rubisco) were not started yet. Consequently, although the total plant nitrogen content at the end of August were low (because of older leaves), measurements in the last-emitted leaf could still have high nitrogen content. This could explain why leaf chlorophyll contents were still high while the rice plants were approximately in the last part of the crop cycle. On the other hand, differences in leaf Chl values are related to the two main rice varieties of the study area. The Senia variety (R1 and L3 ESUs) presents leaves with a high chlorophyll content of about $60 \mu\text{g} \cdot \text{cm}^{-2}$. By contrast, the Bomba variety (D1 and JN ESUs) presented lower chlorophyll content values, about $35\text{--}40 \mu\text{g} \cdot \text{cm}^{-2}$. These Chl values explain the yellow greenish leaf color of the Bomba variety.

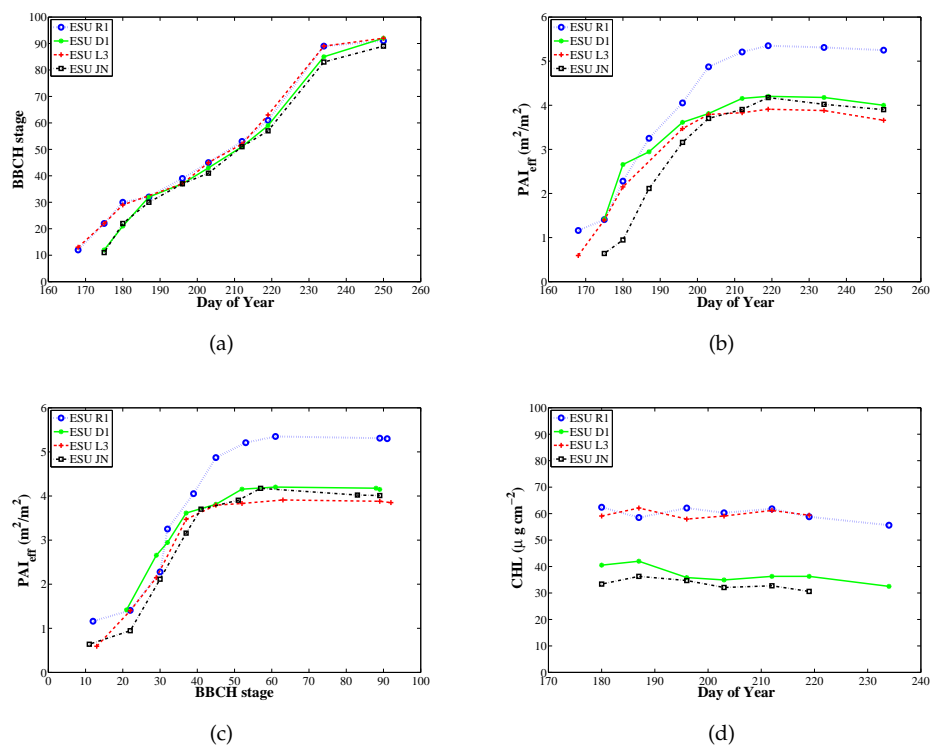


Figure 3. Bio-physical indicators measured in four representative ESUs during the field campaign: (a) Phenology according to the (BBCH) scale; (b) effective plant area index acquired with PocketLAI; (c) effective plant area index related with the BBCH phenological stages; and (d) leaf chlorophyll content measured during the rice season.

3.3. On the PAI_{eff} Measuring Instruments and Maps Comparison

PocketLAI computes the PAI_{eff} of an ESU, averaging each PAI_{eff} calculated from each gap fraction reading on the ESU, while DHP and LAI-2000 first calculate the average gap fraction within an ESU and retrieve PAI_{eff} from it. Since the gap fraction-PAI_{eff} relationship is not linear, it is not equivalent to first average the gap fraction and then estimate the PAI_{eff} than the contrary [17]. This fact may be one of the reasons why PocketLAI generally underestimates PAI_{eff} values (see Figure 4). The PAI_{eff} and PAI_{eff}(57.5°) values obtained with DHP were used to compare estimates from LAI-2000 and PocketLAI with PAI_{eff}, respectively. Different statistics were computed to assess the consistency and performance of the PocketLAI with LAI-2000 and DHP: root-mean-squared error (RMSE) and mean absolute error (MAE) were used to assess the accuracy, Mean error (ME) to evaluate the bias, and coefficient of determination (R^2) to account for the goodness-of-fit and variability between instruments.

Effective PAI values computed with all three instruments are well correlated. The coefficient of determination computed between PAI_{eff} estimates acquired with PocketLAI and classical instruments was $R^2 = 0.95$ and $R^2 = 0.94$ for LAI-2000 and DHP, respectively. Comparisons also reveal high accuracy and small bias between instruments (ME = -0.38 m²/m², MAE = 0.41 m²/m² for LAI-2000, and ME = 0.46 m²/m², MAE = 0.48 m²/m² for DHP). PocketLAI presents a very small negative bias regarding LAI-2000, although a slight positive bias is found at the beginning of the season (PAI_{eff} < 1 m²/m²) (see Figure 4a). PocketLAI is also highly consistent with DHP, although it tends to produce slightly lower values (ME = 0.47) (see Figure 4b). Comparison between LAI-2000 and DHP instruments also shows good results in terms of accuracy, bias, and variability (RMSE = 0.33 m²/m², ME = 0.11 m²/m², $R^2 = 0.94$ m²/m²) (see Figure 4c).

Indirect methods provide a PAI_{eff} associated with several sources of measurement error, including performance of instruments, illumination conditions, simplification of leaf optical properties, suboptimal spatial sampling within an ESU, and saturation of optical signal in dense canopies. Specifically, variability observed when PAI_{eff} values are greater than four typically correspond to rice plants in the reproductive phenology stage. At this point, there is a significant change in the rice morphological structure due to the panicle emergence, leading to an increasing variability of the estimates. Error, Bias, and correlation between instruments are small and do agree with previous studies in different crops [23,53] in which strong correlations ($R^2 = 0.96$ and $R^2 = 0.94$), small bias ($ME \approx 0.2$) and accuracy ($RMSE \approx 0.5$) were found.

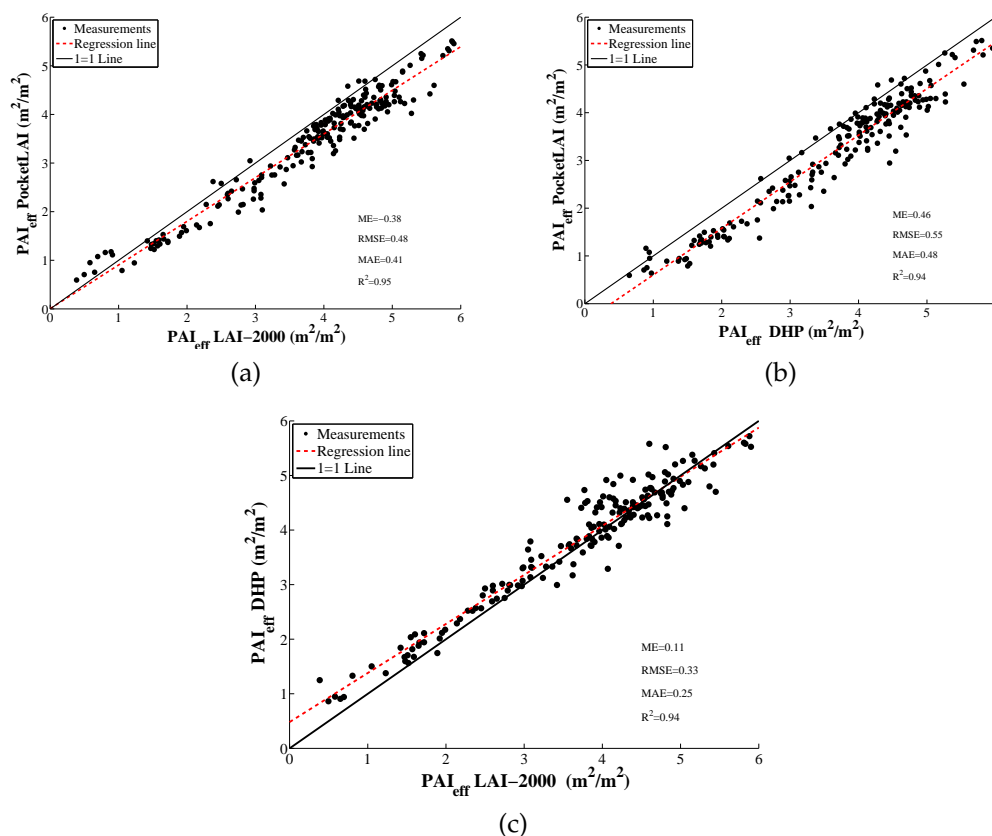


Figure 4. Comparison of effective plant area index measurements collected in the 2014 Spanish ERMES field campaign between: (a) PocketLAI and LAI-2000; (b) PocketLAI and Digital hemispherical photography (DHP) and (c) DHP and LAI-2000.

These *in situ* PAI_{eff} measurements allowed the creation of a transfer function, which was used to derive PAI_{eff} maps. RW and RC errors in function of the selected Landsat-8 combination bands were computed for the TF. The best band combination was (SWIR1,NIR,R,G) in all three cases. This specific band combination reveals $RW = 0.46$ and $RC = 0.50$ in the case of PocketLAI, $RW = 0.51$ and $RC = 0.52$ in the case of LAI-2000, and $RW = 0.50$ and 0.51 in the case of DHP. Statistical indicators of the selected transfer functions showed good correlations and biases in all three cases ($R^2 > 0.93$ and $B < 0.02$). For the sake of brevity we only show one derived PAI_{eff} map per instrument, which corresponds to DoY = 196 and BCCH ≈ 35 (see Figure 5). The three derived maps show similar estimated PAI_{eff} values within the study area. Nevertheless, the estimated PAI_{eff}^{APP} map shows a slight underestimation if it is compared with PAI_{eff}^{DHP} - and $PAI_{eff}^{LAI2000}$ -derived maps. Pixels covering non-interest areas were masked out in blue color. Up-scaled maps derived from this ground dataset are made available for the validation of remote sensing products through

the ImagineS www.fp7-imagines.eu (Implementing Multi-scale aGricultural Indicators Exploiting Sentinels) ground database.

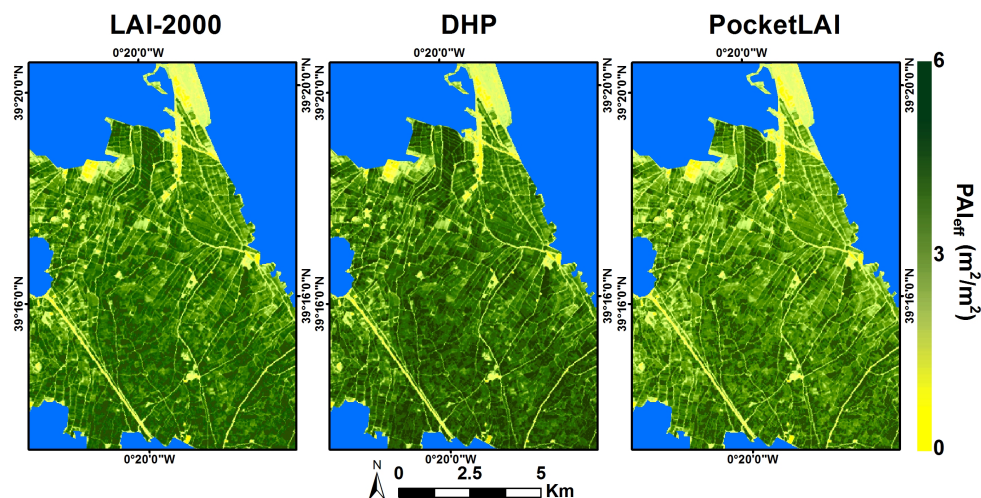


Figure 5. PAI_{eff} -derived maps using a transfer function and Landsat-8 data (DoY = 196, year 2014) using $PAI_{eff}^{LAI2000}$ (left); PAI_{eff}^{DHP} (middle); and PAI_{eff}^{APP} (right) data sets.

Derived maps reveal PAI_{eff} estimates fell within the expected range at that phenological rice state and show high consistency between classical instruments estimates. On the other hand, the map derived with PocketLAI measurements shows less intense greens, corresponding to values slightly lower than the ones retrieved either with DHP or LAI-2000. This fact is in agreement with the low PAI_{eff} underestimation observed in the comparison of the *in situ* PAI_{eff} measurements between the app and the classical instruments. All three estimated maps show expected within-field variations due to the spatial PAI_{eff} variability of rice fields corresponding to different varieties, phenological stages, and low values corresponding to field boundaries and non-vegetated areas such as roads, rice dryers, and agricultural warehouses. Statistical indicators between derived map values showed very high correlations and consistency between the PocketLAI and the classical instruments (see Table 2). Difference between PAI_{eff} maps retrieved with different methods was also computed to explore spatial patterns (see Figure 6). The density scatterplots between derived map values are shown in Figure 7. Difference maps showed no spatial patterns in disagreement between classical methods and the PocketLAI. Nevertheless, the PocketLAI–DHP difference map (Figure 6 (right panel)) revealed higher differences in estimated LAI values, mainly in the condition of dense biomass (high-range values), which suggests an underestimation of the PocketLAI field measurements with respect to the other indirect methods. In addition, the scatterplots between map values showed a proportional underestimation in maps retrieved using PocketLAI data over the rice fields for PAI_{eff} high-range values (*i.e.*, $PAI_{eff} > 4$). This underestimation may reach a maximum of 1 (in PAI_{eff} units) and 0.6 when compared with DHP and LAI-2000, respectively (see Figure 7).

Table 2. Statistical indicators (Root-mean-squared error, RMSE; mean absolute error, MAE; absolute value of the mean error, |ME|; and the coefficient of determination, R^2) between PocketLAI estimated values and classical instruments (DHP and LAI-2000).

Instrument	RMSE	MAE	ME	R^2
DHP	0.67	0.64	0.61	0.94
LAI-2000	0.35	0.33	0.29	0.98

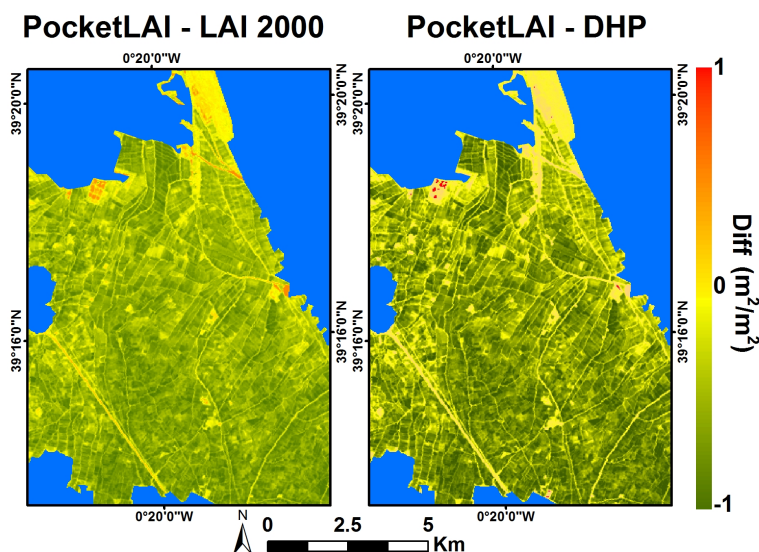


Figure 6. Difference maps between PAI_{eff} derived from PocketLAI and LAI-2000 (left); and DHP (right).

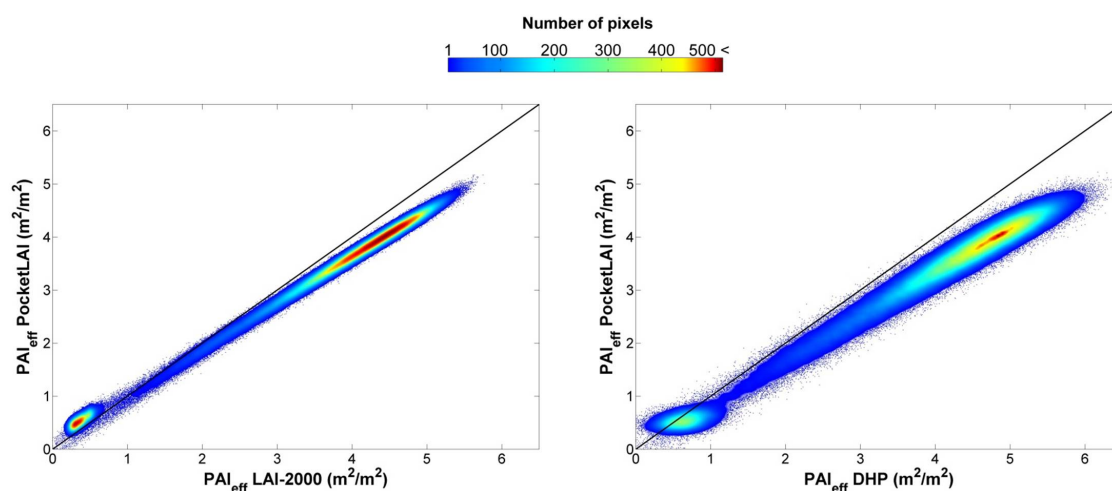


Figure 7. Scatterplots of PAI_{eff} estimates between PocketLAI and LAI-2000 (left) and DHP (right).

4. Conclusions

The results presented in this work bring to light the good performance of a brand new smartphone mobile app called PocketLAI for effective plant area index acquisitions over rice. Ground PocketLAI measurements were compared with those acquired with classical instruments (LAI-2000 and DHP). In this study, the assessment was carried out over paddy fields in Spain during the 2014 ERMES field activities. PocketLAI usually underestimates PAI_{eff} values from LAI-2000 and DHP. Despite that low underestimation, it is found that PAI_{eff} is very well correlated between the app and the classical instruments ($R^2 = 0.94$ for DHP and $R^2 = 0.95$ in the case of LAI-2000). The averaging methods within an ESU are different for PocketLAI, LAI-2000, and DHP. PocketLAI allows farmers to easily monitor crop status during the rice season and to capture within-field spatial variability of its state. The use of the app by smartphones is a very good alternative to classical instruments due to its portability and low-cost. These results suggest that PocketLAI can be used as a plant area index measuring instrument, specially for near-real-time applications. Even so, further studies will include intercomparison with different mobile devices over different crops.

The seasonal PAI_{eff} measurements obtained from this study are in accordance with Committee on Earth Observation Satellites (CEOS) good practice protocols, making it suitable for bio-physical land product validation and up-scaling purposes. As a matter of fact, this data set is being used for validation and can be found on the ImagineS website. This work shows how measures from a smartphone can be used for up-scaling and deriving high-resolution PAI_{eff} maps through a transfer function.

Although this strategy is usually made to build *ad hoc* TF per available measurement-imagery day, this paper proposes the use of a unique global TF made of several measuring dates and their associated reflectances. This approach is more robust to estimate PAI during all stages of the plant season while avoiding overfitting to individual dates.

The map derived from *in situ* PocketLAI measurements was compared with those obtained either from LAI-2000 and DHP. Statistical indicators showed high correlations and consistency when derived map values using PocketLAI acquisitions were compared with DHP and LAI-2000 derived maps.

In this study, the rice monitoring was completed with concomitant *in situ* leaf chlorophyll content and phenology measurements. Leaf chlorophyll content measurements showed no stress situations during the growing season. The rice phenology acquired during this study was used for monitoring of the current growing season and provided useful information to be used in crop models. Specifically, when phenology is $BBCH \geq 50$, which implies that the rice plant is in the reproductive and maturation phases, the PAI_{eff} acquired do not match with the green area index, which is provided by operational products.

This work showed an example of maps derived using one date Landsat data. Multispectral images periodically-recorded from sensors such as Landsat or SPOT5 are commonly used to monitor vegetation status. In this context, and taking the advantage that the study area of this work was recently selected and added for the SPOT5 take5 acquisition plan, future work will consider these free available data for deriving bio-physical parameter maps through the same up-scaling approach used in this study. In the same way, the upcoming dissemination of free Sentinel 2A data will be a good source of data for these purposes due to its similar temporal, spectral, and spatial characteristics.

Acknowledgments: The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 606983. ImagineS and LSA SAF projects are also acknowledged. The authors would like to thank the Spanish rice farmers for allowing collecting *in situ* data within their fields.

Author Contributions: All co-authors of this manuscript significantly contributed to all phases of the investigation. They contributed equally to the preparation, analysis, review and editing of this manuscript.

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

References

1. Whelan, B.; McBratney, A. The “Null Hypothesis” of precision agriculture management. *Precis. Agric.* **2000**, *2*, 265–279.
2. Stafford, J.V. Implementing precision agriculture in the 21st Century. *J. Agr. Eng. Res.* **2000**, *76*, 267–275.
3. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981.
4. Colgan, M.S.; Baldeck, C.A.; Féret, J.B.; Asner, G.P. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data. *Remote Sens.* **2012**, *4*, 3462–3480.
5. D’Oleire-Oltmanns, S.; Marzolf, I.; Peter, K.D.; Ries, J.B. Unmanned Aerial Vehicle (UAV) for monitoring soil erosion in Morocco. *Remote Sens.* **2012**, *4*, 3390–3416.
6. Marshall, M.; Thenkabail, P. Developing *in situ* non-destructive estimates of crop biomass to address issues of scale in remote sensing. *Remote Sens.* **2015**, *7*, 808–835.
7. Son, N.T.; Chen, C.F.; Chen, C.R.; Duc, H.N.; Chang, L.Y. A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam. *Remote Sens.* **2013**, *6*, 135–156.

8. López-Sánchez, J.; Vicente-Guijalba, F.; Ballester-Berman, J.; Cloude, S. Polarimetric response of rice fields at C-Band: Analysis and phenology retrieval. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2977–2993.
9. Nelson, A.; Setiyono, T.; Rala, A.; Quicho, E.; Raviz, J.; Abonete, P.; Maunahan, A.; Garcia, C.; Bhatti, H.; Villano, L.; *et al.* Towards an operational SAR-Based rice monitoring system in ASIA: Examples from 13 demonstration sites across asia in the RIICE project. *Remote Sens.* **2014**, *6*, 10773–10812.
10. Karila, K.; Nevalainen, O.; Krooks, A.; Karjalainen, M.; Kaasalainen, S. Monitoring changes in rice cultivated area from SAR and optical satellite images in Ben Tre and Tra Vinh Provinces in Mekong Delta, Vietnam. *Remote Sens.* **2014**, *6*, 4090–4108.
11. Chen, J.; Black, T. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* **1992**, *15*, 421–429.
12. Vuolo, F.; Neugebauer, N.; Bolognesi, S.F.; Atzberger, C.; D’Urso, G. Estimation of Leaf Area Index using DEIMOS-1 data: Application and transferability of a semi-empirical relationship between two agricultural areas. *Remote Sens.* **2013**, *5*, 1274–1291.
13. Richter, K.; Hank, T.B.; Vuolo, F.; Mauser, W.; D’Urso, G. Optimal exploitation of the Sentinel-2 spectral capabilities for crop leaf area index mapping. *Remote Sens.* **2012**, *4*, 561–582.
14. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352.
15. GCOS. Systematic Observation Requirements for Satellite-Based Products for Climate, 2011 Update, Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update); World Meteorological Organization (WMO): Geneva, Switzerland, 2011.
16. Breda, N.J.J. Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. *J. Exp. Bot.* **2003**, *54*, 2403–2417.
17. Weiss, M.; Baret, F.; Smith, G.; Jonckheere, I.; Coppin, P. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agric. For. Meteorol.* **2004**, *121*, 37–53.
18. Confalonieri, R.; Stroppiana, D.; Boschetti, M.; Gusberti, D.; Bocchi, S.; Acutis, M. Analysis of rice sample size variability due to development stage, nitrogen fertilization, sowing technique and variety using the visual jackknife. *F. Crop. Res.* **2006**, *97*, 135–141.
19. Jonckheere, I.; Fleck, S.; Nackaerts, K.; Muys, B.; Coppin, P.; Weiss, M.; Baret, F. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* **2004**, *121*, 19–35.
20. Warren-Wilson, J. Estimation of foliage denseness and foliage angle by inclined point quadrats. *Aust. J. Bot.* **1963**, *11*, 95–105.
21. Ross, J. *The Radiation Regime and Architecture of Plant Stands*; Springer Science & Business Media: Hague, The Netherlands, 1981.
22. Garrigues, S.; Shabanov, N.; Swanson, K.; Morisette, J.; Baret, F.; Myneni, R. Intercomparison and sensitivity analysis of Leaf Area Index retrievals from LAI-2000, AccuPAR, and digital hemispherical photography over croplands. *Agric. For. Meteorol.* **2008**, *148*, 1193–1209.
23. Verger, A.; Martínez, B.; Camacho, F.; García-Haro, F.J. Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. *Int. J. Remote Sens.* **2009**, *30*, 3849–3849.
24. Confalonieri, R.; Foi, M.; Casa, R.; Aquaro, S.; Tona, E.; Peterle, M.; Boldini, A.; Carli, G.D.; Ferrari, A.; Finotto, G.; *et al.* Development of an app for estimating leaf area index using a smartphone. Trueness and precision determination and comparison with other indirect methods. *Comput. Electron. Agric.* **2013**, *96*, 67–74.
25. Francone, C.; Pagani, V.; Foi, M.; Cappelli, G.; Confalonieri, R. Comparison of leaf area index estimates by ceptometer and PocketLAI smart app in canopies with different structures. *F. Crop. Res.* **2014**, *155*, 38–41.
26. Frazer, G.W.; Trofymow, J.A.; Lertzman, K.P. *A Method for Estimating Canopy Openness, Effective Leaf Area Index, and Photosynthetically Active Photon Flux Density using Hemispherical Photography and Computerized Image Analysis Techniques*; Technical report BC-X-373; Pacific Forestry Centre: Victoria, BC, Canada 1997.
27. Chen, J.; Black, T.; Adams, R. Evaluation of hemispherical photography for determining plant area index and geometry of a forest stand. *Agric. For. Meteorol.* **1991**, *56*, 129–143.

28. Demarez, V.; Duthoit, S.; Baret, F.; Weiss, M.; Dedieu, G. Estimation of leaf area and clumping indexes of crops with hemispherical photographs. *Agric. For. Meteorol.* **2008**, *148*, 644–655.
29. Fang, H.; Li, W.; Wei, S.; Jiang, C. Seasonal variation of leaf area index (LAI) over paddy rice fields in NE China: Intercomparison of destructive sampling, LAI-2200, digital hemispherical photography (DHP), and AccuPAR methods. *Agric. For. Meteorol.* **2014**, *198–199*, 126–141.
30. Baret, F.; de Solan, B.; López-Lozano, R.; Ma, K.; Weiss, M. GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agric. For. Meteorol.* **2010**, *150*, 1393–1401.
31. Lancashire, P.D.; Bleiholder, H.; Boom, T.V.D.; Langelüddeke, P.; Strauss, R.; Weber, E.; Witzengerger, A. A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* **1991**, *119*, 561–601.
32. López-Sánchez, J.; Vicente-Guijalba, F.; Ballester-Berman, J.; Cloude, S. Influence of incidence angle on the coherent copolar polarimetric response of rice at X-Band. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 249–253.
33. Evans, J. Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecolog.* **1989**, *78*, 9–19.
34. Peng, Y.; Gitelson, A.A. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. *Remote Sens. Environ.* **2012**, *117*, 440–448.
35. Clevers, J.; Kooistra, L. Using hyperspectral remote sensing data for retrieving canopy Chlorophyll and nitrogen content. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 574–583.
36. Delegido, J.; Wittenberghe, S.V.; Verrelst, J.; Ortiz, V.; Veroustraete, F.; Valcke, R.; Samson, R.; Rivera, J.P.; Tenjo, C.; Moreno, J. Chlorophyll content mapping of urban vegetation in the city of Valencia based on the hyperspectral NAOC index. *Ecol. Indic.* **2014**, *40*, 34–42.
37. Vincini, M.; Amaducci, S.; Frazzi, E. Empirical estimation of leaf Chlorophyll density in winter wheat canopies using Sentinel-2 spectral resolution. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3220–3235.
38. Jin, X.; Li, Z.; Feng, H.; Xu, X.; Yang, G. Newly combined spectral indices to improve estimation of total leaf Chlorophyll content in cotton. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4589–4600.
39. Hu, H.; Zhang, G.; Zheng, K. Modeling leaf image, Chlorophyll fluorescence, reflectance from SPAD readings. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4368–4373.
40. Mu, X.; Huang, S.; Ren, H.; Yan, G.; Song, W.; Ruan, G. Validating GEOV1 fractional vegetation cover derived from coarse-resolution remote sensing images over croplands. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 439–446.
41. Camacho, F.; Cernicharo, J.; Lacaze, R.; Baret, F.; Weiss, M. GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products. *Remote Sens. Environ.* **2013**, *137*, 310–329.
42. Zhu, L.; Chen, J.; Tang, S.; Li, G.; Guo, Z. Inter-Comparison and validation of the FY-3A/MERSI LAI product over mainland China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 458–468.
43. Fernandes, R.; Plummer, S.; Nightingale, J.; Baret, F.; Camacho, F.; Fang, H.; Garrigues, S.; Gobron, N.; Lang, M.; Lacaze, R.; et al. Global leaf area index product validation good practices. Version 2.0. In *Best Practice for Satellite-Derived Land Product Validation: Land Product Validation Subgroup (WGCV/CEOS)*; Schaepman-Strub, G.; Román, M.; Nickeson, J., Eds.; NASA: Greenbelt, MD, USA, 2014; p. 76.
44. Martínez, B.; García-Haro, F.; de Coca, F.C. Derivation of high-resolution leaf area index maps in support of validation activities: Application to the cropland Barrax site. *Agric. For. Meteorol.* **2009**, *149*, 130–145.
45. Weiss, M.; Baret, F.; Garrigues, S.; Lacaze, R. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: Validation and comparison with MODIS collection 4 products. *Remote Sens. Environ.* **2007**, *110*, 317–331.
46. Cohen, W.B.; Justice, C.O. Validating MODIS terrestrial ecology products: Linking *in situ* and satellite measurements. *Remote Sens. Environ.* **1999**, *70*, 1–3.
47. Nilson, T. A theoretical analysis of the frequency of gaps in plant stands. *Agric. Meteorol.* **1971**, *8*, 25–38.
48. Lemeur, R.; Blad, B.L. Plant modification for more efficient water use a critical review of light models for estimating the shortwave radiation regime of plant canopies. *Agric. Meteorol.* **1974**, *14*, 255–286.
49. Weiss, M.; Baret, F. *CAN-EYE V6.313 User Manual*; INRA: Avignon, France, 2010.
50. Miller, J. A formula for average foliage density. *Aust. J. Bot.* **1967**, *15*, 141–144.

51. Campos-Taberner, M.; García-Haro, F.; Moreno, A.; Gilabert, M.; Martínez, B.; Sánchez-Ruiz, S.; Camps-Valls, G. Development of an earth observation processing chain for crop bio-physical parameters at local scale. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 29–32.
52. Asner, G.P.; Scurlock, J.M.O.; A. Hicke, J. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies. *Global Ecol. Biogeogr.* **2003**, *12*, 191–205.
53. Yilmaz, M.T., Hunt, E.R., Jr.; Goins, L.D.; Ustin, S.L.; Vanderbilt, V.C.; Jackson, T.J. Vegetation water content during SMEX04 from ground data and Landsat 5 Thematic Mapper imagery. *Remote Sens. Environ.* **2008**, *112*, 350–362.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring



Manuel Campos-Taberner^{a,*}, Francisco Javier García-Haro^a, Gustau Camps-Valls^b, Gonçal Grau-Muedra^a, Francesco Nutini^c, Alberto Crema^c, Mirco Boschetti^c

^aDepartament de Física de la Terra i Termodinàmica, Facultat de Física, Universitat de València, Dr. Moliner, 50, Burjassot 46100, València, Spain

^bImage Processing Laboratory (IPL), Universitat de València, Catedrático A. Escardino, 9, Paterna 46980, València, Spain

^cInstitute for Electromagnetic Sensing of the Environment, Italian National Research Council, Via Bassini 15, Milan 20133, Italy

ARTICLE INFO

Article history:

Received 2 May 2016

Received in revised form 16 September 2016

Accepted 1 October 2016

Available online xxxx

Keywords:

Crop monitoring

Rice

Leaf area index (LAI) retrieval

PROSAIL

Smartphone

Gaussian process regression (GPR)

Landsat

SPOT5 Take5

ABSTRACT

This paper presents an operational chain for high-resolution leaf area index (LAI) retrieval from multiresolution satellite data specifically developed for Mediterranean rice areas. The proposed methodology is based on the inversion of the PROSAIL radiative transfer model through the state-of-the-art nonlinear Gaussian process regression (GPR) method. Landsat and SPOT5 data were used for multitemporal LAI retrievals at high-resolution. LAI estimates were validated using time series of *in situ* LAI measurements collected during the rice season in Spain and Italy. Ground LAI data were collected with smartphones using PocketLAI, a specific phone application for LAI estimation. Temporal evolution of the LAI estimates using Landsat and SPOT5 data followed consistently the temporal evolution of the *in situ* LAI measurements acquired on several Mediterranean rice varieties. The estimates had a root-mean-square-error (RMSE) of 0.39 and 0.51 m²/m² in Spain and 0.38 and 0.47 m²/m² in Italy for Landsat and SPOT5 respectively, with a strong correlation ($R^2 > 0.92$) for both cases. Spatial-temporal assessment of the estimated LAI from Landsat and SPOT5 data confirmed the robustness and consistency of the retrieval chain. This paper demonstrates the importance of an adequate characterization of the underlying rice background in order to address changes in background condition related to water management. Results highlight the potential of the proposed chain for deriving multitemporal near real-time decametric LAI maps fundamental for operational rice crop monitoring, and demonstrate the readiness of the proposed method for the processing of data such as the recently launched Sentinel-2.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Green leaf area index (LAI) is a key biophysical parameter which represents half the total intercepting leaf area per unit ground surface area (Chen and Black, 1992). LAI plays an important role in vegetation processes such as photosynthesis and transpiration, and is connected to meteorological/climate and ecological land processes. LAI has been widely used in many agricultural and remote sensing studies (Carlson and Ripley, 1997; Clevers, 1988; Gower et al., 1999; Myneni et al., 1997). Concerning biomass and crop yield estimation, LAI estimates can be assimilated in crop models (Confalonieri et al., 2009) by means of forcing and/or recalibration techniques (Dorigo et al., 2007; Quaife et al., 2008). LAI retrieval from satellite data is among the main goals of the remote sensing community

(Chen et al., 2002; Colombo et al., 2003; Fang and Liang, 2005) as evidenced by the variety and usefulness of operational medium resolution products for vegetation monitoring from satellite sensors, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Myneni et al., 2002) and the *Système Pour l'Observation de la Terre* (SPOT) VEGETATION (Baret et al., 2007, 2013). Nevertheless, higher spatial resolutions (10–30 m) are needed to support crop management activities at a parcel level. In this context, the Landsat Data Continuity Mission (LDCM) (Roy et al., 2014) and the recently European Sentinel-2 Mission (Drusch et al., 2012; Malenovsky et al., 2012) provide valuable high-resolution (HR) information for a wide variety of land applications (Malenovsky et al., 2012) including crop monitoring.

From a methodological point of view, LAI retrieval can be faced following either statistical, physical, or hybrid methods (Camps-Valls et al., 2011; Verrelst et al., 2015a; Campos-Taberner et al., 2015b). Parametric statistical methods have been developed through empirical relations between *in situ* LAI acquisitions and vegetation indices

* Corresponding author.

E-mail address: Manuel.Campos@uv.es (M. Campos-Taberner).

derived from airborne and satellite spectra (Haboudane et al., 2004). Alternatively, non-parametric methods do not assume an explicit (parametric) relation between LAI and the spectral reflectance. Non-parametric models estimate a variable of interest using a training database of biophysical parameter and spectral data pairs. In the last decade, non-parametric techniques excelled in biophysical parameter retrieval, either following pure *statistical* or *hybrid* approaches. Many methods have been used in a wide range of applications (Atzberger and Richter, 2012; Verrelst et al., 2015a). Specifically, current operational vegetation products such as LAI are typically produced with neural networks (NN) (Baret et al., 2007, 2013). Nevertheless, in the recent years, Gaussian process regression (GPR) (Rasmussen and Williams, 2006) provided encouraging results in the framework of biophysical parameter estimation outperforming the rest (Camps-Valls et al., 2016; Lazaro-Gredilla et al., 2014; Verrelst et al., 2012a, 2015b).

In statistical approaches, concomitant *in situ* measurements of the biophysical parameter of interest and the associated spectral data from remote sensing platforms are used as a training database, whereas the hybrid approaches rely on a database generated by a radiative transfer model (RTM). The advantage of hybrid approaches is that a broad range of land cover situations can be simulated (e.g. up to hundred thousands), leading to a data set much bigger than what can be collected during a field campaign. RTMs are based on the physical knowledge describing the interactions between radiation, canopy elements, and the soil surface. RTMs have been used for modeling different types of vegetation, making them suitable for general-purpose retrieval applications (Jacquemoud et al., 2000). In particular, the PROSAIL RTM (Jacquemoud et al., 2009), which results from the PROSPECT leaf optical model (Jacquemoud and Baret, 1990) coupled with the SAIL canopy reflectance model (Verhoef, 1984), has been used in several remote sensing studies. PROSAIL has been successfully applied to a variety of crops (Duan et al., 2014). However, it is worth noting that RTM inversion poses several methodological problems: it may lead to high computational cost, and being an ill-posed problem, it may give rise to unstable results. Prior information related to the distribution of the canopy variables and representative background spectra can be implemented in the RTM to better address the inversion process (Combal et al., 2003; Meroni et al., 2004).

Direct validation comparing LAI predictions with *in situ* LAI measurements is needed to report the accuracy of LAI retrievals. LAI ground measurement methods can be divided into two categories: direct and indirect methods (Breda, 2003; Jonckheere et al., 2004). Direct methods involve destructive techniques and require a field effort in collecting leaf samples. Due to the difficulty of continuous applications of direct methods on a large scale, the use of indirect methods based on measurements of the transmission of radiation through the canopy have been widely used (Weiss et al., 2004). In this context, to exploit new technologies in crop monitoring, smartphones have been used for indirect rice LAI measurements reporting good consistency and performance compared with classical instruments such as LAI-2000 or LAI-2200 Plant Canopy Analyzers (LI-COR, Inc., Nebraska, USA) and digital hemispherical photography (DHP) (Campos-Taberner et al., 2016).

In this paper, we present an operational chain for LAI retrieval from Landsat and SPOT5 satellite data specifically calibrated for rice crops. One specific characteristic of rice cultivation is the land preparation followed by the pre-season flooding, this agronomic practice characterizes worldwide the majority of rice cropping systems (namely, irrigated rice, lowland rain-fed rice and deep-water rice) which account for over 90% of the 154 million ha (Maclean, 2002) cultivated with rice each year. In the European temperate rice cultivation areas, the background of the fields is dry at the beginning of the season, and it remains in this condition until early-May when the fields are starting to be flooded. From this date on, the soil

background is flooded most of the time except in some dates when the farmers pump out the water for agronomic purposes. From a remote sensing point of view, this characteristic determines a strong change in the soil background conditions from dry soil to standing water (Boschetti et al., 2014). Thus, the intermittent flooding in a paddy rice field generates an uncontrolled reflectance signature, which may confound the retrieval of rice LAI. It is therefore relevant to test methods for this specific crop due also to its worldwide importance.

In this study, LAI is estimated using hybrid methods through the generation of a reflectance and associated LAI database from the PROSAIL model and powerful nonlinear inversion methods, such as neural networks, kernel ridge regression (KRR) and Gaussian process regression. Based on experimental performance in our datasets, we focus on GPR. This work uses for the first time GPR as the regression tool for multitemporal LAI production chain during the rice cycle. This study shows results produced in the framework of the ERMES project (<http://www.ermes-fp7space.eu/>) where rice monitoring is performed exploiting seasonal remote sensing data and crop modeling as a demonstration of potential operation system. Actually, Landsat data were processed in near real time and the corresponding LAI maps were provided through the web-based geoportal in the project framework to the crop modelers. Contrarily, SPOT5 Take5 data were analyzed in backcasting to assess the potential use of Sentinel-2 data. Spatial and temporal consistency of the LAI estimates were validated with the available multitemporal *in situ* measurements collected in two ERMES study areas (Italy and Spain) from sowing up to rice LAI peak.

The remainder of the paper is organized as follows. Section 2 describes the study areas for which LAI maps are derived, the *in situ* LAI measurements and the remote sensing surface reflectance data. The parameterization of the PROSAIL model, as well as the theoretical basis of the regression methods, are presented in Sections 3.1 and 3.2 respectively. Section 4 evaluates the spatial-temporal consistency of the obtained LAI estimates and provides an analysis of LAI trends for different varieties and management practices as well as discusses the impact of rice background in multitemporal LAI retrievals. Finally, Section 5 outlines the main conclusions and discusses the utility of the obtained results in the framework of operational rice monitoring systems.

2. Data collection

2.1. Study areas

ERMES aims to develop a prototype of COPERNICUS down-stream services based on the assimilation of Earth observation (EO) and *in situ* data within crop modeling solutions dedicated to the rice sector. In this framework, the ERMES study areas have been selected in Spain (Valencian area), Italy (Piedmont and Lombardy rice district) and Greece (Thessaloniki area), which are the three countries responsible of 85% of total European rice production. In this study, we focus on the Italian and Spanish ERMES local study areas. Within each study area, rice is a common crop with a long tradition and economical value. Rice varieties in the Spanish area are mainly *Bomba* and *Senia* belonging to the Japonica group. Sowing activities are concentrated around May 10–15th and fields are managed by keeping them flooded for most of the time during the rice growing period. In Italy, the rice cropping systems are much more variable. About 180 varieties are cultivated covering both Japonica and Indica groups characterized by short and long cycles with a duration from 120 to >150 days, respectively (Boschetti et al., 2009). For this reason, the sowing date can vary from the beginning of April to mid of May. Moreover, in Italy, the so-called dry seeding technique is increasing year after year. This technique consists of seeding rice in rows with a common seeder in dry soil condition and held without water until

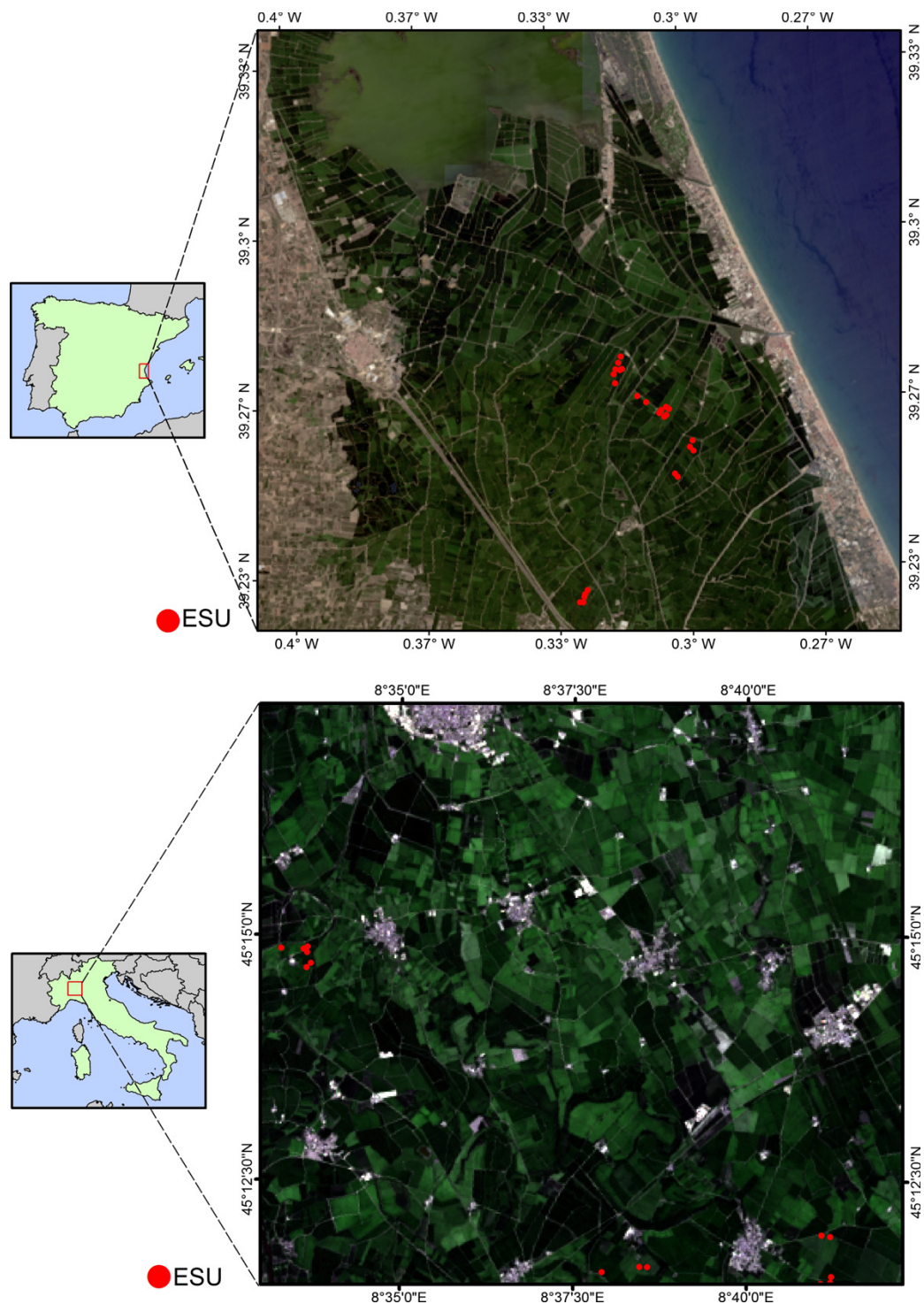


Fig. 1. Study areas: Landsat 8 OLI surface reflectance RGB composite of the Spanish study area on 10 August 2015 (top) and Landsat 8 OLI surface reflectance RGB composite of the Italian study area on 7 August 2015 (bottom). Locations of the ground measurements are displayed as red dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the unfolding of 2–4 leaves, and then crops are flooded. In 2014, it was estimated that more than 30% of the rice cultivated area adopted this technique.

The Spanish study area is located in selected farms within the rice district of Valencia, East of Spain (see Fig. 1 (top)) belonging to the *Albufera Natural Park*, which is included as a special protection area

in the Natura 2000 network by the European Commission allowing only rice crop practices (European Commission, 2011). The area has a typical Mediterranean climate, with an average annual temperature and humidity of 17 °C and 65%, respectively. The mean annual precipitation is approximately 430 mm, mainly recorded in autumn. The Spanish site is a homogeneous rice planting area of approximately 10 km × 20 km extension. The Italian study area is the Lomellina rice district, which is located in the south-western Lombardy region, between the Ticino, Sesia and Po rivers where rice is the dominant crop (>90%). It is the rural part of the Pavia province, which includes 58 municipalities, and is particularly renowned for its rice cultivation (see Fig. 1 (bottom)).

2.2. Ground measurements

In the framework of the 2015 ERMES activities, LAI ground measurements were conducted over the previous study areas. Based on land cover distribution in the areas and information provided by farmers at the very beginning of the rice season, a reliable sampling was achieved selecting ESUs (elementary sampling units) with different rice varieties and sowing dates in order to cover as much as possible the variability of the study areas. In Spain, the total number of ESUs was 40. For some of the ESUs, field acquisitions were made in all field campaigns, whereas in other ESUs the temporal frequency was approximately one every two field campaigns. In the Italian study area, 10 ESUs were fixed and considered during all the field campaigns, increasing the number of ESUs up to 19 for some dates (see Table 1). The same sampling scheme was used over each ESU, following the guidelines and recommendations of the Validation of Land European Remote sensing Instruments (VALERI) protocol (<http://w3.avignon.inra.fr/valeri/>). For the case of row crops, this protocol suggests to take measurements along small transects between rows incorporating some random acquisitions to prevent possible biases in the characterization of the row effect. We adopted the same schema also for fields with broadcast seeding. The size of the ESUs was approximately 20 m × 20 m, and the locations were far from the field borders. In order to characterize the spatial variability within each ESU, a range of 18 to 24 measurements was taken. This number of replicates allows to obtain a statistically significant mean LAI estimate per ESU. The center of the ESU was geo-located using a GPS for later matching and associate the mean LAI estimate with the corresponding satellite spectra.

LAI estimates were acquired in the two countries with smartphones using an app called PocketLAI (Confalonieri et al., 2013). PocketLAI computes indirect LAI measurements through the segmentation of images acquired at 57.5° below the canopy and showed good performance in canopies with different structures (Francone et al., 2014). PocketLAI estimates can reproduce destructive LAI measurements with acceptable results in terms of both reliability and accuracy (Confalonieri et al., 2013). A schematic of a single LAI measurement using PocketLAI and the theoretical background are shown in the supplementary material (see Appendix A). Since the

gap fraction-LAI relationship is not linear, it is not equivalent to first average the gap fraction and then estimate the LAI than the contrary (Weiss et al., 2004). However, specifically over rice crops, we have recently shown that LAI measurements taken with PocketLAI align well with other traditional acquisition techniques, such as plant canopy analyzers and digital cameras for hemispherical photography (Campos-Taberner et al., 2016). Together with PocketLAI data, ancillary information such as rice phenology and flood condition was taken on each ESU to better evaluate the computed LAI maps.

2.3. Landsat imagery

The United States Geological Survey (USGS) facilitates free access to Landsat archive data (Woodcock et al., 2008). In this work, Landsat 8 OLI data (30 m pixel resolution) were downloaded through the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) (<http://espa.cr.usgs.gov/>). The provisional Landsat 8 Surface Reflectance product was used (LaSRC) (Vermote et al., 2016). In order to focus on the rice areas only, the Landsat 8 OLI images were cropped to 1500 × 800 and 480 × 431 pixel size in the case of Spain and Italy, respectively. Images were available every 16 days in Italy. The Spanish study area lies in two Landsat paths within the same row (198/33 and 199/33), increasing the temporal resolution of the images to seven and nine days. Landsat 8 OLI surface reflectance spectral bands were filtered to relate only the blue (B), green (G), red (R), near infrared (NIR), and the two short wave infrared (SWIR1, SWIR2) channels with the ground LAI measurements in the retrieval process.

Cloud contamination is a common problem which limits the utility of passive optical multispectral images. To deal with this problem, Landsat 7 ETM+ data were used for increasing the temporal resolution of cloud free images. Nevertheless, Landsat 7 data can be affected by data gaps causing lost information of about 22% of the pixels in ETM+ images (referred to as SLC-off images) (Arvidson et al., 2006; Ju and Roy, 2008). In this study, we used the Neighborhood Similar Pixel Interpolator (NSPI) algorithm (Chen et al., 2011), which makes use of the neighboring pixels with similar spectral characteristics to predict the value of missing pixels using spatial and temporal information of gap free images (Campos-Taberner et al., 2015a). To compute the spectral similarity between the target pixel and neighboring pixels, only cloud-free observations are used, thus exploiting information about the cloud mask and the temporal trajectory. Fig. 2 shows the available cloud-free Landsat imagery acquired during the 2015 rice season over the study areas.

It is worth noting that Landsat 7 ETM+ and Landsat 8 OLI have different spectral response functions. This leads to slight differences in surface reflectance for each sensor. Nevertheless, differences are low when surface reflectance is simulated (see Fig. S2 of the supplementary material in Appendix A). On the other hand, another source of discrepancies in LAI retrievals is the atmospheric correction used for the generation of Landsat 7 ETM+ and Landsat 8

Table 1
Dates and number of ESUs sampled on the study areas during 2015.

Spain			Italy		
Date	DOY	ESUs	Date	DOY	ESUs
02/06	153	24	26/05	146	10
09/06	160	24	04/06	155	16
17/06	168	22	15/06	166	18
24/06	175	20	24/06	175	19
02/07	183	22	06/07	187	19
09/07	190	22	21/07	202	19
17/07	198	19	07/08	219	19
23/07	204	20			
03/08	215	27			

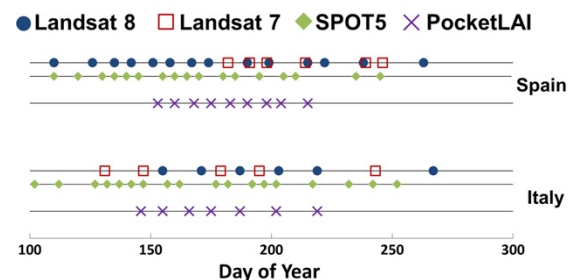


Fig. 2. Remote sensing data acquired during the 2015 rice season over the Spanish and Italian study areas and the *in situ* PocketLAI acquisitions.

OLI surface reflectance data. Although the Landsat 7 ETM+ atmospheric correction (LEDAPS) is less accurate than the Landsat 8 OLI atmospheric correction (Vermote et al., 2016) differences in surface reflectance and LAI estimates of two consecutive images are small (Roy et al., 2016).

2.4. SPOT5 imagery

With the aim of simulating multitemporal data from Sentinel-2 mission, the SPOT5 Take5 experiment provided 10 m pixel resolution data over selected sites every 5 days under constant angles from end-April to early-September 2015, thus covering the majority of the vegetation phase for summer crops in Europe. In this framework, ESA launched a call for proposals of sites selection, and the ERMES study areas were included in the 2015 SPOT5 Take5 acquisition plan. A total of 17 and 18 SPOT5 cloud-free images were acquired over the Spanish and the Italian study areas (see Fig. 2). The imagery were downloaded through the Theia land data center, which provides a top of canopy surface reflectance product (green, red, near infrared and short wave infrared channels) obtained applying an ortho-rectified process (Baillarin et al., 2008) and then computing an atmospheric correction with MACCS Software (Hagolle et al., 2008). Images were spatially cropped to 1440×1293 (Italy) and 4500×2400 (Spain) pixel size in order to match with the extent of the corresponding Landsat imagery.

Note that pixels corresponding to urban areas, water bodies and areas of 'no interest' such as non-rice crops were masked out during the retrieval process in order to avoid meaningless LAI estimates over those surfaces. The masking process was realized using official parcel boundaries and farmers crop declaration for CAP (European Common

Agricultural Policy) obtained from the Valencian government (<http://terrasit.gva.es/>) and the Lombardy regional authorities (<https://www.dati.lombardia.it/browse?category=Agricoltura>).

3. Retrieval methodology

A general outline of the proposed retrieval chain is shown in Fig. 3. The basic ingredients include the PROSAIL model and statistical regression algorithms for model inversion. Once the regression models are developed, we apply them to retrieve high-resolution LAI maps on the study areas from corresponding Landsat and SPOT5 surface reflectance data. In the following subsections we describe these components in detail.

3.1. PROSAIL model

The PROSAIL radiative transfer model (Jacquemoud et al., 2009) was used to build the database for training the retrieval model. It assumes the canopy as a turbid medium for which leaves are randomly distributed. PROSAIL simulates the top of canopy bi-directional reflectance in the range of 400 nm to 2500 nm as a function of input variables related to the structure of the canopy, the leaf optical properties, the background soil reflectance and the sun-view geometry. Leaf optical properties are expressed taking into account the mesophyll structural parameter (N), leaf chlorophyll (C_{ab}), dry matter (C_m), and water (C_w) contents. C_w was tied to the dry matter content ($C_w = C_m \times C_{wREL} / (1 - C_{wREL})$) assuming that green leaves have a relative water content (C_{wREL}) varying within a relatively small range (Baret et al., 2007) (see Table 2). On the other hand, canopy structure is characterized by the average leaf angle

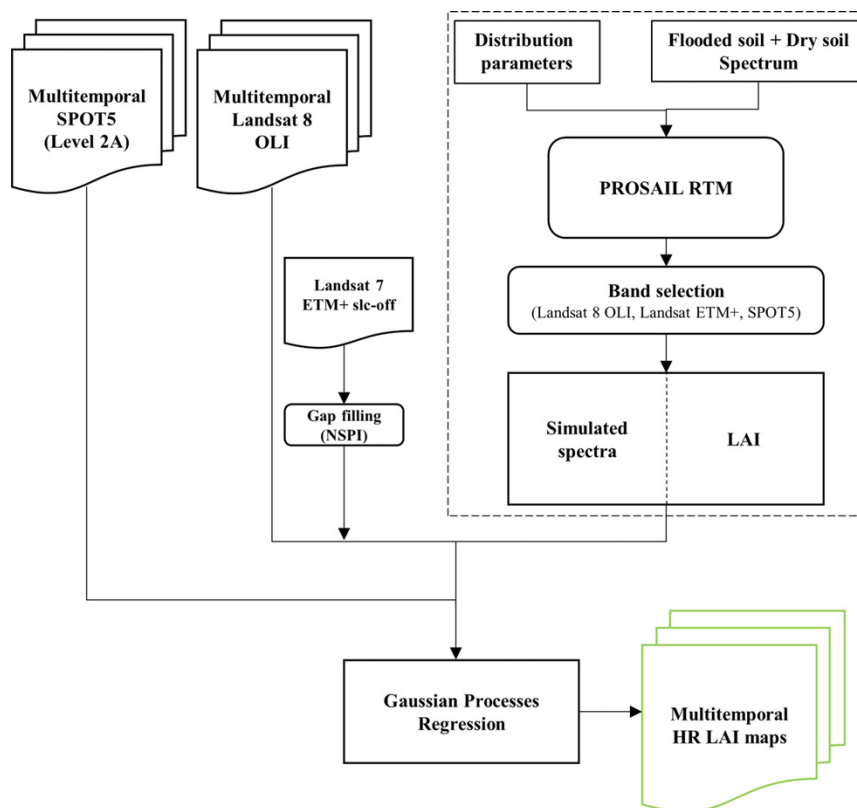


Fig. 3. Operational chain followed in the multitemporal LAI retrieval.

Table 2
Distribution of the canopy, leaf and soil parameters within the PROSAIL RTM.

Parameter		Min	Max	Mode	Std	Type
Canopy	LAI (m ² /m ²)	0	10	3.5	4.5	Gaussian
	ALA (°)	30	80	60	20	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.5	1	1	0.2	Truncated Gaussian
Leaf	N	1.2	2.2	1.5	0.3	Gaussian
	C _{ab} (μg • cm ⁻²)	20	90	45	30	Gaussian
	C _{dm} (g • cm ⁻²)	0.003	0.011	0.005	0.005	Gaussian
	C _{wREL}	0.6	0.8	–	–	Uniform
Soil	β _s	0.3	1.2	0.9	0.25	Gaussian

inclination (ALA), the LAI and the hot-spot parameter (Hotspot). A multiplicative brightness parameter (β_s) was introduced and applied to spectral flooded and dry soil signatures to represent different background reflectance types (Baret et al., 2007; Claverie et al., 2013). The system geometry was described by the solar zenith angle (θ_s), view zenith angle (θ_v), and the relative azimuth angle between both angles (Δθ).

Sub-pixel non-vegetated areas were found in the borders of rice fields but patches of bare/flooded soil, small water stripes and channels were found in paddies as well. These conditions can be due to water drainage, very poor soil fertility, flattening mechanical process and other causes related to agro-practices leading to high yield reduction. The interested reader is referred to Fig. S4 of the supplementary material attached in Appendix A for further details. Therefore, in order to account for these mixed conditions, a pixel can be represented by a linear mixture of vegetation (vCover) and bare/flooded soil (1-vCover) spectra. A linear spectral mixing model was assumed for the sake of simplicity. Taking this heterogeneity into account, the pixel reflectance can be expressed as $R = R_{veg} \times vCover + R_{soil} \times (1 - vCover)$, where R_{veg} and R_{soil} account for pure vegetation reflectance and background, respectively. This simple approach was introduced by Baret et al. (2007) to account for clumping at the landscape level. Note that when validating retrieved LAI of a mixed pixel, it is compared with LAI \times vCover (see distributions of LAI and vCover in Table 2). In this study, vCover was assumed to be independent of LAI, following a truncated Gaussian distribution. In addition, a 5% of pure background spectra (vCover=0) were added to represent situations at the beginning of the season (no vegetation) and large patches of bare/flooded soil present during the rice season.

The leaf and canopy variables as well as the soil brightness parameter were randomly generated following specific distributions (see Table 2). The parameterizations were similar to other studies using high-resolution sensors (Bsaibes et al., 2009; Duveiller et al., 2011). Nevertheless, in this study, a site-specific parameterization of the PROSAIL model based on the available 2014 ERMES field measurements was selected in order to constrain the behavior of the model to Mediterranean rice areas reducing the equifinality of the ill-posed PROSAIL inversion process (Combal et al., 2003). During the ERMES 2014 rice season, leaf chlorophyll content was measured with a SPAD-502 chlorophyll meter (Campos-Taberner et al., 2016) thus allowing to constrain the C_{ab} range to Mediterranean rice values which typically vary from 35 to 70 μg • cm⁻² depending on the rice variety. This range was slightly extended in order to represent rice with high leaf chlorophyll concentration due to fertilization practices, as well as low leaf chlorophyll content caused either by possible diseases, blasts or nitrogen deficits. ALA distribution was selected for accounting specific leaf inclination during rice phenological stages (Zhang et al., 2013). In order to better constrain the retrieval to rice crops, a spectral library of underlying soil background was generated by considering signatures of homogeneous flooded and dry areas identified within rice fields in the study areas with Landsat and SPOT5 imagery, which was spectrally

matched with typical rice background spectra collected by Boschetti et al. (2014). Each background signature was selected randomly from the spectral library and multiplied by the soil brightness factor (β_s), which was assumed to follow a Gaussian distribution (see Table 2). β_s is assumed to be sensitive to soil moisture, roughness and geometrical configuration (Baret et al., 2007). A random Latin hypercube sampling design allowed to populate more evenly the canopy realization space (Mckay et al., 2000). For each sensor, a single dataset of PROSAIL simulations was performed, which included all geometrical configurations. The distributions for the system geometry were randomly generated based on information in imagery metadata.

3.2. Regression methods

In this paper we propose inverting PROSAIL using machine learning statistical algorithms. For this purpose, we used three representative nonlinear regression methods: the familiar artificial neural network, and two related kernel-based regression algorithms: the kernel ridge regression and the Gaussian process regression. This section reviews the three regression methods used and discusses about the implementation issues for the interested reader.

3.2.1. Neural networks (NN)

Artificial neural networks are based on the combination of simple nonlinear processing units, called neurons, into a fully connected hierarchical architecture. The network can model complex, nonlinear input-output relations, and has been the preferred regression and function approximation tool for decades for retrieving biophysical parameters. Actually, the vast majority of hybrid inversion methods consider the use of neural networks (Baret et al., 1995; Baret and Fourty, 1997; Smith, 1993) for retrieval of canopy parameters. Essentially, each neuron in a network performs a linear regression followed by a non-linear activation (sigmoid-like) function. Neurons of different layers are interconnected by weights that are adjusted during the training phase (Haykin, 1999). In order to train the network (i.e. fit the weights), one has to select a cost function (in our case the least squares loss) and an algorithm to do this (in our case the standard back-propagation algorithm). Several hyperparameters are involved as well and impact the solution: essentially, the number of hidden layers and neurons/nodes and the learning rate of the back-propagation algorithm.

3.2.2. Kernel methods

Kernel methods (Shawe-Taylor and Cristianini, 2004) owe their name to the use of kernel functions, which measure similarities between input data examples. We used two related and powerful kernel methods for regression: kernel ridge regression (Shawe-Taylor and Cristianini, 2004), and Gaussian process regression (Rasmussen and Williams, 2006). The KRR is considered as the nonlinear (kernel) version of the canonical least squares linear regression, while the GPR is a probabilistic approximation to nonparametric kernel-based regression, where both a predictive mean (point-wise estimates of LAI) and predictive variance (error bars for the LAI predictions) can be

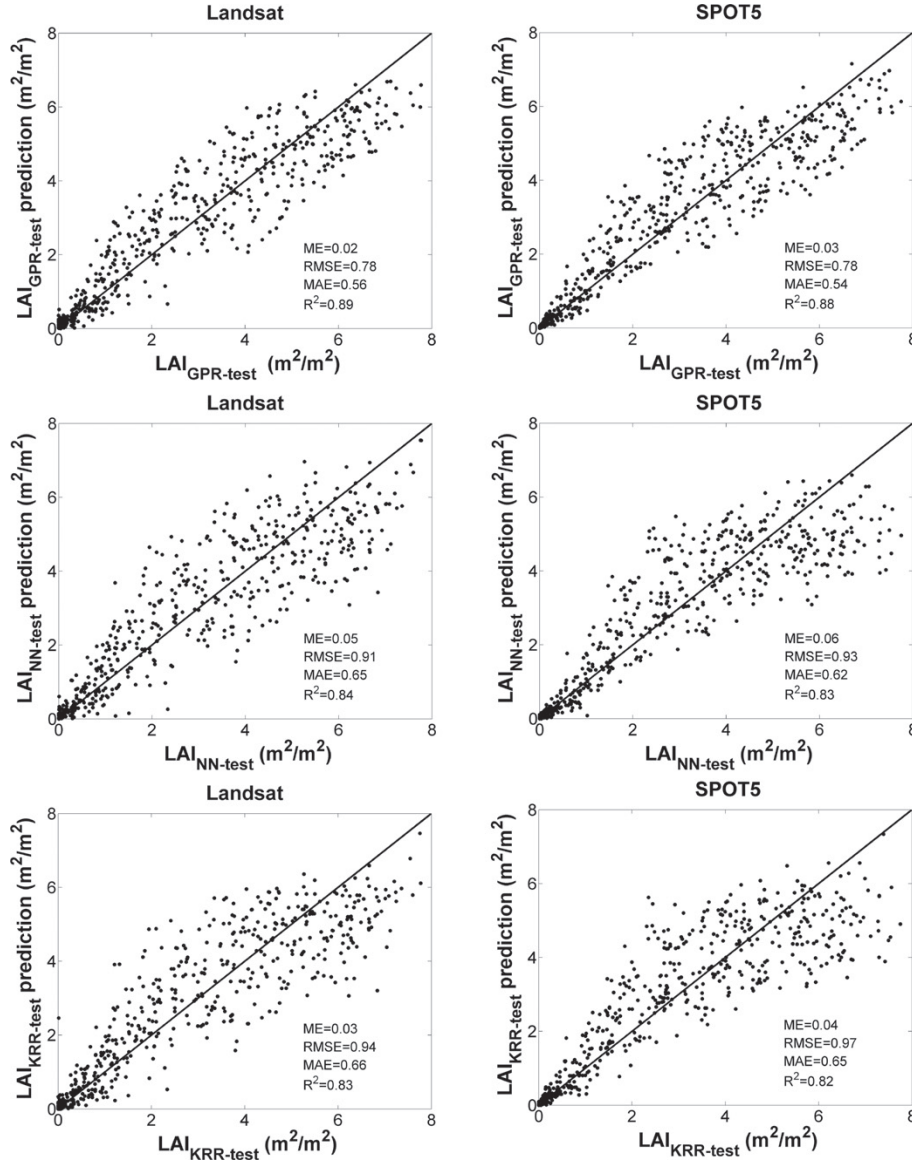


Fig. 4. GPR (top), NN (middle) and KRR (bottom) theoretical performances using Landsat 8 OLI (left) and SPOT5 (right) simulated data. Plots are obtained by applying the trained models on test data. Each method was trained following either a standard fold cross-validation within the training set (KRR and NN) or by maximizing the marginal log-likelihood of the observations, which is analytical (for the GPR).

derived. Notationally, both methods offer the same explicit form of the predictive model, which establishes a relation between the input (e.g., spectral data) $\mathbf{x} = [x^1, \dots, x^B] \in \mathbb{R}^B$ and the output variable (i.e., LAI) $y \in \mathbb{R}$ of the form:

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^N \alpha_i k_{\theta}(\mathbf{x}_i, \mathbf{x}) + \alpha_0, \quad (1)$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the spectra used in the training phase, α_i is the weight assigned to each one of them, α_0 is the bias in the regression function, and k_{θ} is a kernel or covariance function (parametrized by a set of hyperparameters θ) that evaluates the similarity between the test spectrum and all N training spectra.

In order to generate a kernel regression model, one needs to specify a covariance/kernel function k_{θ} , to infer its hyperparameters θ and model weights α . For the KRR prediction model, we used the squared exponential (SE) kernel: $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2))$, which is simply parameterized by $\theta = \sigma$ (also known as the kernel length-scale) that needs to be tuned typically by cross-validation. For the GPR prediction model, we used the so-called automatic relevance determination (ARD) kernel, as an alternative generalization of the isotropic SE prior:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp \left(- \sum_{b=1}^B \frac{(x_i^{(b)} - x_j^{(b)})^2}{2\sigma_b^2} \right) + \sigma_n^2 \delta_{ij}, \quad (2)$$

where ν is a scaling factor, B is the number of bands, and σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b . Model hyperparameters are collectively grouped in $\theta = [\nu, \sigma_n, \sigma_1, \dots, \sigma_B]$, and model weights α_i can be automatically optimized by maximizing the marginal likelihood in the training set (Rasmussen and Williams, 2006; Verrelst et al., 2012b). The obtained weights α_i after optimization give the relevance of each spectrum \mathbf{x}_i , while the inverse of σ_b represents the relevance of each band b . Hence, low values of σ_b indicate a higher informative content of this certain band b to the training function k . In this paper we study both the KRR and the GPR models paying attention to the band relevance conveyed by the inferred σ_b values, and the prediction uncertainty provided by the GPR model.

3.2.3. Model development and source code

Inference of the hyper-parameters for all methods and the weights for doing predictions was done as follows. We first generated 2000 data pairs (reflectances–LAI values) with PROSAIL, and used 70% for model selection (training set), and then evaluated and showed the results in the remaining 30% test set, which was never used or seen in model fitting. Even though 1400 samples could seem to be insufficient for training purposes, increasing this number of samples did not have a significant impact in the accuracy of the retrievals for all methods in the test set, indicating that they did not incur in any overfitting issue and highlighting the good representativity of the simulated data. For neural networks, the hyperparameters to be chosen were the number of hidden layers and neurons (for simplicity we evaluated one hidden layer and 2–30 hidden neurons) and the learning rate between 0.001 and 0.1 in log-scale. The bias input was set to -1 (not adjusted during training) and different initializations of the weights were tested. For the KRR model, we varied the length-scale σ (between 0.1 and 10 times the average distance between all training points in log-scale) and the regularization parameter σ_n (between 10^{-5} and 10^{-2} in log-scale). For the case of the GPR, we inferred the hyperparameters $\theta = [\nu, \sigma_n, \sigma_1, \dots, \sigma_B]$ and model weights using an optimization of the evidence. All experiments were conducted with our SimpleR MATLAB toolbox, freely available at <http://isp.uv.es/>. The toolbox is intended for practitioners with little expertise in machine learning, and that may want to assess advanced methods in their problems easily. The toolbox compares numerically and statistically the algorithms by simply entering the input-output (e.g. reflectances and LAI values) data matrices.

4. Results and validation

This section is devoted to show the theoretical performance of the regression methods and experimental evidence of the performance of the proposed processing scheme. We pay attention to the

derived HR LAI maps, and illustrate the usefulness for multitemporal rice crop monitoring through a temporal evolution analysis of LAI estimates compared to *in situ* measurements in the different test sites.

4.1. Accuracy assessment over the simulated dataset

It was necessary to build a dedicated model for each of the three remote sensing datasets (Landsat 7 ETM+, Landsat 8 OLI and SPOT5) depending on their spectral bands and angular configurations. We first evaluate the GPR performance compared to NN and KRR (see Fig. 4) over the test dataset (600 samples) corresponding to Landsat 7 ETM+, Landsat 8 OLI and SPOT5 datasets (Landsat 7 ETM+ results were similar to Landsat 8 OLI and are not shown for brevity). Hereafter, we will refer simply to ‘Landsat’ estimates, irrespectively of being for Landsat 7 ETM+ or Landsat 8 OLI. Likewise, we refer to GPR-Landsat and GPR-SPOT5 to the GPR models built using Landsat and SPOT5 data, respectively. Results revealed good accuracy and low bias in both Landsat and SPOT5 simulated reflectances. GPR outperformed NN and KRR in all statistical quality measures (see Fig. 4). GPR-Landsat and GPR-SPOT5 performances were robust and very similar, revealing biases of 0.02 and 0.03 m^2/m^2 respectively, while a root mean squared error (RMSE) of 0.78 m^2/m^2 and high determination coefficients ($R^2 > 0.87$) were obtained in both cases. Fig. 5 exhibits the evolution of the average RMSE as a function of the number of predictions, and shows a consistent better performance (i.e. improved robustness to reduced-sized datasets) of GPR over NN and KRR. The curves are the result of averaging a number of realizations, and for each realization, we computed the RMSE with a fixed number of predictions chosen at random. In the limit of 100%, one obtains the results in Fig. 4. This particular plot tries to analyze models’s robustness to local consistency and reliability of the estimated error (Montavon et al., 2013). The shape of the obtained curves are similar for all methods (high variance when few predictions were evaluated, and rapid convergence to stable RMSE), yet better for GPR. These results revealed GPR as the most accurate and robust regression method in both Landsat and SPOT5 test sets.

As discussed earlier (cf. Section 3.2), GPR provides a σ_b parameter whose reciprocal represents the relevance of each band in the regression. Hence, it is possible to identify the most relevant bands used by the GPR model, which it is not affordable when using NN and KRR. Specifically, the most relevant bands in both cases were the green and near infrared bands (see Fig. 6), which showed the theoretical consistency between GPR model behaviors. The identification of the most valuable bands for LAI retrieval was expected since the green and near infrared bands are more related to the greenness of the vegetation (useful for assessing plant vigor) and leaf area index features,

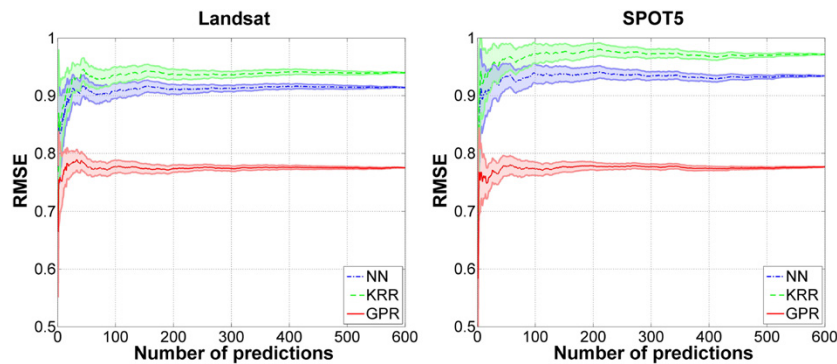


Fig. 5. Performance in the test set averaged over 100 random realizations of the Landsat and SPOT5 training-test data splitting as a function of the number of used predictions.

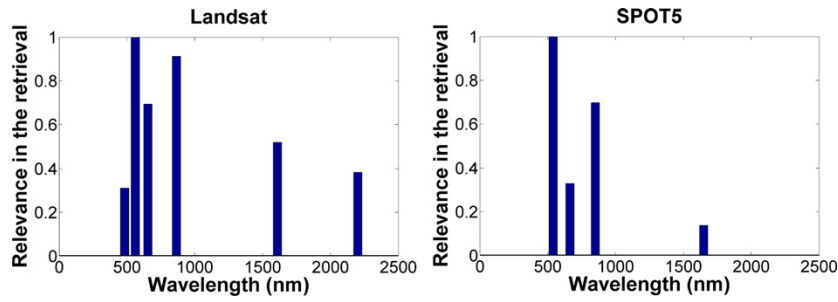


Fig. 6. Relevance band histograms in GPR-Landsat (left) and GPR-SPOT5 (right).

respectively. Similar band ranking was obtained by Verrelst et al. (2015b) although the results are not directly comparable because of the different spatial and spectral resolutions.

In this work, all six Landsat optical bands were used for the GPR-Landsat LAI retrievals although in many studies the blue band is not used because atmospheric effects. Nevertheless, blue band was also considered for LAI retrieval in other studies (Atzberger and Richter, 2012; Borel, 2010; Verrelst et al., 2015b) since the blue spectrum holds information valuable for LAI and phenology (Huete et al., 2002). Actually, GPR is very robust to moderate-to-high dimensional spaces (i.e. few more input variables do not impact results negatively). A comparison of LAI estimates obtained with or without using the Landsat blue band revealed a slight improvement when including the blue band in the GPR retrieval (see Fig. S3 and Table S1 of the supplementary material in Appendix A).

4.2. Accuracy assessment over ground LAI

With the goal of assessing the accuracy of the retrievals, RMSE between the estimates and the *in situ* measurements was computed. RMSE values of 0.39, and 0.38 m^2/m^2 were found in Spain and Italy respectively, showing good accuracy between GPR-Landsat map values and the *in situ* LAI measurements (see Fig. 7). The GPR-SPOT5 retrievals revealed good accuracies as well, showing RMSE values of 0.51 and 0.47 m^2/m^2 in Spain and Italy respectively. Other different statistics, such as the mean error (ME), mean absolute error (MAE) and the coefficient of determination were also computed to evaluate the bias, accuracy, and the goodness-of-fit between GPR-Landsat and GPR-SPOT5 predictions and measurements. A remarkably good correspondence between satellite retrievals and *in situ* measurements was found in the Spanish site, with very low bias for both

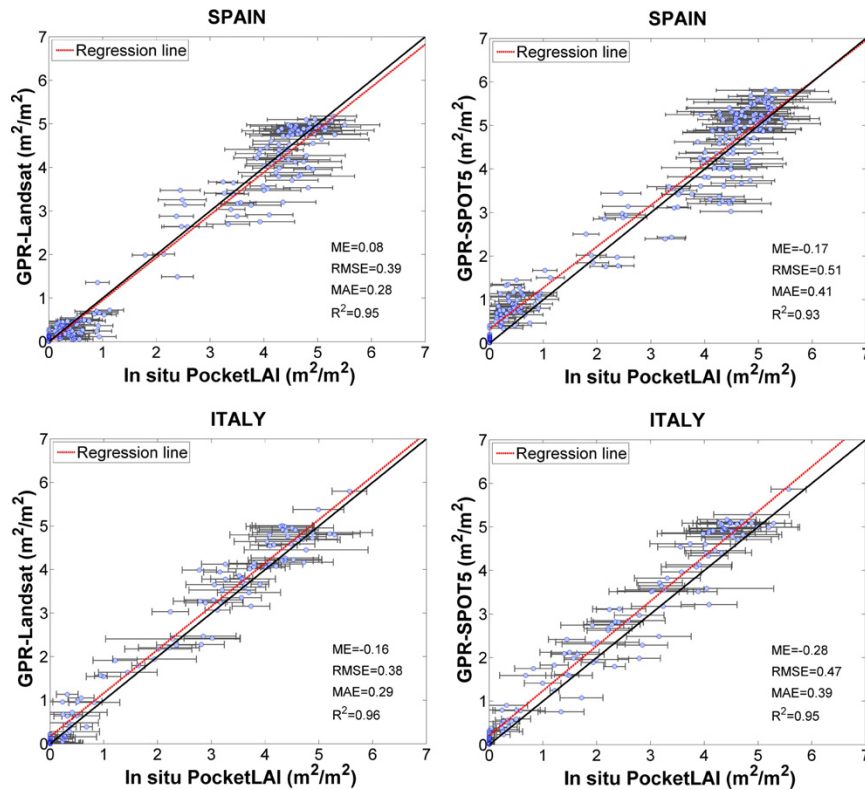


Fig. 7. Scatter plots of estimated LAI values using Landsat (left) and SPOT5 (right) data versus *in situ* LAI measurements acquired with PocketLAI over Spain (up) and Italy (bottom). Standard deviation of measurements is drawn as horizontal error bars as well as bisector line (black line). For the sake of visualization, GPR prediction uncertainty ($\approx \pm 1$) are not shown.

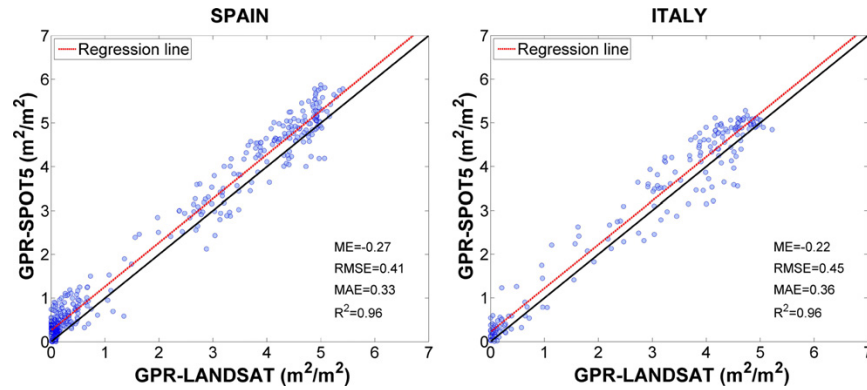


Fig. 8. Scatter plots of the estimated LAI values of the GPR-Landsat and GPR-SPOT5 retrievals in (left) Spain and (right) Italy.

GPR-Landsat and GPR-SPOT5 models. A good agreement and low biases were also observed for the Italian site. In all cases, very high correlations were found with $R^2 > 0.92$. Error bars in Fig. 7 refer to the standard deviation of the field measurements, and are thus related with the heterogeneity of the ESU. In addition, retrievals were also computed using KRR and NN showing slightly less accurate LAI estimates. The interested reader is referred to Fig. S5 of the supplementary material attached in Appendix A.

4.3. Comparison of GPR-Landsat and GPR-SPOT5

Besides the aforementioned theoretical performance of the GPR-Landsat and GPR-SPOT5, the experimental consistency between estimates was assessed with the inter-comparison of multitemporal LAI estimates. The closest acquisition dates between Landsat and SPOT5 were taken into account for the comparison. After the SPOT5 resampling to Landsat resolution, the comparison was achieved averaging the LAI value of the valid estimates computed over the 3×3 pixels if more than 5 out of the 9 pixels were valid (Morissette et al., 2006)

in order to reduce coregistration errors between images and inconsistencies associated to differences in the point spread functions. In general, the estimates provided by both GPRs were highly correlated. However, a slight overestimation of the SPOT5 based retrievals was observed in both study areas (see Fig. 8). In addition, spatial consistency between retrievals was performed for the closest temporal Landsat and SPOT5 images during the rice growing season. The closest useful available images were found in Italy in 2015 July 22nd for Landsat and in 2015 July 21st in the case of SPOT5. Difference ($\Delta = GPR_{SPOT5} - GPR_{Landsat}$) LAI map was computed after the GPR-SPOT5 LAI map resampling into Landsat resolution (30 m). Statistical differences between GPR-SPOT5 and GPR-Landsat from ANOVA (one-way analysis of variance) were computed revealing F -statistic and p -value of 2.860 and 0.001 respectively, highlighting that the two distributions are not statistically different. The obtained results show 77% of the pixels fall within $\pm 0.5 \text{ m}^2/\text{m}^2$ interval and only 1% of the pixels reveal differences higher than $\pm 2 \text{ m}^2/\text{m}^2$ (see Fig. 9 (right)). The prevailing light green color in Fig. 9 (left) demonstrate that the retrievals from the two sensors are coherent and most of the rice areas have

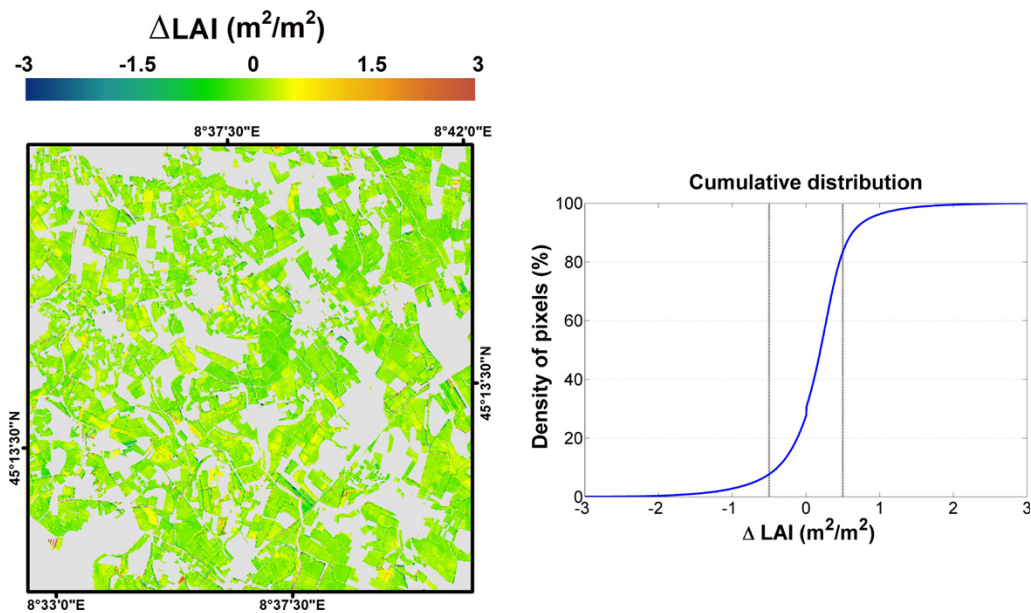


Fig. 9. (Left) LAI differences map between GPR-SPOT5 (2015 July 21st) and GPR-Landsat (2015 July 22nd), and (right) the corresponding cumulative distribution values. Grey mask covers non-rice areas.

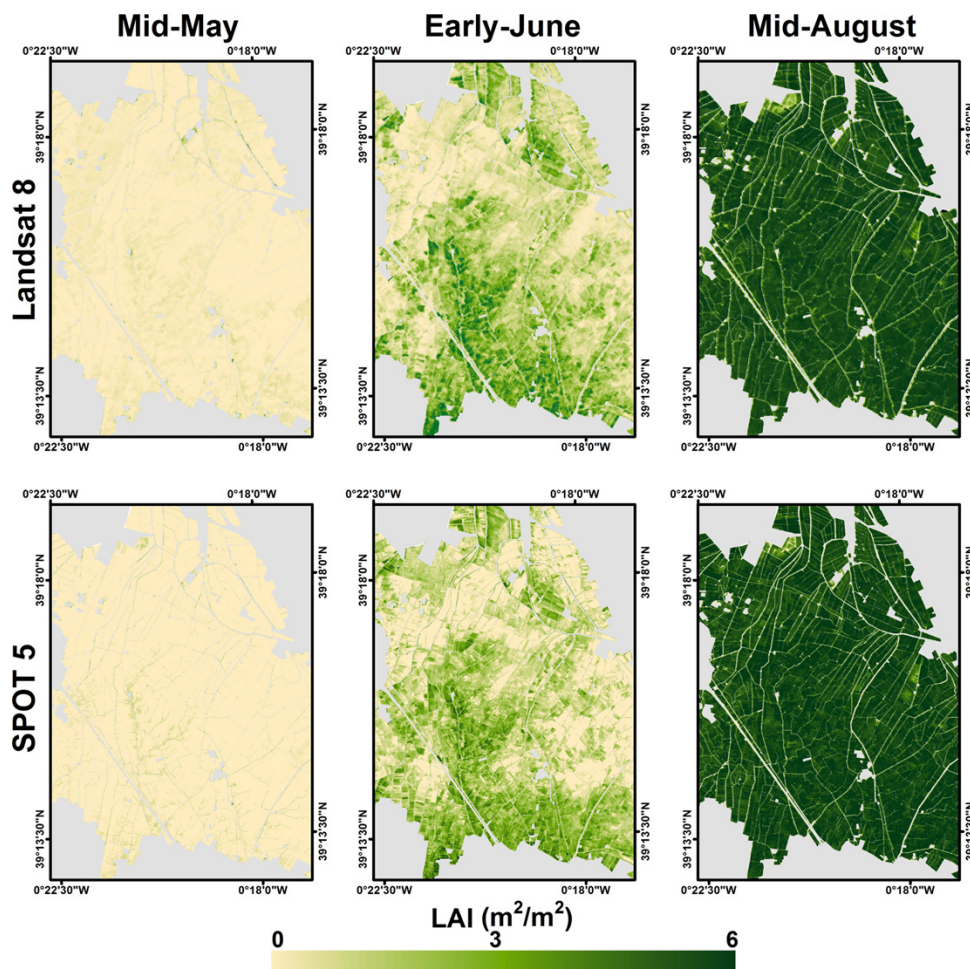


Fig. 10. Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Spanish study area in mid-May (left), early-June (middle) and mid-August (right). Grey mask covers non-rice areas.

LAI differences around zero. Bigger differences can be partially due to the different spatial resolution between Landsat and SPOT5.

4.4. Spatio-temporal analysis

Multitemporal LAI maps were derived for the study areas during the 2015 rice season. Fig. 10 shows the GPR-Landsat and GPR-SPOT5 LAI maps derived for the very beginning of season (mid-May), early growing season (early-June), and maximum leaf rice development (mid-August) in the Spanish site. The corresponding HR estimated LAI maps for the Italian study area are shown in Fig. 11. First inspection of the maps indicates the occurrence of very low LAI values corresponding to mid-May since the sowing dates were around May, 10–15th in Spain. In Italy, for the same period, some rice fields have higher LAI values (>2) because of the early sowing of some rice varieties. In the Spanish site, it can be seen the expected rice emergence in the early-June LAI maps (see Fig. 10 (middle)), while the early-July Italian maps (Fig. 11 (middle)) show already higher LAI estimates due to the advanced phenological growing state corresponding to the rice stem elongation phase. Eventually, the mid-August maps (Figs. 10 and 11 right panels) show the highest LAI estimates because rice plants reached the heading phase and the LAI seasonal peak. All 2015 derived high-resolution maps can be investigated by registering and

joining as local user into the ERMES web-based geo-portal (<http://ermes.dlsi.uji.es/prototype/geoportal/>) or can be found in the ERMES catalogs (<http://get-it.ermes-fp7space.eu/>).

The *in situ* LAI data points allowed us to compare the temporal evolution of field measurements over the study areas (see Fig. 12). In general, LAI estimates derived from the proposed algorithm using GPR-Landsat and GPR-SPOT5 data agree with regard to the seasonal rice phenological cycle in the two countries and followed the temporal dynamics of the ground measurements. The different LAI evolutions that can be observed in Fig. 12, show coherent temporal behaviors as a consequence of either different rice varieties or sowing dates that determine a shift in the development curve. The interested reader is referred to Fig. S6 of the supplementary material attached in Appendix A for further details. It is interesting to note that the difference between Landsat 7 ETM+ and Landsat 8 OLI did not induce large difference in LAI retrieval. For example, LAI estimates of Landsat-8 DOY=190 and Landsat-7 DOY=191 for the Spanish site showed differences lower than $0.25 \text{ m}^2/\text{m}^2$ (see blue profiles in Fig. 12).

On the other hand, both GPR-Landsat and GPR-SPOT5 models also provided a prediction uncertainty for the LAI estimates (see Fig. 12). Although these uncertainties cannot be used as a validation *per se*, they can be useful to draw conclusions about the quality of the

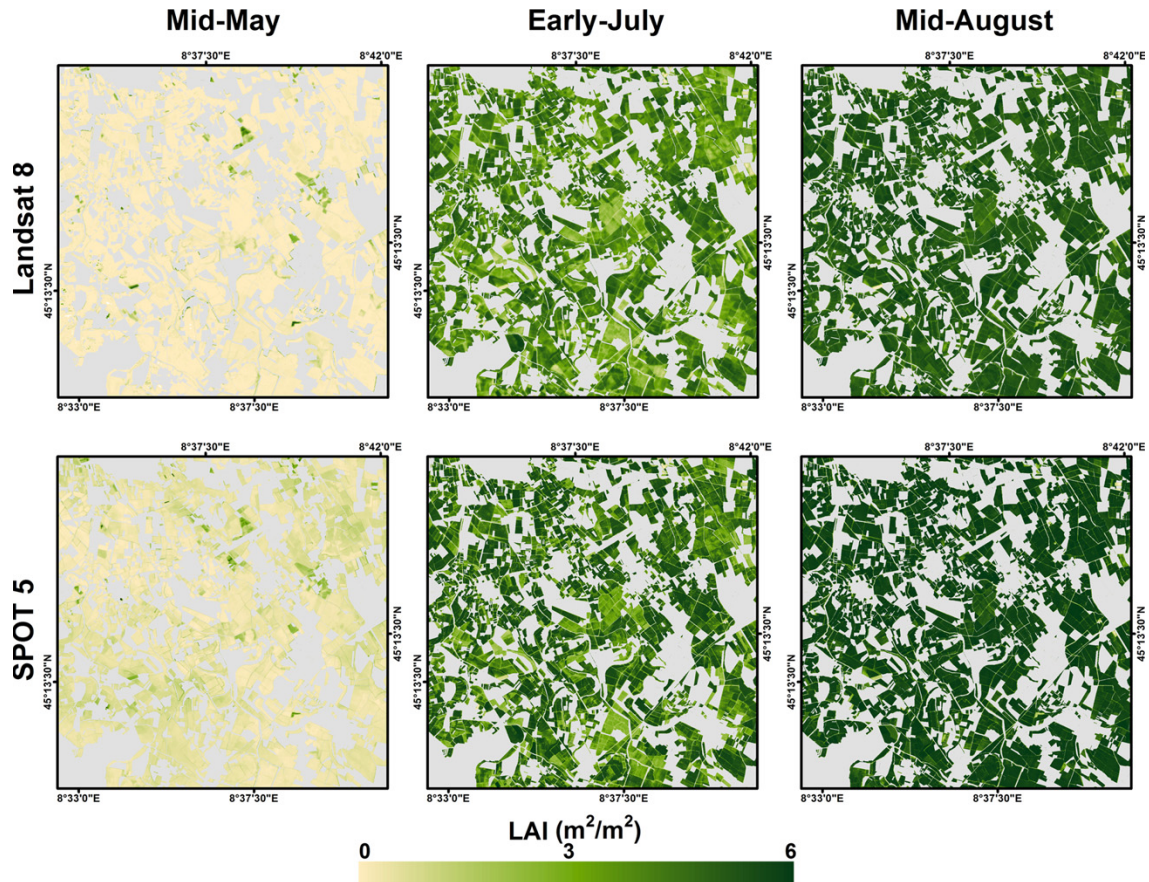


Fig. 11. Estimated LAI maps derived using Landsat (top) and SPOT5 (bottom) images over the Italian study area in mid-May (left), early-July (middle) and mid-August (right). Grey mask covers non-rice areas.

retrievals. Apparently, the prediction uncertainties look constant for the multitemporal LAI estimates, nevertheless a deeper look reveals a change in the behavior: at the beginning of the rice season (no vegetation) values are high (≈ 1.25) and when rice starts to emerge (about day of year 140–150) the uncertainty decreases significantly (≈ 0.95), while during the rice development period it remains virtually constant. This reflects the fact that the simulations include a large amount of cases with intermediate and high LAI values. It is worth mentioning that the GPR prediction uncertainty only depends on the training reflectance. Thus, if the test and training reflectance are deemed similar, the uncertainty for the test set decreases because we are dealing with similar input features. Therefore, prediction uncertainty must be interpreted as a qualitative variable associated to the estimates: the higher the uncertainty, the lower the confidence on the associated estimate. As a matter of fact, this interpretation comes from the PROSAIL training data set used in the GPRs since unrepresented spectra, such as non-vegetated bodies, lead to higher uncertainties for its associated retrievals. In Section 4.5, we analyze the impact of using different background spectra in PROSAIL on LAI prediction uncertainties.

The rate of LAI development depends on the occurrence of specific phenological phases. Green-up corresponds to the end of the tillering phase, the rapid LAI increase occurs during stem the elongation phase and the plateau of maximum LAI is reached in correspondence of heading for all the flowering period. From the analysis of LAI profiles for different ESUs, especially in Italy, it is possible to appreciate

a variability in the occurrence of the mentioned phenological stages due to different sowing dates and cultivated varieties. A deeper inter-comparison of LAI time series derived from GPR-SPOT5 in the Italian case study was conducted over selected monitored fields to better exhibit these behaviors. Fig. 13 reports the average time series of LAI for each monitored field in 2015. A buffer of 5 m was considered in order to get rid of the field border in the average computation. In the figure, the different monitored parcels are indicated in numerical order and different varieties in color (see legend), dotted vertical lines represent the date of sowing and the sowing practices as provided by the farmers (i.e. direct sowing in dry soil or seed broadcasting in flood condition in red and blue color respectively). It is possible to appreciate how rice is usually cultivated in monoculture and sowed in dry bare soil or in water (low/zero LAI value at the beginning of the series). It is noteworthy that anomalous high early LAI values in some fields (fields # 68 or # 384) are indicative of a cover crop preceding the rice season (observation confirmed by farmers communication). An early crop establishment period, about DOY 92 (fields # 50 and # 55), determined crop presence (e.g. LAI value) around DOY 130 to 150 when other fields (fields # 61, # 282 and # 384) were just sown. As a consequence of different sowing dates and varieties, the periods of maximum LAI plateau and senescence occur in different moment. Those fields with an early sowing (DOY 90 to 111) reached the LAI peak around DOY 180 whereas the variety sowed around DOY 140 (Sirio CL, Selenio) reached the LAI peak after DOY 200. The interested reader is referred to Fig. S7 in the

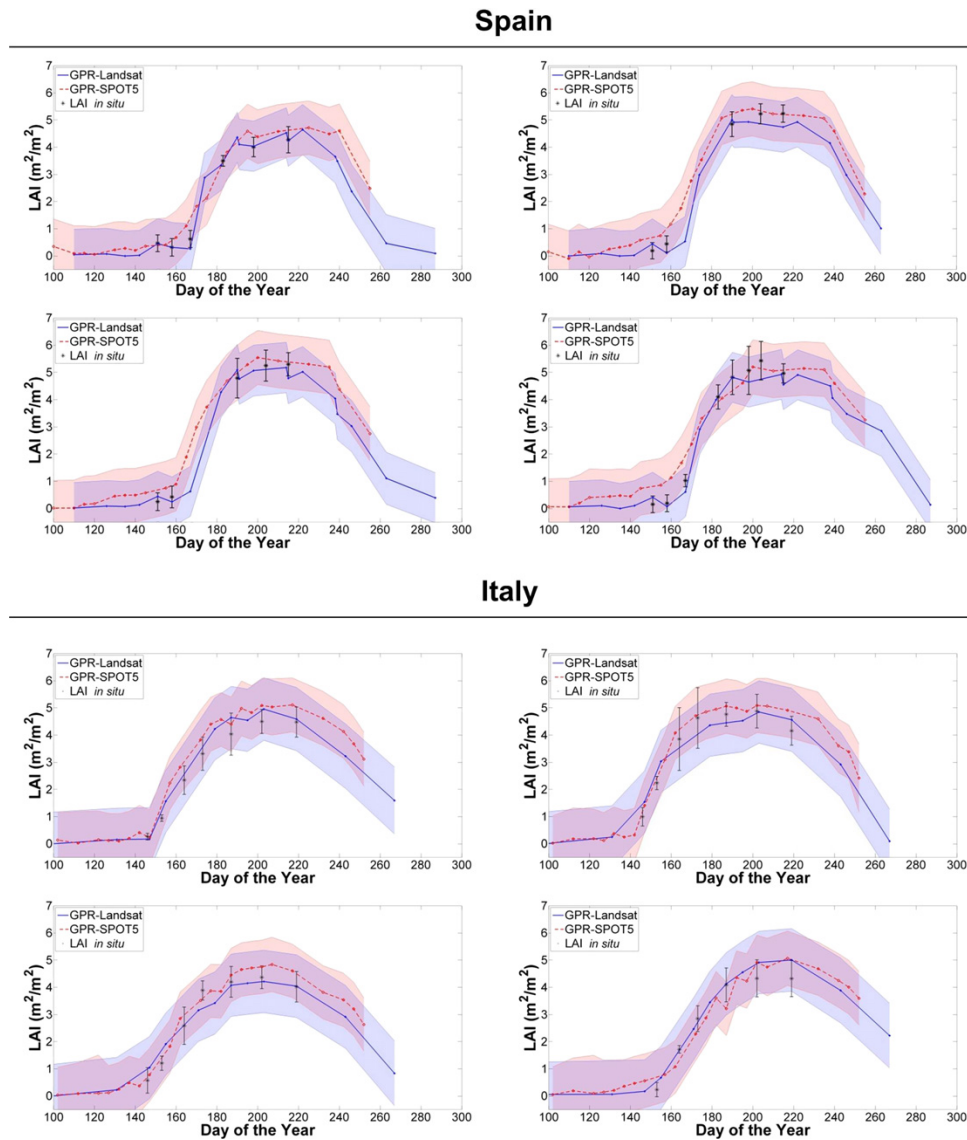


Fig. 12. Temporal dynamics of the Landsat and SPOT5 LAI estimates and the field LAI measurements within four representative ESUs in (top) the Spanish and (bottom) the Italian study areas. The prediction uncertainty provided by the GPRs is shown in shaded blue (GPR-Landsat) and orange (GPR-SPOT5) around the mean prediction; in violet the overlap. The standard deviations of the *in situ* measurements are displayed as error bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

supplementary material attached in Appendix A for further analysis on LAI temporal behavior according to sowing date and cultivated variety.

The analysis of profiles reveals that LAI estimates are in agreement with the qualitative behavior of rice growth in the different conditions. LAI trends do not show artifacts due to changes in spectral background conditions related to water management. When soil changes from dry and flood condition (or vice versa) in the early growing stages (e.g. for dry sowing field # 106 and # 118; for flood sowing field # 62 and # 102) anomalies in LAI retrievals can be appreciated. The small fluctuations within each temporal profile are mainly due to the residual of atmospheric contamination; for example the SPOT5 image acquired at DOY 142 shows a haze pattern from east to west that can justify the small peak in several plots both in flood (# 102) and dry conditions (# 120). The LAI values of

these peaks/drops are anyway within the expected uncertainty of the retrieval.

These results highlight the usefulness of the retrieved LAI time series at high resolution to perform field level crop monitoring and site specific assimilation in crop models.

4.5. Assessing the influence of rice background in LAI retrievals

In this study, we used a representative set of background spectra taking into account all possible expected conditions of the field. The influence of background characteristics in LAI retrievals was assessed introducing different types of background spectra in the training database. Firstly, we trained and inverted the PROSAIL model using typical spectra of dry bare soils which were present at the very beginning of the rice season. Secondly, we used a set of flooded soil

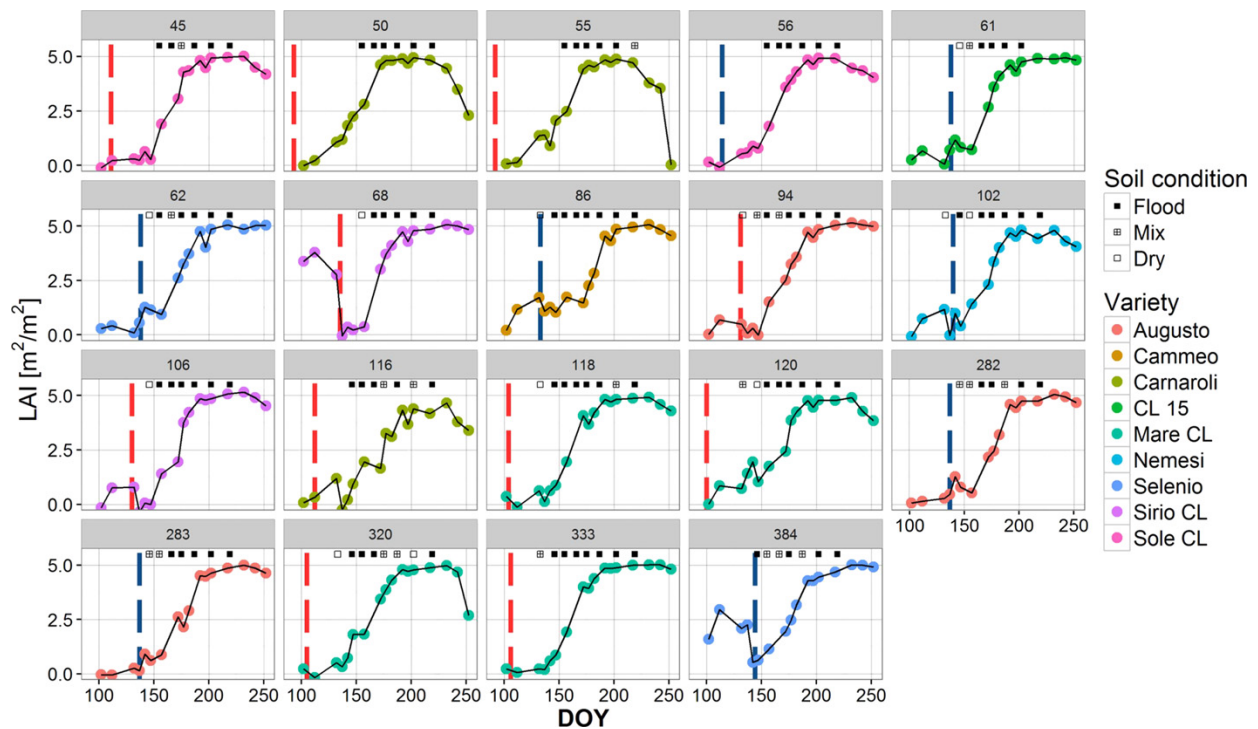


Fig. 13. Average behavior of estimated LAI over the monitored fields. Dotted vertical bars indicate the sowing date in dry soil (red) or in food condition (blue). Black squares indicates the soil condition recorded during the periodical field campaign. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

signatures representing typical background conditions during the rice growing (from sowing to harvest). Eventually, we used both dry and flooded spectra. Note that during these experiments the distributions of PROSAIL parameters for the crop were kept constant, and only the set of background spectra was changed (i.e. dry, flooded and both dry + flooded conditions). Table 3 exhibits the statistics when estimates from Landsat over Spain are compared with *in situ* LAI data. Similar results were obtained in the case of Italy and SPOT5 data (results not shown for brevity). The best results were obtained when a background spectral library composed by flooded and dry soils was used. Conversely, when considering only flooded or dry backgrounds high errors were found during the initial development stages of the plants (during tillering and before panicle formation), as the reflectance of the background constitutes a significant component of the overall spectral signature recorded by the sensor. Thus, a correct characterization of the soil conditions in the training database is mandatory in order to obtain time series of realistic LAI estimates and avoid meaningless values in the first rice growing stages.

Fig. 14 shows a typical example of the SPOT5 LAI evolution retrieved for a rice pixel in a representative Italian and Spanish ESUs. They reflect a common situation in which the soil was dry until DOY

132 (Italian area) and DOY 130 (Spanish area), and flooded immediately before sowing and during plant development. The figure compares the retrievals taking into account both dry and flooded conditions (red line), and the influence of using only dry soils (yellow line) or flooded soils (blue line). It can be observed that the use of a complete and representative spectral database (flooded + dry background) in the PROSAIL simulation provides realistic LAI evolutions (red line) which agree well with field measurements. It should be noted that in this situation LAI remains close to 0 before the rice emergence (from DOY 100 to 150 approx.), irrespectively of the flooded/non-flooded condition of the soil background. In a scenario where the training database does not characterize the flooded conditions (yellow line), the retrieved LAI is unrealistically high after the flooding during the first crop development stages, but as LAI increases canopy closure minimizes the importance of spectral background in the retrieval. The prediction uncertainty reflects this unreliability of the estimates, producing very high values (about 3) in the first crop stages (DOY 125–160). Similarly, when the training database does not characterize the dry conditions (blue line), the retrieved LAI is also unrealistically high before the real flooding of the fields.

In conclusion, the unrealistic effects produced by wrong spectra background can be very critical when time series analysis of rice crop dynamics is conducted. The increase of LAI at the beginning of the season would determine an unrealistic simulation of crop growth and identification of a wrong sowing date/emergence period respectively. This assessment demonstrates the importance of training correctly the RTM to produce a reliable data set for the retrieval process. In addition, retrieval of prediction uncertainty together with the LAI map can be useful to automatically identify non-agronomical areas and targets for which the model was not calibrated and/or performed well.

Table 3

Statistical scores between the Landsat estimated LAI taking into account different background conditions and ground measurements over Spain. In all field conditions, the number of samples used for the generation of the statistics was 171.

	ME	RMSE	MAE	R ²
Dry	−0.50	1.09	0.85	0.81
Flooded	0.10	0.62	0.40	0.90
Dry+flooded	0.08	0.39	0.28	0.95

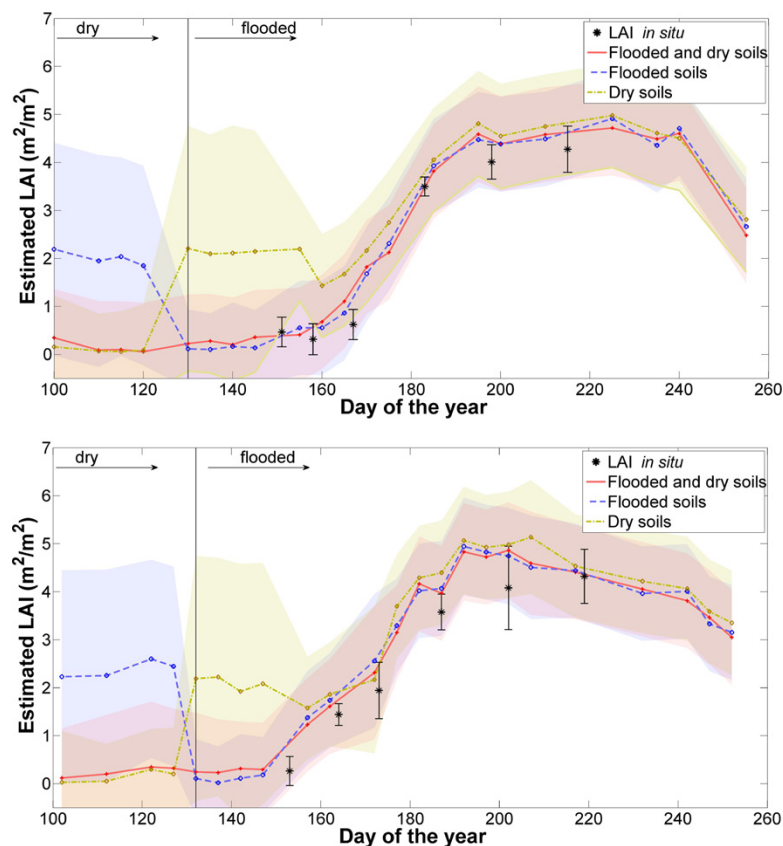


Fig. 14. Temporal evolution of GPR-SPOT5 LAI estimates within an ESU using different soil background in (top) Spain and (bottom) Italy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

In this study, a fully operational chain for deriving high-resolution LAI maps over Mediterranean rice crops is presented. The main novelties of this work include the following: 1) the use of sound machine learning algorithms trained on simulated RTM data specifically generated to characterize rice cropping features, 2) LAI estimates were originally validated with measurements acquired by a smartphone app (at affordable cost both in time, post-processing and human resources), 3) spatially and time-resolved estimations of LAI (not just static estimates) were produced, 4) different areas with particularly different rice crop specificities and spectral background conditions related to water management were analyzed, 5) seasonal remote sensing data from two multispectral sensors were exploited allowing to simulate the future potential of the operational chain with the Sentinel-2 constellation, 6) band ranking for rice monitoring where vegetation and water features make the problem more challenging was studied, and 7) the estimation of prediction uncertainties in a temporal manner were derived from the Gaussian process model and their validity and consistency were assessed. The proposed approach was illustrated in two rice study areas characterized by different rice varieties and agro-practices.

The production of the multitemporal local LAI maps was based on the PROSAIL RTM inversion with Gaussian process regression and Landsat/SPOT5 surface reflectance data. GPR proved to be a highly efficient and accurate method to invert the PROSAIL model, outperforming NN and KRR. However, the flexibility of the processing chain allows its application to any other regression method of choice. Theoretical performances of the regressions were satisfactory, and

a deeper analysis of the GPR-Landsat and GPR-SPOT5 training process showed that both models identified the green and near infrared channels as the most relevant bands in the retrieval process.

The multitemporal HR LAI maps captured the range and the temporal evolution of rice growth and resulted in agreement with corresponding *in situ* LAI measurements obtained from PocketLAI in two different countries. In addition, a comparison of the multitemporal estimates provided by both GPR-Landsat and GPR-SPOT5 retrievals showed good temporal consistency between them. However, a slight overestimation of the GPR-SPOT5 was observed during the period of the rice development.

In the development of the operational chain, the only user-demanding requirement refers to the provision of physical prior knowledge: site-specific characteristics need to be introduced in the model parameterization, such as rice plant and soil characteristics of the interested study area. Results highlight how a correct characterization of the underlying rice background is needed to obtain realistic LAI estimates during the initial development stages of the plants. The use of a training database comprising flooded and dry soil signatures showed to be robust against changes in background condition related to water management.

GPR retrievals also provided associated uncertainty for the predictions as a very valuable side information. This uncertainty is helpful a) for the scientist as a product to understand where potential errors in the retrieval exist, and b) for expert users, such as crop modelers, in order to weight LAI estimates in the assimilation process according to their error/goodness. The prediction uncertainty can also be used for diagnosing the presence of surfaces not addressed in the simulated database. Therefore, inspection of the uncertainty

provided with insights to (1) refine the selection of PROSAIL inputs, such as certain backgrounds not initially included in the preliminary retrieval, and (2) identify non-rice pixels thus identifying possible errors in the rice mask.

The near-real time production of HR LAI maps is useful in planning the management practices (i.e. fertilization) to minimize the yield pattern variability within each parcel. This information is particularly important for precision farming activities, since farmers are expecting to be supported in prescription map production for top-dress fertilization. A dense temporal data set of LAI maps is also fundamental to perform expert crop monitoring and/or improve crop model estimations exploiting assimilation techniques. It is important to mention that, whenever no external information able to represent the real variability in the field is provided for crop modeling, model simulation will provide the same results. This is the case when the aim is to apply crop models in an operational way at a parcel level. It is in fact not possible to obtain micro-meteorological information able to provide information changing from field to field. Moreover, it is not realistic to have detailed soil maps that usually exist at regional level with a scale ranging from 1:25.000 to 1:100.000. In this case, if the sowing date and variety are the same, or slightly different, the only way to capture the real spatio-temporal changes in crop development and production is to assimilate exogenous observation of crop status such as the information provided by EO LAI maps.

These results demonstrate the consistency and robustness of the presented processing chain for rice monitoring, which can be considered satisfactory for the production of HR LAI maps, and suitable to be assimilated by crop models. The proposal is aimed to improve crop monitoring and is specially suited for precision agriculture applications. Future work will include the use of the Sentinel-2 constellation whose spectral and temporal characteristics will make possible to increase the temporal resolution of the LAI estimates due to the combination with Landsat imagery, leading to further high-resolution multi-sensor studies in the framework of crop monitor and management specially in near real time. The presented processing chain has been operational during the 2016 ERMES rice season.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement 606983, the European Research Council (ERC) funding under the ERC-CoG-2014 SEDAL under grant agreement 647423, and the Land-SAF (the EUMETSAT Network of Satellite Application Facilities) project. The authors would like to deeply thank the Spanish and Italian rice farmers for allowing collecting *in situ* data within their fields and Cassandra Lab. for providing the PocketLAI application.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.rse.2016.10.009>.

References

- Arvidson, T., Goward, S., Gasch, J., Williams, D., 2006. Landsat-7 long-term acquisition plan. *Photogramm. Eng. Remote. Sens.* 72, 1137–1146. <http://dx.doi.org/10.14358/PERS.72.10.1137>.
- Atzberger, C., Richter, K., 2012. Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sens. Environ.* 120, 208–218. <http://dx.doi.org/10.1016/j.rse.2011.10.035>.
- Baillarin, S., Gigord, P., Hagolle, O., 2008. Automatic registration of optical images, a stake for future missions: application to ortho-rectification, time series and mosaic products. *Geoscience and Remote Sensing Symposium*, 2008. IGARSS 2008. IEEE International. vol. 2. II-1112–II-1115. <http://dx.doi.org/10.1109/IGARSS.2008.4779194>.
- Baret, F., Clevers, J., Steven, M., 1995. The robustness of canopy gap fraction estimates from red and near-infrared reflectances: a comparison of approaches. *Remote Sens. Environ.* 54, 141–151. [http://dx.doi.org/10.1016/0034-4257\(95\)00136-0](http://dx.doi.org/10.1016/0034-4257(95)00136-0).
- Baret, F., Fourty, T., 1997. Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie* 17, 455–464. <http://dx.doi.org/10.1051/agro:19970903>.
- Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, FAPAR and FCOVER CYCLOPES global products derived from VEGETATION: part 1: principles of the algorithm. *Remote Sens. Environ.* 110, 275–286. <http://dx.doi.org/10.1016/j.rse.2007.02.018>.
- Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholczyk, P., Smets, B., 2013. Geov1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: principles of development and production. *Remote Sens. Environ.* 137, 299–309. <http://dx.doi.org/10.1016/j.rse.2012.12.027>.
- Borel, C.C., 2010. *Vegetative canopy parameter retrieval using 8-band data*. Digital-Globe 8-Band Research Challenge 2010.
- Boschetti, M., Nutini, F., Manfron, G., Brivio, P.A., Nelson, A., 2014. Comparative analysis of normalised difference spectral indices derived from modis for detecting surface water in flooded rice cropping systems. *PLoS one* 9, e88741. <http://dx.doi.org/10.1371/journal.pone.0088741>.
- Boschetti, M., Stroppiana, D., Brivio, P., Bocchi, S., 2009. Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. *Int. J. Remote Sens.* 30, 4643–4662. <http://dx.doi.org/10.1080/01431160802632249>.
- Breda, N.J.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *J. Exp. Bot.* 54, 2403–2417. <http://dx.doi.org/10.1093/jxb/erg263>.
- Bsaibes, A., Courault, D., Baret, F., Weiss, M., Olioso, A., Jacob, F., Hagolle, O., Marloie, O., Bertrand, N., Desfond, V., Kzemipour, F., 2009. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. *Remote Sens. Environ.* 113, 716–729. <http://dx.doi.org/10.1016/j.rse.2008.11.014>.
- Campos-Taberner, M., García-Haro, F.J., Moreno, A., Gilabert, M.A., Martínez, B., Sánchez-Ruiz, S., Camps-Valls, G., 2015a. Development of an earth observation processing chain for crop bio-physical parameters at local scale. *Geoscience and Remote Sensing Symposium (IGARSS)*, 2015 IEEE International. pp. 29–32. <http://dx.doi.org/10.1109/IGARSS.2015.7325689>.
- Campos-Taberner, M., García-Haro, F., Moreno, A., Gilabert, M., Sánchez-Ruiz, S., Martínez, B., Camps-Valls, G., 2015b. Mapping leaf area index with a smartphone and Gaussian processes. *Geosci. Remote Sens. Lett.* IEEE 12, 2501–2505. <http://dx.doi.org/10.1109/LGRS.2015.2488682>.
- Campos-Taberner, M., García-Haro, F.J., Confalonieri, R., Martínez, B., Moreno, A., Sánchez-Ruiz, S., Gilabert, M.A., Camacho, F., Boschetti, M., Busetto, L., 2016. Multitemporal monitoring of plant area index in the valencia rice district with pocketlai. *Remote Sens.* 8, 202. <http://dx.doi.org/10.3390/rs8030202>.
- Camps-Valls, G., Muñoz-Mari, J., Verrelst, J., Mateo, F., Gomez-Dans, J., 2016. A survey on Gaussian processes for Earth observation data analysis. *IEEE Geosci. Remote Sens. Mag.* 4, 58–78. <http://dx.doi.org/10.1109/MGRS.2015.2510084>.
- Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., Malo, J., 2011. *Remote Sensing Image Processing*, vol. 5. Morgan & Claypool Publishers. <http://dx.doi.org/10.2200/S00392ED1V01Y201107IVM012>.
- Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62, 241–252. [http://dx.doi.org/10.1016/S0034-4257\(97\)00104-1](http://dx.doi.org/10.1016/S0034-4257(97)00104-1).
- Chen, J., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S., White, H., Hall, R., Peddle, D., King, D., Trofymow, J., Swift, E., der Sanden, J.V., Pellikka, P., 2002. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sens. Environ.* 80, 165–184. [http://dx.doi.org/10.1016/S0034-4257\(01\)00300-5](http://dx.doi.org/10.1016/S0034-4257(01)00300-5).
- Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in landsat ETM+ SLC-off images. *Remote Sens. Environ.* 115, 1053–1064. <http://dx.doi.org/10.1016/j.rse.2010.12.010>.
- Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. *Plant, Cell Environ.* 15, 421–429. <http://dx.doi.org/10.1111/j.1365-3040.1992.tb00992.x>.
- Claverie, M., Vermote, E.F., Weiss, M., Baret, F., Hagolle, O., Demarez, V., 2013. Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. *Remote Sens. Environ.* 139, 216–230. <http://dx.doi.org/10.1016/j.rse.2013.07.027>.
- Clevers, J., 1988. The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sens. Environ.* 25, 53–69. [http://dx.doi.org/10.1016/0034-4257\(88\)90041-7](http://dx.doi.org/10.1016/0034-4257(88)90041-7).
- Colombo, R., Bellingeri, D., Fasolini, D., Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sens. Environ.* 86, 120–131. [http://dx.doi.org/10.1016/S0034-4257\(03\)00094-4](http://dx.doi.org/10.1016/S0034-4257(03)00094-4).
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* 84, 1–15. [http://dx.doi.org/10.1016/S0034-4257\(02\)00035-4](http://dx.doi.org/10.1016/S0034-4257(02)00035-4).
- Confalonieri, R., Foi, M., Casa, R., Aquaro, S., Tona, E., Peterle, M., Boldini, A., Carli, G.D., Ferrari, A., Finotto, G., Guarneri, T., Manzoni, V., Movedi, E., Nisoli, A., Paleari, L., Radici, I., Suardi, M., Veronesi, D., Bregaglio, S., Cappelli, G., Chiodini, M., Dominoni, P., Francione, C., Frasso, N., Stella, T., Acutis, M., 2013. Development of an app for estimating leaf area index using a smartphone. *Trueness and precision*

- determination and comparison with other indirect methods. *Comput. Electron. Agric.* 96, 67–74. <http://dx.doi.org/10.1016/j.compag.2013.04.019>.
- Confalonieri, R., Rosenmund, A.S., Baruth, B., 2009. An improved model to simulate rice yield. *Agron. Sustain. Dev.* 29, 463–474. <http://dx.doi.org/10.1051/agro/2009005>.
- Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., Schaepman, M., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* 9, 165–193. <http://dx.doi.org/10.1016/j.jag.2006.05.003>.
- Drusch, M., Bello, U.D., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36. <http://dx.doi.org/10.1016/j.rse.2011.11.026>.
- Duan, S.-B., Li, Z.-L., Wu, H., Tang, B.-H., Ma, L., Zhao, E., Li, C., 2014. Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* 26, 12–20. <http://dx.doi.org/10.1016/j.jag.2013.05.007>.
- Duveiller, G., Weiss, M., Baret, F., Defourny, P., 2011. Retrieving wheat green area index during the growing season from optical time series measurements based on neural network radiative transfer inversion. *Remote Sens. Environ.* 115, 887–896. <http://dx.doi.org/10.1016/j.rse.2010.11.016>.
- European Commission, 2011. Natura 2000. (Accessed on May 2016) http://ec.europa.eu/environment/nature/natura2000/index_en.htm.
- Fang, H., Liang, S., 2005. A hybrid inversion method for mapping leaf area index from MODIS data: experiments and application to broadleaf and needleleaf canopies. *Remote Sens. Environ.* 94, 405–424. <http://dx.doi.org/10.1016/j.rse.2004.11.001>.
- Francone, C., Pagani, V., Foi, M., Cappelli, G., Confalonieri, R., 2014. Comparison of leaf area index estimates by ceptometer and pocketlai smart app in canopies with different structures. *Field Crop Res.* 155, 38–41. <http://dx.doi.org/10.1016/j.fcr.2013.09.024>.
- Gower, S.T., Kucharik, C.J., Norman, J.M., 1999. Direct and indirect estimation of leaf area index, FAPAR, and net primary production of terrestrial ecosystems. *Remote Sens. Environ.* 70, 29–51. [http://dx.doi.org/10.1016/S0034-4257\(99\)00056-5](http://dx.doi.org/10.1016/S0034-4257(99)00056-5).
- Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., 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 Sens. Environ.* 90, 337–352. <http://dx.doi.org/10.1016/j.rse.2003.12.013>.
- Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B., Meygret, A., 2008. Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: application to Formosat-2 images. *Remote Sens. Environ.* 112, 1689–1701. <http://dx.doi.org/10.1016/j.rse.2007.08.016>.
- Haykin, S., 1999. *Neural Networks—A Comprehensive Foundation*. 2nd ed., Prentice Hall.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. [http://dx.doi.org/10.1016/S0034-4257\(02\)00096-2](http://dx.doi.org/10.1016/S0034-4257(02)00096-2).
- Jacquemoud, S., Bacour, C., Poilvé, H., Frangi, J.-P., 2000. Comparison of four radiative transfer models to simulate plant canopies reflectance: direct and inverse mode. *Remote Sens. Environ.* 74, 471–481. [http://dx.doi.org/10.1016/S0034-4257\(00\)00139-5](http://dx.doi.org/10.1016/S0034-4257(00)00139-5).
- Jacquemoud, S., Baret, F., 1990. Prospect: A model of leaf optical properties spectra. *Remote Sens. Environ.* 34, 75–91. [http://dx.doi.org/10.1016/0034-4257\(90\)90100-Z](http://dx.doi.org/10.1016/0034-4257(90)90100-Z).
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT + SAIL models: a review of use for vegetation characterization. *Remote Sens. Environ.* 113, Supplement 1, S56–S66. <http://dx.doi.org/10.1016/j.rse.2008.01.026>.
- Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination: part I. Theories, sensors and hemispherical photography. *Agric. For. Meteorol.* 121, 19–35. <http://dx.doi.org/10.1016/j.agrformet.2003.08.027>.
- Ju, J., Roy, D.P., 2008. The availability of cloud-free landsat ETM+ data over the conterminous united states and globally. *Remote Sens. Environ.* 112, 1196–1211. <http://dx.doi.org/10.1016/j.rse.2007.08.011>.
- Lazaro-Gredilla, M., Titsias, M., Verrelst, J., Camps-Valls, G., 2014. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. *Geosci. Remote Sens. Lett.* IEEE 11, 838–842. <http://dx.doi.org/10.1109/LGRS.2013.2279695>.
- Maclean, J.L., 2002. *Rice almanac: source book for the most important economic activity on earth*. Int. Rice Res. Inst.
- Malenovsky, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* 120, 91–101. <http://dx.doi.org/10.1016/j.rse.2011.09.026>.
- Mckay, M.D., Beckman, R.J., Conover, W.J., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42, 55–61. <http://dx.doi.org/10.1080/00401706.2000.10485979>.
- Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sens. Environ.* 92, 195–206. <http://dx.doi.org/10.1016/j.rse.2004.06.005>.
- Montavon, G., Braun, M.L., Krueger, T., Muller, K.R., 2013. Analyzing local structure in kernel-based learning: explanation, complexity, and reliability assessment. *IEEE Signal Process. Mag.* 30, 62–74. <http://dx.doi.org/10.1109/MSP.2013.2249294>.
- Morisette, J., Baret, F., Privette, J., Myneni, R., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sanchez-Azofeifa, G., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iliades, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E., Petersen, G., Seufert, G., Cook, R., 2006. Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Trans. Geosci. Remote Sens.* 44, 1804–1817. <http://dx.doi.org/10.1109/TGRS.2006.872529>.
- Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G., Lotsch, A., Friedl, M., Morisette, J., Votava, P., Nemani, R., Running, S., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* 83, 214–231. [http://dx.doi.org/10.1016/S0034-4257\(02\)00074-3](http://dx.doi.org/10.1016/S0034-4257(02)00074-3).
- Myneni, R., Ramakrishna, R., Nemani, R., Running, S., 1997. Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35, 1380–1393. <http://dx.doi.org/10.1109/36.649788>.
- Quaife, T., Lewis, P., Kauwe, M.D., Williams, M., Law, B.E., Disney, M., Bowyer, P., 2008. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sens. Environ.* 112, 1347–1364. <http://dx.doi.org/10.1016/j.rse.2007.05.020>.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. The MIT Press, New York.
- Roy, D., Kovalsky, V., Zhang, H., Vermote, E., Yan, L., Kumar, S., Egorov, A., 2016. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* 185, 57–70. <http://dx.doi.org/10.1016/j.rse.2015.12.024>.
- 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., Kovalsky, V., Lee, Z., Lymburner, L., Masek, J., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. <http://dx.doi.org/10.1016/j.rse.2014.02.001>.
- Shawe-Taylor, J., Cristianini, N., 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.
- Smith, J., 1993. LAI inversion using backpropagation neural network trained with multiple scattering model. *IEEE Trans. Geosci. Remote Sens.* 31, 1102–1106. <http://dx.doi.org/10.1109/36.263783>.
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.* 16, 125–141. [http://dx.doi.org/10.1016/0034-4257\(84\)90057-9](http://dx.doi.org/10.1016/0034-4257(84)90057-9).
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. <http://dx.doi.org/10.1016/j.rse.2016.04.008>.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G., Moreno, J., 2015a. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties — a review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290. <http://dx.doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., 2012a. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Trans. Geosci. Remote Sens.* 50, 1832–1843. <http://dx.doi.org/10.1109/TGRS.2011.2168962>.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012b. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* 118, 127–139. <http://dx.doi.org/10.1016/j.rse.2011.11.002>.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G., Moreno, J., 2015b. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods — a comparison. *ISPRS J. Photogramm. Remote Sens.* 108, 260–272. <http://dx.doi.org/10.1016/j.isprsjprs.2015.04.013>.
- Weiss, M., Baret, F., Smith, G., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination: part II. Estimation of LAI, errors and sampling. *Agric. For. Meteorol.* 121, 37–53. <http://dx.doi.org/10.1016/j.agrformet.2003.08.001>.
- Woodcock, C., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thakabail, P., Vermote, E., Vogelmann, J., Wulder, M., Wynne, R., 2008. Free access to landsat imagery. *Science* 320, 1011. <http://dx.doi.org/10.1126/science.320.5879.1011a>.
- Zhang, X., Lu, C., Hu, N., Yao, K., Zhang, Q., Dai, Q., 2013. Simulation of canopy leaf inclination angle in rice. *Rice Science* 20, 434–441. [http://dx.doi.org/10.1016/S1672-6308\(13\)60161-4](http://dx.doi.org/10.1016/S1672-6308(13)60161-4).

Article

Exploitation of SAR and Optical Sentinel Data to Detect Rice Crop and Estimate Seasonal Dynamics of Leaf Area Index

Manuel Campos-Taberner ^{1,*}, Francisco Javier García-Haro ¹, Gustau Camps-Valls ²,
 Gonçal Grau-Muedra ¹, Francesco Nutini ³, Lorenzo Busetto ³, Dimitrios Katsantonis ⁴,
 Dimitris Stavrakoudis ⁵, Chara Minakou ⁵, Luca Gatti ⁶, Massimo Barbieri ⁶, Francesco Holecz ⁶,
 Daniela Stroppiana ³ and Mirco Boschetti ³

¹ Department of Earth Physics and Thermodynamics, Faculty of Physics, Universitat de València, Dr. Moliner 50, 46100 Burjassot, València, Spain; j.garcia.haro@uv.es (F.J.G.-H.); gonal.grau@uv.es (G.G.-M.)

² Image Processing Laboratory (IPL), Universitat de València, Catedrático A. Escardino, 46980 Paterna, València, Spain; gustau.camps@uv.es

³ Institute for Electromagnetic Sensing of the Environment, Italian National Research Council, Via Bassini 15, 20133 Milan, Italy; lgatti@sarmap.ch (F.N.); busetto.l@irea.cnr.it (L.B.); stroppiana.d@irea.cnr.it (D.S.); boschetti.m@irea.cnr.it (M.B.)

⁴ Hellenic Agricultural Organization—Demeter, Institute of Plant Breeding and Genetic Resources, 57001 Thessaloniki, Thessaloniki, Greece; dikatsa@cerealinstitute.gr

⁵ Laboratory of Forest Management and Remote Sensing, School of Forestry and Natural Environment, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; jstavrak@auth.gr (D.S.); minakou@for.auth.gr (C.M.)

⁶ Sarmap, Cascine di Barico 10, 6989 Purasca, Switzerland; lgatti@sarmap.ch (L.G.); mbarbieri@sarmap.ch (M.B.); fholecz@sarmap.ch (F.H.)

* Correspondence: manuel.campos@uv.es; Tel.: +34-963-543-256; Fax: +34-963-543-385

Academic Editors: Agnes Begue, Nicolas Baghdadi and Prasad S. Thenkabail

Received: 12 January 2017; Accepted: 4 March 2017; Published: 7 March 2017

Abstract: This paper presents and evaluates multitemporal LAI estimates derived from Sentinel-2A data on rice cultivated area identified using time series of Sentinel-1A images over the main European rice districts for the 2016 crop season. This study combines the information conveyed by Sentinel-1A and Sentinel-2A into a high-resolution LAI retrieval chain. Rice crop was detected using an operational multi-temporal rule-based algorithm, and LAI estimates were obtained by inverting the PROSAIL radiative transfer model with Gaussian process regression. Direct validation was performed with in situ LAI measurements acquired in coordinated field campaigns in three countries (Italy, Spain and Greece). Results showed high consistency between estimates and ground measurements, revealing high correlations ($R^2 > 0.93$) and good accuracies ($RMSE < 0.83$, $rRMSE_m < 23.6\%$ and $rRMSE_r < 16.6\%$) in all cases. Sentinel-2A estimates were compared with Landsat-8 showing high spatial consistency between estimates over the three areas. The possibility to exploit seasonally-updated crop mask exploiting Sentinel-1A data and the temporal consistency between Sentinel-2A and Landsat-7/8 LAI time series demonstrates the feasibility of deriving operationally high spatial-temporal decametric multi-sensor LAI time series useful for crop monitoring.

Keywords: rice map; leaf area index (LAI); Sentinel-1A; Sentinel-2A; Gaussian process regression

1. Introduction

Monitoring Earth's vegetated surfaces is a challenging issue in which the green leaf area index (LAI) has been recognized as an essential climate variable (ECV) by the Global Terrestrial Observation

System (GTOS) and the Global Climate Observing System (GCOS) [1]. Defined as half of the total green leaf area per unit ground surface area [2], LAI plays a key role in land surface models [3] and can be used to address various agricultural issues, such as rice crop monitoring, yield forecasting and crop management [4].

Crop monitoring is necessary to identify the onset of stress conditions, which require actions for reducing their impact on crop yield. Anomaly drops in canopy LAI are indicators of plant stress conditions, which could lead to a decrease of plant production and to an increased senescence rate [5].

Rice yield forecasting is a crucial task for management and planning, and it can be addressed with both statistical and mathematical modeling approaches. Statistical approaches directly link crop biophysical variables, such as LAI, to crop yield. LAI is indeed recognized as the major morphological parameter for determining crop growth, and it is strongly correlated with crop productivity [6]. Crop models are able to simulate rice growing and are used to provide indications on crop status and to predict yields over large areas [7,8]. However, crop models require information on soil, meteorological variable, crop parameters and management practices, which are not always available or practical to be obtained in a spatially distributed way and during the season [9]. Hence, the best way to simulate the real spatial-temporal differences in crop development of fields sowed the same day with the same variety is to assimilate exogenous observation, such as LAI maps, in the modeling solution [10]. Accurate estimation of LAI has been shown to improve the accuracy of grain yield estimates [11], and an operational application of this workflow for rice was successfully demonstrated in Asia in the framework of the RIICE (Remote sensing-based Information and Insurance for Crops in Emerging economies) project (<http://www.riice.org/>) where rice yield is estimated from the Oryza2000 model by assimilating LAI maps derived from synthetic aperture radar (SAR) images [12].

As far as precise crop management is concerned, LAI data have been found to be useful in new approaches for the determination of nitrogen concentration dilution curves, which are traditionally based on plant dry matter (PDM) estimation [13,14]. Remotely-sensed data at decametric resolution (e.g., Sentinel-2A) are the sole source of information available to provide high-resolution (HR) LAI estimation on wide areas to exploit the nitrogen concentration dilution curve approach for optimal crop fertilization.

In the last few years, the scientific community has made big efforts with the goal of providing reliable and accurate LAI estimates at local scales taking advantage of unmanned aerial vehicles and high-resolution sensors, such as Landsat [15] and SPOT5 [16–18]. The recently launched Sentinel-2A satellite [19] provides well-suited spectral and temporal data for LAI retrieval at high-resolution in near real time, useful for assessing crop status and providing support in agro-practices at the parcel level. Many methods have been proposed and implemented in retrieval chains from Earth observation (EO) data to derive LAI estimates [20]. Empirical parametric algorithms have been developed based on calibrated relationships between vegetation indices and canopy biophysical variables [21]. On the other hand, non-parametric algorithms estimate biophysical variables using a training database containing pairs of the biophysical parameter and the associated spectral data [22]. In statistical approaches, concomitant in situ measurements of the biophysical parameter of interest and the associated spectral data from remote sensing platforms are used as a training database, whereas the hybrid approaches rely on a database generated by radiative transfer models (RTMs). RTMs describe the interactions between the incoming radiation, canopy elements and the background soil surface. The PROSAIL RTM has become one of the most popular and widely-used RTMs due to its consistency and robustness [23]. Hybrid approaches retrieve LAI by inverting RTM through machine learning techniques with a large number of methods [24,25]. Among them, neural networks (NN) [26] and kernel-based methods [27,28] are the most popular and used regression methods in remote sensing. State-of-the-art kernel-based methods, such as Gaussian process regression (GPR) [29], provided encouraging results in the framework of biophysical parameter estimation outperforming other kernel methods and NN [18,30].

This study shows Sentinel-2A LAI estimates produced in the framework of the ERMES (an Earth observation Model based RicE information Service) project (<http://www.ermes-fp7space.eu/>), which aims to develop a prototype of Copernicus down-stream services assimilating EO and in situ data on crop modeling dedicated to the rice sector. In this framework, the main objectives of this paper are: (1) to derive multitemporal HR LAI maps based on Sentinel-2A real data acquired during an entire European rice season for supporting agro-practices; (2) to validate the Sentinel-2A LAI estimates by direct comparison with ground LAI measurements conducted in three countries (Italy, Spain and Greece) during the season; and (3) to assess the feasibility of jointly using Sentinel-2A and Landsat-7/8 to assess a decametric high resolution multitemporal time series with an unprecedented frequency. It is worth mentioning that in this study, multitemporal synthetic aperture radar (SAR) data (Sentinel-1A) were used to obtain seasonally-updated rice maps used for focusing optical-based LAI retrievals on rice areas. Sentinel-2A and Landsat-7/8 data were processed in near real time in the three countries, and the corresponding LAI maps were provided through the web-based geo-portal in the project framework to the users and modelers.

The next sections describe the study areas and the remote sensing data used in this study (Section 2), also describing Sentinel-1A rice mapping and Sentinel-2A LAI retrieval methodology (Section 3). Section 4 validates the LAI estimates through direct comparison with in situ measurements and discusses the spatial-temporal consistency between Sentinel-2A and Landsat LAI estimates as a proof of concept for the operational generation of multisensor decametric LAI time series. Finally, Section 5 highlights the main conclusions.

2. Materials

2.1. Study Areas

The study areas have been selected in three Mediterranean countries, which are responsible for 85% of total European rice production: Italy (51.9%), Spain (25.4%) and Greece (7.0%) (see Figure 1). The Spanish study area is located in the rice district of Valencia (east of Spain) belonging to the Albufera Natural Park, which is a special protection area, thus allowing only rice crop practices. The Italian study area belongs to the Lomellina rice district (south-western Lombardy region), where rice is the dominant crop (>90%). The Greek study area is located in the rice district of Thessaloniki, which is the main cultivation area for Greece. Within each study area, rice is a common crop with a long tradition and economic value.

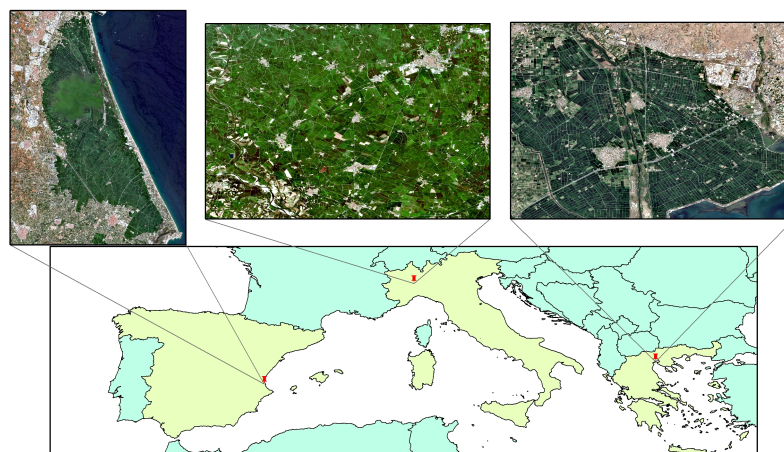


Figure 1. Study areas. Sentinel-2A surface reflectance RGB composite of the Spanish (30 July 2016), Italian (21 July 2016) and Greek (23 July 2016) study areas.

2.2. Field Campaigns

In the framework of the local ERMES field activities, LAI ground measurements were conducted over the study areas. In Spain, Italy and Greece, 32, 16 and 10 ESUs (elementary sampling units) were selected. The temporal frequency of the campaigns was approximately 7–10 days starting from the very beginning of rice emergence (early June) up to the maximum green rice LAI development (mid-August). The field campaigns were planned in such a way as LAI measurements within every ESU were either temporally coincident or 1–2 days apart with respect to the Sentinel-2A images. The same sampling scheme was used over each ESU, following the guidelines and recommendations of the Validation of Land European Remote sensing Instruments (VALERI) protocol (<http://w3.avignon.inra.fr/valeri/>). In order to characterize the spatial variability within each ESU, a range of 18–24 measurements was taken allowing one to obtain a statistically significant mean LAI estimate per ESU. The center of the ESU was geo-located using a GPS for later matching and associating the mean LAI estimate with the corresponding satellite estimation.

LAI measurements were acquired using a dedicated smartphone app (PocketLAI), which uses both smartphone's accelerometer and camera to acquire images at 57.5° below the canopy and compute LAI estimates through an internal segmentation algorithm [31]. PocketLAI estimates can reproduce destructive LAI measurements with acceptable results in terms of both reliability and accuracy [31]. PocketLAI allows the acquisition of in situ LAI measurements at an affordable cost both in computational and human resources and aligned well with estimates obtained using plant canopy analyzers and DHP (digital hemispherical photography) techniques [32].

2.3. Remote Sensing High-Resolution Optical and SAR Data

2.3.1. Sentinel-2A

The Sentinel-2A mission of the Copernicus program provides enhanced continuity of data so far provided by SPOT-4/5 and Landsat-7/8 missions. In this framework, the European Space Agency (ESA) provides free and open access to the Sentinels' data from the Sentinels Scientific Data Hub (<https://scihub.copernicus.eu/>). In this study, Sentinel-2A Level 1C data (top-of-atmosphere reflectances in cartographic geometry) were downloaded from early May up to the end of September in 2016 for the three study areas providing information every 10 days in 13 bands in the visible, near infra-red and short wave infra-red spectrum at a 10-, 20- and 60-m spatial resolution depending on the spectral band. Table S1 in the Supplementary Materials section lists the Sentinel-2A images used over the three areas.

2.3.2. Landsat-7/8

The Landsat Data Continuity Mission (LDCM) provides free access to the Landsat-7/8 archive. In this study, Landsat-8 Operational Land Imager (OLI) and Landsat-7 Enhanced Thematic Mapper (ETM+) surface reflectance data at 30-m spatial resolution were downloaded through the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) during the 2016 rice season over the three study areas. The provisional Landsat-8 Surface Reflectance (LaSRC) [33] and the Landsat-7 ETM+ LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing System) products were used as inputs to derive Landsat-7/8 LAI estimates. Images were available every 16 days in Italy and Greece. On the other hand, since the Spanish rice area lies in two Landsat paths within the same row, the temporal resolution of the images increased up to seven and nine days. See Table S1 in the Supplementary Materials section for further details.

2.3.3. Sentinel-1A

The Sentinel-1A SAR sensor operates at 5.405 GHz (C-band, corresponding to a microwave wavelength of about 5.6 cm) with VH and VV polarizations offering 12 days of revisit time. Sentinel-1A data over the three study areas were downloaded from the ESA Sentinels Scientific Data Hub during

the 2016 European rice season. A fully-automated processing chain was followed in order to convert the multi-temporal SAR single look complex (SLC) data into terrain geocoded σ^0 values. The processing chain is a module within MAPscape-RICE, which included the following steps: strip mosaicking when SAR data are zero-Doppler focused, image co-registration, time-series speckle filtering, terrain geocoding, radiometric calibration and normalization, anisotropic non-linear diffusion (ANLD) filtering and the removal of atmospheric attenuation. The ANLD filter performs a smoothing and a high level of regularization in homogenous areas, while preserving inhomogeneous signal variations (discontinuities), which are typical of features, such as field edges, roads, rivers and other image information that we want to preserve [34]. On the other hand, although microwave signals are able to penetrate clouds, it is possible to observe sudden changes in the dielectric constant at shorter wavelengths (X- and C-band) due to atmospheric moisture changes, dew, rain and severe storms. These events affect the temporal signature of σ^0 in two ways: (i) the thick layer of water vapor generates a strong decrease in σ^0 during the event, followed by a strong increase after the event; (ii) the intense rainfall generates a strong increase in σ^0 during the event, followed by a strong decrease after the event. These effects were removed by identifying anomalous peaks or troughs and correcting the σ^0 values by means of an interpolator.

3. Retrieval Methodology

The PROSAIL radiative transfer model was ran in forward mode for building a database, which was used for training the retrieval model and for mimicking canopy reflectance using the turbid medium assumption, which is particularly well suited for homogeneous canopies like rice [18,35]. It assumes the canopy as a turbid medium for which leaves are randomly distributed. PROSAIL simulates leaf reflectance for the optical spectrum from 400–2500 nm with a 1-nm spectral resolution and as a function of biochemistry and structure of the canopy, its leaves, the background soil reflectance and the sun-view geometry. The mesophyll structural parameter (N), leaf chlorophyll (C_{ab}), dry matter (C_m) and water (C_w) contents are the parameters that determine the leaf optical properties. C_w was tied to the dry matter content ($C_w = C_m \times C_{wREL} / (1 - C_{wREL})$) assuming that green leaves have a relative water content (C_{wREL}) varying within a relatively small range [36]. The canopy structure is characterized by the average leaf angle inclination (ALA), the LAI and the hot-spot parameter (hotspot). A multiplicative brightness parameter (β_s) was introduced and applied to spectral flooded and dry soil signatures to represent different background reflectance types [36,37]. The system geometry was described by the solar zenith angle (θ_s), view zenith angle (θ_v) and the relative azimuth angle between both angles ($\Delta\Theta$).

Sub-pixel rice mixed conditions [18] were represented by a linear mixture of vegetation (vCover) and bare/flooded soil (1-vCover) spectra. In this study, vCover was assumed to follow a truncated Gaussian distribution (see Table 1). The leaf and canopy variables, as well as the soil brightness parameter were randomly generated following the parametrization in [18] in order to constrain the behavior of the model to Mediterranean rice areas (see Table 1). In particular, a spectral library of underlying rice background (flooded and dry) signatures was used to obtain multitemporal LAI retrievals robust against changes in background condition related to water management. The PROSAIL was run in forward mode to build a database of 2000 pairs of multispectral data and LAI.

The inversion of PROSAIL was done using standard regression. In particular, we used a Gaussian process regression (GPR) model [29]. GPR has been recently introduced as a powerful regression tool in remote sensing studies [30]. The GPR model exploits the relations between the input (e.g., bands spectra) $\mathbf{x} = [x^1, \dots, x^B] \in \mathbb{R}^B$ and the output variable (i.e., LAI) $y \in \mathbb{R}$ in a training phase, and the predictive mean function for a new test point \mathbf{x}_* is expressed in the form:

$$\hat{y}_* = f(\mathbf{x}_*) = \sum_{i=1}^N \alpha_i k_{\theta}(\mathbf{x}_i, \mathbf{x}_*) + \alpha_0, \quad (1)$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the N spectra used in the training phase, α_i is the weight assigned to each one of them, α_0 is the bias in the regression function and k_θ is a kernel or covariance function (parametrized by a set of hyperparameters θ) that evaluates the similarity between the test spectrum and all N training spectra.

Table 1. Distribution of the parameters within the PROSAIL RTM at canopy (leaf area index (LAI), leaf angle inclination (ALA), hot-spot parameter (hotspot) and vegetation cover (vCover)) and leaf (mesophyll structural parameter (N), leaf chlorophyll (C_{ab}), dry matter (C_m) and relative water (C_w) contents) levels and the soil brightness parameter (β_s).

	Parameter	Min	Max	Mode	Std. Deviation	Type
Canopy	LAI (m ² /m ²)	0	10	3.5	4.5	Gaussian
	ALA (°)	30	80	60	20	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.5	1	1	0.2	Truncated Gaussian
Leaf	N	1.2	2.2	1.5	0.3	Gaussian
	C_{ab} (μg·cm ^{−2})	20	90	45	30	Gaussian
	C_{dm} (g·cm ^{−2})	0.003	0.011	0.005	0.005	Gaussian
	C_{wREL}	0.6	0.8	-	-	Uniform
Soil	β_s	0.3	1.2	0.9	0.25	Gaussian

In order to generate a kernel regression model, one needs to specify a covariance/kernel function k_θ , to infer its hyperparameters θ and model weights $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_N]$. For the GPR prediction model, we used the so-called automatic relevance determination (ARD) kernel function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp \left(- \sum_{b=1}^B \frac{(x_i^b - x_j^b)^2}{2\sigma_b^2} \right) + \sigma_n^2 \delta_{ij}, \quad (2)$$

where ν is a scaling factor, σ_n accounts for the noise standard deviation and σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b ($b = 1, \dots, B$). Model hyperparameters are collectively grouped into $\theta = [\nu, \sigma_n, \sigma_1, \dots, \sigma_B]$, and model weights α can be automatically obtained by maximizing the marginal likelihood in the training set. The obtained weights α after optimization give the relevance of each datum \mathbf{x}_i , while the inverse of σ_b represents the relevance of each band b . Hence, low values of σ_b indicate a higher informative content of this certain band b to the training function k .

The inversion process was conducted using 70% of the samples for training and the rest for testing with our SimpleR MATLAB toolbox, freely available at the Image Processing Laboratory website <http://isp.uv.es/>. The toolbox is intended for practitioners with little expertise in machine learning and that may want to assess advanced methods in their problems easily.

3.1. Sentinel-2A Processing

In the case of Sentinel-2A, top of canopy reflectances were obtained from Level 1C data after the atmospheric correction (AC) made using the Sen2Cor (Sentinel-2 atmospheric Correction) toolbox [38]. Sen2cor produces an orthoimage of top of canopy corrected reflectance for all bands, excluding the cirrus band, which does not contain surface information.

The atmospheric correction performed by Sen2cor is based on the inversion of the radiative transfer equation through the use of a large set of look-up tables (LUT) of sensor-specific functions, already integrated in the processor [39]. The LUT has been generated using the libRadtran, which contains a wide variety of atmospheric conditions, solar geometries and ground elevations. This database is generated with a high spectral resolution (0.6 nm) and then resampled using Sentinel-2A spectral responses. The Sen2cor algorithm allows the detection of clouds, water, snow and

cloud shadows. The algorithm is based on a series of threshold tests over the Sentinel-2A Level 1C data, band ratios and spectral indexes, which produces a probabilistic cloud mask and snow mask quality indicators [40].

Six Sentinel-2A surface reflectance spectral bands were used during the retrieval process: blue (Band 2), green (Band 3), red (Band 4), near infrared (Band 8) and the two short wave infrared channels (Bands 11 and 12). These channels were selected to enhance the consistency with Landsat-7/8 data [18], allowing thus the creation of a robust multi-sensor retrieval. We have disregarded red-edge bands, despite their potential usefulness for the retrieval of important biophysical parameters, such as chlorophyll, nitrogen content and brown LAI [21,41].

The combined use of Sentinel-2A and Landsat-7/8 poses a number of technical challenges, due to the differences in their orbital, spatial, spectral response functions and image processing chains. In fact, although the radiometric characteristics of these sensors are similar, they are not identical and can lead to slight differences in surface reflectance and retrieval quantities [42], as confirmed by the inspection of the data. In order to reduce this bias in the multi-sensor dataset, the Sentinel-2A images were radiometrically normalized by assuming a linear relationship between the reflectance of Sentinel-2A and Landsat-8 images. For each band, slope and intercept parameters were obtained using orthogonal linear regression. We applied Iteratively Reweighted Multivariate Alteration Detection (IRMAD) transformation [43] to select invariant pixels, which proved to be invariant to changes in sensor gain and offset and radiometric and atmospheric corrections.

3.2. Sentinel-1A Rice Maps

Rice maps derived from Sentinel-1A data were used as masking layer for LAI retrieval. The multi-temporal rule-based rice detection (MSRD) algorithm proposed by [44] for X-band HH SAR data was adapted to C-band VV/VH data and applied to the Sentinel-1A datasets over the three study areas. Processing has been carried out in MAPscape-RICE software. The MRSD algorithm relies on rules applied to the temporal profile of Sentinel-1A σ^0 and defined based on a priori knowledge on the rice calendar, crop practices and agro-ecological conditions of the study areas. The following rules are implemented for each pixel:

1. Rice exclusion condition: average σ^0 lower/higher than expected or average σ^0 below a minimum value for longer than expected or variation of σ^0 larger than expected;
2. Presence of flooding conditions at the start of season (SoS): the temporal series is analyzed starting from the first image acquisition to identify when σ^0 drops below a maximum values for SoS flooding; this time is set as SoS;
3. Confirmation of rice presence: after flooding detection; rice presence is confirmed if σ^0 increase after SoS is consistent with expected value for rice crops (rapid growth of rice biomass after flooding);
4. Late rice condition: when the length of the S1A time series after detected SoS is shorter than expected, rice is classified as being sown late in the season;
5. New rice/crop season: unexpected drop in σ^0 after SoS, which suggest a new flood or a new crop season. Notice that if Rule 2 is not met during the entire temporal series, another rule allows one to determine whether flooding occurred before the first Sentinel-1A acquisition. If σ^0 variation is above a pre-defined value, which is expected for rice and σ^0 decreases after the maximum value, the pixel is classified as early rice.

Crop maps for Italy were produced for the entire Rice district of Piedmont and Lombardy (~210,000 ha). The Lomellina test site (~50,000 ha) where LAI was retrieved belongs to this area.

4. Results and Validation

4.1. Rice Detection from Sentinel-1A

2016 rice maps were produced over the three study areas: Greece (GR), Italy (IT), Spain (SP) from Sentinel-1A (see Figure 2). Validation has been performed by comparison of the rice mapping products with reference observations collected during the season over the three study areas. In particular, in Italy, point data were used as provided by field survey exploiting a Voluntary Geographic Information App (S4A Smart App) [45], while for Greece and Spain, polygons were digitalized over HR optical data based on ground inspection during field campaigns. A confusion matrix [46] was performed at the pixel level (IT: 1833 rice and 1394 no rice pixels; GR: 8181 rice and 7507 no rice pixels; ES: 21792 rice and 7376 no rice pixels), and accuracy measures (overall accuracy (OA) and Kappa coefficient (K)) were derived. Results show that the maps have an OA and K greater than 89 and 0.78 respectively in all cases (IT: OA = 89.1, K = 0.78, GR: OA = 96.7, K = 0.93 and ES: OA = 99.3, K = 0.99). The corresponding confusion matrices are shown in Table S2 of the Supplementary Materials section. In general, the obtained results are more than satisfactory, in particular in Spain and Greece, where an overall accuracy higher than 95% has been achieved. In Italy, a lower overall accuracy, but still more than satisfactory, is reported. The reason is due to the different rice practices (wet and dry sowing) and the presence of other crops (in particular maize and soya), while showing, at the beginning of the crop season, a similar backscattering coefficient.

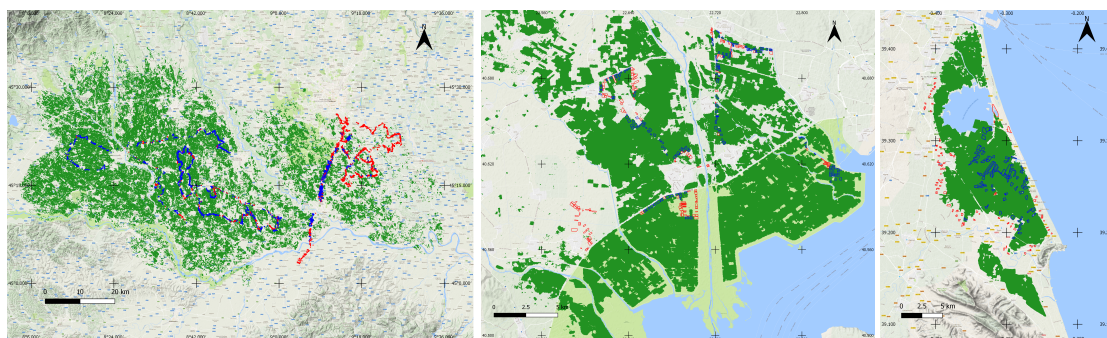


Figure 2. Rice area for Italy (IT) (left), Greece (GR) (middle) and Spain (ES) (right) at the end of the 2016 season and points/field boundaries over rice (red) and not rice (blue) fields.

4.2. Sentinel-2A LAI Estimates

High-resolution (10 m) LAI retrievals were obtained over the three rice areas during the 2016 rice season applying the trained model to Sentinel-2A data. Figure 3 shows the LAI map derived over the Spanish study area using a Sentinel-2A image acquired on 9 August 2016. The high spatial detail of estimates can be seen, which allows identifying significant different values within the same rice field. Those intra-field LAI differences are mainly due to the heterogeneity of the rice field caused by non-homogenous seeding and agro-practices [18,47]. This kind of information provided by LAI estimates at such a resolution is indeed very useful for rice crop users and modelers. In fact, key farmers were able to relate these local maps with their field water flow regime and cultivation practices. From interactions with farmers, it becomes clear that LAI variability maps at dense time points provide relevant information on crop failure of immediate use for improving their agro-practices. Note that non-rice areas were masked out during the retrieval process using the rice maps based on Sentinel-1A data acquired over each study area.

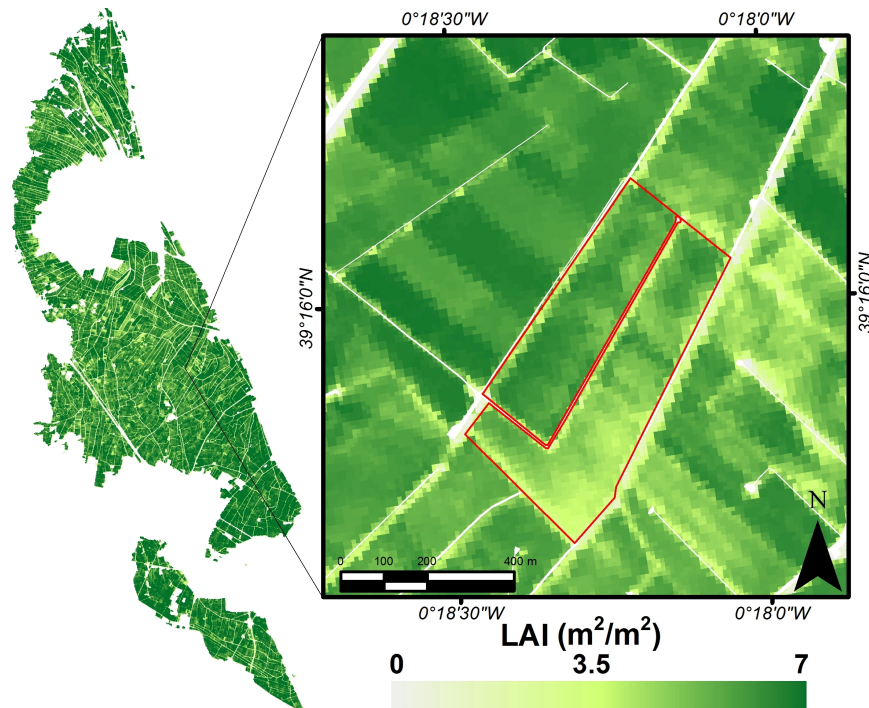


Figure 3. High-resolution LAI map obtained with Sentinel-2A data acquired in the Spanish study area on 9 August 2016. The right panel provides a zoom over a monitored field in the rice season.

These LAI maps can be used to continuously monitor the on-going cropping season and for early detection of crop growth anomaly connected to potential rice damage and yield loss. Preliminary effective application was conducted during the 2016 cropping season in the framework of the ERMES project. LAI maps were produced in near real time and provided through the ERMES geo-portal to farmers and users. In Italy, this information resulted in being useful to identify and assess the impact of the misuse of an herbicide. The maps, computed after the event and compared with the previous ones, were used together with the farmer to clearly delineate which part of the field was mainly impacted and needed recovery. On the other hand, anomalies in Spanish LAI time series were identified and related to fields affected by a rice disease thanks to the interaction with users. These results revealed that operational production of decametric LAI time series can represent a useful screening tools to support field inspection to identify situations likely affected by a crop problem that if not managed will produce yield shortage. This kind of information is also critical in crop modeling, since the identification of anomalous drops in LAI values can improve the accuracy of yield forecasting.

The availability of LAI ground measurements allowed assessing the high-resolution LAI estimates. The root mean squared error (RMSE), mean error (ME), mean absolute error (MAE) and coefficient of determination (R^2) were computed in order to assess the accuracy of the retrievals, bias and goodness-of-fit (see the values in Figure 4 and the metrics in Table S3 of the Supplementary Materials section). Good accuracy and high correlation were found revealing an overall RMSE of $0.69 \text{ m}^2/\text{m}^2$ and a coefficient of determination of 0.95. A slight bias was observed between Sentinel-2A LAI estimates and ground data. More in detail, Table 2 shows the statistical indicators between ground data and Sentinel-2A estimates over each study area. In general, very high correlations were found. Low biases were observed in the case of the Spanish study area, while in Italy and Greece, small biases were observed with respect to the ground data (see Table 2). Biases could be explained due the rice ESU heterogeneity and the slight underestimation of PocketLAI regarding other measurements [32]. Relative RMSE to mean (rRMSE_m) and relative RMS to range (rRMSE_r) are also reported in Table 2 and Figure 4 showing good accuracies and revealing that the Sentinel-2A LAI estimates are suitable to be integrated into crop models [48].

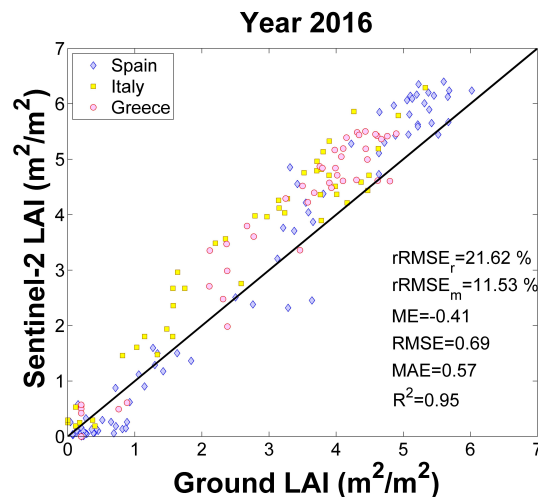


Figure 4. Scatter plots of Sentinel-2A estimated LAI values versus in situ LAI measurements during the 2016 rice season. For the sake of visualization, the standard deviation of measurements and Sentinel-2A prediction uncertainty ($\approx \pm 0.8$) are not shown.

Table 2. Statistical indicators (coefficient of determination (R^2), root mean squared error (RMSE), mean error (ME), mean absolute error (MAE), relative root mean squared error to mean ($rRMSE_m = RMSE/\hat{y}_i$) and relative root mean squared error to range ($rRMSE_r = RMSE/(\max(y_i) - \min(y_i))$) between the ground measurements and LAI estimates during the 2016 rice season over the three study areas.

	Sentinel-2A		
	Spain	Italy	Greece
R^2	0.97	0.94	0.94
RMSE (m^2/m^2)	0.56	0.82	0.77
ME (m^2/m^2)	−0.15	−0.67	−0.62
MAE (m^2/m^2)	0.43	0.69	0.60
$rRMSE_m$ (%)	20.1	23.5	22.7
$rRMSE_r$ (%)	9.3	16.5	15.3

4.3. Spatial-Temporal Consistency between Sentinel-2A and Landsat LAI Estimates

Sentinel-2A LAI estimates were spatially assessed with estimates obtained following the approach described in [18] for Landsat-8 LAI retrieval. Spatial consistency between Sentinel-2A and Landsat-8 LAI estimates was assessed computing the difference map of the pair of the closest acquisition dates between Landsat-8 and Sentinel-2A over each study area (see Figure 5). For this purposes, the Sentinel-2A images were resampled to Landsat-8 resolution, and the comparison was achieved averaging the LAI value computed over a window of 3×3 valid pixels in order to reduce coregistration errors between images [49] and inconsistencies associated with differences in the point spread functions [50]. Good correlations were found in all three study areas, and a notable agreement was found in Italy with practically no bias and low RMSE (see Figure 5). Leaf area index difference maps show the predominance of the light green color, which demonstrates that the retrievals from the two sensors are coherent, exhibiting LAI differences around zero over the majority of the rice areas. Higher differences are mainly due to differences in surface reflectances from Landsat-8 and Sentinel-2A generated by the different atmospheric correction methods. In the case of Spain (Figure 5, top-left), a surrounding halo effect can be discerned. This effect is due a suboptimal Landsat-8 atmospheric correction performance over the edge region between land and sea and could explain the wider dispersion observed in the scatter plot (see Figure 5, bottom-left).

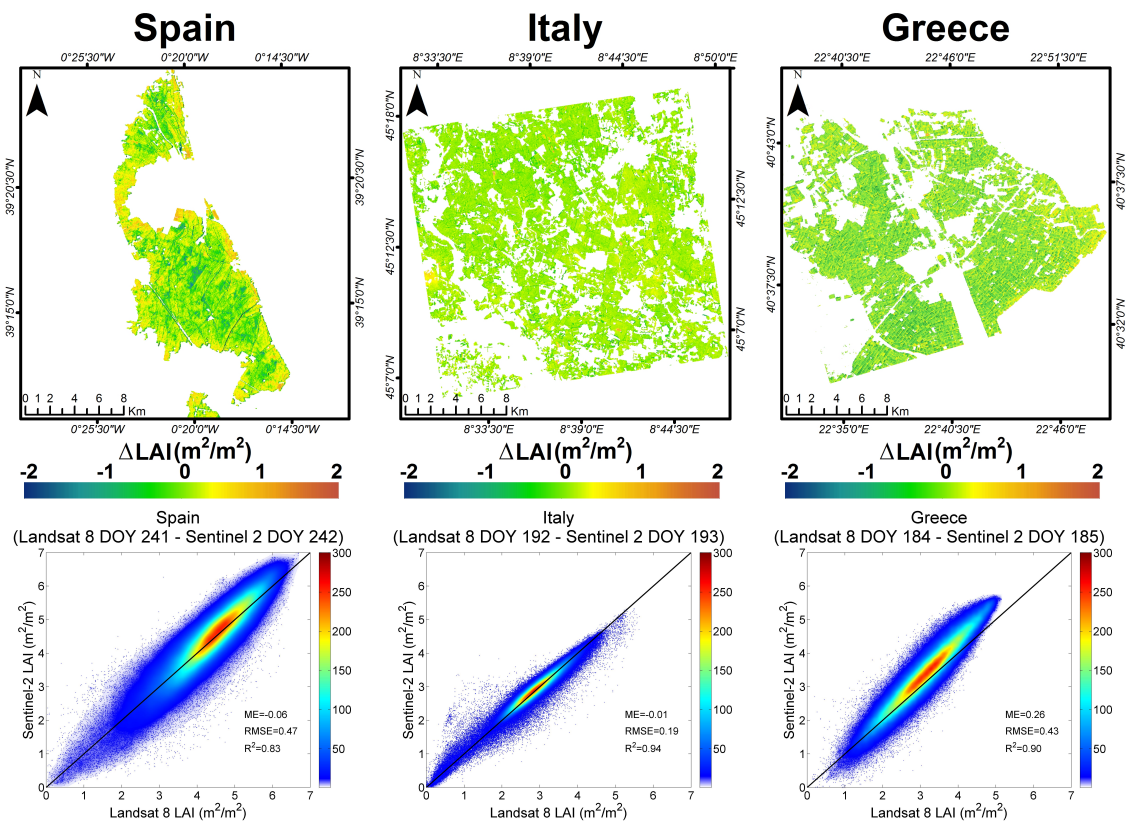


Figure 5. LAI difference maps (Top) and the corresponding scatter plots of the closest Sentinel-2A and Landsat-8 LAI estimates (Bottom) over Spain (Left), Italy (Middle) and Greece (Right).

In addition to the good spatial consistency between Sentinel-2A and Landsat-8 estimates, the 2016 LAI temporal behaviors are shown in Figure 6 which exhibits the temporal consistency of LAI estimates from Sentinel-2A (blue dots) with the ones from Landsat-7/8 (orange/green dots). Each dot is the estimated LAI averaged over a rice field; over each study site. The fields shown in Figure 6 were selected to cover LAI variability as large as possible and, where available, cultivated with local traditional varieties (Bomba and Carnaroli in the top panels of Spain and Italy). All of the other panels show modern rice varieties (Olympiada and Mare CL in Greece, Sirio CL and Selenio in the bottom panels of Spain and Italy, respectively). In Italy, there is larger variability of agro-practices, which is highlighted in the example field of Figure 6 by the significant difference in sowing dates (vertical dashed bars). Moreover, according to in situ observations, Field #68 is characterized by the presence of a cover crop preceding rice sowing, which is also confirmed by LAI profiles (high LAI values before the abrupt drop due to field preparation (ploughing and laser leveling) for rice sowing).

Cloud contamination is a common problem, which limits the utility of passive optical multispectral images. The use of multi-sensor approach allows filling time gaps typical of single sensor time series, due primarily to cloud cover, thus improving the usefulness of the satellite estimation. Figure 6 demonstrates the temporal consistency of multi-source LAI estimates as provided by Sentinel-2A and Landsat-7/8 images, since neither bias nor shifts can be clearly recognized. For example, in Field #36, missing Landsat-7/8 observations at the peak of the season due to cloud cover are filled by Sentinel-2A, thus improving retrieval accuracy of the maximum LAI value at the peak of the season. In detail, the merged time series reaches a maximum LAI value of $5.3 \text{ m}^2/\text{m}^2$ at DOY (day of year) 203, while the Landsat-7/8 series has a maximum LAI value of $4.3 \text{ m}^2/\text{m}^2$ at DOY 215. As in this example, a multi-sensor approach provides more accurate estimates of LAI at key phenological stages related to the shape of the curve given by the temporal profile (e.g., peaks, inflection points). Similar remarks can be done for Field #8267 in Greece, where three Sentinel-2A images are available at the peak of the

season, but only one from Landsat-7/8. The Spanish study site does not suffer a lack of Landsat-7/8 images because it is located in the overlapping region between two Landsat paths, which enhances the frequency of observation up to seven and nine days and the likelihood of clear sky conditions at image acquisition time. These preliminary results suggest that a very frequent time series of LAI at high resolution can be obtained from a multi-sensor approach to better outline rice-growing behavior. The combined curves of LAI could be exploited for: (i) the retrieval of phenological stages for crop modeling purposes [51–53]; (ii) deriving multitemporal training sets for mapping purposes [54]; or (iii) monitoring vegetation production by computing the area under the curve [55].

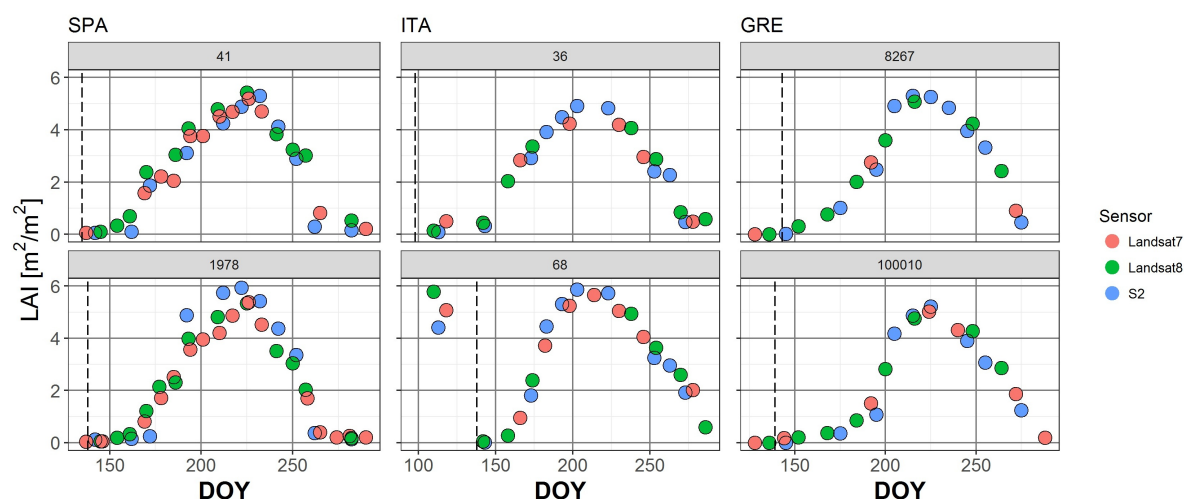


Figure 6. Example of merged series of LAI on three sites. Dotted vertical bars indicate the sowing date. Numbers indicate fields ID. For the sake of visualization, standard deviations are not shown.

4.4. On the Sources of Error and Uncertainty

It is worth mentioning that indirect and non-destructive methods, such as PocketLAI, provide estimates associated with several sources of measurement error, including wrong measurements and suboptimal sampling within an ESU, simplification of canopy leaf optical properties, illumination conditions and the optical signal saturation in dense vegetation. Specifically, PocketLAI measurements over rice crops showed an underestimation with respect to classical instrumentation, such as DHP [32]. The PocketLAI underestimation can be due to the fact that measurement is based on upward photography, which underestimates LAI in rice crops [56], while DHP techniques over rice typically use downward-looking photography due to the characteristics of rice crops (low height and flooded background). In addition, LAI measurements over rice present a saturation during flowering, fruit development and ripening phenological rice stages. Moreover, the specific background conditions of rice crops, which are flooded most of the time, except on some dates, when water is pumped out for agronomic purposes, may affect the temporal trend of non-destructive LAI ground measurements.

In this study, a specific PROSAIL parametrization for simulating rice cropping areas was used obtaining thus a single model, which was applied over the three countries. The sensitivity of the model to each region can be observed in Table 2, where estimates were more accurate and less skewed in Spain; nevertheless, the results for Italy and Greece were also good, revealing the robustness of the model. On the other hand, The GPR model not only provides a mean estimate for the predictions, but also a prediction uncertainty related to the confidence of the estimates (the lower the uncertainty, the higher the confidence). Although this uncertainty is a numerical value, it must be interpreted as a quality-driven indicator [22,57] related, in the case of rice, to (i) an incorrect simulation of rice background, which has a high impact on the reliability of LAI retrievals in rice crops, and (ii) to spectral data under-represented in the PROSAIL training database, such as water bodies and man-made surfaces, for which its associated LAI estimates presented significantly higher prediction uncertainty

(lower confidence). Uncertainty values over the three areas remained constant at a value of about 0.8 during the rice growing period, revealing the robustness of the GPR model. In addition, unexpected estimates (pronounced drops) provided by the GPR were also observed during the season and related to suboptimal atmospheric correction or undetected clouds in Sentinel-2A data, which were labeled as poor-quality estimates and consequently excluded. This feature allows one to create a quality flag indicator useful for users and crop modelers. Figure 7 shows the temporal evolution of Sentinel-2A LAI and uncertainty within a rice field. Note that the unexpected LAI drop reported on 7 September and the associated σ peak were related to a cloudy pixel undetected by Sen2cor.

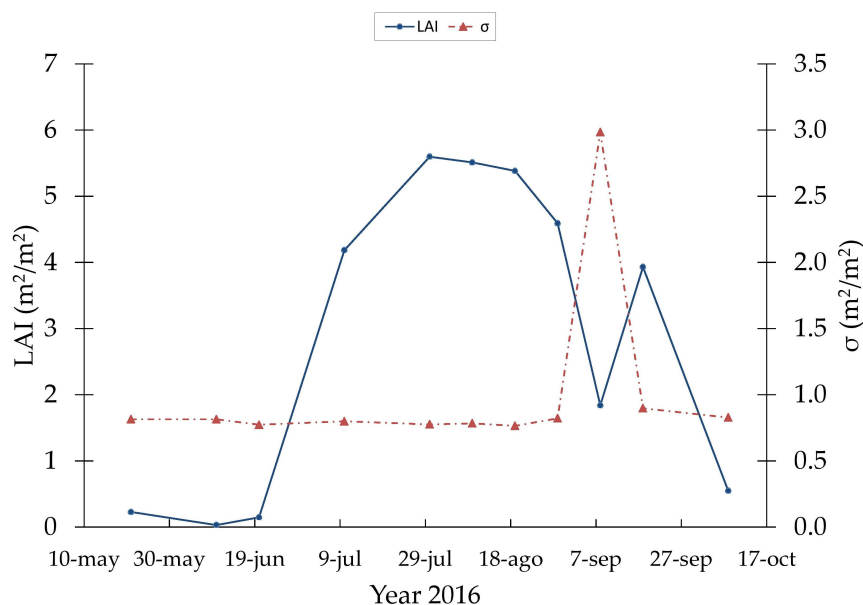


Figure 7. Temporal evolution of LAI and uncertainty (σ) over a rice field in Spain during the 2016 rice season.

5. Conclusions

This study presents 2016 seasonal LAI estimation from Sentinel-2A imagery and validation over the major European rice areas. The approach relies on the inversion of the PROSAIL radiative transfer model with Gaussian process regression on rice field automatically detected using Sentinel-1A data. The proposed approach allows us to retrieve multitemporal high-resolution LAI estimates by exploiting the enhanced spatial and spectral resolutions of the ESA Sentinels constellation.

Sentinel-2A estimates are in agreement with field measurements acquired in the three rice areas during the 2016 crop season. In situ measurements were acquired with smartphones using the PocketLAI app, which allowed conducting coordinated field campaigns from the beginning of the leaf development up to the flowering phase. The spatial and multi-temporal consistency of Sentinel-2A LAI estimates was assessed by comparison with LAI maps obtained from a consolidated and published approach applied to Landsat-7/8 images [18]. The single date pixel-by-pixel agreement shows the goodness of fit of Sentinel-2A and Landsat-8 LAI estimations over the three rice areas (R^2 ranging from 0.83–0.94). In addition, the uncertainty provided by the GPR allows enhancing the detection of artifacts, such as undetected clouds by the Sen2cor module.

Preliminary tests conducted by combining Sentinel-2A and Landsat-7/8 LAI time series highlighted the feasibility of deriving high temporal resolution and multi-sensor LAI estimates. The increased temporal resolution of multi-sensor LAI allows filling gaps mainly due to atmospheric interference to obtain more reliable time series for precision agriculture applications and rice monitoring. A dense temporal dataset of LAI maps is in fact fundamental to perform expert crop monitoring and/or improve crop model estimations exploiting assimilation techniques. The free

availability of Sentinel-2A and Landsat-7/8 data constitutes an obvious advantage in adopting the proposed approach. Future work will focus on the quantitative validation of the improvements brought by multi-sensor LAI time series at the field level for deriving phenological stages, distinguishing rice varieties and estimating green biomass production and yield. Moreover, the proposed processing chain has immediate applicability to oncoming Sentinel-2B data. .

Supplementary Materials: The following are available online at www.mdpi.com/2072-4292/9/3/248/s1, Table S1: Recap of available optical remote sensing data over each study area during the 2016 rice season. Table S2: Confusion matrices related to rice crop maps. Table S3: Metrics of the statistical indicators.

Acknowledgments: The research leading to these results was conducted within the ERMES FP7 project (<http://www.ermes-fp7space.eu/>) which received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement 606983. This work was also supported by the European Research Council (ERC) funding under the ERC Consolidator Grant 2014 SEDAL (Statistical Learning for Earth Observation Data Analysis) project under Grant Agreement 647423 and by the European Organization for the Exploitation of Meteorological Satellites through the Land-SAF (Land Surface Analysis Satellite Applications Facility) project.

Author Contributions: All co-authors of this manuscript significantly contributed to all phases of the investigation. They contributed equally to the preparation, analysis, review and editing of this manuscript.

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

References

1. Gobron, N.; Verstraete, M. ECV T11: Leaf Area Index. In *Assessment of the Status of the Development of Standards for the Terrestrial Essential Climate Variables*; FAO: Rome, Italy, 2009.
2. Chen, J.M.; Black, T.A. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* **1992**, *15*, 421–429.
3. The Global Climate Observing System (GCOS). *Systematic Observation Requirements for Satellite-Based Products for Climate*; GCOS: Geneva, Switzerland, 2011.
4. Busetto, L.; Casteleyn, S.; Granell, C.; Pepe, M.; Barbieri, M.; Campos-Taberner, M.; Casa, R.; Confalonieri, R.; Crema, A.; García-Haro, F.J.; et al. Downstream services for rice crop monitoring in Europe: From regional to local scale. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, accepted.
5. Baret, F.; Houles, V.; Guérif, M. Quantification of plant stress using remote sensing observations and crop models: The case of nitrogen management. *J. Exp. Bot.* **2007**, *58*, 869.
6. Aboelghar, M.; Arafat, S.; Yousef, M.A.; El-Shirbeny, M.; Naeem, S.; Massoud, A.; Saleh, N. Using SPOT data and leaf area index for rice yield estimation in Egyptian Nile delta. *Egypt. J. Remote Sens. Space Sci.* **2011**, *14*, 81–89.
7. Confalonieri, R.; Rosenmund, A.S.; Baruth, B. An improved model to simulate rice yield. *Agron. Sustain. Dev.* **2009**, *29*, 463–474.
8. Launay, M.; Guerif, M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric. Ecosyst. Environ.* **2005**, *111*, 321–339.
9. Mosleh, M.K.; Hassan, Q.K.; Chowdhury, E.H. Application of Remote Sensors in Mapping Rice Area and Forecasting Its Production: A Review. *Sensors* **2015**, *15*, 769–791.
10. Dorigo, W.; Zurita-Milla, R.; de Wit, A.; Brazile, J.; Singh, R.; Schaepman, M. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 165–193.
11. Curnel, Y.; de Wit, A.J.; Duveiller, G.; Defourny, P. Potential performances of remotely sensed {LAI} assimilation in {WOFOST} model based on an {OSS} Experiment. *Agric. For. Meteorol.* **2011**, *151*, 1843–1855.
12. Holecz, F.; Barbieri, M.; Collivignarelli, F.; Gatti, L.; Nelson, A.; Setiyono, T.D.; Boschetti, M.; Manfron, G.; Brivio, P.A.; Quilang, J.; et al. An operational remote sensing based service for rice production estimation at national scale. In *Proceedings of the Living Planet Symposium, Edinburgh, UK, 9–13 September 2013*.
13. Ata-Ul-Karim, S.T.; Zhu, Y.; Yao, X.; Cao, W. Determination of critical nitrogen dilution curve based on leaf area index in rice. *Field Crop. Res.* **2014**, *167*, 76–85.
14. Confalonieri, R.; Debellini, C.; Pirondini, M.; Possenti, P.; Bergamini, L.; Barlassina, G.; Bartoli, A.; Agostoni, E.; Appiani, M.; Babazadeh, L.; et al. A new approach for determining rice critical nitrogen concentration. *J. Agric. Sci.* **2011**, *149*, 633–638.

15. Roy, D.; Wulder, M.; Loveland, T.; Woodcock, C.E.; Allen, R.; Anderson, M.; Helder, D.; Irons, J.; Johnson, D.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172.
16. Gonsamo, A. Leaf area index retrieval using gap fractions obtained from high resolution satellite data: Comparisons of approaches, scales and atmospheric effects. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 233–248.
17. Duan, S.B.; Li, Z.L.; Wu, H.; Tang, B.H.; Ma, L.; Zhao, E.; Li, C. Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 12–20.
18. Campos-Taberner, M.; García-Haro, F.J.; Camps-Valls, G.; Grau-Muedra, G.; Nutini, F.; Crema, A.; Boschetti, M. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. *Remote Sens. Environ.* **2016**, *187*, 102–118.
19. Drusch, M.; Bello, U.D.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36.
20. Verrelst, J.; Camps-Valls, G.; Muñoz-Marí, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.; Moreno, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290.
21. Delegido, J.; Verrelst, J.; Rivera, J.P.; Ruiz-Verdú, A.; Moreno, J. Brown and green LAI mapping through spectral indices. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *35*, 350–358.
22. Campos-Taberner, M.; García-Haro, F.; Moreno, A.; Gilabert, M.; Sanchez-Ruiz, S.; Martinez, B.; Camps-Valls, G. Mapping Leaf Area Index With a Smartphone and Gaussian Processes. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2501–2505.
23. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66.
24. Atzberger, C.; Richter, K. Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sens. Environ.* **2012**, *120*, 208–218.
25. Verrelst, J.; Muñoz, J.; Alonso, L.; Delegido, J.; Rivera, J.P.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* **2012**, *118*, 127–139.
26. Haykin, S. *Neural Networks—A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
27. Shawe-Taylor, J.; Cristianini, N. *Kernel Methods for Pattern Analysis*; Cambridge University Press: Cambridge, UK, 2004.
28. Camps-Valls, G.; Bruzzone, L. *Kernel Methods for Remote Sensing Data Analysis*; Wiley & Sons: Chichester, UK, 2009.
29. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: New York, NY, USA, 2006.
30. Camps-Valls, G.; Muñoz-Marí, J.; Verrelst, J.; Mateo, F.; Gomez-Dans, J. A Survey on Gaussian Processes for Earth Observation Data Analysis. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 58–78.
31. Confalonieri, R.; Foi, M.; Casa, R.; Aquaro, S.; Tona, E.; Peterle, M.; Boldini, A.; Carli, G.D.; Ferrari, A.; Finotto, G.; et al. Development of an app for estimating leaf area index using a smartphone. Trueness and precision determination and comparison with other indirect methods. *Comput. Electron. Agric.* **2013**, *96*, 67–74.
32. Campos-Taberner, M.; García-Haro, F.J.; Confalonieri, R.; Martínez, B.; Moreno, A.; Sánchez-Ruiz, S.; Gilabert, M.A.; Camacho, F.; Boschetti, M.; Busetto, L. Multitemporal Monitoring of Plant Area Index in the Valencia Rice District with PocketLAI. *Remote Sens.* **2016**, *8*, 202.
33. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56.
34. Aspert, F.; Bach-Cuadra, M.; Cantone, A.; Holecz, F.; Thiran, J.P. Time-varying segmentation for mapping of land cover changes. In Proceedings of the ENVISAT Symposium, Montreux, Switzerland, 23–27 April 2007.

35. Darvishzadeh, R.; Matkan, A.A.; Ahangar, A.D. Inversion of a Radiative Transfer Model for Estimation of Rice Canopy Chlorophyll Content Using a Lookup-Table Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1222–1230.
36. Baret, F.; Hagolle, O.; Geiger, B.; Bicheron, P.; Miras, B.; Huc, M.; Berthelot, B.; Nino, F.; Weiss, M.; Samain, O.; et al. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. *Remote Sens. Environ.* **2007**, *110*, 275–286.
37. Claverie, M.; Vermote, E.F.; Weiss, M.; Baret, F.; Hagolle, O.; Demarez, V. Validation of coarse spatial resolution LAI and FAPAR time series over cropland in southwest France. *Remote Sens. Environ.* **2013**, *139*, 216–230.
38. Müller-Wilm, U. *Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual*; Telespazio VEGA Deutschland GmbH: Darmstadt, Germany, 2016.
39. Richter, R.; Schlaepfer, D. *Atmospheric/Topographic Correction for Satellite Imagery: ATCOR-2/3 User Guide Vers. 8.0.2*; DLR—German Aerospace Center, Remote Sensing Data Center: Wessling, Germany, 2011.
40. Louis, J.; Charantonis, A.; Berthelot, B. *Cloud Detection for Sentinel-2*; ESA Special Publication: Noordwijk, The Netherlands, 2010.
41. Clevers, J.; Gitelson, A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–351.
42. Mandanici, E.; Bitelli, G. Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use. *Remote Sens.* **2016**, *8*, 1014.
43. Nielsen, A.A.; Conradsen, K.; Simpson, J.J. Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies. *Remote Sens. Environ.* **1998**, *64*, 1–19.
44. Nelson, A.; Setiyono, T.; Rala, A.; Quicho, E.; Raviz, J.; Abonete, P.; Maunahan, A.; Garcia, C.; Bhatti, H.; Villano, L.; et al. Towards an Operational SAR-Based Rice Monitoring System in Asia: Examples from 13 Demonstration Sites across Asia in the RIICE Project. *Remote Sens.* **2014**, *6*, 10773–10812.
45. Bordogna, G.; Kliment, T.; Frigerio, L.; Brivio, P.A.; Crema, A.; Stroppiana, D.; Boschetti, M.; Sterlacchini, S. A Spatial Data Infrastructure Integrating Multisource Heterogeneous Geospatial Data and Time Series: A Study Case in Agriculture. *ISPRS Int. J. Geoinf.* **2016**, *5*, 73.
46. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
47. Ranghetti, L.; Busetto, L.; Crema, A.; Fasola, M.; Cardarelli, E.; Boschetti, M. Testing estimation of water surface in Italian rice district from MODIS satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 284–295.
48. EL HAJJ, M.; Baghdadi, N.; Cheviron, B.; Belaud, G.; Zribi, M. Integration of remote sensing derived parameters in crop models: Application to the PILOTE model for hay production. *Agric. Water Manag.* **2016**, *176*, 67–79.
49. Storey, J.; Roy, D.P.; Masek, J.; Gascon, F.; Dwyer, J.; Choate, M. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sens. Environ.* **2016**, *186*, 121–122.
50. Morisette, J.; Baret, F.; Privette, J.; Myneni, R.; Nickeson, J.; Garrigues, S.; Shabanov, N.; Weiss, M.; Fernandes, R.; Leblanc, S.; et al. Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1804–1817.
51. Boschetti, M.; Stroppiana, D.; Brivio, P.; Bocchi, S. Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. *Int. J. Remote Sens.* **2009**, *30*, 4643–4662.
52. De Bernardis, C.; Vicente-Guijalba, F.; Martinez-Marin, T.; Lopez-Sanchez, J.M. Particle Filter Approach for Real-Time Estimation of Crop Phenological States Using Time Series of NDVI Images. *Remote Sens.* **2016**, *8*, 610.
53. Boschetti, M.; Busetto, L.; Manfron, G.; Laborte, A.; Asilo, S.; Pazhanivelan, S.; Nelson, A. PhenoRice: A method for automatic extraction of spatio-temporal information on rice crops using satellite data time series. *Remote Sens. Environ.* **2017**, under revision.

54. Nutini, F.; Boschetti, M.; Brivio, P.A.; Bocchi, S.; Antoninetti, M. Land-use and land-cover change detection in a semi-arid area of Niger using multi-temporal analysis of Landsat images. *Int. J. Remote Sens.* **2013**, *34*, 4769–4790.
55. Heumann, B.W.; Seaquist, J.; Eklundh, L.; Jönsson, P. AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982–2005. *Remote Sens. Environ.* **2007**, *108*, 385–392.
56. Fang, H.; Li, W.; Wei, S.; Jiang, C. Seasonal variation of leaf area index (LAI) over paddy rice fields in NE China: Intercomparison of destructive sampling, LAI-2200, digital hemispherical photography (DHP), and AccuPAR methods. *Agric. For. Meteorol.* **2014**, *198*, 126–141.
57. Verrelst, J.; Rivera, J.P.; Moreno, J.; Camps-Valls, G. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 157–167.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Acronyms

ALA	Average Leaf Angle
AVHRR	Advanced Very High Resolution Radiometer
BEF	Broadleaved Evergreen Forest
BELMANIP	Benchmark Land Multisite Analysis and Intercomparison of Products
BDF	Broadleaved Deciduous Forest
BRDF	Bidirectional Reflectance Distribution Function
BRF	Bi-directional Reflectance Factor
CDR	Climate Data Records
CEOS	Committee of Earth Observing Satellites
CI	Clumping Index
CM	Cultivated and Mosaic
DHP	Digital Hemispherical Photography
DOY	Day of the year
EMMAH	<i>Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes</i>
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
ECV	Essential Climate Variable
EPS	EUMETSAT Polar System
ERMES	An Earth obseRvation Model based RicE information Service
ESA	European Space Agency

ESU	Elementary Sampling Unit
ETAL	EPS ten-day albedo product
ETFAPAR	EPS ten-day FAPAR product
ETFVC	EPS ten-day FVC product
ETLAI	EPS ten-day LAI product
ETM+	Enhanced Thematic Mapper
EVI	Enhanced Vegetation Index
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FVC	Fractional Vegetation Cover
GAI	Green Area Index
GCOS	Global Climate Observing System
GP	Gaussian Process
GPR	Gaussian Process Regression
GPS	Global Positioning System
HR	High-Resolution
HSB	Hue Saturation Brightness
INRA	<i>Institut National de la Recherche Agronomique</i>
IPMA	<i>Instituto Portugues do Mar e Atmosfera</i>
KRR	Kernel Ridge Regression
LAI	Leaf Area Index
LANDSAT	LAND observation SATellite
LAND-SAF	LAND surface analysis Satellite Applications Facility
LaSRC	Landsat-8 Surface Reflectance (LaSRC)

LDCM	Landsat Data Continuity Mission
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LIDF	Leaf Inclination Density Function
LODF	Leaf Orientation Density Function
LPV	Land Product Validation
LUT	Look-Up Tables
MAE	Mean Absolute Error
ME	Mean Error
MetOP	Meteorological–Operational
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
NLF	Needleleaved Forest
NN	Neural Network
NOAA	National Oceanographic and Atmospheric Administration
NPP	Net Primary Productivity
NRT	Near Real Time
NWP	Numerical Weather Prediction
PAI	Plant Area Index
PAR	Photosynthetically Active Radiation
PROBA	Project for On-Board Autonomy
PSF	Point Spread Function

RMSE	Root Mean Square Error
RTM	Radiative Transfer Model
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SPOT	Satellite Pour l'Observation de la Terre
SAIL	Scattering by Arbitrary Inclined Leaves
SWIR	Short Wave Infra red
TOA	Top of Atmosphere
TOC	Top of Canopy
OLI	Operational Land Imager
VALERI	Validation of Land European Remote sensing Instruments
VI _s	Vegetation Indices
VITO	<i>Vlaamse Instelling Voor Technologisch Onderzoek</i>

Bibliography

- Allen, W. A., Gausman, H. W., Richardson, A. J., & Thomas, J. R. (1969). Interaction of isotropic light with a compact plant leaf. *JOSA*, 59, 1376–1379.
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote sensing of environment*, 35, 161–173.
- Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J. L., & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. *Remote Sensing of Environment*, 110, 275 – 286.
- Baret, F., Jacquemoud, S., Guyot, G., & Leprieux, C. (1992). Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sensing of Environment*, 41, 133–1142.
- Baret, F., Morisette, J. T., Fernandes, R. A., Champeaux, J. L., Myneni, R. B., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., & Nickeson, J. E. (2006). Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: proposition of the ceos-belmanip. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1794–1803.
- Baret, F., de Solan, B., Lopez-Lozano, R., Ma, K., & Weiss, M. (2010). GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agricultural and Forest Meteorology*, 150, 1393 – 1401.
- Baret, F., Weiss, M., Allard, D., Garrigues, S., Leroy, M., Jeanjean, H., Fernandes, R., Myneni, R., Privette, J., Morisette, J. et al. (2005). VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products. *Remote Sensing of Environment*, 76, 36–39.
- Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholczyk, P., & Smets, B. (2013). Geov1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. *Remote Sensing of Environment*, 137, 299–309.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of photogrammetry and remote sensing*, 58, 239–258.
- Bishop, C. M. (2006). Pattern recognition. *Machine Learning*, 128, 1–58.

- Boschetti, M., Nutini, F., Manfron, G., Brivio, P. A., & Nelson, A. (2014). Comparative analysis of normalised difference spectral indices derived from modis for detecting surface water in flooded rice cropping systems. *PloS one*, 9, e88741.
- Breda, N. J. J. (2003). Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany*, 54, 2403–2417.
- Bunnik, N. J. J. (1978). *The multispectral reflectance of shortwave radiation by agricultural crops in relation with their morphological and optical properties*. Ph.D. thesis Proefschrift Wageningen.
- Busetto, L., Casteleyn, S., Granell, C., Pepe, M., Barbieri, M., Campos-Taberner, M., Casa, R., Collivignarelli, F., Confalonieri, R., Crema, A., García-Haro, F. J., Gatti, L., Gitas, I. Z., González-Pérez, A., Grau-Muedra, G., Guarneri, T., Holecz, F., Katsantonis, D., Minakou, C., Miralles, I., Moredi, E., Nutini, F., Pagani, V., Palombo, A., Paola, F. D., Pascucci, S., Pignatti, S., Rampini, A., Ranghetti, L., Ricciardelli, E., Romano, F., Stavrakoudis, D. G., Stroppiana, D., Viggiano, M., & Boschetti, M. (2017). Downstream services for rice crop monitoring in europe: From regional to local scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP, 1–19.
- Caicedo, J. P. R., Verrelst, J., Munoz-Mari, J., Moreno, J., & Camps-Valls, G. (2014). Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7, 1249–1259.
- Campos-Taberner, M., García-Haro, F. J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Busetto, L., Katsantonis, D., Stavrakoudis, D., Minakou, C., Gatti, L., Barbieri, M., Holecz, F., Stroppiana, D., & Boschetti, M. (2017a). Exploitation of sar and optical sentinel data to detect rice crop and estimate seasonal dynamics of leaf area index. *Remote Sensing*, 9(7), 248.
- Campos-Taberner, M., García-Haro, F. J., Confalonieri, R., Martínez, B., Moreno, I., Sánchez-Ruiz, S., Gilabert, M. A., Camacho, F., Boschetti, M., & Busetto, L. (2016a). Multitemporal monitoring of plant area index in the valencia rice district with pocketlai. *Remote Sensing*, 8(3), 202.
- Campos-Taberner, M., García-Haro, F. J., Moreno, A., Gilabert, M. A., Martínez, B., Sánchez-Ruiz, S., & Camps-Valls, G. (2015a). Development of an earth observation processing chain for crop bio-physical parameters at local scale. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International* (pp. 29–32).
- Campos-Taberner, M., Garcia-Haro, F., Moreno, A., Gilabert, M., Sanchez-Ruiz, S., Martinez, B., & Camps-Valls, G. (2015b). Mapping leaf area index with a smartphone and Gaussian processes. *Geoscience and Remote Sensing Letters, IEEE*, 12, 2501–2505.
- Campos-Taberner, M., García-Haro, F. J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A., & Boschetti, M. (2016b). Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. *Remote Sensing of Environment*, 187, 102 – 118.

- Campos-Taberner, M., García-Haro, F. J., Confalonieri, R., Martínez, B., Moreno, A., Sánchez-Ruiz, S., Gilabert, M., Camacho, F., Boschetti, M., & Busetto, L. (2015c). Intercomparison of instruments for measuring leaf area index over rice. In *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International* (pp. 3389–3392). IEEE.
- Campos-Taberner, M., Grau-Muedra, G., García-Haro, F., Martínez, B., Sánchez-Ruiz, S., Gilabert, M., & Confalonieri, R. (2017b). Use of smartphones as measuring tools for educational purposes in remote sensing. In *INTED2017 Proceedings 11th International Technology, Education and Development Conference* (pp. 5474–5478). IATED.
- Campos-Taberner, M., Romero-Soriano, A., Gatta, C., Camps-Valls, G., Lagrange, A., Saux, B. L., Beaupère, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H., Ferecatu, M., Shimoni, M., Moser, G., & Tuia, D. (2016c). Processing of extremely high-resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS Data Fusion Contest-Part A: 2-D Contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9, 5547–5559.
- Camps-Valls, G., Svendsen, D. H., Martino, L., Muñoz-Marí, J., Laparra, V., Campos-Taberner, M., & Luengo, D. (2017). Physics-aware gaussian processes for earth observation. In P. Sharma, & F. M. Bianchi (Eds.), *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12-14, 2017. Proceedings* (pp. 205–217). Springer International Publishing.
- Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., & Malo, J. (2011). Remote sensing image processing. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 5, 1–192.
- Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62, 241 – 252.
- Chen, J., Menges, C., & Leblanc, S. (2005). Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment*, 97, 447–457.
- Chen, J. M., & Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell and Environment*, 15, 421–429.
- Chen, J. M., & Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote sensing of Environment*, 55, 153–162.
- Cohen, W. B., Maersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84, 561 – 571.
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 84, 1–15.
- Confalonieri, R., Foi, M., Casa, R., Aquaro, S., Tona, E., Peterle, M., Boldini, A., Carli, G. D., Ferrari, A., Finotto, G., Guarneri, T., Manzoni, V., Moredi, E., Nisoli, A., Paleari, L., Radici, I., Suardi, M., Veronesi, D., Bregaglio, S., Cappelli, G., Chiodini, M.,

- Dominoni, P., Francone, C., Frasso, N., Stella, T., & Acutis, M. (2013). Development of an app for estimating leaf area index using a smartphone. trueness and precision determination and comparison with other indirect methods. *Computers and Electronics in Agriculture*, 96, 67 – 74.
- Confalonieri, R., Rosenmund, A. S., & Baruth, B. (2009). An improved model to simulate rice yield. *Agronomy for Sustainable Development*, 29, 463–474.
- Confalonieri, R., Stroppiana, D., Boschetti, M., Gusberti, D., Bocchi, S., & Acutis, M. (2006). Analysis of rice sample size variability due to development stage, nitrogen fertilization, sowing technique and variety using the visual jackknife. *Field Crops Research*, 97, 135 – 141.
- Delgado, J. A., Kowalski, K., & Tebbe, C. (2013). The first nitrogen index app for mobile devices: Using portable technology for smart agricultural management. *Computers and electronics in agriculture*, 91, 121–123.
- Demarez, V., Duthoit, S., Baret, F., Weiss, M., & Dedieu, G. (2008). Estimation of leaf area and clumping indexes of crops with hemispherical photographs. *Agricultural and Forest Meteorology*, 148, 644 – 655.
- Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., & 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, 165–193.
- Drusch, M., Bello, U. D., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25 – 36.
- Fang, H., Li, W., Wei, S., & Jiang, C. (2014). Seasonal variation of leaf area index (lai) over paddy rice fields in NE China: Intercomparison of destructive sampling, LAI-2200, digital hemispherical photography (DHP), and AccuPAR methods. *Agricultural and Forest Meteorology*, 198–199, 126 – 141.
- Fang, H., & Liang, S. (2005). A hybrid inversion method for mapping leaf area index from MODIS data: experiments and application to broadleaf and needleleaf canopies. *Remote Sensing of Environment*, 94, 405 – 424.
- Fernandes, R., Butson, C., Leblanc, S., & Latifovic, R. (2003). Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data. *Canadian Journal of Remote Sensing*, 29, 241–258.
- Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., & Widlowski, J. (2014). Global leaf area index product validation good practices. version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), *Best Practice for Satellite-Derived Land Product Validation: Land Product Validation Subgroup (WGCV/CEOS)* (p. 76).

- Francl, L., & Panigrahi, S. (1997). Artificial neural network models of wheat leaf wetness. *Agricultural and Forest Meteorology*, 88, 57–65.
- Francone, C., Pagani, V., Foi, M., Cappelli, G., & Confalonieri, R. (2014). Comparison of leaf area index estimates by ceptometer and pocketlai smart app in canopies with different structures. *Field Crops Research*, 155, 38 – 41.
- Ganguly, S., Schull, M. A., Samanta, A., Shabanov, N. V., Milesi, C., Nemani, R. R., Knyazikhin, Y., & Myneni, R. B. (2008). Generating vegetation leaf area index earth system data record from multiple sensors. Part 1: Theory. *Remote Sensing of Environment*, 112, 4333–4343.
- Gao, F., Morisette, J., Wolfe, R., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Tan, B., & Nightingale, J. (2008). An algorithm to produce temporally and spatially continuous MODIS-LAI time series. *Geoscience and Remote Sensing Letters, IEEE*, 5, 60–64.
- García-Haro, F., Sommer, S., & Kemper, T. (2005). A new tool for variable multiple endmember spectral mixture analysis (vmesma). *International Journal of Remote Sensing*, 26, 2135–2162.
- Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y., & Yang, W. (2008a). Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research: Biogeosciences*, 113, n/a–n/a. G02028.
- Garrigues, S., Shabanov, N., Swanson, K., Morisette, J., Baret, F., & Myneni, R. (2008b). Intercomparison and sensitivity analysis of leaf area index retrievals from LAI-2000, AccuPAR, and digital hemispherical photography over croplands. *Agricultural and Forest Meteorology*, 148, 1193 – 1209.
- Geiger, B., Carrer, D., Hautecoeur, O., Franchistéguy, L., Roujean, J.-L., Catherine Meurey, X. C., & Jacob, G. (2016). Algorithm theoretical basis document (atbd). land surface albedo products: Lsa-103 (etal). URL: <https://landsaf.ipma.pt/GetDocument.do?id=641>.
- Green, D. S., Erickson, J. E., & Kruger, E. L. (2003). Foliar morphology and canopy nitrogen as predictors of light-use efficiency in terrestrial vegetation. *Agricultural and Forest Meteorology*, 115, 163 – 171.
- Hagolle, O., Lobo, A., Maisongrande, P., Cabot, F., Duchemin, B., & De Pereyra, A. (2005). Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of vegetation 1 and 2 images. *Remote Sensing of Environment*, 94, 172–186.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning* (pp. 485–585). Springer.
- Haykin, S. (1999). *Neural Networks – A Comprehensive Foundation*. (2nd ed.). Prentice Hall.
- Huang, D., Knyazikhin, Y., Dickinson, R. E., Rautiainen, M., Stenberg, P., Disney, M., Lewis, P., Cescatti, A., Tian, Y., Verhoef, W. et al. (2007). Canopy spectral invariants for remote sensing and model applications. *Remote Sensing of Environment*, 106, 106–122.

- Jacquemoud, S., Bacour, C., Poilvé, H., & Frangi, J.-P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, 74, 471–481.
- Jacquemoud, S., & Baret, F. (1990). Prospect: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34, 75–91.
- Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & 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, 19 – 35.
- Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., & Baret, F. (2000). Developments in the 'validation' of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing*, 21, 3383–3390.
- Justice, C. O., Vermote, E., Townshend, J. R., Defries, R., Roy, D. P., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A. et al. (1998). The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1228–1249.
- Kimes, D., Nelson, R., Manry, M., & Fung, A. (1998). Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *Int. Jour. Rem. Sens.*, 19, 2639–2663.
- Knyazikhin, Y., Glassy, J., Privette, J., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J., Votava, P., Myneni, R. et al. (1999). Modis leaf area index (lai) and fraction of photosynthetically active radiation absorbed by vegetation (fpar) product (mod15) algorithm theoretical basis document. *Theoretical Basis Document, NASA Goddard Space Flight Center, Greenbelt, MD, 20771*.
- Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., & Running, S. (1998). Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from modis and misr data. *Journal of geophysical research*, 103, 32257–32276.
- Kuss, M., & Rasmussen, C. (2005). Assessing approximate inference for binary Gaussian process classification. *Machine learning research*, 6, 1679–1704.
- Kuusk, A. (1991). The hot spot effect in plant canopy reflectance. In *Photon-Vegetation Interactions* (pp. 139–159). Springer.
- Lawrence, N. (2005). Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Machine learning research*, 6, 1783–1816.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2014). *Remote sensing and image interpretation*. John Wiley & Sons.
- Luengo, D., Campos-Taberner, M., & Camps-Valls, G. (2016). Latent force models for earth observation time series prediction. In *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on* (pp. 1–6). IEEE.

- Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E., García-Santos, G., Fernandes, R., & Berger, M. (2012). Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, 120, 91 – 101.
- McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., & Nilsson, S. (2009). Satellite-based terrestrial production efficiency modeling. *Carbon balance and management*, 4, 8.
- Miller, J. (1967). A formula for average foliage density. *Australian Journal of Botany*, (p. 141–144).
- Morisette, J., Baret, F., Privette, J., Myneni, R., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sanchez-Azofeifa, G., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E., Petersen, G., Seufert, G., & Cook, R. (2006). Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *Geoscience and Remote Sensing, IEEE Transactions on*, 44, 1804–1817.
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358–371.
- Myneni, R., Ramakrishna, R., Nemani, R., & Running, S. (1997). Estimation of global leaf area index and absorbed par using radiative transfer models. *Geoscience and Remote Sensing, IEEE Transactions on*, 35, 1380–1393.
- Myneni, R., & Williams, D. (1994). On the relationship between fapar and ndvi. *Remote Sensing of Environment*, 49, 200–211.
- Myneni, R. B., & Ross, J. (2012). *Photon-Vegetation interactions: applications in optical remote sensing and plant ecology*. Springer Science & Business Media.
- O'Hagan, A., & Kingman, J. F. C. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40, 1–42.
- Paruelo, J. c., & Tomasel, F. d. (1997). Prediction of functional characteristics of ecosystems: A comparison of artificial neural networks and regression models. *Ecological Modelling*, 98, 173–186.
- Pasquier, H., & Verheyden, A. (1998). The Vegetation Programming Center, a new subsystem in the Spot4 ground segment. In *DASIA 98-Data Systems in Aerospace* (p. 367). volume 422.
- Ranghetti, L., Busetto, L., Crema, A., Fasola, M., Cardarelli, E., & Boschetti, M. (2016). Testing estimation of water surface in italian rice district from MODIS satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 52, 284 – 295.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. New York: The MIT Press.

- Ross, J. (2012). *The radiation regime and architecture of plant stands* volume 3. Springer Science & Business Media.
- Roujean, J.-L. (1996). A tractable physical model of shortwave radiation interception by vegetative canopies. *Journal of Geophysical Research: Atmospheres*, 101, 9523–9532.
- Roujean, J.-L., & Breon, F.-M. (1995). Estimating {PAR} absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51, 375 – 384.
- Roujean, J.-L., & 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, ACL 6–1–ACL 6–14.
- Roujean, J.-L., Leroy, M., & Deschamps, P.-Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research: Atmospheres*, 97, 20455–20468.
- 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., Bindshadler, R., Cohen, W., Gao, F., Hipple, J., Hostert, P., Huntington, J., Justice, C., Kilic, A., Kovalsky, V., Lee, Z., Lymburner, L., Masek, J., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R., & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154 – 172.
- Scholkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.
- Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.
- Snieder, R., & Trampert, J. (1999). Inverse problems in geophysics. In A. Wirgin (Ed.), *Wavefield Inversion* (pp. 119–190). Vienna: Springer Vienna.
- Sterckx, S., Benhadj, I., Duhoux, G., Livens, S., Dierckx, W., Goor, E., Adriaensen, S., Heyns, W., Van Hoof, K., Strackx, G. et al. (2014). The PROBA-V mission: Image processing and calibration. *International Journal of Remote Sensing*, 35, 2565–2588.
- Stokes, G. G. (1860). On the intensity of the light reflected from or transmitted through a pile of plates. *Proceedings of the Royal Society of London*, 11, 545–556.
- Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2016). A note on the temporary misregistration of landsat-8 operational land imager (oli) and sentinel-2 multi spectral instrument (msi) imagery. *Remote Sensing of Environment*, 186, 121 – 122.
- Suits, G. H. (1971–1973). The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment*, 2, 117 – 125.
- Svendsen, D. H., Martino, L., Campos-Taberner, M., García-Haro, J., & Camps-Valls, G. (2017). Joint gaussian processes for biophysical parameter retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, . Submitted.

- Townshend, J. R. (1994). Global data sets for land applications from the Advanced Very High Resolution Radiometer: an introduction. *International Journal of Remote Sensing*, 15, 3319–3332.
- Trigo, I. F., Dacamara, C. C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., de Coca, F. C., Carrer, D., Freitas, S. C., García-Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., & Arboleda, A. (2011). The satellite application facility for land surface analysis. *International Journal of Remote Sensing*, 32, 2725–2744.
- Verger, A., Martínez, B., Camacho, F., & García-Haro, F. J. (2009). Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. *International Journal of Remote Sensing*, 30, 3849–3849.
- Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16, 125–1141.
- Verhoef, W. (1985). Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment*, 17, 165–1178.
- Verhoef, W., & Bach, H. (2003). Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of Environment*, 87, 23–41.
- Verhoef, W., & Bach, H. (2007). Coupled soil–leaf–canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and {TOA} radiance data. *Remote Sensing of Environment*, 109, 166–1182.
- Verhoef, W., Jia, L., Xiao, Q., & Su, Z. (2007). Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. *Geoscience and Remote Sensing, IEEE Transactions on*, 45, 1808–1822.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, 118, 127 – 139.
- Verrelst, J., Rivera, J. P., Moreno, J., & 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.
- Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 260 – 272.
- Wang, Y., Tian, Y., Zhang, Y., El-Saleous, N., Knyazikhin, Y., Vermote, E., & Myneni, R. B. (2001). Investigation of product accuracy as a function of input and model uncertainties: Case study with seawifs and modis lai/fpar algorithm. *Remote Sensing of Environment*, 78, 299–313.
- Warren-Wilson, J. (1963). Estimation of foliage denseness and foliage angle by inclined point quadrats. *Australian Journal of Botany*, 11, 95–105.

- Weiss, M., & Baret, F. (2010). *CAN-EYE V6.313 User Manual*. INRA.
- Weiss, M., Baret, F., Smith, G., Jonckheere, I., & Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology*, 121, 37 – 53.
- Weng, Y.-H., Sun, F.-S., & Grigsby, J. D. (2012). Geotools: An android phone application in geology. *Computers & Geosciences*, 44, 24–30.
- White, M. A., Asner, G. P., Nemani, R. R., Privette, J. L., & Running, S. W. (2000). Measuring fractional cover and leaf area index in arid ecosystems: digital camera, radiation transmittance, and laser altimetry methods. *Remote Sensing of Environment*, 74, 45–57.
- Williams, C. K. I., & Rasmussen, C. E. (1996). Gaussian processes for regression. In *Neural Information Processing Systems, NIPS 8* (pp. 598–604). MIT Press.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Yan, K., Park, T., Yan, G., Chen, C., Yang, B., Liu, Z., Nemani, R. R., Knyazikhin, Y., & Myneni, R. B. (2016). Evaluation of MODIS LAI/FPAR product collection 6. Part 1: Consistency and improvements. *Remote Sensing*, 8, 359.
- Yang, W., Shabanov, N., Huang, D., Wang, W., Dickinson, R., Nemani, R., Knyazikhin, Y., & Myneni, R. (2006). Analysis of leaf area index products from combination of modis terra and aqua data. *Remote Sensing of Environment*, 104, 297–312.
- Zhang, X., Lu, C., Hu, N., Yao, K., Zhang, Q., & Dai, Q. (2013). Simulation of canopy leaf inclination angle in rice. *Rice Science*, 20, 434 – 441.