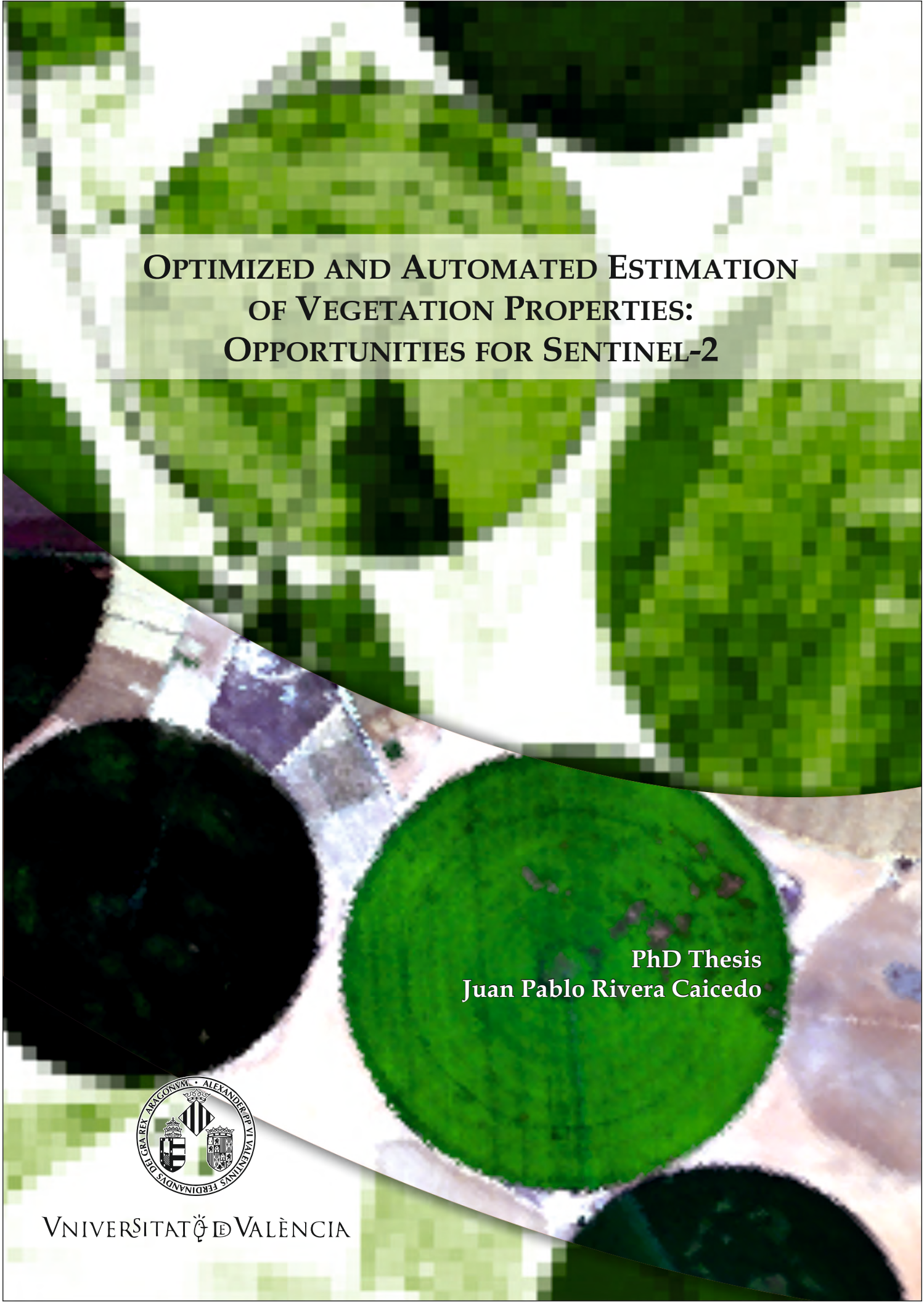
Optimized and Automated Estimation of Vegetation Properties: Opportunities for Sentinel-2



PhD Thesis
Juan Pablo Rivera Caicedo

VNIVERSITAT DE VALÈNCIA

OPTIMIZED AND AUTOMATED ESTIMATION OF VEGETATION PROPERTIES: OPPORTUNITIES FOR SENTINEL-2

By

Juan Pablo Rivera Caicedo

PHD THESIS · TESIS DOCTORAL
Programa de Doctorado en Teledección

THESIS ADVISORS · DIRECTORES DE TESIS
Jochem Verrelst
José F. Moreno

Laboratorio de Procesado de Imágenes
Parque Científico
UNIVERSITAT DE VALÈNCIA - ESTUDI GENERAL
València, Abril 2014

Optimized and automated estimation of vegetation properties: Opportunities for Sentinel-2

Juan Pablo Rivera Caicedo, 2014.

Laboratorio de Procesado de Imágenes
Parque Científico · Universitat de València

D. JOCHEM VERRELST, Doctor en 'Remote Sensing', investigador post-doc en el Laboratorio de Procesado de Imágenes, Parque Científico de la Universitat de València y

D. JOSÉ F. MORENO, Doctor en Ciencias Físicas, Profesor Titular del Departamento de Física de la Tierra y Termodinámica de la Facultad de Física de la Universitat de València

HACEN CONSTAR QUE:

El Ingeniero Agrícola Juan Pablo Rivera Caicedo ha realizado bajo nuestra dirección el trabajo titulado "Optimized and automated estimation of vegetation properties: Opportunities for Sentinel-2", que se presenta en esta memoria para optar al grado de Doctor en Teledetección.

Y para que así conste a los efectos oportunos, firmamos el presente certificado, en Valencia, a 16 de Abril de 2014.

Jochem Verrelst José F. Moreno

Dijo Dios: "Haya luz", y hubo luz.
Génesis 1:3

A lo largo de estos 6 años he comprendido
que escuchas los anhelos de mi corazón
Y en su tiempo, provees todo para que se cumplan.

Gracias por esa luz, que me ha permitido aprender sobre su
interacción con las cubiertas vegetales.

Agradecimientos

Quiero agradecer a mis directores…

Jochem Verrelst por su apoyo y confianza, por la oportunidad que he tenido de trabajar a su lado, de aprender y formar un gran equipo. Esta tesis ha sido posible gracias a Ti. Por todos los momentos que compartí junto a Agata y tu familia.

José Moreno por darme la oportunidad de ver cumplido un anhelo y mostrarme una vida profesional dedicada a la enseñanza y la investigación que deseo seguir. La admiración que siento por ti y el orgullo de trabajar a tu lado crecen siempre.

Al grupo del Laboratorio de Observación de la Tierra (LEO).

Jesús Delegido por su gran apoyo, dedicación y su gran empatía con el país de las tres cordilleras. Gracias por tu familia y todo el cariño que recibí de ustedes.

Luis Alonso por su carácter perfeccionista y su valiosa experiencia. No solo aprendí de lo que me dijiste, sino también de lo que enseñaste a otros.

Jorge Vicent, Neus Sabater y Carolina Tenjo, por poner esa gran chispa que convirtió al grupo LEO en mi familia.

Un especial agradecimiento a Gustavo Camps-Valls y Jordi Muñoz por su gran disposición y ayuda en el desarrollo de ARTMO, por abrirme las puertas del MatLab. Ha sido un orgullo trabajar con ustedes. A Ganna Leonenko por enseñarnos 62 maneras diferentes de comparar espectros y tus palabras de ánimo. A Frank Veroustraete, gracias por tus consejos en el mundo de las relaciones espectrales.

A los profesores del Departamento de Termodinámica y Física de la Tierra, en especial a Soledad Gandía por esa alegría y sonrisa en su rostro, y Javier García Haro por su enseñanza y palabras de ánimo.

A Josep Pardo Pascual y a Luís Ángel Ruiz Fernández y todos los profesores del Grupo de Investigación de Cartografía GeoAmbiental y Teledetección de la Universidad Politécnica de Valencia, por su confianza en mis primeros años de formación. En especial a los profesores Rafael García Bartual y Manuel Pulido por su voto de confianza y permitirme hacer parte del Instituto de Ingeniería del Agua y Medio Ambiente.

Creo que mi mejor manera de agradecerles, es continuar esforzándome en aprender cada día más y continuar el trabajo en equipo que hemos formado sin importar el país donde esté.

A mi casa, mi mamá, mi mamita, Juanda, Toto, Lina y Juanes. Porque son la luz que ha puesto Dios para no desmayar, no perder el rumbo y darle sentido a todo lo que estoy realizando.

A mi preciosa Charito, Rosario, tus palabras de ánimo, tu oración y tu fe me han dado aliento y confianza que se podría lograr. Gracias a mi Familia mexicana, Don Abelino, Doña Dora y tus hermanas y sus familia que nos han apoyado y nos apoyarán en lo que viene.

A mi casa en España, a Yoli e Isa por su amor y aliento en los momentos de soledad.

A don Vicente Hernández V. y sus hijas y familia.

A mis amigos en Valencia que son una bendición de Dios para mi vida.

Juan Pablo Rivera Caicedo

Abstract

Optical Earth observation satellites monitor our planet by acquiring images at different wavelengths of the electromagnetic spectrum. ESA's forthcoming Sentinel-2 Multispectral Instrument foresees to provide continuity to land monitoring services by relying on optical payload with visible, near infrared and shortwave infrared sensors with high spatial and temporal resolution. This unprecedented data availability requires processing techniques, which are reproducible, accurate and fast, when the retrieval of information on plant growth and health status is envisaged. However, optical sensors are inherently incapable to provide directly vegetation properties. Space-based estimation of vegetation properties always requires an intermediate modeling step to transform the measured radiation fluxes reflected from the Earth's surface into biophysical variables. This modeling step can be approached with either statistical (parametric or nonparametric), physical or hybrid retrieval methods.

This Thesis brought together latest retrieval methods presented in the field of vegetation remote sensing. It is admitted that a systematic assessment of these methods in view of the forthcoming Sentinel-2 mission will eventually lead to improved monitoring of vegetation properties. The main objective, therefore, was: *'To analyze, optimize and automate state-of-the-art vegetation properties mapping methods in preparation of forthcoming Sentinel-2 mission'*. The pursued path was to generate both scientific outputs as well scientific software that automates the retrieval routines. A GUI software package called ARTMO (Automated Radiative Transfer Models Operator) has been developed, which consists of a suite of radiative transfer models and retrieval routines in a modular design. ARTMO has been applied to the estimation of leaf area index (LAI) and leaf chlorophyll content (LCC) from simulated Sentinel-2 data, but the majority of investigated methods can essentially be applied to derive any detectable surface variable from any optical airborne or satellite sensor.

In the peer-reviewed Chapters 3 – 6, three different retrieval toolboxes have been presented that assess and apply the main retrieval domains: i.e. (Ch. 3) *parametric regression*, (Ch. 4) *physically-based inversion routines*, and (Ch. 5) *nonparametric regression*. The best performing method has been subsequently applied in a more operational framework (Ch. 6).

Based on the peer-reviewed chapters, this Thesis delivered the following results:

1. A systematic assessment of all possible two-band spectral indices and parametric fitting functions using HyMap data (439 2490 nm). The most sensitive regions have been identified for two-band combinations of green (539-570 nm) with longwave SWIR (2421-2453 nm) for LAI (r^2: 0.83) and far-red (692 nm) with NIR (1340 nm) or shortwave SWIR (1661-1686

nm) for LCC (r^2: 0.93).

2. A systematic assessment of multiple cost functions and regularization options in LUT-based inversion routines using simulated Sentinel-2 data. Introducing noise and opting for the mean of multiple best solutions in the inversion considerably improved retrievals; relative errors can be halved as opposed to without these regularization options. Best LCC retrievals were obtained using a normalized 'L_1-estimate' function with a r^2 of 0.73, while best LAI retrievals were obtained through non-normalized 'least-squares estimator' (LSE) with a r^2 of 0.74.

3. A systematic assessment of a wide range of nonparametric regression algorithms using amongst others simulated Sentinel-2 data. Overall, nonlinear regression algorithms (NN, KRR, GPR) outperformed linear algorithms (PCR, PLSR, DT) in terms of accuracy, bias, and robustness. Most robust results for LAI and LCC retrieval along gradients of training/validation partitioning and noise variance were obtained by KRR, while GPR delivered most accurate estimations with a r^2 of 0.94–0.99.

4. Locally-trained GPR models with extended training dataset reached the 10% precision required by end users, with for LCC a with a r^2 of 0.95-0.99 and for LAI a with a r^2: 0.95–0.96. The developed GPR models were subsequently applied to simulated Sentinel-2 images over various sites across the world. For agricultural sites the associated uncertainty maps were on the same order as the local Barrax test site.

The systematic assessment across the three domains allowed synthesizing the diverse field of biophysical parameter retrieval. The nonlinear nonparametric approaches (i.e. machine learning regression algorithms; MLRAs) led to highest accuracies, while LUT-based inversion routines performed on the whole poorest. The spectral indices approaches led to intermediate results. Especially the kernel-based MLRAs emerged as powerful algorithms. They typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input-output nonlinear mappings. Among the most promising kernel-based MLRAs is Gaussian Processes regression (GPR). Apart from robust retrievals, this Bayesian regression algorithm also provides insight in relevant bands during model development and delivers associated uncertainty estimates. These uncertainties proved to be particularly valuable when transporting the developed model to other sites and conditions. It was demonstrated that for the majority of processed image uncertainties were on the same order as on the local Barrax image. Thus, uncertainty estimates can serve as a convenient quality check when processing images in space and time.

All results were generated with the ARTMO toolbox, which has been made freely available to the remote sensing community. A broader use of ARTMO is foreseen to underpin: (1) an improved understanding in the interactions between light and vegetation properties, (2) improved vegetation properties retrieval algorithms serving forthcoming optical missions, (3) new mapping applications in order to understand better our changing Earth.

Contents

1 Introduction

Contents

1.1 Remote sensing for monitoring a changing Earth

The biosphere is one of the main components of the Earth's system since it regulates exchanges of energy and mass fluxes at the soil, vegetation and atmosphere level. To know the links between vegetation and the terrestrial energy, water and carbon cycles, and how these might change due to eco-physiological responses to elevated CO_2 and changes in land use is of vital importance for the study of the biosphere [IPCC, 2007]. To study these exchanges, several kinds of models (scale and target) have been developed. A model is usually a small version of something larger. At global scale, the main models are General Circulation Models, which are numerical models that represent physical processes in the atmosphere, ocean, cryosphere and land surface are the most advanced tools currently available for simulating the response of the global climate system [Donner *et al.*, 2011]. The carbon cycle model is a representation of the movement of carbon from sources to sinks through chemical and physical transfers [Jungclaus *et al.*, 2010].

At the terrestrial regional-to-global scale, the main models are Dynamic Global Vegetation Models or DGVMs, which are virtual computer models that simulate spatio-temporal changes

in potential vegetation types and their associated biogeochemical and hydrological cycles as a response to changing climate and soil conditions [Malenovský, 2013]. DGVMs use as inputs time series of climate data, topography and soil characteristics and simulate monthly or daily dynamics of ecosystem processes. The models generally combine biogeochemistry, biogeography, and disturbance sub-models. To simplify complex vegetation ecological societies, DGVMs are using so-called plant functional types (PFT) providing a direct link to the leaf-level physiological processes. Each PFT is characterized by specific plant functional traits, e.g. leaf area index (LAI), specific leaf area, potential evapotranspiration, etc. [Malenovský, 2013].

In view of these models, the Global Climate Observing System (GCOS) aims to provide comprehensive information on the total climate system, involving a multidisciplinary range of physical, chemical and biological properties, and atmospheric, oceanic, hydrological, cryospheric and terrestrial processes. Fifty GCOS Essential Climate Variables (ECVs) are required to support the work of the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC) [GCOS, 2014]. Table 1.1 shows the most important ECVs per domain. In this respect, optical remote sensing (RS) provides powerful methods for the estimation of ECVs [Baret *et al.*, 2013; Hollmann *et al.*, 2013].

TABLE 1.1: The 50 GCOS Essential Climate Variables (ECVs) (2010) are required to support the work of the UNFCCC and the IPCC. All ECVs are technically and economically feasible for systematic observation [GCOS, 2014].

Domain	**GCOS Essential Climate Variables**	
Atmospheric	Surface:	Air temperature, Wind speed and direction, Water vapour, Pressure, Precipitation, Surface radiation budget
	Upper-air:	Temperature, Wind speed and direction, Water vapour, Cloud properties, Earth radiation budget (including solar irradiance).
	Composition:	Carbon dioxide, Methane, and other long-lived greenhouse gases, Ozone and Aerosol, supported by their precursors.
Oceanic	Surface:	Sea-surface temperature, Sea-surface salinity, Sea level, Sea state, Sea ice, Surface current, Ocean colour, Carbon dioxide partial pressure, Ocean acidity, Phytoplankton.
	Subsurface:	Temperature, Salinity, Current, Nutrients, Carbon dioxide partial pressure, Ocean acidity, Oxygen, Tracers.
Terrestrial		River discharge, Water use, Groundwater, Lakes, Snow cover, Glaciers and ice caps, Ice sheets, Permafrost, Albedo, Land cover (including vegetation type), Fraction of absorbed photosynthetically active radiation (FAPAR), **Leaf area index (LAI)**, Above-ground biomass, Soil carbon, Fire disturbance, Soil moisture.

In support of these terrestrial models, but also in support of monitoring local-to-global vegetation dynamics, this Thesis focuses on improved estimation of vegetation properties from optical RS data, and more specifically leaf area index (LAI) and leaf chlorophyll content (LCC). Although LCC is currently not considered as an ECV due to the lack of a globally applicable

retrieval algorithm, it is a key variable in vegetation studies [Lichtenthaler, 1987]. Monitoring the distribution and changes of LAI and LCC is important for assessing growth and vigour of vegetation on the planet [Running and Coughlan, 1988]. The quantification of these essential vegetation properties are fundamentally important in land-atmosphere processes and parametrization in climate models [Dorigo *et al.*, 2007]. LAI variable represents the amount of leaf material in ecosystems and controls the links between biosphere and atmosphere through various processes such as photosynthesis, respiration, transpiration and rain interception [Jonckheere *et al.*, 2004]. LCC provides important information about the physiological status of plants and photosynthetic activity, therefore is related to the nitrogen content, water stress and yield forecasting [Gitelson *et al.*, 2005; Schlemmer *et al.*, 2005; Zhang *et al.*, 2008; Peng and Gitelson, 2012]. Optical RS data are able to quantify LAI and LCC at a per-pixel basis, and the processing of a complete image provides insight in their spatial patterns over large areas [e.g., Delegido *et al.*, 2011; Verrelst *et al.*, 2012a].

Two key criteria need to be fulfilled in order to achieve monitoring of vegetation properties across the globe: (1) *accurate optical RS data processing*, and (2) *accurate optical RS data availability*. Accurate data processing is mandatory because an optical sensor only measures the spatially distributed radiation fluxes reflected from the Earth's surface in the direction of the sensor. Optical RS measurements are thus inherently incapable to provide directly vegetation properties. An intermediate step is necessary to transform the RS measurements into estimates of vegetation properties. Fundamentally, the interpretation of RS data always implies the use of a model. This model can be *physical-based* or can be *statistically-based*, or can be a combination of both. During the history of optical RS a panoply of retrieval methods have been proposed by the scientific literature from low to high complexity, and new ones continue to be developed. A diverse variety of them have been brought together in this Thesis.

The second criterion involves *accurate data availability*. The delivery of accurate optical RS data that enable timely and global monitoring vegetation properties on a regular basis, at a high spatial resolution and with high accuracy has always been a major challenge by space agencies. Various missions dedicated to land monitoring have been launched for the last few decades, but only recently optical Earth observation is reaching a mature state with dedicated missions that fulfill these requirements. The European Space Agency (ESA)'s forthcoming Sentinel-2 mission is particularly tailored to the monitoring vegetation properties mapping, with operational monitoring capabilities that goes beyond any existing operational mission. A pair of Sentinel-2 polar-orbiting satellites will provide systematic global acquisitions of high-resolution multispectral imagery (10-60 m) with a high revisit frequency on a free and open data policy basis. With the pair of satellites in operation it has a revisit time of five days at the equator (under cloud-free conditions) and 2–3 days at mid-latitudes [Drusch *et al.*, 2012]. Sentinel-2 images will be used to derive the highly prioritized time series of ECVs such as LAI. Sentinel-2 images will also be used provide various experimental variables, e.g. biochemical variables such as LCC. In preparation of the Sentinel-2 launch an international Sentinels for Science (SEN4SCI) Workshop was organized by ESA in March 2011. The output of the workshop is a list of recommendations that lead to a more comprehensive exploitation of the Sentinel-2

observational capabilities [Malenovsky *et al.*, 2012]. One of the identified top recommendations concerned the following:

> "The bio-physical/-chemical variables of vegetation such as FAPAR and LAI, vegetation canopy cover, leaf chlorophyll/water/dry matter content, and other characteristics describing the canopy structure (e.g., foliage clumping and leaf angle distribution function) should be retrieved jointly using the physically based inversion routines whenever feasible. An evolutionary prototyping approach should be included in their operational retrieval."

This Thesis is dedicated to tackle the stated recommendation and turn it into consolidated guidelines. The undertaken road map was to work on both generating scientific outputs, as well on developing software to automate the retrieval routines. All essential tools to deliver a prototype retrieval approach that could be embedded into an operational Sentinel-2 processing scheme have been prepared into a scientific software package called ARTMO (Automated Radiative Transfer Models Operator). Physically-based approaches but also latest statistically-based methods have been implemented into the software package and systematically evaluated. The retrieval methods have been applied to the estimation of LAI and LCC from simulated Sentinel-2 data, but the majority of investigated methods can essentially be applied to derive any detectable vegetation biochemical or biophysical variable. The fundamentals of ARTMO has been laid during J.P. Rivera's MSc thesis project and has been further developed during the course of my PhD Thesis. The toolbox is built on a suite of radiative transfer models and image processing modules in a modular graphical user interface (GUI) environment. ARTMO has been mainly developed and tested for processing (simulated) Sentinel-2 data in a semiautomatic way, but in principle data from any optical sensor can be processed.

1.2 Objectives and research questions

This Thesis brings together the latest emerging retrieval methods presented in the field of vegetation remote sensing. A synthesis of these methods in view of the performance of forthcoming Sentinel-2 mission will eventually lead to improved monitoring of vegetation properties. The main objective, therefore, is: *'To analyze, optimize and automate state-of-the-art vegetation properties mapping methods in preparation of forthcoming Sentinel-2 mission'*. To reach this objective various specialized retrieval toolboxes have been developed within the ARTMO environment. These toolboxes enabled the systematic assessment of parametric, nonparametric and physically-based retrieval methods. Although in principle the toolboxes can be applied to estimate any surface variables at any location, here they were applied to the estimation of LCC and LAI over the agricultural site Barrax, Spain. The major challenges are addressed by the following research questions:

1. *What is the performance of all possible two-band vegetation indices to enable optimized LCC and LAI estimation?*
2. *What is the performance of physically-based inversion routines given a suite of cost functions and regularization options to enable optimized LCC and LAI estimation?*
3. *What is the performance of a range of nonparametric regression methods and regularization options to enable optimized LCC and LAI estimation?*
4. *Can the best evaluated retrieval method be applied to other sites and conditions in view of operational retrieval of LCC and LAI from forthcoming Sentinel-2 data?*

1.3 Thesis outline

This Thesis continues with a general background chapter and then presents four thematic chapters which are based on peer-reviewed papers (Chapters 3-6). Each of the thematic chapters represent an answer to the research questions presented in section 1.2. Chapter 3 assesses parametric regression approaches, Chapter 4 assesses physically-based inversion routines and Chapter 5 assesses nonparametric regression approaches for the benefit of optimized vegetation properties mapping. Based on the most promising assessed mapping method, Chapter 6 proposes and evaluates a more generally applicable mapping strategy that eventually could be implemented into an operational Sentinel-2 processing chain.

- CHAPTER 2 presents a general background regarding Earth observation of vegetation, addresses state-of-the-art retrieval methods (parametric, nonparametric and physically-based), reviews the diversity of leaf and canopy radiative transfer models and outlines the ARTMO software framework that enables automating and optimizing these retrieval methods.
- CHAPTER 3 [Rivera *et al.*, 2014a] presents ARTMO's 'Spectral Indices' toolbox. With this toolbox all possible two-band indices have been assessed to reach optimized LCC and LAI mapping based on an experimental field dataset and a hyperspectral HyMap imagery over the agricultural site Barrax (Spain).
- CHAPTER 4 [Rivera *et al.*, 2013a] presents ARTMO's 'Inversion' toolbox. In ARTMO the leaf optical model PROSPECT has been coupled with the canopy reflectance model SAIL to generate an extensive LUT. The Inversion toolbox was used to assess the role of cost functions and regularization options. Inversion strategies have been optimized for LAI and LCC retrieval against the same experimental field dataset using simulated Sentinel-2 data.
- CHAPTER 5 [Rivera *et al.*, 2013b] presents ARTMO's 'Machine Learning Regression Algorithms' (MLRA) toolbox. With the MLRA toolbox multiple nonparametric regression algorithms have been assessed to reach optimized LAI and LCC mapping based on the same experimental field dataset. Apart from simulated Sentinel-2 data, also the performances for hyperspectral CHRIS and HyMap data have been assessed.
- CHAPTER 6 [Verrelst *et al.*, 2013b] presents a mapping application based on the best evaluated retrieval algorithm and simulated Sentinel-2 images. It moves away from the local Barrax site and evaluates the mapping performance to other sites and conditions by making use of associated uncertainty estimates. This chapter demonstrates that new type of retrieval approaches holds promises for improved operational mapping and thus refined monitoring of vegetation properties.
- CHAPTER 7 summarizes the accomplished objectives and the main scientific achievements, discusses the main conclusions, and proposes future research lines.

2

Background

Contents

2.1 Remote sensing of vegetation

Since the successful launch and deployment of the first experimental satellite, Sputnik, in 1957, satellites have been used for applications such as surveillance, navigation, communication and Earth observation. Notable applications of remote sensing include those relating to meteorology, agriculture, mining, geology, mapping, ecological monitoring and disaster monitoring. Although primarily electro-optical visible sensors have been used, more recently, the application of thermal imagers, synthetic aperture radar (SAR), light detection and ranging (LIDAR), and hyperspectral imagers has gained increasing attention [Teke *et al.*, 2013].

Sensors can be divided into two broad groups: passive and active. Passive sensors measure ambient levels of existing sources of energy, while active ones provide their own source of energy. The majority of remote sensing is done with passive sensors, for which the sun is the main energy source. The sun radiates energy through space via electromagnetic waves. The

entire region of electromagnetic energy distinguished by wavelength and frequency is called the electromagnetic spectrum (Figure 2.1).

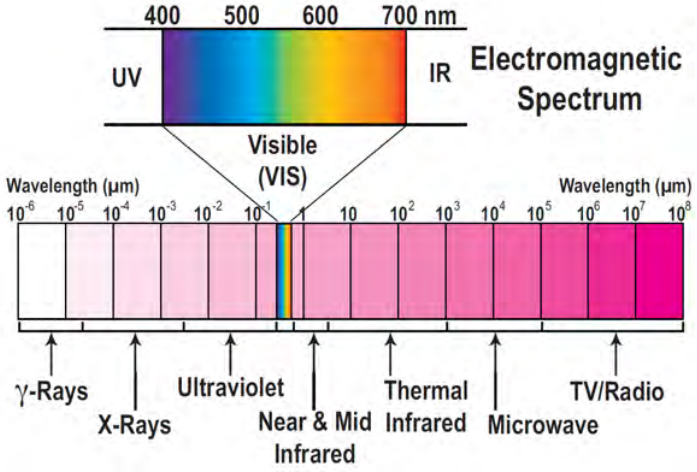


FIGURE 2.1: Electromagnetic spectrum classification based on wavelength range. Note the logarithmic scale.

Interactions between incident radiation and vegetation elements are extremely complex and are described by three main physical mechanisms: absorption, reflectance and transmission. The solar reflected radiation in the optical domain is commonly used in vegetation studies, because most of the diagnostic absorption features of green vegetation are located in this part of the spectrum. Reflectance of vegetation canopies depends on radiative properties of leaves, other non-photosynthetic canopy elements, their spatial organization, and soil background. When inspecting a typical reflectance spectrum of a vegetation canopy in more detail, it can be subdivided into 4 parts: (1) *visible* (400 – 700 nm), (2) *near infrared* NIR (800 – 1300 nm) and, (3) *shortwave-infrared* SWIR (1300 – 2500 nm), as is demonstrated in Figure 2.2.

The *visible* part of the reflectance spectrum (400 – 700 nm) of vegetation is controlled by the pigments in the green leaf chloroplasts that reside in the outer or palisade leaf, the chlorophyll pigments – chlorophyll-a and chlorophyll-b. Chlorophyll is the major absorber of radiation in the visible region and its absorption is dominant in the visible red until red-edge wavelengths (600 – 720 nm); it is called the green pigment and it is common to all photosynthetic cells. Other leaf pigments also have an important impact on the visible part of the spectrum. The carotene (yellow to orange-red pigment responsible for the colour of some flowers, fruits and leaves without chlorophyll) and xantophyll (responsible for the leaf colour in autumn) have strong absorption in the 350 – 500 nm, blue wavelengths.

The *red edge* is a region in the red-NIR transition zone of vegetation reflectance spectrum and marks the boundary between absorption by chlorophyll in the red visible region, and scattering due to leaf internal structure in the NIR region. This transition zone is in the basis of several vegetation indices like NDVI which is the normalized difference between the reflectance in the red visible (700 nm) and the NIR (800 nm) reflectance. Also the red-edge position is used to estimate the chlorophyll content of leaves or over a canopy.

The optical properties in the *near infrared* spectral domain (800 – 1300 nm) are explained by leaf structure. The spongy mesophyll cells located in the interior or back of the leaves reflects NIR light, much of which emerges as strong reflection rays. The intensity of NIR reflectance is commonly greater than most inorganic materials, so vegetation appears bright in NIR wavelengths.

The *shortwave-infrared* region (1300 – 2500 nm) contains information about the absorption of radiation by water, cellulose and lignin and several other biochemical constituents. This region of the vegetation spectrum allows the identification of vegetation stress due to drought.

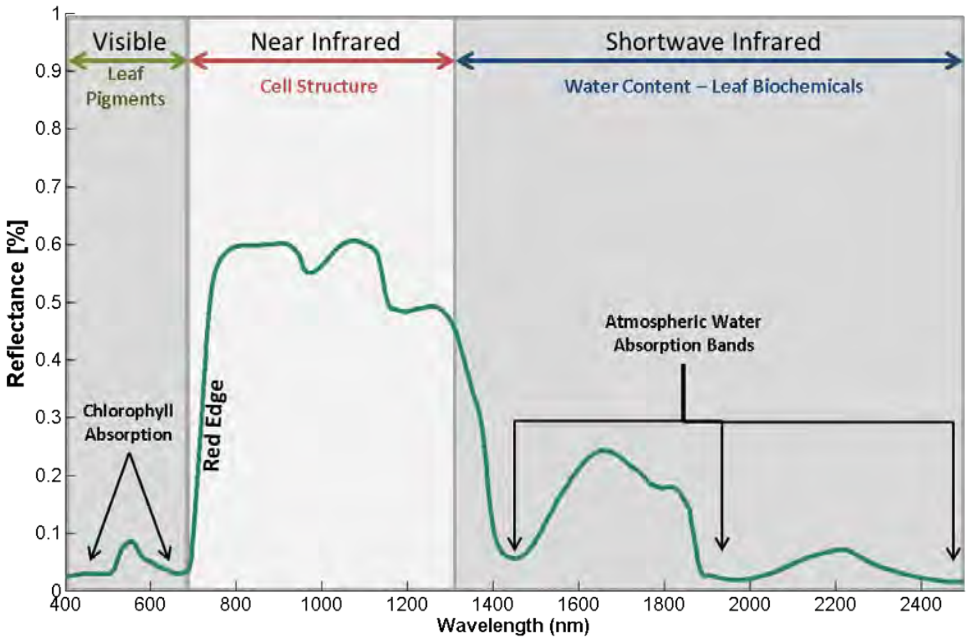


FIGURE 2.2: Electromagnetic spectrum classification based on wavelength range.

2.2 Copernicus space programme

While a few dozen of optical Earth observation missions are currently orbiting around the globe, only a few of them are dedicated to the detection of vegetation reflectance, and even less provide and process recorded data on an operational basis. To fill up this gap, the European Union has recently set up a multi-billion-euro Earth observation project named as Copernicus to monitor the planet.

Copernicus, previously known as GMES (Global Monitoring for Environment and Security), is the most ambitious operational Earth Observation programme to date and will provide global, timely and easily accessible information in application domains such as land, marine, atmosphere, emergency response, climate change and security. Copernicus consists of a complex set of systems which collect data from multiple sources: Earth observation satellites and

in situ sensors such as ground stations, airborne and sea-borne sensors. It processes these data and provides users with reliable and up-to-date information through a set of services related to environmental and security issues [Aschbacher and Milagro-Pérez, 2012]. Copernicus services are based on the processing of environmental data collected from two main sources: (1) a space component, which consists of a constellation of satellites known as the Sentinels, coordinated by European Space Agency (ESA); and, (2) an in situ component, which consist of a multitude of sensors on the ground, at sea or in the air coordinated by European Environment Agency (EEA). A list of the Sentinels is given below.

- The Sentinel-1 constellation is a pair of C-band synthetic aperture radar (SAR) imaging satellites, the first to be launched in April 2014. The SAR sensor will be operated in two main modes: 'Interferometric Wide Swath' and 'Wave', the first having 250 km of swath width and 5*20 m ground resolution. Sentinel-1 data will support applications covering ice/ocean observations, land monitoring/management, hydrology, disaster management, oil spill monitoring, ship detection for maritime security, etc. With two satellites it will be possible to have interferometric image pairs every 6 days, and a coverage every 1–3 days of areas like Europe, Canada and main Northern shipping routess [Torres *et al.*, 2012].
- Sentinel-2 is a polar-orbiting, multispectral high-resolution imaging mission for land monitoring providing, for example, imagery of vegetation, soil and water cover, inland waterways and coastal areas. Sentinel-2 will also deliver information for emergency services [Drusch *et al.*, 2012].
- Sentinel-3 is polar-orbiting, multi-instrument mission to measure variables such as sea-surface topography, sea- and land-surface temperature, ocean colour and land colour with high-end accuracy and reliability [Donlon *et al.*, 2012].
- Sentinel-4 is a payload that will be embarked upon a Meteosat Third Generation-Sounder (MTG-S) satellite in geostationary orbit. Sentinel-4 is dedicated to atmospheric monitoring.
- Sentinel-5 is a payload that will be embarked on a MetOp Second Generation satellite, also known as Post-EPS. Sentinel-5 is dedicated to atmospheric monitoring.
- Sentinel-5 Precursor is a low Earth orbit polar satellite to provide information and services on air quality, climate and the ozone layer in the timeframe 2015–2022. The payload of the mission is the TROPOspheric Monitoring Instrument (TROPOMI) that will measure key atmospheric constituents including ozone, NO_2, SO_2, CO, CH_4, CH_4O and aerosol properties [Veefkind *et al.*, 2012].

2.2.1 Sentinel-2

The Sentinel-2 mission is dedicated to land surface monitoring and therefore most of interest within this Thesis. The key mission objectives for Sentinel-2 are: To provide systematic

global acquisitions of high-resolution multi-spectral imagery with a high revisit frequency, to provide enhanced continuity of multi-spectral imagery provided by the SPOT (Satellite Pour l'Observation de la Terre) series of satellites, and to provide observations for the next generation of operational products such as land-cover maps, land change detection maps, and geobiophysical variables [Drusch *et al.*, 2012].

Sentinel-2 is committed to the full and systematic coverage of land surfaces (including major islands) from 56°S (southern Americas) to 83°N (northern Greenland), providing cloud-free products every 15–30 days. By comparison, the US Landsat-7 and -8 has 16-day revisits and Spot 26-day revisits, and neither provides systematic coverage of land. In order to support operational services for at least 15 years from the launch of the first satellites, a series of four satellites is planned, with two operating in orbit and a third in ground storage as backup. Frequent revisits and high mission availability require two Sentinels operating simultaneously, which dictates a small, cost-effective and low-risk satellite. The orbit is Sun-synchronous at 786 km altitude. The satellite is designed for a 7-year lifetime, with propellant for 12 years of operations, including deorbiting at the end [Drusch *et al.*, 2012].

Sentinel-2's Multi-Spectral Instrument (MSI) features 13 spectral bands from the visible and near-infrared (VNIR) to the short-wave infrared (SWIR), featuring four at 10 m, six at 20 m and three at 60 m resolution (Figure 2.3). The best compromise in terms of user requirements and mission performance, cost and schedule risk, it provides enhanced continuity for Spot and Landsat, with narrower bands for improving identification of features, additional red channels for assessing vegetation, and dedicated bands for improving atmospheric correction and detecting cirrus clouds.

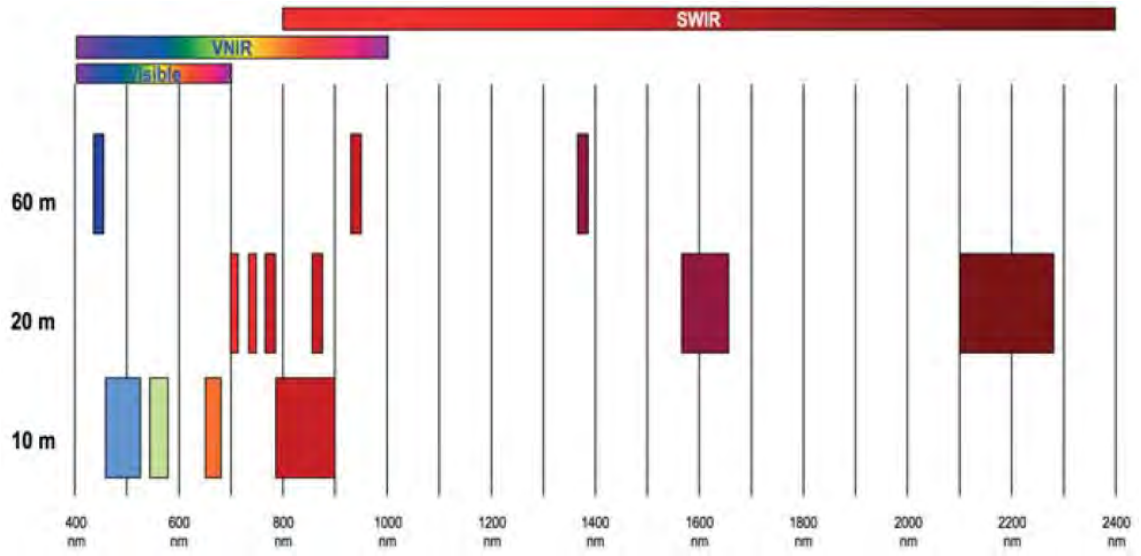


FIGURE 2.3: Sentinel-2 spectral features in relation to spatial resolution (after Drusch *et al.* [2012]).

2.3 Principles of biophysical parameter retrieval methods

While simulated Sentinel-2 data will be the main data source used in this Thesis, vegetation properties mapping can in principle be achieved from any optical sensor. The quantification of surface geobiophysical parameters from optical remote sensing always relies on a model, either *statistically* or *physically* based. This model enables to interpret a spectral observations and to translate into one or multiple surface variables, e.g. diagnostic properties of vegetation health. Using statistical terminology, this is a *regression* problem [Fernandes and Leblanc, 2005]. Over the last decade both fields have greatly expanded and diverged, particularly the statistical approaches, and increasingly elements of both fields have been combined, leading to the so-called *hybrid* approaches. Therefore, there is a need for a more systematic and refined categorization. In all generality, retrieval methods can be subdivided in either one of these 4 categories:

- *Parametric regression methods*: Parametric methods assume an explicit relation between spectral observation and the biophysical parameter. Hence, explicit parametrized expressions are build, typically by relying on some statistical or physical knowledge between the parameter and the spectral response. Typically a band arithmetic formulation is performed (e.g. a vegetation index) and then related to the variable of interest with a curve fitting function (e.g. linear regression).

- *Nonparametric regression methods*: Nonparametric methods construct a regression directly according to information derived from the data. Hence, contrary to parametric regression methods, no explicit choice has to be made about spectral bands, transformation and fitting functions.

- *Physically-based retrieval methods*: Physically-based algorithms follow the physical laws and establish cause-and-effect relationships. They make inferences about model parameters based on general knowledge, such as radiation transfer models (RTMs).

- *Hybrid methods*: Hybrid method combine elements of statistical (typically nonparametric) with physically based methods. They try to exploit the generality of physically-based methods and the flexibility and computational efficiency of nonparametric nonlinear regression methods. The idea is to learn the inverse mapping with a nonparametric model that is being trained using simulated data generated by radiative transfer models.

Because hybrid retrieval methods essentially consists of combining any sort of methods from above categories, they are not further addressed. The following sections will review the most important parametric and nonparametric statistically-based and then the physically-based retrieval methods.

2.3.1 Parametric regression methods

Parametric regression methods build explicit parametrized expressions that relate a few spectral channels with the biophysical parameter of interest. These models assume that the data has come from a type of probability distribution and makes inferences about the parameters of the distribution. Generally speaking, parametric models make more assumptions than nonparametric methods. If those extra assumptions are correct, parametric methods can produce more accurate and precise estimates. However, if assumptions are incorrect, parametric methods can be very misleading. For that reason they are often not considered robust. On the other hand, parametric formulae are often simpler to write down and faster to compute. In some, but definitely not all cases, their simplicity makes up for their non-robustness, especially if care is taken to examine diagnostic statistics. The principle basically entails two steps. It first applies mathematical combinations of spectral channels and then correlates it with the biophysical variable dataset. The correlation occurs through either a linear or nonlinear (e.g. exponential, power, polynomial) fitting function. Many spectral algorithms have been proposed to enhance subtle spectral features and to reduce undesired effects caused by variations in soil reflectance, sun and view geometry, atmospheric composition, and other leaf or canopy properties. These spectral algorithms can be basically categorized into two type of approaches: (1) discrete band formulations and (2) quasi-continuous spectral band formulations. The first category refers to the broad of family of vegetation indices. Vegetation indices are amongst the oldest techniques to interpret spectral data and found their origin in broadband sensors where only from a limited set of band can be chosen from. These are typical two-, three- or four-band formulations according to simple arithmetic expressions. While proven useful in many applications, none of these indices take full advantage of the available spectral information. This under-exploitation is especially prominent when applying to hyperspectral data. It led to the development of parametric techniques that are specialized in exploiting the quasi-continuous signal. Instead of calculating formulas on the basis of a few discrete bands, these techniques aim to extract information from the shape of specific spectral regions and then correlate it with a biophysical parameter. Widely used examples of quasi-continuous spectral band formulations include: (1) red-edge position calculations, (2) derivative-based indices, (3) integral-based indices, and (4) continuum removal.

2.3.2 Nonparametric regression methods

In contrast to parametric regression methods, nonparametric regression methods *learn* the relationship between the input (spectral information) and output (biophysical variables) by fitting a flexible model directly from the observed data. The term *nonparametric* means that no explicit assumption is made about the relations between the data, e.g. no particular distribution (like the Gaussian) is assumed. Nonparametric methods can be categorized into linear or nonlinear methods, depending on the nature of the transformation. Nonparametric regression models have the potential to generate adaptive, robust relationships and, once trained, they are fast to apply. Because full spectra can be directly introduced into nonparametric regressors, these approaches are

increasingly applied to multi- and hyperspectral data. Typically, they are able to cope with the strong nonlinearity of the functional dependence between the biophysical variable and the observed reflected radiance. They are therefore powerful candidates for operational applications. There exists a growing taxonomy of nonparametric regression approaches. While limiting to the most important ones, they can be categorized into either linear or nonlinear nonparametric models.

Linear nonparametric models

Linear nonparametric regression model became quite popular into remote sensing applications for the last decade because of their good performance and simplicity. However, it may not be the best choice when dealing with complex datasets exhibiting nonlinear feature relations, such as it is often the case in multi- and hyperspectral images. Canonical linear regression methods typically rely on the estimation of covariances which can be problematic in the case of very low rates of examples per input data dimensionality. This is well-known in the remote sensing field as *curse of dimensionality*, first described by Hughes [Hughes, 1968]. This is the reason why usually most linear methods are applied after a dimensionality reduction that tries to alleviate the collinearity problems[1]. Examples of linear nonparamtric approaches in vegetation properties mapping include:

- *Principal component regression (PCR)* is a regression analysis that uses principal component analysis (PCA) when estimating regression coefficients. The motivation behind using PCA before applying a linear regression is to avoid the problems that may arise in data that suffer from collinearity, such as in the hyperspectral case. In those cases, standard linear regression obtains unbiased estimations, but with large variances that made them far from true values. The solution adopted in PCR is to perform the regression on the projected data onto the most relevant components (called *scores*) obtained using PCA [Jolliffe, 1986; Wold *et al.*, 1987].

- *Partial least squares regression (PLSR)* deals with the collinearity problem in a different way than PCR. In PCR, the regression is done on the PCA scores. These projections are obtained using only the input patterns, not the outputs. By contrast, PLSR builds the regression model on the projections obtained using Partial Least Squares (PLS) projections, which finds the directions of maximum input-output cross-covariance. Therefore, PLSR takes into account both the input patterns and the output variables to estimate in the projected space, on which the linear regression is carried out [Wold, 1985; Geladi and Kowalski, 1986].

- *Ridge regression (RR)*. As in PCR and PLSR, ridge regression (RR) is a linear least squares regression developed to deal with collinearity problems. As stated before, standard linear regression obtains unbiased estimations, but suffers from large variance. RR

[1] Collinearity is the existence of near-linear relationships among the independent variables.

deals with this problem allowing a degree of bias in the estimates. It does so by adding a small positive value, λ, to the diagonal elements of the correlation matrix of the input data[2]. It can be shown that RR obtains biased estimations compared to standard linear regression, but with less variance [Hastie *et al.*, 2009]. The problem with RR is to find an optimal value for λ. Typically, v-fold approaches are used to obtain a near optimal value.

Nonlinear non-parametric models

In parallel to linear methods, a wide range of nonlinear, nonparametric methods have been developed. Within this family special attention goes to the family of machine learning regression algorithms (MLRAs). The main advantages of MLRAs include its capability of solving non-linear relationships, no assumption about the underlying data distribution, possibility to incorporate *a priori* knowledge, and the ability to combine different types of data into the analysis. In the following subsections we briefly summarize the most commonly used MLRAs.

- *Neural networks:*

  Neural networks (NN) combine both adaptive and non-adaptive elements. They rely on the principle of non-linear, distributed, parallel and local processing and adaptation. A NN is a (potentially fully) connected structure of neurons organized in layers, a neuron being just a linear regression followed by a nonlinear function, $f(\cdot)$. Neurons of different layers are interconnected with the corresponding links (weights). Therefore, in the limit case of using a NN with only one neuron, the results would be similar (or slightly better) than those obtained with linear regression. Roughly speaking, an NN is typically defined by three types of parameters: (1) The interconnection pattern between different layers of neurons; (2) The learning process for updating the weights of the interconnections; and (3) The activation function that converts a neuron's weighted input to its output activation [Haykin, 1999].

- *Decision tree learning:*

  Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. The model structure is formed by a set of nodes that are hierarchically connected: in each node a linear decision is made based on specific input features. Tree models are sometimes referred to as classification or regression trees, depending on the problem at hand. Very often, a simple tree cannot cope with strong nonlinear input-output dependencies. In such cases, the combination of trees can improve the results [Quinlan, 1986].

- *Kernel methods*

  Kernel methods [Shawe-Taylor and Cristianini, 2004] owe their name to the use of kernel functions, which measures similarities between input data examples. Such similarity reproduces a linear dot (scalar) product computed in a possibly higher dimensional feature

[2]RR takes its name because this increase of the diagonal is somehow similar to a ridge.

space, yet without ever computing the coordinates of the data in that space. This approach is known as the kernel trick. The use of kernel methods have been widely adopted in geoscience and remote sensing applications as well [Camps-Valls and Bruzzone, 2009].

- *Support vector machines* (SVMs) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. A version of SVM for regression was proposed in [Vapnik *et al.*, 1997].
- *Kernel ridge regression* (KRR), also known as least squares support vector machines (LS-SVM), are a set of related supervised learning methods that analyze data and recognize patterns, and can be used for classification and regression analysis. In this version one finds the solution by solving a set of linear equations instead of a convex quadratic programming (QP) problem for classical SVMs [Suykens and Vandewalle, 1999].
- *Relevance vector machines* (RVM) use Bayesian inference to obtain parsimonious solutions for regression and classification. The RVM has an identical functional form to the support vector machine, but provides probabilistic classification. Compared to that of SVMs, the Bayesian formulation of the RVM avoids the set of free parameters of the SVM (that usually require cross-validation-based optimizations). However RVMs use an expectation maximization learning method and are therefore at risk of local minima. RVMs provides also more sparse solutions than SVMs [Tipping, 2001].
- *Gaussian processes regression* is based on Gaussian processes (GPs), which are a generalization of Gaussian probability distributions over the space of functions. A Gaussian process is a stochastic process that describes the properties of functions. As in Gaussian distributions, a GP is also described by its mean (which for GPs is a function), and its covariance (a kernel function), which represents the expected covariance between the value of a function at a given point [Rasmussen and Williams, 2006a].

2.4 Principles of physically-based methods: Reflectance Modeling

Whereas statistical approaches are governed by the available data that has been presented during the calibration (parametric) or training (nonparametric) phase, instead, physical approaches are

based on physical laws governing the light-leaf and light-canopy processes of absorption and scattering. The interaction of solar radiation with the canopy is described by means of radiative transfer equations [Myneni and Ross, 1991], which is mathematically implemented in leaf and canopy radiative transfer models (RTMs). In the following sections the main principles of leaf and canopy RTMs are addressed. Only those RTMs that have been implemented into the ARTMO toolbox (c.f. section 5.4) are discussed.

2.4.1 Leaf radiative transfer models

Several leaf optical models with different modeling techniques have been developed for both broadleaf and needleleaf. The basic principles of the most representative leaf models are introduced below.

'Plate' Models

The plate model developed by Allen *et al.* [1969] considers a compact plant leaf as a semi-transparent plate with plane parallel surfaces and initially assumes that the incident light is partially isotropic. That requirement is equivalent to the assumption that the surfaces are rough. Figure 2.4 illustrates incident light interacting with a compact plant leaf: it is partly reflected, partly transmitted and partly absorbed.

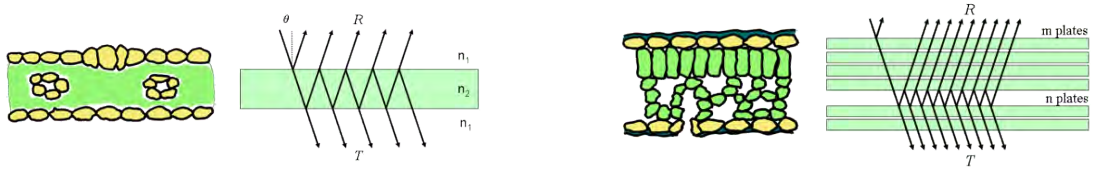


FIGURE 2.4: Schematic representation of a monocot (left) and dicotyledon (right) leaf and multiple reflections produced by a single plate (left), and a set of $N = m + n$ plates (right). Figure after Jacquemoud and Ustin [2001].

The expression for the total reflectance of the plate, R, can be derived by summing the amplitudes of successive reflections and refractions. Allen *et al.* [1970] followed by Breece and Holmes [1971] rapidly extended the plate model to non-compact leaves by introducing the generalized plate model which simply consists of stacking elementary plates. This new model accounts for the development of intercellular spaces in the leaf mesophyll. The leaf is conceptually subdivided into N uniform compact plates separated by N - 1 air spaces. Such a system has been solved for reflectance and transmittance many years ago by the Irish mathematician Stokes (1862) who actually contemplates a set of $N = (m + n)$ plates, and obtains the reflectance $R(m + n)$ and the transmittance $T(m + n)$ of this set in terms of the reflectances and transmittances of the two subsets consisting of m and n plates (Figure 2.4). A well-known leaf reflectance model is the PROSPECT model, first proposed by Jacquemoud and Baret [1990] and modified

later Jacquemoud *et al.* [1995]. Based on PROSPECT, various refinements have been developed, e.g. FluorMODleaf [Pedrós *et al.*, 2010], that also delivers fluorescence emission, and Dorsiventral Leaf Model (DLM) [Stuckens *et al.*, 2009], that distinguishes between adaxial and abaxial optical properties.

2.4.2 Canopy radiative transfer models

Vegetation canopy RTMs have initially been developed for the study of photosynthesis in plant stands. Stimulated by optical remote sensing technology, these models were developed in order to obtain a better insight in the interaction between incident radiation and vegetation canopies. Canopy RTMs can be categorized into four groups [Goel, 1998]:

- *Turbid medium models:*

  The vegetation elements are treated as small absorbing and scattering particles with given optical properties, distributed randomly in horizontal layers and oriented in given directions. The canopy is usually assumed to be homogeneous and its architecture is expressed through leaf area index (LAI) and leaf angle distribution. Quantities such as leaf dimensions, the effective distance between leaves, and the nonrandom distribution of leaves along the horizontal direction are typically neglected. The models in this category use either the Kubelka-Munk theory, such as the widely used SAIL model [Verhoef, 1984], or the radiative transport theory, or discrete models. The turbid medium models are particularly successful in representing the reflectance of denser and more horizontally uniform canopies in which the vegetation elements are smaller in size that the height of the canopy. A representation of a turbid medium model simulation is given in Figure 2.5

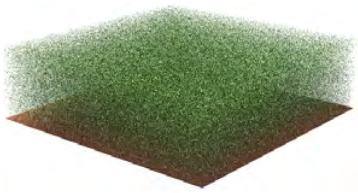


FIGURE 2.5: Representation of a scene generated by a turbid medium model. Figure after RAMI [2014].

- *Geometrical-optical models:*

  Here the canopy is assumed to consist of a ground surface (of known reflective properties) with geometrical objects or protrusions of prescribed shapes (cylinder, sphere, upright and reversed cones, ellipsoid, flat disk etc.) and dimensions and optical properties placed on the ground in a defined manner (regularly or randomly distributed, row structure of trapezoidal, triangular or rectangular cross section). The interception of light and shadowing by the protrusions, and the reflectance from the ground surface are analyzed to determine the reflectance from the canopy. In these models [e.g. Li and Strahler, 1992], the

main assumption is that the geometrical nature of the canopy is the primary factor behind the anisotropy of bidirectional reflectance. A representation of a scene generated by a geometrical-optical model simulation is given in Figure 2.6.

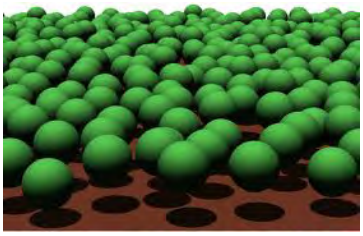


FIGURE 2.6: Representation of a scene generated by a geometrical-optical model. Figure after RAMI [2014].

- *Ray tracing models:*

  In ray tracing models [e.g. Govaerts and Verstraete, 1998; Disney *et al.*, 2006], the arrangement and orientation of vegetation elements are simulated on a computer. Each of these elements is divided into a finite number of areas. A Monte Carlo procedure, involving the selection of random numbers, is used to determine if a given beam of light will hit one of these areas. If it is hit, the direction of scattered radiation is chosen again using a Monte Carlo procedure. Thus, the interception and scattering of radiation is numerically followed, almost on a photon-by-photon basis. These models are computationally intensive, but have the advantage of allowing a more realistic simulation of radiation regime in a canopy. Using these models, one can also investigate the statistical nature of the radiation field, i.e., one can calculate not only the average values but also the higher moments and even the probability distribution of the reflected fluxes. A representation of a scene generated by a ray tracing model simulation is given in Figure 2.7.

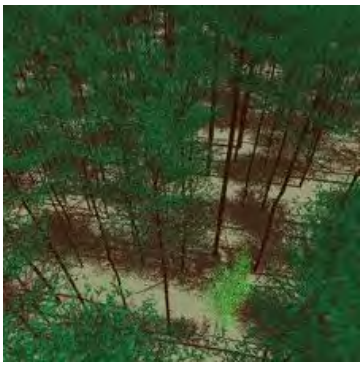


FIGURE 2.7: Representation of a scene generated by a ray-tracing model. Figure after RAMI [2014].

- *Hybrid models:*

  These models combine elements of both the turbid medium, the geometric-optical or the ray tracing models. In hybrid RTMs the canopy is approximated by a distribution

of geometrical-shaped plants except that the multiple scattering is not neglected. The vegetation elements are treated as absorbing and scattering particles. These models are needed for canopies which are neither sparse nor dense. Often they are introduced in a ray tracing environment, e.g. the FLIGHT model [North, 1996]. They are obviously the most complex of geometrical-optical and turbid medium models. These models can be used to represent a wider variety of canopies – sparse to dense with both homogeneous and inhomogeneous tree distributions. However, multiple scattering is not rigorously treated. Like geometrical-optical models, these models also use only one geometrical shape for the trees. A representation of a scene generated by a ray tracing model simulation is given in Figure 2.8.

FIGURE 2.8: Representation of a scene generated by a hybrid model . Figure after RAMI [2014].

2.4.3 Forward running and model inversion

As outlined above, canopy RTMs simulate the reflected radiation for a given observation configuration (e.g. spectral region, sun-target-sensor geometry) and auxiliary variables (e.g. leaf and canopy characteristics). Generally speaking, a discrete forward model to describe such a system is of the form:

$$y=f(X,\theta)+n, \tag{2.1}$$

where y is a set of measurements (such as the expected radiance); X is a matrix containing a set of controllable measurement conditions (such as different combinations of wavelength, viewing direction, time, Sun position, and polarization); θ is a vector of state parameters of the system approximation (i.e. the model input variables); n is an error vector (model and measurement error, noise) [Camps-Valls *et al.*, 2012b]. Conversely, it is also possible to retrieve the model input variables from remote sensing measurements through model inversion. The inversion problem implies the design of algorithms that, departing from the radiation acquired by the sensor (or removed from atmospheric effects, i.e. top-of-canopy reflectance), can calculate back to estimates of the variables of interest through the RTM calculations, thus 'inverting' the RTM. A schematic illustration of the forward and inversion processes is shown in Figure 2.9. In the inversion process, *a priori* information about input parameters or the variables of interest can also be included to improve the performance, such as vegetation properties, geographical location, acquisition time or statistical properties of the data distribution [Baret and Buis, 2008].

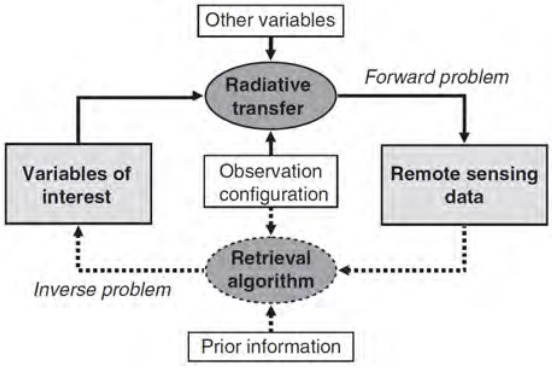


FIGURE 2.9: Forward running and model inversion in variable retrieval. Figure after Baret and Buis [2008].

Unfortunately the inversion process is not what Hadamard [1902] defined as a "well-posed problem". He believed that mathematical models of physical phenomena should have the properties that: (1) a solution exists, (2) the solution is unique, and (3) the solution's behavior changes continuously with the initial conditions. If at least one of the criteria above is not fulfilled the problem is said to be "ill-posed". The problem of inverting the function f is in general undetermined and highly ill-posed: the number of unknowns is generally larger than the number of independent radiometric information remotely sampled by sensors. In general, estimating biophysical parameters from inversion of RTMs against satellite-derived data constitute very difficult problems due to the presence of high levels of uncertainty, such as those associated to sensor calibration, model realism, sun-target-viewing geometry, quality of atmospheric correction [Baret and Buis, 2008; Camps-Valls *et al.*, 2012b].

Since the RTM inversion is highly ill-posed, the inversion technique is based on finding the best match between the measured reflectance values and those either simulated by a RTM or stored within a database made of experimental observations. This bears the consequence that the performances of the approach will both depend on the minimization algorithm itself, the so-called "cost function" and on the level of ill-posedness of the inverse problem as a function of measurement configuration and model and measurement uncertainties. Several minimization techniques have been introduced to circumvent the ill-posed problem: (1) classical iterative optimization, (2) simulated annealing, (3) genetic algorithms, (4) look up tables and (5) Monte Carlo Markov Chains. However, classical iterative optimization techniques and look up tables (LUT) have been the most widely used. LUT-based inversion will be further optimized in this Thesis.

Look Up Tables (LUT)

UT-based inversion is conceptually the simplest inversion technique, although its implementation is not trivial, approach is used to speed up the inversion process. It precomputes the model reflectance for a large range of combinations of parameter values. In this manner, the most

computationally expensive aspect can be completed before the inversion is attempted, and the problem is reduced to searching a LUT for the modeled reflectance set that most resembles the measured set [Liang, 2007]. This is mostly done through a cost function, typically by minimizing the summed differences between simulated and measured reflectances for all wavelengths, but also other cost functions and optimization techniques will be introduced in this Thesis.

2.4.4 Applying RTMS for vegetation properties mapping

The above sections illustrated that different types of RTMs have been developed that range in complexity and computational speed. These differences play a role in the inversion performance, and RTM comparison and evaluation is required. Among the most extensive validation exercises involve the RAdiation transfer Model Inter-comparison (RAMI) exercises, which took place in participation with model developers. Four RAMI exercises have been conducted so far; each time having more advanced models included and more accurately representing real-world situations [Pinty *et al.*, 2001, 2004; Widlowski *et al.*, 2007, 2011]. Although RAMI provides a benchmark for model developers, however, currently no services are available that bring these validated models together on one platform. In fact, many published RT models are not easily accessible. Conversely, some models have a long history in being freely distributed throughout the community and became a standard in RS research. The turbid medium canopy reflectance SAIL model [Verhoef, 1984] is such an example and is currently among the most popular models in RT-based biophysical parameter mapping exercises [Jacquemoud *et al.*, 2009]. A common practice in RS research is opting for one of those available models, regardless of evaluating whether it correctly represent the studied real-world situation. For instance, many studies have inverted SAIL with the purpose of retrieving forest biophysical properties [e.g., Gond *et al.*, 1999; Soudani *et al.*, 2006; le Maire *et al.*, 2008; Zhang *et al.*, 2005]. Nevertheless, SAIL does not simulate 3D structures of trees and associated shadowing. Hence, the risk of oversimplifying the RT processes is introduced when applying a 1D model to forests and so the likelihood of propagating errors in the inversion process. Given this all, there is a need for a platform where multiple 1D and 3D models can be freely accessed so that the user can compare and apply these models for their own purposes.

Along with the absence of a multiple-models platform, another faced inadequacy is that currently no advanced retrieval toolbox is offered in mapping software packages. This can be potentially overcomed. While in the scientific literature a diversity of inversion strategies have been proposed [e.g., Jacquemoud *et al.*, 1995; Chen *et al.*, 1997; Weiss *et al.*, 2000; Combal *et al.*, 2003; Zarco-Tejada *et al.*, 2003], for the majority of approaches the same sequence of processing steps is pursued, such as the generation of a LUT and applying cost functions. Finally, in a recent paper reviewing RT models [Jacquemoud *et al.*, 2009] it was acknowledged that a Graphical User Interface (GUI) would be very helpful to make users understand the value of RTMs in obtaining more accurate information about plant biophysical properties. Furthermore, a GUI can easily be extended to embed all necessary tools to automate model inversion against RS observations, and so eventually compute concurrent maps of biophysical parameters.

To the benefit of RS-related research, it is therefore desirable to bring RT models and essential inversion processing steps together in one GUI toolbox. The toolbox should: (1) consist of a suite of leaf and canopy models, (2) be user friendly, (3) allow easy comparison of different models, (4) automate retrieval strategies to map vegetation properties, and (5) be freely available. A scientific GUI toolbox that aims to cover all these points has been developed during this Thesis project, called ARTMO.

2.5 ARTMO: Software framework for vegetation properties retrieval

ARTMO (Automated Radiative Transfer Models Operator) is a scientific GUI software package developed at the Laboratory of Earth Observation (LEO) at the University of Valencia [Verrelst *et al.*, 2011; Rivera, 2011]. The software structure has been developed according to a three-tier architecture and two connecting levels in MATLAB and connected with a MySQL server running underneath. A detailed description of the architecture can be found in [Rivera, 2011]. Since its first version, the ARTMO package consists of three different conceptual blocks: (1) *radiative transfer models (RTMs)*, (2) *Retrieval toolboxes* and (3) *Tools*. Figure 2.10 shows the RTMs, retrieval algorithms and tools that was implemented in v1.00 [Rivera, 2011].

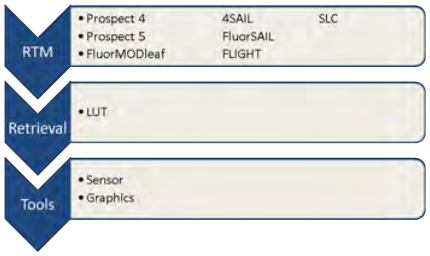


FIGURE 2.10: ARTMO's v.1.00 RTMs, inversion module and tools as developed in [Rivera, 2011].

A major weakness of ARTMO v1.00, however, was that its logical tier structure did not permit the expansion of its modules. Therefore, one of the Thesis' objectives was to improve the conceptual development and implementation of ARTMO as a "modular software framework" by redesigning its structure according to a dynamic extendable modular architecture. It eventually led to the current version 3.03, which is modular and consists of various new modules. By following a basic logical structure that allows communication between the included modules, the software package can be customized and extended by any programmer. The conceptual development of the modular software framework is hereafter presented. A general description of each module along with Manual and Installation Guides can be consulted at ARTMO's website: http://ipl.uv.es/artmo/.

2.5.1 ARTMO's software framework

From v3.00 onwards, ARTMO is built out of four different module types: (1) *Main*, (2) *radiative transfer models (RTMs)*, (3) *Retrieval toolboxes*, and (4) *Tools*. The latter three are now removable modules that are built on top of the *Main* module. Figure 2.11 shows the general outline of the modular software framework. Each module provides the programmer with a standardized set of properties and functions that allow communication between modules. The *main* module is responsible for creating the underlying MySQL database and provides the script that takes care of the communication between the modules and the underlying database. The *RTM* module runs the configured radiative transfer models, the *Retrieval toolboxes* analyze the performance of different biophysical parameter retrieval methods, and *Tools* module provides scripts for post-processing information generated by the *RTM* module. Each module is formed by one or more 'processing units', which defines a set of scripts and GUIs that executes a specific task.

FIGURE 2.11: Module's outline of the ARTMO Framework.

The architecture of all modules has been developed in a structure of three tiers (**client**, **logic** and **database**) and two connection levels.

The **client tier** consists of a set of GUIs developed for entering and reading input and output data to each element of the logical layer. This structure is applied to each modules. The **logic tier** consists of a set of four basic elements: (1) *builder*, (2) *connector*, (3) *core* and (4) *auxiliary*. Each element is formed by one or more scripts; its behavior depends on the module it belongs. The **database tier** contains the logical database element. This element is common to all modules. The database is designed in MySQL. The first level refers to all elements to be installed locally and consists of the presentation (i.e., all GUIs) and logic tier (i.e. all scripts), the second level refers to the communication between the logic tier and the MySQL database. The database can be accessed remotely. Figure 2.12 shows the layout of components and communication between components. The four module types are subsequently described according to the basic elements.

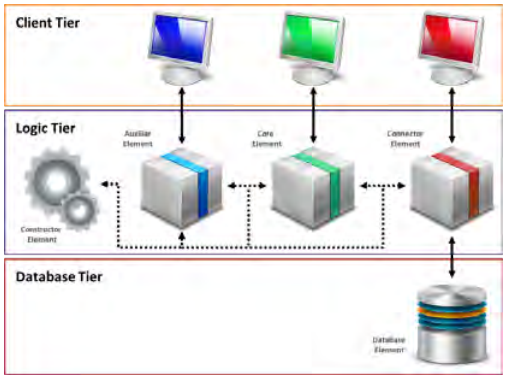


FIGURE 2.12: Visual overview of Three-tiered modules in ARTMO framework.

Main module:

- *Builder*: Stores default information of the general settings and the different main GUI menus of the ARMO Framework.

- *Core*: Is responsible for checking and storing each processing unit configuration data in order to allow the communication among all modules.

- *Connector*: creates the database and contains a set of auxiliary scripts that allows the communication between all modules of logical tier and database.

RTM module:

- *Builder*: Stores all the input and output setup parameters of RTM and the configuration of tables created by the connector element; also creates sub-menus in each menu of the main GUI. Table 2.1 lists ARTMO's implemented RTMs. In this module each implemented RTM is a processing unit.

- *Core*: Runs the main script of the configured RTM.

- *Connector*: Defines the MySQL storage syntaxes of input and output data to each RTM. For instance, Figure 2.13 illustrates a relational database structure for PROSPECT-4.

TABLE 2.1: RTMS implemented into ARTMO

Model	scale	Reference	language
Prospect-4	leaf	Feret *et al.* [2008]	Matlab
Prospect-5	leaf	Feret *et al.* [2008]	Matlab
FluorMODleaf	leaf	Pedrós *et al.* [2010]	Compile file
DLM	leaf	Stuckens *et al.* [2009]	Matlab
4SAIL	canopy	Verhoef *et al.* [2007]	Matlab
FluorSAIL	canopy	Zarco-Tejada *et al.* [2006]	Compile file
FLIGHT	canopy	North [1996]	Compile file
SLC	combined (soil/leaf/canopy)	Verhoef and Bach [2007]	Mex file (Matlab)
SCOPE	combined (soil/leaf/canopy)	Van Der Tol *et al.* [2009]	Matlab

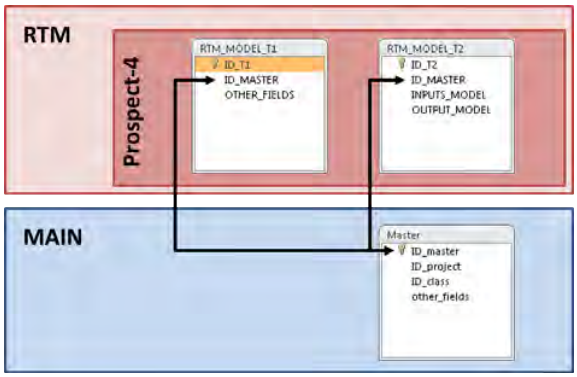


FIGURE 2.13: Table's structure to RTM generic implemented.

Retrieval module:

- *Builder*: Provides basic information to the main module of the retrieval models implemented, creates sub-menus in each menu of the main GUI and stores the settings to run the retrieval model.

- *Core*: Runs the main script of the retrieval model and the different programed scripts in an auxiliary element. Table 2.2 shows retrieval algorithms implemented.

- *Connector*: Defines the MySQL storage syntaxes of input settings of retrieval models and the output statistics analyzed and allows the communication between the core element and the database. Chapters 3 to 5 explain all developed retrieval toolboxes in more detail.

TABLE 2.2: ARTMO's retrieval toolboxes.

Spectral Indices	**LookUp-Table (LUT)**	**Machine learning regression**
Generic spectral index formulations with up to 10 different bands.	63 cost functions according to Leonenko *et al.* [2013]	13 Machine learning regression algorithms according to Camps-Valls *et al.* [2013]

Tools module:

- *Builder*: Provides basic information to the main module, creates access in ARTMO's main GUI window.
- *Core*: Processes the stored information by the processing units of the RTM module in the database.
- *Connector*: Takes care of the communication. The table's syntaxes required by each processing unit is managed by the connector.

In v3.03 the following tools are fully operational:

1. **Graphics tool**: The Graphics tool allows the simulated spectra to be viewed and exported.
2. **Sensor tool**: The Sensor tool enables configuring the band settings of a specific sensor. Once configured, an RTM can then generate outputs according to the bands settings of a chosen sensor.

Finally, it is expected that the development of modular and dynamic tools will contribute to the understanding of the interactions between solar light and vegetation elements. ARTMO's modular framework provides a platform that not only allows researchers, developers and users of RTM to develop new methods or optimize retrieval strategies, but also presents an intuitive toolbox that supports the use and inter-comparison of various RTMs as well possibilities of executing mapping strategies for local-to-regional applications, ranging from precision farming to environmental monitoring issues.

3

Spectral Indices based retrieval

Contents

This chapter is based on:

Juan Pablo Rivera Caicedo, Jochem Verrelst, Jesús Delegido, Frank Veroustraete and José Moreno (2014)

On the semi-automatic retrieval of biophysical parameters based on spectral index optimization.

Remote Sensing (Submitted).

3.1 Abstract

Regression models based on spectral indices are typically empirical formulae enabling the mapping of biophysical parameters derived from Earth Observation (EO) data. Due to its empirical nature, it remains nevertheless uncertain to what extent a selected regression model is the most appropriate one, until all band combinations and curve fitting functions are assessed. This paper describes the application of a Spectral Index (SI) assessment toolbox in the Automated Radiative Transfer Models Operator (ARTMO) package. ARTMO enables semi-automatic retrieval and mapping of biophysical parameters from optical remote sensing observations. The SI toolbox enables the assessment of biophysical parameter retrieval accuracy of established as well as new and generic SIs. For instance, based on the SI formulation used, all possible band combinations of formulations with up to ten bands can be defined and evaluated. Several options are available in the SI assessment: calibration/validation data partitioning, the addition of noise and the definition of curve fitting models. To illustrate its functioning, all two-band combinations according to simple ratio (SR) and normalized difference (ND) formulations as well as various fitting functions (linear, exponential, power, logarithmic, polynomial) have been assessed. HyMap imaging spectrometer (430-2490 nm) data obtained during the SPARC campaign in Barrax, Spain, have been used to extract leaf area index (LAI) and leaf chlorophyll content (LCC) estimates. For both SR and ND formulations the most sensitive regions have been identified for two-band combinations of green (539-570 nm) with longwave SWIR (2421-2453 nm) for LAI (r^2: 0.83) and far-red (692 nm) with NIR (1340 nm) or shortwave SWIR (1661-1686 nm) for LCC (r^2: 0.93). Polynomial, logarithmic and linear fitting functions led to similar best correlations, though spatial differences emerged when applying the functions to HyMap imagery. We suggest that a systematic SI assessment is a strong requirement in the quality assurance approach for accurate biophysical parameter retrieval.

3.2 Introduction

Quantitative and spatially-explicit retrieval of vegetation biophysical parameters is a requirement in a variety of ecological and agricultural applications. Earth observing (EO) satellites, endowed with a high temporal resolution, enable the retrieval and hence monitoring of plant biophysical parameters [Moulin *et al.*, 1998; Báez-González *et al.*, 2002; Dorigo *et al.*, 2007]. With the forthcoming super-spectral Copernicus' Sentinel-2 [Drusch *et al.*, 2012] and Sentinel-3 missions [Donlon *et al.*, 2012] as well as the planned EnMAP [Stuffler *et al.*, 2007], HyspIRI Roberts *et al.* [2012] and PRISMA [Labate *et al.*, 2009] imaging spectrometer missions, an unprecedented data stream becomes available. This data influx requires processing techniques, which are reproducible, accurate and fast, when the retrieval of information on plant growth and health status is envisaged. Typically, the retrieval of biophysical parameters using EO data implies the use of a model [Myneni *et al.*, 1995a], which can be either statistical or physical in nature. A statistical regression model enables the linkage of EO data with biophysical parameters of interest [Gobron *et al.*, 2000] by making use of (*in situ*) calibration data. On the other hand, physically-based approaches allow the retrieval of the biophysical parameters by inversion of a radiative transfer model (RTM) against EO data [Houborg and Boegh, 2008; Rivera *et al.*, 2013a].

Although inversion of a RTM is generally applicable for a wide range of land cover types and sensor configurations [Verrelst *et al.*, 2014b], the approach is often computationally demanding. It usually requires quite some auxiliary information to enable the parameterization of the physical model and the description of the boundary conditions for which the model is valid [Chen and Cihlar, 1996; Myneni, 1997]. This information may not always be available. Moreover, by introducing input parameter uncertainties, the likelihood increases that model inversion may lead to multiple instead of singular solutions in the soil–vegetation–atmosphere matrices (unified theorem of Hadamard well-posedness). In that case, extra steps are required to overcome the ill-posed problem [Combal *et al.*, 2003].

Conversely, statistical models are more easily implementable for parameter retrieval and mapping applications. The basic principle is to correlate mathematical combinations of measured reflectances for different wavelength ranges or broad spectral bands with biophysical vegetation parameters of interest (e.g., leaf area index (LAI), leaf chlorophyll content (LCC), fractional vegetation cover) using a fitting function. This procedure can be considered as an empirical spectral index (SI) modelling approach. Over the past four decades, a large number of SI models has been developed for the retrieval of biophysical parameters [Myneni *et al.*, 1995b; Broge and Mortensen, 2002; Haboudane *et al.*, 2004]. The majority of these SIs and their relationship with biophysical parameters of interest has been developed based on experimental work [Thenkabail *et al.*, 2002; Ustin *et al.*, 2004; Hatfield *et al.*, 2008].

While being successful for many application cases, the empirical approach suffers from drawbacks as well. Typically, a limited portability to different measurement conditions or sensor characteristics can be noted [Baret and Guyot, 1991]. Moreover, the approach is sensitive to perturbing factors such as atmospheric conditions, canopy characteristics and differences in

viewing or solar position geometry [Verrelst *et al.*, 2008, 2010]. In an attempt to generate more generic relationships, RTM-based dataset simulations have been used to assess for the most sensitive SIs [Ceccato *et al.*, 2002; Zarco-Tejada *et al.*, 2001; Haboudane *et al.*, 2004; le Maire *et al.*, 2004; Zarco-Tejada *et al.*, 2005; le Maire *et al.*, 2008]. A spectral index can quite often be used as an estimator for a biophysical parameter using a fitting function through the data; usually by simple linear regression, but also exponential, power, logarithmic and polynomial fitting functions, among others, are commonly applied [Baret and Guyot, 1991; Delegido *et al.*, 2011; Mariotto *et al.*, 2013]. In statistics, this approach can be categorized as a parametric regression modeling approach. This type of model is determined by the introduction of various properties to obtain formulations linking spectral reflectance with the biophysical parameter of interest. In fact, properties that determine the SI regression function can be defined at three levels:

- *Band selection*: Typically, most SIs are mathematical formulations consisting of two or three sensor spectral bands (B). How then, do we evaluate with a high enough scrutiny, whether the most sensitive spectral bands – with respect to biophysical parameter retrieval – have been selected? This question is especially relevant in view of the high number of bands associated with imaging spectrometry [Thenkabail *et al.*, 2011].
- *SI formulation*: Typically, the normalized difference (ND) formulation is applied, i.e., $(B_2\text{-}B_1)/(B_2\text{+}B_1)$. But again, how do we assess, whether the applied ND formulation is the most accurate one with respect to biophysical parameter retrieval? Even given high spectral resolution multi- or hyperspectral reflectance data, there is no reason to assume that a two-band SI formulation leads to the most accurate empirical relationship [Verrelst *et al.*, 2012b].
- *Fitting function*: A regression model is typically reduced to a linear fitting exercise, directly or indirectly by prior transformation to linearity. Also here, the question is, whether the regression function selected is the most accurate one? Typically, saturation effects are common for dense canopies [Chen and Cihlar, 1996; Myneni, 1997].

Accordingly, it is uncertain, whether an established SI – many have been published in the remote sensing literature – will lead to the most accurate biophysical parameter retrieval, particularly when applied to hyperspectral EO data. To ensure that the most sensitive SI regression model is selected for a specific application, a systematic as well as consistent assessment of an exhaustive set of possible band combinations, SI formulations and curve fitting procedures may be required prior to the selection of an SI model for the retrieval of biophysical parameters from (hyperspectral) EO imagery.

Evidently, this is not a trivial task. In view of forthcoming imaging spectrometers, and to obtain an optimized SI model for biophysical parameter retrieval, the assessment of a large number of band combinations is required. To this end, the requirement for a graphical user interface (GUI) toolbox, assisting EO users in the tedious job of evaluating large sets of SI models, is not a luxury. This observation brings us to the following objectives: (1) To develop an 'SI

assessment toolbox' for the semi-automatic retrieval of biophysical parameters; (2) To enable a robust assessment procedure to estimate the impact of different index formulations, spectral band settings and curve fitting functions with respect to their capability of biophysical parameter retrieval; and (3) To apply the most accurate SI regression model, to extract biophysical parameters from EO imagery.

Since spectral indices and their associated regression models can either be developed based on experimental or simulated data, both options should be present in an SI assessment module. To this end, the SI toolbox has been developed within the ARTMO (Automated Radiative Transfer Models Operator) modeling environment. The experimental dataset used for testing has been acquired from ESA′s SPARC campaign (Barrax, Spain).

In the following sections, the most recent version of the ARTMO modeling environment will be briefly described. Subsequently, an introduction to the most important components of the SI assessment toolbox will follow. The dataset used is described and finally, an assessment of selected SIs will be presented. A discussion section is devoted to newly developed spectral indices and a conclusion section finalizes this paper.

3.3 ARTMO software package

ARTMO, developed and running in Matlab (2011a or higher), offers multiple leaf and canopy radiative transfer models (RTMs) accessible in a GUI environment. Dedicated retrieval methods, required for the semi-automatic retrieval of biophysical parameters, are provided. Examples of ARTMO run options are:

- Selection of various forward and invertible leaf and canopy RTMs of a low to high complexity (PROSPECT-4, PROSPECT-5, 4SAIL, FLIGHT);
- The choice to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems;
- The option to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400 to 2500 nm). Look-up tables (LUT) can be generated, which are stored in a relational SQL database management system (MySQL, version 5.5 or higher; local installment required);
- Finally, various retrieval scenarios can be selected and run using EO reflectance datasets.

Figure 5.1 illustrates the ARTMO v3 GUI main window and a systematic overview of the modules. In the GUIs main window, a new project can be defined, a sensor chosen and comments added. All processing modules are accessible by drop-down menus from the top bar. Already in the first version of ARTMO, LUT-based inversion applications were available [Verrelst

et al., 012c; Rivera *et al.*, 2013a]. ARTMO v3 is formally presented in this paper. The software package, as well its manuals, installation guides and tutorials, are freely downloadable at: http://ipl.uv.es/artmo/. Its most important novelties are briefly listed below:

- ARTMO v3 is designed modularly. Its modular architecture offers the possibility for easy addition (or removal) of components, such as RTM models and post-processing modules.
- The MySQL database is organized in such a way that it supports the modular architecture of ARTMO v3. This avoids redundancy and increases the processing speed. For instance, all spectral datasets are stored as binary objects.
- New retrieval toolboxes are incorporated. They are based on parametric and non-parametric regression as well as physically-based inversion using a LUT. This has led to the development of a: (1) 'Spectral Indices assessment toolbox'; (2) 'Machine Learning Regression Algorithm toolbox' Rivera *et al.* [2013b]; and, (3) 'LUT-based inversion toolbox' [Rivera *et al.*, 2013a].

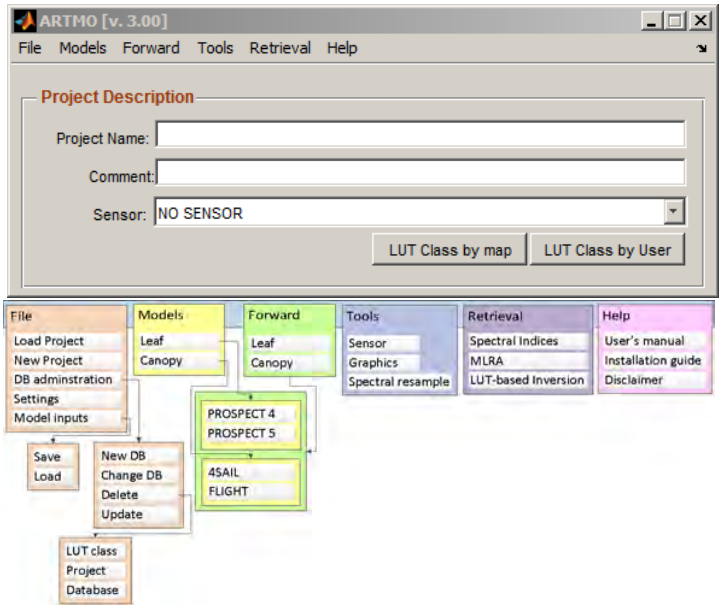


FIGURE 3.1: Screenshot of ARTMO's main window and schematic overview of its drop-down menu.

3.3.1 ARTMO Spectral Indices (SI) toolbox

This paper introduces a 'Spectral Indices assessment toolbox'. Its general architecture is outlined in Figure 3.2. The toolbox enables the loading and applying of existing or user-defined spectral indices. In short, the toolbox encompasses the following utilities. For each new index, a spectral region for each waveband is specified. Every single band of a sensor that falls within

the specified spectral region will then be included in the retrieval capacity assessment. Different types of curve fitting functions have been implemented (e.g., linear, power, exponential, logarithmic, polynomial functions), while new functions can easily be added by the user. Input data originating from RTM runs can be used or user-defined input data can be imported (a plain text file), typically a dataset acquired during a field or flight campaign (i.e., biophysical parameters and associated surface reflectance). Input data can then be partitioned for calibration and validation of the SI–biophysical parameter relationships. Options are available to merge or partition different input datasets; e.g. calibration on the basis of RTM data and validation on the basis of user acquired field data. A land cover map in ENVI format (Exelis Inc.) can be loaded, which allows that for each land cover class distinct SI optimization strategies can be defined (e.g., vegetation indices for vegetated surfaces, water indices for water bodies, etc.). When validation data are available, a large diversity of spectral band combinations as well as regression functions, can be assessed using the validation dataset and statistical analysis results (c.f., section II-D). The most accurate strategy can then be selected, loaded and applied to retrieve biophysical parameters from EO imagery (in ENVI format). Important steps in the outlined procedures are presented in the following sections, they are: (1) *'Add spectral index'*; (2) *'SI settings'*; (3) *'Calibration/validation assessment'*; and, (4) *'Retrieval'*.

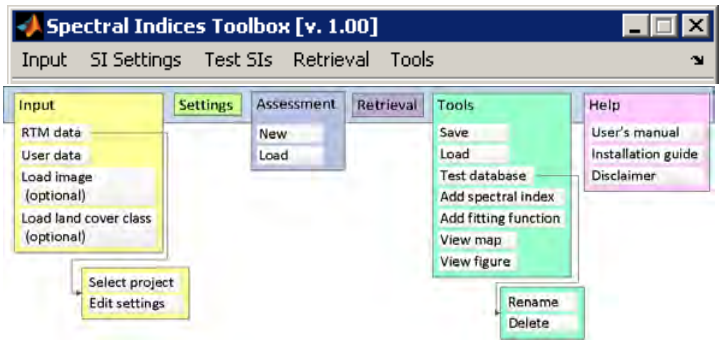


FIGURE 3.2: Screenshot of the ARTMO v3 'Spectral Indices assessment toolbox' and schematic overview of its drop-down menu.

3.3.2 Add spectral index

The 'Add spectral index' window Figure (3.3) in the 'Tools' menu allows for adding a new index to a list with pre-defined SIs. Spectral indices can be created manually using the GUI or they may be imported. SIs are organized in 'SI Groups', according to their similarity in definition, e.g., broadband, narrowband or according to their sensitivity with respect to certain absorption features, e.g. by pigments or water. For instance, the 'Broadband Greenness' SI group consists of the conventional simple ratio (SR: B_2/B_1) and ND (B_2-B_1/B_2+B_1) formulations. A new SI group can also be created, in which multiple SI formulations can be incorporated according to specific input requirements; name, acronym and SI equation definition. When this step is completed, the transfer equation will be assessed for its validity and spectral bands are identified. For each spectral band, a default wavelength or a minimum and maximum of a wavelength

range must be specified. During the subsequent assessment step, all bands within the prespecified spectral region will be automatically evaluated. When kept undefined, every single band of a given sensor will be evaluated according to the specified SI formulation.

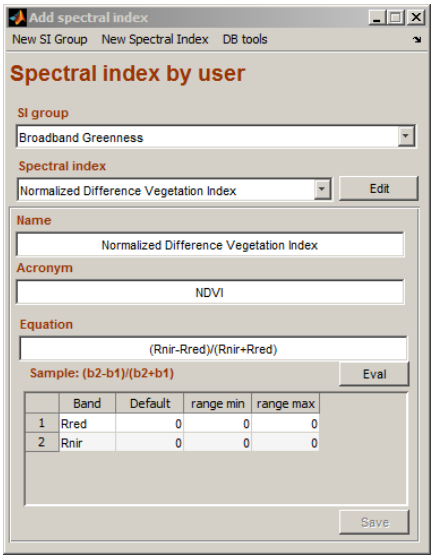


FIGURE 3.3: GUI section used to add a new SI formulation.

3.3.3 SI settings

In order to activate the possibility of defining SI settings, input data are required. Insertion of input data is performed with the 'Input' window. Input data can either consist of a RTM simulated dataset or it can be a field campaign dataset in combination with an EO dataset acquired during the field campaign. The 'Input' GUI helps to perform the data selection steps and checks, if data input is correct (GUI not shown for sake of brevity; consult SI toolbox manual at http://ipl.uv.es/artmo/). Subsequently, the SI settings configure the parametric SI models according to various specifications, which are given in the 'SI settings' GUI (see Figure 3.4). First, if multiple land cover types have been pre-defined (using the GUI's 'Load Image and Class' windows), retrieval strategies can be configured specifically for each land cover class. Second, an SI group must be selected so that its encompassing SIs are subsequently listed. From this list, multiple SIs can be selected for assessment purposes. Third, several pre-defined curve fitting functions can be selected for assessment. Fourth, the option to add Gaussian noise is provided, which is useful for the assessment of environmental as well as instrumental uncertainty estimations. The noise function can be applied both for the parameters to be retrieved, as well as on the reflectance spectra used in the assessment. A range of noise levels can be specified; hence multiple noise scenarios can be assessed with respect to their impact on retrieval accuracy. Fifth, the calibration/validation (cal/val) data partitioning can be defined according to proportional assignment of data to either calibration or validation (i.e. split-sample approach). Both, field and RTM datasets can be combined by assigning a

proportion of each either to calibration or validation. Multiple cal/val partitions can be applied in the assessment.

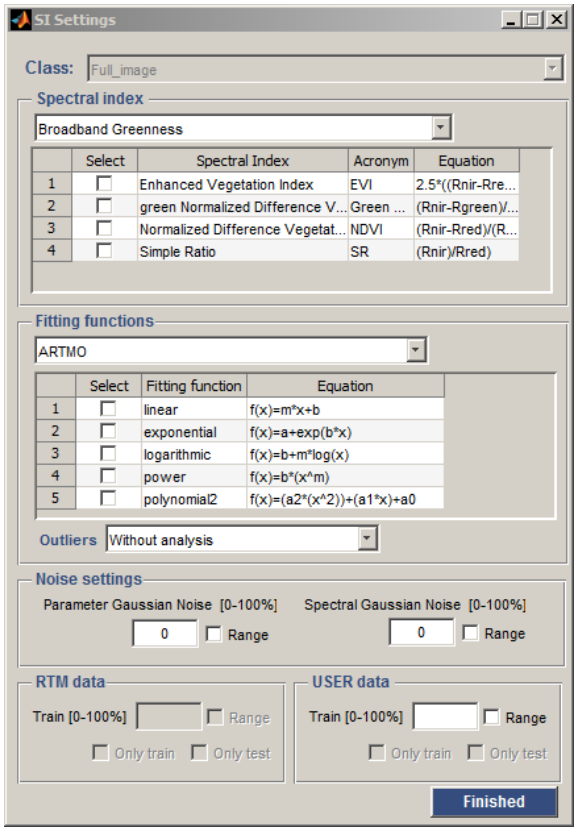


FIGURE 3.4: SI settings window.

3.3.4 Calibration/validation assessment

When input data were correctly imported in ARTMO, cal/val data partitioning has been defined and the configuration of SI settings has been successful, the SI models can be assessed with respect to retrieval accuracy. This analysis starts by assigning a name to a new test dataset in the 'Test SI' menu. The specified retrieval strategies for the SI formulations and spectral band ranges are then analyzed one-by-one. The assessment results are saved by default in the MySQL database. This way, a large number of SIs can be systematically assessed, and results can be stored and queried. An overview table is then assembled, in which assessment results with the highest accuracy scores are dynamically presented; e.g. per land cover type, per parameter, and per cal/val dataset (see Figure 3.5). Assessment is then performed. First, for each SI, all spectral band combinations are correlated against the calibration dataset (results not shown). Second, each SI model obtained according to the chosen fitting function is applied to the estimation of a targeted biophysical parameter. When the dataset has previously been partitioned into a

calibration and validation set, the assessments are generated twice; one time for the calibration dataset and a second time for the validation dataset. The obtained SI models are evaluated with multiple goodness-of-fit measures like the r^2, the root mean squared error (RMSE), the normalized RMSE (NRMSE [%] = RMSE / (value range of parameters as measured in the field) *100), the mean absolute error (MAE) and the mean error (ME). All measures indicate the degree of association between predicted and estimated values of the same parameter and give thus an indication of prediction efficiency. Richter *et al.* [2012a] recommended the combined set of r^2, RMSE and NRMSE, amongst others, for comprehensively quantification the performance of vegetation biophysical models. For a single SI model selected, an option has been foreseen in the GUI, to generate a scatterplot of the retrievals in function of the cal/val measurements (c.f. section 3.5). Additionally, by selecting a statical measure, a 2D correlation matrix can be assembled along 2 spectral dimensions, enabling the visualization of the most sensitive 2-band combinations. Finally, for each retrievable biophysical parameter, an SI model can be selected, based on its accuracy assessment. The selected SI model, e.g. the best performing one, can be accessed through the 'Retrieval' GUI window and may be applied using an EO image for spatially distributed biophysical parameter mapping.

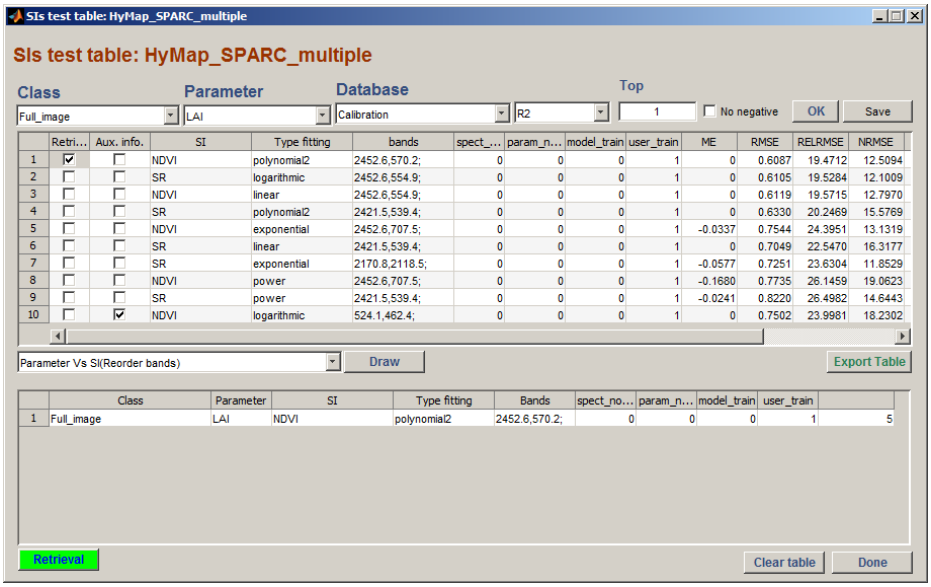


FIGURE 3.5: SIs calibration/validation assessment table.

3.3.5 Retrieval

The 'Retrieval' window enables to run an assessed SI model, or to directly configure an SI model and apply it to map retrieved parameter values (Figure 3.6). The latter option can be useful when dealing with a pre-defined SI model (e.g. when an SI is applied as published in the EO literature) or when cal/val datasets are not available. Hence, the user can select a specific land cover type (if available), a biophysical parameter to be retrieved, an SI group and an SI, a

regression function and the regression function parameters. The user will be able to select one or multiple remote sensing images for which the evaluated SI model will be applied. Biophysical parameter maps are subsequently stored on disc in ENVI format.

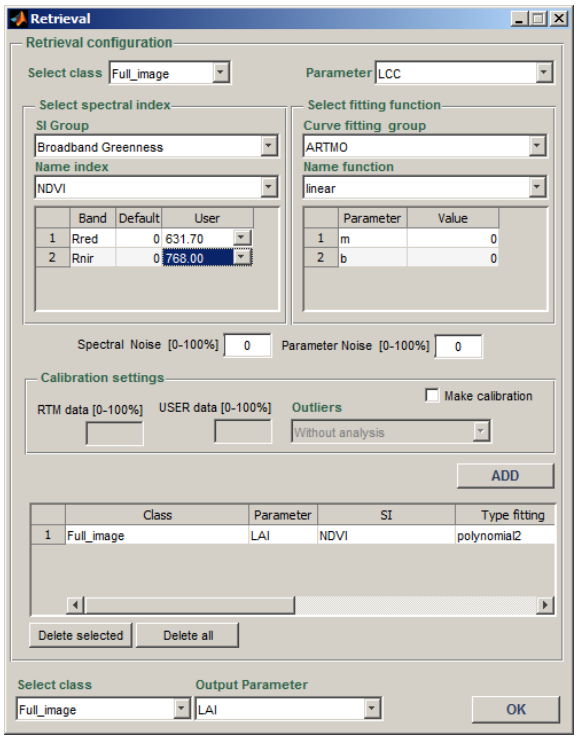


FIGURE 3.6: SIs Retrieval window.

3.4 Assessment and mapping applications

As outlined in the previous section, the SI assessment toolbox is applied to assess the performance of multiple SI regression models with respect to estimation accuracy. Data used are presented first, followed by the experimental setup. Goodness-of-fit results for calibrated SI models are presented subsequently and finally the retrieved spatial maps of biophysical parameters are shown.

3.4.1 Used data

A field dataset, encompassing different crop types, growing phases, canopy geometries and soil conditions has been acquired during the SPectra bARrax Campaign (SPARC). The SPARC-2003 campaign took place from 12 to 14 July in Barrax, La Mancha, Spain (coordinates 30°3'N, 28°6'W, 700 m altitude A.S.L.). Biophysical parameters have been measured within a total of

108 Elementary Sampling Units (ESUs) for different crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat). An ESU refers to a plot, which is sized compatible with pixel dimensions of about 20m $\times$ 20m. Leaf Chlorophyll Content (LCC) was determined by measuring about 50 leaf samples in each ESU with a calibrated CCM-200 Chlorophyll Content Meter [Gandía *et al.*, 2004]. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned a LAI value, obtained as a statistical mean of 24 measurements (8 data readings $\times$ 3 replica) with standard errors ranging from 5 to 10% [Fernández *et al.*, 2005]. Strictly speaking, due to the assumption of a random distribution of foliage, the impact of clumping has been assessed only partially using the LiCor and the corresponding software. Hence, effective LAI is given as an output parameter. For all ESU's, LAI ranges from 0.4 to 6.2 (m^2 single sided leaf surface)/(m^2 ground surface) while LCC ranges from 2 to 51 (μg chlorophyll)/(cm^2 leaf area). Regarding LCC, it should be noted that some clumping in the field data was noted, with many data points between 40-50 and around 20, but no data points between 25-35 $\mu g/cm^2$. No additional bare soil samples were added.

During the campaign, airborne hyperspectral HyMap flightlines have been acquired for the study site, during the month of July 2003. HyMap flew with a configuration of 125 contiguous spectral bands, spectrally positioned between 430 and 2490 nm. Spectral bandwidth varied between 11 and 21 nm. The pixel size during overpass was 5 m. The flight-lines were corrected for radiometric and atmospheric effects according to the procedure that has been implemented in the BEAM CHRIS-Box [Alonso and Moreno, 2005; Guanter *et al.*, 2005]. Finally, a calibration dataset was prepared, which refers to the centre pixel of each ESU and corresponding LCC and LAI values.

3.4.2 Experimental setup

The SI assessment toolbox has been applied to evaluate the retrieval accuracies for different SI formulations, band settings, and fitting functions using the SPARC calibration dataset. Only the two-band results for the most common SI formulations are illustrated, being SR and ND SI combination types. Correlating the possible HyMap two-band SIs according to these combination types and by using five fitting functions (linear, exponential, power, logarithmic and polynomial of the second degree), the SI assessment toolbox did finally automatically assess 155000 (125x124x5x2) SI regression functions for each biophysical parameter. This typical assessment run did not take longer than a few minutes on a 32 bit PC (Windows 7 Intel(R) Core(TM) 2 Quad CPU, 2.4 GHz, 3.00 GB RAM processor).

3.5 Results & Discussion

3.5.1 Optimized SIs

For each SI type, Table 4.4 provides goodness-of-fit statistics (RMSE, NRMSE and r^2) for the best performing calibrated SI models together with its wavelength specifications and regression equation, ranked according to r^2. The following observations can be made:

- With regard to SI formulations, no strong evidence has been found that the widely used ND formulation outperforms the less popular SR formulation. Though, it is also recognized that the main argument for using ND types of indices is not the improved correlation performance in comparison to SRs, but rather the (at least to some extent) improved comparability of different observation times/dates and the possible reduction of effects of varying illumination intensity (shades etc.), due to the inherent normalization of the value range [Rouse *et al.*, 1974]. The ND outperformed the SR for both LAI and LCC only when a second order polynomial is applied as a fitting function. However, when a conventional linear regression is applied, ND performs similar (LCC) or superior (LAI) to SR.

- While the polynomial fit outperforms to some extent the linear regression fitting scenario, the linear fit model is nonetheless quite accurate as well. For instance, the best linear regression fit outperforms the more sophisticated exponential and power curve fitting approaches significantly for both, LAI and LCC.

- For the majority of the most accurate regression functions, about the same spectral band locations were assessed as being optimal for biophysical parameter retrieval. For LAI, two-band formulations are optimal at 570, 555, 539 and 707 nm with the second band at 2453 or 2421 nm. For LCC two-band locations at 692 nm and 632 nm with the second band at 1661-1698 nm or 1200-1340 nm performed best.

It is noteworthy that, for a majority of the most accurate SI models sensitive with respect to LCC, the wavelength location of 692 nm is ranked as the best performing one. This is consistent with earlier studies [e.g. Gitelson *et al.*, 2003; Verrelst *et al.*, 2010], demonstrating that the spectral interval around 700 nm is less prone to saturation than the location at 660 nm (which coincides with the location of the chlorophyll pigment's main absorption peak). Therefore, the far-red at 692 nm ensures a good sensitivity when detecting moderate and high LCC values.

Although Table 4.4 suggests that the optimal fitting functions lead to similar performances in terms of r^2, a closer look may be necessary to assess their impact on the efficiency in mapping applications. Figure 3.7 provides scatterplots of the best performing SI models; the distinct behavior of each fitting function can be observed. For instance, exponential and power function fits can lead to spurious parameter value retrievals at high SI values (or at low SI values in case of the SR-LCC combination). The most accurate fitting function, i.e. the second order

TABLE 3.1: Goodness-of-fit measures (RMSE, NRMSE, r^2) for the best calibrated SI models per SI type and their corresponding used bands and fitting function for LAI and LCC retrieval. In this table, the SI types are sorted according to a decreasing r^2. B_1: spectral band 1, B_2: spectral band 2.

Index type	Fitting function	B_1 (nm)	B_2 (nm)	SI model	abs. RMSE	NRMSE (%)	r^2
LAI:							
ND	polynomial	570	2453	LAI= -1.366ND^2 + 6.154ND + 1.976	0.61	12.51	0.83
SR	logarithmic	555	2453	LAI= 2.291 log(SR) + 1.924	0.61	12.10	0.83
ND	linear	555	2453	LAI= 5.023ND + 1.922	0.61	12.80	0.83
SR	polynomial	539	2421	LAI= -0.621SR^2 + 3.754SR -0.760	0.63	15.58	0.82
SR	linear	539	2421	LAI= 1.751SR + 0.504	0.70	16.32	0.77
SR	exponential	2118	2171	LAI= 0.000238e$^{8.783SR}$	0.72	11.85	0.77
ND	exponential	707	2453	LAI= 0.712e$^{2.721ND}$	0.75	13.13	0.77
ND	power	707	2453	LAI= 4.670ND$^{0.548}$	0.77	19.06	0.77
SR	power	539	2421	LAI= 1.951SR$^{1.112}$	0.82	14.64	0.76
ND	logarithmic	462	524	LAI= 3.888 log(ND) + 6.947	0.85	18.23	0.74
LCC:							
ND	polynomial	692	1686	LCC= -227.499ND^2 + 281.410ND - 36.924	4.21	7.75	0.93
SR	polynomial	692	1661	LCC= -8.114ND^2 + 60.997ND - 61.553	4.30	7.64	0.92
ND	logarithmic	692	1340	LCC= 43.578 log(ND) + 63.272	4.70	10.18	0.91
SR	linear	692	1340	LCC= 7.455SR + 4.707	4.70	20.24	0.91
ND	linear	692	1340	LCC= 90.290ND -14.808	4.21	10.75	0.90
SR	logarithmic	692	1200	LCC= 26.447 log(ND) - 2.990	5.53	11.54	0.87
SR	exponential	692	1215	LCC= 91.164e$^{-4.235SR}$	6.00	11.25	0.87
ND	power	632	1698	LCC= 127.581ND$^{2.119}$	28.06	365.5	0.87
ND	exponential	632	1257	LCC= 2.110e$^{3.827ND}$	6.64	12.80	0.84
SR	power	707	723	LCC= 7.281SR$^{2.938}$	8.28	13.19	0.76

polynomial, is probably the most difficult one to interpret. This is illustrated for the LCC-SR case where the peak of the polynomial fit goes through the LCC data clumping between 40-50 μg/. It implies that a SR above 4 leads to a lower LCC estimate. This type of erratic fits indicates that polynomial fitting - despite the high r^2 and low RMSE - should not be recommended for biophysical parameter retrieval.

Although dating from the times where the campaigns mostly were limited to the silicon-based CCD detector range (400 - 1000 nm), the saturation effect is often reported and is considered as one of the major critical points when using a red-NIR formulation (e.g. NDVI) for LAI retrieval [Verstraete and Pinty, 1996; Stenberg *et al.*, 2004]. A noteworthy observation is that here LAI hardly shows saturation for a linear model fit. Presented results suggest that the use of a band in the SWIR range (2453 nm) causes the saturation effect to be close to negligible. Hence, such a linear regression should be preferred above using the NIR, because saturation violates the assumptions of regression residuals around zero in order for the results to be valid [Draper and Smith, 1998]. Also [Schlerf *et al.*, 2005; Darvishzadeh *et al.*, 2009] found that vegetation indices (VIs) based on bands in the SWIR range tend to be almost linearly related to LAI without showing saturation even at high LAI values (LAI=6). However, it should also be remarked that wavelengths at the end of HyMap sensor' SWIR range (i.e. beyond 2400 nm) face a relatively low signal-to-noise ratio (SNR), i.e. below 400 [Cocks *et al.*, 1998]. Results in this region may therefore be biased by noise influence.

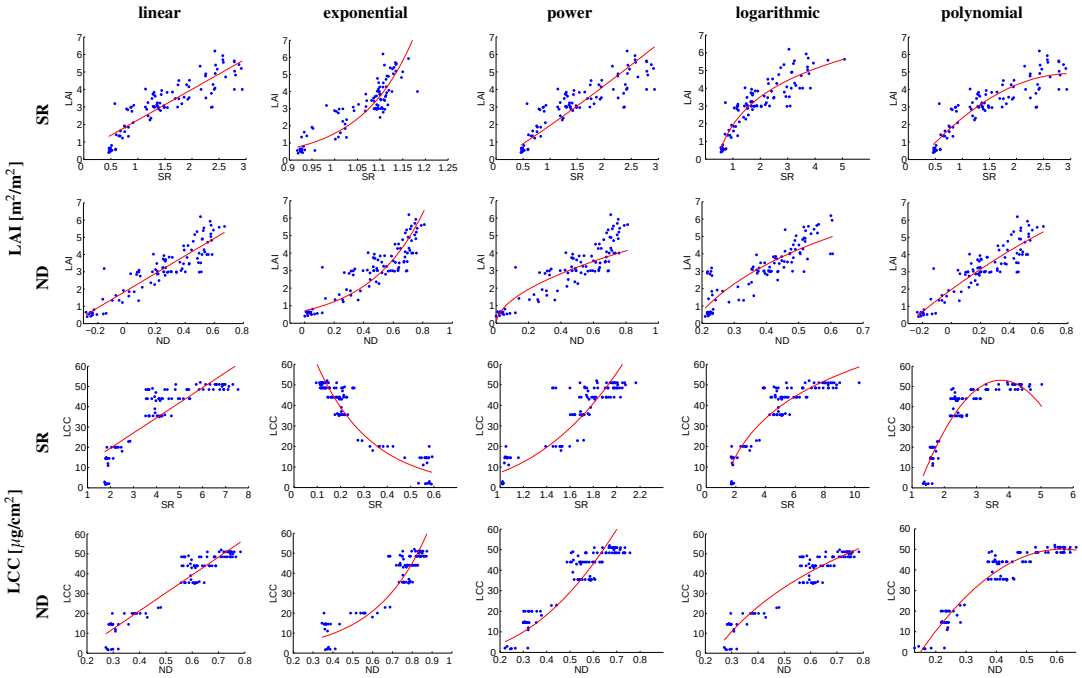


FIGURE 3.7: Fitting functions (red lines) depicting the regression of LAI [m^2/m^2] and LCC [$\mu g/cm^2$] in function of SR and ND formulations as they could be derived from the calibration dataset using optimized band combinations.

3.5.2 Correlation matrices

To fully take account of the prediction efficiency of the spectral domain, figure 8 illustrates all two-band SR and ND combinations according to r^2 correlation between measured and estimated values for the LAI and LCC biophysical parameters for different fitting functions. The correlation matrices for B_1 (430-2490 nm) versus B_2 (430-2490 nm) clearly distinguish the following spectral regions: the visible region (450-700 nm), the red-edge region (700-750 nm), the NIR region (750-1340 nm), the shortwave SWIR region (1470-1800 nm) and the longwave SWIR region (1970-2453 nm). Based on the 2D correlation plots in figure 8, the spectral regions leading to the most accurate indices to retrieve LAI and LCC can be unequivocally identified. Based on tables 1 and 2, the following observations can be made for two-band (B_1, B_2) SIs:

- SR and ND formulations lead to the same spectral regions exhibiting strong correlations between the LAI / LCC biophysical parameters and the two-band SIs. For instance, similar patterns for B_2 in the visible spectral region can be observed. However, the ND power and logarithmic functions fail as good fitting functions for several band combinations. Specifically, they provide lower regression accuracies for B_1 (740-1000 nm) and B_2 (1400-2200 nm).

- Similar patterns with strong and low correlation values are found across the different

curve fitting functions assessed. This suggests that the major impact on retrieval accuracy in EO does not originate from the chosen curve fitting but rather from the spectral dimension. Though, local differences do occur. These local differences are particularly pronounced along exponential, power and logarithmic fittings, characterized by quite low correlation values. Hence, the following sections are limited to the discussion of linear and polynomial curve fits.

- For LAI, Figure 8 shows that the most important spectral region for retrieval is the 539-707 nm region for the B_1band, whereas B_2 is located in the longwave SWIR. Notice the hourglass shape in the SWIR region. A second spectral region of maximal r^2 for $B1$ is the 1300-1500 nm region as well as the region at 1400-1800 nm. These regions are spectrally quite broad, meaning that the spectral indices do not require very narrow and hence precisely positioned spectral bands, at least for the samples that were used in this study for the generation of the empirical relationships.

- For LCC, Figure 3.8 shows that the most important spectral region for retrieval is the visible region, particularly the red region until the red edge. B_2 is spectrally located in the shortwave SWIR. Another sensitive zone, having the same B_1 (around 692 nm) as in the first case, has its second band B_2 in the NIR range as well as in the longwave SWIR. Note that these spectral areas are covered by broad spectral bands.

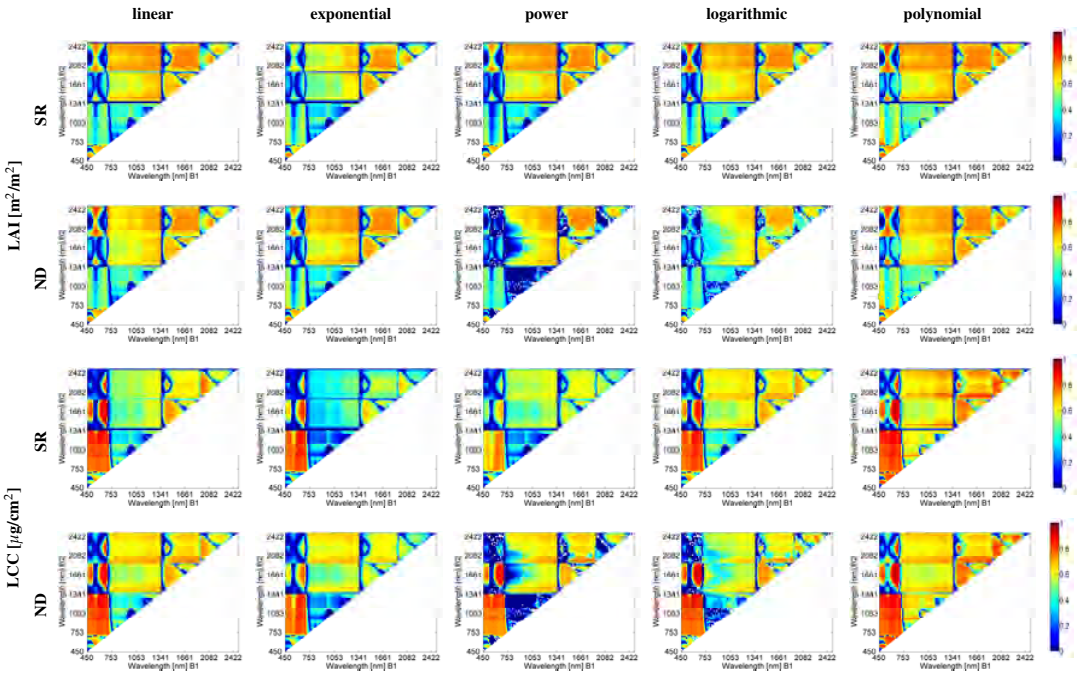


FIGURE 3.8: 2D r^2 correlation matrices calculated from measured vs. estimated values for the calibration results of the LAI [m^2/m^2] and LCC [$\mu g/cm^2$] retrievals according to SR and ND formulations and for different fitting functions. The color bars on the right of this figure give the values of the plotted r^2 in the different graphs.

Considering only the linear and polynomial fitting functions, it is of interest that the regions that were found to be optimal for LAI retrieval in the longwave SWIR region (notice the hourglass shape centered around the green; 550-570 nm) for B_2 oppositely lead to erroneous LCC retrievals. This spectral region can hence be applied for LAI retrieval without being influenced by its sensitivity for LCC. For LCC on the other hand, the most optimal B_1 spectral regions are the NIR and shortwave SWIR in combination with red bands for B_2. Hence, a clearcut distinction can be observed for the sensitive spectrally important regions for the retrieval of LAI versus LCC. Retrieval confusion between LAI and LCC is thereby avoided. The correlation matrices additionally suggest that the SWIR region is preferred above the NIR region, ensuring independent retrievals for LAI and LCC. For instance, the conventional NDVI spectral bands (i.e., B_1 = 680 nm, B_2 = 800 nm) do not appear to be the most optimal ones for LAI nor LCC retrieval.

3.5.3 LAI & LCC mapping

The final step presented in this paper involves LAI and LCC mapping. ARTMO's SI assessment toolbox helps the user to select a SI model (e.g. the most accurate one) and to apply it using EO imagery. The impact of the SI model type on biophysical parameter mapping thus can be compared. For this purpose, it was opted to map LAI and LCC based on the available HyMap data for the most accurate SI model only, being a linear, logarithmic and polynomial of the second order function. Exponential and power functions have not been considered for mapping, not only because of their considerably less performing calibration results, but also because these functions do not consistently lead to maps in conformance with ground measurements. To allow for a comprehensive visual comparison between retrievals and ground samples, LAI maps are color scaled for values between 0 and 6 m^2/m^2. Identically, LCC maps are color scaled for values between 0 and 60 $\mu g/cm^2$. These maps, as presented in Figure 3.9, lead to the following observations:

- For LAI, both, the ND polynomial and ND linear regression models, lead to the same calibration result (RMSE=0.61; r^2=0.83) and accordingly to very similar LAI maps. This example provides confidence that a linear regression is able to deliver adequate LAI maps, when appropriate spectral bands are selected; i.e., green spectral region (B_1: 539-570 nm) in combination with the longwave SWIR (B_2: 1970-2453 nm). Rapid saturation of the LAI in function of the SI is thereby avoided. The map shows areas with a high LAI quite clearly (irrigated circular fields). Areas with a low LAI - senescent crops to bare soil surfaces - elicit a zero value LAI (white color code).

- Conversely, when using SR for LAI mapping, prominent spatial differences appeared for the linear versus polynomial regression models. Particularly, a linear regression exhibits problematic behavior, since LAI variations for senescent regions significantly deviate from the other LAI mapping results. A logarithmic fit function seems to lead to more consistent LAI maps. Nevertheless, although the ND type yields considerably poorer

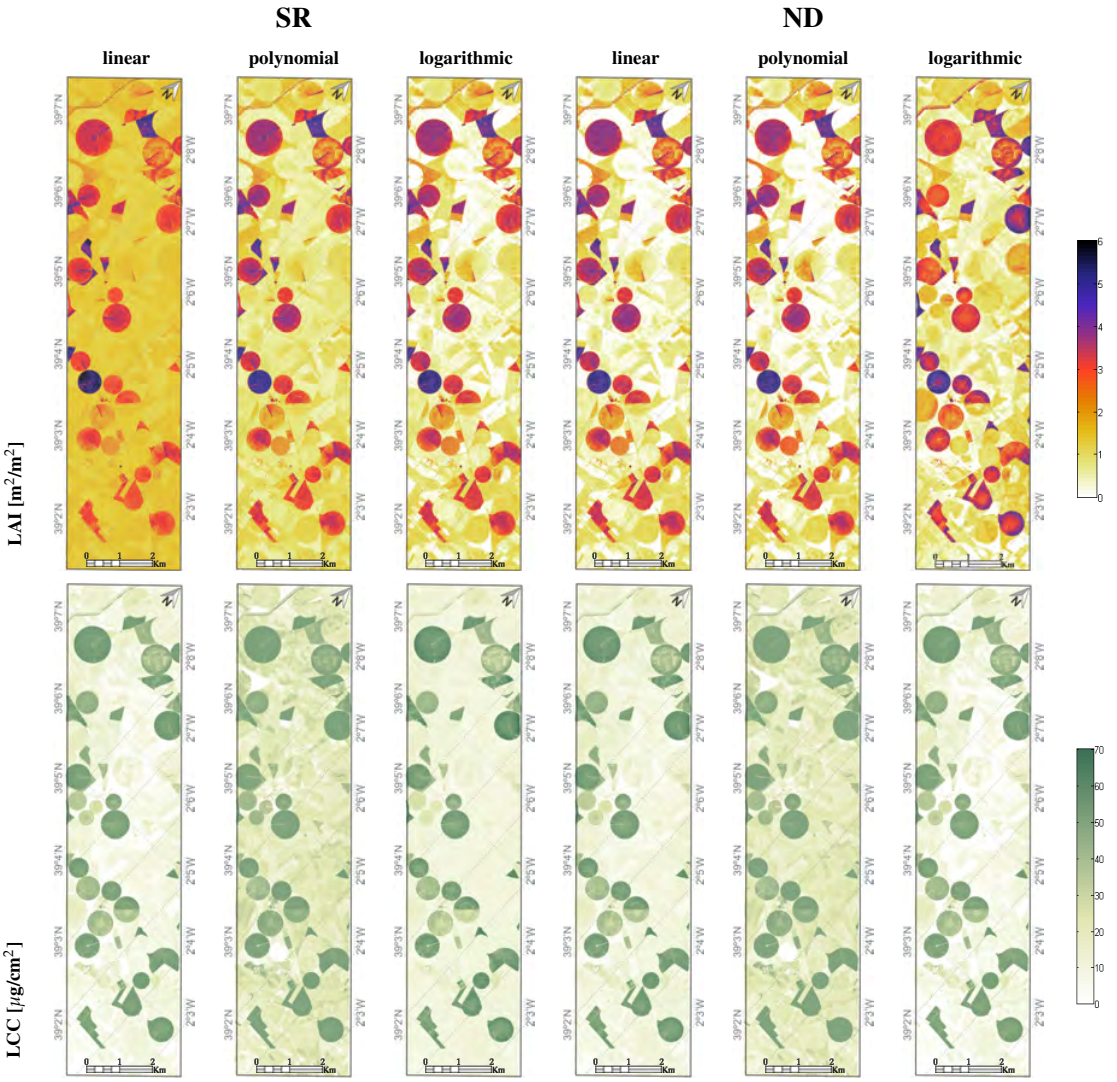


FIGURE 3.9: LAI [m^2/m^2] and LCC [$\mu g/cm^2$] maps based on SR and ND formulations and linear, polynomial and logarithmic fitting functions. The imagery originates from a HyMap flight line over the Barrax agricultural area in Spain. The color bars at the right side of this figure give the values of the LAI and LCC mapped variables.

calibration results, both maps yield similar LAI patterns. The spatial similarity of the three best performing fitting functions (SR, logarithmic, ND linear and ND polynomial) provide confidence in the realism of the maps produced.

- With regard to LCC mapping, the spatial similarity between the SR and ND polynomial fits and the degree of detail is remarkable. However, interpretation becomes difficult for areas earlier identified as having low LAI values in the LAI map and showing high LCC values in the LCC map. One interpretation may be that the nature of a polynomial of second degree fit leads to retrievals of high LCC values at near-zero SI values. Another

problem might be related to the difficulty of upscaling. For the areas where the canopy is very sparse, i.e. the LAI is low, it becomes harder to perceive the chlorophyll pigments in the leaves with a moderate spatial resolution of 1-4m, even if the chlorophyll concentration in the small leaves is relatively high.

- Logarithmic ND fits, along with linear SR and ND fits, elicit the most realistic LCC patterns; with spatial patterns conform to those of the best LAI maps.

Quite remarkably, the different fitting functions lead to considerable local spatial variation. This can be problematic, since it suggests that the prediction efficiency results on their own may be insufficient as an assessment criterion for the quality of the mapping results. An important aspect to be taken into consideration is that an independent validation dataset likewise gives limited assessment information of the complete map due to a typically low spatial cover within a remote sensing image. Hence, the question arises, which maps - as a whole - are quantitatively most trustworthy? This rather enigmatic problem area reveals one of the main weaknesses of SI based biophysical parameter mapping e.g., the absence of an uncertainty (or accuracy) metric. Uncertainty estimates on a pixelwise basis provide an indication of retrieval accuracy and are nowadays a prerequisite in processing chains. In part, this is why more sophisticated regression algorithms have been developed, which do provide associated uncertainty estimates, e.g. in the emerging field of Bayesian approaches such as Gaussian processes regression [Verrelst *et al.*, 2012b]. Nevertheless, if any generalization can be deduced from the maps presented, for LAI the ND polynomial (second degree) fit leads to the most consistent results, while for LCC it is rather the ND logarithmic fit that leads to the most consistent results. Conversely, the dissimilarity across the SR-developed maps suggests that it is a less successful formulation for biophysical parameter mapping. The maps provided here suggest that a normalized difference (ND) formulation can therefore be considered as a more preferable strategy for biophysical parameters mapping.

3.5.4 Towards a new generation of spectral indices

With ARTMO's SI assessment toolbox extensive and systematic SI performance assessments can be conducted, including band combinations, formulations and fitting functions. Its semi-automatic nature enables the reduction of time and resources for analysis and interpretation. By analyzing all bands against each other in a correlation matrix, ARTMO helps identifying redundant bands and overcoming the Hughes' phenomenon or "curse of dimensionality" [Thenkabail *et al.*, 2011]. The case presented in this paper, shows the analysis of all possible 2-band combinations for SR and ND formulation by using four types of curve fittings on hypespectral data. While the relevance of the visible and red edge spectral regions is widely known for LCC and LAI estimation, and has since long been documented for vegetation mapping applications [Filella and Penuelas, 1994; Thenkabail *et al.*, 2002; Delegido *et al.*, 2011; Clevers and Gitelson, 2013; Delegido *et al.*, 2013], the sensitivity in the SWIR is particularly remarkable. Especially for LAI retrieval, but also for that of LCC, the SWIR region was found to be an optimal region

for biophysical parameter mapping, each time in combination with the visible region; for LAI, green (B_1) with longwave SWIR (B_2), for LCC red (B_1) with shortwave SWIR (B_2). The green-SWIR indices do not show an apparent saturation in the value range considered for LAI, so that the transfer function can be defined with a linear regression model. Similar improvements have been reported by previous studies [Brown *et al.*, 2000; Lee *et al.*, 2004; Schlerf *et al.*, 2005; Darvishzadeh *et al.*, 2009; Gonsamo, 2011; Heiskanen *et al.*, 2013]. In the SWIR, leaf reflectance is mainly controlled by water absorption, although leaf biochemical components such as proteins, cellulose, and lignin also contribute to a lesser extent to leaf absorption [e.g. Curran, 1989].

At canopy level, the main drivers of LAI and LCC spatial variation are irrigation regimes at the Barrax test site, leading to high LCC parcels (when irrigated and due to high water absorption) in contrast to non-irrigated low LCC parcels, which are senescent during summertime. Canopy structural differences depending on the crop type are a source of variation related to both biophysical parameters as well [Verrelst *et al.*, 2012a]. The SWIR is sensitive to variations in complex and dissimilar growth stages and growing conditions [Thenkabail *et al.*, 2000, 2004]. Moreover, the SWIR spectral region has the advantage of being sensitive to vegetation, while being less influenced by atmospheric aerosols. This avoids the increase in parameter retrieval uncertainty associated with the atmospheric correction of remotely sensed reflectances required to compensate for aerosol scattering impacts on top-of-atmosphere reflectances, specifically playing a role in the visible part of the solar spectrum [Gonsamo, 2011].

When comparing literature findings, where similar correlation matrices are presented, it is remarkable that identical band combinations are rarely reported for studies based on the use of narrowband (hyperspectral) imaging [Gonsamo, 2011; Heiskanen *et al.*, 2013; Mariotto *et al.*, 2013]. This leads us to suggest that established SI models are more often than not case specific and thus are likely to perform sub-optimal for local hyperspectral mapping applications. Quite often, the traditional red-NIR combination is applied. At the same time, suboptimal SI formulations or regression function selections also lead to less optimal biophysical parameter retrievals. An in-depth analysis is therefore not a luxury. The SI assessment toolbox in ARTMO, as described in this paper, offers a systematic, semi-automatic streamlined and complete procedure for the selection as well an assessment of the most accurate and sensitive SI formulations for biophysical parameter retrieval based on hyperspectral image datasets. The SI model(s) selected can hence confidently be applied for the mapping of biophysical parameters retrieved from hyperspectral imagery.

Finally, it is foreseen that the ARTMO SI assessment toolbox will serve future sensitivity studies. This may lead to a generation of new SIs based on the following research objectives:

- Development and evaluation of generic SIs by using RTMs including viewing and solar geometry. ARTMO includes turbid (e.g., SAIL) as well as explicit 3D RTM's (e.g., FLIGHT). Evidently, the validity of the most optimal models will be assessed using in situ validation data.

- New types of index formulations will be developed and tested across the full optical spectrum. Only SR and ND indices have been assessed here. Yet, alternative mathematical formulations are to be tested as well, e.g. the formulations used for soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI) or more complex ones.
- Although in this paper only two-band formulations have been assessed, the SI toolbox offers the option for a systematic analysis of a full optical spectrum of band combinations with SI formulations of up to ten different bands in the range of 400 to 2500 nm. Since there is no reason to believe that two-band indices lead to the most successful SI models, this multiple band approach may lead to the development of more accurate and sensitive spectral indices.

Regardless of the above-mentioned research objectives, essentially, the SI assessment toolbox provides all necessary tools for quality assurance and quality control (QA/QC), and for a rapid and comprehensive biophysical parameter mapping. The toolbox is intuitive and can assist significantly in precision farming and landscape ecology monitoring applications, thereby allowing to optimize SIs, even per land cover type.

3.6 Conclusions

A newly developed Spectral Indices (SI) assessment toolbox in the ARTMO modeling environment, enables the analysis and assessment of the accuracy of an indefinite number of SI models. Basically, the toolbox offers a systematic but still empirical approach for the assessment of all possible 2, 3 or 4-band SI formulations. Datasets can be partitioned into calibration and validation subsets. These datasets may originate from simulations, e.g. as generated by the optical radiative transfer models in ARTMO, or from field campaigns. Several options have been included in the SI assessment approach, amongst which: (1) The addition of noise and the possibility to select fitting functions (e.g., linear, exponential, power or polynomial functions); (2) the SI toolbox virtually allows for any type of spectral index model to be formulated and evaluated using up to ten spectral bands; and (3) the possibility to assess and apply SIs per land cover class.

Using HyMap data calibrated with field measured data, the predictive power of generic narrowband spectral indices in the 430-2490 nm range to quantify LAI and LCC has been assessed. For LAI retrieval, the B_1 in the green (B_1: 539, 555, 570 nm) combined with longwave SWIR (e.g., B_2: 2421, 2453 nm) has been evaluated as the most accurate approach (RMSE: 0.61; r^2: 0.83). For LCC retrieval, the B_1 in the far-red (692 nm) combined with the B_2 NIR (e.g., 1340 nm) or shortwave SWIR (e.g., 1661, 1686 nm) has been evaluated as the most accurate approach (RMSE: 4.21-4.70, r^2: 0.91- 0.93). In either case, the identification of the SWIR rather than the conventional NIR as an important spectral region reinforces our suggestion that the ARTMO SI assessment toolbox significantly facilitates the development of new and especially better performing spectral indices for biophysical parameter mapping applications.

4

LUT-based RTM inversion

Contents

This chapter is based on:

Juan Pablo Rivera Caicedo, Jochem Verrelst, Ganna Leonenko and José Moreno (2013)
Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model
Remote Sensing 5(7): 3280-3304.
DOI:10.3390/RS5073280

4.1 Abstract

Lookup-table (LUT)-based radiative transfer model inversion is considered a physically-sound and robust method to retrieve biophysical parameters from Earth observation data but regularization strategies are needed to mitigate the drawback of ill-posedness. We systematically evaluated various regularization options to improve leaf chlorophyll content (LCC) and leaf area index (LAI) retrievals over agricultural lands, including the role of 1) cost functions (CFs), 2) added noise, and 3) multiple solutions in LUT-based inversion. Three families of CFs were compared: *information measures*, *M-estimates* and *minimum contrast methods*. We have only selected CFs without additional parameters to be tuned, and thus they can be immediately implemented in processing chains. The coupled leaf/canopy model PROSAIL was inverted against simulated Sentinel-2 imagery at 20 m spatial resolution (8 bands) and validated against field data from the ESA-led SPARC (Barrax, Spain) campaign. For all 18 considered CFs with noise introduction and opting for the mean of multiple best solutions considerably improved retrievals; relative errors can be twice reduced as opposed to those without these regularization options. *M-estimates* were found most successful, but also data normalization influences the accuracy of the retrievals. Here, best LCC retrievals were obtained using a normalized 'L_1-estimate' function with a relative error of 17.6% (r^2: 0.73), while best LAI retrievals were obtained through non-normalized 'least-squares estimator' (LSE) with a relative error of 15.3% (r^2: 0.74).

4.2 Introduction

Leaf area index (LAI) and leaf chlorophyll content (LCC) are essential land biophysical parameters retrievable from optical Earth observation (EO) data [Whittaker and Marks, 1975; Lichtenthaler, 1987; Malenovský *et al.*, 2012]. These parameters give insight in the phenological stage and health status (e.g., development, productivity, stress) of crops and forests [Lichtenthaler *et al.*, 1996; Sampson *et al.*, 2003]. The quantification of these parameters over large areas has become an important aspect in agroecological, environmental and climatic studies [Dorigo *et al.*, 2007]. At the same time, remotely sensed observations are increasingly being applied at a within-field scale for dedicated agronomical monitoring applications [Gianquinto *et al.*, 2011; Delegido *et al.*, 2013].

For the last few decades, various space-based LAI and LCC retrieval approaches have been proposed (see review in [Dorigo *et al.*, 2007]), some of them eventually led to operational retrieval strategies. Particularly LAI proved to be successful parameter for being operationally and globally retrieved at resolutions of 250 m to 1 km (e.g. MODIS, CYCLOPES and Geoland2 products) [Myneni *et al.*, 2002; Baret *et al.*, 2007]. Moreover, a lot of efforts are being undertaken to generate a global LAI product at a 30-m Landsat scale [Ganguly *et al.*, 2012]. However, operationally retrieved land LCC products are scarce. Until the loss of the ENVISAT spacecraft LCC maps were routinely delivered at a medium spatial resolution through the MERIS terrestrial chlorophyll index [Dash and Curran, 2004] or through a trained neural net [Bacour *et al.*, 2006]. But routinely generated LCC maps originated from high spatial resolution images (e.g. $\leq$ 20 m) are absent until now, although some operational approaches have been proposed for SPOT data [Houborg and Boegh, 2008; Houborg *et al.*, 2009]. Meanwhile, new generation of high resolution (i.e. 10-60 m) land monitoring EO missions are being constructed to be launched such as the forthcoming superspectral Sentinel-2 (S2) mission and hyperspectral missions such as Enmap [Stuffler *et al.*, 2007], PRISMA [Labate *et al.*, 2009] and HyspIRI [Roberts *et al.*, 2012]. Such unprecedented richness of high spectral and spatial resolution data streams makes the availability of robust retrieval methods more important than ever.

To implement a method in an operational processing chain the method should be able to deliver accurate estimates with easy implementation in practice. Inversion of physically-based canopy radiative transfer models (RTMs) against actual EO data is generally considered as one of the most robust approaches to map biophysical parameters over terrestrial surfaces [Dorigo *et al.*, 2007; Darvishzadeh *et al.*, 2008]. But this approach is not straightforward. According to Hadamarad postulates, mathematical models of physical phenomena are mathematically invertible if the solution of the inverse problem to be solved exists, is unique and depends continuously on variables [Knyazikhin *et al.*, 1998a]. Unfortunately this assumption is not met. In fact, the inversion of canopy RTMs is by nature an ill-posed problem mainly for two reasons [Durbha *et al.*, 2007]: on the one hand, several combinations of canopy biophysical and leaf biochemical parameters have a mutually compensating effect on canopy reflectance thus leading to very similar solutions. On the other hand, model uncertainties and simplifications (e.g. 1D nature of some models) may induce large inaccuracies in the modelled canopy reflectance.

Several strategies have been proposed to circumvent the drawback of ill-posedness, including Lookup-table (LUT)-based inversion strategies [Combal *et al.*, 2003; Darvishzadeh *et al.*, 2008; Knyazikhin *et al.*, 1998a; Richter *et al.*, 2009; Weiss *et al.*, 2000], hybrid approaches in which LUTs are generated to feed machine learning approaches [Weiss and Baret, 9991; Walthall *et al.*, 2004; Fang and Liang, 2005; Bacour *et al.*, 2006; Durbha *et al.*, 2007; Qu *et al.*, 2008], or LUT-based iterative numerical optimization methods [Jacquemoud *et al.*, 1995]. They all have their strengths and weaknesses in specific situations. But the main advantage of LUT-based inversion approaches is that it can be fast because the most computationally expensive part of the inversion procedure is completed before the inversion itself [Dorigo *et al.*, 2007].

LUT-based inversion in its essential form, i.e. direct comparison of LUT spectra against an observed spectra through a cost function (CF), also in some cases known as distance, merit function, metric or divergence measure, is part of the majority of applied inversion approaches. Such a function yields a value for one or multiple biophysical parameters by minimizing the summed differences between simulated and measured reflectances for all wavelengths [Knyazikhin *et al.*, 1998a]. Various regularization strategies have been proposed to further optimize the robustness of the estimates: 1) the use of prior knowledge about model parameters [Baret and Buis, 2008; Combal *et al.*, 2002, 2003; Darvishzadeh *et al.*, 2008; Dorigo *et al.*, 2009], 2) the use of multiple best solutions in the inversion (instead of the single best solution) [Combal *et al.*, 2002; Koetz *et al.*, 2005; Richter *et al.*, 2009, 2011], 3) adding noise to account for uncertainties attached to measurements and models [Koetz *et al.*, 2005; Richter *et al.*, 2009, 2011], and, 4) the combination of single variables into synthetic variables such as canopy chlorophyll content [Weiss *et al.*, 2000; Bacour *et al.*, 2006; Dorigo *et al.*, 2007; Darvishzadeh *et al.*, 2012]. Nevertheless, aforementioned approaches face limitations when implementing them into a more operational context.

First, in the majority of these studies root mean square error (RMSE) was used as CF between simulated and measured spectra. However, in case of outliers and nonlinearity, the residuals are distorted and therefore the key assumption for using RMSE (maximum likelihood estimation with the Gaussian noise) is violated [Leonenko *et al.*, 2013]. The latter authors suggested that alternative CFs may provide a more robust way to estimate biophysical parameters since they allow retrievals for cases where errors are not normally distributed and allow dealing with nonlinear high-parametric problems. Verrelst *et al.* [2014b] recently demonstrated that alternative CFs, in combination with aforementioned regularization strategies, can considerably improve biophysical parameters retrievals. Yet only three alternative CFs - out of more than 60 - were extensively evaluated so far, which leaves an urgent need to evaluate the performance of other promising CFs.

Second, the majority of these studies focus on a specific vegetation type such as croplands, often identified as a land cover class within an image [Richter *et al.*, 2009, 2011; Atzberger and Richter, 2012]. However, this approach can turn out problematic when applying LUT-based inversion over larger areas. While land cover classification schemes help to split the problem into sub-domains for which prior information is attributed separately [Chen *et al.*, 2002] it assumes that up-to-date knowledge of land cover types is available at high spatial resolution, which is usually not the case in an operational context. These limitations imply that alternatives

have to be sought to enable full-scene LUT-based inversion over heterogeneous areas without relying on predefined land cover maps.

Henceforth, for the sake of realizing robust LCC and LAI retrievals at high spatial resolution, the aim of this work is to invest in a retrieval processing chain that combines LUT-based inversion with different CFs and regularization options. Simulated Sentinel-2 data at 20 m resolution will be used for this exercise but in principle the inversion schemes can be applied to any optical multispectral EO data. This brings us to the following objectives: to evaluate the role of 1) cost functions, and 2) regularization options in LUT-based inversion strategies, such as: *i*) added noise, *ii*) multiple best solutions and *iii*) the role of data normalization.

4.3 Methodology

4.3.1 Cost functions

The estimation of biophysical parameters from satellite data is hampered by uncertainties of very complicated and different nature in every particular study. In many cases it is difficult to study these errors since they have unknown magnitude and distribution. On this basis, we opted to evaluate multiple cost functions that have been introduced in Leonenko *et al.* [2013] and then search for the optimal ones that minimizes the errors between simulated and real reflectances. Different CFs deal with different classes of distributions which allows us to deal with outliers and non-linearities distort in a better manner than commonly used least squares estimation (LSE) distance. Therefore it can provide more accurate results for estimated biophysical parameter. Note, that LSE corresponds to the maximum likelihood estimation with the Gaussian noise.

These distances/metrics came from different fields of mathematics, statistics and physics and they all represent 'closeness' between two functions but the nature of these functions can be different. For this reason, these metrics have been divided into three broad families based on physical properties of functions in consideration: *information measures*, *M-estimates* and *minimum contrast method*. The detailed description of these families can be found in Leonenko *et al.* [2013]. However most of these functions require one or two parameters to be tuned. That may hamper their use in operational processing chains, therefore in this work we chose to analyse only those CFs without additional parameters. Some brief information about them is provided below.

To describe the problem in a statistical way we suppose that $D[P,Q]$ represents a distance between two functions, where $P = (p(\lambda_1), \cdots, p(\lambda_n))$ is satellite and $Q = (q(\lambda_1), \cdots, q(\lambda_n))$ is LUT correspondent reflectances and $\lambda_1, \cdots, \lambda_n$ represent n bands. Also we define η as a biophysical parameter of our interest that we need to estimate (for example LAI or LCC) and $\bar{\zeta} = (\zeta_1, ..., \zeta_r)$ is a vector of other parameters that our function depends on but do not contribute to minimization equation (1) . The classical statistical method of inversion (or estimation) and

finding optimal parameter η^*) can be formulated as a semi-parametric problem

$$\eta^* = arg \min_{\eta} sup_{\bar{\zeta}} D[P(\lambda_j),\ Q_i(\lambda_j)], æ = 1, ..., n. \tag{4.1}$$

The purpose is to find the best estimate for η^* by solving the minimization problem (1) using different statistical distances, as those presented below.

Information measures

This family of measures, also referred to as 'divergence measures', is based on minimization of distances between two probability distributions. In this case reflectances are considered as probability distributions and normalization is required (sum of probabilities is 1) prior to numerical application. Note that normalization has been performed on LUT reflectances as well. This family was first introduced by Kullback and Leibler (KL) [Kullback and Leibler, 1951] and refers to the concept of information divergences, which are non-symmetric between two distributions P and Q. This concept has been further extended in many directions since its initial application in decoding schemes and signal processing and now plays an important role in multimedia classification, neuroscience and cluster analysis. More details and classifications can be found in [Pardo, 2006]. Here is the list of the divergencies that we have used for our study:

1. This measure is called the Kullback Leibler divergence and it also corresponds to the maximum likelihood distance in probabilistic space:

$$D[P, Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) ln \left(\frac{p(\lambda_l)}{q(\lambda_l)} \right). \tag{4.2}$$

2. This measure is called Pearson chi-square:

$$D[P, Q] = \sum_{\lambda_1=1}^{\lambda_n} \frac{(q(\lambda_l) - p(\lambda_l))^2}{p(\lambda_l)}. \tag{4.3}$$

3. Squared-Hellinger measure:

$$D[P, Q] = \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) \left(\sqrt{\frac{p(\lambda_l)}{q(\lambda_l)}} - 1 \right)^2 = \sum_{\lambda_1=1}^{\lambda_n} (\sqrt{p(\lambda_l)} - \sqrt{q(\lambda_l)})^2 \tag{4.4}$$

4. Neyman chi-square divergence:

$$D[P, Q] = \sum_{\lambda_1=1}^{\lambda_n} \frac{(p(\lambda_l) - q(\lambda_l))^2}{q(\lambda_l)}. \tag{4.5}$$

5. Jeffreys-Kullback-Leibler:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} (p(\lambda_l) - q(\lambda_l)) \, (ln(p(\lambda_l)) - ln(q(\lambda_l))) \, . \tag{4.6}$$

6. K-divergence of Lin:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) ln \left(\frac{2p(\lambda_l)}{p(\lambda_l) + q(\lambda_l)} \right) . \tag{4.7}$$

7. L-divergence of Lin is a symmetric version of K-divergence:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) \ln(p(\lambda_l) + q(\lambda_l) \ln(q(\lambda_l) - (p(\lambda_l) + q(\lambda_l)) \ln \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) \tag{4.8}$$

8. The harmonique Toussaint measure:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} \left(p(\lambda_l) - \frac{2p(\lambda_l)q(\lambda_l)}{p(\lambda_l) + q(\lambda_l)} \right) . \tag{4.9}$$

9. The negative exponential disparity measure:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) \left(exp(-\frac{p(\lambda_l) - q(\lambda_l)}{q(\lambda_l)}) - 1 \right) . \tag{4.10}$$

10. Bhattacharyya divergence:

$$D[P,Q] = -log \left(1 + \sum_{\lambda_1=1}^{\lambda_1} \sqrt{p(\lambda_l)q(\lambda_l)} - \frac{1}{2}(p(\lambda_l) + q(\lambda_l)) \right) . \tag{4.11}$$

11. Shannon (1948):

$$\begin{aligned} D(P,Q) = &- \sum_{\lambda_1=1}^{\lambda_n} \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) log \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) + \\ &\frac{1}{2} \left(\sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) log(p(\lambda_l)) + \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) log(q(\lambda_l)) \right) . \end{aligned} \tag{4.12}$$

Nonlinear regression and M-estimates

M-estimates form a broad class of estimators which exhibit certain robust properties and it is obtained as the minima of sums of functions of the data. "M" stands for "maximum likelihood-type" estimates and can be described through a nonlinear regression function that seeks to

find the relationship between one or more independent variables and a dependent one. Certain widely used regression functions, such as least-squares estimator (LSE) have favorable properties if their underlying assumptions are true (for example noise is Gaussian), but can give misleading results if those assumptions are violated. Least-squares estimators and many maximum-likelihood estimators are M-estimators. They are obtained by replacing square loss function into another more general convex function, see [Staudte and Sheather, 1990]. Generally estimates with robust regression methods can be more stable with respect to anomalous errors but the performance of them drops when the parametric family is misspecified. Normalization is not required but it may help improving accuracies.

The classical LSE distance corresponds to the function

$$D[P,Q] = \sum_{\lambda_i=1}^{\lambda_n} (p(\lambda_i) - q(\lambda_i))^2. \tag{4.13}$$

It is well known that LSE method is consistent, asymptotically normal and efficient and errors have Gaussian distribution. Also it can be shown that least squares corresponds to the maximum likelihood criterion and can also be derived as a method of moments estimator. Other examples of *M-estimates* that are used in our research are listed below. 1) More general estimates with Laplace distribution

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} |p(\lambda_l) - q(\lambda_l)| \tag{4.14}$$

are known as L_1-estimate or as least absolute error.

2) The Geman and McClure function tries to reduce the effect of large errors further, but it also cannot guarantee uniqueness.

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \frac{(p(\lambda_l) - q(\lambda_l))^2}{(1 + (p(\lambda_l) - q(\lambda_l))^2)}. \tag{4.15}$$

Minimum Contrast Estimation

This class of estimates considers spectral domain and reflectances in this case can be seen as spectral density functions of some stochastic process. It is close to the class of quasi-likelihood estimators, where instead of independence (which does not hold for many cases) is used asymptotical independence. Under some sets of conditions the minimum contrast estimators are consistent. More information can be found in [Taniguchi, 1979].

The basic idea behind it is to minimize the distance (which also called 'contrast' in this case) between a parametric model and a non-parametric spectral density. Since one can interpret satellite observations as measurements in the spectral domain these distances seem to be a natural choice for analyzing satellite data. Also for this family of CFs normalization is not required but it may help improving accuracies.

Here some of the spectral distances that have been used in our research with correspondent $K(x)$- contrast function.

1. Let $K(x) = logx + \frac{1}{x}$.

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \left\{ log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) + \frac{p(\lambda_j)}{q(\lambda_j)} \right\} - 1. \tag{4.16}$$

2. let $K(x) = -logx + x$, then

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \left\{ -log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) + \frac{q(\lambda_j)}{p(\lambda_j)} \right\} - 1. \tag{4.17}$$

3. let $K(x) = (logx)^2$, then

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \{log(q(\lambda_l)) - log(p(\lambda_l))\}^2. \tag{4.18}$$

4. let $K(x) = xlogx - x$, then

$$D(P,Q) = 1 + \sum_{\lambda_1=1}^{\lambda_n} \frac{q(\lambda_l)}{p(\lambda_l)} \left\{ log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) - 1 \right\}. \tag{4.19}$$

4.3.2 SPARC validation dataset

A diverse field dataset, covering various crop types, growing phases, canopy geometries and soil conditions was collected during SPARC (SPectra bARrax Campaign). The SPARC-2003 and SPARC-2004 campaigns took place in Barrax, La Mancha, Spain (coordinates 30°3'N, 28°6'W, 700 m altitude). The test area has an extent of 5 km $\times$ 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land. The annual rainfall average is about 400 mm. In the 2003 campaign (12-14 July) biophysical parameters were measured within a total of 110 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size of about 20^2 m^2. LCC was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. The calibration took place against field values taken from 50 ESUs. A logarithmic function led to best fit with a r^2 of 0.93 [Gandía *et al.*, 2004]. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned to a LAI value, which was obtained as a statistical mean of 24 measures (8 data readings $\times$ 3 replications) with standard errors between 5 and 10% [Fernández *et al.*, 2005]. Strictly speaking, due to the assumption of a random distribution of foliage, clumping is only partially regarded by the instrument and corresponding software, giving therefore effective

LAI as ouptut. No bare soil samples were added in the validation dataset because inversion of canopy RTMs is only relevant over vegetated land covers. For both years, we have a total of 9 crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat), with field-measured values of LAI that vary between 0.4 and 5.9 (μ: 3.0, SD: 1.5) and LCC between 10 and 52 (μ: 38, SD: 14) μg/cm^2.

4.3.3 Sentinel-2

Although the listed CFs can be applied to any EO data, in preparation to forthcoming Sentinel missions we chose applying them to simulated Sentinel-2 (S2) data. The S2 satellites capitalize on the technology and the vast experience acquired with Landsat and SPOT. S2 will be a polar-orbiting, superspectral high-resolution imaging mission [Drusch *et al.*, 2012]. Each S2 satellite carries a Multi-Spectral Imager (MSI) with a swath of 290 km. It provides a set of 13 spectral bands spanning from the visible and near infrared (VNIR) to the shortwave infrared (SWIR), with four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Table 6.1). S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. The pair of S2 satellites aims to deliver data taken over all land surfaces and coastal zones every five days under cloud-free conditions. To serve the objectives of GMES, S2 satellites will provide imagery for the generation of high-level operational products (level 2b/3) such as land-cover and land-change detection maps and geophysical variables such as LCC, LAI and leaf water content. To ensure that the final product can meet the user requirements, the GMES user committee defined an accuracy goal of the biophysical products of 10% [Drusch *et al.*, 2012].

TABLE 4.1: Sentinel-2 MSI band settings. Bands used in this study are bolded

Band #	B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8a**	B9	B10	B11	B12
Band center (nm)	443	**490**	**560**	**665**	**705**	**740**	**783**	**842**	**865**	945	1375	1610	2190
Band width (nm)	20	**65**	**35**	**30**	**15**	**15**	**20**	**115**	**20**	20	30	90	180
Spatial resolution (m)	60	**10**	**10**	**10**	**20**	**20**	**20**	**10**	**20**	60	60	20	20

Here, S2 MSI imagery was simulated on the basis of Compact High Resolution Imaging Spectrometry (CHRIS) data because of its high spatial and spectral resolution. CHRIS provides high spatial resolution (up to $\sim$17 m) hyperspectral data over the VNIR spectra from 400 to 1050 nm at 5 different viewing angles. It can operate in different modes, balancing the number of spectral bands, site of the covered area and spatial resolution because of on-board memory storage reasons [Barnsley *et al.*, 2004]. We made use of nominal nadir CHRIS observation in Mode 1 (62 bands, maximal spectral information), which were acquired during the SPARC campaign (12/07/2003). CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. CHRIS imagery was processed using ESA's CHRIS-Box available in VISAT/BEAM, which includes radiometric recalibration, coherent-noise reduction, geometric correction and atmospheric correction [Alonso and Moreno, 2005; Guanter *et al.*, 2005].

Constrained by the spectral range of CHRIS, we considered the S2 bands starting from B2 (490 nm) until B8a (865 nm). The bands at a spatial resolution of 10 m have been coarse-grained to 20 m so that in total 8 bands in the visible and NIR at a 20 m resolution are available. S2 bands at 60 m were not considered as these bands are intended for atmospheric applications. These bands are intended for atmospheric applications, such as aerosols correction, water vapor correction and cirrus detection [Drusch *et al.*, 2012] and are unable to deliver TOC reflectances that are interpretable by canopy RT models.

4.3.4 LUT generation

This work was carried out with ARTMO (Automated Radiative Transfer Models Operator) [Verrelst *et al.*, 2011, 012c]. ARTMO is a GUI toolbox written in Matlab that provides essential tools for running and inverting canopy reflectance models with different options. In its latest version (V.3), the toolbox is designed in a modular way, i.e. the radiative transfer models modules along with several post-processing modules. In short, the toolbox enables the user: i) to choose between various leaf optical models (e.g., PROSPECT-4, PROSPECT-5) and canopy reflectance models (e.g., 4SAIL, SLC, FLIGHT), ii) to choose between spectral band settings of various air- and space-borne sensors or to define new sensor band settings (in Sensor module), iii) to simulate a massive amount of spectra and store them in a relational database running underneath ARTMO (main module), iv) to visualize spectra of multiple models in the same window (in Graphics module), and finally, v) to run LUT-based model inversion against EO imagery with selected cost function and with optimization and accuracy options for the estimates (in Inversion module).

From the in ARTMO available models we chose for coupling PROSPECT-4 with 4SAIL because of being fast, invertible and well representing homogeneous plant covers on flat surfaces areas such as those present at Barrax. Both models, commonly referred to as PROSAIL, have been used extensively over the past few years for a variety of applications [Jacquemoud *et al.*, 2009]. PROSPECT-4 calculates leaf reflectance and transmittance over the solar spectrum from 400 to 2500 nm at a 1 nm spectral sampling interval as a function of its biochemistry and anatomical structure. It consists of 4 parameters, being leaf structure, LCC, equivalent water thickness and dry matter content [Feret *et al.*, 2008]. 4SAIL calculates top-of-canopy reflectance. 4SAIL inputs consist of: LAI, leaf angle distribution, ratio diffuse/direct irradiation, a hot spot parameter and sun-target-sensor geometry. Spectral input consists of leaf reflectance and transmittance spectra, here coming from PROSPECT-4, and a a moist and dry soil reflectance spectrum. To obtain these soil spectra, the average of bare soil signature was calculated from bare moist and dry soil pixels identified in the imagery. In 4SAIL a scaling factor, α_{soil}, has been introduced that takes variation in soil brightness into account as a function of these two soil types.

The bounds and distributions of the PROSAIL variables are depicted in Table 4.2. Variable bounds were taken from measurement campaigns and/or other studies working with the same

crops [Richter *et al.*, 2009, 2011]. They were chosen in order to describe the characteristics of all crop types used in the study. Gaussian input distributions were generated for LAI and LCC content in order to put more emphasis on the variable values being present in the actual growth stages of the crops. Sun and viewing conditions correspond to the situation of the satellite overpass. All possible combinations were calculated from the in Table 4.2 defined leaf and canopy input ranges. Since the sum of all these combinations would lead to an unrealistic high number of simulations (about 5 billion), a LUT size of 100000 TOC reflectance realizations was randomly chosen in accordance with [Darvishzadeh *et al.*, 2012]. All input parameters, metadata and associated output simulations were automatically stored in a relational database running underneath ARTMO.

4.3.5 Regularization options

Two regularization options are commonly applied in LUT-based inversion strategies. First, a Gaussian noise is often added to the simulated canopy reflectance to account for uncertainties [Bacour *et al.*, 2006; Baret *et al.*, 2007]. Second, several studies demonstrated that the single best parameter combination corresponding to the very smallest distance as calculated by a cost function does not necessarily lead to the best accuracy [Combal *et al.*, 2002; Darvishzadeh *et al.*, 2008]. Instead the mean of multiple solutions tend to provide more accurate results.

TABLE 4.2: Range and distribution of input parameters used to establish the synthetic canopy reflectance database for use in the LUT.

Model Parameters		Units	Range	Distribution
Leaf parameters: PROSPECT-4				
N	Leaf structure index	unitless	1.3-2.5	Uniform
LCC	Leaf chlorophyll content	[μg/cm^2]	5-75	Gaussian ($\bar{x}$: 35, SD: 30)
C_m	Leaf dry matter content	[g/cm^2]	0.001-0.03	Uniform
C_w	Leaf water content	[cm]	0.002-0.05	Uniform
Canopy variables: 4SAIL				
LAI	Leaf area index	[m^2/m^2]	0.1-7	Gaussian ($\bar{x}$: 3, SD: 2)
α_{soil}	Soil scaling factor	unitless	0-1	Uniform
ALA	Average leaf angle	[$\circ$]	40-70	Uniform
HotS	Hot spot parameter	[m/m]	0.05-0.5	Uniform
skyl	Diffuse incoming solar radiation	[fraction]	0.05	-
θ_s	Sun zenith angle	[$\circ$]	22.3	-
θ_v	View zenith angle	[$\circ$]	20.19	-
ϕ	Sun-sensor azimuth angle	[$\circ$]	0	-

Similar variable ranges/values/distributions were used according to field configurations and related studies [Richter *et al.*, 2009, 2011]. $\bar{x}$: mean, SD: standard deviation.

Because different optimization numbers are reported in literature, the mutual impact of these regularization options has been recently systematically studied for a limited set of CFs [Verrelst *et al.*, 2014b]. In that study a range of Gaussian noise levels was introduced up to 30% with steps of 1%. It appeared however that in some cases this upper limit yielded best accuracies, suggesting that the optimal noise configuration may not have been reached. Therefore, in this study noise levels were examined from 0 to 50% with steps of 2%. The percentage noise is considered as the maximal boundary wherein noise can fall and is wavelength dependent. The same range was applied to multiple solutions, from 0% to 50% with steps of 2%. Note that these wide ranges are only intended to study the behaviour of the CFs, thereby seeking for optima. Further we investigate the impact of spectra normalization in *M-estimates* and *Minimum Contrast Estimates* that have not been done before.

Thus, to summarize, we have:

1. Various standalone cost functions from three mathematically different families.
2. Insertion of Gaussian noise on simulated spectra: 0-50%.
3. Use of multiple sorted best solutions in the inversion: 0-50%.
4. Impact of normalization for CFs in *M-estimates* and *Minimum Contrast Estimates* families.

Given all these factors, their effects on the robustness of LUT-based inversion have been assessed for the retrieval of LCC and LAI at the 20-m S2 resolution. The retrieved predictions were compared against the ground-based SPARC validation dataset using statistics such as coefficient of determination (r^2), absolute RMSE and the normalized RMSE (NRMSE [%] = RMSE/range of the parameters as measured in the field *100). The NRMSE was used to compare the performances across the different CFs and parameters. Lower values indicate less residual variance and thus more successful inversion.

4.4 Results

4.4.1 Evaluation of cost functions and regularization options

Leaf chlorophyll content (LCC) retrieval

Figure 4.1 provides a systematic overview of the impact of different cost functions (matrices), noise levels (x-axis) and multiple solutions (y-axis) on LCC retrieval. To enable comparison across the different parameters, the performance of these inversion strategies were evaluated by calculating the deviations between the retrieved parameter values and the 110 validation points through the NRMSE; further referred to as relative error. When interpreting these error matrices, opting for the very single best solution without added noise in the inversion appeared to be a

poor inversion strategy for each of the CFs. Inversion without these regularization is shown in the left-bottom corner of the matrices and also listed in the rightmost column of Table 4.3. Accuracies further degraded when opting for the very single best solution and adding noise. This pattern appeared for the majority of CFs, which suggests that the strategy of only adding noise in the inversion is to be avoided. The reason of the poor performance of the single best solution lies in the ill-possedness of the problem; multiple parameter combinations lead to the same or very similar spectra. Therefore, picking the sorted most matching spectrum according to a CF may not be the best choice, but rather the mean value of multiple best matching spectra should be considered. Effectively, the inversion clearly benefited from introducing multiple solutions in the inversion for all CFs, preferably in combination with added noise. Overall, with consideration of the mean of all CFs, relative errors more were twice reduced when comparing

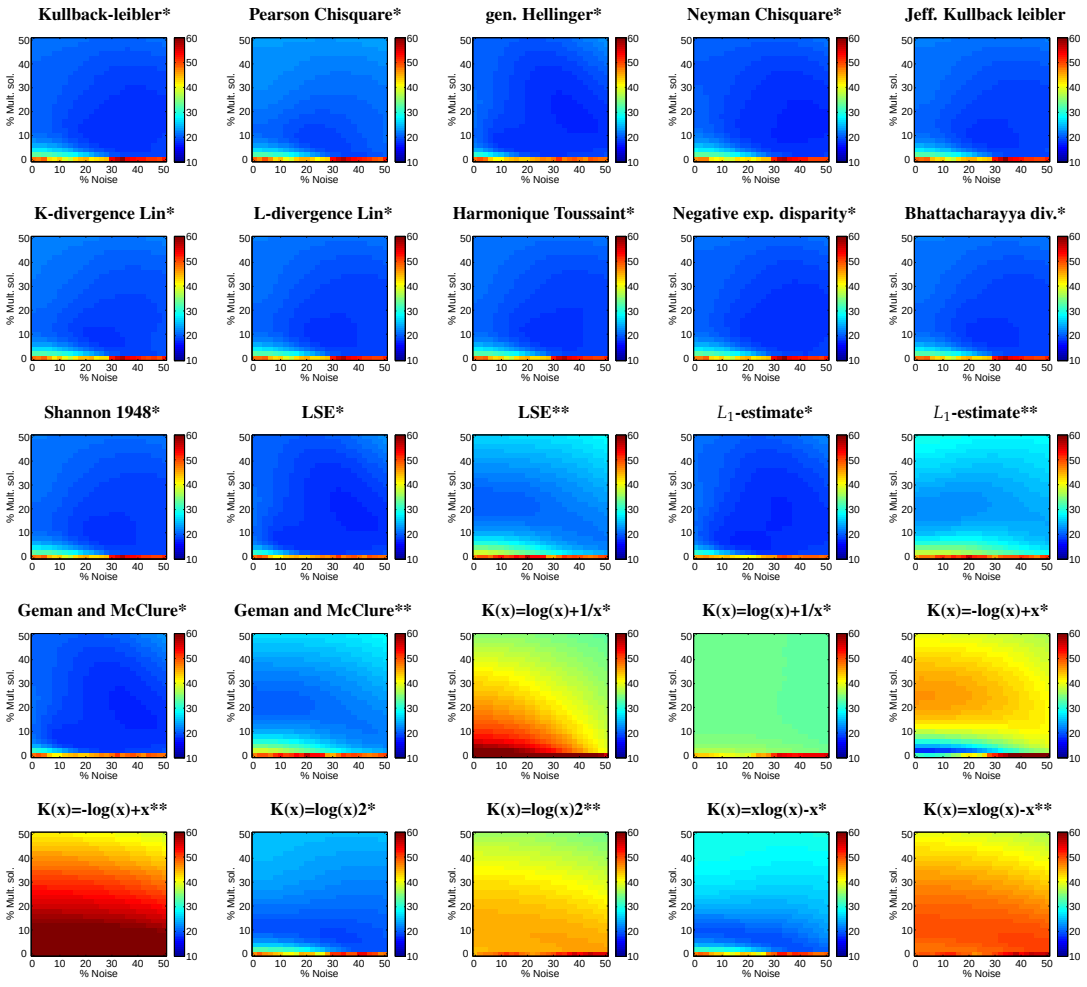


FIGURE 4.1: Normalised RMSE (NRMSE) matrices for LCC retrieval using cost function displaying the impact of % noise (X-axis) against multiple solutions (Y-axis) in LUT-based RTM inversion. *: normalized, **: non-normalized. The more bluish, the lower relative errors and thus the better the inversion.

results without regularization options; $\bar{x}$ NRMSE from 49% to 22% (see Table 4.3).

Moreover, these error matrices reveal the true interactions between the different cost functions and the regularization options. The matrices suggest that each of the tested CFs respond differently to the regularization options. It is interesting to see that in the majority of them a pattern with optimized performances appear (darkest blue), though the location and shape of this pattern may vary per CF. Thereby, not only CFs that are mathematically more alike led to similar trends, but also similarities appeared within the same CF family. This is particularly noticeable for the family of *information measures*, which shows robust LCC estimates across the whole noise and multiple solutions range, but also *M-estimates* CFs show similar behavior.

Interestingly, the matrices also revealed that data normalization governs the success of *M-estimates*. Excellent performances were achieved for the three *M-estimates* (LSE, L_1-estimate, Geman and McClure), with best performances for L_1-estimate (NRMSE of 17.6% at 6% multi solutions and 18% noise). Note that L_1-estimate is proven to be optimal when distribution of errors have Laplace distribution. Also Neyman Chi-square and Generalized Hellinger reached

TABLE 4.3: Statistics (r^2, RMSE , NRMSE) based on best evaluated NRMSE and corresponding mean multiple solutions (%) and noise level (%) for LCC retrieval. *: normalized, **: non-normalized. The '0.0 NRMSE' column represents the results without regularization options (left-bottom corner in matrices). Best inversion strategy is in bold font.

Cost function	**mult. sol. (%)**	**noise (%)**	r^2	**abs. RMSE**	**NRMSE (%)**	**0,0 NRMSE (%)**
Kullback leibler*	10	26	0.71	7.34	19.47	47.54
Chi square*	8	24	0.69	7.57	18.64	48.94
Generalised Hellinger*	20	36	0.74	7.16	17.63	47.55
Neyman Chi square*	16	44	0.73	7.16	17.63	46.62
Jeffreys Kullback leibler*	10	26	0.70	7.41	18.25	48.36
K-divergence Lin*	8	26	0.70	7.43	18.30	48.01
L-divergence Lin*	10	26	0.70	7.39	18.20	48.25
Harmonique Toussaint*	10	26	0.71	7.37	18.15	48.36
Negative exp. disparity*	10	26	0.71	7.31	18.01	47.43
Bhattacharyya divergence*	10	28	0.70	7.40	18.22	48.18
Shannon 1948*	10	26	0.70	7.39	18.20	48.18
LSE*	20	36	0.74	7.16	17.63	47.55
LSE**	22	0	0.68	8.23	20.27	46.99
L_1-estimate*	**6**	**18**	**0.73**	**7.14**	**17.59**	**45.88**
L_1-estimate**	20	12	0.61	9.14	22.52	45.58
Geman and McClure*	20	36	0.74	7.16	17.63	47.55
Geman and McClure**	22	0	0.68	8.24	20.30	46.99
K(x)=log(x)+1/x*	50	50	0.68	13.43	33.09	70.96
K(x)=log(x)+1/x**	16	50	0.62	13.36	32.91	43.86
K(x)=-log(x)+x*	2	0	0.70	7.45	18.34	31.82
K(x)=-log(x)+x**	50	50	0.48	15.76	38.83	83.15
K(x)=log(x)^2*	8	32	0.68	7.76	19.11	46.68
K(x)=log(x)^2**	50	50	0.31	13.87	34.15	44.14
K(x)=x(log(x))-x*	6	30	0.66	7.95	19.58	47.19
K(x)=x(log(x))-x**	50	50	0.30	15.05	37.06	48.99

such degree of accuracy (see Table 4.3). However, *M-estimates* performed considerably poorer (best results around 20%) when data is non-normalized, and matrices show that retrieval results are more affected by the two regularization options. Finally, the *Minimum Contrast Estimation* family yielded more irregular patterns. While in general they benefited from normalizing the dataset, only the 'K(x)=log(x)^2' and 'K(x)=xlog(x)-x' proved to be robust CFs for LCC retrieval. In the absence of normalization best performances of these CFs were obtained at a maximal of tested noise level (i.e., 50%). Because these noise levels led to a considerable degradation of the simulated spectra it seems that these CFs only function well for LCC when data is normalized.

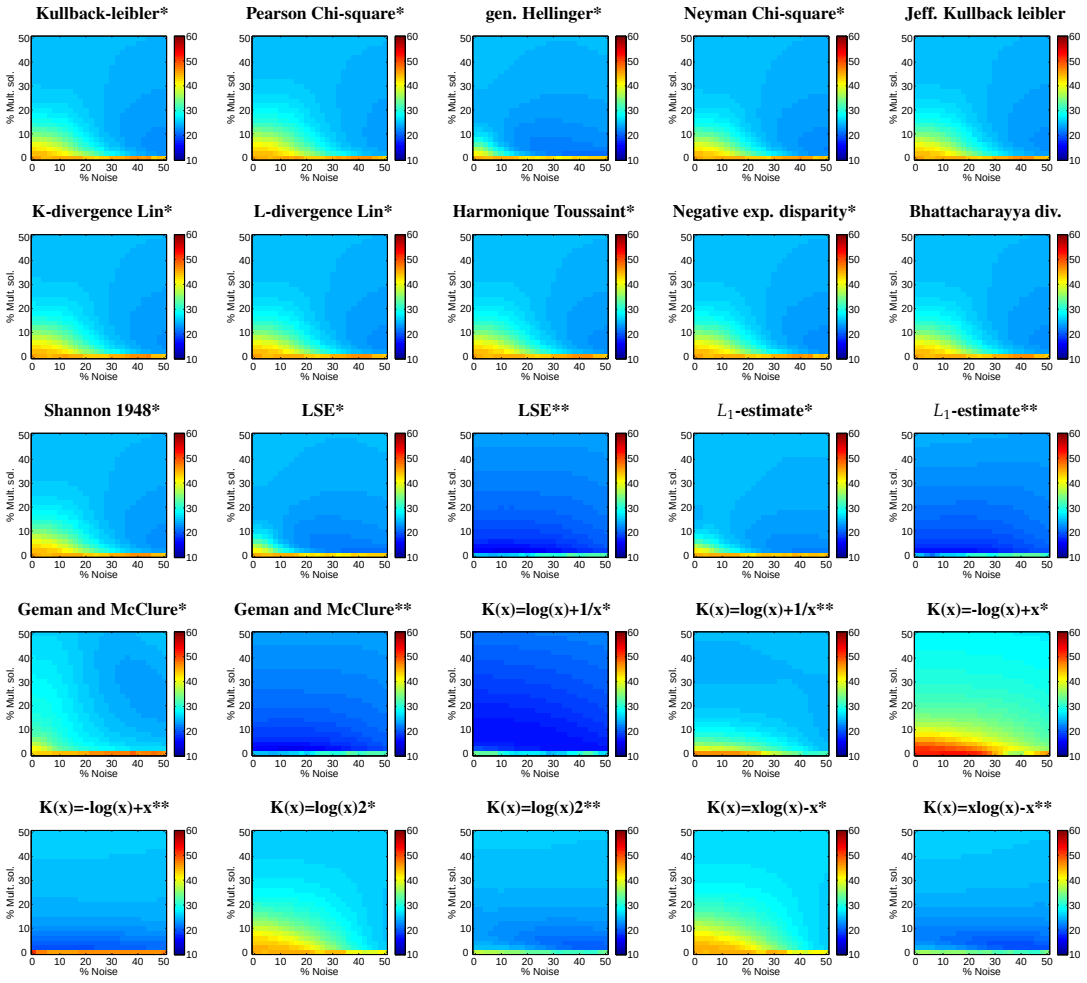


FIGURE 4.2: NRMSE matrices for LAI retrieval using cost function displaying the impact of % noise (X-axis) against multiple solutions (Y-axis) in LUT-based RTM inversion. *: normalized, **: non-normalized. The more bluish, the lower relative errors and thus the better the inversion.

TABLE 4.4: Statistics (r^2, RMSE , NRMSE) based on best evaluated NRMSE and corresponding mean multiple solutions (%) and noise level (%) for LAI retrieval. *: normalized, **: non-normalized. The '0.0 NRMSE' column represents the results without regularization options (left-bottom corner in matrices). Best inversion result is in bold font.

Cost function	mult. sol. (%)	noise (%)	r^2	abs. RMSE	NRMSE (%)	0,0 NRMSE (%)
Kullback leibler*	4	50	0.63	1.25	22.74	45.59
Chi square*	8	50	0.62	1.29	23.53	45.74
Generalised Hellinger*	2	42	0.62	1.17	21.34	44.59
Neyman Chi square*	4	50	0.62	1.24	22.48	45.51
Jeffreys Kullback leibler*	6	50	0.62	1.26	22.93	45.60
K-divergence Lin*	6	50	0.62	1.27	23.06	45.67
L-divergence Lin*	10	26	0.70	1.26	22.92	45.54
Harmonique Toussaint*	6	50	0.62	1.26	22.91	45.60
Negative exp. disparity*	4	50	0.63	1.25	22.72	45.53
Bhattacharyya divergence*	6	50	0.62	1.26	22.93	45.53
Shannon 1948*	6	50	0.62	1.26	22.92	45.54
LSE*	2	42	0.62	1.17	21.34	44.58
LSE**	**2**	**14**	**0.74**	**0.84**	**15.32**	**25.45**
L_1-estimate*	2	50	0.62	1.22	22.25	44.93
L_1-estimate**	2	12	0.73	0.91	16.57	25.31
Geman and McClure*	2	42	0.62	1.17	21.34	44.59
Geman and McClure**	2	14	0.74	0.85	15.39	25.45
K(x)=log(x)+1/x*	2	42	0.63	0.91	16.51	34.02
K(x)=log(x)+1/x**	10	50	0.50	1.35	24.57	46.17
K(x)=-log(x)+x*	50	50	0.52	1.53	27.80	51.69
K(x)=-log(x)+x**	2	0	0.64	1.09	19.86	51.17
K(x)=log(x)^2*	12	50	0.54	1.40	25.42	45.13
K(x)=log(x)^2**	2	46	0.66	1.13	20.47	37.94
K(x)=x(log(x))-x*	6	50	0.63	1.42	25.82	45.11
K(x)=x(log(x))-x**	2	40	0.63	1.08	19.56	35.57

Leaf area index (LAI) retrieval

Introducing noise and mean of multiple solutions considerably improved the LAI accuracies (Figure 4.2). Overall, on average relative errors almost halved: NRMSE from μ 42% to μ 22% (Table 4.4). However, in contrast to LCC retrievals, the *information measures* and the normalized *M-estimates* were no longer able to deliver good inversion results. In fact for the majority of CFs, normalization of the data caused that best results were achieved with the maximal degree of injected noise (i.e., 50%). Since such a spurious high noise level led to a considerable degradation of the simulated spectra it makes us concluding that normalization is not recommended for LAI retrieval. Consequently non-normalized *M-estimates* greatly improved accuracy of LAI retrievals. Particularly the classical LSE demonstrated to be the best CF (NRMSE: 15.3% at multiple solutions of 2% and noise of 14%), closely followed by the other M-estimates (L_1-estimate, Geman and McClure). This means that in our example the errors tend to be Gaussian and solution is stable. Also these results suggest that controlling the normalization option can play an important role in LUT-based inversion. The contrast functions family did not perform

very well and only normalized K(x)=log(x)+1/x is able to retrieve LAI with best relative error of 16.5%. In comparison with LCC (Table 4.3) it can be noted that best LAI retrieval is performing slightly better (e.g., see NRMSE). Though, most important is to realize that different optimized inversion strategies were identified for both parameters. Finally, although not shown here for the sake of brevity, the retrieval capability of canopy-integrated chlorophyll content (CCC: LAI $\times$ LCC) was evaluated. CCC showed analogous patterns as LAI which suggests that this parameter, more than LCC, is the driving parameter of CCC. This can be probably attributed to the fact that LAI influences the whole spectral range, while LCC sensitive spectral range is limited to the beginning of the red edge.

4.4.2 Biophysical parameters mapping

Obtained systematic overview of the various inversion performances shows that deriving simultaneously multiple biophysical parameters using one inversion strategy is not the best choice. When parameters are correlated in a non-linear way it appears that the optimal cost functions are different for each parameter [Leonenko *et al.*, 2013]. Hence, when we try to quantify LCC and LAI simultaneously will at least for one parameter lead to suboptimal performances. While in an operational setting it may be desirable to seek for an inversion strategy that leads to a balanced performance in the generation of all retrievable parameters [Verrelst *et al.*, 2014b], because of their non-linear dependency applying different retrieval strategies to different parameters is generally preferable. That is also the approach pursued here; for each parameter the best performing inversion strategy was applied (i.e., bolded in Tables 4.3, 4.4). For LCC we applied L_1-estimate (with mult. sol.: 4%, noise: 18%) and for LAI non-normalized normal distribution LSE (with mult. sol.: 2%, noise: 14%). Since in each of these strategies the mean of multiple solutions is applied it leads to the mapping of the mean estimates (μ) and associated standard deviation ($\pm\sigma$) (Figure 4.3).

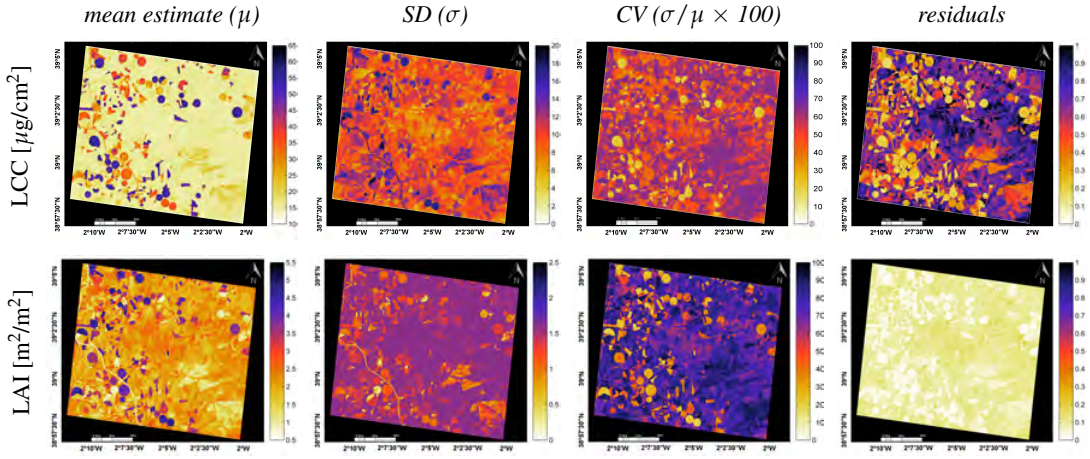


FIGURE 4.3: Mean predictions, standard deviation (SD), coefficient of variation (CV) and residuals for LCC and LAI by using for each parameter best evaluated inversion strategy (see Table 4.3, 4.4).

These maps are briefly interpreted. Starting with the μ-estimate maps, it can be observed that not only the irrigated circular fields with green biomass are clearly marked, but also pronounced within-field variations are notable by both parameters. These irrigated fields are characterized by a LCC above 40 μg/cm^2 and LAI above 3. Areas with low LCC ($\leq$ 25 μg/cm^2) and LAI ($\leq$ 1.5) are mainly bare soils, non-irrigated fallow lands or senescent or harvested cereal fields (wheat, barley). Particularly detailed spatial variation was encountered in the LAI map which suggests that greater sensitivity for this structural parameter is achieved. The $\pm\sigma$-maps can be interpreted as uncertainty of the μ-estimate. The smaller the $\pm\sigma$ for a pixel, the lower that variation in multiple solutions is encountered during the inversion. Note that such $\pm\sigma$ map is also obtained by the MODIS LAI LUT-based retrieval method [Knyazikhin *et al.*, 1998b] and serves as a measure of the solution accuracy; starting from Collection 5, the $\pm\sigma$(LAI) has been provided in the standard products. Here, at a glance, large uncertainty ranges arise when medium to high LCC conditions prevail. Similar studies observed the same large uncertainty ranges over dense canopies and are attributed to the saturation effect in the radiative transfer and the limited information provided in the a generation of the LUT [Pinty *et al.*, 2007; Weiss *et al.*, 2007; Garrigues *et al.*, 2008]. However, since the $\pm\sigma$ also directly depends on the magnitude of μ, it may be more of interest to map the relative uncertainties, i.e. the coefficient of variation (CV; σ/μ). A relative indicator is of interest as it provides more information about the inversion process, i.e. a lower CV means a greater uniformity across the % best solutions, but it also allows a comparison of the inversion performance across all maps (Figure 4.3). For both LCC and LAI the CV maps suggest that the inversion process had least difficulty with the vegetated areas (circular irrigated fields), while most uncertainty was propagated over the fallow land and calcareous bare soils (South-East of image).

At the same time, parameter-independent information regarding the performance of the inversion process can also be obtained by mapping the residuals of the CFs. They provide another indicator on the inversion certainty. For each pixel, it indicates the degree of mismatch between the observed spectrum and the best matching simulated spectra. More than CV, these maps delineate the surfaces where the simulated spectra closely matched the observed spectra and thus a higher probability achieving a successful inversion. The maps suggest that particularly LAI was retrieved without difficulty, with perfect match over the vegetated areas. The inversion process had in all generality more difficulty with LCC, but large differences can be observed, leading to essentially the same earlier observed pattern; a close agreement over vegetated areas as opposed to senescent and non-vegetated lands. On the whole, the uncertainty analysis suggests that the inversion would benefit from a wider LUT with more spectral variation of senescent vegetation and bare soil surfaces.

4.5 Discussion

The upcoming S2 missions open opportunities to implement novel retrieval algorithms in operational processing chains. Specifically, there is a need for retrieval methods that are accurate, robust, and make fully use of the new S2 MSI bands. While in related works vegetation indices

[e.g., Delegido *et al.*, 2011; Clevers and Gitelson, 2013], or machine learning regression algorithms [e.g., Verrelst *et al.*, 2012a; Richter *et al.*, 2012b] have been proposed for S2 biophysical parameter retrieval, alternatively we advocate the method of LUT-based RTM inversion. Although this approach is not new, it has never been exploited to the fullest. Here we systematically analyzed the impact of CFs and regularization options. It led to consolidated findings discussed below.

4.5.1 Cost functions and regularization options

To mitigate the drawback of ill-posedness instead of introducing prior information it was opted to exploit the performances of alternative CFs in combination with regularization options. The rationale for evaluating CFs as opposed to the widely used LSE (or RMSE) is that propagated uncertainties and errors, e.g., due to uncertainties in instrument calibration, variations in atmospheric composition or simplifying assumptions in the representation of canopy and soil background, distort the residuals and in many cases violate a key assumption for using LSE, which corresponds to the maximum likelihood estimation with Gaussian distribution of residuals [Leonenko *et al.*, 2013]. Nevertheless in our presented dataset it was found that LSE outperformed most of the alternative tested CFs for LCC with the condition of data normalization. Only the 'L_1-estimate' function performed slightly better, which suggest that the errors rather tend to be distributed according to a Laplace distribution. Moreover normalization led to better performances with respect of varying added noise and multiple solutions in the inversion. Hence, data normalization provides good accuracy for LCC inversion and can be explained as follows. Variation in reflectance as a function of chlorophyll absorption occurs predominantly in the visible part of the spectrum and declines rapidly once entering red edge [Gitelson *et al.*, 1996], i.e. encompassing S2 bands B2 to B5. Accordingly, in PROSPECT the chlorophyll absorption coefficients start to become negligible from 730 nm onwards. A closer look to the simulated dataset revealed that normalization compresses the dataset and so ensures that the observation spectrum falls within the same range. This particularly plays a role in the visible part of the spectrum. As reflectance of vegetated surfaces in the visible is typically low (often $\leq$ 10%) and variation in this region is narrow, normalization helps reaching a better match in this region between simulated and observed spectra. Conversely, data normalization did not lead to the good CFs performance for LAI, and non-normalized data in combination with multiple solutions and the CF LSE showed to be the most successful. It should herewith be noted that LAI variation (i.e. vegetation density) causes a spectral variation over the whole VNIR spectrum, i.e. S2 bands from B2 to B8. Thereby, the larger magnitude of reflectance in the NIR and greater LAI-induced variation in comparison to the visible part makes NIR an important region for LAI retrieval in LUT-based inversion. Consequently, a spectral observation easily falls within such a broad range of simulated spectral variation, and compression of the simulations through normalization tend to disturb the matching between observed and simulated spectra. A standard inversion scheme proved thus to be most successful for LAI retrieval. Effectively, many publications demonstrated that LAI can be relatively easily retrieved without data normalization [e.g., Darvishzadeh *et al.*, 2008; Richter *et al.*, 2009; Si *et al.*, 2012; Atzberger and

Richter, 2012]. We are not aware of other studies that have systematically evaluated the role of data normalization in LUT-based inversion, though proper usage may greatly improve accuracies. For instance, it explains why in [Verrelst *et al.*, 2014b] RMSE was performing poorly for LAI where this CF was employed in normalized mode only. It may also explain why similar studies reported a poor LCC inversion as there no mentioning of normalizing the data has been made [Darvishzadeh *et al.*, 2008; Si *et al.*, 2012]. Further, as earlier observed in [Verrelst *et al.*, 2014b], this study confirms for 18 tested CFs that introducing some degree of noise and mean of multiple solutions in the inversion can lead to improved inversion performances, though its actual impact strongly depended on the considered parameter and cost function. Hence developing optimized inversion strategies for each single retrievable parameter is strongly recommended.

4.5.2 Inversion performance

The evaluation of various CFs and regularization options led to identified inversion strategies with accuracies of r^2 on the order of 0.73-0.81 and relative errors (NRMSE) of 15-18%. A range of 15%-20% is regarded as the currently achievable accuracy for LAI from EO data [Baret, 2010]. However, a retrieval accuracy of 10% is targeted for the S2 mission [Drusch *et al.*, 2010]. It should thereby be emphasized that the validation dataset consisted of all kinds of crop types, including row crops such as maize, potatoes, onions and vineyards. In related work by [Darvishzadeh *et al.*, 2008; Richter *et al.*, 2009] this degree of accuracy was only achieved over single crops, and it was concluded that PROSAIL fails to invert over multi-species canopies. Thereby, various types of uncertainties have been identified that may lead to suboptimal retrievals, with respect to model usage (e.g., 1D vs. 3D models), parametrization and validation data [Combal *et al.*, 2003]. Suggestions for improvements typically refer to adding more prior information at the level of individual parcels , i.e. through a more specialized LUT [e.g., Knyazikhin *et al.*, 1998a; Combal *et al.*, 2003; Meroni *et al.*, 2004; Atzberger, 2004; Richter *et al.*, 2009; Verrelst *et al.*, 012c]. While such strategies could be beneficial for dedicated sites, site-specific information is usually unavailable in an operational context for larger areas. The proposed regularization options yielded robust inversion schemes that are easily applicable over full scenes covering heterogeneous canopy surfaces. Although most of the CFs within the same family perform alike and are in principle thus replaceable, performance gain is especially reached in combination with applying multiple solutions and noise in the inversion. Moreover, further improvements can be achieved by applying CFs that require tuning of additional parameters. In general tuning parameters in CFs would lead to refined inversion strategies as compared to standalone CFs (e.g. see [Leonenko *et al.*, 2013; Verrelst *et al.*, 2014b]). Further research is planned to have these tunable CFs optimized in an automated fashion.

At the same time, information about the inversion uncertainty on a pixel-by-pixel basis may be as relevant as overall accuracies calculated from a limited set of ground-based validation data. Uncertainty indicators were obtained through mapping of CV and residuals. The CV is an indicator of the uncertainty range around the mean estimate, which tells something about the ill-posedness of the inversion of a retrievable parameter, whereas the residuals tells us how much the observed spectra deviate against that from the LUT spectra in the inversion scheme.

These quality layers allow masking out uncertain estimates.

Both indicators showed a consistent spatial trend for all parameters: pixels of vegetated surfaces matched closely with the synthetic reflectance database while pixels of non-vegetated surfaces faced more difficulties. Two reasons can be identified for this discrepancy: 1) the inversion strategy was optimized against validation data that was exclusively collected on vegetated areas, and 2) PROSAIL is a canopy reflectance model and thus well able to detect variation in vegetation properties. Hence the generated LUT and final inversion scheme were not optimized to detect variations in dried-out fallow and bare soil lands. For retrievals over full images there is thus a need for regulating the inversion strategy both over vegetated targets as well over non-vegetated targets. Therefore, while having inversion over vegetated canopies resolved, adequately processing non-vegetated surfaces remains to be optimized. To start with, further efforts are needed in the generation and evaluation of more generic LUTs, i.e. including spectra of kinds of soil types, man-made surfaces and water bodies. PROSAIL alone is not able to deliver this; hence coupling with a generic soil reflectance model that enables generating a variety of non-vegetated reflectances would be a next step to do. It should hereby also be noted that the 20m-S2 bands B11 and B12 have not been considered in this study because of falling outside the CHRIS range. These bands in the SWIR are known to be sensitive to vegetation structure [Brown *et al.*, 2000] and can better distinguish between dried-out fallow and non-vegetated lands. It is expected that inclusion of the SWIR bands will further improve the retrieval quality. Another promising avenue to be investigated is relying on vegetation indices to spatially constrain the LUTs. For instance, vegetation indices are able to detect bare soil, water bodies, sparsely vegetated areas and densely vegetated areas (e.g., see also [Dorigo *et al.*, 2009]). That information could then be used to constrain LUTs on a per-pixel basis and is currently explored to be implemented into ARTMO.

4.6 Conclusions

LUT-based inversion is considered as a physically-sound retrieval method to quantify biophysical parameters from Earth observation imagery, but the full potential of this method has not been consolidated yet. Here, we have systematically compared 18 different cost functions (CFs) originating from three major statistical families: *information measures*, *M-estimates* and *minimum contrast methods*. All these CFs are standalone functions that can be directly implemented in processing algorithms. These CFs were fully exploited through various regularization options, i.e., adding noise, multiple solutions and normalization, for the benefit of improved LCC and LAI estimations. Inversion of the PROSAIL model against a simulated 20 m Sentinel-2 imagery (8 bands: B2-B8a) over an agricultural site (Barrax, Spain) and against a validation dataset led to the following conclusions:

- All evaluated CFs and biophysical parameters gained from regularization options such as adding some noise and multiple solutions in the inversion. These options with proper adjustment can significantly reduce relative errors.

- With introduction of multiple solutions and noise *information measures* CFs proved to be successful for deriving LCC. However best LCC results were achieved with the M-estimate L_1 (NRMSE of 17.6% at 6% multiple solutions and 18% noise) when data is normalized.

- Data normalization appeared to be unsuccessful for retrieving LAI. Here, the classical LSE yielded best results for non-normalized data; NRMSE of 15.3% at 6% multiple solutions and 18% noise, and 16.4% at 16% multiple solutions and 0% noise, respectively.

Systematic analysis for each biophysical parameter identified different optimized inversion strategy, which was subsequently applied to pixel-by-pixel Sentinel-2 imagery. It provided us with maps of mean estimates and associated statistics and showed insight into the uncertainty of the retrievals (e.g. coefficient of variation and residuals). These indicators showed that inversions were most successful over densely vegetated areas. PROSAIL had more difficulty with processing fallow and non-vegetated lands, but that is expected to be resolved with an adjusted LUT and the addition of SWIR bands in actual Sentinel-2 data.

The bottom line of this work is that, despite common practice, no single inversion strategy was found to be optimal for deriving multiple biophysical parameters. While some general trends with respect to regularization options and CFs was revealed, it is the data distribution that determines the success of the inversion strategy, which is governed by the biophysical parameter, the generated LUT, the spectral bands and the validation data. It is therefore recommended to test different CFs and regularization options before implementing an inversion scheme in a processing chain.

5

MLRA based retrieval

Contents

This chapter is based on:

Juan Pablo Rivera Caicedo, Jochem Verrelst, Jordi Muñoz-Marí, José Moreno and Gustavo Camps-Valls (2014)
Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (In press)
DOI:10.1109/JSTARS.2014.2298752

5.1 Abstract

Biophysical parameters such as leaf chlorophyll content (LCC) and leaf area index (LAI) are standard vegetation products that can be retrieved from Earth observation imagery. This paper introduces a new Machine Learning Regression Algorithms (MLRA) toolbox into the scientific Automated Radiative Transfer Models Operator (ARTMO) software package. ARTMO facilitates retrieval of biophysical parameters from remote observations in a Matlab graphical user interface (GUI) environment. The MLRA toolbox enables analyzing the predictive power of various machine learning regression algorithms in a semiautomatic and systematic manner, and applying a selected MLRA to multispectral or hyperspectral imagery for mapping applications. It contains both linear and nonlinear state-of-the-art regression algorithms, in particular linear feature extraction via principal component regression (PCR) and partial least squares regression (PLSR), decision trees (DT), neural networks (NN), kernel ridge regression (KRR) and Gaussian processes regression (GPR). The performance of multiple implemented regression strategies has been evaluated against the SPARC dataset (Barrax, Spain) and simulated Sentinel-2 (8 bands), CHRIS (62 bands) and HyMap (125 bands) observations. In general, nonlinear regression algorithms (NN, KRR, GPR) outperformed linear techniques (PCR, PLSR) in terms of accuracy, bias, and robustness. Most robust results along gradients of training/validation partitioning and noise variance were obtained by KRR while GPR delivered most accurate estimations. We applied a GPR model to a hyperspectral HyMap flightline to map LCC and LAI. We exploited the associated uncertainty intervals to gain insight in the per-pixel performance of the model.

5.2 Introduction

Leaf area index (LAI) and leaf chlorophyll content (LCC) are essential land biophysical parameters retrievable from optical Earth observation (EO) data [Whittaker and Marks, 1975; Lichtenthaler, 1987; Malenovský *et al.*, 2012]. These parameters provide information about the phenological stage and health status (e.g., development, productivity, stress) of crops and forests [Lichtenthaler *et al.*, 1996]. The quantification of these parameters from space over large areas has become an important aspect in agroecological, environmental and climatic studies [Dorigo *et al.*, 2007]. At the same time, remotely sensed observations are increasingly being applied at a within-field scale for dedicated agronomic monitoring applications [Delegido *et al.*, 2011, 2013; Gianquinto *et al.*, 2011]. With the forthcoming superspectral Sentinel-2 and Sentinel-3 missions and the planned EnMAP and PRISMA imaging spectrometers, the unprecedented data availability requires retrieval processing techniques that are accurate, robust and fast to apply.

Biophysical parameter retrieval always require an intermediate modeling step to transform the measurements into useful estimates [Baret and Buis, 2008]. This modeling step can be approached with either statistical, physical or hybrid methods. In this paper we will focus on the statistical approximation as this field has advanced largely over the last two decades [Camps-Valls and Bruzzone, 2009; Camps-Valls *et al.*, 2012b]. Statistical models can be categorized into either a parametric or nonparametric approaches. *Parametric models* assume an explicit relation between the variables. They rely on physical knowledge of the problem and build explicit parametrized expressions that relate a few spectral channels with the biophysical parameter of interest. Alternatively, *nonparametric models* are adjusted to predict a variable of interest using a training dataset of input-output data pairs, which come from concurrent measurements of the parameter and the corresponding reflectance/radiance observation. Several nonparametric regression algorithms are available in the statistics and machine learning literature, and recently they have been introduced for biophysical parameter retrieval [Baret and Buis, 2008; Verrelst *et al.*, 2012a].

Particularly, the family of machine learning regression algorithms (MLRAs) emerged as a powerful nonparametric approach for delivering biophysical parameters. MLRAs have the potential to generate adaptive, robust relationships and, once trained, they are very fast to apply [Hastie *et al.*, 2009]. Typically, machine learning methods are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. They may therefore be suitable candidates for operational applications. Effectively, algorithms such as neural networks (NNs) are already implemented in operational retrieval chains (e.g. CYCLOPES products) [Bacour *et al.*, 2006; Verger *et al.*, 2008]. It remains still to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insight in the retrievals and evaluating retrieval performances. Besides, training NNs involve tuning several parameters that may greatly impact the final robustness of the model. In part, this why in the recent years alternative and simpler to train regression methods have started replacing NNs. Specifically the family of *kernel methods* [Camps-Valls and Bruzzone, 2009] has emerged as an alternative to NNs in many scenarios. Kernel methods typically involve

few and intuitive hyperparameters to be tuned, and can perform flexible input-output nonlinear mappings. Even though MLRAs are widely recognized as very powerful methods, some questions still remain open, e.g. how robust are these models in case of noisy situations, and how much they depend on changes between the training and testing data distributions. Perhaps more important is the fact that, for the broader remote sensing community, they are also perceived as complicated black boxes with several parameters to be tuned, which requires expertise. Further, until now no user-friendly graphical user interface (GUI) toolbox exists that brings several state-of-the-art MLRAs together. To facilitate and automate the use of MLRAs, in this work we present a novel software package that allows systematically analyzing and applying MLRA-developed models. The so-called MLRA toolbox has been implemented within the in-house developed toolbox called ARTMO (Automated Radiative Transfer Models Operator). ARTMO is a scientific GUI toolbox dedicated to the retrieval of vegetation properties from optical imagery [Verrelst *et al.*, 012c].

This brings us to the objectives of this work, which are: (1) to present the novel MLRA toolbox for semiautomatic retrieval of biophysical parameters; (2) to evaluate the different MLRAs on their performance and robustness; and, (3) to apply the best performing MLRA to EO imagery to test the robustness and accuracy in real scenarios.

The following sections will first briefly describe the considered nonparametric regression algorithms and then the latest status of ARTMO, followed by an introduction of the most important components of the new MLRA toolbox. The used data is subsequently described and an evaluation of six nonparametric regression methods is presented. A discussion on the use of these models for EO processing and a conclusion closes this paper.

5.3 Machine learning regression algorithms

Machine learning regression algorithms learn the relationship between the input (e.g. reflectances) and output (e.g. biophysical parameters) by fitting a flexible model looking at the structure of the data. The hyperparameters of the model are typically adjusted to minimize the prediction error in an independent validation dataset. This way, one looks for the best *generalization* capabilities, not only a good performance in the training set that would give rise to an overfitted solution. In this paper we compare several regression algorithms. A first family of linear methods follow a simple chained approach: first data dimensionality is applied to alleviate collinearity problems which is then followed by canonical least squares linear regression. A second family of methods consists of building nonlinear functions of the data directly. Several state-of-the-art methods are considered here: (1) regression trees (RT), (2) artificial neural networks (NN), (3) kernel ridge regression (KRR), also known as least squares support vector machine, and (4) Gaussian processes regression (GPR). All these regression techniques are popular in various application domains thanks to its relatively fast training, good performance, and robustness to the overfitting problem. In the following subsections we briefly summarize them.

5.3.1 Dimensionality reduction and linear regression

Let us consider a supervised regression problem, and let $\mathbf{X}$ and $\mathbf{y}$ be the input and output centered matrices of sizes $n \times d$ and $n \times 1$, respectively. Here, l is the number of training data points in the problem and d is the data dimension. The objective of standard linear regression is to adjust a linear model for predicting the output variable from the input features, $\widehat{\mathbf{y}} = \mathbf{X}\mathbf{w}$, where $\mathbf{w}$ contains the regression model coefficients (weights) and has size $d \times 1$. Ordinary least-squares (OLS) regression solution is $\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$, where $\mathbf{X}^{\dagger} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}$ is the Moore-Penrose pseudoinverse of $\mathbf{X}$. Highly correlated input variables can result in rank-deficient covariance matrix $\mathbf{C}_{xx} = \frac{1}{n}\mathbf{X}^{\top}\mathbf{X}$, making the inversion unfeasible. The same situation is encountered in the small-sample-size case.

A common approach in statistics to alleviate these problems considers first reducing data dimensionality and then applying the OLS normal equations to the projected data or *scores* [Arenas-García *et al.*, 2013]. These scores reduce to a linear transformation of the original data, $\mathbf{X}' = \mathbf{X}\mathbf{U}$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_f}]$ is referred to as the projection matrix, $\mathbf{u}_i$ being the ith projection vector and n_f the number of extracted features. The best known linear dimensionality reduction method is Principal Component Analysis (PCA) [Jolliffe, 1986] which reduces to solve the eigenproblem:

$$\mathbf{C}_{xx}\mathbf{u} = \lambda\mathbf{u}.$$

An alternative supervised method is Partial Least Squares (PLS) [Wold, 1985] in which we have to solve

$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{xy}^{\top} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

Note that PCA disregards the target data and exploits correlations between the input variables to maximize the variance of the projections, while PLS looks for projections that maximize the covariance between the features and the target data. In both cases, the user selects the dimensionality of the projected data n_f. After projection, the OLS equations are solved using $\mathbf{X}'$. The approaches respectively lead to the so-called principal component regression (PCR) [Wold *et al.*, 1987] and the partial least squares regression (PLSR) [Geladi and Kowalski, 1986] methods. Particularly PLSR emerged as a popular regression technique for interpreting hyperspectral data, with various experimental applications in vegetation properties mapping [e.g., Coops *et al.*, 2003; Gianelle and Guastella, 2007; Cho *et al.*, 2007; Ye *et al.*, 2007].

5.3.2 Regression trees (RT)

Regression trees build predictive models that take the observations as inputs and map them to the target variable. The model structure is made out of nodes (or leaves) and branches. Leaves represent output variable discrete values and branches constitute piece-wise linear decisions. Decision tree learning can be done in several ways and using different algorithms, which mainly vary on the procedure used to determine where to split. In this paper we focused on the standard Breiman's algorithm [Breiman *et al.*, 1984]. Regression trees have several advantages, among

them: 1) They can manage a high number of features and examples in an easy way; 2) They are nonparametric flexible methods so they do not impose a specific functional form to the solution; and 3) The variables, or combination of variables, used at each node to divide the samples into subgroups are the most discriminative features since they assure the lowest estimated error.

The main advantage for analysts in remote sensing applications is that regression trees allow knowledge discovery and full interpretability by analyzing the surrogate and main splits of the tree. They have been successfully used to estimate land surface variables such as LAI, fraction of photosynthetically active radiation (FAPAR) and chlorophyll content from VEGETATION/SPOT4 [Weiss and Baret, 9991], or broadband albedo from the Earth Observing 1 (EO-1) data [Liang *et al.*, 2003], just to name a few applications.

5.3.3 Neural Networks (NN)

The most common approach to develop nonparametric and nonlinear regression is based on artificial neural networks (NN) [Haykin, 1999]. A NN is a (potentially fully) connected structure of neurons organized in layers. A neuron basically performs a linear regression followed by a nonlinear function, $f(\cdot)$. Neurons of different layers are interconnected with the corresponding links (weights). Therefore, in the limit case of using a NN with only one neuron, the results would be similar (or slightly better because of the nonlinearity) than those obtained with OLS regression. Training a NN implies selecting a structure (number of hidden layers and nodes *per* layer), initialize the weights, shape of the nonlinearity, learning rate, and regularization parameters to prevent overfitting. In addition, the selection of a training algorithm and the loss function both have an impact on the final model. In this work, we used the standard multi-layer perceptron, which is a fully-connected network. We selected just one hidden layer of neurons. We optimized the NN structure using the Levenberg-Marquardt learning algorithm with a squared loss function. A cross-validation procedure was employed to avoid overfitting issues. NN weights were initialized randomly according to the Nguyen-Widrow method, and model regularization was done by limiting the maximum number of net weights to half the number of training samples. NNs have been vastly used in biophysical parameter retrieval [e.g., Smith, 1993; Gopal and Woodcock, 1996; Kimes *et al.*, 1998; Bacour *et al.*, 2006; Verger *et al.*, 2011], and are very useful in operational settings [e.g., Baret *et al.*, 2013] because they scale well with the number of training examples.

5.3.4 Kernel Ridge Regression (KRR)

Kernel Ridge Regression (KRR) minimizes the squared residuals in a higher dimensional feature space, and can be considered as the kernel version of the (regularized) OLS linear regression [Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004]. The linear regression model is defined in a Hilbert space, $\mathcal{H}$, of very high dimensionality, where samples have been mapped to through a mapping $\boldsymbol{\phi}(\mathbf{x}_i)$. In matrix notation, the model is given by $\hat{y}_i = \boldsymbol{\phi}(\mathbf{x}_i)^\top \mathbf{w} + b$.

Notationally, we want to solve a regularized OLS problem in Hilbert spaces:

$$\sum_i (y_i - \hat{y}_i)^2 + \lambda \|\mathbf{w}\|^2. \tag{5.1}$$

Taking derivatives with respect to model weights $\mathbf{w}$ and b, and equating them to zero, leads to an equivalent problem depending on the unknown mapping function $\boldsymbol{\phi}$, which in principle is unknown. The problem can be solved by applying the Representer's theorem, by which the weights can be expressed as a linear combination of mapped samples, $\mathbf{w} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}(\mathbf{x}_i)$. The prediction for a test sample $\mathbf{x}_*$ is obtained as a function of the dual weights $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^\top$ (one per sample), as follows:

$$\mathbb{E}[f(\mathbf{x}_*)] = \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y} = \mathbf{k}_{*,:}^\top \boldsymbol{\alpha}, \tag{5.2}$$

where $\mathbf{k}_{*,:}$ contains the (kernel) similarities between the test example and all training data points. Note that for obtaining the model, only the inversion of the Gram (or kernel) matrix $\mathbf{K}$ of size $n \times n$ regularized by λ is needed. We have used the RBF kernel function, whose components $[\mathbf{K}]_{ij}$ are:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2)). \tag{5.3}$$

Therefore, in KRR only the regularization parameter λ and the kernel parameter σ have to be selected. Both parameters were optimized via standard cross-validation. It is worth noting that KRR has been recently used in remote sensing applications [Camps-Valls and Bruzzone, 2009; Camps-Valls *et al.*, 2012a].

5.3.5 Gaussian Processes Regression (GPR)

Gaussian processes regression (GPR) has been recently introduced as a powerful regression tool [Rasmussen and Williams, 2006b], and applied to remote sensing data [Pasolli *et al.*, 2010; Verrelst *et al.*, 2012b, 2013a,b]. The model provides a *probabilistic* approach for learning generic regression problems with kernels. The GPR model establishes a relation between the input and the output variable (biophysical parameter) in the same way as KRR (see Eq. 5.2). However, two main advantages of GPR must be noted.

First, not only a *predictive mean* but also a *predictive variance* can be obtained:

$$\mathbb{V}[f(\mathbf{x}_*)] = \mathbf{k}_{*,*} - \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{k}_{:,*} \tag{5.4}$$

Note that the mean prediction in (5.2) is a linear combination of observations $\mathbf{y} = [y_1, \ldots, y_n]^\top$, while the predictive variance in (5.4) only depends on input data and can be taken as the difference between the prior kernel and the information given by observations about the approximation function.

The second advantage is that one can use very sophisticated kernel functions because hyperparameters can be learned efficiently by maximizing the marginal likelihood in the training

set. See [Rasmussen and Williams, 2006b; Camps-Valls *et al.*, 2009; Verrelst *et al.*, 2012b] for further details. We used a scaled anisotropic Gaussian kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{b=1}^{B} \frac{\left(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)}\right)^2}{2\sigma_b^2} \right), \tag{5.5}$$

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.

Summarizing, three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization gives the relevance of each spectrum $\mathbf{x}_i$. Second, the inverse of σ_b represents the relevance of band b. Intuitively, high values of σ_b mean that relations largely extend along that band, hence suggesting a lower informative content. Finally, a GPR model provides not only a pixel-wise prediction for each spectrum but also an uncertainty (or confidence) level for the prediction.

The previous methods are implemented from the simple Regression toolbox [Camps-Valls *et al.*, 2013], simpleR, freely available at http://www.uv.es/gcamps/code/simpleR.html. The simpleR toolbox contains simple educational code for linear regression (LR), decision trees (TREE), neural networks (NN), support vector regression (SVR), kernel ridge regression (KRR), aka Least Squares SVM, Gaussian Process Regression (GPR), and Variational Heteroscedastic Gaussian Process Regression (VHGPR). The toolbox is not explicitly included in ARTMO, but may be of interest for the reader, as it provides more regression and analysis tools.

5.4 ARTMO

ARTMO brings multiple leaf and canopy radiative transfer models (RTM) together along with essential tools required for semiautomatic retrieval of biophysical parameters in one GUI toolbox. In short, the toolbox permits the user: (1) to choose between various invertible leaf and canopy RTMs of a low to high complexity (e.g., PROSPECT-4, PROSPECT-5, DLM, 4SAIL, FLIGHT); (2) to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems; (3) to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400 to 2500 nm); (4) to generate look-up tables (LUT), which are stored in a relational SQL database management system (MySQL, version 5.5 or higher; local installment required), and finally; (5) to configure and run various retrieval scenarios using EO reflectance datasets for biophysical parameter mapping applications. ARTMO is developed in Matlab (2009 version or higher) and does not require additional Matlab toolboxes. Figure 5.1 presents ARTMO V.3's main window and a systematic overview of the drop-down menu below. To start with, in the main window a new project can be initiated, a sensor chosen and a comment added, whereas all processing modules are accessible through drop-down menus at the top bar.

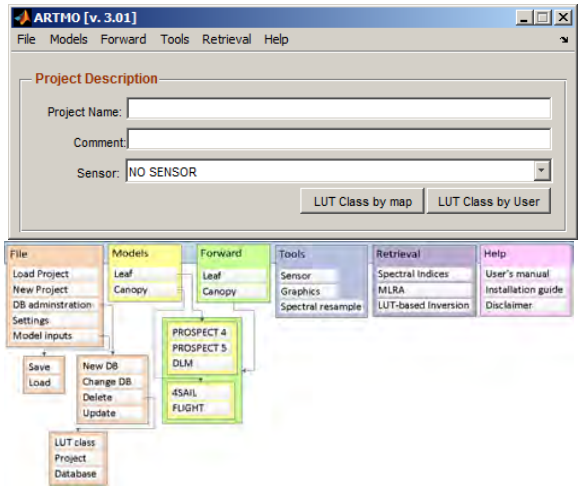


FIGURE 5.1: Screenshot of ARTMO's main window and schematic overview of its drop-down menu.

A first rudimentary version of ARTMO has been used in LUT-based inversion applications [Verrelst *et al.*, 012c, 2014b]. ARTMO v3 is formally presented in this paper. The software package is freely downloadable at: http://ipl.uv.es/artmo. Its most important novelties are briefly listed below:

- ARTMO v3 is designed modularly. Its modular architecture offers the possibility for easy addition (or removal) of components, such as RTM models and post-processing modules.
- The MySQL database is organized in such a way that it supports the modular architecture of ARTMO v3. This avoids redundancy and increases the processing speed. For instance, all spectral datasets are stored as binary objects.
- New retrieval toolboxes are incorporated. They are based on parametric and non-parametric regression as well as physically-based inversion using a LUT. This has led to the development of a: (1) 'Spectral Indices assessment toolbox' [Rivera *et al.*, 2014a]; (2) 'Machine Learning Regression Algorithm toolbox'; and, (3) 'LUT-based inversion toolbox' [Rivera *et al.*, 2013a].

This paper introduces the 'Machine Learning Regression Algorithm module'. Its general architecture is outlined in Figure 5.2.

5.4.1 MLRA Settings module

The following step addresses the analysis of multiple MLRA-based retrieval strategies. A first step to do is inserting input data (i.e. a plain text file), which refers to retrievable biophysical parameters and associated spectra. This is done in 'Input' and can be either RTM-simulated data or can be ground truth data as measured in the field. The GUI will guide the user through

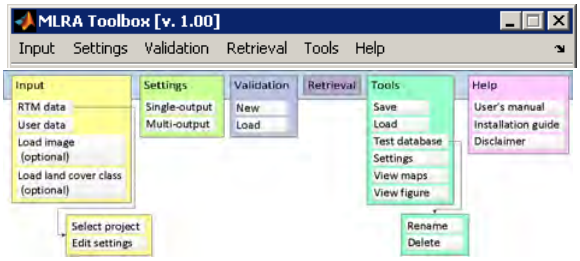


FIGURE 5.2: Screenshot of MLRA's toolbox and schematic overview of its drop-down menu.

the data selection steps and checks if data is properly read (not displayed for brevity). Once data is inserted the 'MLRA settings' module can be configured (Figure 5.3). It can be opted to select either single-output or multi-output regression algorithm. Currently, only PLSR, NN and KRR encompass multi-output capabilities. Obviously these models can also be used for single-output applications.

The 'MLRA settings' module configures the regression algorithms given various options. First, if a land cover map in ENVI format (Exelis Inc.) has been provided then retrieval strategies can be configured *per* land cover class. Second, multiple regression algorithms at once can be selected, which means that they will be analyzed one-by-one. Third, options to add Gaussian noise are provided. Noise can be added both on the parameters to retrieve and on the spectra. A range of noise level can be configured, so that multiple noise scenarios can be evaluated. The injection of noise can be of importance to account for environmental and instrumental uncertainties when synthetic spectra from RTMs are used for training. Fourth, the training/validation data partition can be controlled by setting the percentage of how much data from an RTM or user-defined is assigned to training or to validation (i.e. split-sample approach). Thereby, the user can evaluate the impact of ranging training/validation partitioning by entering a range of training/validation partitions. For each training/validation partition The MLRA toolbox internally divides the defined training set into k subsets using a k-fold cross-validation strategy in order to tune the free parameters of the model.

5.4.2 Validation module

Once that the training/validation data splitting has been defined and MLRA settings configured, a range of scenarios can be run, tested and their performance assessed. This is done by naming a validation set in the 'Validation' module. Each regression model strategy over the configured ranges are one-by-one analyzed through goodness-of-fit measures and validation results are stored in a MySQL database. As such, a large number of results can be stored in a systematic manner, so that they can be easily queried and compared. Validation results are presented in the 'MLRA validation table' (Figure 5.4). The table shows the best performing validation results according to a selected land cover class (if loaded), parameter and statical goodness-of-fit measure. Various options to display the results are provided, e.g., 1:1-line, plotting the band relevance as given by the GPR model, and 2D matrices of performances along ranges of

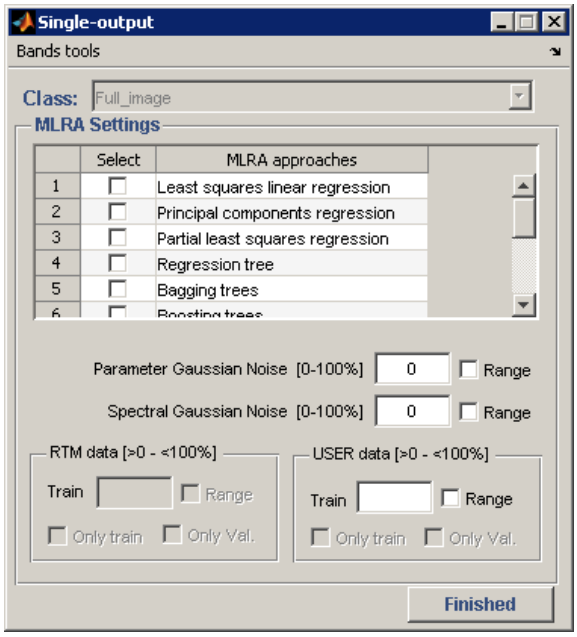


FIGURE 5.3: MLRA's setting window.

noise and varying training/validation distribution (see further in the Results section). Finally, by clicking on 'Retrieval', an analyzed regression function can be selected for each retrievable parameter (e.g. the best one). Such regression function will be accessed in the 'Retrieval' GUI and can then be applied to a remote sensing image.

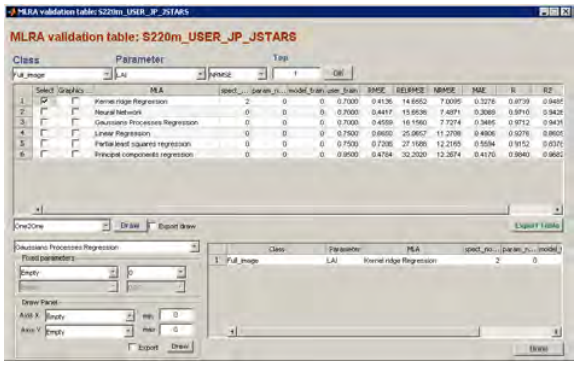


FIGURE 5.4: MLRA's validation window.

5.4.3 Retrieval module

The 'Retrieval' module enables to run an evaluated model or directly configure a model and apply it to an image (provided in standard ENVI file format) to map a parameter (Figure 5.5). Hence, the user can select the required land cover class (if available), the retrievable parameter,

the regression algorithms and training/validation splitting. Similarly, noise can be added to the spectra or parameters and the size of the training data can be selected. The user will then be invited to select one or multiple remote sensing images to which the developed model will be applied. Generated maps are stored in ENVI format.

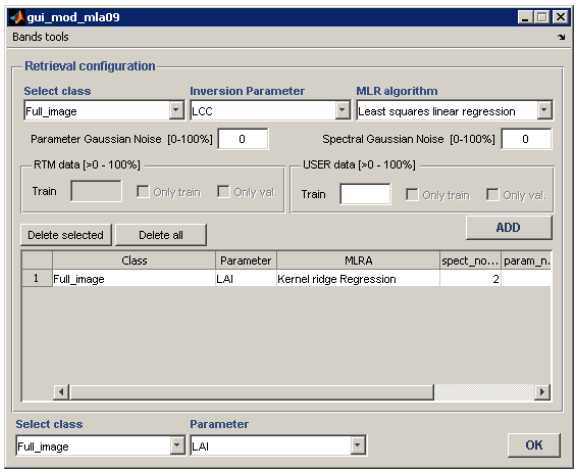


FIGURE 5.5: MLRA's retrieval window.

5.5 Mapping applications

Having the MLRA toolbox presented, it is subsequently applied for evaluating the performance of the six presented nonparametric regression techniques to achieve optimized biophysical parameters estimation. Used data is first outlined, followed by the experimental setup. Results are then presented and a mapping application is shown.

5.5.1 Used data

A diverse field dataset, covering various crop types, growing phases, canopy geometries and soil conditions was collected during SPARC (Spectra bARrax Campaign). The SPARC-2003 campaign took place during 12-14 July in Barrax, La Mancha, Spain (coordinates 30°3'N, 28°6'W, 700 m altitude). Biophysical parameters were measured within a total of 110 Elementary Sampling Units (ESU) among different crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat). ESU refers to a plot size compatible with pixel dimensions of about 20m $\times$ 20m. Leaf chlorophyll content (LCC) was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter [Gandía *et al.*, 2004] Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyser. Each ESU was assigned to a LAI value, which was obtained as a statistical mean of 24 measures (8 data readings $\times$ 3 replications) with standard errors between 5 and 10% [Fernández

et al., 2005]. In total LAI varies between 0.4 and 6.3 and LCC between 2 and 55 μg/cm^2. Additionally, 60 random spectra over bare soils, man-made surfaces and water bodies were added to broaden the dataset to non-vegetation samples (i.e., with a LCC and LAI value of zero), leading to a total of 170 samples.

During the campaign airborne hyperspectral spaceborne CHRIS images and airborne HyMap flightlines were acquired. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, size of the covered area and spatial resolution because of on-board memory storage reasons [Barnsley *et al.*, 2004]. We made use of nominal nadir CHRIS observation in Mode 1 (62 bands, maximal spectral information), which were acquired during the SPARC campaign (12 July 2003). CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 6 to 33 nm depending on the wavelength. CHRIS imagery was processed using ESA's CHRIS-Box available in VISAT/BEAM, which includes radiometric recalibration, coherent-noise reduction, geometric correction and atmospheric correction [Alonso and Moreno, 2005; Guanter *et al.*, 2005]. HyMap was configured with 125 bands between 430 and 2490 nm with bandwidths varying between 11 and 21 nm and a pixel size of 5 m. The same geometric and atmospheric preprocessing as for CHRIS was applied, but given a superior signal-to-noise ratio this sensor provides a better radiometric quality than CHRIS.

5.5.2 Experimental setup

SPARC field data was used for training and validation, and associated spectral data came from CHRIS and HyMap. In view of ESA's forthcoming Sentinel-2 (S2) mission, also S2 data at a spatial resolution of 20 m were additionally generated. S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades. It provides a set of 13 spectral bands spanning between 443 and 2190 nm, four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution [Drusch *et al.*, 2012]. Because of being spaceborne and providing similar pixel size, CHRIS data was resampled to the band settings of S2. Nearest neighbor was used for the spatial resampling and a Gaussian model with full width at halve maximum (FWHM) spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, a dataset of eight bands at 20 m (4 bands at 20 m plus 4 bands at 10 m coarse-grained at 20 m) was prepared, hereafter referred as 'S2-20m'.

The MLRA toolbox was used to evaluate the performance of the different regression algorithms along gradients of changing training/validation distributions (from 5 to 95% training, with steps of 5%; the remaining data goes to validation) and increasing Gaussian noise levels (from 0 to 20% with steps of 2%). By systematically evaluating the performance along those two dimensions in a 2D matrix format, an indication about the robustness of these regression methods can be obtained. Models were developed both for LCC and LAI. The predictive power of the developed models was evaluated with the absolute root-mean-squared error (RMSE), the normalized RMSE (NRMSE [%] = RMSE/ range parameter measurements *100) and the coefficient of determination (r^2) to account for the goodness-of-fit. Here, validation results are

presented in the form of NRMSE, which allows accuracy comparison across different parameters. Typically, remote sensing end users require an error threshold below 10%.

5.6 Results

5.6.1 Regression method evaluation

For each parameter, sensor type and regression algorithm, NRMSE results along varying training/validation distribution and increasing noise levels are presented in 2D matrices (Figure 5.6). The best performing scenario for each matrix is also shown in Table 5.1. When comparing these matrices, the following observations can be made.

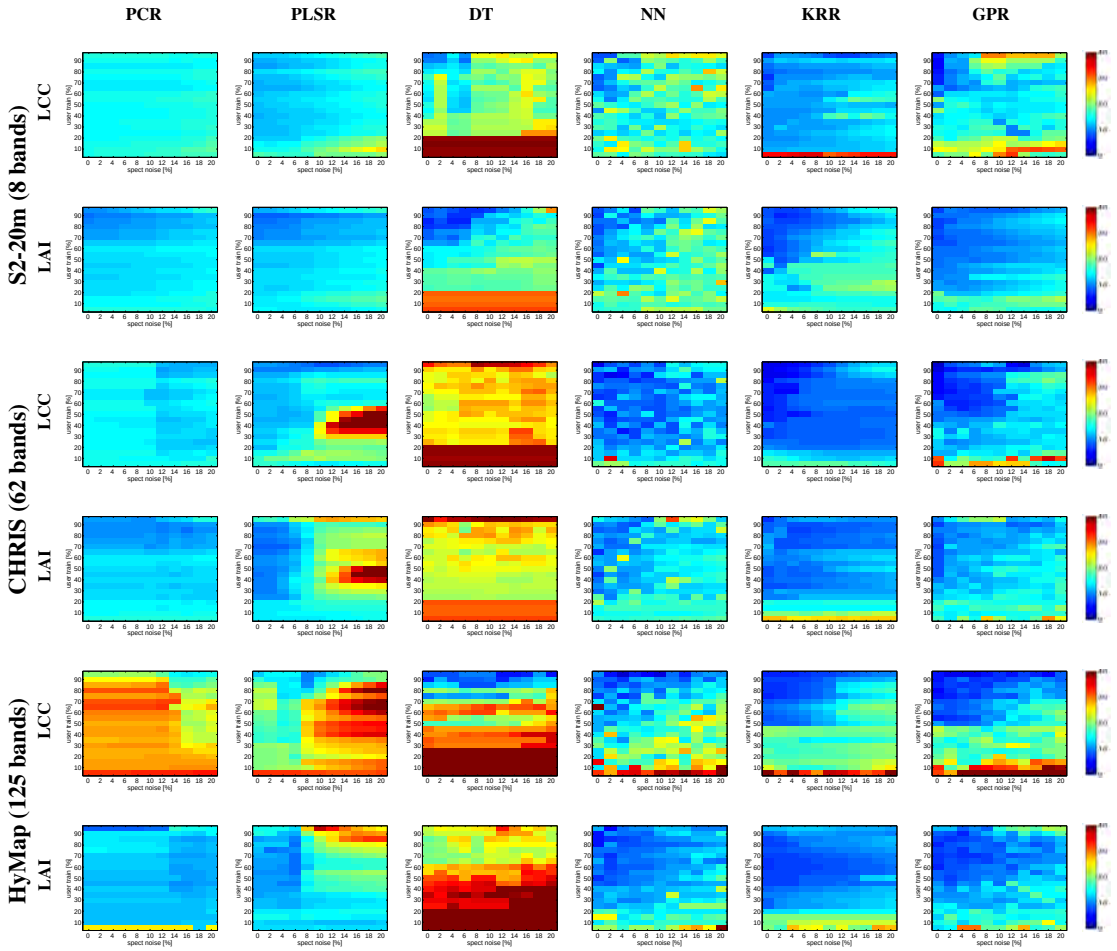


FIGURE 5.6: Normalised RMSE (NRMSE) matrices of validation results for LCC and LAI retrieval using a regression algorithm displaying the impact of % noise (X-axis) against multiple solutions (Y-axis). The more bluish, the lower relative errors and thus the better the retrieval.

Starting with principal component regression (PCR), this method proved to perform rather stable within the matrix space. For S2-20m and CHRIS data, PCR seems to be hardly affected by a varying training/validation partition and noise injection. In fact, injection of some noise rather improved accuracies of CHRIS and HyMap. Hence adding noise can lead to a closer match between training and validation data. However, results were never outstanding, and LCC prediction with HyMap data completely failed. Only for HyMap LAI results improved to r^2 up to 0.97 when 95% of the data was assigned to the training process. Therefore, on the whole, PCR is evaluated as suboptimal performing.

Second, the partial least square regression (PLSR) is an improved version of the PCR and widely used in EO applications. It systematically outperforms PCR in absence of noise. Improvements are particularly notable for LCC (r^2 up to 0.96 for CHRIS and HyMap). But PLSR is also more affected by the injection of noise. Low noise levels led to superior results, but above about 8% accuracies degraded rapidly.

Third, decision trees (DT) yielded on the whole poorest results. Particularly unacceptable poor results were obtained with low training data, and when many bands are involved. This suggests that decision trees would not be a good choice for applying to hyperspectral data unless a large database is available. In fact, only good results were obtained (r^2 up to 0.94 (LCC) and 0.97 (LAI)) in case of 'S2-20m' (8 bands) when more than 80% was used for training and below 8% noise added.

Fourth, neural networks (NN) are characterized by causing erratic patterns in each of the matrices. While being able to deliver very accurate results in some cases (e.g. for LAI using CHRIS and HyMap: r^2 up to 0.96 and 0.99) NN also showed to perform rather unstable, with large probability of delivering poor results. Particularly when more noise is involved and when less data is dedicated for training NN tends to perform more unstable. This erratic behavior can be explained by the complicated training phase whereby a highly specialized model is developed, but therefore easily faces the problem of overfitting. The lack of robustness to noise along with the complexity in training are therefore major drawbacks of NN.

Fifth, from all evaluated regression algorithms, kernel ridge regression (KRR) yielded most robust results. It led to excellent accuracies with r^2 maxima between 0.94 and 0.99 for all datasets, and more importantly also proved to perform very stable with increasing noise levels. Also it should be noted that from all tested nonlinear MLRAs, this regression technique is fastest in developing their models (see [Verrelst *et al.*, 2012a] for a quantitative comparison).

Finally, Gaussian processes regression (GPR) appears to be the most promising regression algorithm. It easily leads to excellent performances, with r^2 maxima between 0.94 and 0.99 for all datasets. Though, in comparison to KRR, GPR is somewhat more affected by noise injection. Note that the predictive mean equations for KRR and GPR are exactly the same so in principle the results should be exactly identical. Nevertheless, in GPR we used a very flexible kernel to account for different lengthscales per feature. While beneficial without noise, this turns to be a curse in these particular experiments because noise affected the marginal likelihood estimation of hyperparameters. For this reason, it can be concluded that GPR performs slightly less robust than KRR.

TABLE 5.1: Validation statistics (noise level (%), training %, RMSE , NRMSE and r^2) sorted according to best evaluated NRMSE for LCC and LAI retrieval. Best NRMSE result per sensor and biophysical parameter is bold typed.

		Regression algorithm	noise (%)	training (%)	RMSE	NRMSE (%)	r^2
S2-20m (8 bands)	**LCC**	Principal component regression	4	65	7.11	13.79	0.87
		Partial least square regression	0	85	5.58	10.90	0.92
		decision tree	6	90	4.56	8.95	0.94
		Neural network	2	95	3.94	7.72	0.96
		kernel ridge regression	0	75	3.48	6.82	0.97
		Gaussian processes regression	**0**	**90**	**2.40**	**4.71**	**0.98**
	LAI	principal component regression	0	90	0.50	9.30	0.93
		partial least square regression	0	90	0.45	8.36	0.95
		decision tree	6	85	0.33	6.09	0.97
		neural network	0	90	0.39	7.17	0.96
		kernel ridge regression	**2**	**90**	**0.36**	**6.03**	**0.98**
		Gaussian processes regression	0	80	0.38	7.03	0.95
CHRIS (62 bands)	**LCC**	principal component regression	12	90	5.68	11.14	0.91
		partial least square regression	12	95	4.07	7.99	0.96
		decision tree	0	55	10.59	20.51	0.71
		neural network	0	95	1.94	3.81	0.99
		kernel ridge regression	0	85	1.95	3.81	0.99
		Gaussian processes regression	**0**	**95**	**1.06**	**3.01**	**0.99**
	LAI	principal component regression	16	85	0.50	9.30	0.92
		partial least square regression	4	70	0.46	8.20	0.93
		decision tree	6	85	1.05	13.32	0.68
		neural network	**0**	**65**	**0.37**	**6.50**	**0.96**
		kernel ridge regression	0	95	0.25	6.70	0.97
		Gaussian processes regression	0	90	0.37	6.88	0.95
HyMap (125 bands)	**LCC**	principal component regression	16	85	6.48	13.18	0.89
		partial least square regression	8	85	3.84	7.82	0.96
		decision tree	12	90	2.25	4.69	0.98
		neural network	4	95	1.19	3.60	0.99
		kernel ridge regression	0	75	2.28	4.64	0.98
		Gaussian processes regression	**0**	**80**	**1.73**	**3.51**	**0.99**
	LAI	principal component regression	8	95	0.39	8.26	0.97
		partial least square regression	6	70	0.55	8.83	0.88
		decision tree	6	85	0.54	10.17	0.88
		neural network	**2**	**95**	**0.23**	**4.74**	**0.99**
		kernel ridge regression	0	60	0.42	6.78	0.94
		Gaussian processes regression	0	80	0.30	5.66	0.95

5.6.2 Biophysical parameter mapping

The developed models can be applied to any EO imagery given the same band settings as those presented during the training phase. From all considered regression techniques, GPR was evaluated as reaching highest accuracies for the majority of cases. Moreover, GPR has unique additional features: (1) it reveals most relevant bands when developing the model; and (2) it

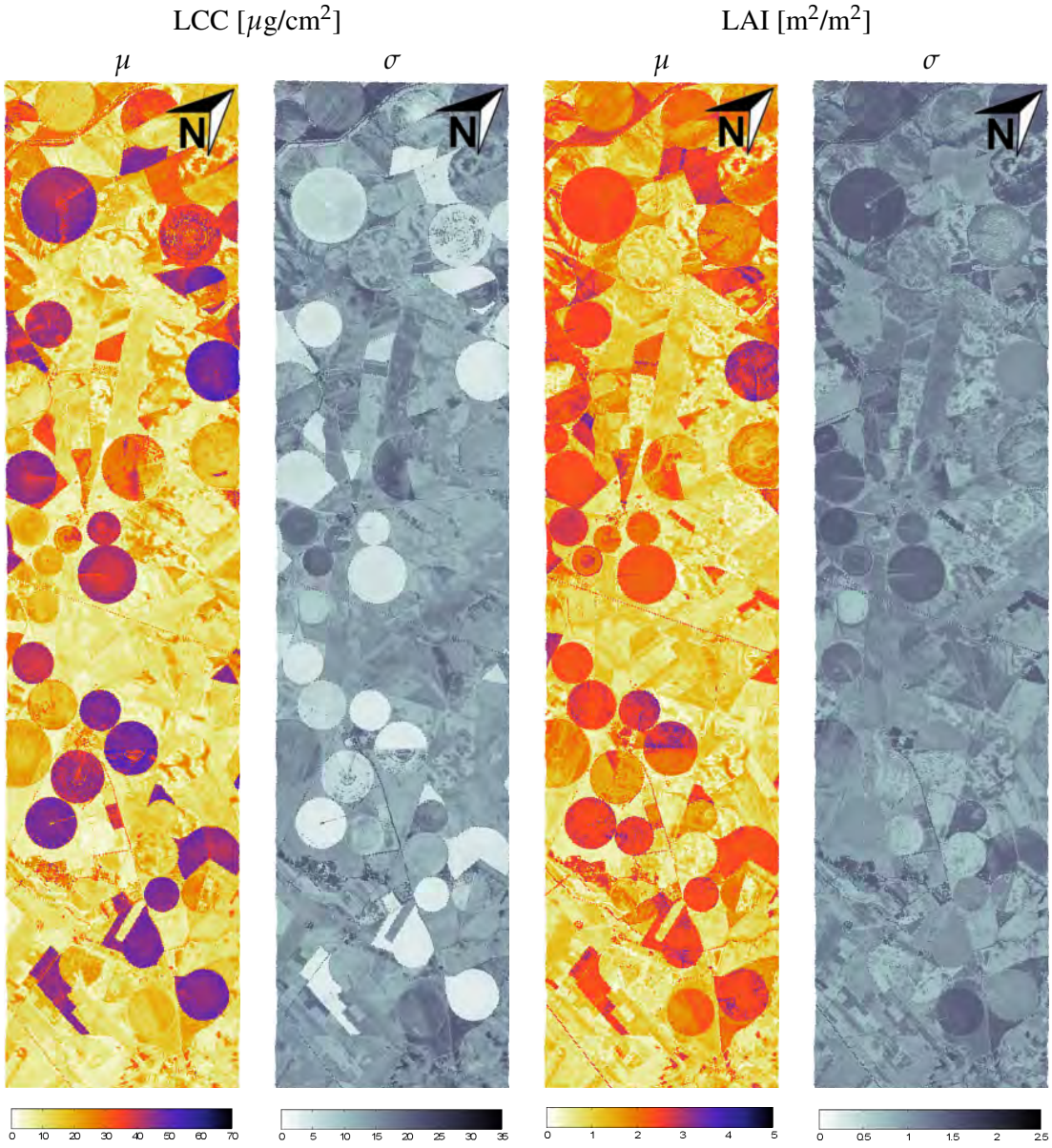


FIGURE 5.7: LCC [μg/cm^2] and LAI [m^2/m^2] mean estimates (μ) and associated uncertainties (σ) for a HyMap flightline over Barrax agricultural area (Spain).

provides uncertainty intervals (σ) associated with the mean predictions (μ). Therefore the consecutive approach was to apply GPR to a hyperspectral HyMap imagery for mapping LCC and LAI estimates along with associated uncertainty intervals (Figure 5.7). In the obtained μ maps, the irrigated circular agricultural fields are clearly differentiated, including within-field variability. In the uncertainty maps, the lower the σ (whiter color) indicate the more certain the retrieval

as processed by the trained model. The delivery of uncertainty estimates allows us to provide insight on a pixelwise basis when applied to any image and so enables the interpretation at which land covers retrievals were associated with great certainty and land covers would benefit from additional sampling. It can be observed that, particularly over the circular agricultural parcels, LCC were processed with high certainty. This is less obvious for LAI retrievals, however it should be kept in mind that σ is also related to the magnitude of the mean estimates (μ). For this reason relative uncertainties (σ / μ) may provide a more meaningful interpretation. These maps can function as a spatial mask that enables displaying only pixels with great certainty. Moreover, uncertainty maps can also give information about the portability of the regression models when applied to images over areas other than the training site [Verrelst *et al.*, 2013a,b].

5.7 Discussion

The hereby presented MLRA toolbox allows evaluating and applying a wide range of regression techniques in a semiautomatic and user-friendly way. As a case study we applied the MLRA toolbox to compare six regression algorithms on their performance and robustness along ranges of varying training/validation distribution and noise variance. These algorithms can be categorized in either data dimensionality transformations (PCR, PLSR) and nonlinear algorithms (DT, NN, KRR, GPR). For all used datasets (S2-20m, CHRIS and HyMap) pronounced differentiation in their best performances emerged. While for PCR, PLSR and DT best accuracies fell within a range of 4.7% to 20.5% (r^2: 0.86-0.97) the MLRA algorithms NN, KPR and GPR yielded higher accuracies, between 3.5% and 7.7% (r^2: 0.94-0.99). Hence, each of these MLRAs reached accuracies below 10%, which is typically demanded in operational products. These excellent performances can be explained by that MLRAs may find the nonlinear feature relations by building more flexible and adaptive models than those restricted to linear projections or regression. The excellent performance of the MLRAs becomes even more apparent when comparing against other classic retrievals methods. The same validation dataset reached r^2 on the order of 0.85 by using vegetation indices from CHRIS data [Verrelst *et al.*, 2012b], and r^2 up to 0.77 for the same S2-20m bands by using look-up table inversion of the PROSAIL radiative transfer (RT) model through cost functions [Verrelst *et al.*, 2014b]. Moreover, apart from the here evaluated algorithms, others can be added relatively easily. Meanwhile a wide array of new MLRAs have already been implemented, among others: support vector regression (SVR), extreme learning machines (ELM) and variational heteroscedastic Gaussian process regression (VHGPR) [Camps-Valls *et al.*, 2013; Lazaro-Gredilla *et al.*, 2014].

An urging open question is about evaluating how well these algorithms perform when being fed by large datasets as generated by canopy RT models. The advantage of RT models is that a broad range of land cover situations can be simulated (e.g. up to hundred thousands), leading to a dataset several times bigger than what can be collected during a field campaign. Operational processing chains typically rely on this hybrid approach [Verger *et al.*, 2008], and similar strategies could be developed by the ARTMO toolbox. It remains however to be investigated how well kernel-based MLRAs perform with large datasets. This is not a trivial point. For

instance, the computational load of the GPR increases exponentially with each added sample, making that this function faces difficulties when being trained by several thousand (in principle distinct) samples. Alternatively, dimensionality reduction techniques may largely overcome the burden of large datasets. Currently, a diversity of linear and nonlinear principal component analysis (PCA) techniques are being implemented (e.g. kernel PCA) in order apply dimensionality reduction. On the other hand, redundancy also takes place along the simulated spectra, e.g., because not all RTM parameters lead to spectral variations, causing redundancy along the dataset. Therefore, the emerging field of redundant data reduction is expected to further reduce the dataset while preserving good performance [Shen *et al.*, 2012; Wang *et al.*, 2012]. The field is also related to active learning approaches [Tuia *et al.*, 2011]. These dimensionality reduction techniques are foreseen to be implemented as well, which will eventually facilitate a smooth coupling between RTM-generated simulated spectra and powerful MLRAs for generic and operational retrieval applications.

5.8 Conclusions

ARTMO's new 'MLRA toolbox' enables applying and analyzing the predictive power of various MLRAs in a semiautomatic manner. Various regularization options have been implemented into the toolbox, e.g., training/validation data splitting, adding noise, and regression models can be developed and evaluated per land cover class. Data can either come from field campaigns or from simulations as generated by radiative transfer models. The predictive power of multiple nonparametric regression algorithms was evaluated across gradients of varying training/validation distribution and increasing noise levels. By using the local SPARC dataset and multispectral simulated sentinel-2 and hyperspectral CHRIS and HyMap imagery over the Barrax (Spain) agricultural area, kernel ridge regression (KRR) and Gaussian processes regression (GPR) emerged as most robust and best performing regression algorithms (r^2 up to 0.94-0.99 and NRMSE down to 7.0-3.0 %). Moreover, GPR provides additional uncertainty estimates on a pixelwise basis, which provides insight in the performance of the model. In all generality, the linear nonparametric algorithms such as the popular partial least squares regression (PLSR) performed systematically poorer than the nonlinear, kernel-based regression algorithms (KRR, GPR).

The presented experimental results demonstrated the utility of the MLRA toolbox, which essentially has been developed to serve efficient and optimized surface properties mapping.

6

GPR uncertainty estimates mapping

Contents

This chapter is based on:

Jochem Verrelst, Juan Pablo Rivera Caicedo, José Moreno and Gustavo Camps-Valls (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
DOI: 10.1016/J.ISPRSJPRS.2013.09.012

6.1 Abstract

ESA's upcoming Sentinel-2 (S2) Multispectral Instrument (MSI) foresees to provide continuity to land monitoring services by relying on optical payload with visible, near infrared and shortwave infrared sensors with high spectral, spatial and temporal resolution. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods, which ideally should provide uncertainty intervals for the predictions. Statistical learning regression algorithms are powerful candidates for the estimation of biophysical parameters from satellite reflectance measurements because of their ability to perform adaptive, nonlinear data fitting. In this paper, we focus on a new emerging technique in the field of Bayesian nonparametric modeling. We exploit Gaussian process regression (GPR) for retrieval, which is an accurate method that also provides uncertainty intervals along with the mean estimates. This distinct feature is not shared by other machine learning approaches. In view of implementing the regressor into operational monitoring applications, here the portability of locally trained GPR models was evaluated. Experimental data came from the ESA-led field campaign SPARC (Barrax, Spain). For various simulated S2 configurations (S2-10m, S2-20m and S2-60m) two important biophysical parameters were estimated: leaf chlorophyll content (LCC) and leaf area index (LAI). Local evaluation of an extended training dataset with more variation over bare soil sites led to improved LCC and LAI mapping with reduced uncertainties. GPR reached the 10% precision required by end users, with for LCC a NRMSE of 3.5-9.2% (r^2: 0.95-0.99) and for LAI a NRMSE of 6.5-7.3% (r^2: 0.95-0.96). The developed GPR models were subsequently applied to simulated Sentinel images over various sites. The associated uncertainty maps proved to be a good indicator for evaluating the robustness of the retrieval performance. The generally low uncertainty intervals over vegetated surfaces suggest that the locally trained GPR models are portable to other sites and conditions.

6.2 Introduction

Biophysical parameter products such as leaf chlorophyll content (LCC) and leaf area index (LAI) have become standard products by space agencies and research institutions. At the same time, proper usage of such products requires that associated uncertainty information needs to be provided [Buermann *et al.*, 2001; Morisette *et al.*, 2006; Fang *et al.*, 2012]. Two categories of product uncertainty information are generally available in the literature, either physical or theoretical [Fang *et al.*, 2012]. Physical uncertainties indicate the departure of product values from hypothetical true values and are obtained through the collection of ground-based validation data. Theoretical uncertainties are caused by uncertainties in the input data and model imperfections and are usually estimated during the retrieval process [Knyazikhin *et al.*, 1998b; Baret *et al.*, 2007; Pinty *et al.*, 2011]. Several operationally delivered products are nowadays accompanied with theoretical uncertainties in the form of a quantitative quality indicator [Knyazikhin *et al.*, 1998b; Pinty *et al.*, 2011]. An alternative method was recently proposed in Fang *et al.* [2012] where authors apply an independent uncertainty calculation through a triple collocation method. As such, independently the uncertainties of MODIS, CYCLOPES, and GLOBCARBON LAI products were evaluated. From these products, it was concluded that particularly CYCLOPES generally meet the quality requirements (± 0.5) proposed by the Global Climate Observing System (GCOS) [GCOS, 2011]. The CYCLOPES algorithm is based on a neural network (NN) trained from the 1D SAIL radiative transfer model (RTM) [Baret *et al.*, 2007].

While NN have proven robust in various operational processing chains, the retrieval schemes are far from being perfect and various limitations have been identified. For instance, LAI estimates from CYCLOPES are less accurate at higher values due to the saturation effect in the radiative transfer simulation and the NN inversion algorithm [Bacour *et al.*, 2006; Weiss *et al.*, 2007]. Furthermore, NN not only behave as a black box model, but they are also relatively unpredictable when used with input spectra that deviate (even slightly) from what has been presented during the training stage [Atzberger, 2004; Baret and Buis, 2008]. It is therefore to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insights in the retrievals and evaluating retrieval performances. Besides, training NNs involve tuning several parameters that may greatly impact the final robustness of the model. For these reasons, alternatives that overcome these limitations are needed. In part, this is why in the recent years NNs are being replaced by other more advanced, simpler to train, machine learning regression algorithms (MLRAs). Actually, during the last two decades, the family of *kernel methods* [Camps-Valls and Bruzzone, 2009] has emerged as an alternative to NNs in many applications. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input-output nonlinear mappings. They are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. Intimately related to the field of kernel methods, here we find the new emerging field of Bayesian non-parametric modeling. The framework of Bayesian nonparametrics gives a Bayesian treatment of statistical inference. This field has given rise to particularly powerful methods, such as relevance vector machines and Gaussian processes. These methods are able to provide high accuracies and at the same time provide uncertainty intervals for

the predictions [e.g. Camps-Valls *et al.*, 2006; Verrelst *et al.*, 2012b]. They may therefore be more suitable candidates for operational applications, especially now that Earth observation is reaching a mature state.

In 2014, the European Space Agency's (ESA) forthcoming Sentinel-2 (S2) mission will start delivering high-resolution optical images on a global scale. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods. Recently, Verrelst *et al.* [2012a] have tested the capabilities of four state-of-the-art MLRAs given different Sentinel-2 and Sentinel-3 band settings. Selected MLRAs were NNs, support vector regression, kernel ridge regression, and Gaussian processes regression (GPR). The methods were compared in terms of accuracy, goodness of fit, robustness to low sample sized scenarios, and computational cost. Training and validation data came from the ESA-led field campaign SPARC, which took place on the agricultural test site Barrax, Spain. The main conclusion of this work was that, in general, GPR outperformed the other regression algorithms in terms of speed and computational costs. At the same time, GPR yielded superior accuracies for the majority of tested cases. Moreover, in contrary to NN, GPR provide directly theoretical uncertainty estimates through Gaussian probability (cf. Section 6.4). These uncertainty estimates opened a new source of information. For instance, they make possible to assess the robustness of the retrievals at various spatial scales. In Verrelst *et al.* [2013a], retrievals from hyperspectral airborne and spaceborne data over the Barrax area were compared. In this way, the uncertainty measure provided information about the upscaling quality, i.e., if the uncertainties are kept constant then the upscaling can be considered stable. Even though retrievals proved to be robust over vegetated areas, high uncertainties appeared over non-vegetated surfaces, which suggested that the training dataset was not representative enough for those land cover types. Furthermore, since statistical approaches are often criticized because of limited generalization and transferability [e.g. Colombo *et al.*, 2003; Meroni *et al.*, 2004], it remains to be questioned how robust the locally-trained GPR models function when applied to other sites and conditions. In this respect, the delivery of additional uncertainty estimates may enable to evaluate the portability of the regression model. Specifically, when uncertainty intervals as produced by a locally trained GPR model over an arbitrary site are on the same order as those produced over the successfully validated reference site, then it can be reasonably assumed that the parameter retrievals are also of the same quality as the retrievals of the reference site. Thus, when successfully validated over a reference imagery then the uncertainty estimates can work as a quality indicator. This concept has been evaluated here, and brings us to the following specific objectives of the present paper: (1) to evaluate the impact of experimental training data on the development of GPR models, particularly in view of improved retrievals over non-vegetated surfaces, and (2) to evaluate the portability of a locally trained GPR model to other sites and conditions by making use of its associated uncertainty intervals.

The remainder of the paper is organized as follows. Section 6.3 briefly describes the Sentinel-2 concept, while section 6.4 revises the Bayesian nonparametric field in general and the Gaussian process regression algorithm in particular. In the methodology (section 6.5) the used training dataset, experimental Sentinel-2 images, and experimental setup are described. Results focus first on a local evaluation of the GPR models (section 6.6.2) and then moves to

the evaluation of algorithm's performance on other sites (section 6.6.3). Section 6.7 discusses main findings and section 6.8 concludes the paper.

6.3 Sentinel-2

ESA's S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades [Drusch *et al.*, 2012]. S2 is a polar-orbiting, superspectral high-resolution imaging mission. The mission is envisaged to fly a pair of satellites with the first planned to launch in 2014. Each S2 satellite carries a Multi-Spectral Imager (MSI) with a swath of 290 km. It provides a versatile set of 13 spectral bands spanning from the visible and near infrared (VNIR) to the shortwave infrared (SWIR), featuring four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Table 6.1). S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. The pair of S2 satellites aims to deliver data taken over all land surfaces and coastal zones every five days under cloud-free conditions, and typically every 15-30 days considering the presence of clouds. To serve the objectives of Copernicus (The European Earth Observation Programme), S2 satellites will provide data for the generation of high-level operational products (level 2b/3) such as land-cover and land-change detection maps and geophysical variables such as LCC, LAI and leaf water content maps. To ensure that the final products can meet user requirements, the Copernicus user committee defined an accuracy goal of 10% [Drusch *et al.*, 2012].

TABLE 6.1: Sentinel-2 MSI band settings.

Band #	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Band center (nm)	443	490	560	665	705	740	783	842	865	945	1375	1610	2190
Band width (nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

6.4 Bayesian nonparametrics and Gaussian processes

Finding a functional relation between input (e.g. reflectances) and ouput (e.g. physical parameter) variables is the main goal of statistical learning. The problem is complex and elusive because possibly an infinite number of functions can be found to fit the data. This problem is known as the excess of capacity of the class of functions implementing the model, and has been referred to the problem of overfitting. The key is to constrain model's capacity, in either one of the following two ways: imposing strong prior knowledge or via regularization schemes that promote simpler solutions. In the last decades, statistical learning (inference) has witnessed an overwhelming interest in kernel methods because they implement nonlinear models

and still rely on linear algebra operations [Camps-Valls and Bruzzone, 2009]. Kernel methods are very appealing for physical parameter retrieval, mainly because they deal efficiently with low-sized datasets of potentially high dimensionality, which are the situations we find in parameter retrieval from multispectral or hyperspectral imagery using models trained by field campaign data. Also, since kernel methods do not assume an explicit prior data distribution but are inherently non-parametric models, they cope well with remote sensing data specificities and complexities [Camps-Valls and Bruzzone, 2009].

In the context of statistical inference and for remote sensing products, one is not solely interested in high accuracies of the algorithm but also in producing uncertainty intervals for the predictions. This calls for a Bayesian treatment of the inference problem. While recently kernel methods have advanced the field of remote sensing data analysis [Camps-Valls and Bruzzone, 2009], the problem of uncertainty estimation with nonparametric models has been elusive in most of the approaches. In this context, the emerging field of *non-parametric Bayesian modeling* constitutes a proper theoretical framework to tackle the problem of physical parameter retrieval [O'Hagan, 1994; Rasmussen and Williams, 2006b; Orbanz and Teh, 2010][1]. Essentially, a nonparametric Bayesian model is a Bayesian model on an infinite-dimensional parameter space, which corresponds to the set of possible patterns, e.g. the class of smooth functions for regression (retrieval). The field has been very active in the last decade, and has delivered successful model instantiations. Some of them has been actually introduced in the field of remote sensing data analysis. For example, the relevance vector machine (RVM) introduced in Tipping [2001] is a nonparametric Bayesian model that assumes a Gaussian prior over the weights in order to enforce sparsity, and uses expectation-maximization to infer the parameters. In Camps-Valls *et al.* [2006] the RVM model was used for oceanic chlorophyll content estimation. The model, however, may incur in too sparse solutions that do not fit well in parts of the space not considered in the retained relevant vectors. Lately, Bayesian nonparametric modeling with Gaussian Processes [Rasmussen and Williams, 2006b] have received much attention in the field of machine learning, and has been also introduced in the remote sensing application field [Verrelst *et al.*, 2012b,a, 2013a]. This paper will focus on GPR.

GPR provides a probabilistic (Bayesian) approach for learning generic regression problems with kernels [Rasmussen and Williams, 2006b]. The GPR model establishes a relation between the input (B-bands spectra) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (canopy parameter) $y \in \mathbb{R}$ of the form:

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}), \tag{6.1}$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the spectra used in the training phase, $\alpha_i \in \mathbb{R}$ is the weight assigned to each one of them, and K is a function evaluating the similarity between the test spectrum $\mathbf{x}$ and all N

[1] Excellent online lectures on Bayesian nonparametrics are available at:
http://videolectures.net/mlss09uk_teh_nbm/ and
http://videolectures.net/mlss09uk_orbanz_fnbm/

training spectra, $\mathbf{x}_i$, $i = 1, \ldots, N$. We used a scaled Gaussian kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{b=1}^{B} \frac{\left(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)}\right)^2}{2\sigma_b^2} \right) + \delta_{ij} \cdot \sigma_n^2, \tag{6.2}$$

where ν is a scaling factor, B is the number of bands, σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b, σ_n is the noise standard deviation and δ_{ij} is the Kronecker's symbol. The kernel is thus parametrized by signal (ν, σ_b) and noise (σ_n) hyperparameters, collectively denoted as $\boldsymbol{\theta} = \{\nu, \sigma_b, \sigma_n\}$.

For training purposes, we assume that the observed variable is formed by noisy observations of the true underlying function $y = f(\mathbf{x}) + \epsilon$. Moreover we assume the noise to be additive independently identically Gaussian distributed with zero mean and variance σ_n. Let us define the stacked output values $\mathbf{y} = (y_1, \ldots, y_n)^\top$, the covariance terms of the test point $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n))^\top$, and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$. From the previous model assumption, the output values are distributed according to:

$$\begin{pmatrix} \mathbf{y} \\ f(x_*) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} \end{pmatrix}\right) \tag{6.3}$$

For prediction purposes, the GPR is obtained by computing the posterior distribution over the unknown output $\mathbf{y}_*$, $p(\mathbf{y}_*|\mathbf{x}_*, \mathcal{D})$, where $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \ldots, N\}$ is the training dataset. Interestingly, this posterior can be shown to be a Gaussian distribution, $p(y_*|\mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*})$, for which one can estimate the *predictive mean* (point-wise predictions):

$$\mu_{\mathrm{GP}*} = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}, \tag{6.4}$$

and the *predictive variance* (confidence intervals):

$$\sigma^2_{\mathrm{GP}*} = k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*. \tag{6.5}$$

The corresponding hyperparameters $\boldsymbol{\theta}$ are typically selected by Type-II Maximum Likelihood, using the marginal likelihood (also called *evidence*) of the observations, which is also analytical. When the derivatives of the log-evidence are also analytical, which is often the case, conjugated gradient ascent is typically used for optimization (see [Rasmussen and Williams, 2006b; Camps-Valls *et al.*, 2009] for further details).

Three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization gives the relevance of each spectrum $\mathbf{x}_i$. The predictive mean is essentially a weighted average of the canopy parameter values associated to the training samples closest to the test sample. Second, the inverse of σ_b represents the relevance of band b. Intuitively, high values of σ_b mean that relations largely extend along that band hence suggesting a lower informative content. These features have been extensively studied in [Verrelst *et al.*, 2012b,a] and proved to be valuable for gaining insight in relevant bands. Finally, a GPR model provides not only a pixelwise prediction for each spectrum but also an

uncertainty (or confidence) level for the prediction. Hence in contrary to other approaches (e.g. NN) uncertainty intervals are directly delivered along with mean estimates. The interested reader is referred to the book by Rasmussen and Williams [2006b] for more details on the theoretical aspects of GPR. A Matlab implementation of GPR is freely available at http://www.gaussianprocess.org/gpml/.

We illustrate the uncertainty intervals estimated by GPR and through standard boostrapping [Wu, 1986] of the nonlinear regression solution in Figure 6.1. Note that GPR basically accounts for a reduction of uncertainty based on the relative local density of the input data points, not the outputs. This is obvious by looking at the predictive variance equation above. On the other hand, the variance of the bootstrap variance estimate quickly vanishes if one moves away from data points in the set. This behavior correctly reflects the fact that the predictions will be practically zero far away from points in the training sets. Gaussian processes indicate that the uncertainty is high because no data has been observed in that area.

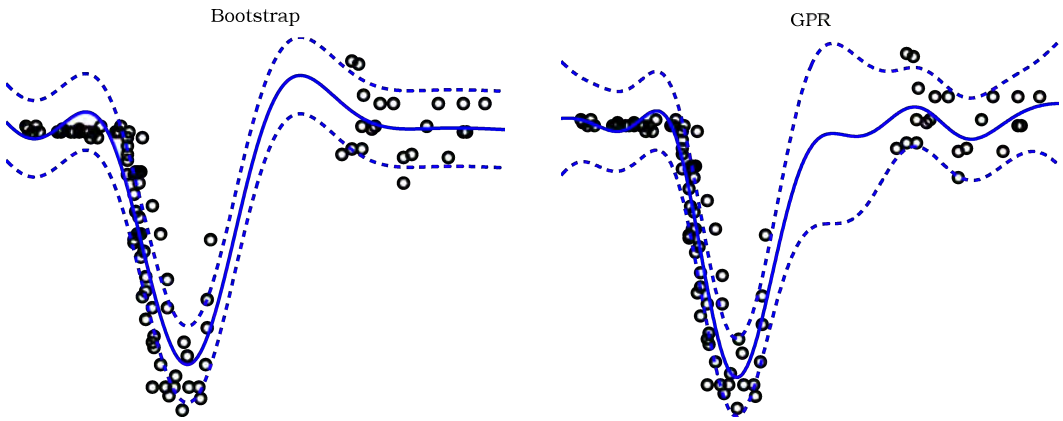


FIGURE 6.1: Toy regression example in 1-D: we predict the y-values from the x-values in a synthetic signal generated by a combination of Gaussian bumps. We show the predictive variance or confidence interval estimates via bootstrap variance estimates (left), and Gaussian process regression (right).

6.5 Methodology

6.5.1 SPARC database

GPR has been trained with a local ground dataset coming from SPARC (SPectra bARrax Campaign). The SPARC-2003 and SPARC-2004 campaigns took place in the Barrax agricultural site in Central Spain (coordinates 39°N, -2°1'E, 700 m altitude). The test area has a rectangular form and an extent of 5 km $\times$ 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. The annual rainfall average is about 400 mm.

In the 2003 campaign (12-14 July) biophysical parameters were measured within a total of 113 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size of about 20^2 m. The same field data were collected in the 2004 campaign (15-16 July) within a total of 18 ESUs among different crops. For both years, within each ESU the averaged leaf LCC was derived by measuring about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned to a LAI value, which was obtained by the average of 24 measures (8 data readings $\times$ 3 replications). In total 9 crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat) were sampled, with field-measured values of LAI that vary between 0.4 and 6.3 and LCC between 2 and 55 μg/cm^2. Further details on the measurements can be found in the data acquisition report [Moreno and participants of the SPARC campaigns, 2004]. Additionally, 30 random bare soil spectra with a biophysical (LCC, LAI) value of zero were added to broaden the dataset to non-vegetated samples. This 'original' dataset used in Verrelst *et al.* [2012b,a, 2014b] in training GPR models and hereafter refers to 'training original' or '$TrOr$'.

In the latter studies, the GPR uncertainty maps showed that mean estimates over vegetated areas were associated to low uncertainties. At the same time, considerably higher uncertainties were encountered over areas of fallow land and bare soils. These higher uncertainties can be attributed to the relatively poor contribution of non-vegetated land cover types in the training dataset. For this reason, we have extended the SPARC training dataset with 60 new spectra that cover all kinds of non-vegetated surfaces, i.e. spectra with an LCC and LAI of zero. Most of the spectra were taken over bare soil surfaces, but also man-made surfaces (e.g., build-up areas, roads) and water bodies have been included. This 'extended' training dataset hereafter refers to 'training extended' or '$TrEx$'.

6.5.2 Reference and other simulated Sentinel-2 images

Because actual S2 data is not available yet, we opted for simulating it on the basis of Compact High Resolution Imaging Spectrometry (CHRIS) data. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, site of the covered area and spatial resolution because of on-board memory storage reasons. The radiometric resolution of CHRIS is 12 bits, which is the same as S2's MSI. We made use of nominal nadir CHRIS observations in Mode 1 (62 bands, maximal spectral information) for the four SPARC campaign days, where field measurements of surface properties were measured in conjunction with satellite overpasses. CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. The images were corrected for atmospheric effects according to the method proposed in Guanter *et al.* [2005]. Since most of ground truth data were collected during the 2003 campaign, the nadir image from 14 July 2003 was used as reference image for spectral and spatial resampling to the settings of S2. The image is shown in Figure 6.2. The majority of ESUs are located on the circular green fields, while non-vegetated samples came from the yellowish-white surfaces. Because configured with

different pixel sizes (10, 20 and 60 m), it is of special interest to simulate S2 bands as a function of pixel's size. A nearest neighbor strategy was used for the spatial resampling and a Gaussian model with full-width-half-maximum spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, experimental data according to the following three Sentinel settings were generated, 'S2-10m': four bands at 10 m, 'S2-20m': eight bands at 20 m (4 bands at 20 m plus the S2-10m bands coarse-grained at 20 m), and 'S2-60m': ten bands at 60 m (2 bands at 60 m plus the earlier bands coarse-grained at 60 m).

Since the objective of this work was to evaluate the ability of transferring Barrax-trained GPR models for the various Sentinel settings to other images, multiple Mode 1 CHRIS images over terrestrial surfaces were arbitrarily collected. The only requirement was that they are predominantly cloud-free and acquired during spring or summertime. The CHRIS images include: multi-temporal images over Barrax (Spain); July 2004, and June 2009. Multitemporal images over Demmin (Germany), May and July 2006, and an image over Los Monegros (Spain), August 2006, Las Tablas (Spain), July 2006 and Sudbury (Canada), August 2007. These sites are described in Table 6.2. All these CHRIS images were corrected for atmospheric effects according to Guanter *et al.* [2005]. A cloud masking over the Sudbury image was applied to remove the pixels with cloud contamination. Although not really necessary for this exercise, the Barrax images were also geometrically corrected according to Alonso and Moreno [2005]. The images were subsequently resampled according to above-described Sentinel settings (S2-10m, S2-20m and S2-60m). The images are displayed in Figure 6.2.

TABLE 6.2: Description of test sites.

Site	**Acquisition time**	**Description**
Barrax, Spain (39°N, -2°E) Barrax, Spain (39°N, -2°E) Barrax, Spain (39°N, -2°E)	14 July 03 16 July 04 19 June 09	The Barrax agricultural area has a rectangular form and an extent of 5 km × 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. The annual rainfall average is about 400 mm.
Demmin, Germany (53.5°N, -13.1°E) Demmin, Germany (53.5°N, -13.1°E)	08 May 06 07 July 06	The Demmin agricultural area is located in Northeast Germany and is based on a group of farms covering app. 25000 ha. The surface is flat at 50 m a.s.l. with some slopes along the Tollense River. The field sizes are large in this area for German standards, about 80-250 ha. The main crops grown are wheat, barley, rape, maize and sugar. The annual rainfall ranges from 500 to 650 nm.
Monegros, Spain (41.2°N, -0.1°1'E)	20 Aug 05	Los Monegros area is a semi-arid region, sparse vegetation and shallow as well as poorly developed soils. Agricultural activities are poor and may trigger land degradation processes. The area is characterized by various small playa lakes. These lakes form in small karstic depression by the dissolution of evaporitic subsurface layers, mainly gypsum and limestone. The playa lakes are usually dry in summer. The annual rainfall average is low, about 350 mm.
Las Tablas, Spain (39.1°N, -3.4°E)	03 July 06	Las Tablas de Daimiel is a Complex of shallow pools and associated marshland, which lies in the great plain of La Mancha. The surface is flat at 600-620 m a.s.l. The site receives floodwater from the permanent freshwater Río Guadiana and the seasonal brackish Río Gigüela, and groundwater from an underground water basin. The surrounding area is used for rain-fed agriculture, mainly olives and wine yards. The annual average rainfall is about 450 mm.
Sudbury, Canada (47.1°N, -81.4°E)	10 August 07	Sudbury is a boreal forested region located in the northern part of Ontario. It is a flat area 350 m a.s.l. The area is predominantly covered by black spruce and aspen stands with an understory of shrubs and herbs. The annual rainfall average is about 820 mm.

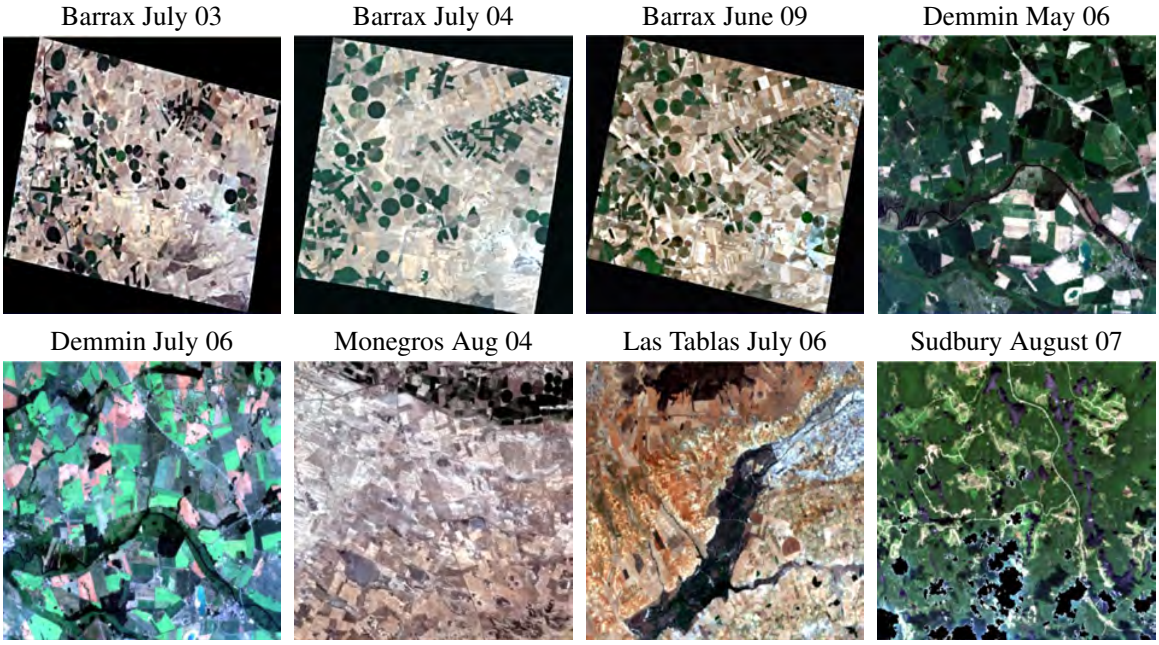


FIGURE 6.2: RGB compositions of CHRIS images used for evaluating the performance of the locally-trained GPR models. Barrax July 03 is the reference image.

6.5.3 Experimental setup

The local *TrOr* and *TrEx* experimental datasets were divided into two subsets: 80% for training and the remaining 20% for validation. The subsequent undertaken approach was straightforward; for each biophysical parameter (LCC, LAI) and each S2 configuration (S2-10m, S2-20m, S2-60m) a *TrOr* and *TrEx* model were trained. The predictive power of the developed models was subsequently validated against the 20% validation data by using the absolute root-mean-squared error (RMSE) and the normalized RMSE (NRMSE [%] = RMSE / range of the parameters as measured in the field *100) to assess accuracy, and the coefficient of determination (r^2) to account for the goodness-of-fit. The NRMSE was used to compare the performances across the different methods and parameters. Once successfully validated, the *TrOr*- and *TrEx*-developed models were applied to the other experimental S2 images and the mean estimates and associated uncertainties were compared. To preserve a physical meaning, negative LCC or LAI estimates were automatically converted to near-zero values (0.0001) during the retrieval process.

6.6 Results

6.6.1 GPR performance with original and extended training dataset

The evaluation of the *TrOr*- and *TrEx*-developed GPR models starts with inspection of the validation results (Table 6.3). Excellent prediction accuracies were obtained for all scenarios with a r^2 between 0.92 and 0.99 and NRMSE between 3.5 and 10.5%. Noteworthy hereby is that excellent LAI accuracies were already achieved with only 4 bands (S2-10m). This is encouraging for LAI mapping at high spatial resolution. The inclusion of more bands only improved accuracies marginally. Conversely, LCC clearly gained from the inclusion of extra bands, with a S2-20m configuration (8 bands: B2-B8a) that managed to reach NRMSE down to 3.5% (r^2: 0.99). It is well known that the inclusion of red-edge bands (B5 and B6) can considerably improve biophysical parameter estimation [e.g. Delegido *et al.*, 2011; Verrelst *et al.*, 2012a; Delegido *et al.*, 2013]. However, the addition of 2 more bands in the blue (B1) and NIR (B9) at a coarser resolution of 60 m (S2-60m) slightly degraded accuracies.

Of more relevance here is the comparison of validation results as achieved by using models developed using *TrOr* and *TrEx*. It can be noticed that *TrEx* yielded slightly improved results. This is especially the case for LAI, where for all S2 scenarios NRMSE dropped with about 2%. Hence, the GPR model is able to incorporate more samples without losing accuracy. In fact the contrary occurred; for all S2 configurations, the NRMSE kept below the threshold of 10%.

Even though successfully validated, it is well understood that a validation dataset is rather limited as compared to the total variability observed in satellite images. Therefore, it is expected that inspection of the mean estimate and associated uncertainty maps on complete scenes will allow us to better understand the performances of both GPR models.

TABLE 6.3: Validation statistics (r^2, absolute RMSE and NRMSE (%)) for the Sentinel configurations and the parameters using *TrOr*- and *TrEx*-developed GPR models.

	r^2		abs. RMSE		NRMSE (%)	
Parameter	*TrOr*	*TrEx*	*TrOr*	*TrEx*	*TrOr*	*TrEx*
S2-10m:						
LCC	0.931	0.949	5.36	4.70	10.50	9.21
LAI	0.910	0.948	0.51	0.39	9.37	7.28
S2-20m:						
LCC	0.991	0.993	1.92	1.77	3.76	3.48
LAI	0.916	0.958	0.49	0.35	9.09	6.51
S2-60m:						
LCC	0.977	0.993	2.37	1.81	6.49	3.55
LAI	0.934	0.959	0.29	0.36	8.42	6.63

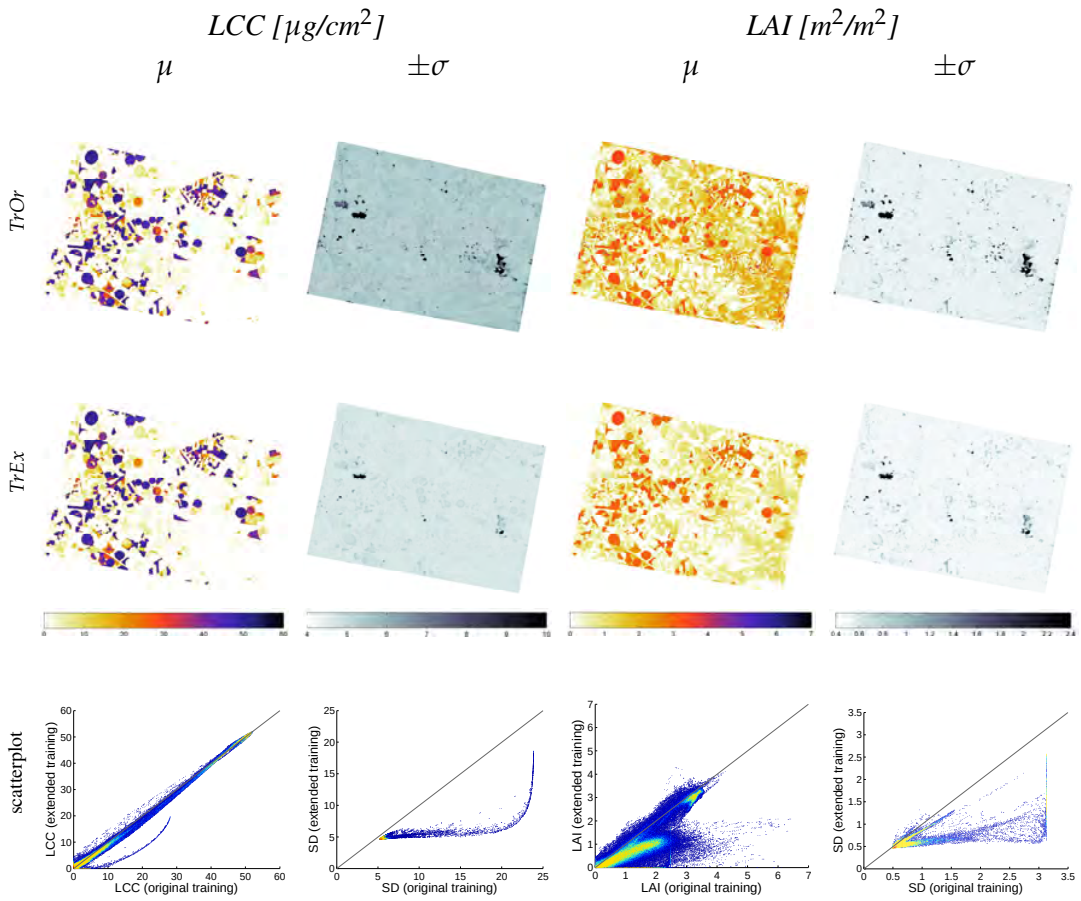


FIGURE 6.3: Mean estimate (μ) and associated uncertainty ($\pm\sigma$) maps for LCC and LAI as generated by the original training dataset ($TrOr$) [top], and extended training dataset ($TrEx$) [middle]. Scatterplots are shown below.

6.6.2 Local mapping over Barrax, Spain

Although LCC and LAI maps were generated for all S2 settings from the 2003 Barrax reference site, for the sake of brevity we only display generated maps at a high spatial resolution of 10 m. In turn the maps will allows us to appreciate the strength of GPR. Biophysical parameter retrieval was thus achieved with 4 bands only: 490, 560, 665 and 842 nm. The retrieval process was completely automated and image-based; the generation of a map was completed almost instantaneously.

Figure 6.3[top] provides mean estimate (μ) maps and associated uncertainties ($\pm\sigma$) over the Barrax test site as generated by the original training dataset ($TrOr$). The mean estimate maps are first briefly interpreted. Within-field variations are clearly detected in both maps. Particularly, the pronounced spatial variation of LCC marks prominently the irrigated circular fields with green biomass. These irrigated fields are characterized by an LCC above 40, and an LAI above 3. Areas with low LCC and LAI (the whitish parts) are mainly bare soils, fallow lands

or rain-fed senescent or harvested cereal fields (wheat, barley). The same maps are provided below, but then generated by the extended training dataset ($TrEx$). When comparing $TrOr$- and $TrEx$-generated LCC maps it can be observed that they are very similar. That similarity is also reflected when correlating both maps in a scatter plot, as displayed below. The large majority of the pixels fall right on the 1:1-line. In turn, more differences between $TrOr$- and $TrEx$-generated LAI maps occurred over non-green vegetated areas (e.g. fallow land and bare soils). The addition of bare soil spectra in the training dataset led thus to more meaningful LAI retrievals for a considerably amount of pixels, as was also observed by the validation dataset in Table 6.3. The scatter plot shows that the $TrEx$-developed LAI model causes that a substantial part of pixels are interpreted towards lower LAI estimates. While this may imply an improved accuracy, the associated uncertainty intervals will manifest the quality of the estimates.

Within these uncertainty maps, areas with reliable retrievals can be clearly distinguished from areas with unreliable retrievals. These differences are more obviously observed in the $TrOr$-generated maps. Reliable retrievals (low $\pm\sigma$) were found on irrigated areas and harvested fields. This is not surprising since the majority of training and validation samples came from these fields. Unreliable retrievals (high $\pm\sigma$) were found on areas with remarkably different spectra, such as bright, whitish calcareous soils (center, right), or harvested rain-fed barley fields with remaining bright straws covering the surface (center). Hence, as earlier noticed in Verrelst *et al.* [2012b,a], a practical implication of the uncertainty maps is the detection of areas that may benefit from a denser ground truth sampling regime. That was also the rationale for the collection of an extended training dataset over non-vegetated targets (i.e. $TrEx$). For both biophysical parameters, $TrEx$ resulted in reduced uncertainties across the whole map (see also statistics in Table 6.4), and regions with large uncertainties have been considerably reduced. That trend is again visualized by the scatter plots displayed down Figure 6.3; the majority of pixels fall below the 1:1-line. Consequently, the extended training dataset demonstrated its superiority; it was validated with high accuracy, yielded realistic LCC and LAI maps, and these maps were delivered with lower uncertainties.

TABLE 6.4: Overview statistics (mean ($\overline{x}$) and standard deviation (SD)) for LCC and LAI μ (mean prediction) and $\pm\sigma$ (associated uncertainty) maps as retrieved by $TrOr$- and $TrEx$-developed GPR models. 'Barrax July 03' is the reference image.

	μ **LCC**		$\pm\sigma$ **LCC**		μ **LAI**		$\pm\sigma$ **LAI**	
	TrOr	*TrEx*	*TrOr*	*TrEx*	*TrOr*	*TrEx*	*TrOr*	*TrEx*
Image	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)	$\overline{x}$ (SD)
Barrax July 03	7.71 (14.89)	7.54 (14.76)	5.53 (1.04)	4.75 (0.41)	1.21 (1.00)	0.82 (0.89)	0.58 (0.26)	0.51 (0.13)
Barrax July 04	9.48 (14.75)	9.22 (14.54)	5.50 (1.03)	4.75 (0.39)	1.37 (1.02)	1.13 (0.91)	0.60 (0.32)	0.51 (0.14)
Barrax June 09	12.44 (11.34)	10.66 (11.49)	6.10 (1.39)	4.98 (0.51)	0.91 (0.99)	0.73 (1.07)	0.73 (0.32)	0.73 (0.32)
Demmin May 06	25.38 (18.35)	24.31 (18.61)	5.75 (0.50)	5.00 (0.25)	2.37 (0.96)	1.70 (1.02)	0.66 (0.19)	0.59 (0.09)
Demmin July 06	26.60 (11.08)	25.69 (11.78)	5.53 (0.23)	4.93 (0.14)	1.85 (0.65)	1.37 (0.58)	0.55 (0.08)	0.55 (0.06)
Monegros Aug 05	2.16 (7.24)	2.42 (6.95)	7.22 (2.05)	4.96 (0.27)	1.81 (1.24)	0.24 (0.46)	1.44 (0.88)	0.67 (0.21)
Las Tablas July 06	7.99 (7.61)	7.42 (7.12)	5.67 (0.88)	4.78 (0.19	0.66 (0.83)	0.24 (0.38)	0.65 (0.35)	0.55 (0.11))
Sudbury Aug 07	6.92 (4.67)	3.65 (4.71)	8.30 (1.40)	5.97 (0.47)	2.32 (0.27)	0.63 (0.44)	1.80 (0.52)	0.96 (0.23)

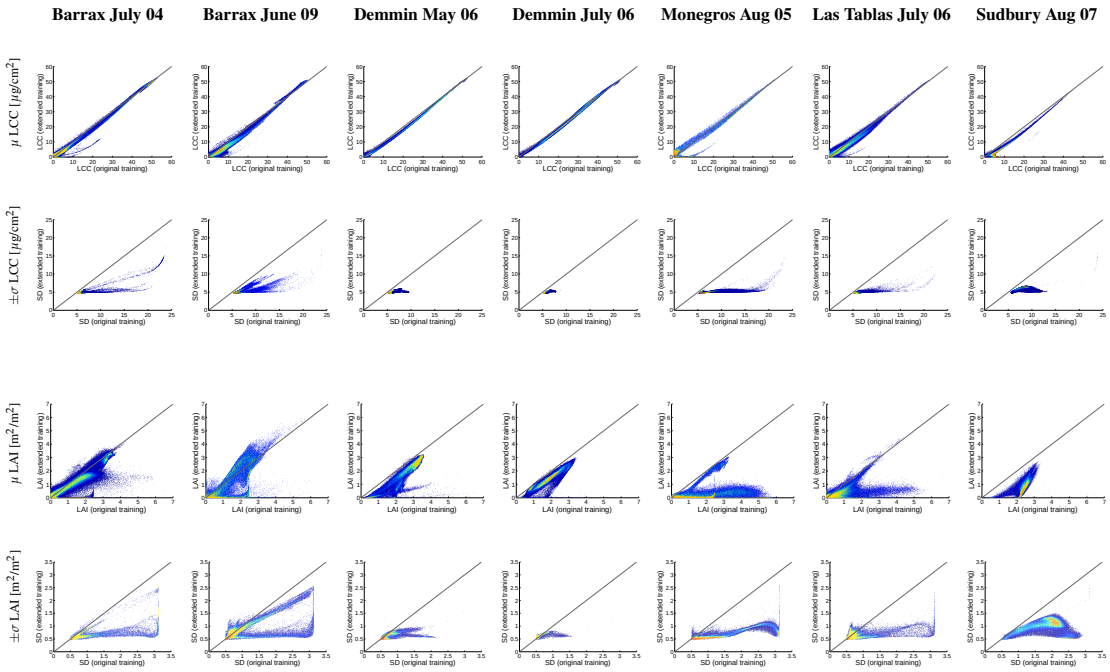


FIGURE 6.4: Scatterplots of maps generated by original training datasets ($TrOr$) vs. extended training dataset ($TrEx$) for μ LCC [top], $\pm\sigma$ LCC [below], μ LCC [below] and $\pm\sigma$ LAI [bottom].

6.6.3 Evaluating portability GPR models to other experimental S2 images by inspecting scatterplots

Even though promising results have been obtained over the reference Barrax site, a key requirement for operational applications is to ensure that GPR models are portable to other regions with similar accuracy. Henceforth, the $TrOr$- and $TrEx$-developed GPR models were applied to simulated S2 images over various sites in Spain, Germany and Canada. Mean estimate and uncertainty $TrOr$ and $TrEx$ maps were generated and again correlated. Scatter plots are displayed in Figure 6.4. Some interesting observations can be made from these scatter plots. For all tested images, LCC mean estimates fell right on the 1:1-line. It confirms the earlier observed trend that the extended training dataset did not lead to erratic LCC predictions. But also the scatter plots of the uncertainty estimates confirms the earlier observed systematic trend; for all images the $TrEx$-developed models yielded a systematic decrease in uncertainty, e.g. LCC $\pm\sigma$ hardly reached above 10 μg/cm^2.

Also with respect to LAI estimates, we observed the same pattern encountered in the reference image, i.e. the majority of $TrEx$-processed pixels largely follows the 1:1-line. However, a considerable amount of pixels tended to deviate towards lower estimates. While this trend appeared only slightly over Barrax and Demmin, the down-estimating occurred strongly over the scarcely vegetated areas of Los Monegros and Las Tablas. Also over Sudbury (Canada), LAI values were underestimated despite being dominated by vegetated surfaces. But most important is that, for all images, a significant and systematic decrease in uncertainties took place.

Particularly those areas that earlier suffered from large uncertainties were now predicted with more certainty.

Overview statistics (mean and standard deviation) for prediction and uncertainty maps are provided in Table 6.4. With regard to portability evaluation, special attention goes to the mean uncertainty statistics along the different images. For both GPR models and LCC and LAI maps uncertainty statistics turned to be almost as good as the reference image. It can thus be concluded that the locally trained GPR model is generally applicable to other sites, thereby reaching uncertainties on the same order of the Barrax 2003 reference site. Only the Sudbury data was processed with considerably less certainty. Moreover, for all images, mean uncertainties have been considerably reduced when comparing $TrEx$ with $TrOr$. Summarizing, the portability of the extended training dataset ($TrEx$) model has been successfully evaluated as mean uncertainties stabilized to about the same level as the reference Barrax image.

6.6.4 S2-10m biophysical parameter mapping

In this section we pay attention to the $TrEx$-generated maps, see Figure 6.5. A first observation across the different sites is that LCC and LAI estimations fall within expected ranges. Variations in land cover are clearly observable and non-vegetated surfaces can be easily distinguished from vegetated surfaces. A second observation is that for most of the sites uncertainty maps show rather low values in general. Nevertheless, noticeable within-image variations are still occurring, particularly for LAI over the Barrax 2009 image. Only for Sudbury (Canada) systematically greater uncertainties appeared. Here a suboptimal atmospheric correction may explain the poorer performance.

While the uncertainty maps provide some information about the robustness of the retrievals, one has nevertheless to be careful with its interpretation. Note that $\pm\sigma$ represents the uncertainty interval around the mean predictions, meaning that they need to be interpreted in relation to the estimates. For instance, an LCC uncertainty interval of about 5 would be more problematic for a mean estimate of 5 μg/cm^2 than of 50 μg/cm^2. Therefore, to evaluate the robustness of the GPR models it requires calculating the coefficient of variation:

$$CV = \frac{\sigma}{\mu} * 100. \tag{6.6}$$

CV maps provide relative uncertainties and are displayed in Figure 6.6. These maps can then be evaluated against an uncertainty threshold, e.g. as proposed by GCOS, i.e. 20% [GCOS, 2011]. Considering the reference Barrax image, dark areas represent retrievals with high uncertainties. These are typically bare soil areas which are characterized by low estimates (close to zero) and a relative high $\pm\sigma$. Conversely, it can be observed that most reliable retrievals occurred on the irrigated agricultural sites. However, only 9.0% of the pixels fell below the GCOS's 20% threshold. This low number seems surprising because the map was earlier excellently validated. It underlines the limited meaning of a (sparse) validation dataset when inspecting a heterogeneous map as a whole.

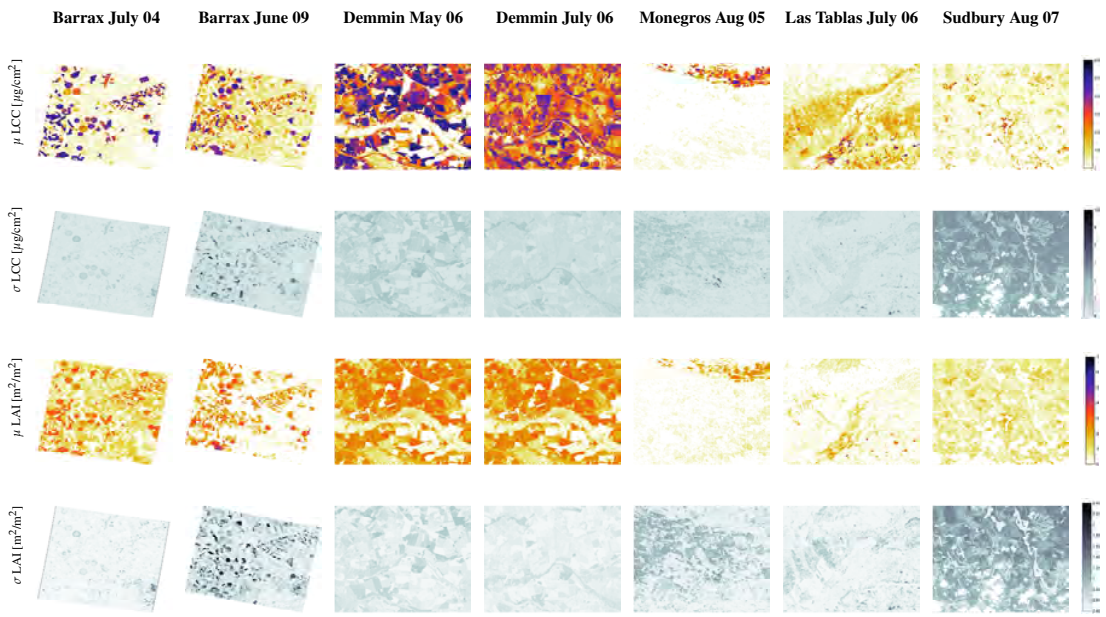


FIGURE 6.5: GPR mean estimation (μ) and uncertainty ($\pm\sigma$) maps for LCC and LAI, for the configurations of S2-10m, S2-20m, S2-60m and S3-300m.

When comparing the CV maps across the different sites it can be observed that particularly over the agricultural areas meaningful estimates were obtained. For instance, the July 2006 Demmin LCC map is processed with low uncertainty over the whole image; 53.5% of pixels fell below the 20% threshold, and 75.6% below the 30% threshold. Also here agricultural parcels were processed with low uncertainty. Only non-vegetated surfaces such as water bodies are flagged with a high relative uncertainty. Spurious relative uncertainties are also observed across the Los Monegros and Las Tablas sites where bare soil dominate. With $TrEx$ it was attempted to account for these bare soils, but due to near-zero mean estimates (0.0001) and, although $\pm\sigma$ is reduced, they still fall above near-zero threshold. CV turned therefore above 100%. Note that these relative uncertainty maps suggest that, at 10 m resolution, there is a greater problem of portability within an image than to other images. Solely the Sudbury maps show systematically higher relative uncertainties. This is probably due to the more difficult atmospheric correction as a consequence of cloud cover and lower sunlight intensity. The Sudbury case underlines the importance of an accurate and consistent atmospheric correction processing chain, which is expected to be provided by ESA (Level 2A Product). From a practical perspective, the CV map serves as a useful quality layer that allows masking out biophysical parameters estimates within an acceptable uncertainty (e.g. CV <20%), and at the same time can mask out non-vegetated surfaces (e.g. CV >100%).

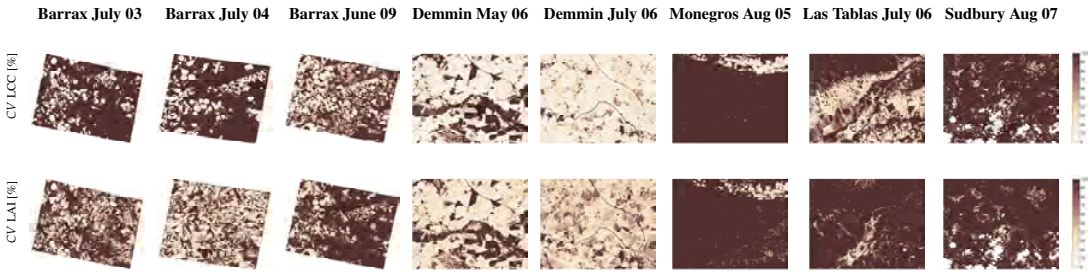


FIGURE 6.6: GPR coefficient of variation (CV) maps for LCC and LAI, for S2-10m configuration.

6.7 Discussion

The forthcoming S2 mission opens opportunities to implement novel retrieval algorithms in operational processing chains. The interest is put on retrieval algorithms that are accurate, fast, robust, and sufficiently flexible to make fully use of the new S2 MSI bands. Machine learning regression algorithms (MLRAs) are able to cope with most of these objectives. In an earlier work, GPR was evaluated as a very promising regressor in terms of processing speed and accuracy when using a local training dataset [Verrelst *et al.*, 2012a]. At the same time, GPR is transparent in terms of model development [Rasmussen and Williams, 2006b]; it may provide a ranking of features (bands) and samples (spectra), thus alleviating the black-box problem. A discussion on its performance in comparison to alternative state-of-the-art retrieval approaches presented literature was provided in Verrelst *et al.* [2012b,a]. In short, in a local setting GPR reached accuracies with S2 MSI band settings comparable (LAI) or superior (LCC) to competitive approaches. Nevertheless, the portability of this statistical approach to other sites remained questionable.

The here presented extended experimental training dataset ($TrEx$) not only further improved performances but also allowed a decrease in theoretical uncertainties. This proof of concept underlines the importance of a broad and diverse training dataset. More importantly, the GPR models were successfully applied to simulated S2 images covering various sites; associated uncertainties were on the same order as those generated by the reference image. The S2-10m examples demonstrated that excellent retrievals can be achieved already with 4 bands at a high spatial resolution of 10 m. Specifically, over fully vegetated surfaces relative uncertainties fell below the 20% requirements proposed by GCOS. This is encouraging for processing data from broadband sensors with a limited number of bands such as SPOT and Landsat or high resolution image such as Ikonos and Quickbird. On the other hand, S2 MSI encompasses additional bands at a coarser spatial resolution of 20 and 60 m. Note hereby that MSI's SWIR bands B11 and B12 have not been considered in this study because of falling outside the CHRIS' spectral range. These SWIR bands are known to be sensitive to vegetation structure [Brown *et al.*, 2000] and can better distinguish between dried-out fallow and non-vegetated lands. It is expected that inclusion of the SWIR bands will further improve the retrieval quality and reduce uncertainties. Moreover, further improvements can be achieved, which may lead to further reduction of uncertainties over non-vegetated surfaces. For instance, one could develop a set of GPRs

each working with different portions of the data or features. The uncertainty intervals could be optimally combined to improve the accuracy and reduce prediction bias. Alternatively, more sophisticated kernel functions that exploit signal-to-noise relations could be eventually considered [Gómez-Chova and Camps-Valls, 2012; Lazaro-Gredilla *et al.*, 2013].

At the same time, it should be noted that in operational biophysical parameter retrievals, pixels over non-vegetated targets are flagged as non-valid and are not considered in uncertainty calculation [see Fang *et al.*, 2012]. Thereby, at coarser pixel's size of MODIS and SPOT/VEGETATION, non-vegetated areas are well-defined by land cover maps, such as desert, inland water body, urban surfaces. In contrast, at 10 m resolution non-vegetated pixels can virtually appear anywhere. Since those kinds of pixels can be easily identified (e.g., mean estimates of ≈ 0, or CV >100%), they could actually just be discarded when interpreting uncertainty estimates. Another remark to bear in mind is that the derived theoretical uncertainties are directly related to what has been presented during the training phase. Theoretical uncertainties are thus not intended to replace the true physical uncertainties of the biophysical parameter products but instead to provide additional complementary information. Physical uncertainties are mandatory to be provided and should be obtained through comprehensive validation datasets collected on various sites, such as that coordinated by the Committee on Earth Observation Satellites (CEOS) Land Product Validation (LPV) community [Morisette *et al.*, 2006].

Finally, it does not escape our attention that only a limited set of images acquired during May-August have been evaluated. In an attempt to make the training data more representative at temporal and global scales, it should be able to cope with the majority of global land cover types along latitudinal gradients and over time. Ideally, ground truth data (biophysical parameters plus associated radiometric data) should be collected over a broad variety of terrestrial surfaces and vegetation types at multiple phenological stages. This, however, is a tedious and expensive job. Alternatively, an urging open question in this respect is to evaluate how robust GPR performs when trained by artificial spectra, e.g. as generated by a radiative transfer model (RTM). The advantage of RTMs is that a broad range of land cover situations can be simulated (e.g. up to hundred thousands), leading to a dataset several times bigger than what can be collected during a field campaign. Operational processing chains typically rely on such a hybrid approach [e.g. Bacour *et al.*, 2006; Baret *et al.*, 2007]. Initial efforts in this direction have been undertaken by implementing a MRLA module into the ARTMO (Automated Radiative Transfer Models Operator) toolbox, which is a suite of RTMs into one GUI toolbox [Verrelst *et al.*, 012c]. As such, GPR and other MLRA models can be automatically developed through lookup tables of simulated spectra and corresponding input parameters (e.g. LCC, LAI). Nevertheless, it should hereby be noted that, while GPR works successful for small training datasets (e.g. <2000 samples), its heavy computational load impedes the use of very large datasets. This limitation has to be resolved when aiming to develop generic models for global applications. For instance, alternative (sparse) versions of GPR have been proposed that can handle large scale datasets, see e.g. http://www.gaussianprocess.org/. Greedy algorithms in active learning settings can be also an alternative: GPRs trained on different data portions can provide a ranked list of the most informative spectra which could be used then to generate a final model. Forthcoming research will move in these directions, ultimately to reach a robust and generic

retrieval processing scheme.

6.8 Conclusions

The delivery of uncertainty information is a prerequisite for the operational use of remote sensing products. Currently only few retrieval approaches provide such estimates. Gaussian Processes Regression (GPR), a machine learning regression algorithm (MLRA) based on Bayesian non-parametric modeling, is one of them. Such retrieval algorithm may be of interest in view of ESA's forthcoming Sentinel-2 (S2) mission. In this work these uncertainty estimates were used to evaluate the robustness and portability of locally-trained GPR models. Hyperspectral CHRIS data was used for the simulation of experimental S2 images, being: S2-10m (4 bands), S2-20m (8 bands) and S2-60m (10 bands). While providing accurate estimates when trained with a local dataset called SPARC ($TrOr$; Barrax, Spain), it was also observed that non-vegetated pixels were processed with great uncertainty. In an attempt to make the LCC and LAI GPR models more robust, an extended training dataset was introduced ($TrEx$), i.e. the original SPARC dataset plus 60 spectra over all kinds of non-vegetated surfaces (e.g., soils, man-made surfaces, water bodies). For the Barrax 2003 reference image, $TrEx$-developed GPR model delivered validation errors below 10% (NRMSE: 3.5-9.2%; r^2: 0.95-0.99), robust LCC and improved LAI estimates, and above all reduced associated uncertainty estimates. These locally-trained GPR models were subsequently applied to other experimental S2 images over various sites across Spain, Germany, Canada. The uncertainty estimates provided insight in the success of the models' performance. Overall, GPR models are portable to other images and uncertainty estimates can thereby function as quality layer to filter out unreliable retrievals. Uncertainty intervals were on the same order as the Barrax 2003 reference image and relative uncertainties over vegetated surfaces were below the 20% requirements proposed by GCOS. However, typically large uncertainty variation within an image was observed due to surface heterogeneity. GPR is concluded as a powerful regressor for remote sensing applications; not only it delivers accurate predictions, it is currently the only MLRA that provides associated uncertainty intervals.

7 Synthesis

Contents

7.1 Main results

The estimation of vegetation properties from optical remote sensing data is a broad field, where a lot of progress has been made in the last three decades. This thesis brought together a wide range of retrieval methods dedicated to the quantification of biophysical variables. In all generality, retrieval methods can be categorized into three main domains: (1) *parametric regression*, (2) *non-parametric regression*, and (3) *physically-based model inversion*. Each of these three domains have been largely automated and fully analyzed in Chapters 3 - 5, eventually leading for each of them to fully operational retrieval toolboxes. In particular, the retrieval toolboxes enable to analyze, compare and apply a variety of state-of-the-art methods in order to reach optimized and automated mapping strategies. These toolboxes have been developed within the scientific ARTMO software package environment.

The principles of vegetation properties mapping from remote sensing data as well the philosophy of the ARTMO toolbox have been explained in Chapter 2. Essentially ARTMO automates the running of RTMs in forward mode, and process them in subsequent retrieval toolboxes. At the same time, these retrieval toolboxes are not restricted to simulated data, but also external data obtained from field campaigns can be processed. These toolboxes enable vegetation properties mapping from optical sensors. To facilitate its usability, ARTMO has been equipped with a list of predefined sensors; established ones, such as Landsat, SPOT, MERIS, MODIS, CHRIS as well forthcoming superspectral and hyperspectral sensors, such as Sentinel-2, -3, EnMAP, FLEX. Emphasis in this thesis was put on the processing of Sentinel-2 data, which is especially designed for land applications, but at the same time mapping strategies have been developed and analyzed for spaceborne CHRIS or airborne HyMAP hyperspectral data. The principles of ARTMO's retrieval toolboxes and relevant findings using these toolboxes are summarized in following sections.

In this section, the main research questions of this thesis, which are elaborated in the core Chapters 3 – 6, are revised and followed by a discussion of the main results.

7.1.1 Research question A: The performance of spectral indices to LAI and LCC mapping

1. *What is the performance of all possible two-band vegetation indices to enable optimized LCC and LAI estimation?*

In Chapter 3, ARTMO's Spectral Indices (SI) assessment toolbox is presented [Rivera *et al.*, 2014a]. This toolbox enables the analysis and assessment of the accuracy of a large number of SI models. Basically, the toolbox offers a systematic but still empirical approach for the assessment of all possible 2, 3 or up to 10-band SI formulations. Datasets can be partitioned into calibration and validation subsets. These datasets may originate from simulations or from field campaigns. Several options have been included in the SI assessment approach, amongst which: (1) The addition of noise and the possibility to select fitting functions (e.g., linear, exponential, power or polynomial functions); (2) the SI toolbox virtually allows for any type of spectral index model to be formulated and evaluated using up to ten spectral bands; and (3) the possibility to assess and apply SIs per land cover class. To illustrate its functioning and predictive power, all two-band combinations according to simple ratio (SR) and normalized difference (ND) formulations as well as various fitting functions (linear, exponential, power, logarithmic, polynomial) have been assessed. HyMap imaging spectrometer [430-2490 nm] data obtained during the SPARC campaign in Barrax, Spain, have been used to extract leaf area index (LAI) and leaf chlorophyll content (LCC) estimates. For both SR and ND formulations, the most sensitive regions have been identified for two-band combinations of green (539-570 nm) with longwave SWIR (2421-2453 nm) for LAI (r^2: 0.83) and far-red (692 nm) with NIR (1340 nm) or shortwave SWIR (1661-1686 nm) for LCC (r^2: 0.93). The identification of the SWIR rather than the conventional NIR as an important spectral region emphasizes the importance of a systematical assessment for accurate vegetation properties mapping.

7.1.2 Research question B: The performance of LUT-based inversion of PROSAIL to LAI and LCC mapping

2. *What is the performance of physically-based inversion routines given a suite of cost functions and regularization options to enable optimized LCC and LAI estimation?*

In Chapter 4, ARTMO's LUT-based inversion toolbox is presented [Rivera *et al.*, 2013a]. Lookup-table (LUT)-based radiative transfer model inversion is considered a physically-sound and robust method to retrieve biophysical parameters from Earth observation data but regularization strategies are needed to mitigate the drawback of ill-posedness. While various LUT-based inversion methods have been proposed in literature, they mostly rely on RMSE as cost function. However RMSE can result in large losses of efficiency when the error distribution is non-Gaussian or non-symmetric. The LUT-based inversion toolbox enables the analysis and assessment of the accuracy of a wide range of alternative cost functions. Apart from over 60 cost functions, several more regularization options have been included, such as the role of (1) added noise; (2) role of data normalization, and (3) mean of multiple solutions in inversion function. Three families of cost functions were compared: information measures, M-estimates and minimum contrast methods. We have selected only cost functions without additional parameters to be tuned, and thus they can be immediately implemented in processing chains. The coupled leaf/canopy model PROSAIL was inverted against simulated Sentinel-2 imagery at 20 m spatial resolution (8 bands) and validated against field data from the ESA-led SPARC (Barrax, Spain) campaign. It was found that introducing noise and opting for the mean of multiple best solutions in the inversion considerably improved retrievals; relative errors can be halved as opposed to without these regularization options. Best LCC retrievals were obtained using a normalized 'L_1-estimate' function with a relative error of 17.6% (r^2: 0.73), while best LAI retrievals were obtained through non-normalized 'least-squares estimator' (LSE) with a relative error of 15.3% (r^2: 0.74). Summarizing, the inversion performance is directly related to the used cost function due to different assumptions on the nature and properties of errors. Regularization strategies play thereby a great role in optimization. It is therefore recommended to evaluate different inversion strategies prior to applying an inversion strategy to the whole image.

7.1.3 Research question C: The performance of nonparametric regression algorithms to LAI and LCC mapping

3. *What is the performance of a range of nonparametric regression methods and regularization options to enable optimized LCC and LAI estimation?*

In Chapter 5, ARTMO's Machine Learning Regression Algorithms (MLRA) toolbox is presented [Rivera *et al.*, 2013b]. The MLRA toolbox enables analyzing the predictive power of various machine learning regression algorithms in a semiautomatic and systematic manner and applying a selected MLRA to multispectral or hyperspectral imagery for mapping applications. It contains both linear and nonlinear state-of-the-art regression algorithms, such as linear

feature extraction via principal component regression (PCR) and partial least squares regression (PLSR), decision trees (DT), neural networks (NN), kernel ridge regression (KRR) and Gaussian processes regression (GPR). Similar to the aforementioned toolboxes, various regularization options have been implemented into the MLRA toolbox, e.g., training/validation data splitting, adding noise, and regression models can be developed and evaluated per land cover class. Training and validation data can either come from field campaigns or from simulations as generated by radiative transfer models. The performance of multiple implemented regression strategies has been evaluated for LAI and LCC retrieval against the SPARC dataset (Barrax, Spain) and simulated Sentinel-2 (8 bands), CHRIS (62 bands) and HyMap (125 bands) observations. In general, nonlinear regression algorithms (NN, KRR, GPR) outperformed linear techniques (PCR, PLSR, DT) in terms of accuracy, bias, and robustness. Most robust results for LAI and LCC retrieval along gradients of training/validation partitioning and noise variance were obtained by KRR, while GPR delivered most accurate estimations (r^2: 0.94-0.99 and NRMSE: 7.0-3.0 %). Apart from excellent performance, GPR provides additional uncertainty estimates on a pixelwise basis. A GPR model was applied to a hyperspectral HyMap flightline to map LAI and LCC.

7.1.4 Research question D: The utility of GPR uncertainty estimates in broader LAI and LCC mapping

4. *Can the best evaluated retrieval method be applied to other sites and conditions in view of operational retrieval of LCC and LAI from forthcoming Sentinel-2 data?*

In Chapter 6, a mapping application is presented that moves away from the local Barrax test site [Verrelst *et al.*, 2013b]. Because of their ability to perform adaptive, nonlinear data fittings, statistical learning regression algorithms are perceived as powerful candidates for the estimation of biophysical parameters from satellite reflectance measurements. However, these methods have often been criticized for their lack of generality and portability. In this work, we analyzed the applicability of Gaussian process regression (GPR) for retrieving LAI and LCC from simulated Sentinel-2 (S2) images over landscapes varying in space and time. GPR was chosen because it is not only a powerful regression function, but also provides uncertainty intervals along with the mean estimates. We used those uncertainties to analyze the portability of locally trained GPR models. Experimental data came from the ESA-led field campaign SPARC (Barrax, Spain). Local evaluation of an extended training dataset with more variation over bare soil sites led to improved LCC and LAI mapping with reduced uncertainties. GPR reached the 10% precision required by end users, with for LCC a NRMSE of 3.5-9.2% (r^2: 0.95-0.99) and for LAI a NRMSE of 6.5-7.3% (r^2: 0.95-0.96). The developed GPR models were subsequently applied to simulated Sentinel images over various sites across the world. The associated uncertainty maps were compared relative to the local Barrax test site. The generally low uncertainty intervals over vegetated surfaces suggest that the locally trained GPR models are portable to other sites and conditions. GPR is concluded as a powerful regression algorithm

for remote sensing applications; not only it delivers accurate predictions; it is currently the only MLRA that provides valuable associated uncertainty intervals.

7.2 General conclusions

This Thesis brought together and analyzed a wide variety of retrieval methods within the scientific ARTMO toolbox. The implemented methods have been evaluated in view of optimized LAI and LCC mapping. The main conclusions drawn from this work are:

1. The ARTMO toolbox has been developed during this thesis. This toolbox consists of a suite of RTMS and retrieval toolboxes in a Matlab GUI environment and with a MySQL database running underneath, and is freely available to the broader remote sensing community (http://ipl.uv.es/artmo/). ARTMO has been modularly designed; RTMs or toolboxes can be seamlessly added or removed. Essentially, it automates the running of RTMs and assists in the development of retrieval strategies. At the same time, its modular design facilitates developers to implement new tools or toolboxes based on available templates. ARTMO can underpin and strengthen advances in remote sensing of vegetation.

2. The calculation of vegetation indices is a widely-known, simple and fast way to develop parametric regression models for biophysical variable mapping. However, no evidence was found that the well established indices (e.g. NDVI, PRI) are best performing. In fact, by assessing the predictive power of all possible two-band combinations more sensitive spectral regions were found with one band in the SWIR region. Moreover, there is no reason to believe that a two-band index is best performing. Initial results indicate that performances can still improve significantly when a well-selected combination of multiple bands is entered into an index formulation.

3. Inversion of a RTM is generally considered as a physically-sound approach to realize retrievals. However, the approach is not straightforward. Model assumptions, boundary conditions and applied inversion strategy are all factors impacting the retrieval performance. Typically a RMSE is applied as cost function, and the same inversion strategy is applied to retrieve multiple variables. It was found that evaluating alternative cost functions and regularization options can halve relative errors as compared to a non-optimized, standard inversion strategy. It was also found that optimized inversion strategies differ per variable due to the nature of its data; hence retrieval of multiple variables is suboptimal and not recommended.

4. For the last few years most progress regarding mapping algorithms has been made in the field of machine learning regression approaches (MLRAs). The majority of these advanced nonparametric regression algorithms have hardly made it into the field of vegetation mapping. Only partial least square regression (PLSR) and neural networks (NN) received broad attention in many mapping applications. However, we found that PLSR

performs inferior and NN performs less robust than the non-linear, kernel-based MLRAs. Moreover, the family of kernel-based MLRAs (KRR, GPR) not only proved to deliver accurate estimates, but also more robust to varying noise and training/validation splitting.

5. Regarding novel kernel-based MLRAs, the most promising one is probably GPR. Apart from robust retrievals, this Bayesian regression algorithm also provides insight in relevant bands during model development and delivers associated uncertainty estimates. These uncertainties proved to be particularly valuable when transporting the developed model to other sites and conditions. We have demonstrated that for the majority of processed images uncertainties were on the same order as on the local image (Barrax, Spain). Thus, associated uncertainty estimates as delivered by a GPR model can serve as a convenient quality check when processing images in space and time.

7.3 Reflection

In Chapters 3 – 5, three different toolboxes have been presented that assess and apply the main retrieval domains: i.e. parametric regression, LUT-based inversion strategies and non-parametric regression. State-of-the-art methods of each retrieval domain have been implemented into their respective toolboxes, and in this thesis these methods have been systematically analyzed. This extensive assessment across the three domains allows, for the first time, synthesizing the field of biophysical parameter retrieval. Such comparison is possible, because for all chapters the dataset from the same campaign (SPARC, Barrax, Spain) was used and the same variables were retrieved (LAI and LCC). Before entering into conclusions, the following points should be remarked:

- For the spectral indices analysis airborne HyMap imaging spectrometer (430-2490 nm) data was used, while for the inversion analysis simulated Sentinel-2 (20 m; 8 bands) data was used. However, for the MLRA analysis both simulated Sentinel-2 (20 m; 8 bands), CHRIS and airborne HyMap data were analyzed.
- For the spectral indices analysis and for the inversion analysis all field data (the SPARC dataset) was used for either calibration (spectral indices) or validation (inversion). However, for the MLRA analysis, the SPARC dataset was split into training-validation subsets.

It is recognized that these inconsistencies affect accuracies, e.g. more bands, a broader spectral range, better SNR, airborne scale (HyMap) may favor retrieval performances. Nevertheless, some general observations can be made:

- Spectral indices: best two-band vegetation indices applied to HyMAP data (430-2490 nm) led to a r^2 prediction efficiency of 0.93 for LCC (NRMSE: 7.75%) and r^2 of 0.83 for LAI (NRMSE: 12.51%). The SI models are designed by the user thus completely transparent. Computational cost of development as well image processing is minimal.

The downside of these rather simple methods is that they suffer from poor generality and portability.

- LUT-based inversion: best LCC retrievals for Sentinel-2 (20 m; 8 bands) data were obtained using a normalized 'L_1-estimate' function with a r^2 prediction efficiency of 0.73 (NRMSE: 17.6%), while best LAI retrievals were obtained through non-normalized 'least-squares estimator' (LSE) with a r^2 of 0.74 (NRMSE:15.3%). In principle the retrieval procedure is transparent since RTMs are based on physical laws, and minimization cost functions. However, the many model assumptions and parametrization steps makes that transparency somewhat obscured, and expert knowledge is required. An advantage of this method is that insight in retrieval uncertainty is obtained, e.g. through inversion residuals. Because the inversion problem is resolved against a large LUT table on a pixel-by-pixel basis, its computational burden is a serious drawback. Another important point to be remarked is that inversion of a RTM essentially means calculating back to its input variables. Hence, this inherently implies that only so-called 'state-variables', i.e. RTM input variables that have strong impact on reflectance through physical laws, can be retrieved.

- Machine learning regression algorithms (MLRA): KRR and GPR delivered most accurate estimates with a r^2 prediction efficiency between 0.94-0.99 and NRMSE down to 7.0-3.0% for both LCC and LAI and regardless of the different analyzed sensor configurations. These methods, although very powerful, are mainly criticized for the lack of model transparency, i.e. so-called 'black box' models. Nevertheless, some of latest MLRAs include features that give some hints in model development. For instance, GPR provides information of band relevance during model development.

Comparing best obtained results, the MLRA approaches led to highest accuracies, while LUT-based inversion routines performed on the whole poorest. The spectral indices approaches led to intermediate results. The relatively poorer performance of the LUT-based inversion can be largely explained by the fact that this method relies on simulated data. While inverting a radiative transfer model makes this approach generally applicable, simulated data does not always properly represent image observations (e.g. due to simplicity of the model, poor model parametrization, imperfect atmospheric correction, etc.).

Table 7.1 summarizes the most important features of the three retrieval approaches.

TABLE 7.1: Synthesis key features of the three main retrieval approaches. Best features are bolded.

	Parametric regression: Vegetation Indices	**LUT-based inversion**	**Non-parametric regression: MLRAs**
Surface variables	**Any**	Only state variables	**Any**
Model performance	Good to excellent	good	**Excellent**
Insight in model	**Yes**	partly	partly
Mapping speed	**Instantly**	slow	**fast**
Uncertainties	no	**Yes (e.g. residuals)**	**Yes**

Altogether, taken both performance and processing speed into account, MLRA approaches seem to be most promising for new generation of mapping methods. Moreover, some MLRAs

provide additional uncertainties (e.g. GPR). Based on this synthesis the following recommendations can be drawn:

- When a broad field dataset of the target area is available that can be applied for training, statistical nonparametric mapping should be preferred. These methods are accurate, fast and for GPR also associated uncertainties are provided.
- Robust earlier generated GPR models (e.g. from SPARC dataset) can be applied to other sites with similar conditions as shown in GPR study. The uncertainties provide insight in quality of performance. As a guideline, the Global Climate Observing System (GCOS) proposed uncertainties below 20% are considered as reliable, but also below 30% should be still acceptable (see Chapter 6).
- If little field data is available that prevents developing a robust leading to non-robust models (e.g. images processed with uncertainties above 30% for the large majority of the map) then LUT-based inversion strategies may be an attractive alternative. However, here it should be kept in mind that expert knowledge is required and only state variables can be retrieved.

7.4 Outlook

7.4.1 Retrieval toolboxes

The retrieval toolboxes paved the path for various new mapping applications, pushing those fields into superior performances and new application opportunities. They have been made freely available to the broader remote sensing community. During the course of this work several suggestions have been identified to boost the field of vegetation properties mapping forward:

Spectral Indices toolbox:

- In Chapter 3, two-band vegetation indices according to SR and NDVI formulations have been optimized. However, the SI toolbox permits the development and assessment of new vegetation indices with all possible band combinations of up to 10 different bands. A follow-up study with spectral indices on the basis of simulated Sentinel-2 data for all possible four bands has been conducted [Verrelst *et al.*, 2013c], with optimized indices that pushed accuracies with a few points r^2 up. It is to be expected that new index formulations with more bands involved will further improve accuracies.
- Another path to be explored concerns the analysis of generic indices based on simulated reflectance data. In this thesis, optimized vegetation indices have been derived from the SPARC dataset. Given the absence of uncertainties, it remains however unknown how

general optimized models can be applied to other sites and conditions. For this reason it would be valuable to have indices analyzed on the basis of simulated reflectance data. ARTMO has all the tools operational to undertake this research, not only with RTMs both at leaf or canopy level, but also for different vegetation types, i.e. 1D (e.g. SAIL) or 3D forest RTMs (e.g. FLIGHT).

LUT-based inversion toolbox:

- In Chapter 4, only 18 stand-alone cost functions out of more than 60 functions have been analyzed. The majority of cost functions require one or more parameters to be tuned, which in principle should lead to refined performances compared to the stand-alone cost functions. The inversion toolbox offers the possibility to loop over those parameters in order to seek for optimum. These features have not been explored yet, leaving opportunities for further optimizing the inversion scheme.

- LUT-based inversion relies on a large spectral database that have been generated by RTMs. Until now only PROSAIL has been applied to generate the LUT because of relatively few input variables involved (about 8). However, ARTMO enables the development of spectral databases originating from more advanced RTMs such as FLIGHT, SLC, SCOPE. While those models require more parametrization, once having them configured, running the subsequent inversion process is just the same.

MLRA toolbox:

- In Chapter 5, only six available nonparametric regression algorithms have been analyzed (PCR, PLSR, DT, NN, KRR, GPR). Meanwhile, several more novel MLRAs have been implemented into the toolbox, with some of them never being applied in the context of vegetation properties mapping, being bagging trees, boosting trees, extreme machine learning, relevance vector machine, variational heteroscedastic GPR. Recent results indicate that particularly relevance vector machine, variational heteroscedastic GPR and extreme machine learning are among the top performing regression algorithms [Verrelst *et al.*, 2013c].

- Hybrid methods try to combine the generality of physically-based methods with the flexibility and computational efficiency of MLRAs. The idea is to learn the inverse mapping with a nonparametric model that is being trained using simulated data generated by RTMs. This approach has long been restricted to training a neural net by simulated data from a RTM (e.g. PROSAIL), which has been proven successful in an operational processing scenes (e.g. GEOLAND products). However, given the recent advances made in the field of MLRAs, it can be questioned whether a neural network is really the best choice. In Chapter 5 it was demonstrated that novel MLRAs (e.g. GPR) tend to outperform neural networks. A MLRA such as GPR, with the feature of providing uncertainties, holds promises to develop new generation of retrieval methods that can be implemented into operational processing chains.

7.4.2 Future work ARTMO: beyond the retrieval toolboxes

Coupling with atmospheric models

Although ARTMO is already at an advanced stage of development, currently the integration with an atmospheric model is missing. This means that only top-of-canopy (TOC) reflectance can be simulated. It also bears the consequences that an atmospheric correction has to be applied when aiming to invert a RTM against remote sensing data. Alternatively, when having an atmospheric model implemented into ARTMO then canopy simulations can be upscaled to top-of-atmosphere (TOA) radiance. This holds the advantage that it would allow us to develop retrieval strategies of atmospheric and biophysical variables directly from the satellite observations, i.e. without the need to correct for atmospheric effects. Potential atmospheric models to be implemented into ARTMO of a low to high complexity would be 6S and MODTRAN. Efforts in coupling a MODTRAN-generated LUT with an ARTMO-generated LUT are already underway, e.g. as part of a scene generator module.

Scene generator module (SGM)

In support of ESA's Earth Explorer 8 candidate mission FLEX (FLuorescence EXplorer), a scene generator module (SGM) is currently under development within the ARTMO environment [Rivera *et al.*, 2014b]. Essentially, the SGM is in charge of simulating the scene to be observed by an instrument. A scene is understood as a TOA radiance map, e.g. as observed by a satellite mission. The generation of these synthetic scenes includes the distribution of biogeophysical and atmospheric parameters over the scene map. In addition, the SGM takes also into account environmental conditions such as surface topography and observation/illumination geometry. All these parameters and environmental conditions serve as input for the generation of TOA radiance maps through the use of canopy RTMs (as those provided by ARTMO) coupled with an atmospheric model (e.g. MODTRAN) or from external radiometric data (e.g. reflectance, fluorescence and/or TOA radiance spectral databases or external image files). The SGM will ultimately become part of a complete End-to-End FLEX mission performance simulator software, that is being developed in collaboration with partners and industry. At the same time, a simplified but more generic version of the SGM (i.e. applicable for any optical sensor) is foreseen to be implemented as an ARTMO tool.

Global sensitivity analysis (GSA) tool

Another ongoing study within the ARTMO environment is the development of a global sensitivity analysis GSA tool. Sensitivity analysis evaluates the relative importance of each input parameter and can be used to identify the most (and least) influential variables in determining the variability of model outputs. In contrast to a local sensitivity analysis that evaluates one factor at a time, a GSA explores the full input parameter space, i.e., all input parameters are tested together. In support of FLEX an advanced RTM that also provides fluorescence outputs,

i.e. SCOPE, has recently been analyzed using GSA [Verrelst *et al.*, 2014a]. In this way the driving variables determining variations in reflectance and fluorescence outputs have been derived. Apart from these FLEX-related activities, it is foreseen that in the near future a fully operational GSA toolbox will be implemented into ARTMO that is able analyzing all available RTMs, as well enabling sensitivity analysis of coupled RTMs, e.g. leading to the unraveling the radiative transfer of a complete soil-leaf-canopy-atmosphere system.

Remaining forthcoming activities and final considerations

Finally, apart from the above-mentioned ongoing activities, the modular design of ARTMO opens opportunities to implement new RTMs, tools and toolboxes as plug-ins into the existing framework. To stay within the context of vegetation properties mapping – but in principle applications can go in any direction – the following activities are planned: (1) a classifier toolbox, (2) time series toolbox, (3) unmixing toolbox, and (4) data assimilation toolbox.

Ultimately these activities will not only facilitate the use of advanced remote sensing techniques to a broader community, the developed tools and toolboxes will also boost progress towards: (1) an improved understanding in the interactions between light and vegetation properties, (2) improved retrieval algorithms serving forthcoming optical missions such as Sentinel-2 -3, EnMAP, FLEX, (3) new mapping applications to understand better our changing Earth.

For instance, currently no operational global LCC mapping algorithm exists, which prevents this variable to be considered as an essential climate variable (ECV). Currently only absorbed photosynthetically active radiation (FAPAR) is considered as an ECV [GCOS, 2014], and often used as a substitute of LCC because of closely related with LCC [Gitelson *et al.*, 2003] and carbon assimilation [Sellers, 1987; Sellers *et al.*, 1992]. However, FAPAR is a component of the land-surface radiation budget and also influenced by recording time and direct versus indirect/diffuse solar radiation. The robustness of FAPAR as an indicator of LCC may therefore be questioned and more robust retrieval schemes are desirable. An operational LCC retrieval algorithm can be developed within ARTMO, e.g. based on the one presented in Chapter 6. A precise knowledge of LCC and LAI is also indispensable in view of the FLEX mission in order to properly calculate and interpret solar-induced chlorophyll fluorescence and relate it to photosynthesis activity of vegetation [Verrelst *et al.*, 2013a]. Furthermore, ARTMO can serve improved or new vegetation products, which in turn can serve as input into assimilation procedures for higher level products such as plant traits, photosynthesis, plant diversity; e.g. by coupling with Dynamic Global Vegetation Models or General Circulation Models. Altogether, these activities will eventually open new avenues to improved local-to-global (agro)ecosystem processes monitoring.

7.5 Achievements and relevance

The results generated during this thesis have been presented on several conferences and published as research papers on different international journals. The following list summarizes the published journal and conference papers directly related to this Thesis:

International journal papers:

1. Van Wittenberghe, S., Verrelst, J., **Rivera, J.P.**, Alonso, J., F., Moreno, J., Samson, R., (2014). Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. *Journal of Photochemistry and Photobiology B: Biology*, (In press).

2. **Rivera, J.P.**, Verrelst, J., Delegido, J., Veroustraete, F., Moreno, J., (2014). On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. *Remote Sensing*, (Submitted).

3. Delegido, J., Van Wittenberghe, S., Verrelst, J., Ortiz, V., Veroustraete, F., Valcke, R., Samson, R., **Rivera, J.P.**, Tenjo, C., Moreno, J., (2014). Chlorophyll content mapping of urban vegetation in the city of Valencia based on the hyperspectral NAOC index. *Ecological Indicators*, 40, p. 34-42.

4. **Rivera, J.P.**, Verrelst, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G., (2014). Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Tocics in Applied Earth Observation and Remote Sensing*, (In press).

5. Verrelst, J., **Rivera, J.P.**, Leoneko, G., Alonso, L., Moreno, J. (2014). Optimizing LUT-based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 data: Role of Cost Functions. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), p. 257-269.

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

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

8. Delegido, J., Verrelst, J., Meza, C.M., **Rivera, J.P.**, Alonso L., Moreno J., (2013). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy*, 46, p. 45-52.

9. Verrelst, J., Alonso, L., **Rivera, J.P.**, Moreno, J. Camps-Valls, G., (2013). Guassian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE Journal of Selected Tocics in Applied Earth Observation and Remote Sensing*, 6(2), Part 3.

10. Verrelst, J., Muñoz-Marí, 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, p127-139.

Conference contributions:

1. **Rivera, J.P.**, Sabater, N., Tenjo, C., Vicent, J., Alonso, L., Moreno, J., (2014). Synthetic scene simulator for hyperspectral spaceborne passive optical sensors. Application to ESA's FLEX/Sentinel-3 tandem mission. In: 6th workshop on hyperspectral image and signal processing 2014. 24-27 June, Lausanne, Switzerland.

2. Verrelst, J., **Rivera, J.P.**, Van Der Tol, C., Magnani, F., Mohammed, G., Moreno, J., (2014). Global sensitivity analysis of the A-SCOPE model in support of future FLEX fluorescence retrievals. In: 6th workshop on hyperspectral image and signal processing 2014. 24-27 June, Lausanne, Switzerland.

3. Vicent, J., Acarreta, J.R., Franco, R., Mandazo, M., **Rivera, J.P.**, Sabater, N., Tenjo, C., Verrelst, J., Moreno, J., (2014). FLEX-E: ESA's FLEX End-to-End Mission Performance Simulator. Architecture design, current status and preliminary results. In: 5th Int. Workshop on Remote Sensing of Vegetation Fluorescence 2014, 22-24 April, Paris, France.

4. Sabater, N., **Rivera, J.P.**, Tenjo, C., Vicent, J., Alonso, L., Verrelst, J., Moreno, J., (2014). The scientific modules of FLEX End-to-End simulator: the Scene Generator Module and the Level-2 Retrieval Module. In: 5th Int. Workshop on Remote Sensing of Vegetation Fluorescence 2014, 22-24 April, Paris, France.

5. Verrelst, J., **Rivera, J.P.**, Van Der Tol, C., Magnani, F., Mohammed, G., Moreno, J., (2014). A-SCOPE: automating fluorescence modeling in support of FLEX. In: 5th Int. Workshop on Remote Sensing of Vegetation Fluorescence 2014, 22-24 April, Paris, France.

6. Vicente, J., Sabater, N., Alonso, L., Verrelst, J., **Rivera, J.P.**, Tenjo, C., Moreno, J. (2013). FLEX End-to-End Mission Performance Simulator Architecture Design using a Generic Reference Architecture for EO Missions. In: ESA Living Planet Symposium 2013, 09-13 September, Edinburgh, UK.

7. Verrelst, J., **Rivera, J.P.**, Sabater, N., Vicente, J., Tenjo, C., Alonso, L., Moreno, J. (2013). Assessing the Benefit of the Sentinel-3/FLEX Tandem Mission Concept for Retrieval of Fluorescence and Biophysical Parameters. In: ESA Living Planet Symposium 2013, 09-13 September, Edinburgh, UK.

8. Verrelst, J., **Rivera, J.P.**, Camps-Valls, G., Moreno, J. (2013). Recent advances in Biophysical Parameter Retrieval Methods - Opportunities for Sentinel-2. In: ESA Living Planet Symposium 2013, 09-13 September, Edinburgh, UK.

9. Moreno, J., Alonso, L., Delegido, J., **Rivera, J.P.**, Sabater, N., Tenjo, C., Verrelst., J., Vicent, J., (2013). Misión FLEX (Fluorescence Explorer): observación de la fluorescencia por teledetección como nueva técnica de estudio del estado de la vegetación terrestre a escala global. In: XV Congreso de la Asociación Española de Teledetección (AET) 2013, 22-24 October, Madrid, Spain.

10. **Rivera, J.P.**, Verrelst, J., Leonenko, G., Moreno, J. (2013). Evaluación de múltiples funciones de mérito en la inversión de parámetros biofísicos a través del modelo PROSAIL usando tablas de búsqueda. In: XV Congreso de la Asociación Española de Teledetección (AET) 2013, 22-24 October, Madrid, Spain.

11. **Rivera, J.P.**, Verrelst, J., Delegido, J., Moreno, J. (2013). Herramienta informática para el diseño y evaluación de índices espectrales genéricos para la inversioń de parámetros biofísicos. In: XV Congreso de la Asociación Española de Teledetección (AET) 2013, 22-24 October, Madrid, Spain.

12. Verrelst, J., **Rivera, J.P.**, Muñoz, J., Moreno, J. Camps-Valls, G. (2013). ARTMO's new Machine Learning Regression Algorithm (MLRA) module for semiautomatic mapping of biophysical parameters. In: EARSeL 8th SIG-Imaging Spectroscopy Workshop 2013, 08-10 April, Nantes, France.

13. Verrelst, J., **Rivera, J.P.**, Guadalarjara, A., Delegido, J., Moreno, J. (2013). ARTMO's new Spectral Indices (SI) module to rapidly evaluate a multitude of SIs for mapping of biophysical parameters. In: EARSeL 8th SIG-Imaging Spectroscopy Workshop 2013, 08-10 April, Nantes, France.

14. Verrelst, J., **Rivera, J.P.**, Guanter, L., Alonso, L., Moreno, J. (2012). Evaluating machine learning regression algorithms for operation retrieval of biophysical parameters: Opportunities for Sentinel-3. In: Sentinel-3 Symposium 2012, 15-19 October, Frascati, Italy.

15. Verrelst, J., **Rivera, J.P.**, Alonso, L., Linstrot, R., Moreno, J. (2012). Synergy of Sentinel-3 OLCI and FLEX-FLORIS leads to improved retrieval of biophysical parameters: a model study. In: Sentinel-3 symposium, 15-19 October, Frascati, Italy.

16. Verrelst, J., **Rivera, J.P.**, Alonso, L., Linstrot, R., Moreno, J. (2012). Potential retrieval of biophysical parameters from FLORIS, S3-OLCI and its synergy. In: IGARSS 2012, 22-27 July, Munich Germany.

17. Verrelst, J., **Rivera, J.P.**, Leonenko, G., Alonso, L., Moreno, J. (2012). Optimizing LUT-based radiative transfer model inversion for retrieval of biophyiscal parameters using hyperspectral data. In: IGARSS 2012, 22-27 July, Munich Germany.

18. Delegido, J., Van Wittenberghe, S., Verrelst, J., Alonso, L., **Rivera, J.P.**, Sabater, J., Vicent, J.,Ortiz, V., Moreno, J. (2012). Chlorophyll fluorescence: the signal hidden behind leaf reflectance. A new technique for the study of plant status by remote sensing. IN SELPER 2012, 19-23 November, Cayenne, French Guyana. In: XV symposium of SELPER.

19. Delegido, J., Ortiz, V., Verrelst, J., Van Wittenberghe, S., **Rivera, J.P.**, Moreno, J. (2012). SPAD calibration in trees of Valencia city (Spain) and chlorophyll map of urban vegetation from hyperspectral index NAOC. IN SELPER 2012, 19-23 November, Cayenne, French Guyana. In: XV symposium of SELPER.

20. Verrelst, J., **Rivera, J.P.**, Leonenko, G., Alonso, L., Moreno, J. (2012). Using the ARTMO toolbox for automated retrieval of biophysical parameters through radiative transfer model inversion: Optimizing LUT-based inversion. In: EGU General Assembly 2012, 22-27 April, Vienna, Austria.

21. Verrelst, J., **Rivera, J.P.**, Guanter, L., Alonso, L., Moreno, J. (2012). Evaluating machine learning regression algorithms for operation retrieval of biophysical parameters: Opportunities for Sentinel. In: Sentinel-2 Symposium 2012, 23-27 April, Frascati, Italy.

22. Verrelst, J., **Rivera, J.P.**, Leonenko, G., Alonso, L., Moreno, J. (2012). ARTMO: a toolbox for automated retrieval of biophysical parameters through inversion of plant radiative transfer models. In: 2nd Terrabites Symposium 2012, 6-8 February, Frascati, Italy.

23. Verrelst, J., **Rivera, J.P.**, Guanter, L., Moreno, J. (2012). Evaluating the robustness of machine learning algorithms for operational retrieval of biophysical parameters. In: 2nd Terrabites Symposium 2012, 6-8 February, Frascati, Italy.

24. Verrelst, J., Muñoz-Marí, J., Alonso, L., **Rivera J.P.**, Moreno, J. (2011). Gaussian Processes regression for biophysical parameter retrieval from imaging spectroscopy data: opportunities for Sentinel-2 and -3. In: EARSeL 7th SIG-Imaging Spectroscopy Workshop 2011, 11-13 April, Edinburgh, UK.

25. Verrelst, J., **Rivera J.P.**, Alonso, L., Moreno, J. (2011). ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. In: EARSeL 7th SIG-Imaging Spectroscopy Workshop 2011, 11-13 April, Edinburgh, UK.

26. Verrelst, J., Muñoz-Marí, J., Alonso, L., **Rivera J.P.**, Moreno, J. (2011). Gaussian Processes regression for biophysical parameter retrieval from Sentinel-2 and Sentinel-3 data. In: European Geosciences Union (EGU) General Assembly 2011, 03-08 April, Vienna, Austria.

27. Verrelst, J., **Rivera J.P.**, Alonso, L., Moreno, J. (2011). An Automated Radiative Transfer Models Operator (ARTMO) toolbox for automated retrieval of biophysical parameters through model inversion. In: European Geosciences Union (EGU) General Assembly 2011, 03-08 November, Vienna, Austria.

28. Verrelst, J., Alonso, L., Delegido, J., **Rivera J.P.**, Moreno, J. (2010). Gaussian Processes regression for robust retrieval of chlorophyll content from imaging spectroscopy data. In: 4th Int. Workshop on Remote Sensing of Vegetation Fluorescence, 15-17 November, Valencia, Spain.

29. Verrelst J., Romijn, E., Kooistra L., **Rivera, J.P.**, Alonso, L., Moreno, J. (2010). Mapping vegetation density properties in a river floodplain and agricultural ecosystem using CHRIS-PROBA data. In: The Third International Symposium On Recent Advances In Quantitative Remote Sensing (RAQRS) 2010, 27 September - 1 October, Valencia, Spain.

Additionally, the outcomes of this work are relevant for the research carried out by the author and her colleagues at the University of València in the context of different research projects in which they are involved. A list of the related projects in which the author of this Thesis has participated is provided:

- Spanish Ministry for Science and Innovation: Estimación de parametros biofisicos para la misión FLEX a partir de Sentinel-3 (AYA2010-21432-C02-01/ESP), Ministerio de Educación y Ciencia (Spain) [2011-2013].
- STEREO II research project: Biomonitoring of urban habitat quality with airborne hyperspectral observations (BIOHYPE), Belgian Science Policy Office (BELSPO) in the frame of the STEREO II program-project BIOHYPE (SR/00/131) [2011-2014].
- ESA research project: Atmospheric Corrections for Fluorescence Signal and Surface Pressure Retrieval Over Land (FLUSS), ESA-ESTEC Contract No. 4000102733 [2011-2013].
- ESA research project: FLEX - Performance Analysis and Requirements Consolidation Study (PARCS), ESA-ESTEC Contract No. 4000105078/11/NL/AF [2012-2014].
- ESA research project: FLEX/Sentinel-3 Tandem Mission Photosynthesis Study (PS), ESA-ESTEC Contract No. 7088/12/NL/AF [2012-2014].
- ESA research project: FLEX simulator high level definition, ESA-ESTEC Contract No. 4000108364/ 13/NL/BJ [2012-2014].

7.6 Acknowledgments

Special gratitude is expressed to the co-authors and anonymous reviewers who revised the peer-reviewed Chapters 3-6. Their numerous critical comments greatly contributed to improving the Thesis as a whole. This Thesis has been possible thanks to the financial support of the following projects:

- Spanish Ministry for Science and Innovation: Estimación de parametros biofisicos para la misión FLEX a partir de Sentinel-3 (AYA2010-21432-C02-01/ESP), Ministerio de Educación y Ciencia (Spain) [2011-2013].
- ESA research project: FLEX - Performance Analysis and Requirements Consolidation Study (PARCS), ESA-ESTEC Contract No. 4000105078/11/NL/AF [2012-2014].

A

Summary in Spanish

Contents

A.1 Introducción

La Biosfera es uno de los principales sistemas que conforman la Tierra. Su estudio permite comprender la relación entre la vegetación y el ciclo del carbono y cómo éste puede ser afectado por los cambios en los niveles de CO2 y los usos de suelo [IPCC, 2007]. Para el estudio de estas dinámicas a escala global y local, han sido desarrollados diversos modelos que son representaciones de la realidad en una escala y complejidad más simple.

Los modelos de circulación general son algoritmos numéricos avanzados que en la actualidad se usan para la simulación de los complejos procesos en la atmósfera, la cubierta terrestre y los océanos a escala global [Donner *et al.*, 2011]. Los modelos globales de la dinámica vegetal - DGVM (por sus siglas en inglés Dynamic Global Vegetation Models) simulan los cambios espacio-temporales en la cubiertas vegetales asociados a la biogeología, hidrología y climatología [Malenovský, 2013].

Parte de las variables de entrada de estos modelos son obtenidas mediante medidas de teledetección gracias al Global Climate Observing System (GCOS), que ha determinado un conjunto de 50 variables climáticas esenciales que contribuyen a los estudios de cambio climático que lidera la Convención Marco de las Naciones Unidas y el Panel Intergubernamental del Cambio Climático [GCOS, 2014]. En esta lista está incluido el índice de área foliar (LAI). El contenido de clorofila en hoja (LCC) es otro parámetro biofísico clave para los estudios de biosfera.

La relevancia de estas dos variables se refleja en el papel que juegan en los estudios de evaluación del vigor y el crecimiento de la vegetación, la parametrización de las dinámicas vegetación-atmósfera en los modelos climáticos, estado fisiológico de las plantas y su actividad fotosintética [Lichtenthaler, 1987; Running and Coughlan, 1988; Dorigo *et al.*, 2007; Jonckheere *et al.*, 2004; Gitelson *et al.*, 2005; Schlemmer *et al.*, 2005; Zhang *et al.*, 2008; Peng and Gitelson, 2012]. La teledetección proporciona métodos para la estimación de éstas y otras variables climáticas esenciales [Baret *et al.*, 2013; Hollmann *et al.*, 2013].

Para contribuir al estudio de la biosfera a escala local como es el caso de la agricultura de precisión, modelos de producción agrícola o de los estudios ecológicos, esta tesis se enfoca en la estimación de propiedades vegetales a partir de datos espectrales obtenidos por plataformas espaciales, concretamente los parámetros LAI y LCC. El estudio de las propiedades de la vegetación desde el espacio requiere: (1) Métodos óptimos para el procesamiento y la estimación de la información y, (2) Disponibilidad de datos espaciales.

Los métodos de procesado y estimación de parámetros biofísicos son necesarios ya que el sensor solo mide los flujos de energía reflejados por las cubiertas vegetales distribuidos espacialmente. Por ello, han sido desarrollados diversos modelos, que van desde complejos modelos con base física hasta modelos estadísticos o la combinación de los anteriores. En el desarrollo de esta tesis se ha reunido una amplia variedad de ellos.

La disponibilidad de datos de calidad ha sido siempre uno de los retos de las agencia espaciales ya que el monitoreo de la vegetación requiere de datos espectrales de calidad,

adquiridos de manera regular y con un amplio histórico para analizar la evolución espacio-temporal de la vegetación. Para ello, la Agencia Espacial Europea (ESA) ha desarrollado la misión Sentinel-2 que está especialmente diseñada para el monitoreo de las propiedades de la vegetación, con las capacidades operativas que cumplen los requerimientos espectrales, espaciales y temporales [Drusch *et al.*, 2012]. Los datos que proporcionará la misión Sentinel-2 permitirán garantizar la continuidad de las misiones presentes Spot y Landsat, aportando un tiempo de revisita menor, mejora de la amplitud de barrido, mayor resolución espectral y una mejor calibración y calidad de imagen.

Para el procesamiento y la extracción de información de parámetros biofísicos han sido desarrollados diferentes paquetes computacionales por diversos grupos de investigación. Esta tesis pretende suministrar un conjunto de herramientas computacionales, dinámicas y flexibles que permitan automatizar y evaluar el potencial de los diferentes métodos que en la actualidad han sido publicados y están disponibles para su libre uso. Presenta los resultados científicos de la evaluación del impacto de diferentes parámetros de ajuste en los principales métodos de estimación de parámetros biofísicos, centrándonos en datos simulados del satélite Sentinel-2, previsto para ser lanzado en 2015.

Para dicho trabajo se han reunido los principales métodos de estimación que van desde las simples relaciones espectrales hasta los complejos modelos de transferencia radiativa (RTM). Para esto, hemos implementado un conjunto de herramientas informáticas que permiten el diseño y evaluación de diversas estrategias de regularización como son la normalización de los datos, la sinergia entre datos simulados por RTM y datos de campañas de campo o de laboratorio, adición de modelos de ruido a los datos simulados y un amplio conjunto de métodos de regresión tanto paramétricos como no paramétricos.

Este trabajo constituye la continuación de mi trabajo Final del Máster de Teledetección, donde se desarrolló una herramienta informática llamado ARTMO (por sus siglas en inglés Automated Radiative Transfer Models Operator) que reunió los RTM de la familia Prospect, SAIL y FLIGTH. Se implementó el método de estimación por tablas de búsqueda (LUT). Esta tesis presenta la evolución de ARTMO que pasa de ser una herramienta informática rígida que no permitía de manera sencilla la ampliación de sus funciones, a un flexible marco de desarrollo (framework software), donde ARTMO se convierte en una plataforma de soporte de diversos módulos implementados de manera independiente.

Esta nueva versión de ARTMO permite a cualquier grupo de investigación desarrollar y compartir nuevas funciones, algoritmos y métodos de estimación de parámetros biofísicos. Además, hemos establecido las bases para la creación de una red tanto de usuarios como de desarrolladores en torno al estudio de las propiedades de la vegetación, sirviendo de apoyo para el estudio de nuevos algoritmos de estimación, diseño de nuevos sensores ópticos o para su uso en el campo de la educación.

A.2 Estructura

La presente tesis doctoral se presenta en el formato de compendio de publicaciones. Para ello, según la normativa vigente, se exige que la tesis sea estructurada alrededor del trabajo de, al menos, tres artículos científicos publicados. La tesis se ha dividido en siete capítulos:

1. El Capítulo 1 presenta la justificación de la temática elegida, se presentan los objetivos de este trabajo y se explica la aportación original del autor en cada uno de los artículos seleccionados.

2. El Capítulo 2 muestra una síntesis de los últimos avances de la teledetección en el estudio de la vegetación y en especial los métodos de estimación de parámetros biofísicos (paramétricos, no paramétricos y con base física). Se presenta, además, la estructura del novedoso ARTMO como un marco de desarrollo o framework.

3. El Capítulo 3, corresponde a la publicación [Rivera *et al.*, 2014a] en donde, utilizando el módulo de 'índices espectrales' de ARTMO, se han evaluado los diferentes parámetros de ajuste como son el porcentaje de datos para la construcción del modelo, la combinación de bandas y el método de regresión usado para relacionar el parámetro biofísico con su respectivo índice espectral.

4. El Capítulo 4, corresponde a la publicación [Rivera *et al.*, 2013a] en donde se presenta el módulo de estimación por tablas de búsqueda (LUT) de ARTMO para la inversión de parámetros biofísicos con datos simulados a partir de RTM. En este artículo se analiza el impacto de la función de mérito y tres opciones de regularización: normalización de datos, adición de ruido Gaussiano y media de los mejores resultados.

5. En el Capítulo 5 [Rivera *et al.*, 2013b] se presenta el módulo de MLRA (por sus siglas en inglés Machine Learning Retrieval Algorithm) de ARTMO que se especializa en la inversión de parámetros biofísicos con algoritmos de regresión de tipo 'Machine Learning'. Este artículo explora el impacto de seis algoritmos de regresión en la inversión de los parámetros de LAI y LCC.

6. El Capítulo 6 corresponde a la publicación [Verrelst *et al.*, 2013b] que aborda el tema de la incertidumbre en la estimación de parámetros biofísicos usando MLRA. Evalúa el uso de modelos no paramétricos para la estimación de propiedades de la vegetación en zonas diferentes a donde se han tomado los datos para el desarrollo del modelo.

7. El Capítulo 7 presenta la síntesis y análisis de los resultados obtenidos, se muestran las conclusiones y se proponen líneas futuras de investigación.

A.3 Objetivo y metodología

El objetivo de esta tesis es analizar, optimizar y automatizar diferentes métodos de estimación de parámetros biofísicos por teledetección enfocándose en los datos que generará la misión Sentinel-2. Para lograr este objetivo se ha desarrollado un conjunto de tres herramientas informáticas especializadas en la inversión de parámetros biofísicos o toolbox. Cada toolbox se ha focalizado en cada uno de los tres grupos de métodos para la estimación de parámetros biofísicos: índices espectrales, métodos estadísticos de regresión y tablas de búsqueda. Se ha actualizado el software ARTMO para convertirlo en una plataforma donde se integran diversos módulos y herramientas que permiten el estudio de las propiedades de la vegetación.

Para la validación de resultados, se ha usado una amplia base de datos de campo que abarca diversos tipos de cultivos y árboles, en distintas etapas fisiológicas y con diferentes tipos y condiciones del suelo. Estos datos se obtuvieron durante las campañas SPARC (Spectra bARrax Campaign) llevadas a cabo por la ESA los años 2003 y 2004. Los parámetros biofísicos fueron medidos en 110 unidades elementales de muestreo o ESU (por sus siglas en inglés Elementary Sampling Units) que representa un pixel de 20 m x 20 m [Gandía *et al.*, 2004; Fernández *et al.*, 2005]. Durante la campaña se contó con imágenes del sensor CHRIS sobre la plataforma espacial PROBA y datos del sensor HyMAP sobre una plataforma aerotransportada. La base de datos se completó con 60 espectros de suelo desnudo, superficies artificiales y cuerpos de agua para un total de 170 muestras.

Se realizó un remuestreo espacial y espectral de los datos de SPARC a la configuración de la misión Sentinel-2 para los canales con una resolución espacial menor a 20 m.

Se analizaron dos familias de índices espectrales, Ratio Simple (SR) y diferencias normalizadas (ND), para los cuales se estudiaron todas las posibles combinaciones espectrales y se evaluaron el impacto de 4 funciones de ajuste de tipo lineal, logarítmico, exponencial y polinómica de grado 2.

Se generó una LUT de 100,000 datos simulados con el RTM PROSAIL. Se evaluó el impacto de la función de mérito (FM) que permite discriminar la similitud entre los espectros simulados, o medidos en campo o laboratorio (Q) y los espectros medidos por un sensor óptico (P), donde D[P,Q] representa la distancia entre dos espectros. Evaluamos 18 FM clasificadas en tres familias: f-divergence, M-estimates y mínimo contraste, seleccionadas según el trabajo desarrollado por Leonenko *et al.* [2013]. Tambien se evaluó el impacto de tres opciones de regularización diferentes: ruido Gaussiano añadido a los datos, normalización de los espectros y promedio de múltiples soluciones.

Se evaluaron 6 métodos de regresión: Mínimos cuadrados parciales (PLS), Componentes principales (PCR), Árboles de decisión (DT), Redes neuronales (NN), Kernel Rigde Regression (KRR) y Gaussian processes regression (GPR). Éstos fueron entrenados con los datos de la campaña SPARC para un conjunto de tres sensores: Sentinel-2 (8 bandas), CHRIS (62 bandas) y HyMap (125 bandas). Se evaluó el impacto que tiene en la estimación de LAI y LCC el tamaño de la muestra de entrenamiento, donde se evaluó entre el 5% y el 95% con un incremento de 5%, analizando además el nivel de ruido Gaussiano añadido a las muestras de entrenamiento

variando entre un 0% y un 20%, con un incremento del 2%.

Se ha hecho una evaluación de la capacidad para extrapolar los modelos construidos a partir de datos de entrenamiento local, con el objetivo de estimar parámetros de vegetación a partir de imágenes con diferente configuración espectral, distinta localización y tiempo de adquisición. Para determinar el nivel de incertidumbre en la estimación de estos parámetros, se usó GPR. Se contó con una base de datos con medidas de campo y 8 imágenes adquiridas en España, Alemania y Canadá entre los años 2003 y 2009. GPR se entrenó con los datos de la campaña SPARC tomando un 80 % para entrenamiento y el restante para el test del modelo desarrollado.

A.4 Resultados

La estimación de los parámetros biofísicos de la vegetación a partir de datos espectrales obtenidos por teledetección, ha sido objeto de numerosos estudios debido al gran número de aplicaciones climáticas, agronómicas y medioambientales que posee. Por ello, los métodos desarrollados para su estudio han logrado grandes avances en las últimas tres décadas. En esta tesis se reúnen los principales y más novedosos métodos para la estimación de parámetros biofísicos por teledetección, que se pueden clasificar en tres familias: (1) Paramétricos, (2) No paramétricos y (3) Métodos con base física. Cada una de estas familias han sido evaluadas sistemáticamente y optimizadas en los capítulos del 3 al 5 y se han desarrollado herramientas informáticas operativas para su automatización. Cada una de estas herramientas permite analizar y comparar una amplia gama de métodos de inversión y generar de manera automática mapas de parámetros biofísicos a partir de imágenes obtenidas por teledetección.

A continuación se presentan los principales resultados obtenidos en el desarrollo de esta tesis, para cada uno de los principales métodos de estimación de parámetros biofísicos: paramétricos, no paramétricos y con base física, pues se han evaluado de manera sistemática los principales parámetros en cada método.

Índices Espectrales: Desde los inicios de la teledetección, han sido utilizados y desarrollados diferentes índices espectrales en el estudio de la vegetación. El capítulo 3 presenta la herramienta informática implementada en ARTMO para evaluar el funcionamiento de los parámetros involucrados en el diseño de índices espectrales:

- Combinación de bandas: La herramienta permite el diseño de relaciones espectrales combinando hasta 10 bandas diferentes.
- Método de ajuste: se han implementado las funciones de ajuste lineal, exponencial, logarítmico y polinómico.
- Parámetros de regularización: permite la adición de nivel de ruido Gaussiano en los datos espectrales y los parámetros biofísicos.

Para la estimación de LAI, se han evaluado dos índices muy usados en teledetección: Ratio Simple (SR) y Diferencias Normalizadas (ND). Los mejores resultados para el índice espectral

SR se han obtenido con la configuración de bandas centradas en 570 nm y 2453 nm y ajuste de tipo polinomio de grado 2 con un valor del RMSE de 0,6. Para el índice espectral tipo ND, la mejor configuración de bandas es la de 555 nm y 2453 nm mediante un ajuste de tipo logarítmico obteniendo el mismo valor de RMSE. Para la estimación de LCC con el índice espectral tipo SR, el mejor resultado se obtuvo con la combinación de bandas de 692 nm y 1689 nm mediante un ajuste de un polinomio de grado 2, con un valor de RMSE de 4,2 μg/cm^2 mientras que para el índice espectral tipo ND la combinación de bandas fue 692 nm y 1661 nm con un ajuste de tipo polinomio de grado 2 y con un valor de RMSE de 4,3 μg/cm^2.

Tablas de búsqueda: El artículo del capítulo 4 presenta la herramienta informática para la evaluación sistemática de diferentes funciones de mérito, y 3 opciones de regularización: normalización de datos espectrales, adición de ruido Gaussiano a los datos sintéticos y valor medio de las mejores múltiples soluciones. Se evaluaron 18 funciones de mérito pertenecientes a las familias: information measures, M-estimates y mínimo contraste. Los mejores resultados en la estimación de LCC se obtuvieron con la función de mérito 'L1-estimate' perteneciente a la familia de M-estimates con los valores normalizados, con un error relativo de 17,6 % (r^2 = 0,73). Para LAI se obtuvieron con la función de mérito 'least-squares estimator' perteneciente a la familia de M-estimates con valores no normalizados y con un error relativo del 15,3 % (r^2 = 0,74).

Métodos no paramétricos: El artículo del capítulo 5 presenta la herramienta informática que se especializa en la inversión de parámetros biofísicos con métodos estadísticos paramétricos y no paramétricos. Permite la comparación de métodos estadísticos de inversión y la evaluación de manera sistemática de los ajustes generales de estos métodos según: 1. El tamaño de la base de datos de entrenamiento, 2. Trabajar con datos de campo y datos de simulación con RTM y 3. Adición de ruido Gaussiano a los datos espectrales y los parámetros biofísicos. Del conjunto de estimadores estadísticos implementados se evaluaron siete: PLSR, PCR, RT, NN, KRR y GPR. En términos de robustez del modelo, los mejores resultados en la estimación de LCC se obtuvieron con KRR, ya que con menos datos de entrenamiento se obtuvieron los menores errores relativos de 6,8 %, 3,8 % y 4,6 % para las imágenes de Sentinel-2, CHRIS y HyMAP respectivamente. En la estimación de LAI los métodos más robustos fueron DT con un error relativo del 6,1 %, NN con un error relativo de 6,5 % y KRR con un error relativo de 6,8 % para las imágenes de Sentinel-2, CHRIS y HyMAP respectivamente.

El capítulo 6 explora el uso de métodos estadísticos no paramétricos para la estimación de parámetros biofísicos entrenados con datos que no pertenecen a la zona de interés. Una de las grandes limitaciones de los métodos estadísticos es la necesidad de realizar un entrenamiento previo para desarrollar el modelo. Esto limita su uso en zonas donde no se cuenta con datos de entrenamiento. El método estadístico seleccionado fue GPR debido a su robustez, como ya se mostró en el capítulo 5. Los resultados muestran que GPR alcanzó el resultado óptimo menor al 10 % de error relativo para los dos parámetros en estudio, pues para la estimación de LCC su error relativo estuvo entre 3,5 % a 9,2 % y para el LAI se obtuvo un error relativo entre 6,5 % y 7,3 %. Además se encontraron bajos valores de incertidumbre en la estimación en las zonas de cultivos, aumentando la incertidumbre en píxeles de suelos desnudos o bosques.

A.5 Conclusiones

La herramienta informática ARTMO, que ha sido ampliada y mejorada en esta tesis, es una potente herramienta que permite a investigadores y usuarios de la teledetección el uso de los principales métodos para la estimación de parámetros biofísicos de la vegetación. Además, al ser capaz de utilizar de manera simultánea un elevado conjunto de métodos estadísticos y modelos de inversión, permite evaluar la configuración idónea para futuras misiones o la selección del mejor método para un sensor determinado. En concreto, al evaluar datos simulados de Sentinel-2, se obtienen las siguientes conclusiones:

1. No hay evidencia de que las regiones espectrales usadas en las clásicas combinaciones espectrales sean las óptimas para la estimación de parámetros como LCC y LAI. La región del SWIR muestra una gran sensibilidad para el estudio de LCC y LAI, y aunque no son regiones espectrales donde dichas propiedades interaccionen de manera directa, sin embargo permiten mejorar la estimación utilizando índices. La toolbox de índices espectrales es una herramienta robusta para el desarrollo de nuevos índices para sensores hiperespectrales ya que permite la combinación de hasta 10 bandas.

2. En la estimación con LUT, existe una relación entre la función de mérito utilizada, la estrategia de regularización y los parámetros biofísicos, que hace necesario que su inversión se analice de manera independiente. La clásica función de mérito RMSE no se muestra como la más precisa. Como alternativa se presentan los estimadores robustos de la familia M-estimates, que junto con la óptima estrategia de regularización, mejora las estimaciones de LAI y LCC. La normalización de los datos optimiza la estimación de LAI pero no la de LCC. El promediar los mejores resultados es una estrategia efectiva para mitigar el problema mal planteado inherente en la inversión de parámetros biofísicos.

3. Los novedosos algoritmos de regresión de tipo Kernel, como los GPR y KRR, sobresalen respecto al resto de algoritmos tanto paramétricos como no paramétricos, por su capacidad de adaptabilidad y precisión en estimación de LAI y LCC.

4. El método estadístico GPR permite, además, obtener información acerca de la incertidumbre de la estimación, convirtiéndolo en una herramienta potente y robusta para la extrapolación de los modelos desarrollados a nuevas zonas diferentes, rompiendo de alguna manera la idea de la necesidad de datos de entrenamiento local para la obtención de resultados aceptables.

El conjunto de herramientas informáticas desarrolladas en este trabajo incorporando los principales métodos de estimación de los parámetros biofísicos, permitirá el desarrollo de nuevas estrategias para el estudio de la vegetación y dise sensores ópticos. Además, la novedosa estructura de programación de ARTMO mejorada en este trabajo, nos permite plantearnos las siguientes líneas de trabajos:

- Sinergia con modelos atmosféricos: la integración de los modelos de transferencia radiativos de la vegetación, la atmósfera y el suelo permitirán evaluar los diferentes métodos de inversión con datos espectrales en el techo de la atmósfera y desarrollar estrategias de estimación de parámetros biofísicos que puedan ser aplicados a diferentes localidades, distintas condiciones atmosféricas y diferentes tipos de suelo.

- Generación de escenarios multitemporales: el estudio temporal de las propiedades de la vegetación permitirá el desarrollo de novedosas estrategias de inversión. Los diferentes módulos de ARTMO conforman la base para la desarrollo de generadores de escenas genéricos que puedan explorar la información que se puede extraer de la evolución espacio temporal de las cubiertas vegetales.

El trabajo presentado en esta tesis, ha permitido avanzar en el estudio de las propiedades de la vegetación. Los objetivos alcanzados y las nuevas líneas en las que se están trabajando, como contar con una comunidad de usuarios especializados y desarrolladores de herramientas informáticas, aportarán nuevas técnicas al estudio de la vegetación y el ciclo de carbono y permitirá mejorar y facilitar el conocimiento de la vegetación a diferentes escalas por técnicas de teledetección.

Acronyms and Abbreviations

ARTMO Automated Radiative Transfer Models Operator

BRDF Bidirectional Reflectance Distribution Function

CCC Canopy-integrated Chlorophyll Content

CF Cost Function

CHRIS Compact High Resolution Imaging Spectrometer

CV Coefficient of Variation

DGVM Dynamic Global Vegetation Models

DT Decision Trees

EO Earth Observation

ECV Essential Climate Variables

ESA European Space Agency

ESU Elementary Sampling Units

FAPAR Fraction of absorbed photosynthetically active radiation

FLEX Fluorescence Explorer

FWHM Full-Width at Half-Maximum

GCOS Global Climate Observing System

GPR Gaussian processes regression

GSA Global Sensitivity Analysis

GUI Graphical User Interface

IPCC Intergovernmental Panel on Climate Change

KRR Kernel Ridge regression

LAI Leaf Area Index

LCC leaf chlorophyll content

LSE least squares estimation

LUT Look-Up Table

MLRA Machine Learning Regression Algorithm

MODTRAN MODerate resolution TRANsmittance

MSI Multi-Spectral Instrument

ND Normalized Difference

NDVI Normalized Difference Vegetation Index

NIR Near–infraRed

NN Neural networks

PCA Principal Components Analysis

PCR Principal component regression

PFT Plant Functional Types

PLSR Partial least squares regression

RMSE Root Mean Squared Error

RTM Radiation Transfer Model

RVM Relevance Vector Machines

RR Ridge regression

RS Remote Sensing

RT Regression Trees

S2 Sentinel-2

SGM Scene Generator Module

SI Spectral Index

SPARC SPectra bARrax Campaign

SR simple ratio

SVR Support vector machines

SWIR ShortWave-InfraRed

TOA Top Of Atmosphere

TOC Top Of Canopy

VNIR Visible and Near InfraRed

References

Allen, W., Gausman, H., Richardson, A., and Thomas, J. (1969). Interaction of isotropic light with a compact plant leaf. *J Opt Soc Amer*, **59**(10), 1376–1379.

Allen, W., Gausman, H., and Richardson, A. (1970). Mean effective optical constants of cotton leaves. *Journal of the Optical Society of America*, **60**(4), 542–547.

Alonso, L. and Moreno, J. (2005). Advances and limitations in a parametric geometric correction of CHRIS/PROBA data. *Proceedings of the 3rd CHRIS/Proba Workshop, ESA/ESRIN, Frascati, Italy.*

Arenas-García, J., Petersen, K. B., Camps-Valls, G., and Hansen, L. K. (2013). Kernel multivariate analysis framework for supervised subspace learning. *IEEE Signal Processing Magazine*, page 1.

Aschbacher, J. and Milagro-Pérez, M. (2012). The European Earth monitoring GMES programme: Status and perspectives. *Remote Sensing of Environment*, **120**, 3–8.

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

Atzberger, C. and Richter, K. (2012). Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. *Remote Sensing of Environment*, **120**, 208–218.

Bacour, C., Baret, F., Béal, D., Weiss, M., and Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAI$\times$Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment*, **105**(4), 313–325.

Báez-González, A., Chen, P.-Y., Tiscareño López, M., and Srinivasan, R. (2002). Using satellite and field data with crop growth modeling to monitor and estimate corn yield in mexico. *Crop Science*, **42**(6), 1943–1949.

Baret, F. (2010). Biophysical vegetation variables retrieval from remote sensing observations. *Proceedings of Remote Sensing for Agriculture, Ecosystems, and Hydrology XII*, **7824**, 17–19.

Baret, F. and Buis, S. (2008). Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. *In S. Liang (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application, Springer*, pages 171–200.

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

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J., and Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm. *Remote Sensing of Environment*, **110**(3), 275–286.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., and 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**(0), 299 – 309.

Barnsley, M. J., Settle, J. J., Cutter, M. A., Lobb, D. R., and Teston, F. (2004). The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(7), 1512–1520.

Breece, I. H. and Holmes, R. (1971). Biderectional scattering characteristics of healthy green soybean and corn leaves in vivo. *Applied Optics*, **10**(1), 119–127.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.

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

Brown, L., Chen, J., Leblanc, S., and Cihlar, J. (2000). A shortwave infrared modification to the simple ratio for lai retrieval in boreal forests: An image and model analysis. *Remote Sensing of Environment*, **71**(1), 16–25.

Buermann, W., Dong, J., Zeng, X., Myneni, R. B., and Dickinson, R. E. (2001). Evaluation of the utility of satellite-based vegetation leaf area index data for climate simulations. *Journal of Climate*, **14**(17), 3536–3550.

Camps-Valls, G. and Bruzzone, L., editors (2009). *Kernel methods for Remote Sensing Data Analysis*. Wiley & Sons, UK.

Camps-Valls, G., Gómez-Chova, L., Vila-Francés, J., Amorós-López, J., Muñoz-Marí, J., and Calpe-Maravilla, J. (2006). Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sensing of Enviroment*, **105**(1), 23–33.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., and Calpe-Maravilla, J. (2009). Biophysical parameter estimation with adaptive Gaussian Processes. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2009*, Capetown, South Africa.

Camps-Valls, G., Muñoz-Marí and, J., Gómez-Chova, L., Guanter, L., and Calbet, X. (2012a). Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1759–1769.

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., and Malo, J. (2012b). Remote sensing image processing. *Synthesis Lectures on Image, Video, and Multimedia Processing*, **12**, 1–194.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Lázaro-Gredilla, M., and Verrelst, J. (2013). simpleR: A simple educational Matlab toolbox for statistical regression. V2.1.

Ceccato, P., Flasse, S., and Grégoire, J.-M. (2002). Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sensing of Environment*, **82**(2-3), 188–197.

Chen, J. and Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using landsat tm images. *Remote Sensing of Environment*, **55**(2), 153–162.

Chen, J., Rich, P., Gower, S., Norman, J., and Plummer, S. (1997). Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research D: Atmospheres*, **102**(24), 29429–29443.

Chen, J., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S., White, H., Hall, R., Peddle, D., King, D., Trofymow, J., Swift, E., Van Der Sanden, J., and Pellikka, P. (2002). Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sensing of Environment*, **80**(1), 165–184.

Cho, M. b., Skidmore, A. b., Corsi, F., van Wieren, S., and Sobhan, I. b. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, **9**(4), 414–424.

Clevers, J. and Gitelson, A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, **23**(1), 344–351.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., and Shields, T. (1998). The HyMap airborne hyperspectral sensor: The system, calibration and performance. *Proc. 1st EARSeL Workshop on Imaging Spectroscopy*, pages 37–42.

Colombo, R., Bellingeri, D., Fasolini, D., and Marino, C. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment*, **86**(1), 120–131.

Combal, B., Baret, F., and Weiss, M. (2002). Improving canopy variables estimation from remote sensing data by exploiting ancillary information. case study on sugar beet canopies. *Agronomie*, **22**(2), 205–215.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., and 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), 1–15.

Coops, N. C., Smith, M.-L., Martin, M., and Ollinger, S. V. (2003). Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1338–1346.

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

Darvishzadeh, R., Skidmore, A., Schlerf, M., and Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, **112**(5), 2592–2604.

Darvishzadeh, R., Atzberger, C., Skidmore, A., and Abkar, A. (2009). Leaf area index derivation from hyperspectral vegetation indicesand the red edge position. *International Journal of Remote Sensing*, **30**(23), 6199–6218.

Darvishzadeh, R., Matkan, A. A., and Dashti Ahangar, A. (2012). Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **PP**(99), 1 –9.

Dash, J. and Curran, P. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, **25**(23), 5403–5413.

Delegido, J., Verrelst, J., Alonso, L., and Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, **11**(7), 7063–7081.

Delegido, J., Verrelst, J., Meza, C., Rivera, J., Alonso, L., and Moreno, J. (2013). A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy*, **46**, 42–52.

Disney, M., Lewis, P., and Saich, P. (2006). 3D modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. *Remote Sensing of Environment*, **100**(1), 114–132.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., and Sciarra, R. (2012). The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment*, **120**, 37–57.

Donner, L., Wyman, B., Hemler, R., Horowitz, L., Ming, Y., Zhao, M., Golaz, J.-C., Ginoux, P., Lin, S.-J., Schwarzkopf, M., Austin, J., Alaka, G., Cooke, W., Delworth, T., Freidenreich, S., Gordon, C., Griffies, S., Held, I., Hurlin, W., Klein, S., Knutson, T., Langenhorst, A., Lee, H.-C., Lin, Y., Magi, B., Malyshev, S., Milly, P., Naik, V., Nath, M., Pincus, R., Ploshay, J., Ramaswamy, V., Seman, C., Shevliakova, E., Sirutis, J., Stern, W., Stouffer, R., Wilson, R., Winton, M., Wittenberg, A., and Zeng, F. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, **24**(13), 3484–3519.

Dorigo, W., Richter, R., Baret, F., Bamler, R., and Wagner, W. (2009). Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach. *Remote Sensing*, **1**(4), 1139–1170.

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

Draper, N. and Smith, H. (1998). Applied regression analysis. *3rd Ed.; Wiley Series in Probability and Statistics*.

Drusch, M., Gascon, F., and Berger, M. (2010). Sentinel-2 mission requirements document.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, **120**, 25–36.

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

Fang, H. and 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**(3), 405–424.

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

Feret, J. B., François, C., Asner, G. P., Gitelson, A. A., Martin, R. E., Bidel, L. P. R., Ustin, S. L., le Maire, G., and Jacquemoud, S. (2008). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, **112**(6), 3030–3043.

Fernandes, R. and Leblanc, S. (2005). Parametric (modified least squares) and non-parametric (theil-sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. *Remote Sensing of Environment*, **95**(3), 303–316.

Fernández, G., Moreno, J., Gandía, S., Martínez, B., Vuolo, F., and Morales, F. (2005). Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns. *Proceedings of the SPARC Workshop*.

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

Gandía, S., Fernández, G., Garcia, J. C., and Moreno, J. (2004). Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign. *Proceedings of the 2nd CHRIS/PROBAWorkshop*.

Ganguly, S., Nemani, R., Zhang, G., Hashimoto, H., Milesi, C., Michaelis, A., Wang, W., Votava, P., Samanta, A., Melton, F., Dungan, J., Vermote, E., Gao, F., Knyazikhin, Y., and Myneni, R. (2012). Generating global leaf area index from landsat: Algorithm formulation and demonstration. *Remote Sensing of Environment*, **122**, 185–202.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., Fernandes, R., Plummer, S., Shabanov, N., Myneni, R., Knyazikhin, Y., and Yang, W. (2008). Validation and intercomparison of global leaf area index products derived from remote sensing data. *Journal of Geophysical Research G: Biogeosciences*, **113**(2).

GCOS (2011). *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). Available online at http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf*, page 138.

GCOS (2014). Gcos essential climate variables. Page visited in March 2014.

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

Gianelle, D. and Guastella, F. b. (2007). Nadir and off-nadir hyperspectral field data: Strengths and limitations in estimating grassland biophysical characteristics. *International Journal of Remote Sensing*, **28**(7), 1547–1560.

Gianquinto, G., Orsini, F., Fecondini, M., Mezzetti, M., Sambo, P., and Bona, S. (2011). A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield. *European Journal of Agronomy*, **35**(3), 135–143.

Gitelson, A., Gritz, Y., and Merzlyak, M. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, **160**(3), 271–282.

Gitelson, A., Vina, A., Ciganda, V., Rundquist, D., and Arkebauer, T. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, **32**(8), 1–4.

Gitelson, A. A., Merzlyak, M. N., and Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, **148**(3-4), 501–508.

Gobron, N., Pinty, B., Verstraete, M., and Widlowski, J.-L. (2000). Advanced vegetation indices optimized for up-coming sensors: design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(6), 2489–2505.

Goel, N. S. & Grier, T. (1998). a model for radiative transfer in heterogeneous three-dimensional canopies. In T. Guyenne and J. Hunt, editors, *Spectral Signatures of Objects in Remote Sensing*.

Gómez-Chova, L. and Camps-Valls, G. (2012). Learning with the kernel signal-to-noise ratio. In *IEEE Workshop on Machine Learning for Signal Processing, MLSP'12*, Santander, Spain.

Gond, V., De Pury, D., Veroustraete, F., and Ceulemans, R. (1999). Seasonal variations in leaf area index, leaf chlorophyll, and water content; scaling-up to estimate fapar and carbon balance in a multilayer, multispecies temperate forest. *Tree Physiology*, **19**(10), 673–679.

Gonsamo, A. (2011). Normalized sensitivity measures for leaf area index estimation using three-band spectral vegetation indices. *International Journal of Remote Sensing*, **32**(7), 2069–2080.

Gopal, S. and Woodcock, C. (1996). Remote sensing of forest change using artificial neural networks. *IEEE Trans. Geosci. Remote Sens*, **34**, 398–404.

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

Guanter, L., Alonso, L., and Moreno, J. (2005). A method for the surface reflectance retrieval from PROBA/CHRIS data over land: Application to ESA SPARC campaigns. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(12), 2908–2917.

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

Hadamard, J. (1902). Sur les problemes aux derivees partielles et leur signification physique. *Bull. Univ. Princeton*, **13**, 49–52.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer-Verlag, 2nd edition.

Hatfield, J., Gitelson, A., Schepers, J., and Walthall, C. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, **100**(3 SUPPL.), S117–S131.

Haykin, S. (1999). *Neural Networks – A Comprehensive Foundation*. Prentice Hall, 2nd edition.

Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., and Vesanto, V.-H. (2013). Sensitivity of narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species composition. *ISPRS Journal of Photogrammetry and Remote Sensing*, **78**, 1–14.

Hollmann, R., Merchant, C., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., Van Roozendael, M., and Wagner, W. (2013). The esa climate change initiative: Satellite data records for essential climate variables. *Bulletin of the American Meteorological Society*, **94**(10), 1541–1552.

Houborg, R. and Boegh, E. (2008). Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and spot reflectance data. *Remote Sensing of Environment*, **112**(1), 186–202.

Houborg, R., Anderson, M., and Daughtry, C. (2009). Utility of an image-based canopy reflectance modeling tool for remote estimation of lai and leaf chlorophyll content at the field scale. *Remote Sensing of Environment*, **113**(1), 259–274.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory*, **14**(1), 55–63.

IPCC (2007). Climate change 2007: the physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, page 996.

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

Jacquemoud, S. and Ustin, S. L. (2001). Leaf optical properties: A state of the art. In *8th International Symposium of Physical Measurements & Signatures in Remote Sensing*, pages 223–332. CNES.

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

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

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer-Verlag, Berlin; New York.

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

Jungclaus, J., Lorenz, S., Timmreck, C., Reick, C., Brovkin, V., Six, K., Segschneider, J., Giorgetta, M., Crowley, T., Pongratz, J., Krivova, N., Vieira, L., Solanki, S., Klocke, D., Botzet, M., Esch, M., Gayler, V., Haak, H., Raddatz, T., Roeckner, E., Schnur, R., Widmann, H., Claussen, M., Stevens, B., and Marotzke, J. (2010). Climate and carbon-cycle variability over the last millennium. *Climate of the Past*, **6**(5), 723–737.

Kimes, D. S., Nelson, R. F., Manry, M. T., and Fung, A. K. (1998). Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, **19**(14), 2639–2662.

Knyazikhin, Y., Kranigk, J., Myneni, R., Panfyorov, O., and Gravenhorst, G. (1998a). Influence of small-scale structure on radiative transfer and photosynthesis in vegetation canopies. *Journal of Geophysical Research D: Atmospheres*, **103**(D6), 6133–6144.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., and Running, S. (1998b). 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 D: Atmospheres*, **103**(D24), 32257–32275.

Koetz, B., Baret, F., Poilvé, H., and Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, **95**(1), 115–124.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Labate, D., Ceccherini, M., Cisbani, A., De Cosmo, V., Galeazzi, C., Giunti, L., Melozzi, M., Pieraccini, S., and Stagi, M. (2009). The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronautica*, **65**(9-10), 1429–1436.

Lazaro-Gredilla, M., Titsias, M., Verrelst, J., and Camps-Valls, G. (2013). Retrieval of biophysical parameters with heteroscedastic gaussian processes. *Geoscience and Remote Sensing Letters, IEEE*, **PP**(99), 1–5.

Lazaro-Gredilla, M., Titsias, M., Verrelst, J., and Camps-Valls, G. (2014). Retrieval of biophysical parameters with heteroscedastic gaussian processes. *IEEE Geoscience and Remote Sensing Letters*, **11(4)**, 838–842.

le Maire, G., François, C., and Dufrene, E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, **89**(1), 1–28.

le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailler, J.-Y., Bréda, N., Genet, H., Davi, H., and Dufrêne, E. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, **112**(10), 3846–3864.

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

Leonenko, G., North, P., and Los, S. (2013). Statistical distances and their applications to biophysical parameter estimation: Information measures, M-estimates, and minimum contrast methods. *Remote Sensing*, **5**(3), 1355–1388.

Li, X. and Strahler, A. H. (1992). Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: Effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(2), 276–292.

Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*, **31**(5), 501–516.

Liang, S., Fang, H., Kaul, M., Van Niel, T. G., Mcvicdr, T. R., Pearlman, J., Walthall, C. L., Daughtry, C., and Huemmrich, K. F. (2003). Estimation of land surface broadband albedos and leaf area index from EO-1 ALI data and validation. *IEEE Trans. Geosci. Rem. Sens.*, **41**, 1260–1268.

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol.*, **148**, 350–382.

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., and Miehe, J. A. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology*, **148**(5), 599–612.

Malenovský, Z. (2013). Brief information and list of dynamic global vegetation models (dgvms). Technical note for esa's photosynthesis study (ps), Global Change Research Centre, Czech Republic.

Malenovsky, Z., Rott, H., Cihlar, J., Schaepman, M., Garcia, J. C.a-Santos, G., Fernandes, R., and 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.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M., García-Santos, G., Fernandes, R., and 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.

Mariotto, I., Thenkabail, P., Huete, A., Slonecker, E., and Platonov, A. (2013). Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission. *Remote Sensing of Environment*, **139**, 291–305.

Meroni, M., Colombo, R., and Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, **92**(2), 195–206.

Moreno, J. and participants of the SPARC campaigns (2004). SPARC data acquisition report. Contract no: 18307/04/NL/FF, University Valencia.

Morisette, J., Baret, F., Privette, J., Myneni, R., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sánchez-Azofeifa, G., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iiames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E., Petersen, G., Seufert, G., and Cook, R. (2006). Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(7), 1804–1814.

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

Myneni, R. (1997). Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, **35**(6), 1380–1393.

Myneni, R. and Ross, J. (1991). Photon-vegetation interactions.

Myneni, R., Maggion, S., Iaquinta, J., Privette, J., Gobron, N., Pinty, B., Kimes, D., Verstraete, M., and Williams, D. (1995a). Optical remote sensing of vegetation: Modeling, caveats, and algorithms. *Remote Sensing of Environment*, **51**(1), 169–188.

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., and Running, S. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, **83**(1-2), 214–231.

Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L. (1995b). Interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, **33**(2), 481–486.

North, P. (1996). Three-dimensional forest light interaction model using a monte carlo method. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(4), 946–956.

O'Hagan, A. (1994). *Bayesian Inference*, volume 2B. Kendall's Advanced Theory of Statistics. Arnold, London, UK.

Orbanz, P. and Teh, Y.-W. (2010). *Bayesian Nonparametric Models*. Encyclopedia of Machine Learning. Springer.

Pardo, L. (2006). *Statistical Inference Based on Divergence Measures, Statistics: a series Textbooks and monographs, Chapman and Hall/CRC*.

Pasolli, L., Melgani, F., and Blanzieri, E. (2010). Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *Geoscience and Remote Sensing Letters, IEEE*, **PP**(99), 464–468.

Pedrós, R., Goulas, Y., Jacquemoud, S., Louis, J., and Moya, I. (2010). Fluormodleaf: A new leaf fluorescence emission model based on the prospect model. *Remote Sensing of Environment*, **114**(1), 155–167.

Peng, Y. and Gitelson, A. (2012). Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. *Remote Sensing of Environment*, **117**, 440–448.

Pinty, B., Gobron, N., Widlowski, J.-L., Gerstl, S., Verstraete, M., Antunes, M., Bacour, C., Gascon, F., Gastellu, J.-P., Goel, N., Jacquemoud, S., North, P., Qin, W., and Thompson, R. (2001). Radiation transfer model intercomparison (rami) exercise. *Journal of Geophysical Research D: Atmospheres*, **106**(D11), 11937–11956.

Pinty, B., Widlowski, J.-L., Taberner, M., Gobron, N., Verstraete, M., Disney, M., Gascon, F., Gastellu, J.-P., Jiang, L., Kuusk, A., Lewis, P., Li, X., Ni-Meister, W., Nilson, T., North, P., Qin, W., Su, L., Tang, S., Thompson, R., Verhoef, W., Wang, H., Wang, J., Yan, G., and Zang, H. (2004). Radiation transfer model intercomparison (rami) exercise: Results from the second phase. *Journal of Geophysical Research D: Atmospheres*, **109**(6), D06210 1–19.

Pinty, B., Lavergne, T., Vobbeck, M., Kaminski, T., Aussedat, O., Giering, R., Gobron, N., Taberner, M., Verstraete, M., and Widlowski, J.-L. (2007). Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products. *Journal of Geophysical Research D: Atmospheres*, **112**(10).

Pinty, B., Andredakis, I., Clerici, M., Kaminski, T., Taberner, M., Verstraete, M., Gobron, N., Plummer, S., and Widlowski, J.-L. (2011). Exploiting the MODIS albedos with the Two-Stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties. *Journal of Geophysical Research D: Atmospheres*, **116**(9).

Qu, Y., Wang, J., Wan, H., Li, X., and Zhou, G. (2008). A bayesian network algorithm for retrieving the characterization of land surface vegetation. *Remote Sensing of Environment*, **112**(3), 613–622.

Quinlan, J. (1986). Induction of decision trees. *Machine Learning*, **1**(1), 81–106.

RAMI (2014). RAdiation transfer Model Intercomparison RAMI. Page visited in March 2014.

Rasmussen, C. E. and Williams, C. K. I. (2006a). *Gaussian Processes for Machine Learning*. The MIT Press, New York.

Rasmussen, C. E. and Williams, C. K. I. (2006b). *Gaussian Processes for Machine Learning*. The MIT Press, New York.

Richter, K., Atzberger, C., Vuolo, F., W., and P., D'Urso, G. (2009). Experimental assessment of the sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Canadian Journal of Remote Sensing*, **35**(3), 230–247.

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

Richter, K., Atzberger, C., Hank, T., and Mauser, W. (2012a). Derivation of biophysical variables from earth observation data: Validation and statistical measures. *Journal of Applied Remote Sensing*, **6**(1).

Richter, K., Hank, T., Vuolo, F., Mauser, W., and D'Urso, G. (2012b). Optimal exploitation of the Sentinel-2 spectral capabilities for crop leaf area index mapping. *Remote Sensing*, **4**(3), 561–582.

Rivera, J. (2011). *Desarollo de una herramiento informatica para el estudio de la interaccion de la radiacion solar con las cubiertas vegetales*. Master's thesis, Image Processing Laboratory (IPL), University of Valencia, Spain.

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

Rivera, J., Verrelst, J., Alonso, L., Moreno, J., and Camps-Valls, G. (2013b). Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, page In press.

Rivera, J., Verrelst, J., Delegido, J., Veroustraete, F., and Moreno, J. (2014a). On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. *Remote Sensing*, **(In review)**.

Rivera, J., Sabater, N., Tenjo, C., Vicent, J., Alonso, L., and Moreno, J. (2014b). Syntretrievals simulator for hyperspectral spaceborne passive optical sensors. Application to ESA's FLEX/Sentinel-3 tandem mission. 5th International Workshop on Remote Sensing of Vegetation Fluorescence, 22 - 24 April 2014, Paris, France.

Roberts, D., Quattrochi, D., Hulley, G., Hook, S., and Green, R. (2012). Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. *Remote Sensing of Environment*, **117**, 83–101.

Rouse, J., Haas, R., Schell, J., Deering, D., and Harlan, J. (1974). Monitoring the vernal advancement of retrogradation of natural vegetation. *NASA tech. doc.*

Running, S. and Coughlan, J. (1988). A general model of forest ecosystem processes for regional applications i. hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modelling*, **42**(2), 125–154.

Sampson, P. H., Zarco-Tejada, P. J., Mohammed, G. H., Miller, J. R., and Noland, T. L. (2003). Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods. *Forest Science*, **49**(3), 381–391.

Schlemmer, M., Francis, D., Shanahan, J., and Schepers, J. (2005). Remotely measuring chlorophyll content in corn leaves with differing nitrogen levels and relative water content. *Agronomy Journal*, **97**(1), 106–112.

Schlerf, M., Atzberger, C., and Hill, J. (2005). Remote sensing of forest biophysical variables using hymap imaging spectrometer data. *Remote Sensing of Environment*, **95**(2), 177–194.

Schölkopf, B. and Smola, A. (2002). *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press Series.

Sellers, P. (1987). Canopy reflectance, photosynthesis, and transpiration, ii. the role of biophysics in the linearity of their interdependence. *Remote Sensing of Environment*, **21**(2), 143–183.

Sellers, P., Berry, J., Collatz, G., Field, C., and Hall, F. (1992). Canopy reflectance, photosynthesis, and transpiration. iii. a reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sensing of Environment*, **42**(3), 187–216.

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

Shen, X., Wu, H., and Zhu, Q. (2012). Training support vector machine through redundant data reduction. pages 25–28.

Si, Y., Schlerf, M., Zurita-Milla, R., Skidmore, A., and Wang, T. (2012). Mapping spatio-temporal variation of grassland quantity and quality using meris data and the prosail model. *Remote Sensing of Environment*, **121**, 415–425.

Smith, J. (1993). LAI inversion using backpropagation neural network trained with multiple scattering model. *IEEE Trans. Geosc. Rem. Sens.*, **31**(5), 1102–1106.

Soudani, K., Francesco, V.ois, C., le Maire, G., Le Dantec, V., and Dufrene, E.ne, E. (2006). Comparative analysis of ikonos, spot, and etm+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. *Remote Sensing of Environment*, **102**(1-2), 161–175.

Staudte, R. and Sheather, S. (1990). Robust estimation and testing, wiley, new york.

Stenberg, P., Rautiainen, M., Manninen, T., Voipio, P., and Smolander, H. (2004). Reduced simple ratio better than NDVI for estimating LAI in Finnish pine and spruce stands. *Silva Fennica*, **38**(1), 3–14.

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

Stuffler, T., Kaufmann, C., Hofer, S., Förster, K., Schreier, G., Mueller, A., Eckardt, A., Bach, H., Penné, B., Benz, U., and Haydn, R. (2007). The EnMAP hyperspectral imager-an advanced optical payload for future applications in earth observation programmes. *Acta Astronautica*, **61**(1-6), 115–120.

Suykens, J. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, **9**(3), 293–300.

Taniguchi, M. (1979). On estimation of parameters of gaussian stationary processes. *Journal of Applied Probability*, **16**, 575–591.

Teke, M., Deveci, H., Haliloglu, O., Gurbuz, S., and Sakarya, U. (2013). A short survey of hyperspectral remote sensing applications in agriculture. In *Recent Advances in Space Technologies (RAST), 2013 6th International Conference on*, pages 171–176.

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

Thenkabail, P., Smith, R., and De Pauw, E. (2002). Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing*, **68**(6), 607–621.

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

Thenkabail, P., Lyon, G., and Huete, A. (2011). Advances in hyperspectral remote sensing of vegetation and agricultural croplands. *Hyperspectral Remote Sensing of Vegetation*, pages 3–29.

Tipping, M. E. (2001). The relevance vector machine. *Journal of Machine Learning Research*, **1**, 211–244.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F. (2012). Gmes sentinel-1 mission. *Remote Sensing of Environment*, **120**(0), 9 – 24. The Sentinel Missions - New Opportunities for Science.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., and Muñoz-Marí, J. (2011). A survey of active learning algorithms for supervised remote sensingimage classification. *IEEE Journal on Selected Topics in Signal Processing*, **4**, 606–617.

Ustin, S., Roberts, D., Gamon, J., Asner, G., and Green, R. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, **54**(6), 523–534.

Van Der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., and Su, Z. (2009). An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences*, **6**(12), 3109–3129.

Vapnik, V., Golowich, S., and Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, **9**, 281–287.

Veefkind, J., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H., de Haan, J., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, **120**(0), 70–83.

Verger, A., Baret, F., and Weiss, M. (2008). Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment*, **112**(6), 2789–2803.

Verger, A., Baret, F., and Camacho, F. (2011). Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with chris/proba observations. *Remote Sensing of Environment*, **115**(2), 415–426.

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

Verhoef, W. and 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**(2), 166–182.

Verhoef, W., Jia, L., Xiao, Q., and Su, Z. (2007). Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(6), 1808–1822.

Verrelst, J., Schaepman, M., Koetz, B., and Kneubuhler, M. (2008). Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sensing of Environment*, **112**(5), 2341–2353.

Verrelst, J., Schaepman, M. E., Malenovský, Z., and Clevers, J. G. P. W. (2010). Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sensing of Environment*, **114**(3), 647–656.

Verrelst, J., Rivera, J., Alonso, L., and Moreno, J. (2011). ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. *Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop 2011,11-13 April, Edinburgh, UK*.

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

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., and Moreno, J. (2012b). Retrieval of vegetation biophysical parameters using gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5 PART 2), 1832–1843.

Verrelst, J., Romijn, E., and Kooistra, L. (2012c). Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. *Remote Sensing*, **4**(9), 2866–2889.

Verrelst, J., Alonso, L., Rivera Caicedo, J., Moreno, J., and Camps-Valls, G. (2013a). Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **6**(2), 867–874.

Verrelst, J., Rivera, J., Moreno, J., and Camps-Valls, G. (2013b). 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., Camps-Valls, G., and Moreno, J. (2013c). Recent advances in biophysical parameter retrieval methods - opportunites for sentinel-2. ESA Living Planet Symposium 2013, 09-13 September, Edinbrugh, UK.

Verrelst, J., Rivera, J., van der Tol, C., Magnani, F., Mohammed, G., and Moreno, J. (2014a). Global sensitivity Analysis of the A-SCOPE model in support of future FLEX fluorescence retrievals. 5th International Workshop on Remote Sensing of Vegetation Fluorescence, 22 - 24 April 2014, Paris, France.

Verrelst, J., Rivera, J., Leonenko, G., Alonso, L., and Moreno, J. (2014b). Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. *IEEE Transactions on Geoscience and Remote Sensing*, **52(1)**, 257–269.

Verstraete, M. and Pinty, B. (1996). Designing optimal spectral indexes for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(5), 1254–1265.

Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., and Liang, S. (2004). A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from landsat ETM+ imagery. *Remote Sensing of Environment*, **92**(4), 465–474.

Wang, S., Li, Z., and Zhang, X. (2012). Bootstrap sampling based data cleaning and maximum entropy svms for large datasets. volume 1, pages 1151–1156.

Weiss, M. and Baret, F. (19991). Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sensing of Environment*, **70**, 293–306.

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

Weiss, M., Baret, F., Garrigues, S., and Lacaze, R. (2007). LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. *Remote Sensing of Environment*, **110**(3), 317–331.

Whittaker, R. H. and Marks, P. L. (1975). Methods of assessing terrestrial productivity. *Primary Productivity of the Biosphere*, pages 55–118.

Widlowski, J.-L., Taberner, M., Pinty, B., Bruniquel-Pinel, V., Disney, M., Fernandes, R., Gastellu-Etchegorry, J.-P., Gobron, N., Kuusk, A., Lavergne, T., Leblanc, S., Lewis, P., Martin, E., Mättus, M., North, P., Qin, W., Robustelli, M., Rochdi, N., Ruiloba, R., Soler, C., Thompson, R., Verhoef, W., Verstraete, M., and Xie, D. (2007). Third Radiation Transfer Model Intercomparison (RAMI) exercise: Documenting progress in canopy reflectance models. *Journal of Geophysical Research D: Atmospheres*, **112**(9).

Widlowski, J.-L., Pinty, B., Clerici, M., Dai, Y., De Kauwe, M., De Ridder, K., Kallel, A., Kobayashi, H., Lavergne, T., Ni-Meister, W., Olchev, A., Quaife, T., Wang, S., Yang, W., Yang, Y., and Yuan, H. (2011). RAMI4PILPS: An intercomparison of formulations for the partitioning of solar radiation in land surface models. *Journal of Geophysical Research G: Biogeosciences*, **116**(2).

Wold, H. (1985). Partial least squares. *Encyclopedia of Statistical Sciences*, **6**, 581–591.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**(1-3), 37–52.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, **14**(4), 1261–1295.

Ye, X., Sakai, K., Manago, M., Asada, S.-I., and Sasao, A. (2007). Prediction of citrus yield from airborne hyperspectral imagery. *Precision Agriculture*, **8**(3), 111–125.

Zarco-Tejada, P., Miller, J., Noland, T., Mohammed, G., and Sampson, P. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(7), 1491–1507.

Zarco-Tejada, P., Rueda, C., and Ustin, S. (2003). Water content estimation in vegetation with modis reflectance data and model inversion methods. *Remote Sensing of Environment*, **85**(1), 109–124.

Zarco-Tejada, P., Miller, J., Pedrós, R., Verhoef, W., and Berger, M. (2006). Fluormodgui v3.0: A graphic user interface for the spectral simulation of leaf and canopy chlorophyll fluorescence. *Computers and Geosciences*, **32**(5), 577–591.

Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., González, M. R., and Frutos, A. D. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, **99**(3), 271–287.

Zhang, J., Blackmer, A., Ellsworth, J., and Koehler, K. (2008). Sensitivity of chlorophyll meters for diagnosing nitrogen deficiencies of corn in production agriculture. *Agronomy Journal*, **100**(3), 543–550.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F., and Moore III, B. (2005). Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using modis data and a radiative transfer model. *Remote Sensing of Environment*, **99**(3), 357–371.

Appendices: Published Chapters 3 – 6

Submitted to *Remote Sens.*. Pages 1 - 24.

Article

On the semi-automatic retrieval of biophysical parameters based on spectral index optimization

Juan Pablo Rivera [1], Jochem Verrelst [1]*, Jesús Delegido [1], Frank Veroustraete [2] and José Moreno [1]

[1] Image Processing Laboratory (IPL), Parc Científic, Universitat de València, 46980 Paterna, València, Spain

[2] Department of Bioscience Engineering, Faculty of Sciences, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

* Author to whom correspondence should be addressed; jochem.verrelst@uv.es, Tel: +34 96 354 40 67, Fax: +34 96 354 32 61.

Version February 20, 2014 submitted to RemoteSensing. *Typeset by LaTeX using class file mdpi.cls*

Abstract: Regression models based on spectral indices are typically empirical formulae enabling the mapping of biophysical parameters derived from Earth Observation (EO) data. Due to its empirical nature, it remains nevertheless uncertain to what extent a selected regression model is the most appropriate one, until all band combinations and curve fitting functions are assessed. This paper describes the application of a Spectral Index (SI) assessment toolbox in the Automated Radiative Transfer Models Operator (ARTMO) package. ARTMO enables semi-automatic retrieval and mapping of biophysical parameters from optical remote sensing observations. The SI toolbox enables the assessment of biophysical parameter retrieval accuracy of established as well as new and generic SIs. For instance, based on the SI formulation used, all possible band combinations of formulations with up to ten bands can be defined and evaluated. Several options are available in the SI assessment: calibration/validation data partitioning, the addition of noise and the definition of curve fitting models. To illustrate its functioning, all two-band combinations according to simple ratio (SR) and normalized difference (ND) formulations as well as various fitting functions (linear, exponential, power, logarithmic, polynomial) have been assessed. HyMap imaging spectrometer [430-2490 nm] data obtained during the SPARC campaign in Barrax, Spain, have been used to extract leaf area index (LAI) and leaf chlorophyll content (LCC) estimates. For both SR and ND formulations the most sensitive regions have been identified for two-band combinations of green (539-570 nm) with longwave SWIR (2421-2453 nm) for LAI (r^2: 0.83) and far-red (692 nm) with NIR (1340 nm) or shortwave SWIR (166-1686

nm) for LCC (r^2: 0.93). Polynomial, logarithmic and linear fitting functions led to similar best correlations, though spatial differences emerged when applying the functions to HyMap imagery. We suggest that a systematic SI assessment is a strong requirement in the quality assurance approach for accurate biophysical parameter retrieval.

Keywords: Spectral indices; Empirical regression models; Biophysical parameter retrieval; GUI toolbox; leaf area index; leaf chlorophyll content; HyMap

1. Introduction

Quantitative and spatially-explicit retrieval of vegetation biophysical parameters is a requirement in a variety of ecological and agricultural applications. Earth observing (EO) satellites, endowed with a high temporal resolution, enable the retrieval and hence monitoring of plant biophysical parameters [1–3]. With the forthcoming super-spectral Copernicus' Sentinel-2 [4] and Sentinel-3 missions [5] as well as the planned EnMAP [6], HyspIRI [7] and PRISMA [8] imaging spectrometer missions, an unprecedented data stream becomes available. This data influx requires processing techniques, which are reproducible, accurate and fast, when the retrieval of information on plant growth and health status is envisaged. Typically, the retrieval of biophysical parameters using EO data implies the use of a model [9], which can be either statistical or physical in nature. A statistical regression model enables the linkage of EO data with biophysical parameters of interest [10] by making use of (*in situ*) calibration data. On the other hand, physically-based approaches allow the retrieval of the biophysical parameters by inversion of a radiative transfer model (RTM) against EO data [11,12].

Although inversion of a RTM is generally applicable for a wide range of land cover types and sensor configurations [13], the approach is more often than not computationally demanding. It usually requires quite some auxiliary information to enable the parameterization of the physical model and the description of the boundary conditions for which the model is valid [14,15]. This information may not always be available. Moreover, by introducing input parameter uncertainties, the likelihood increases that model inversion may lead to multiple instead of singular solutions in the soil–vegetation–atmosphere matrices (unified theorem of Hadamard well-posedness). In that case, extra steps are required to overcome the ill-posed problem [16].

Conversely, statistical models are more easily implementable for parameter retrieval and mapping applications. The principle basically is to correlate mathematical combinations of measured reflectances for different wavelength ranges or broad spectral bands with biophysical vegetation parameters of interest (e.g., leaf area index (LAI), leaf chlorophyll content (LCC), fractional vegetation cover) using a fitting function. This procedure can be considered as an empirical spectral index (SI) modelling approach. Over the past four decades, a large number of SI models has been developed for the retrieval of biophysical parameters [9,17,18]. The majority of these SIs and their relationship with biophysical parameters of interest has been developed based on experimental work [19–21].

While being successful for many application cases, the empirical approach suffers from drawbacks as well. Typically, a limited portability to different measurement conditions or sensor characteristics can be noted [22]. Moreover, the approach is sensitive to perturbing factors such as atmospheric conditions, canopy characteristics and differences in viewing or solar position geometry [23,24]. In an attempt to generate more generic relationships, RTM-based dataset simulations have been used to assess for the most sensitive SIs [18,25–29]. A spectral index can quite often be used as an estimator for a biophysical parameter using a fitting function through the data; usually by simple linear regression, but also exponential, power, logarithmic and polynomial fitting functions, among others, are commonly applied [22,30,31]. In statistics, this approach can be categorized as a parametric regression modeling approach. This type of model is determined by the introduction of various boundary conditions to obtain formulations linking spectral reflectance with the biophysical parameter of interest. In fact, boundary conditions can be defined at three levels of SI regression modeling:

- *Band selection*: Typically, most SIs are mathematical formulations consisting of two or three sensor spectral bands (B). How then, do we evaluate with a high enough scrutiny, whether the most sensitive spectral bands - with respect to biophysical parameter retrieval - have been selected? This question is especially relevant in view of the high number of bands associated with imaging spectrometry [32].

- *SI formulation*: Typically, the normalized difference (ND) formulation is applied, i.e., (B_2-B_1)/(B_2+B_1). But again, how do we assess, whether the applied ND formulation is the most accurate one with respect to biophysical parameter retrieval? Even given high spectral resolution multi- or hyperspectral reflectance data, there is no reason to assume that a two-band SI formulation leads to the most accurate empirical relationship [33].

- *Fitting function*: A regression model is typically reduced to a linear fitting exercise, directly or indirectly by prior transformation to linearity. Also here, the question is, whether the regression function selected is the most accurate one? Typically, saturation effects are common for dense canopies [14,15].

Accordingly, it is uncertain, whether an established SI - many have been published in the remote sensing literature - will lead to the most accurate biophysical parameter retrieval, particularly when applied to hyperspectral EO data. To ensure that the most sensitive SI regression model is selected for a specific application, a systematic as well as consistent assessment of an exhaustive set of possible band combinations, SI formulations and curve fitting procedures may be required prior to the selection of an SI model for the retrieval of biophysical parameters from (hyperspectral) EO imagery.

Evidently, this is not a trivial task. In view of forthcoming imaging spectrometers, and to obtain an optimized SI model for biophysical parameter retrieval, the assessment of a large number of band combinations is required. To this end, the requirement for a graphical user interface (GUI) toolbox, assisting EO users in the tedious job of evaluating large sets of SI models, is not a luxury. This observation brings us to the following objectives: (1) To develop an SI assessment toolbox for the semi-automatic retrieval of biophysical parameters; (2) To enable a robust assessment procedure to estimate the impact of different index formulations, spectral band settings and curve fitting functions

with respect to their capability of biophysical parameter retrieval; and (3) To apply the most accurate SI regression model, to extract biophysical parameters from EO imagery.

Since spectral indices and their associated regression models can either be developed based on experimental or simulated data, both options should be present in an SI assessment module. To this end, the SI toolbox has been developed within the ARTMO (Automated Radiative Transfer Models Operator) modeling environment. The experimental dataset used for testing has been acquired from ESA′s SPARC campaign (Barrax, Spain).

In the following sections, the most recent version of the ARTMO modeling environment will be briefly described. Subsequently, an introduction to the most important components of the SI assessment toolbox will follow. The dataset used is described and finally, an assessment of selected SIs will be presented. A discussion section is devoted to newly developed spectral indices and a conclusion section finalizes this paper.

2. ARTMO software package

ARTMO, developed and running in Matlab (2011a or higher), offers multiple leaf and canopy radiative transfer models (RTMs) accessible in a GUI environment. Dedicated retrieval methods, required for the semi-automatic retrieval of biophysical parameters, are provided. Examples of ARTMO run options are:

- Selection of various forward and invertible leaf and canopy RTMs of a low to high complexity (PROSPECT-4, PROSPECT-5, 4SAIL, FLIGHT);
- The choice to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems;
- The option to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400 to 2500 nm). Look-up tables (LUT) can be generated, which are stored in a relational SQL database management system (MySQL, version 5.5 or higher; local installment required);
- Finally, various retrieval scenarios can be selected and run using EO reflectance datasets.

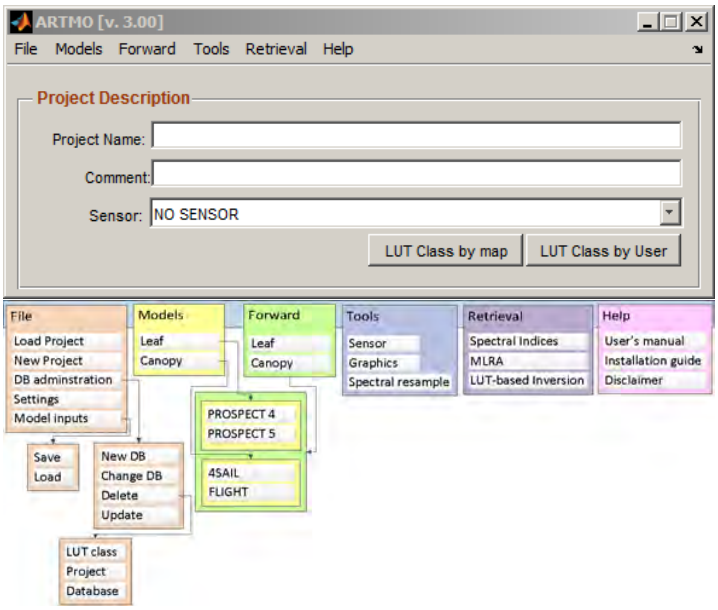
Figure 1 illustrates the ARTMO v3 GUI main window and a systematic overview of the modules. In the GUIs main window, a new project can be defined, a sensor chosen and comments added. All processing modules are accessible by drop-down menus from the top bar. Already in the first version of ARTMO, LUT-based inversion applications were available [12,13,34]. ARTMO v3 is formally presented in this paper. The software package, as well its manuals, installation guides and tutorials, are freely downloadable at: http://ipl.uv.es/artmo/. Its most important novelties are briefly listed below:

- ARTMO v3 is designed modularly. Its modular architecture offers the possibility for easy addition (or removal) of components, such as RTM models and post-processing modules.

- The MySQL database is organized in such a way that it supports the modular architecture of ARTMO v3. This avoids redundancy and increases the processing speed. For instance, all spectral datasets are stored as binary objects.

- New retrieval toolboxes are incorporated. They are based on parametric and non-parametric regression as well as physically-based inversion using a LUT. This has led to the development of a: (1) 'Spectral Indices assessment toolbox'; (2) 'Machine Learning Regression Algorithm toolbox' [35]; and, (3) 'LUT-based inversion toolbox' [12].

Figure 1. Screenshot of ARTMO's main window and schematic overview of its drop-down menu.

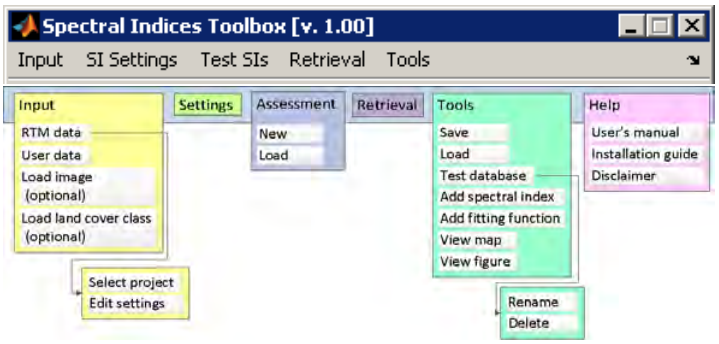


2.1. ARTMO Spectral Indices (SI) toolbox

This paper introduces a 'Spectral Indices assessment toolbox'. Its general architecture is outlined in Figure 2. The toolbox enables the loading and application of existing or user-defined spectral indices. In short, the toolbox encompasses the following utilities. For each new index, a spectral region for each waveband is specified. Every single band of a sensor that falls within the specified spectral region will then be included in the retrieval capacity assessment. Different types of curve fitting functions have been implemented (e.g., linear, power, exponential, logarithmic, polynomial functions), while new functions can easily be added by the user. Input data originating from RTM runs can be used or user-defined input data can be imported (a plain text file), typically a dataset acquired during a field or flight campaign (i.e., biophysical parameters and associated surface reflectance). Input data can then be partitioned for calibration and validation of the SI–biophysical parameter relationships. Options are available to merge or partition different input datasets; e.g. calibration on the basis of RTM data and validation on the basis of user acquired field data. A land cover map in ENVI format (Exelis Inc.) can be loaded, which

allows that for each land cover class distinct SI optimization strategies can be defined (e.g., vegetation indices for vegetated surfaces, water indices for water bodies, etc.). When validation data are available, a large diversity of spectral band combinations as well as regression functions, can be assessed using the validation dataset and statistical analysis results (c.f., section II-D). The most accurate strategy can then be selected, loaded and applied to retrieve biophysical parameters from EO imagery (in ENVI format). Important steps in the outlined procedures are presented in the following sections, they are: (1) *Add spectral index*; (2) *SI settings*; (3) *Calibration/validation assessment*; and, (4) *Retrieval*.

Figure 2. Screenshot of the ARTMO v3 'Spectral Indices assessment toolbox' and schematic overview of its drop-down menu.



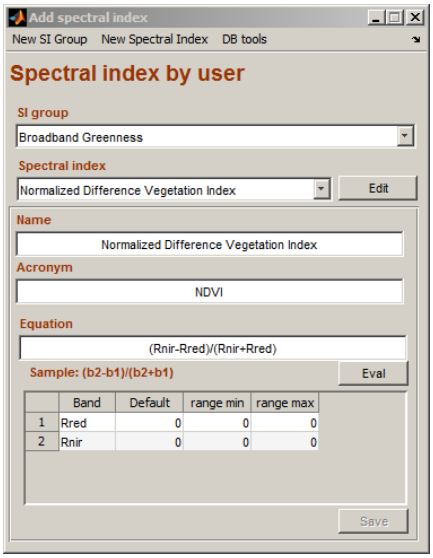
2.2. Add spectral index

The 'Add spectral index' window (Figure 3) in the 'Tools' menu allows for adding a new index to a list with pre-defined SIs. Spectral indices can be created manually using the GUI or they may be imported. SIs are organized in 'SI Groups', according to their similarity in definition, e.g., broadband, narrowband or according to their sensitivity with respect to certain absorption features, e.g. by pigments or water. For instance, the Broadband Greenness SI group consists of the conventional simple ratio (SR: B_2/B_1) and ND (B_2-B_2/B_2+B_1) formulations. A new SI group can also be created, in which multiple SI formulations can be incorporated according to specific input requirements; name, acronym and SI equation definition. When this step is completed, the transfer equation will be assessed for its validity and spectral bands are identified. For each spectral band, a default wavelength or a minimum and maximum of a wavelength range must be specified. During the subsequent assessment step, all bands within the prespecified spectral region will be automatically evaluated. When kept undefined, every single band of a given sensor will be evaluated according to the specified SI formulation.

2.3. SI settings

In order to activate the possibility of defining SI settings, input data are required. Insertion of input data is performed with the 'Input' window. Input data can either consist of a RTM simulated dataset or it can be a field campaign dataset in combination with an EO dataset acquired during the field campaign. The 'Input' GUI helps to perform the data selection steps and checks, if data input is correct (GUI not shown for sake of brevity; consult SI toolbox manual at http://ipl.uv.es/artmo/). Subsequently, the SI

Figure 3. GUI section used to add a new SI formulation.

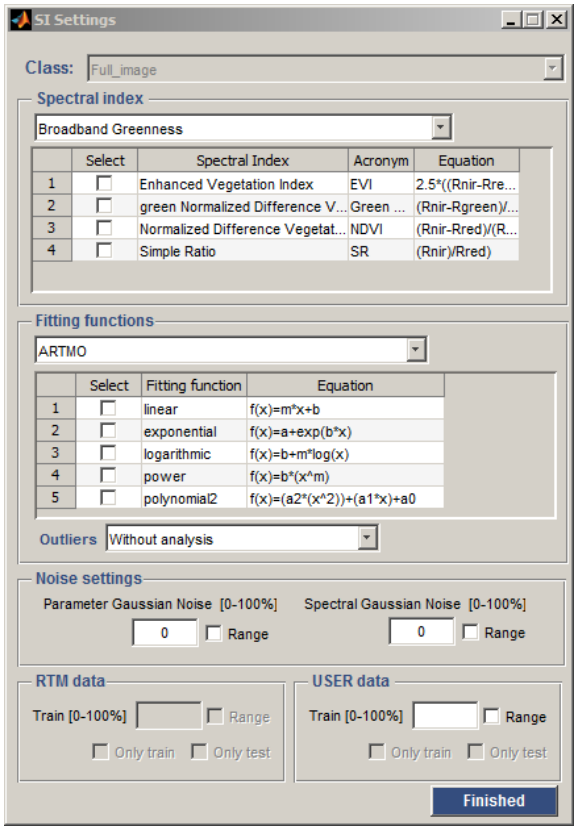


settings configure the parametric SI models according to various specifications, which are given in the 'SI settings' GUI (see Figure 4). First, if multiple land cover types have been pre-defined (using the GUI's 'Load Image and Class' windows), retrieval strategies can be configured specifically for each land cover class. Second, an SI group must be selected so that its encompassing SIs are subsequently listed. From this list, multiple SIs can be selected for assessment purposes. Third, several pre-defined curve fitting functions can be selected for assessment. Fourth, the option to add Gaussian noise is provided, which is useful for the assessment of environmental as well as instrumental uncertainty estimations. The noise function can be applied both for the parameters to be retrieved, as well as on the reflectance spectra used in the assessment. A range of noise levels can be specified; hence multiple noise scenarios can be assessed with respect to their impact on retrieval accuracy. Fifth, the calibration/validation (cal/val) data partitioning can be defined according to proportional assignment of data to either calibration or validation (i.e. split-sample approach). Both, field and RTM datasets can be combined by assigning a proportion of each either to calibration or validation. Multiple cal/val partitions can be applied in the assessment.

2.4. Calibration/validation assessment

When input data were correctly imported in ARTMO, cal/val data partitioning has been defined and the configuration of SI settings has been successful, the SI models can be assessed with respect to retrieval accuracy. This analysis starts by assigning a name to a new test dataset in the 'Test SI' menu. The specified retrieval strategies for the SI formulations and spectral band ranges are then analyzed one-by-one. The assessment results are saved by default in the MySQL database. This way, a large number of SIs can be systematically assessed, and results can be stored and queried. An overview table is then assembled, in which assessment results with the highest accuracy scores are dynamically presented; e.g. per land cover type, per parameter, and per cal/val dataset (see Figure 5). Assessment is then performed. First, for each SI, all spectral band combinations are correlated against the calibration dataset (results not shown). Second, each SI model obtained according to the chosen fitting function

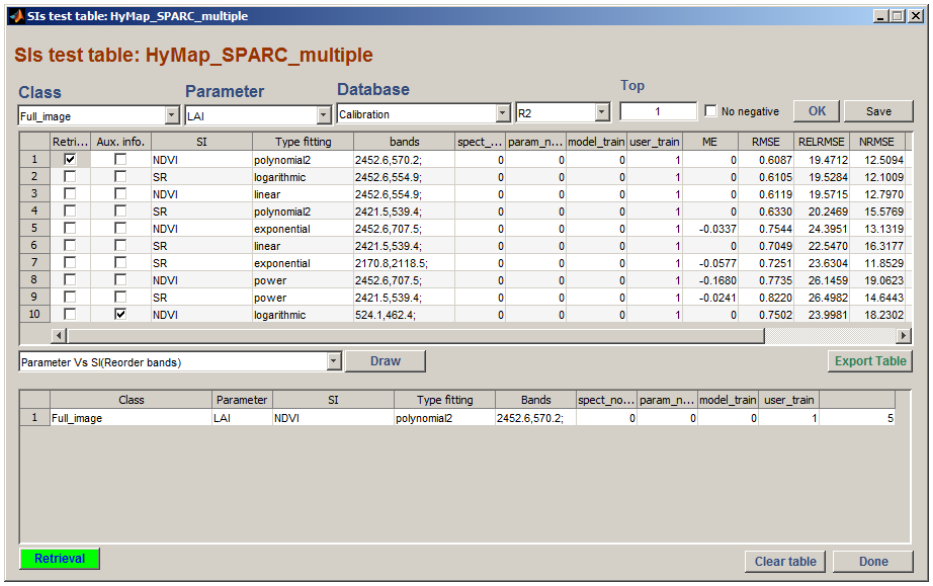
Figure 4. SI settings window.



is applied to the estimation of a targeted biophysical parameter. When the dataset has previously been partitioned into a calibration and validation set, the assessments are generated twice; one time for the calibration dataset and a second time for the validation dataset. The obtained SI models are evaluated with multiple goodness-of-fit measures like the r^2, the root mean squared error (RMSE), the normalized RMSE (NRMSE [%] = RMSE / (value range of parameters as measured in the field) *100), the mean absolute error (MAE) and the mean error (ME). All measures indicate the degree of association between predicted and estimated values of the same parameter and give thus an indication of prediction efficiency. Richter *et al.* [36] recommended the combined set of r^2, RMSE and NRMSE, amongst others, for comprehensively quantification the performance of vegetation biophysical models. For a single SI model selected, an option has been foreseen in the GUI, to generate a scatterplot of the retrievals in function of the cal/val measurements (c.f. section 4). Additionally, by selecting a statical measure, a 2D correlation matrix can be assembled along 2 spectral dimensions, enabling the visualization of the most sensitive 2-band combinations. Finally, for each retrievable biophysical parameter, an SI model can be selected, based on its accuracy assessment. The selected SI model, e.g. the best performing one, can be accessed through the 'Retrieval' GUI window and may be applied using an EO image for spatially distributed biophysical parameter mapping.

2.5. Retrieval

Figure 5. SI's calibration/validation assessment table.



The 'Retrieval' window enables to run an assessed SI model, or to directly configure an SI model and apply it to map retrieved parameter values (Figure 6). The latter option can be useful when dealing with a pre-defined SI model (e.g. when an SI is applied as published in the EO literature) or when cal/val datasets are not available. Hence, the user can select a specific land cover type (if available), a biophysical parameter to be retrieved, an SI group and an SI, a regression function and the regression function parameters. The user will be able to select one or multiple remote sensing images for which the evaluated SI model will be applied. Biophysical parameter maps are subsequently stored on disc in ENVI format.

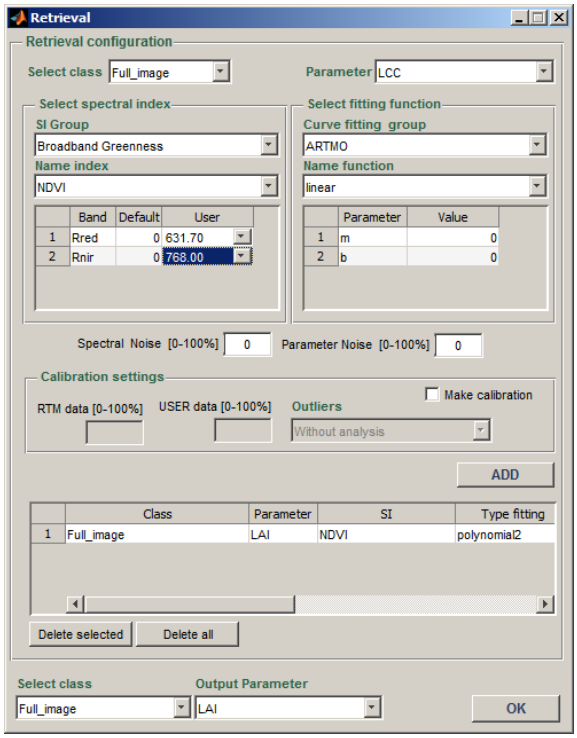
3. Assessment and mapping applications

As outlined in the previous section, the SI assessment toolbox is applied to assess the performance of multiple SI regression models with respect to estimation accuracy. Data used are presented first, followed by the experimental setup. Goodness-of-fit results for calibrated SI models are presented subsequently and finally the retrieved spatial maps of biophysical parameters are shown.

3.1. Used data

A field dataset, encompassing different crop types, growing phases, canopy geometries and soil conditions has been acquired during the SPectra bARrax Campaign (SPARC). The SPARC-2003 campaign took place from 12 to 14 July in Barrax, La Mancha, Spain (coordinates 30°3'N, 28°6'W, 700 m altitude A.S.L.). Biophysical parameters have been measured within a total of 108 Elementary Sampling Units (ESUs) for different crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat). An ESU refers to a plot, which is sized compatible with pixel dimensions of about 20m $\times$ 20m. Leaf Chlorophyll Content (LCC) was determined by measuring about 50 leaf samples in each ESU with a calibrated CCM-200 Chlorophyll Content Meter [37]. Green LAI was derived from

Figure 6. SI's Retrieval window.



canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned a LAI value, obtained as a statistical mean of 24 measurements (8 data readings $\times$ 3 replica) with standard errors ranging from 5 to 10% [38]. Strictly speaking, due to the assumption of a random distribution of foliage, the impact of clumping has been assessed only partially using the LiCor and the corresponding software. Hence, effective LAI is given as an output parameter. For all ESUs, LAI ranges from 0.4 to 6.2 (m^2 single sided leaf surface)/(m^2 ground surface) while LCC ranges from 2 to 51 (μg chlorophyll)/(cm^2 leaf area). Regarding LCC, it should be noted that some clumping in the field data was noted, with many data points between 40-50 and around 20, but no data points between 25-35 $\mu g/cm^2$. No additional bare soil samples were added.

During the campaign, airborne hyperspectral HyMap flightlines have been acquired for the study site, during the month of July 2003. HyMap flew with a configuration of 125 contiguous spectral bands, spectrally positioned between 430 and 2490 nm. Spectral bandwidth varied between 11 and 21 nm. The pixel size during overpass was 5 m. The flight-lines were corrected for radiometric and atmospheric effects according to the procedure that has been implemented in the BEAM CHRIS-Box

3.2. Experimental setup

The SI assessment toolbox has been applied to evaluate the retrieval accuracies for different SI formulations, band settings, and fitting functions using the SPARC calibration dataset. Only the two-band results for the most common SI formulations are illustrated, being SR and ND SI combination types. Correlating the possible HyMap two-band SIs according to these combination types and by using five fitting functions (linear, exponential, power, logarithmic and polynomial of the second degree), the

SI assessment toolbox did finally automatically assess 155000 (125x124x5x2) SI regression functions for each biophysical parameter. This typical assessment run did not take longer than a few minutes on a 32 bit PC (Windows 7 Intel(R) Core(TM) 2 Quad CPU, 2.4 GHz, 3.00 GB RAM processor).

4. Results & Discussion

4.1. Optimized SI's

For each SI type, Table 1 provides goodness-of-fit statistics (RMSE, NRMSE and r^2) for the best performing calibrated SI models together with its wavelength specifications and regression equation, ranked according to r^2. The following observations can be made:

- With regard to SI formulations, no strong evidence has been found that the widely used ND formulation outperforms the less popular SR formulation. Though, it is also recognized that the main argument for using ND types of indices is not the improved correlation performance in comparison to SRs, but rather the (at least to some extent) improved comparability of different observation times/dates and the possible reduction of effects of varying illumination intensity (shades etc.), due to the inherent normalization of the value range [39]. The ND outperformed the SR for both LAI and LCC only when a second order polynomial is applied as a fitting function. However, when a conventional linear regression is applied, ND performs similar (LCC) or superior (LAI) to SR.

- While the polynomial fit outperforms to some extent the linear regression fitting scenario, the linear fit model is nonetheless quite accurate as well. For instance, the best linear regression fit outperforms the more sophisticated exponential and power curve fitting approaches significantly for both, LAI and LCC.

- For the majority of the most accurate regression functions, about the same spectral band locations were assessed as being optimal for biophysical parameter retrieval. For LAI, two-band formulations are optimal at 570, 555, 539 and 707 nm with the second band at 2453 or 2421 nm. For LCC two-band locations at 692 nm and 632 nm with the second band at 1661-1698 nm or 1200-1340 nm performed best.

It is noteworthy that, for a majority of the most accurate SI models sensitive with respect to LCC, the wavelength location of 692 nm is ranked as the best performing one. This is consistent with earlier studies (e.g., [24,40]), demonstrating that the spectral interval around 700 nm is less prone to saturation than the location at 660 nm (which coincides with the location of the chlorophyll pigments main absorption peak). Therefore, the far-red at 692 nm ensures a good sensitivity when detecting moderate and high LCC values.

Although Table 1 suggests that the most optimal fitting functions lead to similar performances in terms of r^2, a closer look may be necessary to assess their impact on the efficiency in mapping applications. Figure 7 provides scatterplots of the best performing SI models; the distinct behavior of each fitting function can be observed. For instance, exponential and power function fits can lead to spurious

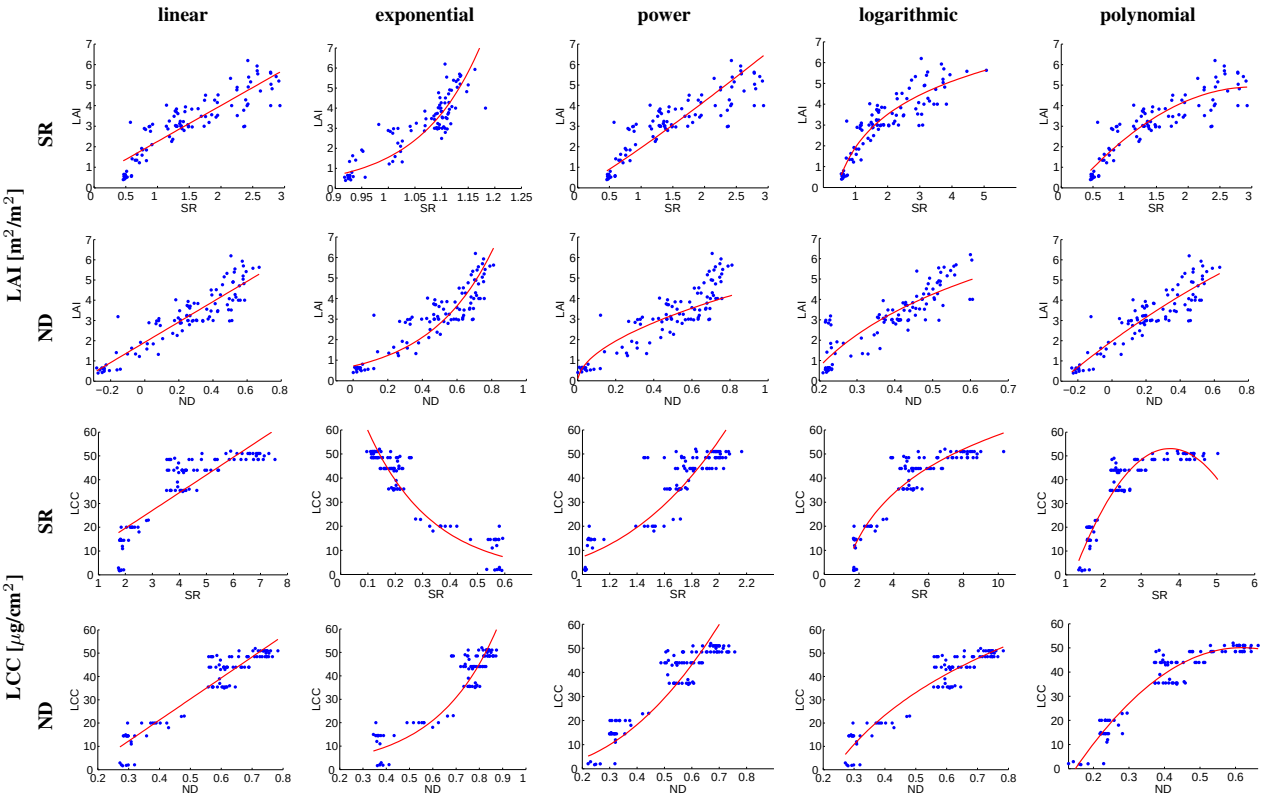
Table 1. Goodness-of-fit measures (RMSE, NRMSE, r^2) for the best calibrated SI models per SI type and their corresponding used bands and fitting function for LAI and LCC retrieval. In this table, the SI types are sorted according to a decreasing r^2. B_1: spectral band 1, B_2: spectral band 2.

Index type	Fitting function	B_1 (nm)	B_2 (nm)	SI model	abs. RMSE	NRMSE (%)	r^2
LAI:							
ND	polynomial	570	2453	LAI= -1.366ND^2 + 6.154ND + 1.976	0.61	12.51	0.83
SR	logarithmic	555	2453	LAI= 2.291 log(SR) + 1.924	0.61	12.10	0.83
ND	linear	555	2453	LAI= 5.023ND + 1.922	0.61	12.80	0.83
SR	polynomial	539	2421	LAI= -0.621SR^2 + 3.754SR -0.760	0.63	15.58	0.82
SR	linear	539	2421	LAI= 1.751SR + 0.504	0.70	16.32	0.77
SR	exponential	2118	2171	LAI= 0.000238e$^{8.783SR}$	0.72	11.85	0.77
ND	exponential	707	2453	LAI= 0.712e$^{2.721ND}$	0.75	13.13	0.77
ND	power	707	2453	LAI= 4.670ND$^{0.548}$	0.77	19.06	0.77
SR	power	539	2421	LAI= 1.951SR$^{1.112}$	0.82	14.64	0.76
ND	logarithmic	462	524	LAI= 3.888 log(ND) + 6.947	0.85	18.23	0.74
LCC:							
ND	polynomial	692	1686	LCC= -227.499ND^2 + 281.410ND - 36.924	4.21	7.75	0.93
SR	polynomial	692	1661	LCC= -8.114ND^2 + 60.997ND - 61.553	4.30	7.64	0.92
ND	logarithmic	692	1340	LCC= 43.578 log(ND) + 63.272	4.70	10.18	0.91
SR	linear	692	1340	LCC= 7.455SR + 4.707	4.70	20.24	0.91
ND	linear	692	1340	LCC= 90.290ND -14.808	4.21	10.75	0.90
SR	logarithmic	692	1200	LCC= 26.447 log(ND) - 2.990	5.53	11.54	0.87
SR	exponential	692	1215	LCC= 91.164e$^{-4.235SR}$	6.00	11.25	0.87
ND	power	632	1698	LCC= 127.581ND$^{2.119}$	28.06	365.5	0.87
ND	exponential	632	1257	LCC= 2.110e$^{3.827ND}$	6.64	12.80	0.84
SR	power	707	723	LCC= 7.281SR$^{2.938}$	8.28	13.19	0.76

parameter value retrievals at high SI values (or at low SI values in case of the SR-LCC combination). The most accurate fitting function, i.e. the second order polynomial, is probably the most difficult one to interpret. This is illustrated for the LCC-SR case where the peak of the polynomial fit goes through the LCC data clumping between 40-50 μg/. It implies that a SR above 4 leads to a lower LCC estimate. This type of erratic fits indicates that polynomial fitting - despite the high r^2 and low RMSE - should not be recommended for biophysical parameter retrieval.

Although dating from the times where the campaigns mostly were limited to the silicon detector range (400 - 1000 nm), the saturation effect is often reported and is considered as one of the major critical points when using a red-NIR formulation (e.g. NDVI) for LAI retrieval [41,42]. A noteworthy observation is that here LAI hardly elicits saturation for a linear model fit. Presented results suggest that the use of a band in the SWIR range (2453 nm) causes the saturation effect to be close to negligible. Hence, such a linear regression should be preferred above using the NIR, because saturation violates the assumptions of regression residuals around zero in order for the results to be valid [43]. Also [44,45] found that vegetation indices (VIs) based on bands in the SWIR range tend to be almost linearly related to LAI without showing saturation even at high LAI values (LAI=6). However, it should also be remarked that wavelengths at the end of HyMap sensor' SWIR range (i.e. beyond 2400 nm) face a relatively low signal-to-noise ratio (SNR), i.e. below 400 [46]. Results in this region may therefore be biased by noise influence.

Figure 7. Fitting functions (red lines) depicting the regression of LAI [m^2/m^2] and LCC [$\mu g/cm^2$] in function of SR and ND formulations as they could be derived from the calibration dataset using optimized band combinations.



4.2. Correlation matrices

To fully take account of the prediction efficiency of the spectral domain, figure 8 illustrates all two-band SR and ND combinations according to r^2 correlation between measured and estimated values for the LAI and LCC biophysical parameters for different fitting functions. The correlation matrices for B_1 (430-2490 nm) versus B_2 (430-2490 nm) clearly elicit the following spectral regions: the visible region (450-700 nm), the red-edge region (700-750 nm), the NIR region (750-1340 nm), the shortwave SWIR region (1470-1800 nm) and the longwave SWIR region (1970-2453 nm). Based on the 2D correlation plots in figure 8, the spectral regions leading to the most accurate indices to retrieve LAI and LCC can be unequivocally identified. Based on tables 1 and 2, the following observations can be made for two-band (B_1, B_2) SIs:

- SR and ND formulations lead to the same spectral regions exhibiting strong correlations between the LAI / LCC biophysical parameters and the two-band SIs. For instance, similar patterns for B_2 in the visible spectral region can be observed. However, the ND power and logarithmic functions fail as good fitting functions for several band combinations. Specifically, they provide lower regression accuracies for B_1 (740-1000 nm) and B_2 (1400-2200 nm).

- Similar patterns with strong and low correlation values are elicited across the different curve fitting functions assessed. This suggests that the major impact on retrieval accuracy in EO does not originate from the chosen curve fitting but rather from the spectral dimension. Though, local differences do occur. These local differences are particularly pronounced along exponential, power and logarithmic fittings, characterized by quite low correlation values. Hence, the following sections are limited to the discussion of linear and polynomial curve fits.

- For LAI, Figure 8 shows that the most important spectral region for retrieval is the 539-707 nm region for the B_1band, whereas B_2 is located in the longwave SWIR. Notice the hourglass shape in the SWIR region. A second spectral region of maximal r^2 for $B1$ is the 1300-1500 nm region as well as the region at 1400-1800 nm. These regions are spectrally quite broad, meaning that the spectral indices do not require very narrow and hence precisely positioned spectral bands, at least for the samples that were used in this study for the generation of the empirical relationships.

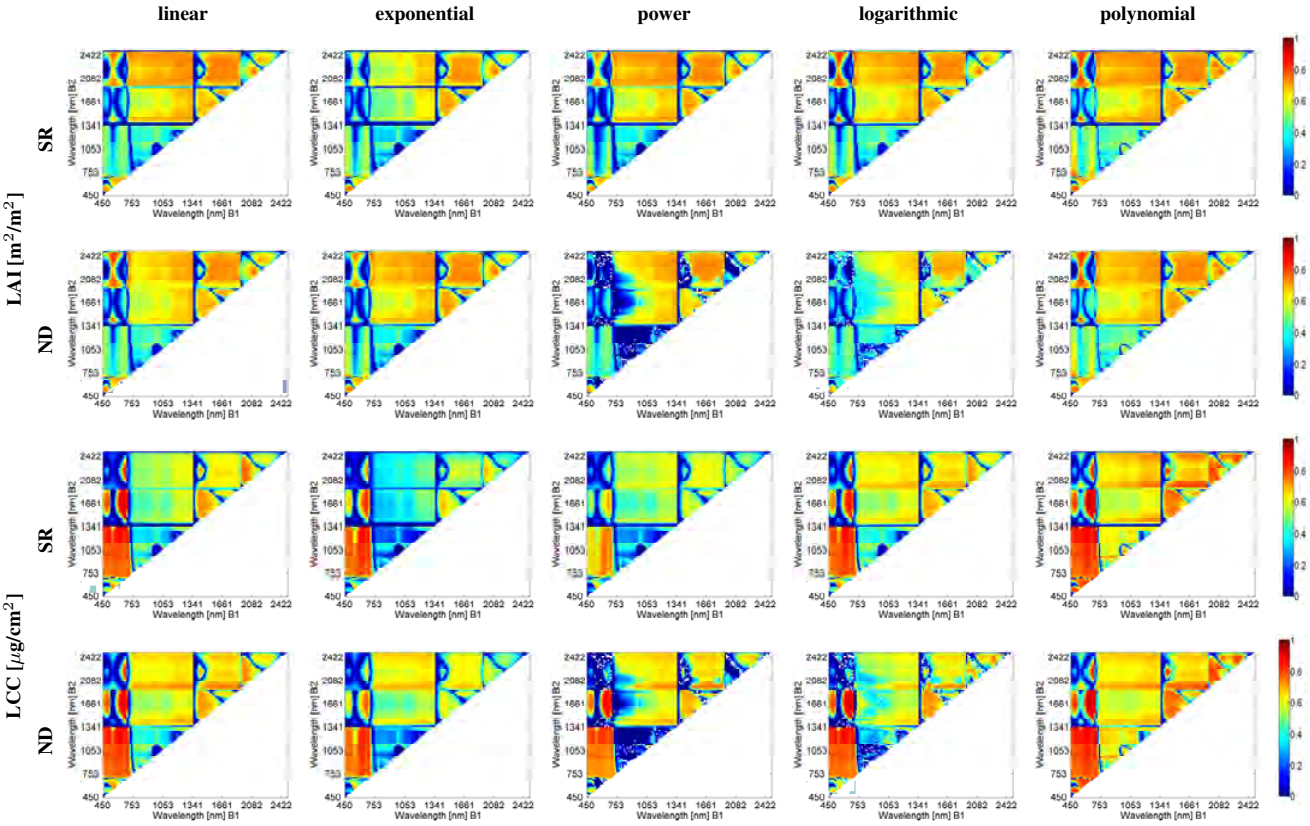
- For LCC, Figure 8 shows that the most important spectral region for retrieval is the visible region, particularly the red region until the red edge. B_2 is spectrally located in the shortwave SWIR. Another sensitive zone, having the same B_1 (around 692 nm) as in the first case, has its second band B_2 in the NIR range as well as in the longwave SWIR. Note that these spectral areas are covered by broad spectral bands.

Considering only the linear and polynomial fitting functions, it is of interest that the regions that were found to be optimal for LAI retrieval in the longwave SWIR region (notice the hourglass shape centered around the green; 550-570 nm) for B_2 oppositely lead to erroneous LCC retrievals. This spectral region can hence be applied for LAI retrieval without being influenced by its sensitivity for LCC. For LCC on the other hand, the most optimal B_1 spectral regions are the NIR and shortwave SWIR in combination with red bands for B_2. Hence, a clearcut distinction can be observed for the sensitive spectrally important regions for the retrieval of LAI versus LCC. Retrieval confusion between LAI and LCC is thereby avoided. The correlation matrices additionally suggest that the SWIR region is preferred above the NIR region, ensuring independent retrievals for LAI and LCC. For instance, the conventional NDVI spectral bands (i.e., B_1 = 680 nm, B_2 = 800 nm) do not appear to be the most optimal ones for LAI nor LCC retrieval.

4.3. LAI & LCC mapping

The final step presented in this paper involves LAI and LCC mapping. ARTMO's SI assessment toolbox helps the user to select a SI model (e.g. the most accurate one) and to apply it using EO imagery. The impact of the SI model type on biophysical parameter mapping thus can be compared. For this purpose, it was opted to map LAI and LCC based on the available HyMap data for the most accurate SI model only, being a linear, logarithmic and polynomial of the second order function. Exponential and power functions have not been considered for mapping, not only because of their considerably less performing calibration results, but also because these functions do not consistently lead to maps conform with ground measurements. To allow for a comprehensive visual comparison between retrievals and ground samples, LAI maps are color scaled for values between 0 and 6 m^2/m^2. Identically, LCC maps

Figure 8. 2D r^2 correlation matrices calculated from measured vs. estimated values for the calibration results of the LAI [m^2/m^2] and LCC [$\mu g/cm^2$] retrievals according to SR and ND formulations and for different fitting functions. The color bars on the right of this figure give the values of the plotted r^2 in the different graphs.

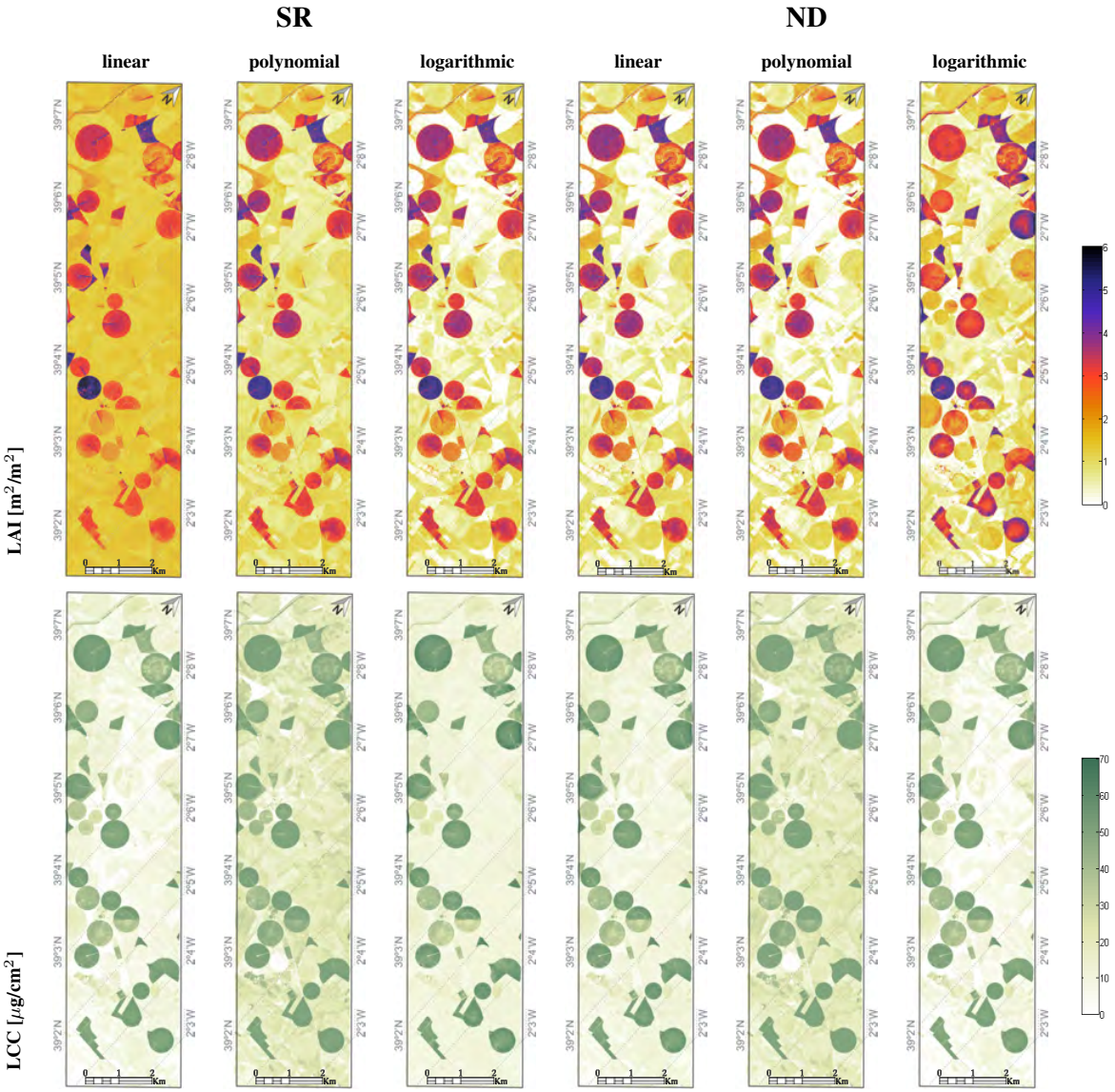


are color scaled for values between 0 and 60 $\mu g/cm^2$. These maps, as presented in Figure 9, lead to the following observations:

- For LAI, both, the ND polynomial and ND linear regression models, lead to the same calibration result (RMSE=0.61; r^2=0.83) and accordingly to precisely identical LAI maps. This example provides confidence that a linear regression is able deliver adequate LAI maps, when appropriate spectral bands are selected; i.e., green spectral region (B_1: 539-570 nm) in combination with the longwave SWIR (B_2: 1970-2453 nm). Rapid saturation of the LAI in function of the SI is thereby avoided. The map shows areas with a high LAI quite clearly (irrigated circular fields). Areas with a low LAI - senescent crops to bare soil surfaces - elicit a zero value LAI (white color code).

- Conversely, when using SR for LAI mapping, prominent spatial differences appeared for the linear versus polynomial regression models. Particularly, a linear regression exhibits problematic behavior, since LAI variations for senescent regions significantly deviate from the other LAI mapping results. A logarithmic fit function seems to lead to more consistent LAI maps. Nevertheless, although the ND type yields considerably poorer calibration results, both maps yield similar LAI patterns. The spatial similarity of the three best performing fitting functions

Figure 9. LAI [m^2/m^2] and LCC [$\mu g/cm^2$] maps based on SR and ND formulations and linear, polynomial and logarithmic fitting functions. The imagery originates from a HyMap flight line over the Barrax agricultural area in Spain. The color bars at the right side of this figure give the values of the LAI and LCC mapped variables.



(SR, logarithmic, ND linear and ND polynomial) provide confidence in the realism of the maps produced.

- With regard to LCC mapping, the spatial similarity between the SR and ND polynomial fits and the degree of detail is remarkable. However, interpretation becomes difficult for areas earlier identified as having low LAI values in the LAI map and showing high LCC values in the LCC map. One interpretation may be that the nature of a polynomial of second degree fit leads to retrievals of high LCC values at near-zero SI values. Another problem might be related to the difficulty of upscaling. For the areas where the canopy is very sparse, i.e. the LAI is low, it becomes harder to perceive the chlorophyll pigments in the leaves with a moderate spatial resolution of 1-4m, even if the chlorophyll concentration in the small leaves is relatively high.

- Logarithmic ND fits, along with linear SR and ND fits, elicit the most realistic LCC patterns; with spatial patterns conform to those of the best LAI maps.

Quite remarkably, the different fitting functions lead to considerable local spatial variation. This can be problematic, since it suggests that the prediction efficiency results on their own may be insufficient as an assessment criterion for the quality of the mapping results. An important aspect to be taken into consideration is that an independent validation dataset likewise gives limited assessment information of the complete map due to a typically low spatial cover within a remote sensing image. Hence, the question arises, which maps - as a whole - are quantitatively most trustworthy? This rather enigmatic problem area reveals one of the main weaknesses of SI based biophysical parameter mapping e.g., the absence of an uncertainty (or accuracy) metric. Uncertainty estimates on a pixelwise basis provide an indication of retrieval accuracy and are nowadays a prerequisite in processing chains. In part, this is why more sophisticated regression algorithms have been developed, which do provide associated uncertainty estimates, e.g. in the emerging field of Bayesian approaches such as Gaussian processes regression [33]. Nevertheless, if any generalization can be deduced from the maps presented, for LAI the ND polynomial (second degree) fit leads to the most consistent results, while for LCC it is rather the ND logarithmic fit that leads to the most consistent results. Conversely, the dissimilarity across the SR-developed maps suggests that it is a less successful formulation for biophysical parameter mapping. The maps provided here suggest that a normalized difference (ND) formulation can therefore be considered as a more preferable strategy for biophysical parameters mapping.

4.4. Towards a new generation of spectral indices

With ARTMO's SI assessment toolbox extensive and systematic SI performance assessments can be conducted, including band combinations, formulations and fitting functions. Its semi-automatic nature enables the reduction of time and resources for analysis and interpretation. By analyzing all bands against each other in a correlation matrix, ARTMO helps identifying redundant bands and overcoming the Hughes phenomenon or "curse of dimensionality" [32]. The case presented in this paper, shows the analysis of all possible 2-band combinations for SR and ND formulation by using four types of curve fittings on hypespectral data. While the relevance of the visible and red edge spectral regions is widely known for LCC and LAI estimation, and has since long been documented for vegetation mapping

applications [19,30,47–49], the sensitivity in the SWIR is particularly remarkable. Especially for LAI retrieval, but also for that of LCC, the SWIR region was found to be an optimal region for biophysical parameter mapping, each time in combination with the visible region; for LAI, green (B_1) with longwave SWIR (B_2), for LCC red (B_1) with shortwave SWIR (B_2). The green-SWIR indices do not show an apparent saturation in the value range considered for LAI, so that the transfer function can be defined with a linear regression model. Similar improvements have been reported by previous studies [44,45,50–53]. In the SWIR, leaf reflectance is mainly controlled by water absorption, although leaf biochemical components such as proteins, cellulose, and lignin also contribute to a lesser extent to leaf absorption (e.g., [54]).

At canopy level, the main drivers of LAI and LCC spatial variation are irrigation regimes at the Barrax test site, leading to high LCC parcels (when irrigated and due to high water absorption) in contrast to non-irrigated low LCC parcels, which are senescent during summertime. Canopy structural differences depending on the crop type are a source of variation related to both biophysical parameters as well [55]. The SWIR is sensitive to variations in complex and dissimilar growth stages and growing conditions [56,57]. Moreover, the SWIR spectral region has the advantage of being sensitive to vegetation, while being less influenced by atmospheric aerosols. This avoids the increase in parameter retrieval uncertainty associated with the atmospheric correction of remotely sensed reflectances required to compensate for aerosol scattering impacts on top-of-atmosphere reflectances, specifically playing a role in the visible part of the solar spectrum [52].

When comparing literature findings, where similar correlation matrices are presented, it is remarkable that identical band combinations are rarely reported for studies based on the use of narrowband (hyperspectral) imaging [31,52,53]. This leads us to suggest that established SI models are more often than not case specific and thus are likely to perform sub-optimal for local hyperspectral mapping applications. Quite often, the traditional red-NIR combination is applied. At the same time, suboptimal SI formulations or regression function selections also lead to less optimal biophysical parameter retrievals. An in-depth analysis is therefore not a luxury. The SI assessment toolbox in ARTMO, as described in this paper, offers a systematic, semi-automatic streamlined and complete procedure for the selection as well an assessment of the most accurate and sensitive SI formulations for biophysical parameter retrieval based on hyperspectral image datasets. The SI model(s) selected can hence confidently be applied for the mapping of biophysical parameters retrieved from hyperspectral imagery.

Finally, it is foreseen that the ARTMO SI assessment toolbox will serve future sensitivity studies. This may lead to a generation of new SIs based on the following research objectives:

- Development and evaluation of generic SIs by using RTMs including viewing and solar geometry. ARTMO includes turbid (e.g., SAIL) as well as explicit 3D RTMs (e.g., FLIGHT). Evidently, the validity of the most optimal models will be assessed using in situ validation data.

- New types of index formulations will be developed and tested across the full optical spectrum. Only SR and ND indices have been assessed here. Yet, alternative mathematical formulations are to be tested as well, e.g. the formulations used for soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI) or more complex ones.

- Although in this paper only two-band formulations have been assessed, the SI toolbox offers the option for a systematic analysis of a full optical spectrum of band combinations with SI formulations of up to ten different bands in the range of 400 to 2500 nm. Since there is no reason to believe that two-band indices lead to the most successful SI models, this multiple band approach may lead to the development of more accurate and sensitive spectral indices.

Regardless of the above-mentioned research objectives, essentially, the SI assessment toolbox provides all necessary tools for quality assurance and quality control (QA/QC), and for a rapid and comprehensive biophysical parameter mapping. The toolbox is intuitive and can assist significantly in precision farming and landscape ecology monitoring applications, thereby allowing to optimize SIs, even per land cover type.

5. Conclusions

A newly developed Spectral Indices (SI) assessment toolbox in the ARTMO modeling environment, enables the analysis and assessment of the accuracy of an indefinite number of SI models. Basically, the toolbox offers a systematic but still empirical approach for the assessment of all possible 2, 3 or 4-band SI formulations. Datasets can be partitioned into calibration and validation subsets. These datasets may originate from simulations, e.g. as generated by the optical radiative transfer models in ARTMO, or from field campaigns. Several options have been included in the SI assessment approach, amongst which: (1) The addition of noise and the possibility to select fitting functions (e.g., linear, exponential, power or polynomial functions); (2) the SI toolbox virtually allows for any type of spectral index model to be formulated and evaluated using up to ten spectral bands; and (3) the possibility to assess and apply SIs per land cover class.

Using HyMap data calibrated with field measured data, the predictive power of generic narrowband spectral indices in the 430-2490 nm range to quantify LAI and LCC has been assessed. For LAI retrieval, the B_1 in the green (B_1: 539, 555, 570 nm) combined with longwave SWIR (e.g., B_2: 2421, 2453 nm) has been evaluated as the most accurate approach (RMSE: 0.61; r^2: 0.83). For LCC retrieval, the B_1 in the far-red (692 nm) combined with the B_2 NIR (e.g., 1340 nm) or shortwave SWIR (e.g., 1661, 1686 nm) has been evaluated as the most accurate approach (RMSE: 4.21-4.70, r^2: 0.91- 0.93). In either case, the identification of the SWIR rather than the conventional NIR as an important spectral region reinforces our suggestion that the ARTMO SI assessment toolbox significantly facilitates the development of new and especially better performing spectral indices for biophysical parameter mapping applications.

Acknowledgements

This paper has been partially supported by the Spanish Ministry for Science and Innovation under projects: AYA2010-21432-C02-01 and CSD2007-00018.

References

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

2. Báez-González, A.; Chen, P.Y.; Tiscareño López, M.; Srinivasan, R. Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. *Crop Science* **2002**, *42*, 1943–1949.
3. Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.W.; Brazile, J.; Singh, R.; Schaepman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* **2007**, *9*, 165–193.
4. Drusch, M.; Del Bello, U.; 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. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **2012**, *120*, 25–36.
5. Donlon, C.; Berruti, B.; Buongiorno, A.; Ferreira, M.H.; Fmnias, P.; Frerick, J.; Goryl, P.; Klein, U.; Laur, H.; Mavrocordatos, C.; Nieke, J.; Rebhan, H.; Seitz, B.; Stroede, J.; Sciarra, R. The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sensing of Environment* **2012**, *120*, 37–57.
6. Stuffler, T.; Kaufmann, C.; Hofer, S.; Frster, K.; Schreier, G.; Mueller, A.; Eckardt, A.; Bach, H.; Penn, B.; Benz, U.; Haydn, R. The EnMAP hyperspectral imager-An advanced optical payload for future applications in Earth observation programmes. *Acta Astronautica* **2007**, *61*, 115–120.
7. Roberts, D.; Quattrochi, D.; Hulley, G.; Hook, S.; Green, R. Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. *Remote Sensing of Environment* **2012**, *117*, 83–101.
8. Labate, D.; Ceccherini, M.; Cisbani, A.; De Cosmo, V.; Galeazzi, C.; Giunti, L.; Melozzi, M.; Pieraccini, S.; Stagi, M. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronautica* **2009**, *65*, 1429–1436.
9. Myneni, R.; Maggion, S.; Iaquinta, J.; Privette, J.; Gobron, N.; Pinty, B.; Kimes, D.; Verstraete, M.; Williams, D. Optical remote sensing of vegetation: Modeling, caveats, and algorithms. *Remote Sensing of Environment* **1995**, *51*, 169–188.
10. Gobron, N.; Pinty, B.; Verstraete, M.; Widlowski, J.L. Advanced vegetation indices optimized for up-coming sensors: design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing* **2000**, *38*, 2489–2505.
11. Houborg, R.; Boegh, E. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sensing of Environment* **2008**, *112*, 186–202.
12. Rivera, J.; Verrelst, J.; Leonenko, G.; Moreno, J. Multiple Cost Functions and Regularization Options for Improved Retrieval of Leaf Chlorophyll Content and LAI through Inversion of the PROSAIL Model. *Remote Sensing* **2013**, *5*, 3280–3304.
13. Verrelst, J.; Rivera, J.; Leonenko, G.; Alonso, L.; Moreno, J. Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *In press*.

14. Chen, J.; Cihlar, J. Retrieving leaf area index of boreal conifer forests using landsat TM images. *Remote Sensing of Environment* **1996**, *55*, 153–162.
15. Myneni, R. Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing* **1997**, *35*, 1380–1393.
16. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Macé, D.; Pragnère, A.; Myneni, R.; Knyazikhin, Y.; Wang, L. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment* **2003**, *84*, 1–15.
17. Broge, N.; Mortensen, J. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote Sensing of Environment* **2002**, *81*.
18. Haboudane, D.; Miller, J.; Pattey, E.; Zarco-Tejada, P.; Strachan, I. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* **2004**, *90*.
19. Thenkabail, P.; Smith, R.; De Pauw, E. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing* **2002**, *68*, 607–621.
20. Ustin, S.; Roberts, D.; Gamon, J.; Asner, G.; Green, R. Using imaging spectroscopy to study ecosystem processes and properties. *BioScience* **2004**, *54*, 523–534.
21. Hatfield, J.; Gitelson, A.; Schepers, J.; Walthall, C. Application of spectral remote sensing for agronomic decisions. *Agronomy Journal* **2008**, *100*, S117–S131. cited By (since 1996)67.
22. Baret, F.; Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* **1991**, *35*, 161–173.
23. Verrelst, J.; Schaepman, M.; Koetz, B.; Kneubhler, M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sensing of Environment* **2008**, *112*, 2341–2353.
24. Verrelst, J.; Schaepman, M.E.; Malenovský, Z.; Clevers, J.G.P.W. Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sensing of Environment* **2010**, *114*, 647–656.
25. Ceccato, P.; Flasse, S.; Grégoire, J.M. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sensing of Environment* **2002**, *82*, 188–197.
26. Zarco-Tejada, P.; Miller, J.; Noland, T.; Mohammed, G.; Sampson, P. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* **2001**, *39*, 1491–1507.
27. Le Maire, G.; Franois, C.; Dufrłne, E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment* **2004**, *89*, 1–28.
28. Zarco-Tejada, P.; Berjón, A.; López-Lozano, R.; Miller, J.; Martn, P.; Cachorro, V.; Gonzlez, M.; De Frutos, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy

reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment* **2005**, *99*, 271–287.

29. le Maire, G.; François, C.; Soudani, K.; Berveiller, D.; Pontailler, J.Y.; Bréda, N.; Genet, H.; Davi, H.; Dufrêne, E. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment* **2008**, *112*, 3846–3864.
30. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* **2011**, *11*, 7063–7081.
31. Mariotto, I.; Thenkabail, P.; Huete, A.; Slonecker, E.; Platonov, A. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission. *Remote Sensing of Environment* **2013**, *139*, 291–305.
32. Thenkabail, P.; Lyon, G.; Huete, A. Advances in Hyperspectral Remote Sensing of Vegetation and Agricultural Croplands. *Hyperspectral Remote Sensing of Vegetation* **2011**, pp. 3–29.
33. Verrelst, J.; Alonso, L.; Camps-Valls, G.; Delegido, J.; Moreno, J. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing* **2012**, *50*, 1832–1843.
34. Verrelst, J.; Romijn, E.; Kooistra, L. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. *Remote Sensing* **2012c**, *4*, 2866–2889.
35. Rivera Caicedo, J.; Verrelst, J.; Alonso, L.; Moreno, J.; Camps-Valls, G. Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2013**, p. In press.
36. Richter, K.; Atzberger, C.; Hank, T.; Mauser, W. Derivation of biophysical variables from earth observation data: Validation and statistical measures. *Journal of Applied Remote Sensing* **2012**, *6*.
37. Gandía, S.; Fernández, G.; Garcia, J.C.; Moreno, J. Retrieval of Vegetation Biophysical Variables from CHRIS/PROBA Data in the SPARC Campaign. *Proceedings of the 2nd CHRIS/PROBA Workshop* **2004**.
38. Fernández, G.; Moreno, J.; Gandía, S.; Martínez, B.; Vuolo, F.; Morales, F. Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns. *Proceedings of the SPARC Workshop* **2005**.
39. Rouse, J.; Haas, R.; Schell, J.; Deering, D.; Harlan, J. *Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation* **1974**.
40. Gitelson, A.; Gritz, Y.; Merzlyak, M. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology* **2003**, *160*, 271–282.
41. Stenberg, P.; Rautiainen, M.; Manninen, T.; Voipio, P.; Smolander, H. Reduced simple ratio better than NDVI for estimating LAI in Finnish pine and spruce stands. *Silva Fennica* **2004**, *38*, 3–14.
42. Verstraete, M.; Pinty, B. Designing optimal spectral indexes for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing* **1996**, *34*, 1254–1265.

43. Draper, N.; Smith, H. Applied Regression Analysis. *3rd Ed.; Wiley Series in Probability and Statistics* **1998**.
44. Darvishzadeh, R.; Atzberger, C.; Skidmore, A.; Abkar, A. Leaf Area Index derivation from hyperspectral vegetation indicesand the red edge position. *International Journal of Remote Sensing* **2009**, *30*, 6199–6218.
45. Schlerf, M.; Atzberger, C.; Hill, J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment* **2005**, *95*, 177–194.
46. Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I.; Shields, T. The HyMap airborne hyperspectral sensor: The system, calibration and performance. *Proc. 1st EARSeL Workshop on Imaging Spectroscopy* **1998**, pp. 37–42.
47. Filella, I.; Penuelas, J. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing* **1994**, *15*, 1459–1470.
48. Clevers, J.; Gitelson, A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *23*, 344–351.
49. Delegido, J.; Verrelst, J.; Meza, C.; Rivera, J.; Alonso, L.; Moreno, J. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *European Journal of Agronomy* **2013**, *46*, 42–52.
50. Brown, L.; Chen, J.; Leblanc, S.; Cihlar, J. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sensing of Environment* **2000**, *71*, 16–25.
51. Lee, K.S.; Cohen, W.; Kennedy, R.; Maiersperger, T.; Gower, S. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment* **2004**, *91*, 508–520.
52. Gonsamo, A. Normalized sensitivity measures for leaf area index estimation using three-band spectral vegetation indices. *International Journal of Remote Sensing* **2011**, *32*, 2069–2080.
53. Heiskanen, J.; Rautiainen, M.; Stenberg, P.; Mõttus, M.; Vesanto, V.H. Sensitivity of narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species composition. *ISPRS Journal of Photogrammetry and Remote Sensing* **2013**, *78*, 1–14.
54. Curran, P. Remote sensing of foliar chemistry. *Remote Sensing of Environment* **1989**, *30*, 271–278.
55. Verrelst, J.; Muñoz, J.; Alonso, L.; Delegido, J.; Rivera, J.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment* **2012a**, *118*, 127–139.
56. Thenkabail, P.; Smith, R.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment* **2000**, *71*, 158–182.
57. Thenkabail, P.; Enclona, E.; Ashton, M.; Van Der Meer, B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment* **2004**, *91*, 354–376.

Remote Sens. **2013**, *5*, 3280-3304; doi:10.3390/rs5073280

OPEN ACCESS

Remote Sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Multiple Cost Functions and Regularization Options for Improved Retrieval of Leaf Chlorophyll Content and LAI through Inversion of the PROSAIL Model

Juan Pablo Rivera [1], **Jochem Verrelst** [1,*], **Ganna Leonenko** [2] **and José Moreno** [1]

[1] Image Processing Laboratory (IPL), Parc Científic, Universitat de València, E-46980 Paterna, Spain; E-Mails: rijuanpa@uv.es (J.P.R.); jose.moreno@uv.es (J.M.)

[2] Department Psychological Med., Cardiff University, Cardiff CF14 4XN, UK; E-Mail: leonenkog1@cardiff.ac.uk

* Author to whom correspondence should be addressed; E-Mail: jochem.verrelst@uv.es; Tel.: +34-96-354-40-67; Fax: +34-96-354-32-61.

Received: 27 April 2013; in revised form: 25 June 2013 / Accepted: 26 June 2013 / Published: 9 July 2013

Abstract: Lookup-table (LUT)-based radiative transfer model inversion is considered a physically-sound and robust method to retrieve biophysical parameters from Earth observation data but regularization strategies are needed to mitigate the drawback of ill-posedness. We systematically evaluated various regularization options to improve leaf chlorophyll content (LCC) and leaf area index (LAI) retrievals over agricultural lands, including the role of (1) cost functions (CFs); (2) added noise; and (3) multiple solutions in LUT-based inversion. Three families of CFs were compared: *information measures*, *M-estimates* and *minimum contrast methods*. We have only selected CFs without additional parameters to be tuned, and thus they can be immediately implemented in processing chains. The coupled leaf/canopy model PROSAIL was inverted against simulated Sentinel-2 imagery at 20 m spatial resolution (8 bands) and validated against field data from the ESA-led SPARC (Barrax, Spain) campaign. For all 18 considered CFs with noise introduction and opting for the mean of multiple best solutions considerably improved retrievals; relative errors can be twice reduced as opposed to those without these regularization options. *M-estimates* were found most successful, but also data normalization influences the accuracy of the retrievals. Here, best LCC retrievals were obtained using a normalized "L_1-estimate" function with a relative error of 17.6% (r^2: 0.73), while best LAI retrievals were obtained through non-normalized "least-squares estimator" (LSE) with

a relative error of 15.3% (r^2: 0.74).

Keywords: biophysical parameters; LUT-based inversion; cost functions; radiative transfer models; PROSAIL; Sentinel-2

1. Introduction

Leaf area index (LAI) and leaf chlorophyll content (LCC) are essential land biophysical parameters retrievable from optical Earth observation (EO) data [1–3]. These parameters give insight in the phenological stage and health status (e.g., development, productivity, stress) of crops and forests [4,5]. The quantification of these parameters over large areas has become an important aspect in agroecological, environmental and climatic studies [6]. At the same time, remotely sensed observations are increasingly being applied at a within-field scale for dedicated agronomical monitoring applications [7,8].

For the last few decades, various space-based LAI and LCC retrieval approaches have been proposed (see review in [6]), some of them eventually led to operational retrieval strategies. Particularly LAI proved to be successful parameter for being operationally and globally retrieved at resolutions of 250 m to 1 km (e.g., MODIS, CYCLOPES and Geoland2 products) [9,10]. Moreover, a lot of efforts are being undertaken to generate a global LAI product at a 30-m Landsat scale [11]. However, operationally retrieved land LCC products are scarce. Until the loss of the ENVISAT spacecraft LCC maps were routinely delivered at a medium spatial resolution through the MERIS terrestrial chlorophyll index [12] or through a trained neural net [13]. But routinely generated LCC maps originated from high spatial resolution images (e.g., $\leq$20 m) are absent until now, although some operational approaches have been proposed for SPOT data [14,15].

Meanwhile, new generation of high resolution (*i.e.*, 10–60 m) land monitoring EO missions are being constructed to be launched such as the forthcoming superspectral Sentinel-2 (S2) mission and hyperspectral missions such as Enmap [16], PRISMA [17] and HyspIRI [18].

Such unprecedented richness of high spectral and spatial resolution data streams makes the availability of robust retrieval methods more important than ever.

To implement a method in an operational processing chain the method should be able to deliver accurate estimates with easy implementation in practice. Inversion of physically-based canopy radiative transfer models (RTMs) against actual EO data is generally considered as one of the most robust approaches to map biophysical parameters over terrestrial surfaces [6,19]. But this approach is not straightforward. According to Hadamarad postulates, mathematical models of physical phenomena are mathematically invertible if the solution of the inverse problem to be solved exists, is unique and depends continuously on variables [20]. Unfortunately this assumption is not met. In fact, the inversion of canopy RTMs is by nature an ill-posed problem mainly for two reasons [21]: on the one hand, several combinations of canopy biophysical and leaf biochemical parameters have a mutually compensating effect on canopy reflectance thus leading to very similar solutions. On the other hand,

model uncertainties and simplifications (e.g., 1D nature of some models) may induce large inaccuracies in the modelled canopy reflectance.

Several strategies have been proposed to circumvent the drawback of ill-posedness, including Lookup-table (LUT)-based inversion strategies [19,20,22–24], hybrid approaches in which LUTs are generated to feed machine learning approaches [13,21,25–28], or LUT-based iterative numerical optimization methods [29]. They all have their strengths and weaknesses in specific situations. But the main advantage of LUT-based inversion approaches is that it can be fast because the most computationally expensive part of the inversion procedure is completed before the inversion itself [6].

LUT-based inversion in its essential form, *i.e.*, direct comparison of LUT spectra against an observed spectra through a cost function (CF), also in some cases known as distance, merit function, metric or divergence measure, is part of the majority of applied inversion approaches. Such a function yields a value for one or multiple biophysical parameters by minimizing the summed differences between simulated and measured reflectances for all wavelengths [20]. Various regularization strategies have been proposed to further optimize the robustness of the estimates: (1) the use of prior knowledge about model parameters [19,22,30–32]; (2) the use of multiple best solutions in the inversion (instead of the single best solution) [23,31,33,34]; (3) adding noise to account for uncertainties attached to measurements and models [23,33,34]; and (4) the combination of single variables into synthetic variables such as canopy chlorophyll content [6,13,24,35]. Nevertheless, aforementioned approaches face limitations when implementing them into a more operational context.

First, in the majority of these studies root mean square error (RMSE) was used as CF between simulated and measured spectra. However, in case of outliers and nonlinearity, the residuals are distorted and therefore the key assumption for using RMSE (maximum likelihood estimation with the Gaussian noise) is violated [36]. The latter authors suggested that alternative CFs may provide a more robust way to estimate biophysical parameters since they allow retrievals for cases where errors are not normally distributed and allow dealing with nonlinear high-parametric problems. Verrelst *et al.* [37] recently demonstrated that alternative CFs, in combination with aforementioned regularization strategies, can considerably improve biophysical parameters retrievals. Yet only three alternative CFs—out of more than 60—were extensively evaluated so far, which leaves an urgent need to evaluate the performance of other promising CFs.

Second, the majority of these studies focus on a specific vegetation type such as croplands, often identified as a land cover class within an image [23,34,38]. However, this approach can turn out problematic when applying LUT-based inversion over larger areas. While land cover classification schemes help to split the problem into sub-domains for which prior information is attributed separately [39] it assumes that up-to-date knowledge of land cover types is available at high spatial resolution, which is usually not the case in an operational context. These limitations imply that alternatives have to be sought to enable full-scene LUT-based inversion over heterogeneous areas without relying on predefined land cover maps.

Henceforth, for the sake of realizing robust LCC and LAI retrievals at high spatial resolution, the aim of this work is to invest in a retrieval processing chain that combines LUT-based inversion with different CFs and regularization options. Simulated Sentinel-2 data at 20 m resolution will be used for this exercise but in principle the inversion schemes can be applied to any optical multispectral EO data. This brings

us to the following objectives: to evaluate the role of (1) cost functions; and (2) regularization options in LUT-based inversion strategies, such as: (i) added noise; (ii) multiple best solutions; and (iii) the role of data normalization.

2. Methodology

2.1. Cost Functions

The estimation of biophysical parameters from satellite data is hampered by uncertainties of very complicated and different nature in every particular study. In many cases it is difficult to study these errors since they have unknown magnitude and distribution. On this basis, we opted to evaluate multiple cost functions that have been introduced in Leonenko *et al.* [36] and then search for the optimal ones that minimizes the errors between simulated and real reflectances. Different CFs deal with different classes of distributions which allows us to deal with outliers and non-linearities distort in a better manner than commonly used least squares estimation (LSE) distance. Therefore it can provide more accurate results for estimated biophysical parameter. Note, that LSE corresponds to the maximum likelihood estimation with the Gaussian noise.

These distances/metrics came from different fields of mathematics, statistics and physics and they all represent "closeness" between two functions but the nature of these functions can be different. For this reason, these metrics have been divided into three broad families based on physical properties of functions in consideration: information measures, M-estimates and minimum contrast method. The detailed description of these families can be found in Leonenko *et al.* [36]. However most of these functions require one or two parameters to be tuned. That may hamper their use in operational processing chains, therefore in this work we chose to analyse only those CFs without additional parameters. Some brief information about them is provided below.

To describe the problem in a statistical way we suppose that $D[P, Q]$ represents a distance between two functions, where $P = (p(\lambda_1), \cdots, p(\lambda_n))$ is satellite and $Q = (q(\lambda_1), \cdots, q(\lambda_n))$ is LUT correspondent reflectances and $\lambda_1, \cdots, \lambda_n$ represent n bands. Also we define η as a biophysical parameter of our interest that we need to estimate (for example LAI or LCC) and $\bar{\zeta} = (\zeta_1, ..., \zeta_r)$ is a vector of other parameters that our function depends on but do not contribute to minimization Equation (1) . The classical statistical method of inversion (or estimation) and finding optimal parameter (η^*) can be formulated as a semi-parametric problem

$$\eta^* = arg \min_{\eta} \sup_{\bar{\zeta}} D[P(\lambda_j),\, Q_i(\lambda_j)], \mathrm{J} = 1, ..., n \tag{1}$$

The purpose is to find the best estimate for η^* by solving the minimization problem (1) using different statistical distances, as those presented below.

2.1.1. Information Measures

This family of measures, also referred to as "divergence measures", is based on minimization of distances between two probability distributions. In this case reflectances are considered as probability distributions and normalization is required (sum of probabilities is (1) prior to numerical application.

Note that normalization has been performed on LUT reflectances as well. This family was first introduced by Kullback and Leibler (KL) [40] and refers to the concept of information divergences, which are non-symmetric between two distributions P and Q. This concept has been further extended in many directions since its initial application in decoding schemes and signal processing and now plays an important role in multimedia classification, neuroscience and cluster analysis. More details and classifications can be found in [41]. Here is the list of the divergencies that we have used for our study:

1. This measure is called the Kullback Leibler divergence and it also corresponds to the maximum likelihood distance in probabilistic space:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) ln \left(\frac{p(\lambda_l)}{q(\lambda_l)} \right) \tag{2}$$

2. This measure is called Pearson chi-square:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} \frac{(q(\lambda_l) - p(\lambda_l))^2}{p(\lambda_l)} \tag{3}$$

3. Squared-Hellinger measure:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) \left(\sqrt{\frac{p(\lambda_l)}{q(\lambda_l)}} - 1 \right)^2 = \sum_{\lambda_1=1}^{\lambda_n} (\sqrt{p(\lambda_l)} - \sqrt{q(\lambda_l)})^2 \tag{4}$$

4. Neyman chi-square divergence:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} \frac{(p(\lambda_l) - q(\lambda_l))^2}{q(\lambda_l)} \tag{5}$$

5. Jeffreys-Kullback-Leibler:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} (p(\lambda_l) - q(\lambda_l)) \left(ln(p(\lambda_l)) - ln(q(\lambda_l)) \right) \tag{6}$$

6. K-divergence of Lin:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) ln \left(\frac{2p(\lambda_l)}{p(\lambda_l) + q(\lambda_l)} \right) \tag{7}$$

7. L-divergence of Lin is a symmetric version of K-divergence:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) \ln(p(\lambda_l) + q(\lambda_l) \ln(q(\lambda_l) - (p(\lambda_l) + q(\lambda_l)) \ln \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) \tag{8}$$

8. The harmonique Toussaint measure:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} \left(p(\lambda_l) - \frac{2p(\lambda_l)q(\lambda_l)}{p(\lambda_l) + q(\lambda_l)} \right) \tag{9}$$

9. The negative exponential disparity measure:

$$D[P,Q] = \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) \left(exp(-\frac{p(\lambda_l) - q(\lambda_l)}{q(\lambda_l)}) - 1 \right) \tag{10}$$

10. Bhattacharyya divergence:

$$D[P,Q] = -log \left(1 + \sum_{\lambda_1=1}^{\lambda_1} \sqrt{p(\lambda_l)q(\lambda_l)} - \frac{1}{2}(p(\lambda_l) + q(\lambda_l)) \right) \tag{11}$$

11. Shannon (1948):

$$\begin{aligned} D(P,Q) = &- \sum_{\lambda_1=1}^{\lambda_n} \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) log \left(\frac{p(\lambda_l) + q(\lambda_l)}{2} \right) + \\ &\frac{1}{2} \left(\sum_{\lambda_1=1}^{\lambda_n} p(\lambda_l) log(p(\lambda_l)) + \sum_{\lambda_1=1}^{\lambda_n} q(\lambda_l) log(q(\lambda_l)) \right) \end{aligned} \tag{12}$$

2.1.2. Nonlinear Regression and M-Estimates

M-estimates form a broad class of estimators which exhibit certain robust properties and it is obtained as the minima of sums of functions of the data. "M" stands for "maximum likelihood-type" estimates and can be described through a nonlinear regression function that seeks to find the relationship between one or more independent variables and a dependent one. Certain widely used regression functions, such as least-squares estimator (LSE) have favorable properties if their underlying assumptions are true (for example noise is Gaussian), but can give misleading results if those assumptions are violated. Least-squares estimators and many maximum-likelihood estimators are M-estimators. They are obtained by replacing square loss function into another more general convex function, see [42]. Generally estimates with robust regression methods can be more stable with respect to anomalous errors but the performance of them drops when the parametric family is misspecified. Normalization is not required but it may help improving accuracies.

The classical LSE distance corresponds to the function

$$D[P,Q] = \sum_{\lambda_i=1}^{\lambda_n} (p(\lambda_i) - q(\lambda_i))^2 \tag{13}$$

It is well known that LSE method is consistent, asymptotically normal and efficient and errors have Gaussian distribution. Also it can be shown that least squares corresponds to the maximum likelihood criterion and can also be derived as a method of moments estimator. Other examples of *M-estimates* that are used in our research are listed below. (1) More general estimates with Laplace distribution

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} |p(\lambda_l) - q(\lambda_l)| \tag{14}$$

are known as L_1-estimate or as least absolute error. (2) The Geman and McClure function tries to reduce the effect of large errors further, but it also cannot guarantee uniqueness.

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \frac{(p(\lambda_l) - q(\lambda_l))^2}{(1 + (p(\lambda_l) - q(\lambda_l))^2)} \tag{15}$$

2.1.3. Minimum Contrast Estimation

This class of estimates considers spectral domain and reflectances in this case can be seen as spectral density functions of some stochastic process. It is close to the class of quasi-likelihood estimators, where instead of independence (which does not hold for many cases) is used asymptotical independence. Under some sets of conditions the minimum contrast estimators are consistent. More information can be found in [43].

The basic idea behind it is to minimize the distance (which also called "contrast" in this case) between a parametric model and a non-parametric spectral density. Since one can interpret satellite observations as measurements in the spectral domain these distances seem to be a natural choice for analyzing satellite data. Also for this family of CFs normalization is not required but it may help improving accuracies.

Here some of the spectral distances that have been used in our research with correspondent $K(x)$- contrast function.

1. Let $K(x) = logx + \frac{1}{x}$,

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \left\{ log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) + \frac{p(\lambda_j)}{q(\lambda_j)} \right\} - 1 \tag{16}$$

2. let $K(x) = -logx + x$, then

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \left\{ -log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) + \frac{q(\lambda_j)}{p(\lambda_j)} \right\} - 1 \tag{17}$$

3. let $K(x) = (logx)^2$, then

$$D(P,Q) = \sum_{\lambda_1=1}^{\lambda_n} \{log(q(\lambda_l)) - log(p(\lambda_l))\}^2 \tag{18}$$

4. let $K(x) = xlogx - x$, then

$$D(P,Q) = 1 + \sum_{\lambda_1=1}^{\lambda_n} \frac{q(\lambda_l)}{p(\lambda_l)} \left\{ log\left(\frac{q(\lambda_l)}{p(\lambda_l)}\right) - 1 \right\} \tag{19}$$

2.2. SPARC Validation Dataset

A diverse field dataset, covering various crop types, growing phases, canopy geometries and soil conditions was collected during SPARC (SPectra bARrax Campaign). The SPARC-2003 and SPARC-2004 campaigns took place in Barrax, La Mancha, Spain (coordinates 30°3′N, 28°6′W, 700 m altitude). The test area has an extent of 5 km × 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land. The annual rainfall average is about 400 mm. In the 2003 campaign (12–14 July) biophysical parameters were measured within a total of 110 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size of about 20^2 m^2.

LCC was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. The calibration took place against field values taken from 50 ESUs. A logarithmic function led to best fit with a r^2 of 0.93 [44]. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned to a LAI value, which was obtained as a statistical mean of 24 measures (8 data readings $\times$ 3 replications) with standard errors between 5% and 10% [45]. Strictly speaking, due to the assumption of a random distribution of foliage, clumping is only partially regarded by the instrument and corresponding software, giving therefore effective LAI as ouptut. No bare soil samples were added in the validation dataset because inversion of canopy RTMs is only relevant over vegetated land covers. For both years, we have a total of 9 crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat), with field-measured values of LAI that vary between 0.4 and 5.9 (μ: 3.0, SD: 1.5) and LCC between 10 and 52 (μ: 38, SD: 14) μg/cm^2.

2.3. Sentinel-2

Although the listed CFs can be applied to any EO data, in preparation to forthcoming Sentinel missions we chose applying them to simulated Sentinel-2 (S2) data. The S2 satellites capitalize on the technology and the vast experience acquired with Landsat and SPOT. S2 will be a polar-orbiting, superspectral high-resolution imaging mission [46]. Each S2 satellite carries a Multi-Spectral Imager (MSI) with a swath of 290 km. It provides a set of 13 spectral bands spanning from the visible and near infrared (VNIR) to the shortwave infrared (SWIR), with four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Table 1). S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. The pair of S2 satellites aims to deliver data taken over all land surfaces and coastal zones every five days under cloud-free conditions. To serve the objectives of GMES, S2 satellites will provide imagery for the generation of high-level operational products (level 2b/3) such as land-cover and land-change detection maps and geophysical variables such as LCC, LAI and leaf water content. To ensure that the final product can meet the user requirements, the GMES user committee defined an accuracy goal of the biophysical products of 10% [46].

Table 1. Sentinel-2 MSI band settings. Bands used in this study are bolded.

Band #	B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8a**	B9	B10	B11	B12
Band center (nm)	443	**490**	**560**	**665**	**705**	**740**	**783**	**842**	**865**	945	1375	1610	2190
Band width (nm)	20	**65**	**35**	**30**	**15**	**15**	**20**	**115**	**20**	20	30	90	180
Spatial resolution (m)	60	**10**	**10**	**10**	**20**	**20**	**20**	**10**	**20**	60	60	20	20

Here, S2 MSI imagery was simulated on the basis of Compact High Resolution Imaging Spectrometry (CHRIS) data because of its high spatial and spectral resolution. CHRIS provides high spatial resolution (up to $\sim$17 m) hyperspectral data over the VNIR spectra from 400 to 1,050 nm at 5 different viewing angles. It can operate in different modes, balancing the number of spectral bands, site of the covered area and spatial resolution because of on-board memory storage reasons [47]. We made use of nominal nadir CHRIS observation in Mode 1 (62 bands, maximal spectral information), which were acquired

during the SPARC campaign (12/07/2003). CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. CHRIS imagery was processed using ESA's CHRIS-Box available in VISAT/BEAM, which includes radiometric recalibration, coherent-noise reduction, geometric correction and atmospheric correction [48,49].

Constrained by the spectral range of CHRIS, we considered the S2 bands starting from B2 (490 nm) until B8a (865 nm). The bands at a spatial resolution of 10 m have been coarse-grained to 20 m so that in total 8 bands in the visible and NIR at a 20 m resolution are available. S2 bands at 60 m were not considered as these bands are intended for atmospheric applications. These bands are intended for atmospheric applications, such as aerosols correction, water vapor correction and cirrus detection [46] and are unable to deliver TOC reflectances that are interpretable by canopy RT models.

2.4. LUT Generation

This work was carried out with ARTMO (Automated Radiative Transfer Models Operator) [50,51]. ARTMO is a GUI toolbox written in Matlab that provides essential tools for running and inverting canopy reflectance models with different options. In its latest version (V.3), the toolbox is designed in a modular way, *i.e.*, the radiative transfer models modules along with several post-processing modules. In short, the toolbox enables the user: (i) to choose between various leaf optical models (e.g., PROSPECT-4, PROSPECT-5) and canopy reflectance models (e.g., 4SAIL, SLC, FLIGHT); (ii) to choose between spectral band settings of various air- and space-borne sensors or to define new sensor band settings (in Sensor module); (iii) to simulate a massive amount of spectra and store them in a relational database running underneath ARTMO (main module); (iv) to visualize spectra of multiple models in the same window (in Graphics module); and finally, (v) to run LUT-based model inversion against EO imagery with selected cost function and with optimization and accuracy options for the estimates (in Inversion module).

From the in ARTMO available models we chose for coupling PROSPECT-4 with 4SAIL because of being fast, invertible and well representing homogeneous plant covers on flat surfaces areas such as those present at Barrax. Both models, commonly referred to as PROSAIL, have been used extensively over the past few years for a variety of applications [52]. PROSPECT-4 calculates leaf reflectance and transmittance over the solar spectrum from 400 to 2,500 nm at a 1 nm spectral sampling interval as a function of its biochemistry and anatomical structure. It consists of 4 parameters, being leaf structure, LCC, equivalent water thickness and dry matter content [53]. 4SAIL calculates top-of-canopy reflectance. 4SAIL inputs consist of: LAI, leaf angle distribution, ratio diffuse/direct irradiation, a hot spot parameter and sun-target-sensor geometry. Spectral input consists of leaf reflectance and transmittance spectra, here coming from PROSPECT-4, and a a moist and dry soil reflectance spectrum. To obtain these soil spectra, the average of bare soil signature was calculated from bare moist and dry soil pixels identified in the imagery. In 4SAIL a scaling factor, α_{soil}, has been introduced that takes variation in soil brightness into account as a function of these two soil types.

The bounds and distributions of the PROSAIL variables are depicted in Table 2. Variable bounds were taken from measurement campaigns and/or other studies working with the same crops [23,34]. They were chosen in order to describe the characteristics of all crop types used in the study. Gaussian

input distributions were generated for LAI and LCC content in order to put more emphasis on the variable values being present in the actual growth stages of the crops. Sun and viewing conditions correspond to the situation of the satellite overpass. All possible combinations were calculated from the in Table 2 defined leaf and canopy input ranges. Since the sum of all these combinations would lead to an unrealistic high number of simulations (about 5 billion), a LUT size of 100,000 TOC reflectance realizations was randomly chosen in accordance with [35]. All input parameters, metadata and associated output simulations were automatically stored in a relational database running underneath ARTMO.

Table 2. Range and distribution of input parameters used to establish the synthetic canopy reflectance database for use in the LUT.

Model Parameters		Units	Range	Distribution
Leaf parameters: PROSPECT-4				
N	Leaf structure index	unitless	1.3–2.5	Uniform
LCC	Leaf chlorophyll content	(μg/cm^2)	5–75	Gaussian ($\bar{x}$: 35, SD: 30)
C_m	Leaf dry matter content	(g/cm^2)	0.001–0.03	Uniform
C_w	Leaf water content	(cm)	0.002–0.05	Uniform
Canopy variables: 4SAIL				
LAI	Leaf area index	(m^2/m^2)	0.1–7	Gaussian ($\bar{x}$: 3, SD: 2)
α_{soil}	Soil scaling factor	unitless	0–1	Uniform
ALA	Average leaf angle	(○)	40–70	Uniform
HotS	Hot spot parameter	(m/m)	0.05–0.5	Uniform
skyl	Diffuse incoming solar radiation	(fraction)	0.05	-
θ_s	Sun zenith angle	(○)	22.3	-
θ_v	View zenith angle	(○)	20.19	-
ϕ	Sun-sensor azimuth angle	(○)	0	-

Similar variable ranges/values/distributions were used according to field configurations and related studies [23,34]. $\bar{x}$: mean, SD: standard deviation.

2.5. Regularization Options

Two regularization options are commonly applied in LUT-based inversion strategies. First, a Gaussian noise is often added to the simulated canopy reflectance to account for uncertainties [9,13]. Second, several studies demonstrated that the single best parameter combination corresponding to the very smallest distance as calculated by a cost function does not necessarily lead to the best accuracy [19,31]. Instead the mean of multiple solutions tend to provide more accurate results. Because different optimization numbers are reported in literature, the mutual impact of these regularization options has been recently systematically studied for a limited set of CFs [37]. In that study a range of Gaussian noise levels was introduced up to 30% with steps of 1%. It appeared however that in some cases this upper limit yielded best accuracies, suggesting that the optimal noise configuration may not have been reached. Therefore, in this study noise levels were examined from 0% to 50% with steps of 2%. The percentage noise is considered as the maximal boundary wherein noise can fall and is wavelength dependent.

The same range was applied to multiple solutions, from 0% to 50% with steps of 2%. Note that these wide ranges are only intended to study the behaviour of the CFs, thereby seeking for optima. Further we investigate the impact of spectra normalization in *M-estimates* and *Minimum Contrast Estimates* that have not been done before.

Thus, to summarize, we have:

1. Various standalone cost functions from three mathematically different families.
2. Insertion of Gaussian noise on simulated spectra: 0–50%.
3. Use of multiple sorted best solutions in the inversion: 0–50%.
4. Impact of normalization for CFs in *M-estimates* and *Minimum Contrast Estimates* families.

Given all these factors, their effects on the robustness of LUT-based inversion have been assessed for the retrieval of LCC and LAI at the 20-m S2 resolution. The retrieved predictions were compared against the ground-based SPARC validation dataset using statistics such as coefficient of determination (r^2), absolute RMSE and the normalized RMSE (NRMSE [%] = RMSE/range of the parameters as measured in the field * 100). The NRMSE was used to compare the performances across the different CFs and parameters. Lower values indicate less residual variance and thus more successful inversion.

3. Results

3.1. Evaluation of Cost Functions and Regularization Options

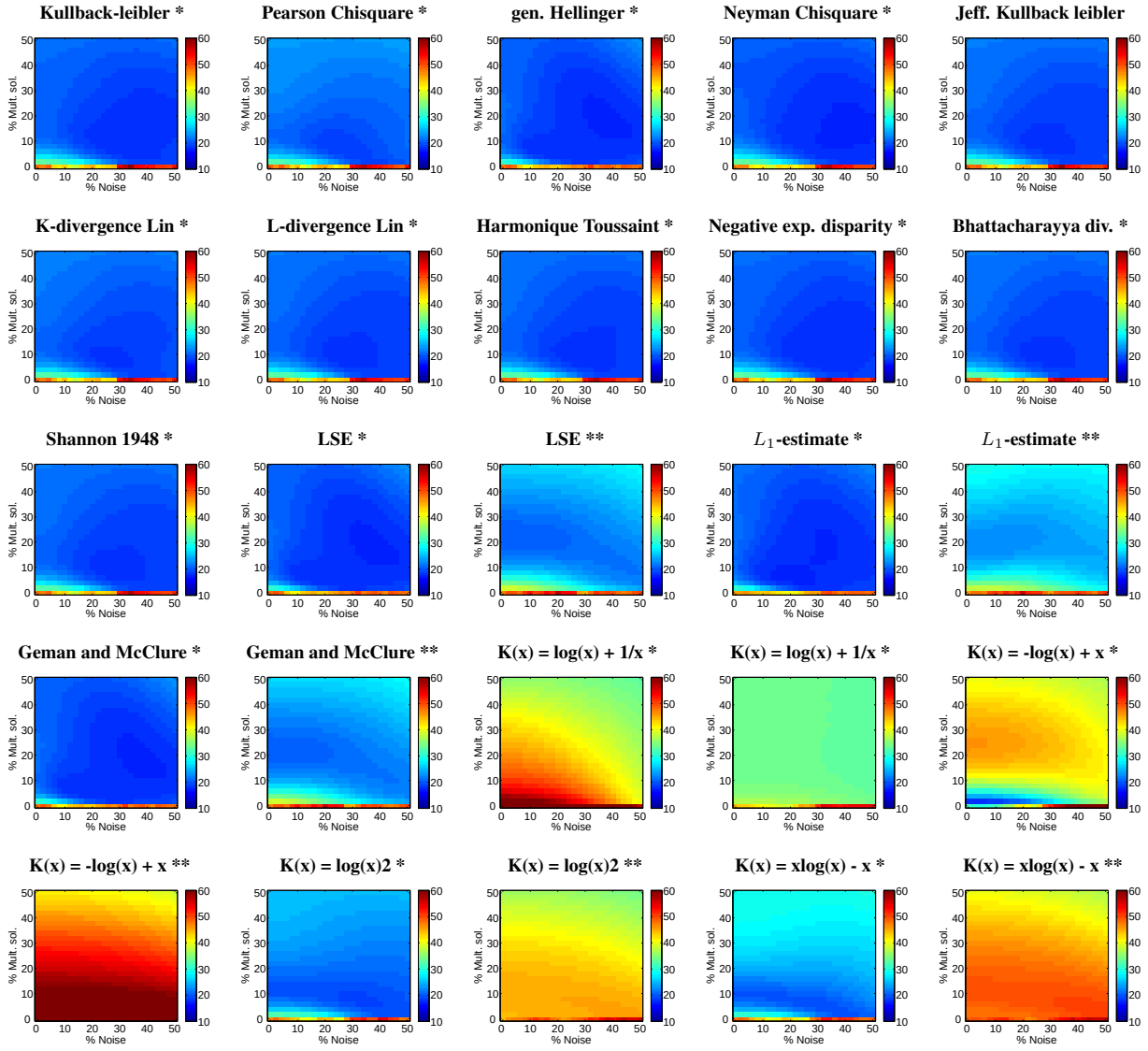
3.1.1. Leaf Chlorophyll Content (LCC) Retrieval

Figure 1 provides a systematic overview of the impact of different cost functions (matrices), noise levels (x-axis) and multiple solutions (y-axis) on LCC retrieval. To enable comparison across the different parameters, the performance of these inversion strategies were evaluated by calculating the deviations between the retrieved parameter values and the 110 validation points through the NRMSE; further referred to as relative error. When interpreting these error matrices, opting for the very single best solution without added noise in the inversion appeared to be a poor inversion strategy for each of the CFs. Inversion without these regularization is shown in the left-bottom corner of the matrices and also listed in the rightmost column of Table 3. Accuracies further degraded when opting for the very single best solution and adding noise. This pattern appeared for the majority of CFs, which suggests that the strategy of only adding noise in the inversion is to be avoided. The reason of the poor performance of the single best solution lies in the ill-possedness of the problem; multiple parameter combinations lead to the same or very similar spectra. Therefore, picking the sorted most matching spectrum according to a CF may not be the best choice, but rather the mean value of multiple best matching spectra should be considered. Effectively, the inversion clearly benefited from introducing multiple solutions in the inversion for all CFs, preferably in combination with added noise. Overall, with consideration of the mean of all CFs, relative errors more were twice reduced when comparing results without regularization options; $\bar{x}$ NRMSE from 49% to 22% (see Table 3).

Moreover, these error matrices reveal the true interactions between the different cost functions and the regularization options. The matrices suggest that each of the tested CFs respond differently to the

regularization options. It is interesting to see that in the majority of them a pattern with optimized performances appear (darkest blue), though the location and shape of this pattern may vary per CF. Thereby, not only CFs that are mathematically more alike led to similar trends, but also similarities appeared within the same CF family. This is particularly noticeable for the family of *information measures*, which shows robust LCC estimates across the whole noise and multiple solutions range, but also *M-estimates* CFs show similar behavior.

Figure 1. Normalised RMSE (NRMSE) matrices for LCC retrieval using cost function displaying the impact of % noise (X-axis) against multiple solutions (Y-axis) in LUT-based RTM inversion. * : normalized; ** : non-normalized. The more bluish, the lower relative errors and thus the better the inversion.



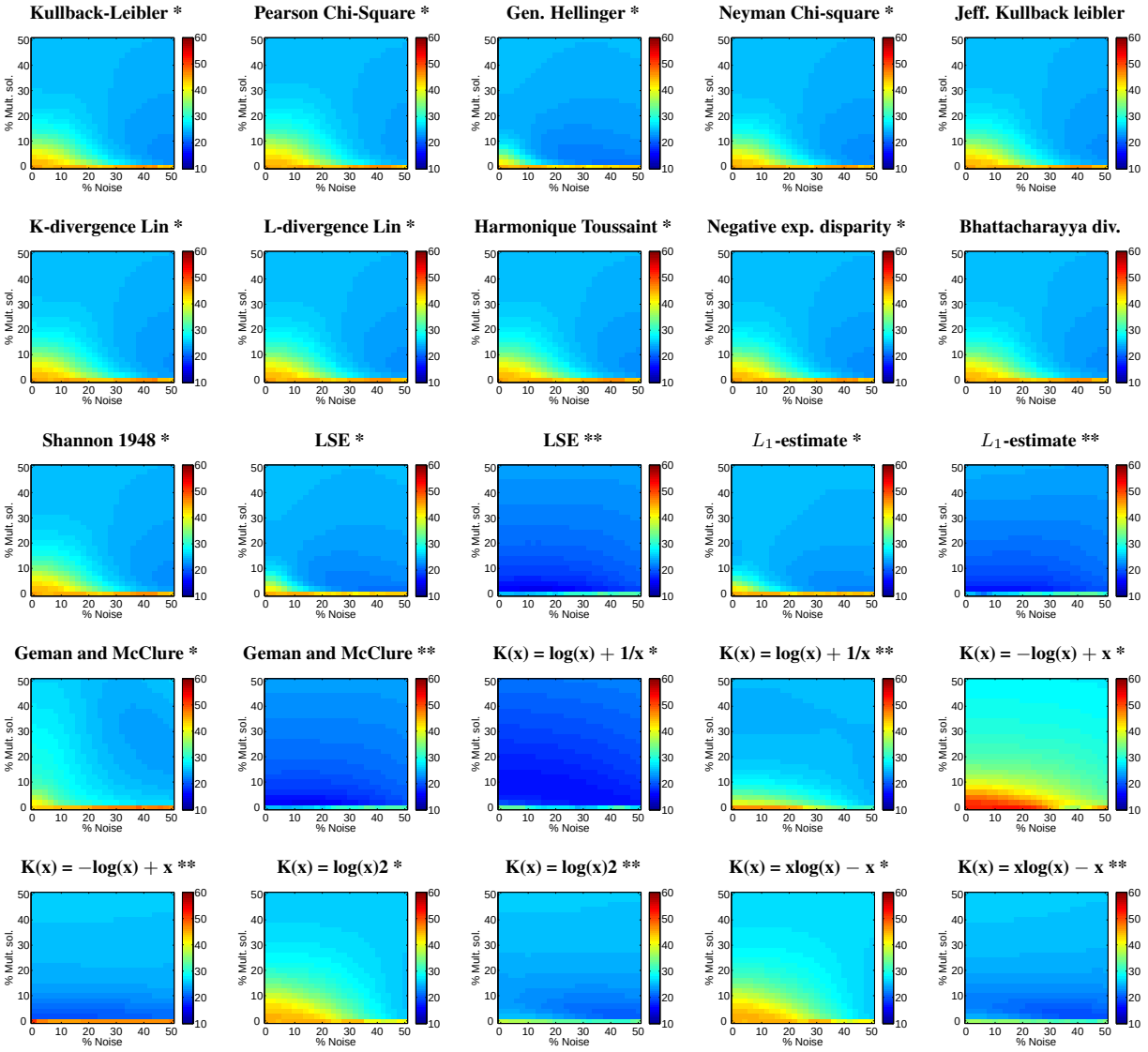
Interestingly, the matrices also revealed that data normalization governs the success of *M-estimates*. Excellent performances were achieved for the three *M-estimates* (LSE, L_1-estimate, Geman and McClure), with best performances for L_1-estimate (NRMSE of 17.6% at 6% multi solutions and

18% noise). Note that L_1-estimate is proven to be optimal when distribution of errors have Laplace distribution. Also Neyman Chi-square and Generalized Hellinger reached such degree of accuracy (see Table 3). However, *M-estimates* performed considerably poorer (best results around 20%) when data is non-normalized, and matrices show that retrieval results are more affected by the two regularization options. Finally, the *Minimum Contrast Estimation* family yielded more irregular patterns. While in general they benefited from normalizing the dataset, only the "K(x) = log(x)^2" and "K(x) = xlog(x) − x" proved to be robust CFs for LCC retrieval. In the absence of normalization best performances of these CFs were obtained at a maximal of tested noise level (*i.e.*, 50%). Because these noise levels led to a considerable degradation of the simulated spectra it seems that these CFs only function well for LCC when data is normalized.

Table 3. Statistics (r^2, RMSE , NRMSE) based on best evaluated NRMSE and corresponding mean multiple solutions (%) and noise level (%) for LCC retrieval. *: normalized; **: non-normalized. The "0.0 NRMSE" column represents the results without regularization options (left-bottom corner in matrices). Best inversion strategy is in bold font.

Cost Function	Mult. Sol. (%)	Noise (%)	r^2	abs. RMSE	NRMSE (%)	0,0 NRMSE (%)
Kullback leibler *	10	26	0.71	7.34	19.47	47.54
Chi square *	8	24	0.69	7.57	18.64	48.94
Generalised Hellinger *	20	36	0.74	7.16	17.63	47.55
Neyman Chi square *	16	44	0.73	7.16	17.63	46.62
Jeffreys Kullback leibler *	10	26	0.70	7.41	18.25	48.36
K-divergence Lin *	8	26	0.70	7.43	18.30	48.01
L-divergence Lin *	10	26	0.70	7.39	18.20	48.25
Harmonique Toussaint *	10	26	0.71	7.37	18.15	48.36
Negative exp. disparity *	10	26	0.71	7.31	18.01	47.43
Bhattacharyya divergence *	10	28	0.70	7.40	18.22	48.18
Shannon 1948 *	10	26	0.70	7.39	18.20	48.18
LSE *	20	36	0.74	7.16	17.63	47.55
LSE **	22	0	0.68	8.23	20.27	46.99
L_1-estimate *	**6**	**18**	**0.73**	**7.14**	**17.59**	**45.88**
L_1-estimate **	20	12	0.61	9.14	22.52	45.58
Geman and McClure *	20	36	0.74	7.16	17.63	47.55
Geman and McClure **	22	0	0.68	8.24	20.30	46.99
K(x) = log(x) + 1/x *	50	50	0.68	13.43	33.09	70.96
K(x) = log(x) + 1/x **	16	50	0.62	13.36	32.91	43.86
K(x) = −log(x) + x *	2	0	0.70	7.45	18.34	31.82
K(x) = −log(x) + x **	50	50	0.48	15.76	38.83	83.15
K(x) = log(x)^2 *	8	32	0.68	7.76	19.11	46.68
K(x) = log(x)^2 **	50	50	0.31	13.87	34.15	44.14
K(x) = x(log(x)) − x *	6	30	0.66	7.95	19.58	47.19
K(x) = x(log(x)) − x **	50	50	0.30	15.05	37.06	48.99

Figure 2. NRMSE matrices for LAI retrieval using cost function displaying the impact of % noise (X-axis) against multiple solutions (Y-axis) in LUT-based RTM inversion. * : normalized; ** : non-normalized. The more bluish, the lower relative errors and thus the better the inversion.



3.1.2. Leaf Area Index (LAI) Retrieval

Introducing noise and mean of multiple solutions considerably improved the LAI accuracies (Figure 2). Overall, on average relative errors almost halved: NRMSE from μ 42% to μ 22% (Table 4). However, in contrast to LCC retrievals, the *information measures* and the normalized *M-estimates* were no longer able to deliver good inversion results. In fact for the majority of CFs, normalization of the data caused that best results were achieved with the maximal degree of injected noise (*i.e.*, 50%). Since such a spurious high noise level led to a considerable degradation of the simulated spectra it makes us concluding that normalization is not recommended for LAI retrieval. Consequently non-normalized *M-estimates* greatly improved accuracy of LAI retrievals. Particularly the classical LSE demonstrated to

be the best CF (NRMSE: 15.3% at multiple solutions of 2% and noise of 14%), closely followed by the other M-estimates (L_1-estimate, Geman and McClure). This means that in our example the errors tend to be Gaussian and solution is stable. Also these results suggest that controlling the normalization option can play an important role in LUT-based inversion. The contrast functions family did not perform very well and only normalized K(x) = log(x) + 1/x is able to retrieve LAI with best relative error of 16.5%. In comparison with LCC (Table 3) it can be noted that best LAI retrieval is performing slightly better (e.g., see NRMSE). Though, most important is to realize that different optimized inversion strategies were identified for both parameters. Finally, although not shown here for the sake of brevity, the retrieval capability of canopy-integrated chlorophyll content (CCC: LAI × LCC) was evaluated. CCC showed analogous patterns as LAI which suggests that this parameter, more than LCC, is the driving parameter of CCC. This can be probably attributed to the fact that LAI influences the whole spectral range, while LCC sensitive spectral range is limited to the beginning of the red edge.

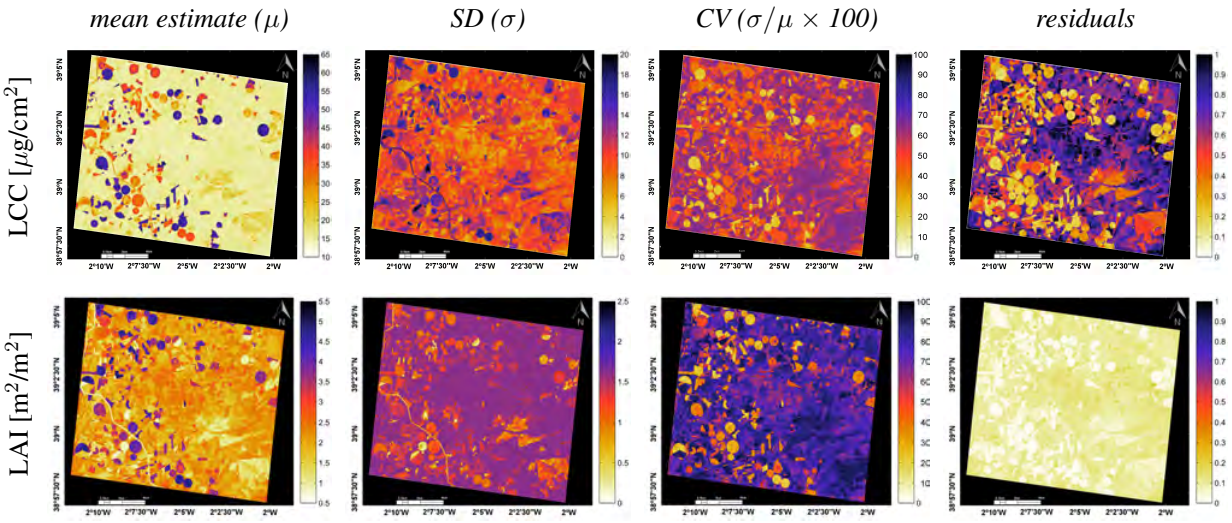
Table 4. Statistics (r^2, RMSE , NRMSE) based on best evaluated NRMSE and corresponding mean multiple solutions (%) and noise level (%) for LAI retrieval. *: normalized; **: non-normalized. The '0.0 NRMSE' column represents the results without regularization options (left-bottom corner in matrices). Best inversion result is in bold font.

Cost Function	Mult. Sol. (%)	Noise (%)	r^2	abs. RMSE	NRMSE (%)	0,0 NRMSE (%)
Kullback leibler *	4	50	0.63	1.25	22.74	45.59
Chi square *	8	50	0.62	1.29	23.53	45.74
Generalised Hellinger *	2	42	0.62	1.17	21.34	44.59
Neyman Chi square *	4	50	0.62	1.24	22.48	45.51
Jeffreys Kullback leibler *	6	50	0.62	1.26	22.93	45.60
K-divergence Lin *	6	50	0.62	1.27	23.06	45.67
L-divergence Lin *	10	26	0.70	1.26	22.92	45.54
Harmonique Toussaint *	6	50	0.62	1.26	22.91	45.60
Negative exp. disparity *	4	50	0.63	1.25	22.72	45.53
Bhattacharyya divergence *	6	50	0.62	1.26	22.93	45.53
Shannon 1948 *	6	50	0.62	1.26	22.92	45.54
LSE *	2	42	0.62	1.17	21.34	44.58
LSE **	**2**	**14**	**0.74**	**0.84**	**15.32**	**25.45**
L_1-estimate *	2	50	0.62	1.22	22.25	44.93
L_1-estimate **	2	12	0.73	0.91	16.57	25.31
Geman and McClure *	2	42	0.62	1.17	21.34	44.59
Geman and McClure **	2	14	0.74	0.85	15.39	25.45
K(x) = log(x) + 1/x *	2	42	0.63	0.91	16.51	34.02
K(x) = log(x) + 1/x **	10	50	0.50	1.35	24.57	46.17
K(x) = −log(x) + x *	50	50	0.52	1.53	27.80	51.69
K(x) = −log(x) + x **	2	0	0.64	1.09	19.86	51.17
K(x) = log(x)^2 *	12	50	0.54	1.40	25.42	45.13
K(x) = log(x)^2 **	2	46	0.66	1.13	20.47	37.94
K(x) = x(log(x)) − x *	6	50	0.63	1.42	25.82	45.11
K(x) = x(log(x)) − x **	2	40	0.63	1.08	19.56	35.57

3.2. Biophysical Parameters Mapping

Obtained systematic overview of the various inversion performances shows that deriving simultaneously multiple biophysical parameters using one inversion strategy is not the best choice. When parameters are correlated in a non-linear way it appears that the optimal cost functions are different for each parameter [36]. Hence, when we try to quantify LCC and LAI simultaneously will at least for one parameter lead to suboptimal performances. While in an operational setting it may be desirable to seek for an inversion strategy that leads to a balanced performance in the generation of all retrievable parameters [37], because of their non-linear dependency applying different retrieval strategies to different parameters is generally preferable. That is also the approach pursued here; for each parameter the best performing inversion strategy was applied (*i.e.*, bolded in Tables 3 and 4). For LCC we applied L_1-estimate (with mult. sol.: 4%, noise: 18%) and for LAI non-normalized normal distribution LSE (with mult. sol.: 2%, noise: 14%). Since in each of these strategies the mean of multiple solutions is applied it leads to the mapping of the mean estimates (μ) and associated standard deviation ($\pm\sigma$) (Figure 3).

Figure 3. Mean predictions, standard deviation (SD), coefficient of variation (CV) and residuals for LCC and LAI by using for each parameter best evaluated inversion strategy (see Tables 3 and 4).



These maps are briefly interpreted. Starting with the μ-estimate maps, it can be observed that not only the irrigated circular fields with green biomass are clearly marked, but also pronounced within-field variations are notable by both parameters. These irrigated fields are characterized by a LCC above 40 μg/cm^2 and LAI above 3. Areas with low LCC ($\leq$25 μg/cm^2) and LAI ($\leq$1.5) are mainly bare soils, non-irrigated fallow lands or senescent or harvested cereal fields (wheat, barley). Particularly detailed spatial variation was encountered in the LAI map which suggests that greater sensitivity for this structural parameter is achieved. The $\pm\sigma$-maps can be interpreted as uncertainty of the μ-estimate. The smaller the $\pm\sigma$ for a pixel, the lower that variation in multiple solutions is encountered during the inversion. Note that such $\pm\sigma$ map is also obtained by the MODIS LAI LUT-based retrieval method [20] and serves

as a measure of the solution accuracy; starting from Collection 5, the $\pm\sigma$(LAI) has been provided in the standard products. Here, at a glance, large uncertainty ranges arise when medium to high LCC conditions prevail. Similar studies observed the same large uncertainty ranges over dense canopies and are attributed to the saturation effect in the radiative transfer and the limited information provided in the a generation of the LUT [54–56]. However, since the $\pm\sigma$ also directly depends on the magnitude of μ, it may be more of interest to map the relative uncertainties, *i.e.*, the coefficient of variation (CV; σ/μ). A relative indicator is of interest as it provides more information about the inversion process, *i.e.*, a lower CV means a greater uniformity across the % best solutions, but it also allows a comparison of the inversion performance across all maps (Figure 3). For both LCC and LAI the CV maps suggest that the inversion process had least difficulty with the vegetated areas (circular irrigated fields), while most uncertainty was propagated over the fallow land and calcareous bare soils (South-East of image).

At the same time, parameter-independent information regarding the performance of the inversion process can also be obtained by mapping the residuals of the CFs. They provide another indicator on the inversion certainty. For each pixel, it indicates the degree of mismatch between the observed spectrum and the best matching simulated spectra. More than CV, these maps delineate the surfaces where the simulated spectra closely matched the observed spectra and thus a higher probability achieving a successful inversion. The maps suggest that particularly LAI was retrieved without difficulty, with perfect match over the vegetated areas. The inversion process had in all generality more difficulty with LCC, but large differences can be observed, leading to essentially the same earlier observed pattern; a close agreement over vegetated areas as opposed to senescent and non-vegetated lands. On the whole, the uncertainty analysis suggests that the inversion would benefit from a wider LUT with more spectral variation of senescent vegetation and bare soil surfaces.

4. Discussion

The upcoming S2 missions open opportunities to implement novel retrieval algorithms in operational processing chains. Specifically, there is a need for retrieval methods that are accurate, robust, and make fully use of the new S2 MSI bands. While in related works vegetation indices (e.g., [57,58]), or machine learning regression algorithms (e.g., [59,60]) have been proposed for S2 biophysical parameter retrieval, alternatively we advocate the method of LUT-based RTM inversion. Although this approach is not new, it has never been exploited to the fullest. Here we systematically analyzed the impact of CFs and regularization options. It led to consolidated findings discussed below.

4.1. Cost Functions and Regularization Options

To mitigate the drawback of ill-posedness instead of introducing prior information it was opted to exploit the performances of alternative CFs in combination with regularization options. The rationale for evaluating CFs as opposed to the widely used LSE (or RMSE) is that propagated uncertainties and errors, e.g., due to uncertainties in instrument calibration, variations in atmospheric composition or simplifying assumptions in the representation of canopy and soil background, distort the residuals and in many cases violate a key assumption for using LSE, which corresponds to the maximum likelihood estimation with Gaussian distribution of residuals [36]. Nevertheless in our presented dataset it was found that

LSE outperformed most of the alternative tested CFs for LCC with the condition of data normalization. Only the "L_1-estimate" function performed slightly better, which suggest that the errors rather tend to be distributed according to a Laplace distribution. Moreover normalization led to better performances with respect of varying added noise and multiple solutions in the inversion. Hence, data normalization provides good accuracy for LCC inversion and can be explained as follows. Variation in reflectance as a function of chlorophyll absorption occurs predominantly in the visible part of the spectrum and declines rapidly once entering red edge [61], *i.e.*, encompassing S2 bands B2 to B5. Accordingly, in PROSPECT the chlorophyll absorption coefficients start to become negligible from 730 nm onwards. A closer look to the simulated dataset revealed that normalization compresses the dataset and so ensures that the observation spectrum falls within the same range. This particularly plays a role in the visible part of the spectrum. As reflectance of vegetated surfaces in the visible is typically low (often $\leq$10%) and variation in this region is narrow, normalization helps reaching a better match in this region between simulated and observed spectra. Conversely, data normalization did not lead to the good CFs performance for LAI, and non-normalized data in combination with multiple solutions and the CF LSE showed to be the most successful. It should herewith be noted that LAI variation (*i.e.*, vegetation density) causes a spectral variation over the whole VNIR spectrum, *i.e.*, S2 bands from B2 to B8. Thereby, the larger magnitude of reflectance in the NIR and greater LAI-induced variation in comparison to the visible part makes NIR an important region for LAI retrieval in LUT-based inversion. Consequently, a spectral observation easily falls within such a broad range of simulated spectral variation, and compression of the simulations through normalization tend to disturb the matching between observed and simulated spectra. A standard inversion scheme proved thus to be most successful for LAI retrieval. Effectively, many publications demonstrated that LAI can be relatively easily retrieved without data normalization (e.g., [19,23,38,62]). We are not aware of other studies that have systematically evaluated the role of data normalization in LUT-based inversion, though proper usage may greatly improve accuracies. For instance, it explains why in [37] RMSE was performing poorly for LAI where this CF was employed in normalized mode only. It may also explain why similar studies reported a poor LCC inversion as there no mentioning of normalizing the data has been made [19,62]. Further, as earlier observed in [37], this study confirms for 18 tested CFs that introducing some degree of noise and mean of multiple solutions in the inversion can lead to improved inversion performances, though its actual impact strongly depended on the considered parameter and cost function. Hence developing optimized inversion strategies for each single retrievable parameter is strongly recommended.

4.2. Inversion Performance

The evaluation of various CFs and regularization options led to identified inversion strategies with accuracies of r^2 on the order of 0.73–0.81 and relative errors (NRMSE) of 15%–18%. A range of 15%–20% is regarded as the currently achievable accuracy for LAI from EO data [63]. However, a retrieval accuracy of 10% is targeted for the S2 mission [64]. It should thereby be emphasized that the validation dataset consisted of all kinds of crop types, including row crops such as maize, potatoes, onions and vineyards. In related work by [19,23] this degree of accuracy was only achieved over single crops, and it was concluded that PROSAIL fails to invert over multi-species canopies. Thereby, various

types of uncertainties have been identified that may lead to suboptimal retrievals, with respect to model usage (e.g., 1D *vs*. 3D models), parametrization and validation data [22]. Suggestions for improvements typically refer to adding more prior information at the level of individual parcels, *i.e.*, through a more specialized LUT (e.g., [20,22,23,51,65,66]). While such strategies could be beneficial for dedicated sites, site-specific information is usually unavailable in an operational context for larger areas. The proposed regularization options yielded robust inversion schemes that are easily applicable over full scenes covering heterogeneous canopy surfaces. Although most of the CFs within the same family perform alike and are in principle thus replaceable, performance gain is especially reached in combination with applying multiple solutions and noise in the inversion. Moreover, further improvements can be achieved by applying CFs that require tuning of additional parameters. In general tuning parameters in CFs would lead to refined inversion strategies as compared to standalone CFs (e.g., see [36,37]). Further research is planned to have these tunable CFs optimized in an automated fashion.

At the same time, information about the inversion uncertainty on a pixel-by-pixel basis may be as relevant as overall accuracies calculated from a limited set of ground-based validation data. Uncertainty indicators were obtained through mapping of CV and residuals. The CV is an indicator of the uncertainty range around the mean estimate, which tells something about the ill-posedness of the inversion of a retrievable parameter, whereas the residuals tells us how much the observed spectra deviate against that from the LUT spectra in the inversion scheme. These quality layers allow masking out uncertain estimates.

Both indicators showed a consistent spatial trend for all parameters: pixels of vegetated surfaces matched closely with the synthetic reflectance database while pixels of non-vegetated surfaces faced more difficulties. Two reasons can be identified for this discrepancy: (1) the inversion strategy was optimized against validation data that was exclusively collected on vegetated areas; and (2) PROSAIL is a canopy reflectance model and thus well able to detect variation in vegetation properties. Hence the generated LUT and final inversion scheme were not optimized to detect variations in dried-out fallow and bare soil lands. For retrievals over full images there is thus a need for regulating the inversion strategy both over vegetated targets as well over non-vegetated targets. Therefore, while having inversion over vegetated canopies resolved, adequately processing non-vegetated surfaces remains to be optimized. To start with, further efforts are needed in the generation and evaluation of more generic LUTs, *i.e.*, including spectra of kinds of soil types, man-made surfaces and water bodies. PROSAIL alone is not able to deliver this; hence coupling with a generic soil reflectance model that enables generating a variety of non-vegetated reflectances would be a next step to do. It should hereby also be noted that the 20 m-S2 bands B11 and B12 have not been considered in this study because of falling outside the CHRIS range. These bands in the SWIR are known to be sensitive to vegetation structure [67] and can better distinguish between dried-out fallow and non-vegetated lands. It is expected that inclusion of the SWIR bands will further improve the retrieval quality. Another promising avenue to be investigated is relying on vegetation indices to spatially constrain the LUTs. For instance, vegetation indices are able to detect bare soil, water bodies, sparsely vegetated areas and densely vegetated areas (e.g., see also [32]). That information could then be used to constrain LUTs on a per-pixel basis and is currently explored to be implemented into ARTMO.

5. Conclusions

LUT-based inversion is considered as a physically-sound retrieval method to quantify biophysical parameters from Earth observation imagery, but the full potential of this method has not been consolidated yet. Here, we have systematically compared 18 different cost functions (CFs) originating from three major statistical families: *information measures*, *M-estimates* and *minimum contrast methods*. All these CFs are standalone functions that can be directly implemented in processing algorithms. These CFs were fully exploited through various regularization options, *i.e.*, adding noise, multiple solutions and normalization, for the benefit of improved LCC and LAI estimations. Inversion of the PROSAIL model against a simulated 20 m Sentinel-2 imagery (8 bands: B2-B8a) over an agricultural site (Barrax, Spain) and against a validation dataset led to the following conclusions:

- All evaluated CFs and biophysical parameters gained from regularization options such as adding some noise and multiple solutions in the inversion. These options with proper adjustment can significantly reduce relative errors.
- With introduction of multiple solutions and noise *information measures* CFs proved to be successful for deriving LCC. However best LCC results were achieved with the M-estimate L_1 (NRMSE of 17.6% at 6% multiple solutions and 18% noise) when data is normalized.
- Data normalization appeared to be unsuccessful for retrieving LAI. Here, the classical LSE yielded best results for non-normalized data; NRMSE of 15.3% at 6% multiple solutions and 18% noise, and 16.4% at 16% multiple solutions and 0% noise, respectively.

Systematic analysis for each biophysical parameter identified different optimized inversion strategy, which was subsequently applied to pixel-by-pixel Sentinel-2 imagery. It provided us with maps of mean estimates and associated statistics and showed insight into the uncertainty of the retrievals (e.g., coefficient of variation and residuals). These indicators showed that inversions were most successful over densely vegetated areas. PROSAIL had more difficulty with processing fallow and non-vegetated lands, but that is expected to be resolved with an adjusted LUT and the addition of SWIR bands in actual Sentinel-2 data.

The bottom line of this work is that, despite common practice, no single inversion strategy was found to be optimal for deriving multiple biophysical parameters. While some general trends with respect to regularization options and CFs was revealed, it is the data distribution that determines the success of the inversion strategy, which is governed by the biophysical parameter, the generated LUT, the spectral bands and the validation data. It is therefore recommended to test different CFs and regularization options before implementing an inversion scheme in a processing chain.

Acknowledgements

This paper has been partially supported by the Spanish Ministry for Science and Innovation under projects: AYA2010-21432-C02-01 and CSD2007-00018. Three anonymous reviewers are thanked for providing comments that helped with improving the quality of the original manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Whittaker, R.H.; Marks, P.L. Methods of Assessing Terrestrial Productivity. In *Primary Productivity of the Biosphere*; Springer Verlag: Berlin/Heidelberg, Germany, 1975; pp. 55–118.
2. Lichtenthaler, H.K. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol.* **1987**, *148*, 350–382.
3. Malenovský, Z.; Rott, H.; Cihlar, J.; Schaepman, M.; García-Santos, G.; Fernandes, R.; Berger, M. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* **2012**, *120*, 91–101.
4. Lichtenthaler, H.K.; Lang, M.; Sowinska, M.; Heisel, F.; Miehe, J.A. Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.* **1996**, *148*, 599–612.
5. Sampson, P.H.; Zarco-Tejada, P.J.; Mohammed, G.H.; Miller, J.R.; Noland, T.L. Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods. *For. Sci.* **2003**, *49*, 381–391.
6. Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.W.; Brazile, J.; Singh, R.; Schaepman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 165–193.
7. Delegido, J.; Verrelst, J.; Meza, C.; Rivera, J.; Alonso, L.; Moreno, J. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *Eur. J. Agron.* **2013**, *46*, 42–52.
8. Gianquinto, G.; Orsini, F.; Fecondini, M.; Mezzetti, M.; Sambo, P.; Bona, S. A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield. *Eur. J. Agron.* **2011**, *35*, 135–143.
9. Baret, F.; Hagolle, O.; Geiger, B.; Bicheron, P.; Miras, B.; Huc, M.; Berthelot, B.; Niço, 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.
10. Myneni, R.; Hoffman, S.; Knyazikhin, Y.; Privette, J.; Glassy, J.; Tian, Y.; Wang, Y.; Song, X.; Zhang, Y.; Smith, G.; *et al*. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* **2002**, *83*, 214–231.
11. Ganguly, S.; Nemani, R.; Zhang, G.; Hashimoto, H.; Milesi, C.; Michaelis, A.; Wang, W.; Votava, P.; Samanta, A.; Melton, F.; *et al*. Generating global leaf area index from landsat: Algorithm formulation and demonstration. *Remote Sens. Environ.* **2012**, *122*, 185–202.
12. Dash, J.; Curran, P. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413.
13. Bacour, C.; Baret, F.; Béal, D.; Weiss, M.; Pavageau, K. Neural network estimation of LAI, fAPAR, fCover and LAI $\times$ Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens. Environ.* **2006**, *105*, 313–325.

14. Houborg, R.; Boegh, E. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sens. Environ.* **2008**, *112*, 186–202.
15. Houborg, R.; Anderson, M.; Daughtry, C. Utility of an image-based canopy reflectance modeling tool for remote estimation of LAI and leaf chlorophyll content at the field scale. *Remote Sens. Environ.* **2009**, *113*, 259–274.
16. Stuffler, T.; Kaufmann, C.; Hofer, S.; Förster, K.; Schreier, G.; Mueller, A.; Eckardt, A.; Bach, H.; Penné, B.; Benz, U.; *et al*. The EnMAP hyperspectral imager-An advanced optical payload for future applications in Earth observation programmes. *Acta Astronaut.* **2007**, *61*, 115–120.
17. Labate, D.; Ceccherini, M.; Cisbani, A.; De Cosmo, V.; Galeazzi, C.; Giunti, L.; Melozzi, M.; Pieraccini, S.; Stagi, M. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronaut.* **2009**, *65*, 1429–1436.
18. Roberts, D.; Quattrochi, D.; Hulley, G.; Hook, S.; Green, R. Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) decadal survey mission. *Remote Sens. Environ.* **2012**, *117*, 83–101.
19. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604.
20. Knyazikhin, Y.; Kranigk, J.; Myneni, R.; Panfyorov, O.; Gravenhorst, G. Influence of small-scale structure on radiative transfer and photosynthesis in vegetation canopies. *J. Geophys. Res. D* **1998**, *103*, 6133–6144.
21. Durbha, S.; King, R.; Younan, N. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sens. Environ.* **2007**, *107*, 348–361.
22. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Mace, D.; Pragnère, A.; Myneni, R.; Knyazikhin, Y.; Wang, L. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* **2003**, *84*, 1–15.
23. Richter, K.; Atzberger, C.; Vuolo, F.; Weihs, P.; D'Urso, G. Experimental assessment of the Sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Can. J. Remote Sens.* **2009**, *35*, 230–247.
24. Weiss, M.; Baret, F.; Myneni, R.; Pragnère, A.; Knyazikhin, Y. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* **2000**, *20*, 3–22.
25. Fang, H.; Liang, S. A hybrid inversion method for mapping leaf area index from MODIS data: Experiments and application to broadleaf and needleleaf canopies. *Remote Sens. Environ.* **2005**, *94*, 405–424.
26. Qu, Y.; Wang, J.; Wan, H.; Li, X.; Zhou, G. A Bayesian network algorithm for retrieving the characterization of land surface vegetation. *Remote Sens. Environ.* **2008**, *112*, 613–622.
27. Walthall, C.; Dulaney, W.; Anderson, M.; Norman, J.; Fang, H.; Liang, S. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sens. Environ.* **2004**, *92*, 465–474.

28. Weiss, M.; Baret, F. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sens. Environ.* **1999**, *70*, 293–306.
29. Jacquemoud, S.; Baret, F.; Andrieu, B.; Danson, F.; Jaggard, K. Extraction of vegetation biophysical parameters by inversion of the PROSPECT+SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote Sens. Environ.* **1995**, *52*, 163–172.
30. Baret, F.; Buis, S. Estimating Canopy Characteristics from Remote Sensing Observations. Review of Methods and Associated Problems. In *Advances in Land Remote Sensing: System, Modeling, Inversion and Application*; Liang, S., Ed.; Springer: Berlin, Germany, 2008; pp. 173–201.
31. Combal, B.; Baret, F.; Weiss, M. Improving canopy variables estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies. *Agronomie* **2002**, *22*, 205–215.
32. Dorigo, W.; Richter, R.; Baret, F.; Bamler, R.; Wagner, W. Enhanced automated canopy characterization from hyperspectral data by a novel two step radiative transfer model inversion approach. *Remote Sens.* **2009**, *1*, 1139–1170.
33. Koetz, B.; Baret, F.; Poilvé, H.; Hill, J. Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sens. Environ.* **2005**, *95*, 115–124.
34. Richter, K.; Atzberger, C.; Vuolo, F.; D'Urso, G. Evaluation of Sentinel-2 spectral sampling for radiative transfer model based LAI estimation of wheat, sugar beet, and maize. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 458–464.
35. Darvishzadeh, R.; Matkan, A.A.; Dashti Ahangar, A. 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**, *PP*, 1–9.
36. Leonenko, G.; North, P.; Los, S. Statistical distances and their applications to biophysical parameter estimation: Information measures, M-estimates, and minimum contrast methods. *Remote Sens.* **2013**, *5*, 1355–1388.
37. Verrelst, J.; Rivera, J.; Leonenko, G.; Moreno, J. Optimizing LUT-based RTM inversion for semiautomatic mapping of crop biophysical parameters from Sentinel-2 and -3 data: Role of cost functions. *IEEE Trans. Geosci. Remote Sens.* **2013**, in press.
38. 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.
39. Chen, J.; Pavlic, G.; Brown, L.; Cihlar, J.; Leblanc, S.; White, H.; Hall, R.; Peddle, D.; King, D.; Trofymow, J.; *et al*. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sens. Environ.* **2002**, *80*, 165–184.
40. Kullback, S.; Leibler, R. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
41. Pardo, L. *Statistical Inference Based on Divergence Measures, Statistics: A Series Textbooks and Monographs*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006.
42. Staudte, R.; Sheather, S. *Robust Estimation and Testing*; Wiley: New York, NY, USA, 1990.

43. Taniguchi, M. On estimation of parameters of Gaussian stationary processes. *J. Appl. Probab.* **1979**, *16*, 575–591.
44. Gandía, S.; Fernández, G.; Garcia, J.C.; Moreno, J. Retrieval of Vegetation Biophysical Variables from CHRIS/PROBA Data in the SPARC Campaign. In Proceedings of the 2nd CHRIS/PROBAWorkshop, Frascati, Italy, 28–30 April 2004.
45. Fernández, G.; Moreno, J.; Gandía, S.; Martínez, B.; Vuolo, F.; Morales, F. Statistical Variability of Field Measurements of Biophysical Parameters in SPARC-2003 and SPARC-2004 Campaigns. In Proceedings of the SPARC Workshop WPP-250, Enschede, The Netherlands, 4–5 July 2005.
46. Drusch, M.; Del Bello, U.; 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.
47. Barnsley, M.J.; Settle, J.J.; Cutter, M.A.; Lobb, D.R.; Teston, F. The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1512–1520.
48. Alonso, L.; Moreno, J. Advances and Limitations in A Parametric Geometric Correction of CHRIS/PROBA Data. In Proceedings of the 3rd CHRIS/Proba Workshop, Frascati, Italy, 21–23 March 2005.
49. Guanter, L.; Alonso, L.; Moreno, J. A method for the surface reflectance retrieval from PROBA/CHRIS data over land: Application to ESA SPARC campaigns. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2908–2917.
50. Verrelst, J.; Rivera, J.; Alonso, L.; Moreno, J. ARTMO: An Automated Radiative Transfer Models Operator Toolbox for Automated Retrieval of Biophysical Parameters through Model Inversion. In Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK, 11–13 April 2011.
51. Verrelst, J.; Romijn, E.; Kooistra, L. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. *Remote Sens.* **2012**, *4*, 2866–2889.
52. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.; Asner, G.; François, C.; Ustin, S. PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66.
53. Feret, J.B.; François, C.; Asner, G.P.; Gitelson, A.A.; Martin, R.E.; Bidel, L.P.R.; Ustin, S.L.; le Maire, G.; Jacquemoud, S. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.* **2008**, *112*, 3030–3043.
54. Garrigues, S.; Lacaze, R.; Baret, F.; Morisette, J.; Weiss, M.; Nickeson, J.; Fernandes, R.; Plummer, S.; Shabanov, N.; Myneni, R.; *et al*. Validation and intercomparison of global leaf area index products derived from remote sensing data. *J. Geophys. Res.-Biogeosci.* **2008**, *113*, doi:10.1029/2007JG000635.
55. Pinty, B.; Lavergne, T.; Vobbeck, M.; Kaminski, T.; Aussedat, O.; Giering, R.; Gobron, N.; Taberner, M.; Verstraete, M.; Widlowski, J.L. Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products. *J. Geophy. Res.-Atmos.* **2007**, *112*; doi:10.1029/2006JD008105.

56. 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.
57. Clevers, J.; Gitelson, A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *ITC J.* **2013**, in press.
58. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* **2011**, *11*, 7063–7081.
59. Verrelst, J.; Muñoz, J.; Alonso, L.; Delegido, J.; Rivera, J.; 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.
60. Richter, K.; Hank, T.; 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.
61. Gitelson, A.; Merzlyak, M.; Lichtenthaler, H. Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* **1996**, *148*, 501–508.
62. Si, Y.; Schlerf, M.; Zurita-Milla, R.; Skidmore, A.; Wang, T. Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote Sens. Environ.* **2012**, *121*, 415–425.
63. Baret, F. Biophysical vegetation variables retrieval from remote sensing observations. *Proc. SPIE* **2010**, *7824*, 17–19.
64. Drusch, M.; Gascon, F.; Berger, M. Sentinel-2 Mission Requirements Document, 2010. Available online: http://esamultimedia.esa.int/docs/GMES/Sentinel-2_MRD.pdf (accessed on 27 April 2013).
65. Atzberger, C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sens. Environ.* **2004**, *93*, 53–67.
66. Meroni, M.; Colombo, R.; Panigada, C. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sens. Environ.* **2004**, *92*, 195–206.
67. Brown, L.; Chen, J.; Leblanc, S.; Cihlar, J. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sens. Environ.* **2000**, *71*, 16–25.

Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters

Juan Pablo Rivera Caicedo, Jochem Verrelst, Jordi Muñoz-Marí, *Member*, *IEEE*, José Moreno, *Member*, *IEEE*, and Gustavo Camps-Valls, *Senior Member*, *IEEE*

***Abstract*—Biophysical parameters such as leaf chlorophyll content (LCC) and leaf area index (LAI) are standard vegetation products that can be retrieved from Earth observation imagery. This paper introduces a new machine learning regression algorithms (MLRAs) toolbox into the scientific Automated Radiative Transfer Models Operator (ARTMO) software package. ARTMO facilitates retrieval of biophysical parameters from remote observations in a MATLAB graphical user interface (GUI) environment. The MLRA toolbox enables analyzing the predictive power of various MLRAs in a semiautomatic and systematic manner, and applying a selected MLRA to multispectral or hyperspectral imagery for mapping applications. It contains both linear and nonlinear state-of-the-art regression algorithms, in particular linear feature extraction via principal component regression (PCR), partial least squares regression (PLSR), decision trees (DTs), neural networks (NNs), kernel ridge regression (KRR), and Gaussian processes regression (GPR). The performance of multiple implemented regression strategies has been evaluated against the SPARC dataset (Barrax, Spain) and simulated Sentinel-2 (8 bands), CHRIS (62 bands) and HyMap (125 bands) observations. In general, nonlinear regression algorithms (NN, KRR, and GPR) outperformed linear techniques (PCR and PLSR) in terms of accuracy, bias, and robustness. Most robust results along gradients of training/validation partitioning and noise variance were obtained by KRR while GPR delivered most accurate estimations. We applied a GPR model to a hyperspectral HyMap flightline to map LCC and LAI. We exploited the associated uncertainty intervals to gain insight in the per-pixel performance of the model.**

***Index Terms*—Biophysical parameter retrieval, CHRIS, graphical user interface (GUI) toolbox, HyMap, leaf area index (LAI), leaf chlorophyll content (LCC), machine learning, nonparametric regression, Sentinel-2 (S2).**

I. Introduction

LEAF AREA INDEX (LAI) and leaf chlorophyll content (LCC) are essential land biophysical parameters retrievable from optical Earth observation (EO) data [1]–[3]. These parameters provide information about the phenological stage and health status (e.g., development, productivity, and stress) of crops and forests [4]. The quantification of these parameters from space over large areas has become an important aspect in agroecological, environmental, and climatic studies [5]. At the same time, remotely sensed observations are increasingly being applied at a within-field scale for dedicated agronomic monitoring applications [6]–[8]. With the forthcoming superspectral Sentinel-2 (S2) and Sentinel-3 missions and the planned EnMAP and PRISMA imaging spectrometers, the unprecedented data availability requires retrieval processing techniques that are accurate, robust, and fast to apply.

Biophysical parameter retrieval always requires an intermediate modeling step to transform the measurements into useful estimates [9]. This modeling step can be approached with either statistical, physical, or hybrid methods. In this paper, we will focus on the statistical approximation as this field has advanced largely over the last two decades [10], [11]. Statistical models can be categorized into either parametric or nonparametric approaches. *Parametric models* assume an explicit relation between the variables. They rely on the physical knowledge of the problem and build explicit parametrized expressions that relate a few spectral channels with the biophysical parameter of interest. Alternatively, *nonparametric models* are adjusted to predict a variable of interest using a training dataset of input–output data pairs, which come from concurrent measurements of the parameter and the corresponding reflectance/radiance observation. Several nonparametric regression algorithms are available in the statistics and machine learning literature, and recently they have been introduced for biophysical parameter retrieval [9], [12].

Particularly, the family of machine learning regression algorithms (MLRAs) emerged as a powerful nonparametric approach for delivering biophysical parameters. MLRAs have the potential to generate adaptive, robust relationships, and once trained, they are very fast to apply [13]. Typically, machine learning methods are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. They may therefore be suitable candidates for operational applications. Effectively, algorithms such as neural networks (NNs) are already implemented in operational retrieval chains (e.g., CYCLOPES products) [14], [15]. It still remains to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insight in the retrievals, and evaluating retrieval performances. Besides, training NNs involve tuning several parameters that may greatly impact the final robustness of the model. In part, this why in the recent years alternative and simpler to train regression methods have started replacing NNs. Specifically, the family of *kernel methods* [10] has emerged as an alternative to NNs in many

Manuscript received June 14, 2013; revised December 13, 2013; accepted December 31, 2013. This work was supported in part by the Spanish Ministry of Science and Innovation under project AYA2010-21432-C02-01, in part by the Spanish Ministry of Economy and Competitiveness (MINECO) under project TIN2012-38102-C03-01 (LIFE-VISION), and in part by ESA PARCS project (4000105078/11/NL/AF).

The authors are with the Image Processing Laboratory (IPL), Universitat de València, València 46980, Spain (e-mail: juan.rivera@uv.es; jochem.verrelst@uv.es; jordi.munoz; jose.moreno@uv.es; gustavo.camps@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/JSTARS.2014.2298752

scenarios. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input–output nonlinear mappings. Even though MLRAs are widely recognized as very powerful methods, some questions still remain open, e.g., how robust are these models in case of noisy situations, and how much they depend on changes between the training and testing data distributions. Perhaps more important is the fact that, for the broader remote sensing community, they are also perceived as complicated black boxes with several parameters to be tuned, which requires expertise. Further, until now no user-friendly graphical user interface (GUI) toolbox exists that brings several state-of-the-art MLRAs together. To facilitate and automate the use of MLRAs, in this work, we present a novel software package that allows systematically analyzing and applying MLRA-developed models. The so-called MLRA toolbox has been implemented within the in-house developed toolbox called Automated Radiative Transfer Models Operator (ARTMO). ARTMO is a scientific GUI toolbox dedicated to the retrieval of vegetation properties from optical imagery [16].

This brings us to the objectives of this work that are: 1) to present the novel MLRA toolbox for semiautomatic retrieval of biophysical parameters, 2) to evaluate the different MLRAs on their performance and robustness, and 3) to apply the best performing MLRA to EO imagery to test the robustness and accuracy in real scenarios.

The following sections will first briefly describe the considered nonparametric regression algorithms and then the latest status of ARTMO, followed by an introduction of the most important components of the new MLRA toolbox. The used data is subsequently described and an evaluation of six nonparametric regression methods is presented. A discussion on the use of these models for EO processing and a conclusion closes this paper.

II. Machine Learning Regression Algorithms

MLRAs learn the relationship between the input (e.g., reflectances) and output (e.g., biophysical parameters) by fitting a flexible model looking at the structure of the data. The hyperparameters of the model are typically adjusted to minimize the prediction error in an independent validation dataset. This way, one looks for the best *generalization* capabilities, not only a good performance in the training set that would give rise to an overfitted solution. In this paper, we compare several regression algorithms. A first family of linear methods follow a simple chained approach: first data dimensionality is applied to alleviate collinearity problems which is then followed by canonical least squares linear regression (LR). A second family of methods consists of building nonlinear functions of the data directly. Several state-of-the-art methods are considered here: 1) regression trees (RTs); 2) artificial NNs; 3) kernel ridge regression (KRR), also known as least squares support vector machine; and 4) Gaussian processes regression (GPR). All these regression techniques are popular in various application domains, thanks to its relatively fast training, good performance, and robustness to the overfitting problem. In the following sections, we briefly summarize them.

A. Dimensionality Reduction and LR

Let us consider a supervised regression problem, and let $\mathbf{X}$ and $\mathbf{y}$ be the input and output centered matrices of sizes $n \times d$ and $n \times l$, respectively. Here, l is the number of training data points in the problem and d is the data dimension. The objective of standard LR is to adjust a linear model for predicting the output variable from the input features $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}$, where $\mathbf{w}$ contains the regression model coefficients (weights) and has size $d \times 1$. Ordinary least-squares (OLS) regression solution is $\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$, where $\mathbf{X}^{\dagger} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}$ is the Moore–Penrose pseudoinverse of $\mathbf{X}$. Highly correlated input variables can result in rank-deficient covariance matrix $\mathbf{C}_{xx} = \frac{1}{n}\mathbf{X}^{\top}\mathbf{X}$, making the inversion unfeasible. The same situation is encountered in the small-sample-size case.

A common approach in statistics to alleviate these problems considers first in reducing the data dimensionality and then applying the OLS normal equations to the projected data or *scores* [17]. These scores reduce to a linear transformation of the original data $\mathbf{X}' = \mathbf{X}\mathbf{U}$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_f}]$ is referred to as the projection matrix, $\mathbf{u}_i$ being the ith projection vector and n_f the number of extracted features. The best known linear dimensionality reduction method is principal component analysis (PCA) [18] which reduces to solve the eigenproblem

$$\mathbf{C}_{xx}\mathbf{u} = \lambda\mathbf{u}.$$

An alternative supervised method is partial least squares (PLS) [19] in which we have to solve

$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{xy}^{\top} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

Note that PCA disregards the target data and exploits correlations between the input variables to maximize the variance of the projections, while PLS looks for projections that maximize the covariance between the features and the target data. In both cases, the user selects the dimensionality of the projected data n_f. After projection, the OLS equations are solved using $\mathbf{X}'$. The approaches respectively lead to the so-called principal component regression (PCR) [20] and the partial least squares regression (PLSR) [21] methods. Particularly, PLSR emerged as a popular regression technique for interpreting hyperspectral data with various experimental applications in vegetation properties mapping (e.g., [22]–[25]).

B. Regression Trees (RTs)

RTs build predictive models that take the observations as inputs and map them to the target variable. The model structure is made out of nodes (or leaves) and branches. Leaves represent output variable discrete values and branches constitute piecewise linear decisions. Decision tree (DT) learning can be done in several ways and using different algorithms, which mainly vary on the procedure used to determine where to split. In this paper, we focused on the standard Breiman's algorithm [26]. RTs have several advantages, among them: 1) they can manage a high number of features and examples in an easy way; 2) they are nonparametric flexible methods so they do not impose a specific functional form to the solution; and 3) the variables, or

combination of variables, used at each node to divide the samples into subgroups are the most discriminative features since they assure the lowest estimated error.

The main advantage for analysts in remote sensing applications is that RTs allow knowledge discovery and full interpretability by analyzing the surrogate and main splits of the tree. They have been successfully used to estimate land surface variables such as LAI, fraction of photosynthetically active radiation (FAPAR) and chlorophyll content from VEGETATION/SPOT4 [27], or broadband albedo from the Earth Observing 1 (EO-1) data [28], just to name a few applications.

C. Neural Networks (NNs)

The most common approach to develop nonparametric and nonlinear regression is based on artificial NNs [29]. An NN is a (potentially fully) connected structure of neurons organized in layers. A neuron basically performs a LR followed by a nonlinear function $f(\cdot)$. Neurons of different layers are interconnected with the corresponding links (weights). Therefore, in the limit case of using an NN with only one neuron, the results would be similar (or slightly better because of the nonlinearity) than those obtained with OLS regression. Training an NN implies selecting a structure (number of hidden layers and nodes *per* layer), initialize the weights, shape of the nonlinearity, learning rate, and regularization parameters to prevent overfitting. In addition, the selection of a training algorithm and the loss function both have an impact on the final model. In this work, we used the standard multi-layer perceptron, which is a fully connected network. We selected just one hidden layer of neurons. We optimized the NN structure using the Levenberg–Marquardt learning algorithm with a squared loss function. A cross-validation procedure was employed to avoid overfitting issues. NN weights were initialized randomly according to the Nguyen–Widrow method, and model regularization was done by limiting the maximum number of net weights to half the number of training samples. NNs have been vastly used in biophysical parameter retrieval (e.g., [14] and [30]–[33]), and are very useful in operational settings (e.g., [34]) because they scale well with the number of training examples.

D. Kernel Ridge Regression (KRR)

KRR minimizes the squared residuals in a higher dimensional feature space, and can be considered as the kernel version of the (regularized) OLS LR [35], [36]. The LR model is defined in a Hilbert space $\mathcal{H}$ of a very high dimensionality, where samples have been mapped to through a mapping $\phi(\mathbf{x}_i)$. In matrix notation, the model is given by $\hat{y}_i = \phi(\mathbf{x}_i)^{\mathrm{T}}\mathbf{w} + b$. Notationally, we want to solve a regularized OLS problem in Hilbert spaces

$$\sum_i (y_i - \hat{y}_i)^2 + \lambda \|\mathbf{w}\|^2. \tag{1}$$

Taking derivatives with respect to model weights $\mathbf{w}$ and b, and equating them to zero, leads to an equivalent problem depending on the unknown mapping function ϕ, which in principle is unknown. The problem can be solved by applying the Representer's theorem, by which the weights can be expressed as a linear combination of mapped samples, $\mathbf{w} = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)$. The prediction for a test sample $\mathbf{x}_*$ is obtained as a function of the dual weights $\alpha = [\alpha_1, \ldots, \alpha_n]^\top$ (one per sample), as follows:

$$\mathbb{E}[f(\mathbf{x}_*)] = \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y} = \mathbf{k}_{*,:}^\top \boldsymbol{\alpha} \tag{2}$$

where $\mathbf{k}_{*,:}$ contains the (kernel) similarities between the test example and all training data points. Note that for obtaining the model, only the inversion of the Gram (or kernel) matrix $\mathbf{K}$ of size $n \times n$ regularized by λ is needed. We have used the RBF kernel function, whose components $[\mathbf{K}]_{ij}$ are

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2)). \tag{3}$$

Therefore, in KRR only the regularization parameter λ and the kernel parameter σ have to be selected. Both parameters were optimized via standard cross-validation. It is worth noting that KRR has been recently used in remote sensing applications [10], [37].

E. Gaussian Processes Regression (GPR)

GPR has been recently introduced as a powerful regression tool [38] and applied to remote sensing data [39]–[42]. The model provides a *probabilistic* approach for learning generic regression problems with kernels. The GPR model establishes a relation between the input and the output variables (biophysical parameter) in the same way as KRR (see Eq. 2). However, two main advantages of GPR must be noted.

First, not only a *predictive mean* but also a *predictive variance* can be obtained

$$\mathbb{V}[f(\mathbf{x}_*)] = \mathbf{k}_{*,*} - \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{k}_{:,*}. \tag{4}$$

Note that the mean prediction in (2) is a linear combination of observations $\mathbf{y} = [y_1, \ldots, y_n]^\top$, whereas the predictive variance in (4) only depends on input data and can be taken as the difference between the prior kernel and the information given by observations about the approximation function.

The second advantage is that one can use very sophisticated kernel functions because hyperparameters can be learned efficiently by maximizing the marginal likelihood in the training set. See [38], [43], and [40] for further details. We used a scaled anisotropic Gaussian kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{b=1}^{B} \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2} \right) \tag{5}$$

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.

Summarizing, three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization give the relevance of each spectrum $\mathbf{x}_i$. Second, the inverse of σ_b represents the relevance of band b. Intuitively, high values of σ_b mean that relations largely extend along that band, hence suggesting a lower informative content. Finally, a GPR model provides not only a pixel-wise prediction for each

spectrum but also an uncertainty (or confidence) level for the prediction.

The previous methods are implemented from the simple regression toolbox [44], simpleR, freely available at http://www.uv.es/gcamps/code/simpleR.html. The simpleR toolbox contains simple educational code for LR, DTs (TREE), NNs, support vector regression (SVR), KRR, aka least squares SVM, Gaussian process regression (GPR), and variational heteroscedastic Gaussian process regression (VHGPR). The toolbox is not explicitly included in ARTMO, but may be of interest for the reader, as it provides more regression and analysis tools.

III. ARTMO

ARTMO brings multiple leaf and canopy radiative transfer models (RTMs) together along with essential tools required for semiautomatic retrieval of biophysical parameters in one GUI toolbox. In short, the toolbox permits the user: 1) to choose between various invertible leaf and canopy RTMs of a low to high complexity (e.g., PROSPECT-4, PROSPECT-5, DLM, 4SAIL, and FLIGHT); 2) to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems; 3) to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400–2500 nm); 4) to generate look-up tables (LUTs), which are stored in a relational SQL database management system (MySQL, version 5.5 or higher; local installment required), and finally; 5) to configure and run various retrieval scenarios using EO reflectance datasets for biophysical parameter mapping applications. ARTMO is developed in MATLAB (2009 version or higher) and does not require additional MATLAB toolboxes. Fig. 1 presents ARTMO v3's main window and a systematic overview of the drop-down menu below. To start with, in the main window, a new project can be initiated, a sensor chosen and a comment added, whereas all processing modules are accessible through drop-down menus at the top bar.

A first rudimentary version of ARTMO has been used in LUT-based inversion applications [45], [46]. ARTMO v3 is formally presented in this paper. The software package is freely downloadable at: http://ipl.uv.es/artmo. Its most important novelties are briefly listed below.

1) ARTMO v3 is designed modularly. Its modular architecture offers the possibility for easy addition (or removal) of components, such as RTM models and post-processing modules.
2) The MySQL database is organized in such a way that it supports the modular architecture of ARTMO v3. This avoids redundancy and increases the processing speed. For instance, all spectral datasets are stored as binary objects.
3) New retrieval toolboxes are incorporated. They are based on parametric and nonparametric regression as well as physically based inversion using a LUT. This has led to the development of a: 1) "Spectral Indices assessment toolbox" [47]; 2) "MLRA toolbox"; and 3) "LUT-based inversion toolbox" [48].

This paper introduces the "MLRA module." Its general architecture is outlined in Fig. 2.

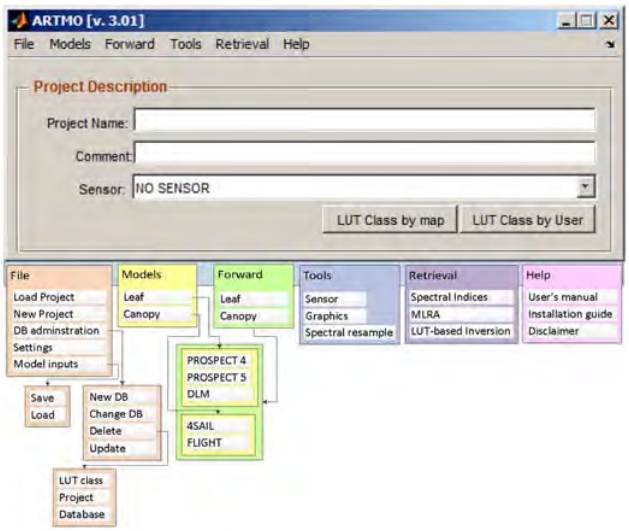


Fig. 1. Screenshot of ARTMO's main window and schematic overview of its drop-down menu.

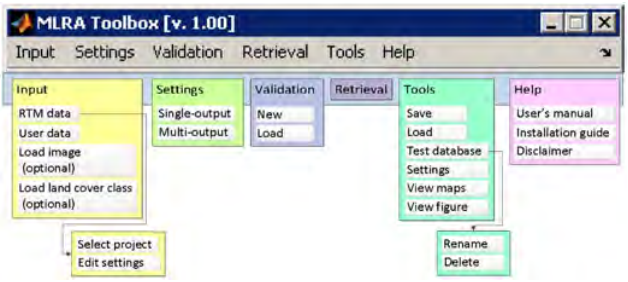


Fig. 2. Screenshot of MLRA's toolbox and schematic overview of its drop-down menu.

A. MLRA Settings Module

The following step addresses the analysis of multiple MLRA-based retrieval strategies. A first step to do is inserting input data (i.e., a plain text file), which refers to retrievable biophysical parameters and associated spectra. This is done in "Input" and can be either RTM-simulated data or can be ground truth data as measured in the field. The GUI will guide the user through the data selection steps and checks if data is properly read (not displayed for brevity). Once data is inserted the "MLRA settings" module can be configured (Fig. 3). It can be opted to select either single-output or multi-output regression algorithm. Currently, only PLSR, NN, and KRR encompass multi-output capabilities. Obviously, these models can also be used for single-output applications.

The "MLRA settings" module configures the regression algorithms given various options. First, if a land cover map in ENVI format (Exelis Inc.) has been provided then retrieval strategies can be configured *per* land cover class. Second, multiple regression algorithms at once can be selected, which means that they will be analyzed one-by-one. Third, options to add Gaussian noise are provided. Noise can be added both on the parameters to retrieve and on the spectra. A range of noise level can be configured, so that multiple noise scenarios can be evaluated.

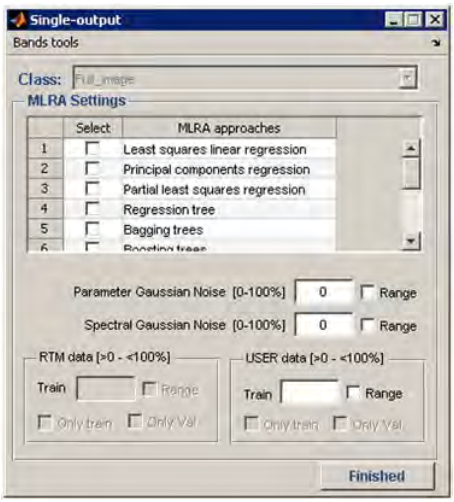


Fig. 3. MLRA's setting window.

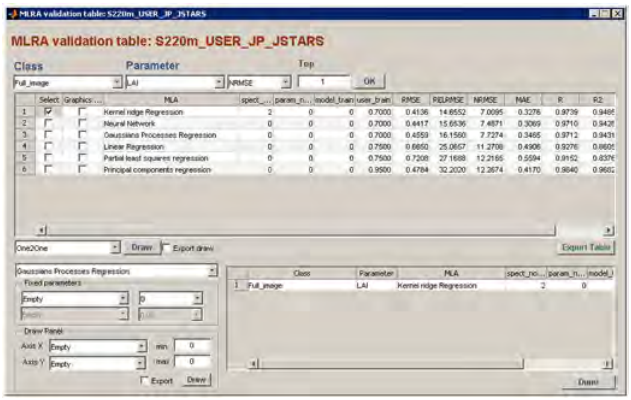


Fig. 4. MLRA's validation window.

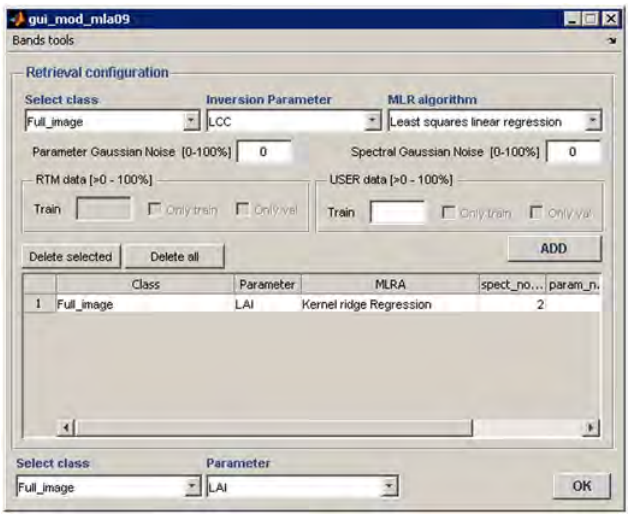


Fig. 5. MLRA's retrieval window.

The injection of noise can be of importance to account for environmental and instrumental uncertainties when synthetic spectra from RTMs are used for training. Fourth, the training/validation data partition can be controlled by setting the percentage of how much data from an RTM or user-defined is assigned to training or to validation (i.e., split-sample approach). Thereby, the user can evaluate the impact of ranging training/validation partitioning by entering a range of training/validation partitions. For each training/validation partition, the MLRA toolbox internally divides the defined training set into k subsets using a k-fold cross-validation strategy in order to tune the free parameters of the model.

B. Validation Module

Once that the training/validation data splitting has been defined and MLRA settings configured, a range of scenarios can be run, tested and their performance assessed. This is done by naming a validation set in the "Validation" module. Each regression model strategy over the configured ranges are one-by-one analyzed through goodness-of-fit measures and validation results are stored in a MySQL database. As such, a large number of results can be stored in a systematic manner, so that they can be easily queried and compared. Validation results are presented in the "MLRA validation table" (Fig. 4). The table shows the best performing validation results according to a selected land cover class (if loaded), parameter, and statical goodness-of-fit measure. Various options to display the results are provided, e.g., 1:1-line, plotting the band relevance as given by the GPR model, and 2-D matrices of performances along ranges of noise and varying training/validation distribution (see Section V). Finally, by clicking on "Retrieval," an analyzed regression function can be selected for each retrievable parameter (e.g., the best one). Such regression function will be accessed in the "Retrieval" GUI and can then be applied to a remote sensing image.

C. Retrieval Module

The "Retrieval" module enables to run an evaluated model or directly configure a model and apply it to an image (provided in standard ENVI file format) to map a parameter (Fig. 5). Hence, the user can select the required land cover class (if available), the retrievable parameter, the regression algorithms and training/validation splitting. Similarly, noise can be added to the spectra or parameters and the size of the training data can be selected. The user will then be invited to select one or multiple remote sensing images to which the developed model will be applied. Generated maps are stored in ENVI format.

IV. Mapping Applications

Having the MLRA toolbox presented, it is subsequently applied for evaluating the performance of the six presented nonparametric regression techniques to achieve optimized biophysical parameters estimation. Used data is first outlined, followed by the experimental setup. Results are then presented and a mapping application is shown.

A. Used Data

A diverse field dataset, covering various crop types, growing phases, canopy geometries, and soil conditions was collected

during SPARC (Spectra bARrax Campaign). The SPARC-2003 campaign took place during 12–14 July in Barrax, La Mancha, Spain (coordinates $30°3'\mathrm{N}$, $28°6'W$, 700 m altitude). Biophysical parameters were measured within a total of 110 elementary sampling units (ESUs) among different crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard, and wheat). ESU refers to a plot size compatible with pixel dimensions of about $20\ \mathrm{m} \times 20\ \mathrm{m}$. LCC was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter [49] Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyser. Each ESU was assigned to an LAI value, which was obtained as a statistical mean of 24 measures ($8\ \mathrm{data\ reading} \times 3\ \mathrm{replications}$) with standard errors between 5% and 10% [50]. In total, LAI varies between 0.4 and 6.3 and LCC between 2 and $55\ \mu\mathrm{g/cm}^2$. Additionally, 60 random spectra over bare soils, man-made surfaces, and water bodies were added to broaden the dataset to nonvegetation samples (i.e., with an LCC and LAI value of zero), leading to a total of 170 samples.

During the campaign airborne hyperspectral spaceborne CHRIS images and airborne HyMap flightlines were acquired. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, size of the covered area, and spatial resolution because of on-board memory storage reasons [51]. We made use of nominal nadir CHRIS observation in Mode 1 (62 bands, maximal spectral information), which were acquired during the SPARC campaign (12 July 2003). CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 6 to 33 nm depending on the wavelength. CHRIS imagery was processed using ESA's CHRIS-Box available in VISAT/BEAM, which includes radiometric recalibration, coherent-noise reduction, geometric correction and atmospheric correction [52], [53]. HyMap was configured with 125 bands between 430 and 2490 nm with bandwidths varying between 11 and 21 nm and a pixel size of 5 m. The same geometric and atmospheric preprocessing as for CHRIS was applied, but given a superior signal-to-noise ratio this sensor provides a better radiometric quality than CHRIS.

B. Experimental Setup

SPARC field data were used for training and validation, and associated spectral data came from CHRIS and HyMap. In view of ESA's forthcoming S2 mission, also S2 data at a spatial resolution of 20 m were additionally generated. S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades. It provides a set of 13 spectral bands spanning between 443 and 2190 nm, four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution [54]. Because of being spaceborne and providing similar pixel size, CHRIS data were resampled to the band settings of S2. Nearest neighbor was used for the spatial resampling and a Gaussian model with full width at half maximum (FWHM) spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, a dataset of eight bands at 20 m (4 bands at 20 m plus 4 bands at 10 m coarse-grained at 20 m) was prepared, hereafter referred as "S2-20 m."

The MLRA toolbox was used to evaluate the performance of the different regression algorithms along gradients of changing training/validation distributions (from 5% to 95% training, with steps of 5%; the remaining data go to validation) and increasing Gaussian noise levels (from 0% to 20% with steps of 2%). By systematically evaluating the performance along those two dimensions in a 2-D matrix format, an indication about the robustness of these regression methods can be obtained. Models were developed both for LCC and LAI. The predictive power of the developed models was evaluated with the absolute root-mean-squared error (RMSE), the normalized RMSE ($\mathrm{NRMSE}\ [\%] = \mathrm{RMSE}/\mathrm{range\ parameter\ measurements} * 100$) and the coefficient of determination (r^2) to account for the goodness-of-fit. Here, validation results are presented in the form of NRMSE, which allows accuracy comparison across different parameters. Typically, remote sensing end users require an error threshold below 10%.

V. Results

A. Regression Method Evaluation

For each parameter, sensor type and regression algorithm, NRMSE results along varying training/validation distribution and increasing noise levels are presented in 2-D matrices (Fig. 6). The best performing scenario for each matrix is also shown in Table I. When comparing these matrices, the following observations can be made.

Starting with PCR, this method proved to perform rather stable within the matrix space. For S2-20 m and CHRIS data, PCR seems to be hardly affected by a varying training/validation partition and noise injection. In fact, injection of some noise rather improved accuracies of CHRIS and HyMap. Hence, adding noise can lead to a closer match between training and validation data. However, results were never outstanding, and LCC prediction with HyMap data completely failed. Only for HyMap LAI results improved to r^2 up to 0.97 when 95% of the data were assigned to the training process. Therefore, on the whole, PCR is evaluated as suboptimal performing.

Second, the partial least square regression (PLSR) is an improved version of the PCR and widely used in EO applications. It systematically outperforms PCR in absence of noise. Improvements are particularly notable for LCC (r^2 up to 0.96 for CHRIS and HyMap). But PLSR is also more affected by the injection of noise. Low noise levels led to superior results, but above about 8% accuracies degraded rapidly.

Third, DTs yielded on the whole poorest results. Particularly, unacceptable poor results were obtained with low training data, and when many bands are involved. This suggests that DTs would not be a good choice for applying to hyperspectral data unless a large database is available. In fact, only good results were obtained [r^2 up to 0.94 (LCC) and 0.97 (LAI)] in case of "S2-20 m" (8 bands) when more than 80% was used for training and below 8% noise added.

Fourth, NNs are characterized by causing erratic patterns in each of the matrices. While being able to deliver very accurate results in some cases (e.g., for LAI using CHRIS and HyMap: r^2 up to 0.96 and 0.99) NN also showed to perform rather unstable, with large probability of delivering poor results. Particularly,

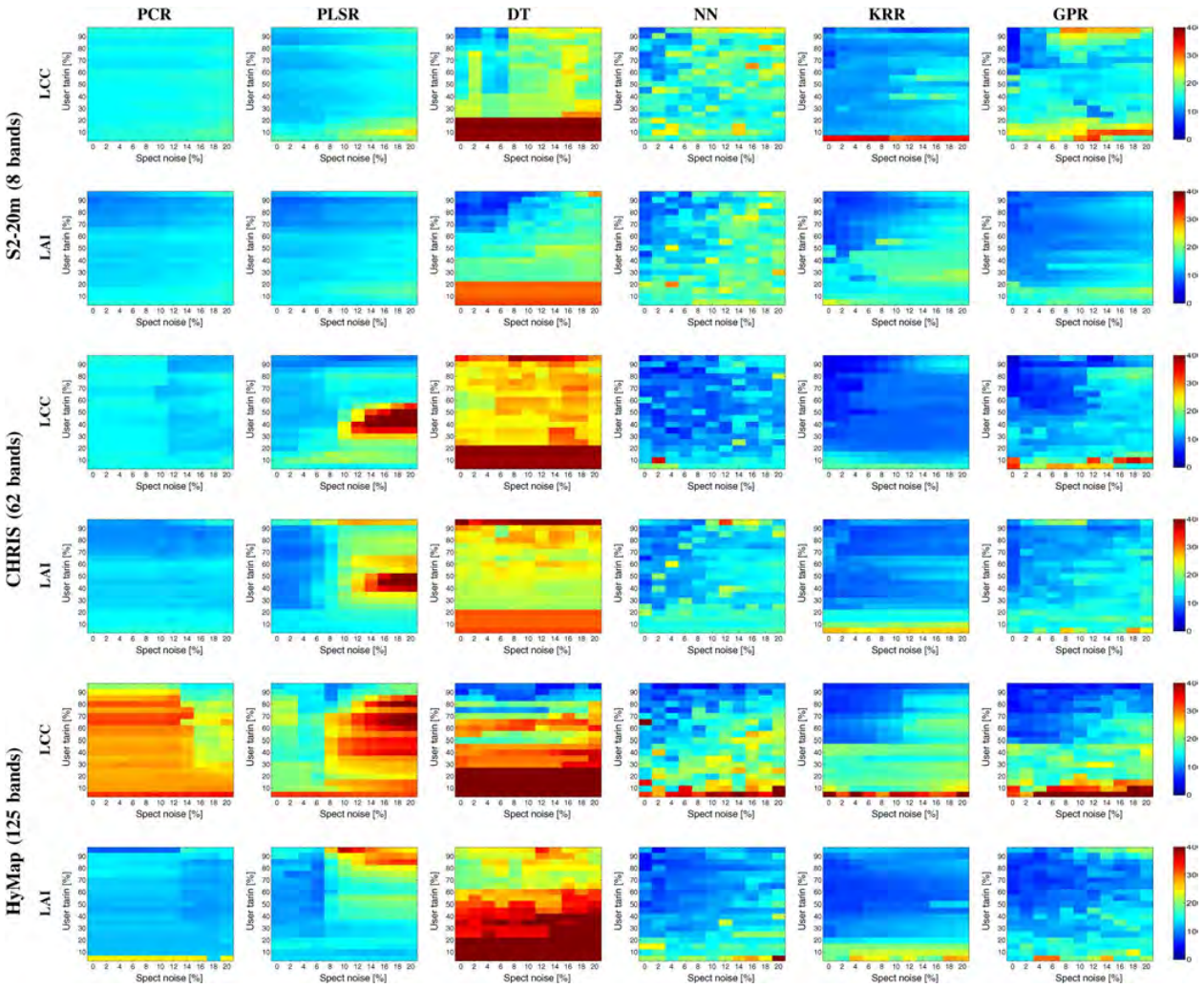


Fig. 6. NRMSE matrices of validation results for LCC and LAI retrieval using a regression algorithm displaying the impact of percentage noise (X-axis) against multiple solutions (Y-axis). The more bluish, the lower relative errors and thus the better the retrieval.

when more noise is involved and when less data are dedicated for training NN tends to perform more unstable. This erratic behavior can be explained by the complicated training phase whereby a highly specialized model is developed, but therefore easily faces the problem of overfitting. The lack of robustness to noise along with the complexity in training are therefore major drawbacks of NN.

Fifth, from all evaluated regression algorithms, KRR yielded most robust results. It led to excellent accuracies with r^2 maxima between 0.94 and 0.99 for all datasets, and more importantly also proved to perform very stable with increasing noise levels. Also, it should be noted that from all tested nonlinear MLRAs, this regression technique is fastest in developing their models (see [12] for a quantitative comparison).

Finally, GPR appears to be the most promising regression algorithm. It easily leads to excellent performances, with r^2 maxima between 0.94 and 0.99 for all datasets. Though, in comparison to KRR, GPR is somewhat more affected by noise injection. Note that the predictive mean equations for KRR and GPR are exactly the same so in principle the results should be exactly identical. Nevertheless, in GPR, we used a very flexible kernel to account for different lengthscales per feature. Although beneficial without noise, this turns to be a curse in these particular experiments because noise affected the marginal likelihood estimation of hyperparameters. For this reason, it can be concluded that GPR performs slightly less robust than KRR.

B. Biophysical Parameter Mapping

The developed models can be applied to any EO imagery given the same band settings as those presented during the training phase. From all considered regression techniques, GPR was evaluated as reaching highest accuracies for the majority of cases. Moreover, GPR has unique additional features: 1) it reveals most relevant bands when developing the model and 2) it provides uncertainty intervals (σ) associated with the mean predictions (μ). Therefore the consecutive approach was to apply GPR to a hyperspectral HyMap imagery for mapping LCC and LAI estimates along with associated uncertainty intervals (Fig. 7). In the obtained μ maps, the irrigated circular agricultural fields are clearly differentiated, including within-field variability.

TABLE I
VALIDATION STATISTICS [NOISE LEVEL (%), TRAINING (%), RMSE, NRMSE, AND r^2] SORTED ACCORDING TO BEST EVALUATED NRMSE FOR LCC AND LAI RETRIEVAL

		Regression algorithm	Noise (%)	Training (%)	RMSE	NRMSE (%)	r^2
S2-20m (8 bands)	LCC	Principal component regression	4	65	7.11	13.79	0.87
		Partial least square regression	0	85	5.58	10.90	0.92
		Decision tree	6	90	4.56	8.95	0.94
		Neural network	2	95	3.94	7.72	0.96
		Kernel ridge regression	0	75	3.48	6.82	0.97
		Gaussian processes regression	**0**	**90**	**2.40**	**4.71**	**0.98**
	LAI	Principal component regression	0	90	0.50	9.30	0.93
		Partial least square regression	0	90	0.45	8.36	0.95
		Decision tree	6	85	0.33	6.09	0.97
		Neural network	0	90	0.39	7.17	0.96
		Kernel ridge regression	**2**	**90**	**0.36**	**6.03**	**0.98**
		Gaussian processes regression	0	80	0.38	7.03	0.95
CHRIS (62 bands)	LCC	Principal component regression	12	90	5.68	11.14	0.91
		Partial least square regression	12	95	4.07	7.99	0.96
		Decision tree	0	55	10.59	20.51	0.71
		Neural network	0	95	1.94	3.81	0.99
		Kernel ridge regression	0	85	1.95	3.81	0.99
		Gaussian processes regression	**0**	**95**	**1.06**	**3.01**	**0.99**
	LAI	Principal component regression	16	85	0.50	9.30	0.92
		Partial least square regression	4	70	0.46	8.20	0.93
		Decision tree	6	85	1.05	13.32	0.68
		Neural network	**0**	**65**	**0.37**	**6.50**	**0.96**
		Kernel ridge regression	0	95	0.25	6.70	0.97
		Gaussian processes regression	0	90	0.37	6.88	0.95
HyMap (125 bands)	LCC	Principal component regression	16	85	6.48	13.18	0.89
		Partial least square regression	8	85	3.84	7.82	0.96
		Decision tree	12	90	2.25	4.69	0.98
		Neural network	4	95	1.19	3.60	0.99
		Kernel ridge regression	0	75	2.28	4.64	0.98
		Gaussian processes regression	**0**	**80**	**1.73**	**3.51**	**0.99**
	LAI	Principal component regression	8	95	0.39	8.26	0.97
		Partial least square regression	6	70	0.55	8.83	0.88
		Decision tree	6	85	0.54	10.17	0.88
		Neural network	**2**	**95**	**0.23**	**4.74**	**0.99**
		Kernel ridge regression	0	60	0.42	6.78	0.94
		Gaussian processes regression	0	80	0.30	5.66	0.95

Best NRMSE result per sensor and biophysical parameter is bold typed.

In the uncertainty maps, the lower the σ (whiter color) indicate the more certain the retrieval as processed by the trained model. The delivery of uncertainty estimates allows us to provide insight on a pixelwise basis when applied to any image and so enables the interpretation at which land covers retrievals were associated with great certainty and land covers would benefit from additional sampling. It can be observed that, particularly over the circular agricultural parcels, LCC was processed with high certainty. This is less obvious for LAI retrievals; however, it should be kept in mind that σ is also related to the magnitude of the mean estimates (μ). For this reason, relative uncertainties (σ/μ) may provide a more meaningful interpretation. These maps can function as a spatial mask that enables displaying only pixels with great certainty. Moreover, uncertainty maps can also give information about the portability of the regression models when applied to images over areas other than the training site [41], [42].

VI. DISCUSSION

The hereby presented MLRA toolbox allows evaluating and applying a wide range of regression techniques in a semiautomatic and user-friendly way. As a case study, we applied the MLRA toolbox to compare six regression algorithms on their performance and robustness along ranges of varying training/validation distribution and noise variance. These algorithms can be categorized in either data dimensionality transformations (PCR, PLSR) and nonlinear algorithms (DT, NN, KRR, GPR). For all used datasets (S2-20 m, CHRIS and HyMap) pronounced differentiation in their best performances emerged. While for PCR, PLSR and DT best accuracies fell within a range of 4.7%–20.5% (r^2: 0.86–0.97) the MLRA algorithms NN, KPR and GPR yielded higher accuracies between 3.5% and 7.7% (r^2: 0.94–0.99). Hence, each of these MLRAs reached accuracies below 10%, which is typically demanded in operational products. These excellent performances can be explained by that MLRAs may find the nonlinear feature relations by building more flexible and adaptive models than those restricted to linear projections or regression. The excellent performance of the MLRAs becomes even more apparent when comparing against other classic retrievals methods. The same validation dataset reached r^2 on the order of 0.85 by using vegetation indices from CHRIS data [40], and r^2 up to 0.77 for the same S2-20 m bands by using LUT inversion of the PROSAIL radiative transfer (RT) model through cost functions [46]. Moreover, apart from the evaluated algorithms here, others can be added relatively easily. Meanwhile, a

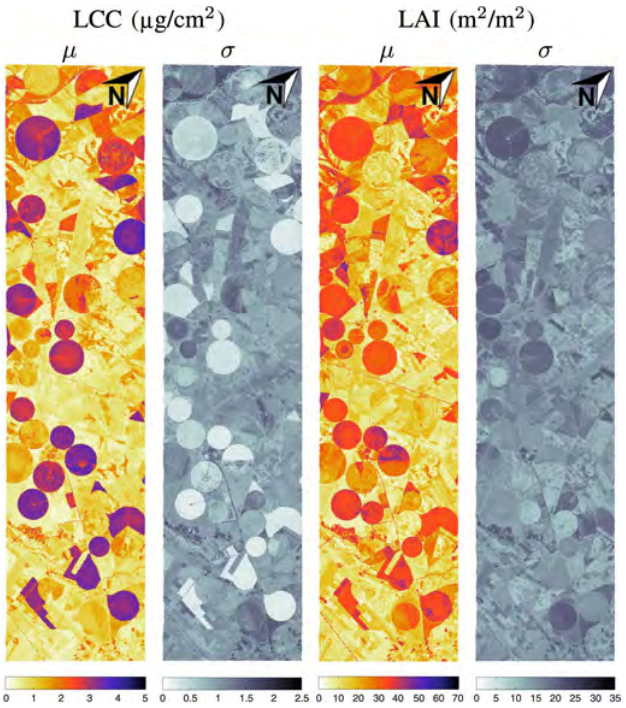


Fig. 7. LCC ($\mu g/cm^2$) and LAI (m^2/m^2) mean estimates (μ) and associated uncertainties (σ) for a HyMap flightline over Barrax agricultural area (Spain).

wide array of new MLRAs have already been implemented, among others: SVR, extreme learning machines (ELMs), and VHGPR [44], [55].

An urging open question is about evaluating how well these algorithms perform when being fed by large datasets as generated by canopy RT models. The advantage of RT models is that a broad range of land cover situations can be simulated (e.g., up to hundred thousands), leading to a dataset several times bigger than what can be collected during a field campaign. Operational processing chains typically rely on this hybrid approach [15], and similar strategies could be developed by the ARTMO toolbox. It remains, however, to be investigated how well kernel-based MLRAs perform with large datasets. This is not a trivial point. For instance, the computational load of the GPR increases exponentially with each added sample, making that this function faces difficulties when being trained by several thousand (in principle distinct) samples. Alternatively, dimensionality reduction techniques may largely overcome the burden of large datasets. Currently, a diversity of linear and nonlinear PCA techniques are being implemented (e.g., kernel PCA) in order to apply dimensionality reduction. On the other hand, redundancy also takes place along the simulated spectra, e.g., because not all RTM parameters lead to spectral variations, causing redundancy along the dataset. Therefore, the emerging field of redundant data reduction is expected to further reduce the dataset while preserving good performance [56], [57]. The field is also related to active learning approaches [58]. These dimensionality reduction techniques are foreseen to be implemented as well, which will eventually facilitate a smooth coupling between RTM-generated simulated spectra and powerful MLRAs for generic and operational retrieval applications.

VII. Conclusion

ARTMO's new "MLRA toolbox" enables applying and analyzing the predictive power of various MLRAs in a semi-automatic manner. Various regularization options have been implemented into the toolbox, e.g., training/validation data splitting, adding noise, and regression models can be developed and evaluated per land cover class. Data can either come from field campaigns or from simulations as generated by RTMs. The predictive power of multiple nonparametric regression algorithms was evaluated across gradients of varying training/validation distribution and increasing noise levels. By using the local SPARC dataset and multispectral simulated S2 and hyperspectral CHRIS and HyMap imagery over the Barrax (Spain) agricultural area, KRR and GPR emerged as most robust and best performing regression algorithms (r^2 up to 0.94–0.99 and NRMSE down to 7.0%–3.0%). Moreover, GPR provides additional uncertainty estimates on a pixelwise basis, which provides insight in the performance of the model. In all generality, the linear nonparametric algorithms such as the popular PLSR performed systematically poorer than the nonlinear, kernel-based regression algorithms (KRR, GPR).

The presented experimental results demonstrated the utility of the MLRA toolbox, which essentially has been developed to serve efficient and optimized surface properties mapping.

Acknowledgment

The authors would like to thank the three anonymous reviewers for their valuable comments.

References

[1] R. H. Whittaker and P. L. Marks, "Methods of assessing terrestrial productivity," in *Primary Productivity of the Biosphere*, New York, NY, USA: Springer, 1975, pp. 55–118.
[2] H. K. Lichtenthaler, "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," *Methods Enzymol.*, vol. 148, pp. 350–382, 1987.
[3] Z. Malenovský, H. Rott, J. Cihlar, M. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, "Sentinels for science: Potential of sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, 2012.
[4] H. K. Lichtenthaler, M. Lang, M. Sowinska, F. Heisel, and J. A. Miehe, "Detection of vegetation stress via a new high resolution fluorescence imaging system," *J. Plant Physiol.*, vol. 148, no. 5, pp. 599–612, 1996.
[5] W. A. Dorigo, R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman, "A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 2, pp. 165–193, 2007.
[6] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.
[7] J. Delegido, J. Verrelst, C. Meza, J. Rivera, L. Alonso, and J. Moreno, "A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems," *Eur. J. Agronomy*, vol. 46, pp. 42–52, 2013.
[8] G. Gianquinto, F. Orsini, M. Fecondini, M. Mezzetti, P. Sambo, and S. Bona, "A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield," *Eur. J. Agronomy*, vol. 35, no. 3, pp. 135–143, 2011.
[9] F. Baret and S. Buis, "Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems," in *Advances in Land Remote Sensing: System, Modeling, Inversion and Application*, New York, NY, USA: Springer, 2008, pp. 172–301.
[10] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*. New York, NY, USA: Wiley, 2009.
[11] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, "Remote sensing image processing," in *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 12, pp. 1–194, 2012.

[12] J. Verrelst, J. Muñoz, L. Alonso, J. Delegido, J. Rivera, G. Camps-Valls, and J. Moreno, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3," *Remote Sens. Environ.*, vol. 118, pp. 127–139, 2012.

[13] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

[14] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and $\mathrm{LAI} \times \mathrm{Cab}$, from top of canopy MERIS reflectance data: Principles and validation," *Remote Sens. Environ.*, vol. 105, no. 4, pp. 313–325, 2006.

[15] A. Verger, F. Baret, and M. Weiss, "Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2789–2803, 2008.

[16] J. Verrelst, J. Rivera, L. Alonso, and J. Moreno, "ARTMO: An automated radiative transfer models Operator toolbox for automated retrieval of biophysical parameters through model inversion," in *Proc. EARSeL 7th SIG-Imag. Spectrosc. Workshop*, Edinburgh, U.K., 2011.

[17] J. Arenas-García, K. B. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 16–29, 2013.

[18] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[19] H. Wold, "Partial least squares," *Encyclopedia Statist. Sci.*, vol. 6, pp. 581–591, 1985.

[20] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.

[21] P. Geladi and B. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chim. Acta*, vol. 185, no. C, pp. 1–17, 1986.

[22] N. C. Coops, M.-L. Smith, M. Martin, and S. V. Ollinger, "Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1338–1346, Jun. 2003.

[23] D. Gianelle and F. Guastella, "Nadir and off-nadir hyperspectral field data: Strengths and limitations in estimating grassland biophysical characteristics," *Int. J. Remote Sens.*, vol. 28, no. 7, pp. 1547–1560, 2007.

[24] M. A. Cho, A. K. Skidmore, F. Corsi, S. E. van Wieren, and I. Sobhan, "Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 4, pp. 414–424, 2007.

[25] X. Ye, K. Sakai, M. Manago, S.-I. Asada, and A. Sasao, "Prediction of citrus yield from airborne hyperspectral imagery," *Precis. Agric.*, vol. 8, no. 3, pp. 111–125, 2007.

[26] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA, USA: Wadsworth and Brooks, 1984.

[27] M. Weiss and F. Baret, "Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data," *Remote Sens. Environ.*, vol. 70, pp. 293–306, 1999.

[28] S. Liang, H. Fang, M. Kaul, T. G. Van Niel, T. R. Mcvicdr, J. Pearlman *et al.*, "Estimation of land surface broadband albedos and leaf area index from EO-1 ALI data and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1260–1268, Jun. 2003.

[29] S. Haykin, *Neural Networks—A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, Oct. 1999.

[30] J. Smith, "LAI inversion using backpropagation neural network trained with multiple scattering model," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 5, pp. 1102–1106, Sep. 1993.

[31] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 398–404, Mar. 1996.

[32] D. S. Kimes, R. F. Nelson, M. T. Manry, and A. K. Fung, "Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2639–2662, 1998.

[33] A. Verger, F. Baret, and F. Camacho, "Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 415–426, 2011.

[34] F. Baret, M. Weiss, R. Lacaze, F. Camacho, H. Makhmara, P. Pacholcyzk, and B. Smets, "GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 1: Principles of development and production," *Remote Sens. Environ.*, vol. 137, pp. 299–309, 2013.

[35] B. Schölkopf and A. Smola, *Learning with Kernels—Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[36] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1759–1769, May 2012.

[38] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. New York, NY, USA: MIT Press, 2006.

[39] L. Pasolli, F. Melgani, and E. Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 464–468, Jul. 2010.

[40] 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, pt. 2, pp. 1832–1843, May 2012.

[41] J. Verrelst, L. Alonso, J. Rivera Caicedo, J. Moreno, and G. Camps-Valls, "Gaussian process retrieval of chlorophyll content from imaging spectroscopy data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 867–874, Apr. 2013.

[42] J. Verrelst, J. 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.

[43] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, and J. Calpe-Maravilla, "Biophysical parameter estimation with adaptive Gaussian Processes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* (IGARSS'2009), Capetown, South Africa, 2009.

[44] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, M. Lázaro-Gredilla, and J. Verrelst. (Jun. 2013). "simpleR: A simple educational MATLAB toolbox for statistical regression, v 2.1" [Online]. Available: http://www.uv.es/gcamps/code/simpleR.html.

[45] J. Verrelst, E. Romijn, and L. Kooistra, "Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data," *Remote Sens.*, vol. 4, no. 9, pp. 2866–2889, 2012.

[46] J. Verrelst, J. Rivera, G. Leonenko, L. Alonso, and J. Moreno, "Optimizing LUT-based RTM inversion for semiautomatic mapping of crop biophysical parameters from Sentinel-2 and -3 data: Role of cost functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 257–269, Jan. 2014.

[47] J. Rivera, J. Verrelst, J. Delegido, F. Veroustraete, and J. Moreno, "On the semi-automatic retrieval of biophysical parameters based on spectral index optimization," *Remote Sens.,* 2014 (in review).

[48] J. Rivera, J. Verrelst, G. Leonenko, and J. Moreno, "Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the prosail model," *Remote Sens.*, vol. 5, no. 7, pp. 3280–3304, 2013.

[49] S. Gandía, G. Fernández, J. C. Garcia, and J. Moreno, "Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign," in *Proc. 2nd CHRIS/PROBAWorkshop*, Frascati, Italy: ESA/ESRIN, 2004.

[50] G. Fernández, J. Moreno, S. Gandía, B. Martínez, F. Vuolo, and F. Morales, "Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns," in *Proc. SPARC Workshop*, Frascati, Italy: ESA/ESRIN, 2005.

[51] M. J. Barnsley, J. J. Settle, M. A. Cutter, D. R. Lobb, and F. Teston, "The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1512–1520, Jul. 2004.

[52] L. Alonso and J. Moreno, "Advances and limitations in a parametric geometric correction of CHRIS/PROBA data," in *Proc. 3rd CHRIS/Proba Workshop*, Frascati, Italy: ESA/ESRIN, 2005.

[53] L. Guanter, L. Alonso, and J. Moreno, "A method for the surface reflectance retrieval from PROBA/CHRIS data over land: Application to ESA SPARC campaigns," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2908–2917, Dec. 2005.

[54] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[55] 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.

[56] X. Shen, H. Wu, and Q. Zhu, "Training support vector machine through redundant data reduction," in *ACM Int. Conf. Proc. Ser.,* 2012 pp. 25–28.

[57] S. Wang, Z. Li, and X. Zhang, "Bootstrap sampling based data cleaning and maximum entropy SVMs for large datasets," in *Proc. Int. Conf. Tools Artif. Intell.* (ICTAI), 2012, vol. 1, pp. 1151–1156.

[58] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 4, pp. 606–617, 2011.

Juan Pablo Rivera Caicedo received the B.Sc. degree in agricultural engineering from the National University of Colombia, Bogotá, Colombia and University of Valle, Cali, Colombia, in 2001, the Master degree in irrigation engineering from the Centro de Estudios y Experimentación de Obras Publicas (CEDEX), Spain, in 2003, and the M.Sc. degree in remote sensing from the University of Valencia, Valencia, Spain.

Currently, he is involved in preparatory activities of the Fluorescence Explorer (FLEX). His research interests include retrieval of vegetation properties using airborne and satellite data, leaf and canopy radiative transfer modeling, and hyperspectral data analysis. His research activities led to the scientific toolbox ARTMO: http://ipl.uv.es/artmo.

Since January 2011, he is a member of the Laboratory for Earth Observation (LEO), the Image Processing Laboratory, University of Valencia, as a Ph.D. candidate.

Jochem Verrelst received the M.Sc. degree in tropical land use and geo-information science, and the Ph.D. degree in remote sensing from Wageningen University, Wageningen, Netherlands, in 2005 and 2010, respectively. The subject of his dissertation was on the space-borne spectrodirectional estimation of forest properties.

From 2010 to 2012, he has been a Marie Curie Postdoctoral Fellow at the Laboratory for Earth Observation (LEO), the Image Processing Laboratory, University of Valencia, Valencia, Spain. Currently, he is involved in preparatory activities of the Fluorescence Explorer (FLEX). His research interests include retrieval of vegetation properties using airborne and satellite data, canopy radiative transfer modeling, multi-angular and hyperspectral data analysis. His research activities led to the scientific toolbox ARTMO: http://ipl.uv.es/artmo.

Jordi Muñoz-Marí (M'11) received the B.Sc. degree in physics, the M.Sc. degree in electronics engineering, and the Ph.D. degree in electronics engineering (2003) from the Universitat de València, València, Spain, in 1993, 1996, and 2003, respectively.

Currently, he is an Associate Professor with the Electronics Engineering Department, Universitat de València, where he teaches electronic circuits and, programmable logical devices, digital electronic systems, and microprocessor electronic systems. His research interests are mainly related to the development of machine learning algorithms for signal and image processing. Visit http://www.uv.es/jordi/ for more information.

José Moreno (A'89–M'09) is currently with the Department of Earth Physics and Thermodynamics, Faculty of Physics of the University of Valencia, Valencia, Spain, as a Professor of earth physics, teaching and working on different projects related to remote sensing and space research as responsible for the Laboratory for Earth Observation. His main work is related to the modeling and monitoring of land surface processes by using remote sensing techniques. He has been involved in many international projects and research networks, including the preparatory activities and exploitation programmes of several satellite missions (ENVISAT, CHRIS/PROBA, GMES/Sentinels, and SEOSAT) and the Fluorescence Explorer (FLEX), and as a Candidate ESA Earth Explorer Mission. He is a Director of the Laboratory for Earth Observation (LEO) at the Image Processing Laboratory/Scientific Park.

Dr. Moreno has served as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (1994–2000) and has been a Member of the ESA Earth Sciences Advisory Committee (1998–2002), the Space Station Users Panel, and other international advisory committees.

Gustavo Camps-Valls (M'04–SM'07) received the Ph.D. degree in physics from the Universitat de València, Valencia, Spain, in 2002.

Currently, he is an Associate Professor with the Department of Electronics Engineering and a Leading Researcher with the Image Processing Laboratory (IPL), Universitat de València. Recently included in the ISI lists as a highly cited researcher. He co-edited the books "Kernel methods in bioengineering, signal and image processing" (IGI, 2007) and "Kernel methods for remote sensing data analysis" (Wiley & Sons, 2009).

Dr. Camps-Valls serves on the Program Committees of International Society for Optical Engineers (SPIE) Europe, International Geoscience and Remote Sensing Symposium (IGARSS), Machine Learning for Signal Processing (MLSP), and International Conference on Image Processing (ICIP). Since 2007, he is a Member of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society and since 2009, he is a Member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society. He is involved in the EUMETSAT MTG-IRS Science Team. He is an Associate Editor of the "IEEE TRANSACTIONS ON SIGNAL PROCESSING," "IEEE SIGNAL PROCESSING LETTERS," and "IEEE GEOSCIENCE AND REMOTE SENSING LETTERS." Visit http://www.uv.es/gcamps for more information.

ISPRS Journal of Photogrammetry and Remote Sensing 86 (2013) 157–167

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval

Jochem Verrelst *, Juan Pablo Rivera, José Moreno, Gustavo Camps-Valls

Image Processing Laboratory (IPL), Parc Científic, Universitat de València, 46980 Paterna, València, Spain

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form 24 September 2013
Accepted 27 September 2013
Available online 29 October 2013

Keywords:
Uncertainty estimates
Gaussian processes regression
Biophysical parameters
Sentinel-2
Chlorophyll content
Leaf area index

ABSTRACT

ESA's upcoming Sentinel-2 (S2) Multispectral Instrument (MSI) foresees to provide continuity to land monitoring services by relying on optical payload with visible, near infrared and shortwave infrared sensors with high spectral, spatial and temporal resolution. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods, which ideally should provide uncertainty intervals for the predictions. Statistical learning regression algorithms are powerful candidats for the estimation of biophysical parameters from satellite reflectance measurements because of their ability to perform adaptive, nonlinear data fitting. In this paper, we focus on a new emerging technique in the field of Bayesian nonparametric modeling. We exploit Gaussian process regression (GPR) for retrieval, which is an accurate method that also provides uncertainty intervals along with the mean estimates. This distinct feature is not shared by other machine learning approaches. In view of implementing the regressor into operational monitoring applications, here the portability of locally trained GPR models was evaluated. Experimental data came from the ESA-led field campaign SPARC (Barrax, Spain). For various simulated S2 configurations (S2-10m, S2-20m and S2-60m) two important biophysical parameters were estimated: leaf chlorophyll content (LCC) and leaf area index (LAI). Local evaluation of an extended training dataset with more variation over bare soil sites led to improved LCC and LAI mapping with reduced uncertainties. GPR reached the 10% precision required by end users, with for LCC a NRMSE of 3.5–9.2% (r^2: 0.95–0.99) and for LAI a NRMSE of 6.5–7.3% (r^2: 0.95–0.96). The developed GPR models were subsequently applied to simulated Sentinel images over various sites. The associated uncertainty maps proved to be a good indicator for evaluating the robustness of the retrieval performance. The generally low uncertainty intervals over vegetated surfaces suggest that the locally trained GPR models are portable to other sites and conditions.

1. Introduction

Biophysical parameter products such as leaf chlorophyll content (LCC) and leaf area index (LAI) have become standard products by space agencies and research institutions. At the same time, proper usage of such products requires that associated uncertainty information needs to be provided (Buermann et al., 2001; Morisette et al., 2006; Fang et al., 2012). Two categories of product uncertainty information are generally available in the literature, either physical or theoretical (Fang et al., 2012). Physical uncertainties indicate the departure of product values from hypothetical true values and are obtained through the collection of ground-based validation data. Theoretical uncertainties are caused by uncertainties in the input data and model imperfections and are usually estimated during the retrieval process (Baret et al., 2007; Knyazikhin et al., 1998; Pinty et al., 2011). Several operationally delivered products are nowadays accompanied with theoretical uncertainties in the form of a quantitative quality indicator (Knyazikhin et al., 1998; Pinty et al., 2011). An alternative method was recently proposed in (Fang et al., 2012) where authors apply an independent uncertainty calculation through a triple collocation method. As such, independently the uncertainties of MODIS, CYCLOPES, and GLOBCARBON LAI products were evaluated. From these products, it was concluded that particularly CYCLOPES generally meet the quality requirements (±0.5) proposed by the Global Climate Observing System (GCOS) (GCOS, 2011). The CYCLOPES algorithm is based on a neural network (NN) trained from the 1D SAIL radiative transfer model (RTM) (Baret et al., 2007).

While NNs have proven robust in various operational processing chains, the retrieval schemes are far from being perfect and various limitations have been identified. For instance, LAI estimates from CYCLOPES are less accurate at higher values due to the saturation effect in the radiative transfer simulation and the

* Corresponding author. Tel.: +34 963544067.
E-mail address: jochem.verrelst@uv.es (J. Verrelst).

0924-2716/$ - see front matter © 2013 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.isprsjprs.2013.09.012

NN inversion algorithm (Bacour et al., 2006; Weiss et al., 2007). Furthermore, NNs not only behave as a black box model, but they are also relatively unpredictable when used with input spectra that deviate (even slightly) from what has been presented during the training stage (Atzberger, 2004; Baret and Buis, 2008). It is therefore to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insights in the retrievals and evaluating retrieval performances. Besides, training NNs involve tuning several parameters that may greatly impact the final robustness of the model. For these reasons, alternatives that overcome these limitations are needed. In part, this is why in the recent years NNs are being replaced by other more advanced, simpler to train, machine learning regression algorithms (MLRAs). Actually, during the last two decades, the family of *kernel methods* (Camps-Valls and Bruzzone, 2009) has emerged as an alternative to NNs in many applications. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input–output nonlinear mappings. They are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. Intimately related to the field of kernel methods, here we find the new emerging field of Bayesian nonparametric modeling. The framework of Bayesian nonparametrics gives a Bayesian treatment of statistical inference. This field has given rise to particularly powerful methods, such as relevance vector machines and Gaussian processes. These methods are able to provide high accuracies and at the same time provide uncertainty intervals for the predictions (e.g. Camps-Valls et al., 2006; Verrelst et al., 2012b). They may therefore be more suitable candidates for operational applications, especially now that Earth observation is reaching a mature state.

In 2014, the European Space Agency's (ESA) forthcoming Sentinel-2 (S2) mission will start delivering high-resolution optical images on a global scale. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods. Recently, Verrelst et al. (2012a) have tested the capabilities of four state-of-the-art MLRAs given different Sentinel-2 and Sentinel-3 band settings. Selected MLRAs were NNs, support vector regression, kernel ridge regression, and Gaussian processes regression (GPR). The methods were compared in terms of accuracy, goodness of fit, robustness to low sample sized scenarios, and computational cost. Training and validation data came from the ESA-led field campaign SPARC, which took place on the agricultural test site Barrax, Spain. The main conclusion was that, in general, GPR outperformed the other regressors in terms of speed and computational costs. At the same time, GPR yielded superior accuracies for the majority of tested cases. Moreover, in contrary to NNs, GPR provide directly theoretical uncertainty estimates through Gaussian probability (cf. Section 3). These uncertainty estimates opened a new source of information. For instance, they make possible to assess the robustness of the retrievals at various spatial scales. In Verrelst et al. (2013), retrievals from hyperspectral airborne and spaceborne data over the Barrax area were compared. In this way, the uncertainty measure provided information about the upscaling quality, i.e., if the uncertainties are kept constant then the upscaling can be considered stable. Even though retrievals proved to be robust over vegetated areas, high uncertainties appeared over non-vegetated surfaces, which suggested that the training dataset was not representative enough for those land cover types. Furthermore, since statistical approaches are often criticized because of limited generalization and transferability (e.g. Colombo et al., 2003; Meroni et al., 2004), it remains to be questioned how robust the locally-trained GPR models function when applied to other sites and conditions. In this respect, the delivery of additional uncertainty estimates may enable to evaluate the portability of the regression model. Specifically, when uncertainty intervals as produced by a locally trained GPR model over an arbitrary site are on the same order as those produced over the successfully validated reference site, then it can be reasonably assumed that the parameter retrievals are also of the same quality as the retrievals of the reference site. Thus, when successfully validated over a reference imagery then the uncertainty estimates can work as a quality indicator. This concept has been evaluated here, and brings us to the following specific objectives of the present paper: (1) to evaluate the impact of experimental training data on the development of GPR models, particularly in view of improved retrievals over non-vegetated surfaces, and (2) to evaluate the portability of a locally trained GPR model to other sites and conditions by making use of its associated uncertainty intervals.

The remainder of the paper is organized as follows. Section 2 briefly describes the Sentinel-2 concept, while Section 3 revises the Bayesian nonparametric field in general and the Gaussian process regression algorithm in particular. In the methodology (Section 4) the used training dataset, experimental Sentinel-2 images, and experimental setup are described. Results focus first on a local evaluation of the GPR models (Section 5.2) and then moves to the evaluation of algorithm's performance on other sites (Section 5.3). Section 6 discusses main findings and Section 7 concludes the paper.

2. Sentinel-2

ESA's S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades (Drusch et al., 2012). S2 is a polar-orbiting, superspectral high-resolution imaging mission. The mission is envisaged to fly a pair of satellites with the first planned to launch in 2014. Each S2 satellite carries a Multi-Spectral Imager (MSI) with a swath of 290 km. It provides a versatile set of 13 spectral bands spanning from the visible and near infrared (VNIR) to the shortwave infrared (SWIR), featuring four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Table 1). S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. The pair of S2 satellites aims to deliver data taken over all land surfaces and coastal zones every five days under cloud-free conditions, and typically every 15–30 days considering the presence of clouds. To serve the objectives of Copernicus (The European Earth Observation Programme), S2 satellites will provide data for the generation of high-level operational products (level 2b/3) such as land-cover and land-change detection maps and geophysical variables such as LCC, LAI and leaf water content maps. To ensure that the final products can meet user requirements, the Copernicus user committee defined an accuracy goal of 10% (Drusch et al., 2012).

3. Bayesian nonparametrics and Gaussian processes

Finding a functional relation between input (e.g. reflectances) and output (e.g. physical parameter) variables is the main goal of statistical learning. The problem is complex and elusive because possibly an infinite number of functions can be found to fit the data. This problem is known as the excess of capacity of the class of functions implementing the model, and has been referred to the problem of overfitting. The key is to constrain model's capacity, in either one of the following two ways: imposing strong prior knowledge or via regularization schemes that promote simpler solutions. In the last decades, statistical learning (inference) has witnessed an overwhelming interest in kernel methods because they implement nonlinear models and still rely on linear algebra operations (Camps-Valls and Bruzzone, 2009). Kernel methods are very appealing for physical parameter retrieval, mainly because they deal efficiently with low-sized datasets of potentially high dimensionality, which are the situations we find in parameter

Table 1
Sentinel-2 MSI band settings.

Band #	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Band center (nm)	443	490	560	665	705	740	783	842	865	945	1375	1610	2190
Band width (nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

retrieval from multispectral or hyperspectral imagery using models trained by field campaign data. Also, since kernel methods do not assume an explicit prior data distribution but are inherently nonparametric models, they cope well with remote sensing data specificities and complexities (Camps-Valls and Bruzzone, 2009).

In the context of statistical inference and for remote sensing products, one is not solely interested in high accuracies of the algorithm but also in producing uncertainty intervals for the predictions. This calls for a Bayesian treatment of the inference problem. While recently kernel methods have advanced the field of remote sensing data analysis (Camps-Valls and Bruzzone, 2009), the problem of uncertainty estimation with nonparametric models has been elusive in most of the approaches. In this context, the emerging field of *nonparametric Bayesian modeling* constitutes a proper theoretical framework to tackle the problem of physical parameter retrieval (O'Hagan, 1994; Rasmussen and Williams, 2006; Orbanz and Teh, 2010).[1] Essentially, a nonparametric Bayesian model is a Bayesian model on an infinite-dimensional parameter space, which corresponds to the set of possible patterns, e.g. the class of smooth functions for regression (retrieval). The field has been very active in the last decade, and has delivered successful model instantiations. Some of them has been actually introduced in the field of remote sensing data analysis. For example, the relevance vector machine (RVM) introduced in Tipping (2001) is a nonparametric Bayesian model that assumes a Gaussian prior over the weights in order to enforce sparsity, and uses expectation-maximization to infer the parameters. In Camps-Valls et al. (2006) the RVM model was used for oceanic chlorophyll content estimation. The model, however, may incur in too sparse solutions that do not fit well in parts of the space not considered in the retained relevant vectors. Lately, Bayesian nonparametric modeling with Gaussian Processes (Rasmussen and Williams, 2006) have received much attention in the field of machine learning, and has been also introduced in the remote sensing application field (Verrelst et al., 2012a,b, 2013). This paper will focus on GPR.

GPR provides a probabilistic (Bayesian) approach for learning generic regression problems with kernels (Rasmussen and Williams, 2006). The GPR model establishes a relation between the input (B-bands spectra) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (canopy parameter) $y \in \mathbb{R}$ of the form:

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}), \tag{1}$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the spectra used in the training phase, $\alpha_i \in \mathbb{R}$ is the weight assigned to each one of them, and we intentionally dropped the bias term of the regression assuming centered observations, and K is a function evaluating the similarity between the test spectrum $\mathbf{x}$ and all N training spectra, $\mathbf{x}_i$, $i = 1, \ldots, N$. We used a scaled Gaussian kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = v \exp\left(-\sum_{b=1}^{B} \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2} \right) + \delta_{ij} \cdot \sigma_n^2, \tag{2}$$

where v is a scaling factor, B is the number of bands, σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b, σ_n is the noise standard deviation and δ_{ij} is the Kronecker's symbol. The kernel is thus parametrized by signal (v, σ_b) and noise (σ_n) hyperparameters, collectively denoted as $\boldsymbol{\theta} = \{v, \sigma_b, \sigma_n\}$.

For training purposes, we assume that the observed variable is formed by noisy observations of the true underlying function $y = f(\mathbf{x}) + \epsilon$. Moreover we assume the noise to be additive independently identically Gaussian distributed with zero mean and variance σ_n. Let us define the stacked output values $\mathbf{y} = (y_1, \ldots, y_n)^\top$, the covariance terms of the test point $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n))^\top$, and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$. From the previous model assumption, the output values are distributed according to:

$$\begin{pmatrix} \mathbf{y} \\ f(x_*) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} \end{pmatrix} \right) \tag{3}$$

For prediction purposes, the GPR is obtained by computing the posterior distribution over the unknown output $\mathbf{y}_*$, $p(\mathbf{y}_*|\mathbf{x}_*, \mathcal{D})$, where $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \ldots, N\}$ is the training dataset. Interestingly, this posterior can be shown to be a Gaussian distribution, $p(y_*|\mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*})$, for which one can estimate the *predictive mean* (point-wise predictions):

$$\mu_{\mathrm{GP}*} = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}, \tag{4}$$

and the *predictive variance* (confidence intervals):

$$\sigma^2_{\mathrm{GP}*} = k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*. \tag{5}$$

The corresponding hyperparameters $\boldsymbol{\theta}$ are typically selected by Type-II Maximum Likelihood, using the marginal likelihood (also called *evidence*) of the observations, which is also analytical. When the derivatives of the log-evidence are also analytical, which is often the case, conjugated gradient ascent is typically used for optimization (see (Rasmussen and Williams, 2006; Camps-Valls et al., 2009) for further details).

Three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization gives the relevance of each spectrum $\mathbf{x}_i$. The predictive mean is essentially a weighted average of the canopy parameter values associated to the training samples closest to the test sample. Second, the inverse of σ_b represents the relevance of band b. Intuitively, high values of σ_b mean that relations largely extend along that band hence suggesting a lower informative content. These features have been extensively studied in (Verrelst et al., 2012a,b) and proved to be valuable for gaining insight in relevant bands. Finally, a GPR model provides not only a pixelwise prediction for each spectrum but also an uncertainty (or confidence) level for the prediction. Hence in contrary to other approaches (e.g. NN) uncertainty intervals are directly delivered along with mean estimates. The interested reader is referred to the book by Rasmussen and Williams (2006) for more details on the theoretical aspects of GPR. A Matlab implementation of GPR is freely available at http://www.gaussianprocess.org/gpml/.

We illustrate the uncertainty intervals estimated by GPR and through standard boostrapping (Wu, 1986) of the nonlinear regression solution in Fig. 1. Note that GPR basically accounts for a reduction of uncertainty based on the relative local density of the input data points, not the outputs. This is obvious by looking at the

[1] Excellent online lectures on Bayesian nonparametrics are available at: http://videolectures.net/mlss09uk_teh_nbm/ and http://videolectures.net/mlss09uk_orbanz_fnbm/.

Fig. 1. Toy regression example in 1-D: we predict the *y*-values from the *x*-values in a synthetic signal generated by a combination of Gaussian bumps. We show the predictive variance or confidence interval estimates via bootstrap variance estimates (left), and Gaussian process regression (right).

predictive variance equation above. On the other hand, the variance of the bootstrap variance estimate quickly vanishes if one moves away from data points in the set. This behavior correctly reflects the fact that the predictions will be practically zero far away from points in the training sets. Gaussian processes indicate that the uncertainty is high because no data has been observed in that area.

4. Methodology

4.1. SPARC database

GPR has been trained with a local ground dataset coming from SPARC (SPectra bARrax Campaign). The SPARC-2003 and SPARC-2004 campaigns took place in the Barrax agricultural site in Central Spain (coordinates 39°N, −2°1′E, 700 m altitude). The test area has a rectangular form and an extent of 5 km × 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. The annual rainfall average is about 400 mm.

In the 2003 campaign (12–14 July) biophysical parameters were measured within a total of 113 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size of about 20^2 m. The same field data were collected in the 2004 campaign (15–16 July) within a total of 18 ESUs among different crops. For both years, within each ESU the averaged leaf LCC was derived by measuring about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned to a LAI value, which was obtained by the average of 24 measures (8 data readings × 3 replications). In total 9 crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat) were sampled, with field-measured values of LAI that vary between 0.4 and 6.3 and LCC between 2 and 55 μg/cm^2. Further details on the measurements can be found in the data acquisition report (Moreno and Participants of the SPARC campaigns, 2004). Additionally, 30 random bare soil spectra with a biophysical (LCC, LAI) value of zero were added to broaden the dataset to non-vegetated samples. This 'original' dataset used in Verrelst et al. (2012a,b, 2013) in training GPR models and hereafter refers to 'training original' or '*TrOr*'.

In the latter studies, the GPR uncertainty maps showed that mean estimates over vegetated areas were associated to low uncertainties. At the same time, considerably higher uncertainties were encountered over areas of fallow land and bare soils. These higher uncertainties can be attributed to the relatively poor contribution of non-vegetated land cover types in the training dataset. For this reason, we have extended the SPARC training dataset with 60 new spectra that cover all kinds of non-vegetated surfaces, i.e. spectra with an LCC and LAI of zero. Most of the spectra were taken over bare soil surfaces, but also man-made surfaces (e.g., build-up areas, roads) and water bodies have been included. This 'extended' training dataset hereafter refers to 'training extended' or '*TrEx*'.

4.2. Reference and other simulated Sentinel-2 images

Because actual S2 data is not available yet, we opted for simulating it on the basis of Compact High Resolution Imaging Spectrometry (CHRIS) data. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, site of the covered area and spatial resolution because of on-board memory storage reasons. The radiometric resolution of CHRIS is 12 bits, which is the same as S2's MSI. We made use of nominal nadir CHRIS observations in Mode 1 (62 bands, maximal spectral information) for the four SPARC campaign days, where field measurements of surface properties were measured in conjunction with satellite overpasses. CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. The images were corrected for atmospheric effects according to the method proposed in Guanter et al. (2005). Since most of ground truth data were collected during the 2003 campaign, the nadir image from 14 July 2003 was used as reference image for spectral and spatial resampling to the settings of S2. The image is shown in Fig. 2. The majority of ESUs are located on the circular green fields, while non-vegetated samples came from the yellowish-white[2] surfaces. Because configured with different pixel sizes (10, 20 and 60 m), it is of special interest to simulate S2 bands as a function of pixel's size. A nearest neighbor strategy was used for the spatial resampling and a Gaussian model with full-width-half-maximum spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, experimental data according to the following three Sentinel settings were generated, 'S2-10m': four bands at 10 m, 'S2-20m': eight bands at 20 m (4 bands at 20 m plus the S2-10m bands coarse-grained at 20 m), and 'S2-60m': ten bands at 60 m (2 bands at 60 m plus the earlier bands coarse-grained at 60 m).

Since the objective of this work was to evaluate the ability of transferring Barrax-trained GPR models for the various Sentinel settings to other images, multiple Mode 1 CHRIS images over terrestrial surfaces were arbitrarily collected. The only requirement was that they are predominantly cloud-free and acquired during spring or summertime. The CHRIS images include: multi-temporal images over Barrax (Spain); July 2004, and June 2009. Multitemporal images over Demmin (Germany), May and July 2006, and an image over Los Monegros (Spain), August 2006, Las Tablas (Spain), July 2006 and Sudbury (Canada), August 2007. The sites are

[2] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

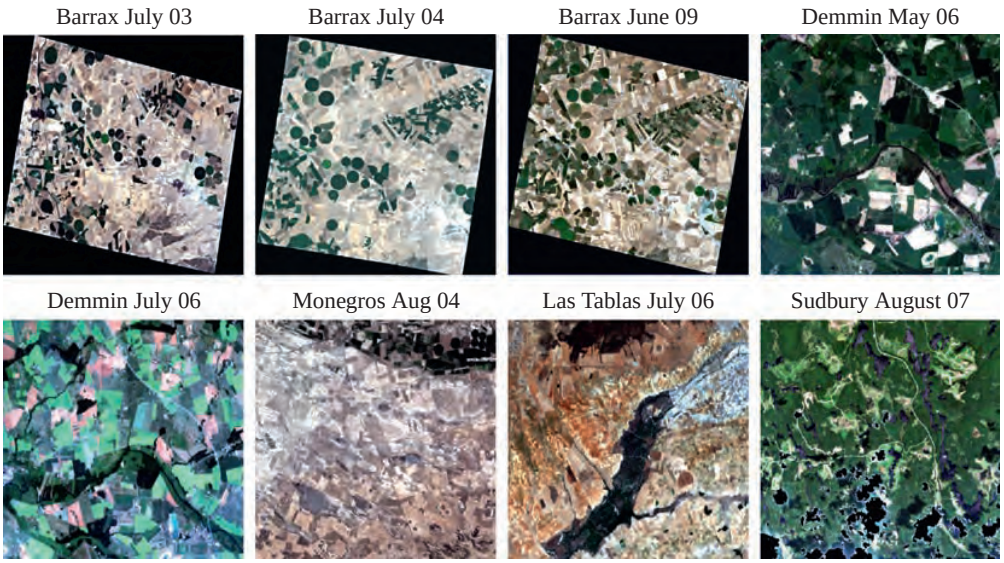


Fig. 2. RGB compositions of CHRIS images used for evaluating the performance of the locally-trained GPR models. Barrax July 03 is the reference image.

Table 2
Description of test sites.

Site	Acquisition time	Description
Barrax, Spain (39°N, −2°E)	14 July 03	The Barrax agricultural area has a rectangular form and an extent of 5 km × 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. The annual rainfall average is about 400 mm
Barrax, Spain (39°N, −2°E)	16 July 04	
Barrax, Spain (39°N, −2°E)	19 June 09	
Demmin, Germany (53.5°N, −13.1°E)	08 May 06	The Demmin agricultural area is located in Northeast Germany and is based on a group of farms covering app. 25,000 ha. The surface is flat at 50 m a.s.l. with some slopes along the Tollense River. The field sizes are large in this area for German standards, about 80–250 ha. The main crops grown are wheat, barley, rape, maize and sugar. The annual rainfall ranges from 500 to 650 nm
Demmin, Germany (53.5°N, −13.1°E)	07 July 06	
Monegros, Spain (41.2°N, −0.1°1′E)	20 August 05	Los Monegros area is a semi-arid region, sparse vegetation and shallow as well as poorly developed soils. Agricultural activities are poor and may trigger land degradation processes. The area is characterized by various small playa lakes. These lakes form in small karstic depression by the dissolution of evaporitic subsurface layers, mainly gypsum and limestone. The playa lakes are usually dry in summer. The annual rainfall average is low, about 350 mm
Las Tablas, Spain (39.1°N, −3.4°E)	03 July 06	Las Tablas de Daimiel is a Complex of shallow pools and associated marshland, which lies in the great plain of La Mancha. The surface is flat at 600–620 m a.s.l. The site receives floodwater from the permanent freshwater Río Guadiana and the seasonal brackish Río Gigüela, and groundwater from an underground water basin. The surrounding area is used for rain-fed agriculture, mainly olives and wine yards. The annual average rainfall is about 450 mm
Sudbury, Canada (47.1°N, −81.4°E)	10 August 07	Sudbury is a boreal forested region located in the northern part of Ontario. It is a flat area 350 m a.s.l. The area is predominantly covered by black spruce and aspen stands with an understory of shrubs and herbs. The annual rainfall average is about 820 mm

described in Table 2. All these CHRIS images were corrected for atmospheric effects according to Guanter et al. (2005). A cloud masking over the Sudbury image was applied to remove the pixels with cloud contamination. Although not really necessary for this exercise, the Barrax images were also geometrically corrected according to Alonso and Moreno (2005). The images were subsequently resampled according to above-described Sentinel settings (S2-10m, S2-20m and S2-60m). The images are displayed in Fig. 2.

4.3. Experimental setup

The local *TrOr* and *TrEx* experimental datasets were divided into two subsets: 80% for training and the remaining 20% for validation. The subsequent undertaken approach was straightforward; for each biophysical parameter (LCC, LAI) and each S2 configuration (S2-10m, S2-20m, S2-60m) a *TrOr* and *TrEx* model were trained. The predictive power of the developed models was subsequently validated against the 20% validation data by using the absolute root-mean-squared error (RMSE) and the normalized RMSE (NRMSE [%] = RMSE/range of the parameters as measured in the field ∗100) to assess accuracy, and the coefficient of determination (r^2) to account for the goodness-of-fit. The NRMSE was used to compare the performances across the different methods and parameters. Once successfully validated, the *TrOr*- and *TrEx*-developed models were applied to the other experimental S2 images and the mean estimates and associated uncertainties were compared. To preserve a physical meaning, negative LCC or LAI estimates were automatically converted to near-zero values (0.0001) during the retrieval process.

5. Results

5.1. GPR performance with original and extended training dataset

The evaluation of the *TrOr*- and *TrEx*-developed GPR models starts with inspection of the validation results (Table 3). Excellent

prediction accuracies were obtained for all scenarios with a r^2 between 0.92 and 0.99 and NRMSE between 3.5% and 10.5%. Noteworthy hereby is that excellent LAI accuracies were already achieved with only 4 bands (S2-10m). This is encouraging for LAI mapping at high spatial resolution. The inclusion of more bands only improved accuracies marginally. Conversely, LCC clearly gained from the inclusion of extra bands, with a S2-20m configuration (8 bands: B2-B8a) that managed to reach NRMSE down to 3.5% (r^2: 0.99). It is well known that the inclusion of red-edge bands (B5 and B6) can considerably improve biophysical parameter estimation (e.g. Verrelst et al., 2012a; Delegido et al., 2011, 2013). However, the addition of 2 more bands in the blue (B1) and NIR (B9) at a coarser resolution of 60 m (S2-60m) slightly degraded accuracies.

Of more relevance here is the comparison of validation results as achieved by using models developed using *TrOr* and *TrEx*. It can be noticed that *TrEx* yielded slightly improved results. This is especially the case for LAI, where for all S2 scenarios NRMSE dropped with about 2%. Hence, the GPR model is able to incorporate more samples without losing accuracy. In fact the contrary occurred; for all S2 configurations, the NRMSE kept below the threshold of 10%.

Even though successfully validated, it is well understood that a validation dataset is rather limited as compared to the total variability observed in satellite images. Therefore, it is expected that inspection of the mean estimate and associated uncertainty maps on complete scenes will allow us to better understand the performances of both GPR models.

5.2. Local mapping over Barrax, Spain

Although LCC and LAI maps were generated for all S2 settings from the 2003 Barrax reference site, for the sake of brevity we only display generated maps at a high spatial resolution of 10 m. In turn the maps will allows us to appreciate the strength of GPR. Biophysical parameter retrieval was thus achieved with 4 bands only: 490, 560, 665 and 842 nm. The retrieval process was completely automated and image-based; the generation of a map was completed almost instantaneously.

Fig. 3[top] provides mean estimate (μ) maps and associated uncertainties ($\pm\sigma$) over the Barrax test site as generated by the original training dataset (*TrOr*). The mean estimate maps are first briefly interpreted. Within-field variations are clearly detected in both maps. Particularly, the pronounced spatial variation of LCC marks prominently the irrigated circular fields with green biomass. These irrigated fields are characterized by an LCC above 40, and an LAI above 3. Areas with low LCC and LAI (the whitish parts) are mainly bare soils, fallow lands or rain-fed senescent or harvested cereal fields (wheat, barley). The same maps are provided below, but then generated by the extended training dataset (*TrEx*). When comparing *TrOr*- and *TrEx*-generated LCC maps it can be observed that they are very similar. That similarity is also reflected when correlating both maps in a scatter plot, as displayed below. The large majority of the pixels fall right on the 1:1 line. In turn, more differences between *TrOr*- and *TrEx*-generated LAI maps occurred over non-green vegetated areas (e.g. fallow land and bare soils). The addition of bare soil spectra in the training dataset led thus to more meaningful LAI retrievals for a considerably amount of pixels, as was also observed by the validation dataset in Table 3. The scatter plot shows that the *TrEx*-developed LAI model causes that a substantial part of pixels are interpreted towards lower LAI estimates. While this may imply an improved accuracy, the associated uncertainty intervals will manifest the quality of the estimates.

Within these uncertainty maps, areas with reliable retrievals can be clearly distinguished from areas with unreliable retrievals. These differences are more obviously observed in the *TrOr*-generated maps. Reliable retrievals (low $\pm\sigma$) were found on irrigated areas and harvested fields. This is not surprising since the majority of training and validation samples came from these fields. Unreliable retrievals (high $\pm\sigma$) were found on areas with remarkably different spectra, such as bright, whitish calcareous soils (center, right), or harvested rain-fed barley fields with remaining bright straws covering the surface (center). Hence, as earlier noticed in Verrelst et al. (2012a,b), a practical implication of the uncertainty maps is the detection of areas that may benefit from a denser ground truth sampling regime. That was also the rationale for the collection of an extended training dataset over non-vegetated targets (i.e. *TrEx*). For both biophysical parameters, *TrEx* resulted in reduced uncertainties across the whole map (see also statistics in Table 4), and regions with large uncertainties have been considerably reduced. That trend is again visualized by the scatter plots displayed down Fig. 3; the majority of pixels fall below the 1:1 line. Consequently, the extended training dataset demonstrated its superiority; it was validated with high accuracy, yielded realistic LCC and LAI maps, and these maps were delivered with lower uncertainties.

Table 3
Validation statistics (r^2, absolute RMSE and NRMSE (%)) for the Sentinel configurations and LCC and LAI as generated by the *TrOr*- and *TrEx*-developed GPR models.

Parameter	r^2		abs. RMSE		NRMSE (%)	
	TrOr	*TrEx*	*TrOr*	*TrEx*	*TrOr*	*TrEx*
S2-10m:						
LCC	0.931	0.949	5.36	4.70	10.50	9.21
LAI	0.910	0.948	0.51	0.39	9.37	7.28
S2-20m:						
LCC	0.991	0.993	1.92	1.77	3.76	3.48
LAI	0.916	0.958	0.49	0.35	9.09	6.51
S2-60m:						
LCC	0.977	0.993	2.37	1.81	6.49	3.55
LAI	0.934	0.959	0.29	0.36	8.42	6.63

5.3. Evaluating portability GPR models to other experimental S2 images by inspecting scatterplots

Even though promising results have been obtained over the reference Barrax site, a key requirement for operational applications is to ensure that GPR models are portable to other regions with similar accuracy. Henceforth, the *TrOr*- and *TrEx*-developed GPR models were applied to simulated S2 images over various sites in Spain, Germany and Canada. Mean estimate and uncertainty *TrOr* and *TrEx* maps were generated and again correlated. Scatter plots are displayed in Fig. 4. Some interesting observations can be made from these scatter plots. For all tested images, LCC mean estimates fell right on the 1:1 line. It confirms the earlier observed trend that the extended training dataset did not lead to erratic LCC predictions. But also the scatter plots of the uncertainty estimates confirms the earlier observed systematic trend; for all images the *TrEx*-developed models yielded a systematic decrease in uncertainty, e.g. LCC $\pm\sigma$ hardly reached above 10 $\mu g/cm^2$.

Also with respect to LAI estimates, we observed the same pattern encountered in the reference image, i.e. the majority of *TrEx*-processed pixels largely follows the 1:1 line. However, a considerable amount of pixels tended to deviate towards lower estimates. While this trend appeared only slightly over Barrax and Demmin, the down-estimating occurred strongly over the scarcely vegetated areas of Los Monegros and Las Tablas. Also over Sudbury (Canada), LAI values were underestimated despite being dominated by vegetated surfaces. But most important is that, for all images, a significant and systematic decrease in uncertainties took

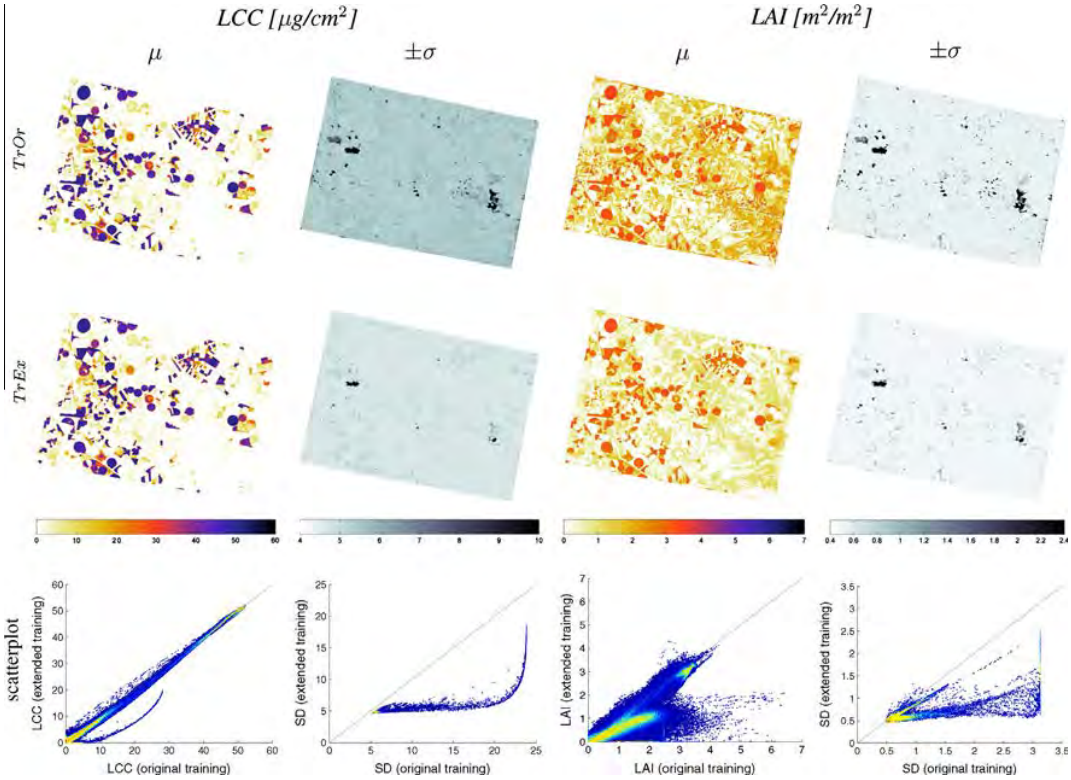


Fig. 3. Mean estimate (μ) and associated uncertainty ($\pm\sigma$) maps for LCC and LAI as generated by the S2-10m original training dataset (*TrOr*) [top], and extended training dataset (*TrEx*) [middle]. Scatterplots of *TrOr* vs. *TrEx* are shown below.

Table 4
Overview statistics (mean ($\bar{x}$) and standard deviation (SD)) for LCC and LAI μ (mean prediction) and $\pm\sigma$ (associated uncertainty) S2-10m maps as retrieved by *TrOr*- and *TrEx*-developed GPR models. 'Barrax July 03' is the reference image.

Image	μ LCC		$\pm\sigma$ LCC		μ LAI		$\pm\sigma$ LAI	
	TrOr $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)
Barrax July 03	7.71 (14.89)	7.54 (14.76)	5.53 (1.04)	4.75 (0.41)	1.21 (1.00)	0.82 (0.89)	0.58 (0.26)	0.51 (0.13)
Barrax July 04	9.48 (14.75)	9.22 (14.54)	5.50 (1.03)	4.75 (0.39)	1.37 (1.02)	1.13 (0.91)	0.60 (0.32)	0.51 (0.14)
Barrax June 09	12.44 (11.34)	10.66 (11.49)	6.10 (1.39)	4.98 (0.51)	0.91 (0.99)	0.73 (1.07)	0.73 (0.32)	0.73 (0.32)
Demmin May 06	25.38 (18.35)	24.31 (18.61)	5.75 (0.50)	5.00 (0.25)	2.37 (0.96)	1.70 (1.02)	0.66 (0.19)	0.59 (0.09)
Demmin July 06	26.60 (11.08)	25.69 (11.78)	5.53 (0.23)	4.93 (0.14)	1.85 (0.65)	1.37 (0.58)	0.55 (0.08)	0.55 (0.06)
Monegros Aug 05	2.16 (7.24)	2.42 (6.95)	7.22 (2.05)	4.96 (0.27)	1.81 (1.24)	0.24 (0.46)	1.44 (0.88)	0.67 (0.21)
Las Tablas July 06	7.99 (7.61)	7.42 (7.12)	5.67 (0.88)	4.78 (0.19	0.66 (0.83)	0.24 (0.38)	0.65 (0.35)	0.55 (0.11))
Sudbury Aug 07	6.92 (4.67)	3.65 (4.71)	8.30 (1.40)	5.97 (0.47)	2.32 (0.27)	0.63 (0.44)	1.80 (0.52)	0.96 (0.23)

place. Particularly those areas that earlier suffered from large uncertainties were now predicted with more certainty.

Overview statistics (mean and standard deviation) for prediction and uncertainty maps are provided in Table 4. With regard to portability evaluation, special attention goes to the mean uncertainty statistics along the different images. For both GPR models and LCC and LAI maps uncertainty statistics turned to be almost as good as the reference image. It can thus be concluded that the locally trained GPR model is generally applicable to other sites, thereby reaching uncertainties on the same order of the Barrax 2003 reference site. Only the Sudbury data was processed with considerably less certainty. Moreover, for all images, mean uncertainties have been considerably reduced when comparing *TrEx* with *TrOr*. Summarizing, the portability of the extended training dataset (*TrEx*) model has been successfully evaluated as mean uncertainties stabilized to about the same level as the reference Barrax image.

5.4. S2-10m biophysical parameter mapping

In this section we pay attention to the *TrEx*-generated maps, see Fig. 5. A first observation across the different sites is that LCC and LAI estimations fall within expected ranges. Variations in land cover are clearly observable and non-vegetated surfaces can be easily distinguished from vegetated surfaces. A second observation is that for most of the sites uncertainty maps show rather low values in general. Nevertheless, noticeable within-image variations are still occurring, particularly for LAI over the Barrax 2009 image. Only for Sudbury (Canada) systematically greater uncertainties appeared. Here a suboptimal atmospheric correction may explain the poorer performance.

While the uncertainty maps provide some information about the robustness of the retrievals, one has nevertheless to be careful with its interpretation. Note that $\pm\sigma$ represents the uncertainty interval around the mean predictions, meaning that they need to

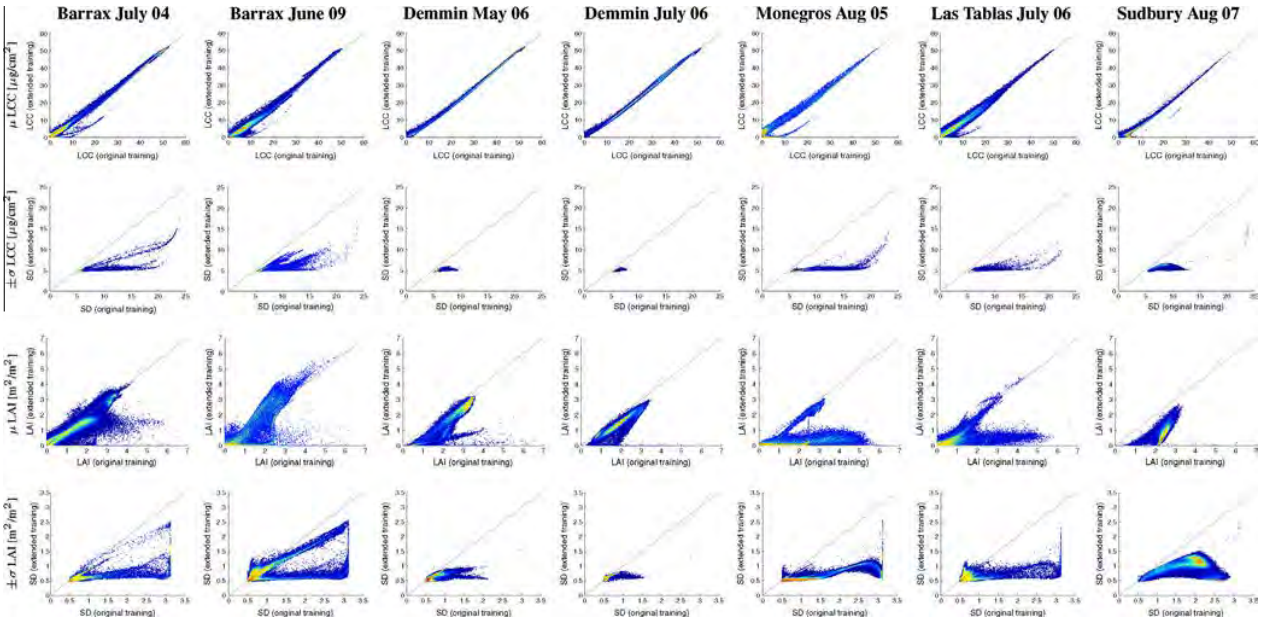


Fig. 4. Scatterplots of maps generated by the S2-10m original training datasets (*TrOr*) vs. extended training dataset (*TrEx*) for μ LCC [top], $\pm\sigma$ LCC [below], μ LCC [below] and $\pm\sigma$ LAI [bottom].

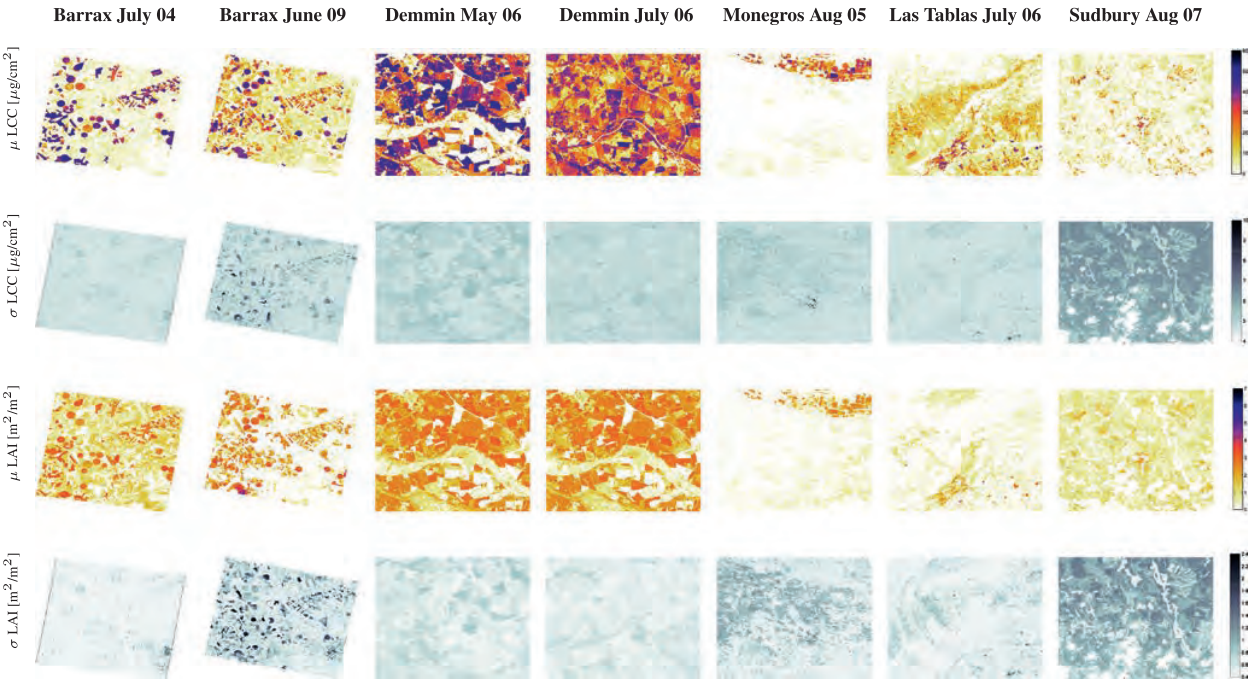


Fig. 5. Mean estimation (μ) and uncertainty ($\pm\sigma$) S2-10m maps for LCC and LAI as generated by the *TrEx*-developed GPR model.

be interpreted in relation to the estimates. For instance, an LCC uncertainty interval of about 5 would be more problematic for a mean estimate of 5 μg/cm^2 than of 50 μg/cm^2. Therefore, to evaluate the robustness of the GPR models it requires calculating the coefficient of variation:

$$CV = \frac{\sigma}{\mu} * 100. \quad (6)$$

CV maps provide relative uncertainties and are displayed in Fig. 6. These maps can then be evaluated against an uncertainty threshold, e.g. as proposed by GCOS, i.e. 20% (GCOS, 2011). Considering the reference Barrax image, dark areas represent retrievals with high uncertainties. These are typically bare soil areas which are characterized by low estimates (close to zero) and a relative high $\pm\sigma$. Conversely, it can be observed that most reliable retrievals occurred on the irrigated agricultural sites. However, only 9.0% of the pixels fell below the GCOS's 20% threshold. This low number seems surprising because the map was earlier excellently validated. It underlines the limited meaning of a (sparse) validation dataset when inspecting a heterogeneous map as a whole.

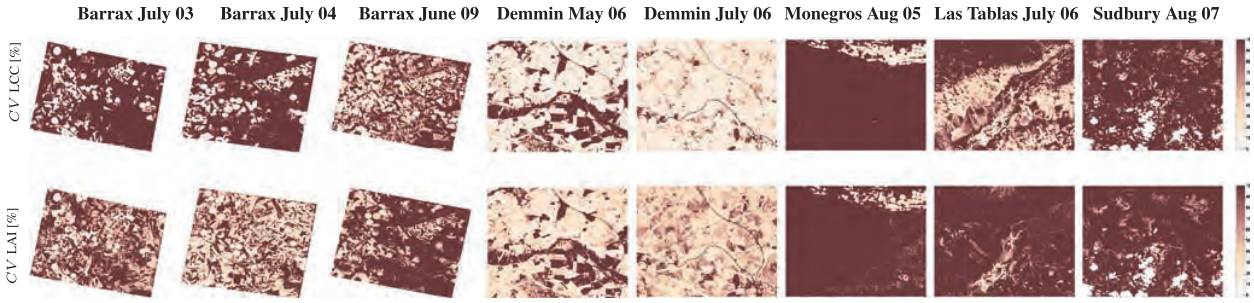


Fig. 6. Coefficient of variation (*CV*) S2-10m maps for LCC [top] and LAI [bottom] as generated by the *TrEx*-developed GPR model.

When comparing the *CV* maps across the different sites it can be observed that particularly over the agricultural areas meaningful estimates were obtained. For instance, the July 2006 Demmin LCC map is processed with low uncertainty over the whole image; 53.5% of pixels fell below the 20% threshold, and 75.6% below the 30% threshold. Also here agricultural parcels were processed with low uncertainty. Only non-vegetated surfaces such as water bodies are flagged with a high relative uncertainty. Spurious relative uncertainties are also observed across the Los Monegros and Las Tablas sites where bare soil dominate. With *TrEx* it was attempted to account for these bare soils, but due to near-zero mean estimates (0.0001) and, although $\pm\sigma$ is reduced, they still fall above near-zero threshold. *CV* turned therefore above 100%. Note that these relative uncertainty maps suggest that, at 10 m resolution, there is a greater problem of portability within an image than to other images. Solely the Sudbury maps show systematically higher relative uncertainties. This is probably due to the more difficult atmospheric correction as a consequence of cloud cover and lower sunlight intensity. The Sudbury case underlines the importance of an accurate and consistent atmospheric correction processing chain, which is expected to be provided by ESA (Level 2A Product). From a practical perspective, the *CV* map serves as a useful quality layer that allows masking out biophysical parameters estimates within an acceptable uncertainty (e.g. $CV < 20\%$), and at the same time can mask out non-vegetated surfaces (e.g. $CV > 100\%$).

6. Discussion

The forthcoming S2 mission opens opportunities to implement novel retrieval algorithms in operational processing chains. The interest is put on retrieval algorithms that are accurate, fast, robust, and sufficiently flexible to make fully use of the new S2 MSI bands. Machine learning regression algorithms (MLRAs) are able to cope with most of these objectives. In an earlier work, GPR was evaluated as a very promising regressor in terms of processing speed and accuracy when using a local training dataset (Verrelst et al., 2012a). At the same time, GPR is transparent in terms of model development (Rasmussen and Williams, 2006); it may provide a ranking of features (bands) and samples (spectra), thus alleviating the black-box problem. A discussion on its performance in comparison to alternative state-of-the-art retrieval approaches presented literature was provided in Verrelst et al. (2012a,b). In short, in a local setting GPR reached accuracies with S2 MSI band settings comparable (LAI) or superior (LCC) to competitive approaches. Nevertheless, the portability of this statistical approach to other sites remained questionable.

The here presented extended experimental training dataset (*TrEx*) not only further improved performances but also allowed a decrease in theoretical uncertainties. This proof of concept underlines the importance of a broad and diverse training dataset. More importantly, the GPR models were successfully applied to simulated S2 images covering various sites; associated uncertainties were on the same order as those generated by the reference image. The S2-10m examples demonstrated that excellent retrievals can be achieved already with 4 bands at a high spatial resolution of 10 m. Specifically, over fully vegetated surfaces relative uncertainties fell below the 20% requirements proposed by GCOS. This is encouraging for processing data from broadband sensors with a limited number of bands such as SPOT and Landsat or high resolution image such as Ikonos and Quickbird. On the other hand, S2 MSI encompasses additional bands at a coarser spatial resolution of 20 and 60 m. Note hereby that MSI's SWIR bands B11 and B12 have not been considered in this study because of falling outside the CHRIS' spectral range. These SWIR bands are known to be sensitive to vegetation structure (Brown et al., 2000) and can better distinguish between dried-out fallow and non-vegetated lands. It is expected that inclusion of the SWIR bands will further improve the retrieval quality and reduce uncertainties. Moreover, further improvements can be achieved, which may lead to further reduction of uncertainties over non-vegetated surfaces. For instance, one could develop a set of GPRs each working with different portions of the data or features. The uncertainty intervals could be optimally combined to improve the accuracy and reduce prediction bias. Alternatively, more sophisticated kernel functions that exploit signal-to-noise relations could be eventually considered (Gómez-Chova and Camps-Valls, 2012; Lazaro-Gredilla et al., 2013).

At the same time, it should be noted that in operational biophysical parameter retrievals, pixels over non-vegetated targets are flagged as non-valid and are not considered in uncertainty calculation (see Fang et al., 2012). Thereby, at coarser pixel's size of MODIS and SPOT/VEGETATION, non-vegetated areas are well-defined by land cover maps, such as desert, inland water body, urban surfaces. In contrast, at 10 m resolution non-vegetated pixels can virtually appear anywhere. Since those kinds of pixels can be easily identified (e.g., mean estimates of ≈ 0, or $CV > 100\%$), they could actually just be discarded when interpreting uncertainty estimates. Another remark to bear in mind is that the derived theoretical uncertainties are directly related to what has been presented during the training phase. Theoretical uncertainties are thus not intended to replace the true physical uncertainties of the biophysical parameter products but instead to provide additional complementary information. Physical uncertainties are mandatory to be provided and should be obtained through comprehensive validation datasets collected on various sites, such as that coordinated by the Committee on Earth Observation Satellites (CEOS) Land Product Validation (LPV) community (Morisette et al., 2006).

Finally, it does not escape our attention that only a limited set of images acquired during May-August have been evaluated. In an attempt to make the training data more representative at temporal and global scales, it should be able to cope with the majority of global land cover types along latitudinal gradients and over time.

Ideally, ground truth data (biophysical parameters plus associated radiometric data) should be collected over a broad variety of terrestrial surfaces and vegetation types at multiple phenological stages. This, however, is a tedious and expensive job. Alternatively, an urging open question in this respect is to evaluate how robust GPR performs when trained by artificial spectra, e.g. as generated by a radiative transfer model (RTM). The advantage of RTMs is that a broad range of land cover situations can be simulated (e.g. up to hundred thousands), leading to a dataset several times bigger than what can be collected during a field campaign. Operational processing chains typically rely on such a hybrid approach (e.g. Baret et al., 2007; Bacour et al., 2006). Initial efforts in this direction have been undertaken by implementing a MRLA module into the ARTMO (Automated Radiative Transfer Models Operator) toolbox, which is a suite of RTMs into one GUI toolbox (Verrelst et al., 2012c). As such, GPR and other MRLA models can be automatically developed through lookup tables of simulated spectra and corresponding input parameters (e.g. LCC, LAI). Nevertheless, it should hereby be noted that, while GPR works successful for small training datasets (e.g. <2000 samples), its heavy computational load impedes the use of very large datasets. This limitation has to be resolved when aiming to develop generic models for global applications. For instance, alternative (sparse) versions of GPR have been proposed that can handle large scale datasets, see e.g. http://www.gaussianprocess.org/. Greedy algorithms in active learning settings can be also an alternative: GPRs trained on different data portions can provide a ranked list of the most informative spectra which could be used then to generate a final model. Forthcoming research will move in these directions, ultimately to reach a robust and generic retrieval processing scheme.

7. Conclusions

The delivery of uncertainty information is a prerequisite for the operational use of remote sensing products. Currently only few retrieval approaches provide such estimates. Gaussian Processes Regression (GPR), a machine learning regression algorithm (MLRA) based on Bayesian nonparametric modeling, is one of them. Such retrieval algorithm may be of interest in view of ESA's forthcoming Sentinel-2 (S2) mission. In this work these uncertainty estimates were used to evaluate the robustness and portability of locally-trained GPR models. Hyperspectral CHRIS data was used for the simulation of experimental S2 images, being: S2-10m (4 bands), S2-20m (8 bands) and S2-60m (10 bands). While providing accurate estimates when trained with a local dataset called SPARC (*TrOr*; Barrax, Spain), it was also observed that non-vegetated pixels were processed with great uncertainty. In an attempt to make the LCC and LAI GPR models more robust, an extended training dataset was introduced (*TrEx*), i.e. the original SPARC dataset plus 60 spectra over all kinds of non-vegetated surfaces (e.g., soils, man-made surfaces, water bodies). For the Barrax 2003 reference image, *TrEx*-developed GPR model delivered validation errors below 10% (NRMSE: 3.5–9.2%; r^2: 0.95–0.99), robust LCC and improved LAI estimates, and above all reduced associated uncertainty estimates. These locally-trained GPR models were subsequently applied to other experimental S2 images over various sites across Spain, Germany, Canada. The uncertainty estimates provided insight in the success of the models' performance. Overall, GPR models are portable to other images and uncertainty estimates can thereby function as quality layer to filter out unreliable retrievals. Uncertainty intervals were on the same order as the Barrax 2003 reference image and relative uncertainties over vegetated surfaces were below the 20% requirements proposed by GCOS. However, typically large uncertainty variation within an image was observed due to surface heterogeneity. GPR is concluded as a powerful regressor for remote sensing applications; not only it delivers accurate predictions, it is currently the only MLRA that provides associated uncertainty intervals.

Acknowledgements

This paper has been partially supported by the Spanish Ministry for Science and Innovation under Projects AYA2010-21432-C02-01 and TIN2012-38102-C03-01. Dr. Luis Guanter at the Free University of Berlin is thanked for providing the atmospherically-corrected CHRIS images. Two anonymous reviewers are thanked for providing comments that helped to improve the quality of the original manuscript.

References

Alonso, L., Moreno, J., 2005. Advances and limitations in a parametric geometric correction of CHRIS/PROBA data. In: Proceedings of the 3rd CHRIS/Proba Workshop, ESA/ESRIN, Frascati, Italy. <http://earth.esa.int/workshops/chris_proba_05/papers/06_alonso.pdf>.

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

Bacour, C., Baret, F., Béal, D., Weiss, M., Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: principles and validation. Remote Sensing of Environment 105 (4), 313–325.

Baret, F., Buis, S., 2008. Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. In: Liang, S. (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application. Springer, pp. 171–200.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: principles of the algorithm. Remote Sensing of Environment 110 (3), 275–286.

Brown, L., Chen, J., Leblanc, S., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. Remote Sensing of Environment 71 (1), 16–25.

Buermann, W., Dong, J., Zeng, X., Myneni, R.B., Dickinson, R.E., 2001. Evaluation of the utility of satellite-based vegetation leaf area index data for climate simulations. Journal of Climate 14 (17), 3536–3550.

Camps-Valls, G., Bruzzone, L. (Eds.), 2009. Kernel Methods for Remote Sensing Data Analysis. Wiley & Sons, UK.

Camps-Valls, G., Gómez-Chova, L., Vila-Francés, J., Amorós-López, J., Muñoz-Marí, J., Calpe-Maravilla, J., 2006. Retrieval of oceanic chlorophyll concentration with relevance vector machines. Remote Sensing of Environment 105 (1), 23–33.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Calpe-Maravilla, J., 2009. Biophysical parameter estimation with adaptive Gaussian Processes. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2009. Capetown, South Africa.

Colombo, R., Bellingeri, D., Fasolini, D., Marino, C., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sensing of Environment 86 (1), 120–131.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11 (7), 7063–7081.

Delegido, J., Verrelst, J., Meza, C., Rivera, J., Alonso, L., Moreno, J., 2013. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. European Journal of Agronomy 46, 42–52.

Drusch, M., Del Bello, U., 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., Wei, S., Jiang, C., Scipal, K., 2012. Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. Remote Sensing of Environment 124, 610–621.

GCOS, 2011. 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), pp. 138. <http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf>.

Gómez-Chova, L., Camps-Valls, G., September 2012. Learning with the kernel signal-to-noise ratio. In: IEEE Workshop on Machine Learning for Signal Processing, MLSP'12. Santander, Spain.

Guanter, L., Alonso, L., Moreno, J., 2005. A method for the surface reflectance retrieval from PROBA/CHRIS data over land: application to ESA SPARC campaigns. IEEE Transactions on Geoscience and Remote Sensing 43 (12), 2908–2917.

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 D: Atmospheres 103 (D24), 32257–32275.

Lazaro-Gredilla, M., Titsias, M., Verrelst, J., Camps-Valls, G., 2013. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. Geoscience and Remote Sensing Letters, IEEE PP (99), 1–5.

Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. Remote Sensing of Environment 92 (2), 195–206.

Moreno, J., Participants of the SPARC campaigns, 2004. SPARC Data Acquisition Report. Contract No: 18307/04/NL/FF. University Valencia. <https://earth.esa.int/c/document_library/get_file?folderId=21020&name=DLFEs-386.pdf>.

Morisette, J., Baret, F., Privette, J., Myneni, R., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sánchez-Azofeifa, G., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iiames, 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 Transactions on Geoscience and Remote Sensing 44 (7), 1804–1814.

O'Hagan, A., 1994. Bayesian Inference. Kendall's Advanced Theory of Statistics, vol. 2B. Arnold, London, UK.

Orbanz, P., Teh, Y.-W., 2010. Bayesian Nonparametric Models. Encyclopedia of Machine Learning. Springer.

Pinty, B., Andredakis, I., Clerici, M., Kaminski, T., Taberner, M., Verstraete, M., Gobron, N., Plummer, S., Widlowski, J.-L., 2011. Exploiting the MODIS albedos with the two-stream inversion package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties. Journal of Geophysical Research D: Atmospheres 116 (9).

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press, New York, US.

Tipping, M.E., 2001. The relevance vector machine. Journal of Machine Learning Research 1, 211–244.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J., Camps-Valls, G., Moreno, J., 2012a. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sensing of Environment 118, 127–139.

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., 2012b. Retrieval of vegetation biophysical parameters using Gaussian process techniques. IEEE Transactions on Geoscience and Remote Sensing 50 (5 PART 2), 1832–1843.

Verrelst, J., Romijn, E., Kooistra, L., 2012c. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. Remote Sensing 4 (9), 2866–2889.

Verrelst, J., Alonso, L., Rivera Caicedo, J., Moreno, J., Camps-Valls, G., 2013. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (submitted for publication).

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote Sensing of Environment 110 (3), 317–331.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14 (4), 1261–1295.