Suivi de l'indice foliaire (LAI) à l'échelle globale : amélioration de la définition, de la continuité et de la cohérence des estimations de LAI à partir d'observations satellitaires kilometriques

Sivasathivel Kandasamy

THÈSE

Présentée par

Sivasathivel KANDASAMY

Pour obtenir le titre de

DOCTEUR DE L'UNIVERSITE D'AVIGNON ET DES PAYS DE VAUCLUSE

Spécialité : Télédétection de la Végétation

Leaf Area Index (LAI) monitoring at global scale: improved definition, continuity and consistency of LAI estimates from kilometric satellite observations

---

Date de soutenance :

Composition du jury :

Laurence HUBERT-MOY	Université de Rennes 2	Rapporteur
Richard FERNANDES	CCRS, Ottawa, Canada	Rapporteur
Roselyne LACAZE	HYGEOS, Toulouse	Examinateur
Raul LOPEZ-LOZANO	JRC Ispra, Italie	Examinateur
Philippe NEVEUX	Université d'Avignon et des Pays des Vaucluse	Examinateur
Aleixandre VERGER	INRA-EMMAH, Avignon	Co-directeur de thèse
Frédéric BARET	INRA-EMMAH, Avignon	Directeur de thèse

Thèse préparée au sein du Laboratoire de Environnement Méditerranéen et Modélisation des Agro-Hydro systèmes (EMMAH), INRA-PACA, Avignon, France

Acknowledgements

The experience of working on Ph.D. Thesis has been a wonderful and, in more times than one, an overwhelming experience. These three-and-half years of my thesis provided me with new learning opportunities, and experiences in conducting experiments and writing articles that are sure to shape my future career. The thesis has also been a motivating experience for further research in this area of remote sensing. The successful completion of thesis involves the efforts for many more and I would like to take this opportunity to thank them for their contributions.

First, I would like to thank my supervisor Dr. Frédéric Baret, for providing me an opportunity to work in this project. You have been a constant source of guidance and inspiration in all these years, even amidst your busy schedule. I thank for all fruitful the discussions we have had during these years and for your patience even at stressful times. This thesis would not have come to fruition but for support and guidance.

I would like to thank my co-supervisor Dr. Aleixandre Verger who has been instrumental in the successful completion of the project. I thank you for the numerous weekends and evenings that you had spent correcting my papers. I thank you for the excellent discussions we have had and for constantly inspiring me to expand my horizons of knowledge. I thank you for your friendship, patience and support.

I thank Dr. Samuel Buis, without whose help many of the thesis objectives would not have been achieved. You have helped us implement our programs in the different computer platforms and have supported us throughout the thesis. I thank for your friendship and support.

I thank Dr. Phillippe Neveux, for his contributions in helping us understand a new time-series processing method and in adapting it for our project. Without your help, we would have never succeeded in understanding and adapting the Singular Spectrum Analysis method for processing remotely senses vegetation data.

I also thank Dr. Maries Weiss and Dr. Eric Vermote for providing us the data required for the thesis and Dr. Peter North for his comments and suggestions for using his FLIGHT radiative transfer model in our study. I also thank Dr. Phillipe Clastre and Mr. Régis Marcel-Auda for helping us with technical issues during this period.

I also thank my teachers I have had over the years. I would like to start with Dr. Audrey Minghelli-Roman who was my Masters (VIBOT) thesis supervisor and my first contact in the subject of remote sensing. I thank you for introducing me to Dr. Baret and for being a

constant source of support and encouragement over all these years. I also would like to thank the VIBOT coordinators for selecting me for their Masters program, which was a great experience that prepared me for this thesis. I also would like to thank Dr. Nakkeeran Rangaswamy and Dr. Srinivasan Emperumal, for their encouragement and guidance throughout my engineering degree and after, and for the many fruitful and motivating discussions over the years. I also thank Rev.Fr. J. Paul and Mr. Balasubramanium for their encouragement to pursue a research career, during my school years.

I thank my friends Dr. Lucia Andreini, Dr. Claudia Dussaubat, Dr. Anita Simic, Dr. Tonya Lander and Dr. Alfredo Alessandrini encouraging and helping me in all these years and for the fun times. I also thank all the members of the unit (INRA-EMMAH) for providing me the most conducive and friendly environment during my thesis. I also thank my family for their patience and support.

Abstract

Monitoring biophysical variables at a global scale over long time periods is vital to address the climate change and food security challenges. Leaf Area Index (LAI) is a structure variable giving a measure of the canopy surface for radiation interception and canopy-atmosphere interactions. LAI is an important variable in many ecosystem models and it has been recognized as an Essential Climate Variable. This thesis aims to provide global and continuous estimates of LAI from satellite observations in near-real time according to user requirements to be used for diagnostic and prognostic evaluations of vegetation state and functioning. There are already some available LAI products which show however some important discrepancies in terms of magnitude and some limitations in terms of continuity and consistency. This thesis addresses these important issues. First, the nature of the LAI estimated from these satellite observations was investigated to address the existing differences in the definition of products. Then, different temporal smoothing and gap filling methods were analyzed to reduce noise and discontinuities in the time series mainly due to cloud cover. Finally, different methods for near real time estimation of LAI were evaluated. Such comparison assessment as a function of the level of noise and gaps were lacking for LAI.

Results achieved within the first part of the thesis show that the effective LAI is more accurately retrieved from satellite data than the actual LAI due to leaf clumping in the canopies. Further, the study has demonstrated that multi-view observations provide only marginal improvements on LAI retrieval. The study also found that for optimal retrievals the size of the uncertainty envelope over a set of possible solutions to be approximately equal to that in the reflectance measurements. The results achieved in the second part of the thesis found the method with locally adaptive temporal window, depending on amount of available observations and Climatology as background estimation to be more robust to noise and missing data for smoothing, gap-filling and near real time estimations with satellite time series.

Keywords: Leaf Area Index, radiative transfer model inversion, satellite, reflectance, time series processing, near real-time estimation

Résumé

Le suivi des variables biophysiques à l'échelle globale sur de longues périodes de temps est essentielle pour répondre aux nouveaux enjeux que constituent le changement climatique et la sécurité alimentaire. L'indice foliaire (LAI) est une variable de structure définissant la surface d'interception du rayonnement incident et et d'échanges gazeux avec l'atmosphère. Le LAI est donc une variable importante des modèles d'écosystèmes et a d'ailleurs été reconnue comme variable climatique essentielle (ECV). Cette thèse a pour objectif de fournir des estimations globales et continues de LAI à partir d'observations satellitaires en temps quasi-réel en réponse aux besoins des utilisateurs pour fournir des diagnostiques et pronostiques de l'état et du

fonctionnement de la végétation. Quelques produits LAI sont déjà disponibles mais montrent des désaccords et des limitations en termes de cohérence et de continuité. Cette thèse a pour objectif de lever ces limitations. Dans un premier temps, on essaiera de mieux définir la nature des estimations de LAI à partir d'observations satellitaires. Puis, différentes méthodes de lissage te bouchage des séries temporelles ont été analysées pour réduire le bruit et les discontinuités principalement liées à la couverture nuageuse. Finalement quelques méthodes d'estimation temps quasi réel ont été évaluées en considérant le niveau de bruit et les données manquantes.

Les résultats obtenus dans la première partie de cette thèse montrent que la LAI effectif et bien mieux estimé que la valeur réelle de LAI du fait de l'agrégation des feuilles observée au niveau du couvert. L'utilisation d'observations multidirectionnelles n'améliore que marginalement les performances d'estimation. L'étude montre également que les performances d'estimation optimales sont obtenues quand les solutions sont recherchées à l'intérieur d'une enveloppe définie par l'incertitude associée aux mesures radiométriques. Dans la deuxième partie consacrée à l'amélioration de la continuité et la cohérence des séries temporelles, les méthodes basées sur un fenêtre temporelle locale mais de largeur dépendant du nombre d'observations présentes, et utilisant la climatologie comme information a priori s'avèrent les plus intéressantes autorisant également l'estimation en temps quasi réel.

Mots-clés: Indice foliaire, inversion, modèle de transfert radiatif, satellite, séries temporelle, réflectance, estimation en temps quasi-réel.

Table of Contents

1 Introduction

In this introductory chapter the objectives of the thesis are put in perspective. The need and current initiatives for global monitoring of vegetation are first discussed. The significance and challenges associated with the estimation of LAI, specifically from remote sensing observations for global monitoring are briefly dealt. Methods for the estimation of LAI as well as temporal filtering methods for smoothing and gap filling satellite estimates are introduced. Finally, the specific objectives of the thesis are outlined along with the structure of the document.

1.1 Global monitoring of surface vegetation

Global monitoring of vegetation variables is of importance for the description and understanding earth dynamic processes, particularly in the context of the on-going global climate change to design policies for ensuring food security, conserving biodiversity and mitigate climate change effects through efficient management of resources. Remote Sensing observations from satellites, particularly from medium resolution satellite sensors, provide global observations over long-time periods. These observations would provide the policy makers of the different countries with the information that would help them to make their decisions. Consequently, vegetation monitoring at global and regional scales plays an important role in the economic development and well-being of the nations. Many initiatives have been taken over the years to enable continuous and consistent observations to the users. Some of the notable initiatives are:

1. Global Climate Observing System, *GCOS* (GCOS, 2012), established in 1992, is a joint initiative by the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Program (UNEP) and the International Council for Science (ISCU).GCOS aims to provide comprehensive information pertaining to earth climate processes for long-time periods. Information on vegetation and other terrestrial process are provided by the Global Terrestrial Observing System (GTOS, 2007).
2. Group on Earth observations, GEO (GEO, 2012) is a voluntary partnership of governments and international organizations to provide a global framework for projects and efforts aimed at providing Earth observations. GEO is coordinating the international effort to build Global Earth Observation System of Systems (GEOSS), a global public infrastructure for Earth observations consisting of flexible and distributed network of content providers. This framework was established in 2005, in response to the calls for sustainable development of G8

countries. As of January 2008, 72 Governments and the European Commission have been participating in this initiative. In addition, 52 intergovernmental, international and regional organizations with a mandate in Earth observation have also been recognized as Participating Organizations. GEO also coordinates the "GEO-GLAM" initiative, for agricultural monitoring at the continental and global scales.

3. Global Monitoring for Environment and Security, GMES (GMES, 2012) is a joint initiative by the European Commission and European Space Agency with an attempt at creating an autonomous and operational Earth observation capacity from multiple-sources. The main users of the GMES initiative are the policy-makers. The services of GMES can broadly be classified into: land, marine and atmospheric services, emergency and security services, and climate change services. The GMES uses its satellites and non-space assets to provide services to its users. Geoland 2, the context of project comes under the land, marine and atmospheric services of the GMES.

GEOLAND2(Geoland2, 2011a), a GMES services framework for land monitoring, aims at building user-driven operational land monitoring services. It derives from earlier projects of the European Commission and European Space Agency for the development of time- and cost-effective geo-information. It realizes its objectives through two of its layers, namely, the core mapping services and core information services (Geoland2, 2011b).

- The Core Mapping Services produce the basic earth observations on land-cover, land-use, the annual and seasonal vegetation cycles, radiation budget and water cycles from satellite observations.

- The Core Information Services are a set of thematic elements built upon the Core Mapping Services to generate products pertaining to European Environmental Policies and International treaties. Core Information Services also compares and evaluates the products of Core Mapping Services with products from other common approaches from an operational perspective. The Core Information Services include spatial planning, Agri-Environment Monitoring, Water Monitoring, Forest Monitoring, Land Carbon assessment, Natural Resource Monitoring in Africa and Global Crop Monitoring.

The vegetation biophysical products are part of the BioPar core mapping service. Thus, this thesis contributes to the interests of the BioPar core mapping service of the GEOLAND 2.

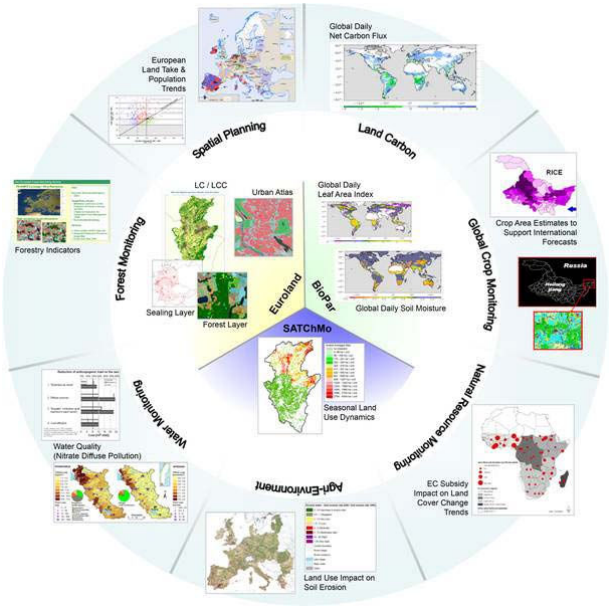

Figure 1. Operational Structure of GEOLAND2

This thesis is conducted under the premises of GEOLAND2 and aims at the characterization of vegetation at global scale from medium resolution satellite sensors to provide near-real time estimates of vegetation variables and the leaf area index (LAI) in particular.

1.1.1 Leaf Area Index in Ecological perspective

Leaf Area Index (LAI), is a biophysical variable accessible from remote sensing and identified to be essential for monitoring the earth climate processes (GCOS, 2010). LAI is defined as half the area of the green elements per unit horizontal area (Chen and Black, 1992). It provides an estimate of the mass-energy interface between the surface vegetation and the atmosphere. Thus, it is of interest to many studies including climate change, global carbon fluxes, land cover and crop monitoring.

LAI is one of the primary canopy variables that control the photosynthetic activity in the canopy. LAI along with canopy clumping index was found to be an important parameter for modeling plant growth and carbon cycle (Chen et al., 2003). LAI directly influences the photosynthetic capacity of the canopies and consequently, is a primary biophysical control on the net ecosystem productivity (Barr et al., 2004). LAI controls fAPAR, i.e. the fraction of photosynthetically active radiation (400-700 nm) absorbed by the canopy. The big-leaf model used for the canopy photosynthesis estimation requires empirical curvature factors in response to irradiance. This curvature factor, for any given distribution of leaf nitrogen content in the canopy, is dependent on LAI and leaf nitrogen content (De Pury and Farquhar, 1997). This observation led (De Pury and Farquhar, 1997) to treat the sunlit and shaded leaves separately to obtain improved estimates of photosynthesis as compared to using fAPAR. Further, from studies on Maize canopies

the LAI is found to be predominant factor for describing the canopy processes where water stress varies gradually over long periods (Steduto and Hsiao, 1998).

LAI is also an important parameter for hydrological studies related to water stress. (Boulet et al., 2000) proposed a water-energy balance model using LAI estimated from remote sensing observations to describe the canopy structure. LAI is also one of the primary input parameters for hydrological models used for evapotranspiration and energy-balance studies (Chen et al., 2005;Anderson et al., 2008;Vinukollu et al., 2011;Yan et al., 2012;Qin et al., 2010;Consoli et al., 2006). The Vegetation Cover, fCover, plays a vital role in discriminating the contribution from soil and vegetation in the models used for the estimation of evapotranspiration. fCover is largely determined by LAI.

At the regional and global scales, (Running and Coughlan, 1988) established the importance of LAI as indicators of the effects of climate change on vegetation. Earlier studies had also identified LAI as a single important biophysical variable to describe vegetation characteristics in the context of mass and energy exchange (Botkin, 1986;Wittwer, 1983). More recently, (Buermann et al., 2001) demonstrated the significance of this variable for climate variations. (Lucht et al., 2002) explained the early spring budburst to be largely due to the increase in surface temperature using LAI as input for the primary production and the carbon exchange. These studies along with other similar studies have identified the LAI to be one of the essential climate variables for studies on the effects of climate change.

This thesis will focus on LAI due to its significance in understanding the soil-vegetation-atmospheric processes as briefly outlined earlier. It is also an indicator of the health of vegetation.

1.2 Estimation of LAI

LAI can be estimated from ground measurements based on direct or indirect methods (Weiss et al., 2004;Jonckheere et al., 2004). Direct methods like area harvesting for destructive sampling, using allometric relations and litter traps are measurement intensive. Indirect methods like LICOR-2000, TRAC or hemispherical photography (‘CanEye’) use the penetration of light through the canopies to estimate the LAI (Lang and McMurtrie, 1992). Although much efficient than destructive methods, they still need intensive ground sampling in case of coverage of large domains. Therefore, ground-based methods cannot be used for estimations at global scale or for long periods due to economic and labor constraints.

Remote sensing observations from satellite sensors provide an opportunity to estimate LAI on a global scale for long periods at high revisit frequency. Medium resolution satellite sensors like MODIS (Knyazikhin et al., 1999), AVHRR (Baret et al., 2011a) and

VEGETATION (Baret et al., 2007;Weiss et al., 2007a) provide long term estimation of LAI. Canopy reflectance at various wavelengths observed by the sensors is then used to retrieve LAI. Canopy reflectance results from the radiative transfer processes within the canopy and includes the contribution from the soil background. While forward modeling establishes a causal relation between the biophysical variable of interest, LAI, with the observed reflectance measurements the retrieval is an inversion problem, which is usually ill-posed (Combal et al., 2003), i.e. several possible solutions can correspond to a given set of measurements.

The literature is rich with methods to retrieve LAI from reflectance measurements. These retrieval methods may consist in a simple empirical method that estimates the LAI by comparing the observed reflectance values to an experimental dataset, or may consist in a more complex method involving the inversion of a radiative transfer model. The different methods proposed in the literature for the retrieval of LAI could broadly be classified into biophysical variables driven methods and radiometric data driven methods (Baret and Buis, 2008).

1.2.1 The biophysical variable driven methods

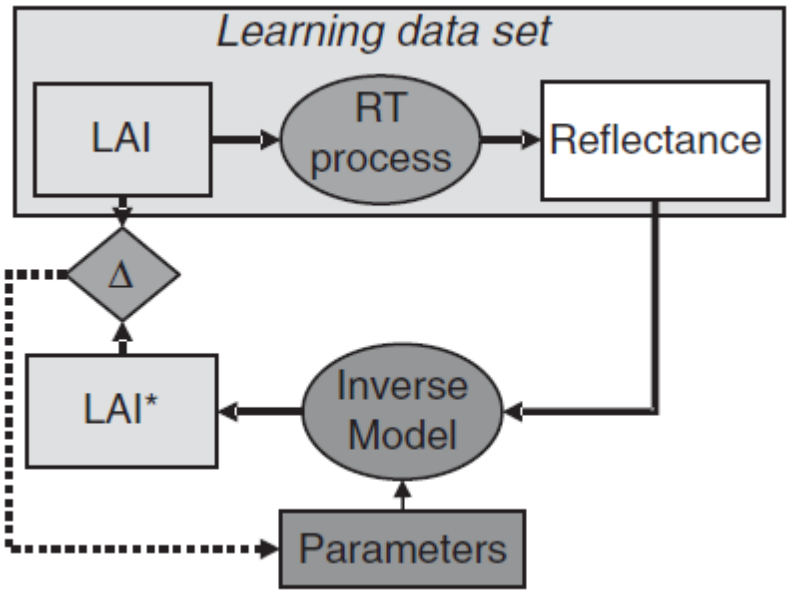

Figure 2. Calibration of the Inverse Model for the estimation of LAI. 'Δ' is the cost function to be minimized (Baret and Buis, 2008)

The *biophysical variable driven methods* consist in fitting an inverse parametric model between the LAI and the reflectance values using a learning dataset. This inverse model can then be used to retrieve LAI estimates from observed reflectance measurements. The learning dataset with LAI and reflectance values are generated from ground-based experiments or from radiative transfer models (RTMs). The early methods for the

retrieval of LAI were based on calibration of the parametric models over an experimental dataset. The reflectance values at different spectral bands are combined to produce Vegetation Indices (VIs) to reduce the influence of soil reflectance and atmospheric effects (Baret and Guyot, 1991). These VIs are then calibrated over experimental datasets to estimate LAI (Price, 1993;Huete, 1988;Wiegand et al., 1990;Richardson et al., 1992;Wiegand et al., 1992). These VIs were later used to derive LAI maps at global (Feng et al., 2006) and regional scales (Chen et al., 2002a;Pandya et al., 2006;Chen et al., 2002b). However, the need for a well representative training dataset sampling the whole set of possible vegetation types and observational conditions limits the use of this approach. This method is also prone to measurement errors (Dong et al., 2006;Fernandes and G. Leblanc, 2005) and errors due to scaling between ground and satellite measurements (Morisette et al., 2006). These limitations were largely overcome by the use of RTMs to simulate canopy reflectances as simulations covering a wide range of situations could be easily generated. Many operational algorithms use RTMs to simulate canopy reflectance for producing LAI estimates based on VIs (Gobron et al., 2000;Knyazikhin et al., 1999;Roujean and Lacaze, 2002). However, the limitation in this approach arises from the limited use of spectral bands to compute the VIs. Recently, artificial neural networks (ANN) are used to retrieve the biophysical variables from the satellite reflectance data (Verger et al., 2008;Verger et al., 2011a). ANNs are preferred over the VIs because they allow including more bands as well as their ability to provide better interpolation. ANNs are very efficient from the perspective of operational algorithms as the radiative transfer model is run offline.

1.2.2 The radiometric data driven methods

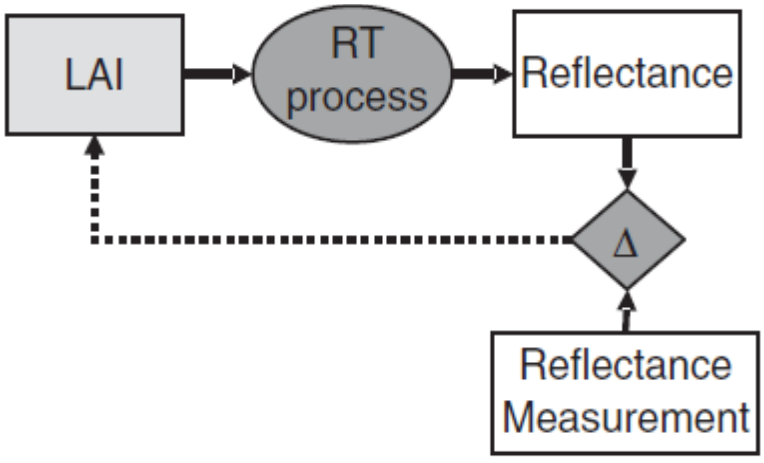

Figure 3. Radiometric based approach estimating LAI from the training and observed reflectances data. 'Δ' is the cost function to be minimized (Baret and Buis, 2008)

The radiometric data driven approaches, aim at finding the best match between the observed reflectance and the one simulated by the radiative transfer model. The training dataset may be generated using RTMs or by other experiments. The performances of

these approaches depend on the minimization algorithms and the uncertainties associated with the canopy configurations and measurements. Several methods have been proposed to find the best match for a set of observed reflectances: classical iterative optimization (OPT), simulated annealing (Bacour, 2001), genetic algorithms (Fang et al., 2003), Look-Up Tables (LUTs) and Monte-Carlo Markov Chains (Zhang et al., 2005). Of these methods, OPTs and LUTs are the most widely used approaches for the estimation of LAI. The OPT approaches use classical optimization algorithms like Quasi-Newton optimization to minimize a cost function defined over the observed and simulated reflectance. The cost functions span from simple least squares to more complex functions involving observational configurations. (Goel and Deering, 1985) using the SAIL canopy reflectance model (Verhoef, 1984) and least square metrics, found that the retrieved LAI values better match the model when the maximum view zenith angle is lower than 50°. (Jacquemoud et al., 1995) coupled the PROSPECT model for leaf optical properties (Jacquemoud and Baret, 1990) and the SAIL canopy reflectance model to retrieve the biophysical properties of sugar-beet canopies. It was found that the canopy or leaf structural variables such as the leaf mesophyll structure parameter, the LAI, the leaf inclination angle and the hot-spot size, were poorly estimated. However, by assigning one of the structural parameters to a given value, better LAI estimates were achieved. (Bicheron and Leroy, 1999) studied the retrieval of LAI from global satellite observations and found the LAI retrieval to have a root mean square error of about 0.7. Further, (Combal et al., 2002) showed that the LAI retrievals could be improved by exploiting auxiliary information on the canopy. In spite of its success, the OPT approaches may get trapped in a local minima of the cost function and need large computational resources. The need for computational resources is further compounded by the need to process large images making these approaches highly unsuitable for operational purposes. The Look-Up table, on the other hand, is conceptually simple. In addition, the search for the best fit between the observed reflectance and the simulated one is global (over the ensemble of simulated cases), thereby overcoming the local minima problem. Algorithm for the retrieval of LAI from MODIS observations is based on this approach (Knyazikhin et al., 1999). (Weiss et al., 2000) further showed that the accuracy of the model used to generate the LUT affects the performance more than the radiometric noise associated to the measurements. These LUT approaches require the sampling to better represent the canopy characteristics (Combal et al., 2003;Weiss et al., 2000).

Irrespective of the LAI retrieval approach, the estimates are not always accurate due to the uncertainties involved in the inversion process (Combal et al., 2003;Jacquemoud et al., 2009). Further, the selection of a radiative transfer model should result from a balance between realism and complexity: realism will generally be associated to more complex radiative transfer models with additional unknown variables making the inverse problem largely under-determined. One of the specific objectives of his thesis will be to

better understand the nature of the LAI retrievals from remote sensing observations on the assumptions embedded on canopy structure in the radiative transfer models.

1.3 Time Series Estimates of LAI

Many applications require the monitoring of LAI dynamics, i.e. considering time series of LAI estimates: (Maass et al., 1995;González-Sanpedro et al., 2008;Verbesselt et al., 2010) (Gitelson et al., 2003) for biomass estimation, (Ma et al.) for yield forecasting, (Kobayashi and Dye, 2005;Lucht et al., 2002;Piao et al., 2006;Buermann et al., 2001) for climate change, (Yan et al., 2012;Bacour et al., 2006a;Thenkabail et al., 2005;Roerink et al., 2000) land cover and (Hansen et al., 2002a;Moody and Johnson, 2001;Roy et al., 2008;Hilker et al., 2009) for land cover change. However, many of these applications require long, continuous and consistent time series of LAI.

In the last decades, LAI estimates have been generated using the observations from the medium resolution satellite sensors like MODIS (Knyazikhin et al., 1999), AVHRR (Baret et al., 2011b), VEGETATION (Baret et al., 2007;Weiss et al., 2007a), MERIS (Bacour et al., 2006a). However, these time series estimates are affected by discontinuities mainly due to cloud cover and are also noisy due to the uncertainties involved in the retrieval algorithms or sensor noise. Hence, it becomes necessary to process these time series estimates to fill the gaps (missing data) and smooth them.

The significance of high quality time series estimates of LAI had prompted many efforts to improve the quality of the estimates obtained from the satellite observations. These methods can be broadly classified into the smoothing methods and the gap-filling methods.

- ***The gap-filling methods*** intend to provide the most probable estimate for the missing observations by exploiting the characteristics of the LAI time series estimates; fusing estimates from different sensors (Roy et al., 2008;Hilker et al., 2009;Verger et al., 2011b), or by filling gaps by other empirical means. Some of the methods that exploit the characteristics of the time series estimates to fill gaps include: regression(Eilers, 2003), climatology (inter-annual mean/median) (Verger et al., 2011c), multi-resolution decomposition based methods(Kandasamy et al., 2012a;Sakamoto et al., 2010;Sakamoto et al., 2005), etc. The other empirical methods exploit the spatial characteristics or the understanding of the canopy processes to provide the most probable estimates for the gaps (missing data) (Chen et al., 2011;Zhu et al., 2010). Many methods for processing the LAI time series usually combine the gap-filling and smoothing methods as many smoothing methods could fail in the presence of larger amount and length of gaps.

- ***The smoothing methods*** aim at removing the noise in the estimates and to provide the temporal evolution of the LAI as would be expected from the studies of physical processes involved in the canopies. Some of the popular methods used for the smoothing of vegetation biophysical estimates include: median smoothing(Reed et al., 1994), curve fitting(Hermance et al., 2007;Bradley et al., 2007), Fourier analysis based methods(Bacour et al., 2006b;Roerink et al., 2000;Moody and Johnson, 2001), polynomial filtering (Chen et al., 2004;Verger et al., 2011b), function fitting(Jonsson and Eklundh, 2002;Jönsson and Eklundh, 2004), multi-resolution decomposition(Kandasamy et al., 2012a;Sakamoto et al., 2010;Sakamoto et al., 2005), penalized least square regression(Eilers, 2003), etc. Some of these smoothing methods could also fill gaps. For example, the Asymmetric Gaussian Functions could fill gaps when the length of the gaps is less than 0.2 years (Jonsson and Eklundh, 2002;Jönsson and Eklundh, 2004) and Savitzky-Golay smoothing method by (Chen et al., 2004) uses linear interpolation for the purpose. However, in these methods the objective was to realize a smoother time series.

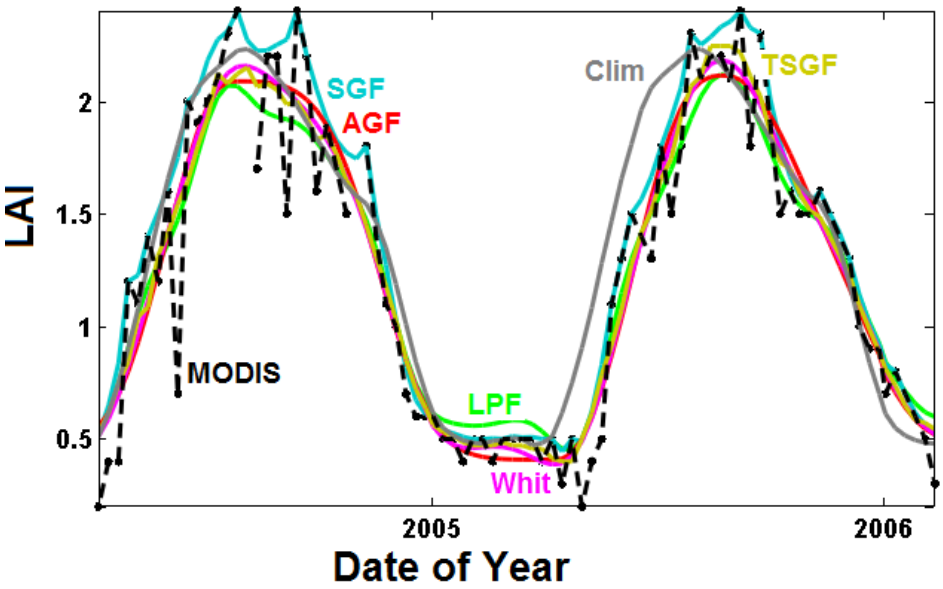

Figure 4. Illustration of the smoothing and gap-filling of MODIS 8-day LAI estimates of deciduous broadleaf forest BELMANIP2 site by some methods. Methods: SGF (Chen et al., 2004), AGF (Jonsson and Eklundh, 2002;Jönsson and Eklundh, 2004), LPF (Bacour et al., 2006b), Whit (Eilers, 2003), TSGF (Verger et al., 2011b) and Clim (climatology)

Moreover, for operational purposes continuous time series estimates are required to be produced in near real-time. In absence of observations from sensors due to cloud cover or system failure, the historical estimates are used to predict the current estimates. These near real-time estimations are challenging as compared to the gap-filling methods because only the past estimates could be used. Until recently not many efforts have been found in the literature. Some of the recent efforts, in this regard, include: real-time

estimation of MODIS LAI (Zhiqiang et al., 2009) using Ensemble Kalman Filter and a SARIMA (Seasonal Auto Regressive Integrated Moving Average (Box et al., 2008b)) model and the comparison of three statistical methods (Dynamic Harmonic Regression, Seasonal Trend Decomposition and SARIMA) to forecast LAI estimates using MODIS data (Jiang et al., 2010a). However, these studies fall short at examining the performances of the methods in relation to the discontinuities in the time series estimates and under global conditions.

1.4 Objectives and structure of the Thesis

The thesis is organized in three chapters according to the three following objectives:

1. **Effective Nature of LAI retrievals**: Many operational systems such as MODIS, AVHRR, or VEGETATION use Radiative Transfer Models (RTMs) to retrieve LAI estimates from medium resolution satellite observations. Consequently, many RTMs have been proposed varying from 1D to 3D characterization of canopies. However, the LAI retrieved by inversion of a 1D RTM model may differ from the actual LAI (LAI_{act}) of the canopies due to non-random distribution of leaves and the lack of structure details on canopy. The non-random distributions of leaves in the canopies give rise to the 'effective Leaf Area Index (LAI_{eff})'. LAI_{eff} of the canopies can be defined as the value of LAI_{act} that would produce the same directional variation of gap fraction (transmittance of the canopy assuming leaves as black) as that of a canopy with random foliage distribution. The LAI_{act} and LAI_{eff} are related by a parameter called 'clumping index'. However, due to the uncertainties involved in the inversion process and the violated assumptions on canopy structure in the RTMs used, the estimated LAI differs from both the LAI_{act} and LAI_{eff}. This retrieved LAI is an 'apparent' estimate for the canopy. From our initial study (Kandasamy et al., 2010) based on the SAIL (1D model) (Verhoef, 1984) and GEOSAIL (2.5D model) (Huemmrich, 2001), the apparent LAI retrieved was generally closer to the LAI_{eff} than to the LAI_{act}. This study will be here extended to include a 3D RTM, FLIGHT (North, 1996a).
2. **Comparison of the smoothing and gap-filling methods for MODIS LAI time series** to better select the most promising approaches. This specific objective is achieved by comparing 8 time series processing methods (Kandasamy et al., 2011;Kandasamy et al., 2012a): Iterative "Caterpillar" Singular Spectrum Analysis (ICSSA) (Kandasamy et al., 2012b), Empirical Mode Decomposition (EMD) (Huang et al., 1998), Low Pass filtering (Bacour et al., 2006b), Savitzky-Golay based smoothing (Chen et al., 2004), Asymmetric Gaussian Function (Jonsson and Eklundh, 2002;Jönsson and Eklundh, 2004), Whittaker Method, Time Series Smoothing and Gap filling (Verger et al., 2011b), Whittaker method (Eilers, 2003) and Climatology. This study used the MOD15A2 (Knyazikhin et al., 1999) LAI products of years 2000-2008 from BELMANIP2 sites (Baret et al., 2006) since

MODIS LAI product was known to incorporate a significant level of noise at the pixel level. The performances of the methods were evaluated based on their ability to faithfully reproduce the available observations, their ability to accurately predict missing observations and their ability to preserve phenology during reconstructions.

3. **Near real-time estimation of LAI.** From the methods assessed previously for smoothing and gap-filling, three methods were selected based on their ability to handle missing observations and make predictions: the Whittaker method (WM) (Eilers, 2003), the Gaussian Process model (GPM) (Rasmussen and Williams, 2005) and the GEOV2 algorithm (Baret et al., 2012). These methods were compared using the AVHRR daily LAI estimates (Baret et al., 2011b) for the period 1996-2000 for a set of globally representative sites. The AVHRR LAI time series estimates presented significant amount of noise and gaps and hence was chosen for the study. Further, AVHRR estimates are of interest to GEOLAND2 to extend back in time the VEGETATION observations. The performances of the methods to make near real-time estimates were studied by making realistic simulation of gaps, as a function of the amount and the length of the gaps.

2 Inversion of Spectral Reflectance for the estimation of biophysical variables

LAI is the biophysical variable of focus in this thesis, which is identified as one of the essential climatic variables for studies on earth vegetation. It is a canopy structure variable, whose estimation is affected by the highly ill-posed nature of the inversion process. The several validation results of the currently existing LAI products have demonstrated large discrepancies between products as well as with ground measurements (Weiss et al., 2007b;Camacho et al., 2010;Garrigues et al., 2008). In addition to the deviation from in situ measurements (which is affected by many other factors including scaling issues), the clumping and the differences in the definition of LAI highly explain the existing discrepancies between the several satellite products. In this context, it is of interest to better understand the nature of the estimates made and thereby better define them. This chapter of the thesis focuses on the nature of LAI retrievals made from satellite sensors by simulating the canopy reflectances using different radiative transfer models (RTMs).

For the purpose of analysis of the nature of LAI retrievals, a Look-Up Table (LUT) based approach for inversion is used to estimate LAI values, in this study. This approach was chosen due to its conceptual simplicity and its ability to retrieve solutions without getting trapped in local optima. This study uses three radiative transfer models for simulating the canopy reflectances at satellite sensors. They are: SAIL(Verhoef, 1984), GEOSAIL(Huemmrich, 2001) and FLIGHT(North, 1996b). These models differ in their assumption of the canopies and ranges from simple to complex representation of the canopy for the reflectance simulations. The SAIL model assumes the foliage in the canopy to be randomly distributed while the other models (GEOSAIL and FLIGHT) increasingly introduces details on the crown placement, shape factors, crown clumping, etc. Since, the canopy reflectances are also dependent on the foliage inclinations and their optical properties, and the soil optical properties many canopies with various structural and optical characteristics (leaves and soil) were simulated using the RTMs. A portion of which, are considered as the test cases to represent the satellite measurement of canopy reflectances. For the test cases to better represent the satellite measurements, these reflectances are corrupted by additive and multiplicative band dependent and independent Gaussian noise. Since the accuracy and the nature of the estimate are dependent on the choice of the possible solutions this chapter explores the influence of the uncertainty envelope on the radiometric observations over LAI retrievals at different view configurations. From this study, the nature of the LAI estimations and the influence on the uncertainty envelope are better understood.

ON THE EFFECTIVE NATURE OF LAI AS OBSERVED FROM REMOTE SENSING

Sivasathivel Kandasamy, Frédéric Baret and Aleixandre Verger
INRA-EMMAH, Site Agroparc, 84 914 Avignon, France

Abstract
Biophysical variables, particularly Leaf Area Index (LAI), estimated from medium resolution satellite observations provides valuable insights on the canopy structures and plays a vital role in understanding the vegetation-atmosphere interactions. LAI estimated over long time periods, at a global scale, are necessary to better monitor the impacts of climate change on vegetation and describe the earth-dynamic processes. These LAI estimates are usually estimated by inverting radiative transfer models (RTMs) from the radiometric data observed at the satellite sensor. However, due to the 'violated' assumptions of observed canopies, the uncertainty in the canopy parameters used as inputs for the RTMs, and the assumptions/uncertainties in retrieval algorithms of choice the retrieved LAI often deviate from the actual LAI (LAI_{act}). The retrieved LAI is also different from effective LAI (LAI_{eff}), which is related to the LAI_{act} through the 'clumping index' used to quantify the non-randomness of the foliage distribution in the canopies. Hence, the retrieved LAI is only an 'apparent' LAI estimate (LAI_{app}) of the canopy for an observed reflectance measurement and incorporates the uncertainty due the assumptions in the retrieval algorithm. In this study, the dependency of the LAI retrieval on the uncertainty envelope over the potential solutions for an observed reflectance is explored through the inversion of 3D RTM, FLIGHT. The study is conducted by simulating 'pseudo-reflectances', optical properties not a function of any particular wavelength, of the canopies and then estimating the canopy reflectances at SENTINEL 2 spectral bands over three view angles using the spectral invariance property of the RTM. The study follows a LUT based approach for the inversion of a 3D radiative transfer model, FLIGHT. The size of the uncertainty envelope is defined as a factor of the noise level in the observed reflectances and the factor is referred to as the 'Confidence Interval (CI)'. This study finds that optimum retrieval performances are obtained for CI values around 1. Further, the use of multi-view configuration is found to have little advantage over the single-view configuration in the estimation of LAI. However, use of multi-view configuration is found to seriously deteriorate the number of successful retrievals, though the RMSE is found to be slightly better in many retrieval configurations (combinations of assumptions of the LUTs and Test cases). The study also finds that LAI_{eff} of canopies (both with random and clumped foliage) are better retrieved when the canopies in the LUT are with random foliage distribution.

Index Terms—Leaf Area Index, FLIGHT, Look-Up Table, RTM Inversion.

1 INTRODUCTION

Monitoring terrestrial vegetation at global and regional scale is vital for addressing modern challenges due to climate change and to ensure food security. Many missions such as Moderate resolution Imaging Spectroradiometer (MODIS) (Myneni et al., 2002), Advanced Very High Resolution Radiometer (AVHRR) (Ganguly et al., 2008), VEGETATION (Feng et al., 2006;Baret et al., 2007), etc., have been launched, over the decade, to provide timely vegetation estimates at global scale for long time periods for the purpose. Due to the role of vegetation in regulating the mass-energy exchange between surface and atmosphere, their dynamics are of interest for studies on the effect of climate change (Pettorelli et al., 2005;Kobayashi et al., 2007), on global carbon fluxes (Wylie et al., 2007;Schubert et al., 2010), crop monitoring (Kastens et al., 2005;Dente et al., 2008;Becker-Reshef et al., 2010), etc. The Global Climate Observing System (GCOS), setup with the intention of providing comprehensive information on earth climate processes, has identified three vegetation biophysical properties as essential climate variables, namely, Land Cover, Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) and Leaf Area Index (LAI) (GCOS, 2010). Of these variables, LAI is a structural variable describing the foliage placement in the canopy. Due to its significance in the ecosystem processes and the challenges involved in its estimation, the variable has attracted considerable interest from the scientific community on both its application and improvement.

LAI defined as half the green area per unit horizontal surface (Chen and Black, 1992b), defines the surface of the interface for vegetation-atmospheric mass-energy exchanges. Many studies have established its significance from the ecological perspective. Studies have established LAI to be an important parameter to model plant growth and carbon cycle (Chen et al., 2003). LAI defines the area for photosynthesis in the canopies and hence exhibit a direct biophysical control on the net ecosystem productivity (Barr et al., 2004) and fAPAR (Fraction of Absorbed Photosynthetically Active Radiation) (De Pury and Farquhar, 1997). LAI has been identified by many studies as an important parameter in the studies on water stress (Steduto and Hsiao, 1998) and in water-energy models (Boulet et al., 2000). This variable is an important parameter for hydrological models used for evapotranspiration and energy-balance studies (Anderson et al., 2008;Chen et al., 2005a;Consoli et al., 2006;Qin et al., 2010;Vinukollu et al., 2011;Yan et al., 2012). Studies have established LAI to be an important biophysical variable for studies on climate change (Buermann et al., 2001;Lucht et al., 2002;Running and Coughlan, 1988) and mass-energy exchanges (Wittwer, 1983). Hence, there is a clear interest in the observation of this variable at both regional and global scales.

For studies at global-scale, LAI estimates produced from the observed radiometric data by the different the sensors (MODIS, AVHRR, VEGETATION, etc.) onboard the earth orbiting satellites are valuable due to their global coverage and high revisit frequency over long time periods. The canopy reflectances observed by these sensors at multiple wavelengths are used to estimate the LAI of surface viewed. This estimation of LAI is an ill-posed inverse problem (Combal et al., 2003) i.e. several possible estimates of LAI could be realized for an observed reflectance over a range of potential canopies. Due to the significance of LAI and its ill-posed nature, many methods have been proposed in the literature for estimating the LAI from satellite observations. These methods could broadly be classified into: biophysical variable driven methods and radiometric data driven methods (Baret and Buis, 2008).

The biophysical variable driven methods estimate LAI using a parametric model between observed reflectances and LAI, calibrated over a training data set generated either from field measurements or simulations using radiative transfer models (RTMs). Typically, these methods combine the reflectance values at different spectral bands to produce Vegetation Indices (VIs). This step is done to reduce the effect of soil in the observed reflectances (Baret and Guyot, 1991). The VIs of the training dataset is used to calibrate the inverse parametric models, to be used for the estimation of LAI (Huete, 1988;Wiegand et al., 1990;Price, 1993). However, these methods require a well representative set of training samples sampling the whole set of possible vegetation and observational conditions, which is difficult in many situations. These approaches are also prone to measurement and scaling errors (Fernandes and G. Leblanc, 2005;Dong et al., 2006;Morisette et al., 2006). Many of the operational algorithms currently uses RTMs to overcome these limitations. However the use of VIs, constructed using a few of the spectral bands, greatly limits the usefulness of these approaches. Currently, artificial neural networks (ANNs) are increasingly being used for the estimation of LAI for their ability to include more spectral bands and provide better interpolation (Verger et al., 2008;Verger et al., 2011).

The radiometric driven methods estimate the LAI by finding one or more matches between the observed reflectances and that in the training data set. These training datasets may be obtained through RTM simulations or through ground measurements. Many methods have been proposed to find the best matching reflectances based on classical iterative optimization (OPT), simulated annealing (Bacour, 2001), genetic algorithms (Fang et al., 2003), Look-Up Tables (LUT) and Monte-Carlo Markov Chains (Zhang et al., 2005). Many cost functions have also been proposed, in the literature, ranging from simple least squares to more complex functions involving observational configurations. The performances of these methods are largely dependent on the choice of the cost-function, the efficiency of the optimization methods used and the associated uncertainties. The methods based on OPTs are further disadvantaged by the possibility of the algorithm getting trapped in local-minima and are computationally expensive. The LUTs, on the other hand, are conceptually simple and the search for the matching reflectances is global. For methods based on LUTs, the accuracy of the model used to generate the LUT affects the performance of the LAI retrieval more than the radiometric noise in the different spectral bands (Weiss et al., 2000). The LUT based approaches also require the LUT to be well representative of the canopies being estimated (Weiss et al., 2000;Combal et al., 2003).

Irrespective of the method used, the retrieved LAI from satellite observations almost always deviate from the actual LAI (LAI_{act}) of the canopies. This deviation could be attributed to the non-random distribution of canopy foliage. Spatial grouping of canopy foliage elements occur at several levels: leaf level, shoot level, branch level, crown levels, etc. (Ganguly et al., 2008;Chen et al., 2005b;Chen and Black, 1992a). This grouping may be due to biological and environmental factors, and would affect the light reflected by the canopies resulting in an observation that corresponds to a lower value for LAI (LAI_{act}). The LAI estimated from sensor measurements, effective LAI (LAI_{eff}), is related to the LAI_{act} through a parameter called 'Clumping Index' (Chen et al., 2005b;Chen and Black, 1992b). For a 'clumping index' of 1, i.e. for canopies with purely random foliage distribution, the LAI_{eff} is same as that of LAI_{act}. Thus, LAI_{eff} can be defined as the value of LAI that would produce the same directional radiometric variation as observed in a canopy with random foliage distribution. In this study, the clumping is restricted to level of crowns. However, the LAI estimated by inverting RTMs or parametric models deviate also from LAI_{eff} due to the assumptions and uncertainties in the retrieval algorithm and the observational configurations. Hence, it is clearly of interest to understand LAI retrieved from remote sensing observations. This study extends our previous efforts (Kandasamy et al., 2010) using 1D (SAIL (Verhoef, 1984)) and 2.5D (GEOSAIL (Huemmrich, 2001)) RTMs to 3D (FLIGHT (North, 1996)) RTMs with multiple view angles.

2 Approach, method and models

The study uses a 3D radiative transfer model, FLIGHT (North, 1996) for simulating the canopy reflectance at multiple view angles. FLIGHT RTM uses Monte-Carlo simulations and is computationally expensive. It restricts the number of possible simulations when limited computational resources are available. This restriction is partially overcome by exploiting the spectral invariant property of the RTMs (Huang et al., 2007;Knyazikhin et al., 2011;Schull et al., 2011), through which a wide range of canopy reflectance could be simulated for a given canopy architecture and observation geometry from its pseudo-reflectances. That is, instead of simulating the canopy reflectances for every possible leaf optical properties, canopy reflectances are computed for a set of leaf reflectance and transmittance. These 'pseudo-reflectances' of the canopiesdon't necessarilly correspond to actual leaf optical properties in given spectral bands such as the SENTINEL 2 ones but would allow to compute canopy reflectance for any leaf optical properties through interpolation process.

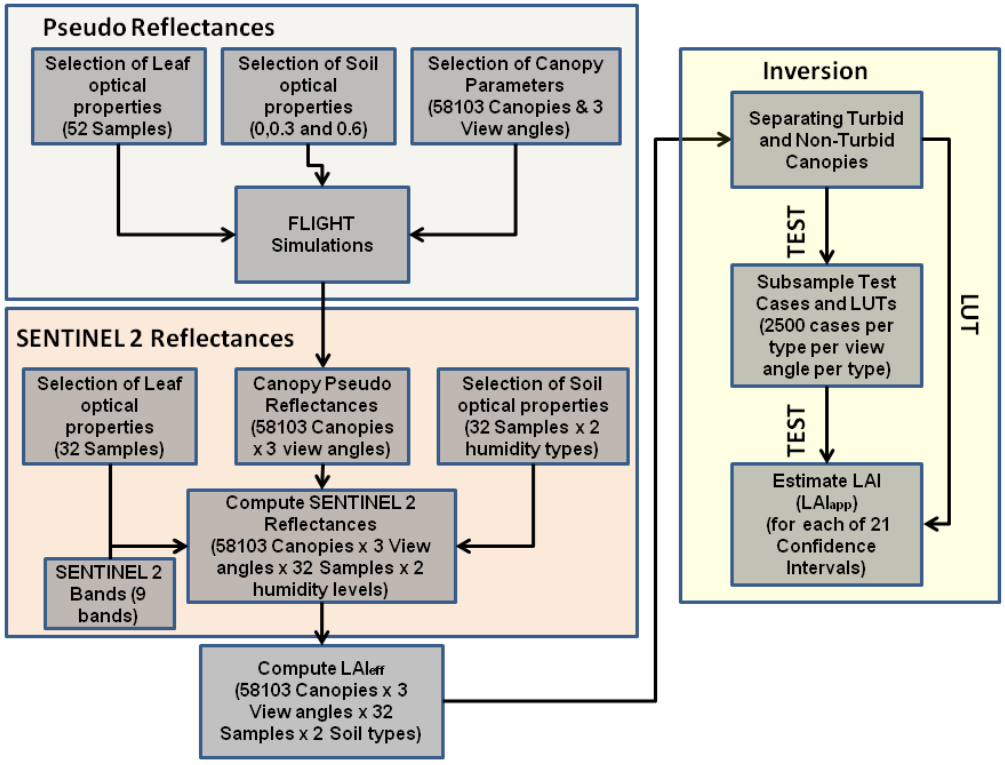

Figure 1. Overall Approach to the Study

Figure 1 shows the overall approach followed in this study. As could be seen from the figure, the study is organized in three phases: Creation of 'Pseudo-Reflectances', Creation of SENTINEL2 Reflectances and Study of Inversion. Of these, Creation of 'Pseudo Reflectances' and Creation of SENTINEL2 Reflectances is concerned with the creation of training and test data for studies on inversion.

2.1 Creation of Pseudo Reflectances for the Canopies

The simulation of canopy reflectance for a given canopy structure and geometrical configuration requires knowledge of leaf reflectance and transmittance as well as soil reflectance. In this study, the canopy reflectance for any actual leaf reflectance and transmittance values and soil reflectance values will be computed by interpolating within a limited set of cases simulated with FLIGHT for specific leaf reflectance and transmittance as well as soil reflectance values. In a first step, a sample of leaf reflectance and transmittance values will be proposed to allow accurate interpolation between the sampled values. Then the same will achieved for soil reflectance. Finally the way actual canopy reflectance is computed will be presented.

2.1.1 Selection of Leaf Optical properties

PROSPECT Leaf RTM (Jacquemoud and Baret, 1990) is used for simulating the leaf reflectances and transmittances for wavelengths ranging from 400nm to 2500nm. This model is based on the "plate model" of leaves and simulates leaf reflectances and transmittances by adjusting the input parameters: the mesophyll structure (N), chlorophyll concentration (C_{ab}), amount of dry matter (C_{dm}), water content (C_w), and brown pigment concentration (C_{bp}). The leaf reflectances and transmittances were obtained for a set of realistic PROSPECT input parameters (Table 1) for wavelengths 400nm to 2500nm. The water content $\mathbf{C_w}$ is computed from relative water content $\mathbf{C_{w\_rel}}$ as $\mathbf{C_w = C_{dm}/(1- C_{w\_rel})}$. Figure 2 (black dots) shows the distribution of the space of leaf realization in the leaf reflectance and transmittance feature plan.

Table 1. PROSPECT input parameters

Parameter	Minimum value	Steps	Maximum value
N	1.0	0.1	2.5
C_{ab} ($\mu g/$	15	10	100

cm^2)			
$C_{dm}(g/cm^2)$	0.003	0.005	0.025
$C_{w\ rel}$	0.65	0.05	0.85
C_{bp}	0	0	0

A set of 52 leaf optical properties (leaf reflectances and transmittances) is selected from the entire set to allow accurate interpolation in the space generated by the leaf reflectances and transmittances(Figure 2).

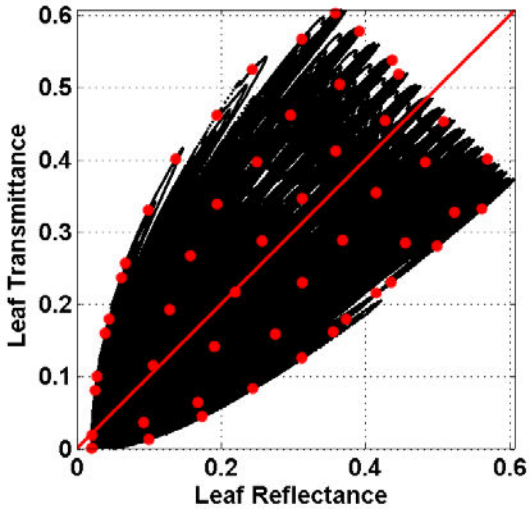

Figure 2. The leaf optical property space (black) with the selected set marked in red (dots).

These 52 leaf reflectances and transmittance values correspond to 52 'pseudo-bands'.

The accuracy of the interpolation at the canopy reflectance level was tested for different sample size.. The accuracy is found to increase quickly and continuously with the number of samples (Figure 3) and obviously decreases with the amount of leaves observed in the canopy and driven by the crown cover. For 52 Canopy reflectance was simulated with a RMSE better than 0.005 which was assumed satisfactorily according to the known uncertainties associated to canopy reflectance measurements.

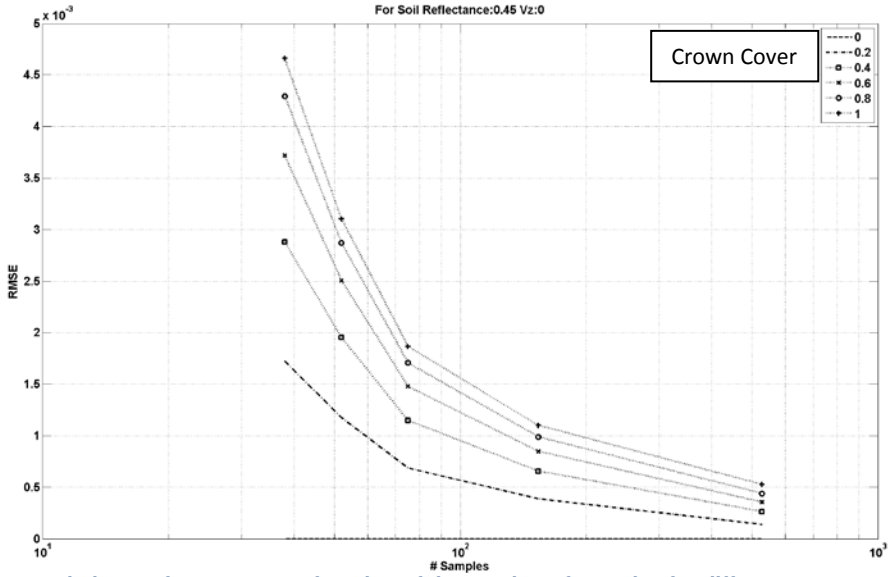

Figure 3. Interpolation performance as a function of the number of samples for different crown cover values

2.1.2 Selection of soil reflectance

Three soil reflectances 0, 0.3 and 0.6 were chosen to simulate the 'pseudo-reflectances' of the canopies. The reflectance values were chosen to facilitate the estimation of canopy reflectances for any actual soils. For this purpose, simulation of canopy over black soils (soil reflectance = 0) is required, as is indicated in 2.2.2 to be able to get the 'black soil' canopy reflectance required in the interpolation process. The other reflectances correspond to the mid-value and maximum value of observed reflectance over all wavelengths and humidity levels (Weidong et al., 2002). It should be noted that, similarly to the pseudo bands for leaf reflectance and transmittance, the soil reflectances correspond to no particular wavelength but were chosen to accurately compute canopy reflectance for any soil reflectance values from the canopy reflectance simulated for the 3 soil reflectance using the principle exposed later in section 2.2.2

2.1.3 Selection of Canopy Parameters

FLIGHT model (North, 1996) simulates the reflectance of a canopy described by the placement of crowns, crown and soil properties, and source-sensor configuration. Figure 4 shows a typical canopy in FLIGHT simulations.

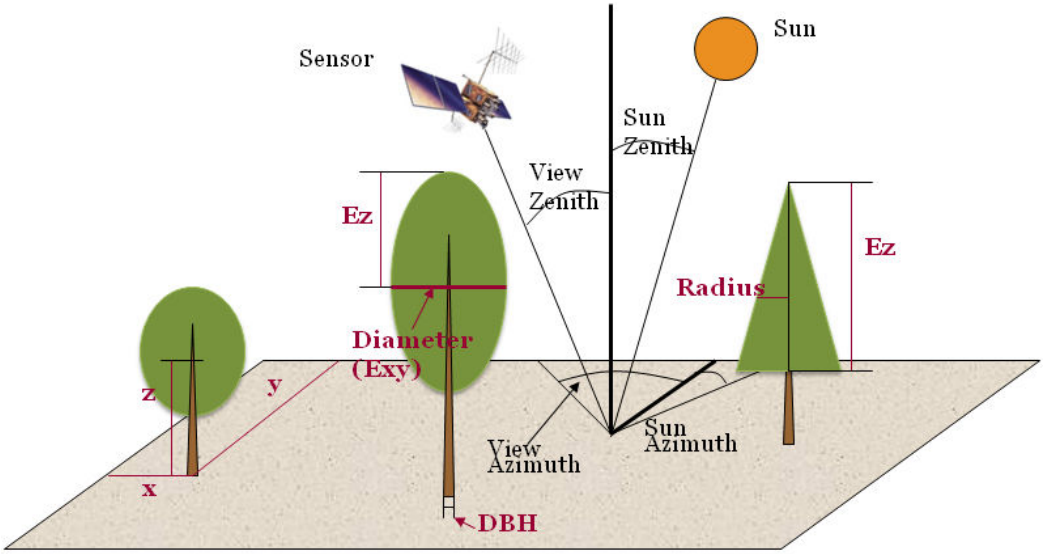

Figure 4. FLIGHT Canopy Scene

As could be seen from Figure 4, FLIGHT simplifies the canopies by using geometric primitives. In this study, For the sake of simplicity, the tree barks were neglected by assigning a nul trunk diameter. The crown shapes could be either conical or ellipsoidal through the selection of crown parameters. The sun zenith is fixed at 45° corresponding to a median configuration. The relative azimuth is assumed at 60°, which corresponds to the typical observational configurations of polar satellites. The canopy reflectances are simulated for a wide range of view zenith angles: 0°, 30° and 57°.

Parameter values and their allometric relations used for describing canopies vary with vegetation types. However, the distributions of these parameters are not purely random across canopies. The shape of the crowns, determined by the botanical and environmental factors, of the different vegetation types could easily be classified into circular (Crown Height = Crown Diameter), ellipsoidal (Crown Height > Crown Diameter) and oblong (Crown Height < Crown Diameter). We neglected the cases corresponding to conical crown shapes. Typical canopy parameter values were selected to represent a wide but realistic range of canopy structure, including forest as well as crop and grassland types. Some parameters are chosen as primary parameters for canopy struicture: crown height (Ez), Leaf area Index (LAI), Crown Cover (CC), Number of crowns in the scene, Average Leaf Angle (ALA) and Leaf size. The other parameters crown diameter (Exy), crown height, Leaf area density (LAD) are defined relatively to the

primary parameters through allometric relationships. The selection of the canopy parameters is detailed in the following steps:

1. Draw a value of canopy height (Ez) between 0.1m and 20m, which are typical height of vegetation
2. Draw value of the ratio Exy/Ez (ration between crown diameter and crown height), such that 0.5 < Exy/Ez < 2. This correspo,nds to typical crown shapes: elongated (Exy/Ez =0.5), spherical (Exy/Ez =1) and flat crown (Exy/Ez =2). Exy is then computed from this ratio.
3. Draw the ratio of height of trunk (Ht) to crown height (Ez), such that 0 < Ht/Ez < 0.54 corresponding to typical values, and compute Ht from this ratio.
4. Set DBH = 0, to neglect the trunks.
5. Draw a value of scene LAI between 0 and 10 which corresponds to a wide range of vegetation amount.
6. Draw the Crown Cover (CC) from 0.1 to 1.0. This parameter is important, as it drives most of the clumping at the stand level.
7. Compute Leaf area density (LAD): The scenes were simulated with a fixed number of trees (crowns). 80 trees was selected as a compromise between computation efficiency and statistical representativeness. The tree density (D_{tree} , number of trees per unit area) was therefore varied by adjusting the size of the scene (XDim²): D_{tree} = 80/ XDim² . Thus, the leaf area density was computed as: LAD = LAI/D_{tree}/Crown Volume. To avoid unrealistic silulations in, the following constraints were further applied: leaf area density lower than 10m²/m^3 and the average LAI under a tree (LAI/D_{tree}/(π Exy²/4) lower than 20 m²/m².
8. Draw Average Leaf Angle Distribution (ALAD) from 30° to 80° and compute the corresponding LIDF using the ellipsoidal distribution function (Campbell, 1990)
9. Draw the leaf size (Ls, leaf radius) from 0.005 m to 0.1 m, which are typical leaf sizes.

A total of 58103 valid canopies were created .The distribution of the canopy parameters and their relative variations are shown in Figure 5.

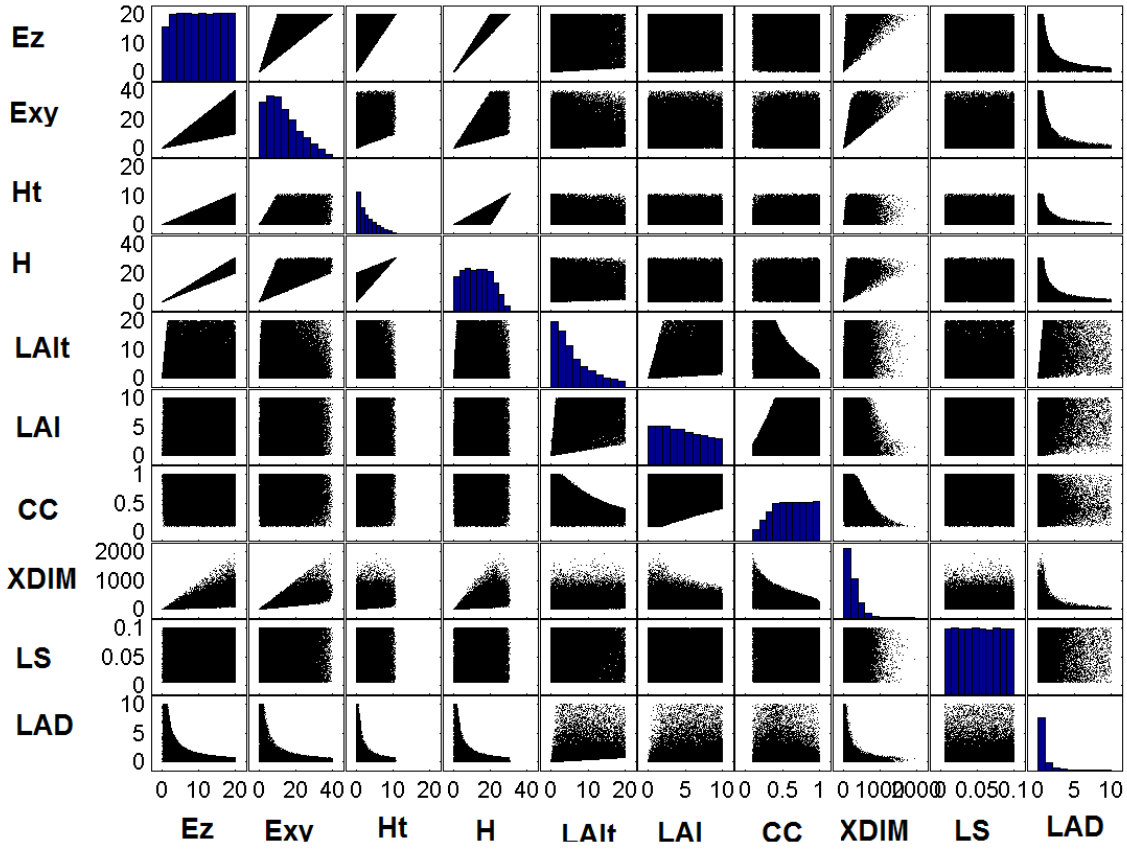

Figure 5. Distribution of Canopy parameters

To get more homogenous (while random) tree distribution patterns, the scene is divided evenly into 16 cells (4x4) in each of which 5 crowns are placed at random in x and y dimensions. Because of possible overlap between crowns, the dimension of the scene is computed from the relationship between the ratio of Crown diameter to surface dimension (Exy/XDim) and the crown cover (CC) (Figure 6). This relation is obtained by in silico experiment: crowns were placed randomly according to the scheme proposed for given scene dimensions and the corresponding crown cover (CC) was computed. . The relation between Exy/XDim is then summarized using a polynomial function. Using the polynomial, the ratio of 'Exy/XDim' could be computed for a given value of CC. With 'Exy' known, the scene dimension 'XDim' is computed. This process allows getting relatively even but randomly distributed crowns, focusing mostly on the clumping at the crown level.

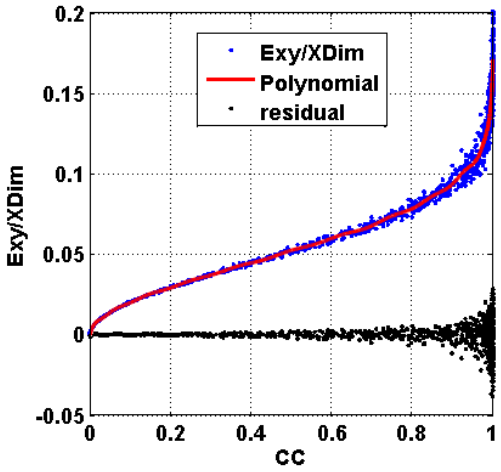

Figure 6. Relation to compute Scene dimensions

2.2 Computation of SENTINEL 2 canopy reflectance

With the canopy 'pseudo reflectances' simulated for for 52 optical properties (§2.1.1) and 3 soil properties (§2.1.2), the canopy reflectances for any wavelengths and leaf and soil characteristics could be computed. In addition, canopy reflectance corresponding to black leaves (reflectance=0, transmittance=0) and white soil (soil reflectance=1) was also simulated to get an estimate of the effective leaf area as we will see later. We should now select the actual Leaf and Soil optical properties desired to represent the expected range of variability. The main 9 bands of SENTINEL 2 satellite will be considered.

2.2.1 Selection of Leaf and Soil Optical Properties

As in 2.1.1, PROSPECT leaf optical model was used to generate a set of 32 leaf reflectances and transmittance for the SENTINEL 2 spectral bands. The sampling was achieved using a hypercube latin scheme, assuming the independency between each pair of parameters. The input parameters for the model are given in Table 2 and their relative distributions are shown in Figure 7.

Table 2. Input parameters for leaf optical properties in SENTINEL 2 bands

	Variable	Minimum	Maximum	Nb_Class	Distribution
Leaf Properties	**N**	1,0	2,5	2	uniform
	Cab $(\mu g/cm^2)$	30,0	100,0	2	uniform
	Cdm (g/cm^2)	0,002	0,02	2	uniform
	Cw_rel	0,65	0,9	2	uniform
Nb_Sims for each Canopy & observational configuration				**32**	

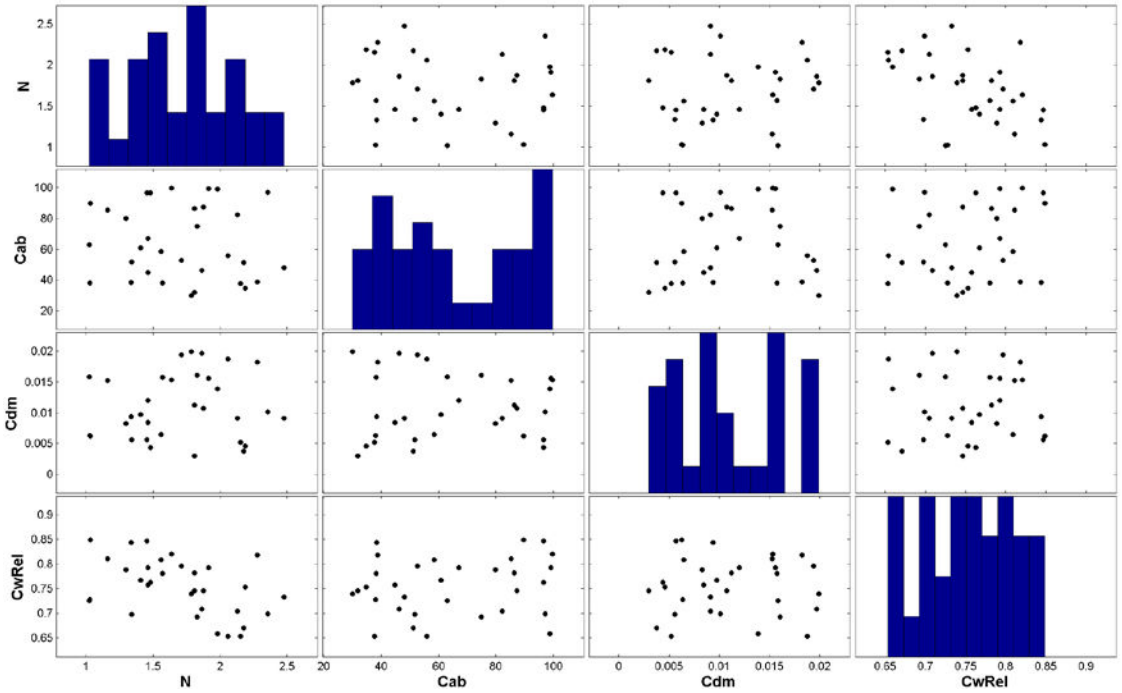

Figure 7. Distribution of Leaf Input Parameters to simulate actual leaf Reflectance and transmittance for SENTINEL 2 bands

From Figure 7, it could be seen that the parameters are uniformly distributed in their respective valid ranges and are independent of one another. The canopy reflectances are to be computed for the leaf reflectances and transmittances for each of these 32 combinations of input parameters.

The soil reflectances for the SENTINEL 2 bands are selected from the soil reflectance library (Weidong et al., 2002) for 2 humidity levels: dry and slightly moist soils. From a set of 104 soil types, 32 soils were chosen randomly for each of the humidity levels (Figure 8).

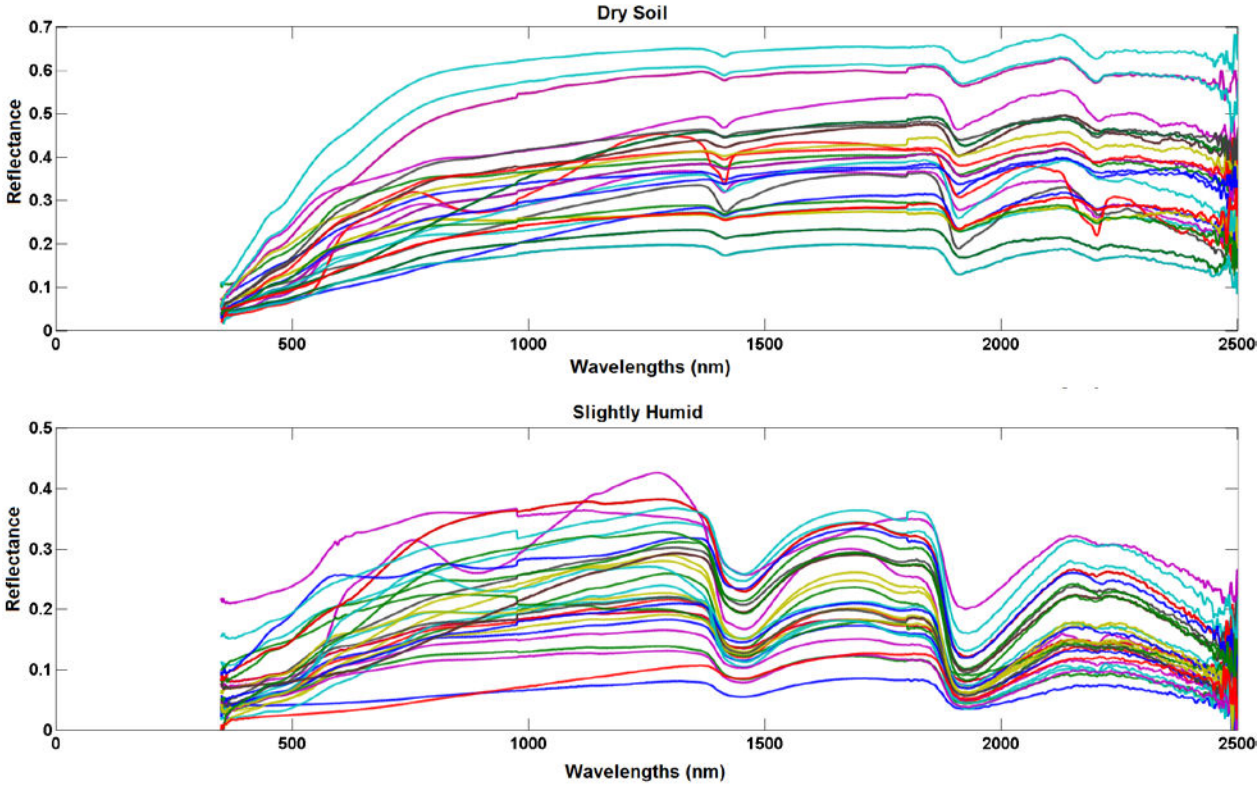

Figure 8. Spectra of the Selected Soils

To reduce the noise in the bands, the soil reflectances for SENTINEL 2 bands are computed as the mean of the reflectances at ±10nmcentered over the Sentinel 2 bands. The relative variations of the selected soils at the two humidity levels are shown in Figure 9 and Figure 10, respectively.

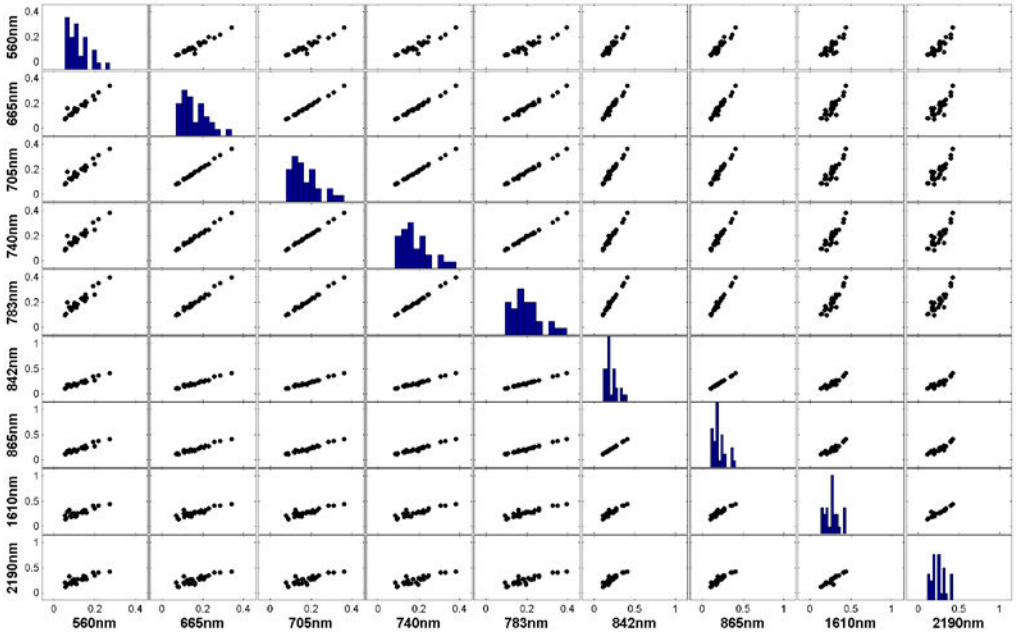

Figure 9. Soil Reflectance at different SENTINEL 2 spectral bands for dry soils

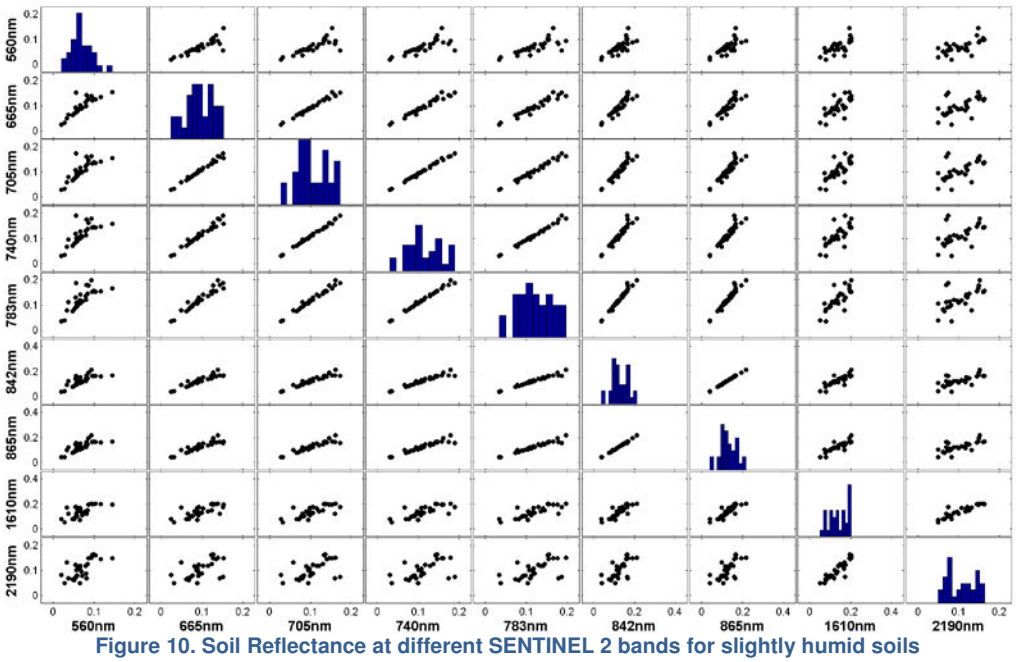

Figure 10. Soil Reflectance at different SENTINEL 2 bands for slightly humid soils

From Figure 9 and Figure 10, the soil reflectances for some spectral bands appear to be correlatedas expected from the soil-line concept of the soil reflectance (Baret et al., 1993b, a).

2.2.2 Computation of SENTINEL 2 Reflectances

Computation of canopy reflectances for the actal leaf and soil optical properties (2.2.1) is done in two steps:

- Computation of canopy reflectances for actual leaf optical properties and,
- extending this canopy reflectance for the actual soil reflectance.

The spectral invariance property states that canopy reflectance is a monotonic and continuous function of leaf optical properties for a given canopy structure and observational configuration (Lewis and Disney, 2007). The actual canopy reflectance $\rho_c(S, \Omega, [\rho_l, \tau_l], \rho_s)$ corresponding to the actual canopy reflectance and transmittance $[\rho_l, \tau_l]$ for a given canopy structure ($S = [LAI, CC, Ez, Exy, Ht, D_{tree}, LIDF, Ls]$), observational configuration (Ω) and soil reflectance ρ_s is therefore approximated by a linear interpolation between the closest values of the neighbor pseudo reflectance values $\rho_c(S, \Omega, [\rho_i, \tau_i], \rho_s)$:

$$\rho_c(S, \Omega, [\rho_l, \tau_l], \rho_s) = \rho_c^{closest}(S, \Omega, [\rho_l, \tau_l], \rho_s) + \left(\rho_l - \rho_l^{closest}\right)\frac{\Delta\rho_c}{\Delta\rho_l} + \left(\tau_l - \tau_l^{closest}\right)\frac{\Delta\rho_c}{\Delta\tau_l} \quad (1)$$

Where $\rho_c^{closest}$ is the pseudo canopy reflectance value corresponding to the pseudo leaf optical properties $\left[\rho_l^{closest}, \tau_l^{closest}\right]$ the closest to the actual ones, $[\rho_l, \tau_l]$, $\frac{\Delta\rho_c}{\Delta\rho_l}$ and $\frac{\Delta\rho_c}{\Delta\tau_l}$ being respectively the local dericative of canopy reflectrance to leaf reflectance and transmittance. Expression (1) corresponds to the first order term of a Taylor development of canopy reflectance. It therefore assumes only local linearity of canopy reflectance with leaf optical properties.

Soil forms the background and, therefore, affects the canopy reflectances. However, the relation between the canopy reflectance and the soil is not linear. Hence the spectral invariant property is best used in canopies with a dark background or dense canopies (GC1) (Knyazikhin et al., 2011). The canopy reflectances for the new soil reflectances are obtained by considering the radiant flux in the canopies. In this case, the canopy reflectance could be expressed as a sum of canopy reflectance with dark soil (soil reflectance = 0), reflectance from the exposed soil and a function of the multiple soil-canopy interactions (2).

$$\rho_c = \rho_c(S, \Omega, [\rho_l, \tau_l], 0) + \rho_s(\rho_c(S, \Omega\ , [0,0], 1)\rho_c(S, \Omega_{sun}, [0,0], 1) + (\frac{A}{1 - B\rho_s}) \quad (2)$$

Where $\rho_c(S, \Omega, [\rho_l, \tau_l], 0)$ is the black soil term of canopy reflectance (i.e. canopy reflectance when $\rho_s = 0$, and A and B two coefficients to be adjusted using the system of equations (2) applied to the 2 soil reflectance values for which canopy reflectance was computed: $\rho_s = 0.3$ and $\rho_s = 0.6$. The term $\rho_c(S, \Omega_{view}, [0,0], 1)\rho_c(S, \Omega_{sun}, [0,0], 1)$ corresponds to the fraction of soil viewed and illuminated, and the term $\frac{A}{1-B\rho_s}$ corresponds to the interaction between the soil and the canopy.

2.3 Computation of Effective LAI of the Canopies

The effective LAI (LAI_{eff}) is defined as the value of LAI that would produce the directional gap fraction variation as observed in a canopy with random foliage distribution. Hence, for a canopy with random foliage distribution, the gap fraction and the reflectances of a hypothetical canopy with black foliage and white soil should be equal, provided the source and sensor are in the same direction. However, in our study the sun is fixed at an zenith angle of 45° and the canopy reflectances were simulated for 0°, 30° and 57° zenith angles. Hence, the reflectance in a given view direction for the canopy would be the product of the soil illuminated by the sun and that could be viewed since the view direction is far from the source direction (no hotspot configuration). Thus, the LAI_{eff}, could be computed by iteratively minimizing the following cost function :

$$J = \sum_{\Omega}^{\Omega} \left(Po\left(LAI_{eff,} ALA, \Omega\ \right) Po\left(LAI_{eff,} ALA, \Omega_{sun}\right) - \rho_c(S, \Omega\ , [0,0], 1)\rho_c(S, \Omega_{sun}, [0,0], 1) \right)^2 \quad (3)$$

Where $Po\left(LAI_{eff,} ALA, \Omega\right)$ is the gap fraction in direction Ω for a turbid medium canopy that is computed classically as:

$$Po\left(LAI_{eff,} ALA, \Omega\right) = e^{-\frac{G(ALA,\Omega)}{\cos(\Omega)} LAI_{eff,}}$$

Where $G(ALA, \Omega)$ is the projection function, i.e. the projection of a unit leaf area in direction Ω.

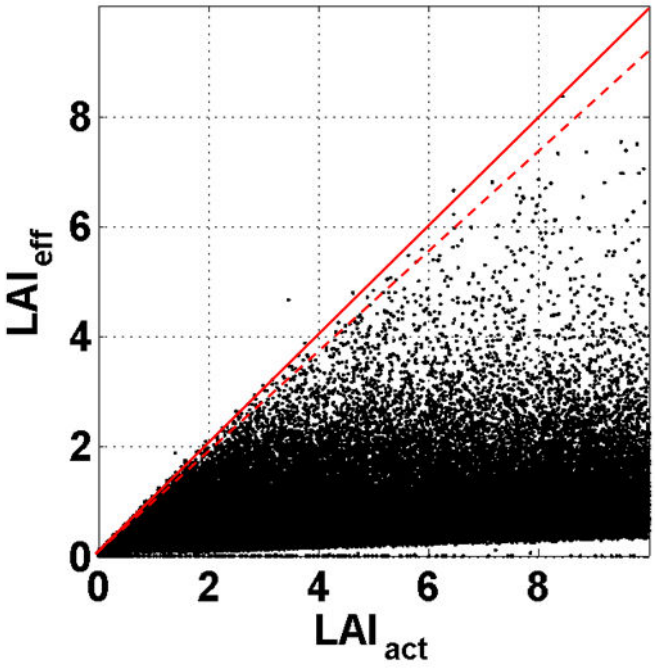

Figure 11. Relation between the computed LAI_{eff} and LAI_{act}. The solid red line corresponds to pure turbid medium cases and the dashed red line corresponds to cases close to turbid medium (LAIact-LAIeff <10%).

Figure 11 shows as expected that LAI_{eff} is generally an underestimation of the actual LAI (LAI_{act}). However for a few cases $LAI_{eff} \approx LAI_{act}$, which corresponds to canopies with an almost random distribution of leaves. This proposed way to estimate LAI_{eff} appears to be a credible approximation.

2.4 Inversion process

2.4.1 Creation of Look-Up Tables and Test Cases

The simulated canopies were first separated into 'Turbid' (when LAI_{act}-LAI_{eff} <10%) and 'Non-Turbid' cases (when LAI_{act}-LAI_{eff} >10%) showing a significant degree of clumpiness at the stand level according to FLIGHT assumptions. For both turbid and non-turbid situations, 2500 cases were chosen randomly as test cases. The remaining cases were used to build the LUTs. Similarly, 2 LUTs were generated, corresponding respectively to turbid and non-turbid situations.

The reflectances for the test cases were corrupted with additive and multiplicative noises that are band dependent and independent, to mimic the measurement uncertainties. The multiplicative noise levels are set to 2% while the additive noise levels are fixed to 1% (3).

$$R_\lambda^{Test} = R_\lambda^{Canopy}\left(1 + (MD_\lambda + MI)\right) + AD_\lambda + AI \quad (4)$$

Where:

- MD , corresponds to band dependent multiplicative noise, set to 2%
- MI , corresponds to band independent multiplicative noise, set to 2%
- AD , corresponds to band dependent additive noise, set to 1%
- AI , corresponds to band independent additive noise, set to 1%

2.4.2 Retrieval Algorithm

A LUT contains a set of reflectances simulated by the FLIGHT RTM for various combinations of canopy parameters. The reflectances of the test cases contaminated with the radiometric uncertainties, are compared with those stored in the LUT. The solutions correspond to the cases in the LUT for which:

$$\rho_c^{test}\left(S^{test},\Omega_i,\left[{\rho_l}^{test}(\lambda_j),{\tau_l}^{test}(\lambda_j)\right],{\rho_s}^{test}(\lambda_j)\right)-\rho_c^{LUT}\left(S^{LUT},\Omega_i,\left[{\rho_l}^{LUT}(\lambda_j),{\tau_l}^{LUT}(\lambda_j)\right],{\rho_s}^{LUT}(\lambda_j)\right)<CI.\sigma(\lambda_j) \quad (5)$$

Where $\sigma(\lambda_j)$ correspond to the radiometric uncertainties for band λ_j and view direction Ω_i considered in the test cases (as well as in the LUTs!), and CI (confidence interval) is a multiplicative factor that should be close to 1.0 (Kandasamy et al., 2010). This means that all the cases that are in the LUT within a distance to the test case considered smaller than a confidence interval associated to the radiometric values are considered as solutions. The performances of the estimate depends obviously on the width of the confidence interval. in this study, CI was varied within $0.5 < CI < 1.5$. The solution is finally computed as the median of the LAI values corresponding to the individual cases verifying (5).

The inversion is completed for all the combinations of the LUT and Test Cases for turbid and 'all' (turbid and non-turbid) situations (Table 3). Not that for turbid situations, no distinction is made between LAI_{act} and LAI_{eff} since $LAI_{act} \approx LAI_{eff}$.

Table 3. The various inversion configurations considered in the study

			Look-Up Tables (LUTs)		
			Turbid	***ALL***	
			LAI_{act}	LAI_{act}	LAI_{eff}
Test Cases	***Turbid***	LAI_{act}	LUT(Turb)-Test(Turb)	LUT(All)-Test(Turb)	LUT(Turb)-Test(All)
	ALL	LAI_{act}	LUT(Turb)-Test(All)	LUT(All)-Test(All)	LUT(All)-Test(All)
		LAI_{eff}	LUT(Turb)-Test(All)	LUT(All)-Test(All)	LUT(All)-Test(All)

2.4.3 Metrics for performance evaluation

The performances of the estimates is evaluated based on the root mean square error metrics between the known actual or effective LAI values of the test cases (LAI(test)) and the retrieved LAI value called here estimated LAI (LAI(est)).

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(LAI(est)-LAI(test))^2}{n}} \quad (6)$$

Where $n=2500$ is the number of test cases. Because the LUT is not necessarily very dense, and because the criterion used for selecting the solution is relatively restrictive (all bands and view deirections considered must be within $CI.\sigma$), in a significant number of cases, no cases within the LUT were found to verify these conditions. The ability of the method to successfully estimate the LAI values for a given CI value is also proposed to evaluate the % success:

$$\%Success \quad =\left(\frac{No.of\ retrievals}{n}\right)\times 100 \quad (7)$$

3 Results and Discussions

3.1 Selection of the optimal CI

Performances as measured by the RMSE metrics as well as the $\%Success$ are first analyzed to define the optimal CI value. As a matter of fact, the RMSE generally decreases with CI then reaches a minimum before increasing with CI (Figure 12).This is very consistent with the previous observations of (Kandasamy et al., 2010) and corresponds first to a 'regularization' through an increasing number of solutions and then to solutions very different from the actual case when the CI is too large. RMSE will asymptotically tend to a fixed value defined by the variance of the LUT cases and the difference between the test case value considered and the median of the LUT values. The CI values corresponding to the minimum RMSE, is obtained generally around CI=0.8.

On the other hand, the $\%Success$ increases continuously with CI since the probability to get a case that verifies (5) increases with the size of the confidence interval (Figure 12). For CI=0.8, the $\%Success$ is relatively small and generally lower than 50%, particularly for the multi-view configuration, i.e. when considering an observational configuration where the 3 view angles are acquired concurrently (Figure 13). For this reason, we selected $CI = 1.0$ as being optimal, i.e. providing small RMSE values while having $\%Success$ relatively large and at least $\%Success > 50\%$ for single view configurations.

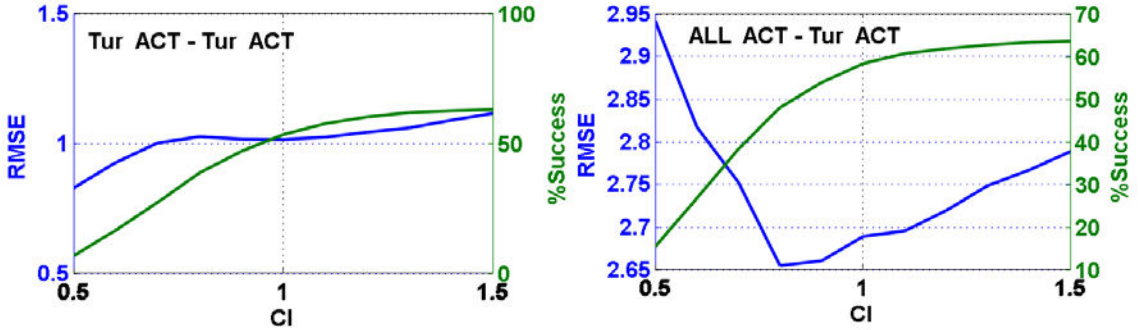

Figure 12. Variation of Retrieval Performance and its success as a function of CI for typical cases investigated for single view configurations.

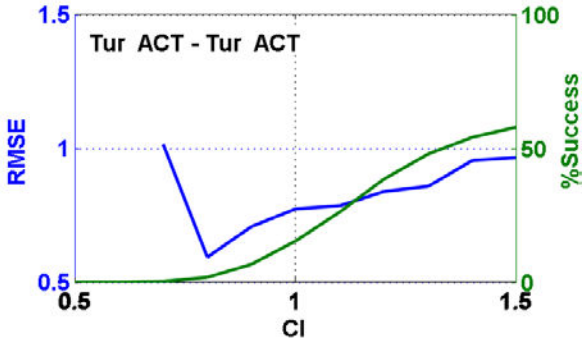

Figure 13. Variation of Retrieval Performance and its success as a function of CI for multiple view configurations.

3.2 Evaluation of performances for CI=1.0

The performances of the LAI retrievals are studied in both single view and multi-view configurations, in this study. In the single view configuration, the LAI retrievals are investigated for each of the 3 view directions considered. In the multi-view configuration, the three view directions are combined for the estimation of LAI values.

3.2.1 Single View Configuration

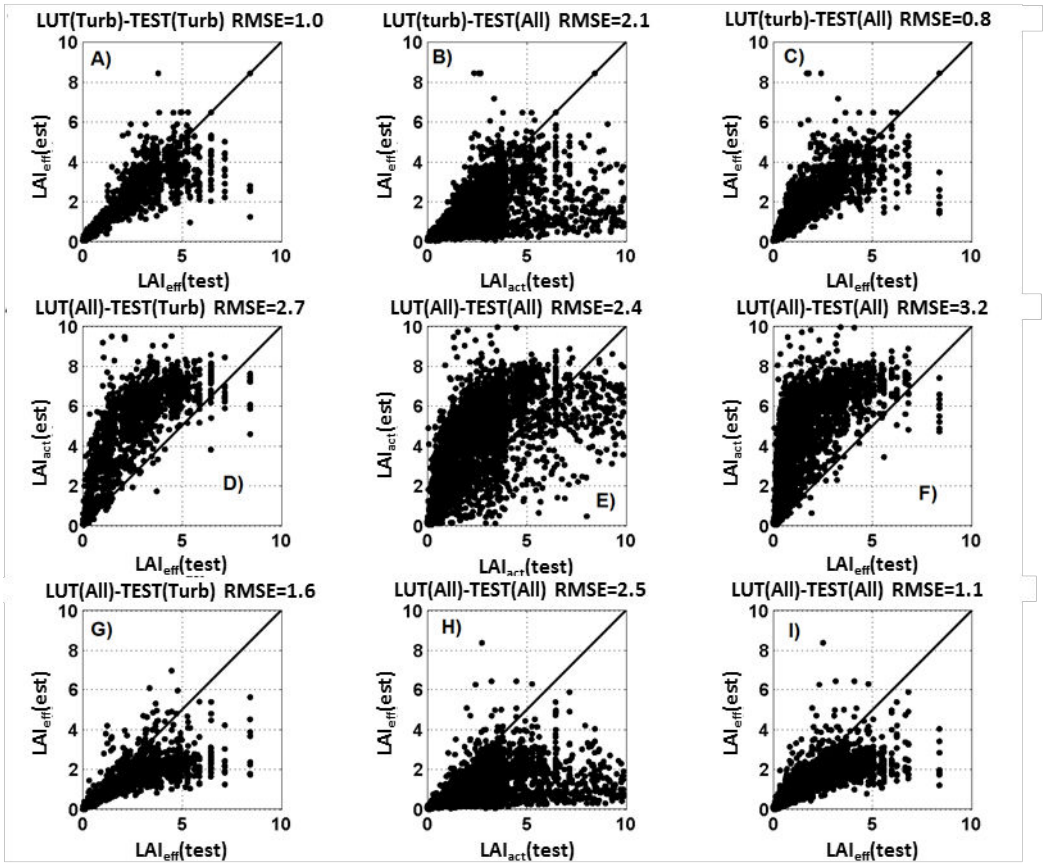

Figure 14. Estimated LAI for Actual (LAI_{act}(est)) and Effective LAI (LAI_{eff}(est)) as a function of the test values for Actual (LAI_{act}(test)) and Effective LAI (LAI_{eff}(test)). The several conbinations of LUT (Turbid or All) and Test (Turbid or All) presented in Table 3 are displayed here. Single view zenith of 0° and CI = 1.

The results of the estimation using the LUT approach is shown in Figure 14 for the nadir view configuration. The following comments may be done:

- The estimation of actual LAI based on LUT based on actual LAI including all the cases (panel E), a large overestimation if observed for most of the cases with LAI_{act}(test)<6. This situation corresponds to a generic algorithm applied to cases representing the whole possible variability of canopy structure. For each test case, a wide variability in the solutions may be encountered, depending on the other structural variables that drive the spatial distribution of leaves within the canopy volulme.
- Conversely, when a specific algorithm (LUT(Turb)) is applied to the same specific situations (Test(Turb)) as in panel A, the retrieval performances appear much better and similar to the theoretical performances of current algorithms (Baret et al. 2010). However, this may be due either to the restricted number of cases and/or to the fact that the effective LAI, LAI_{eff}, was targeted here.

- Panel I) shows that the effective LAI is better estimated than the actual one providing that the LUT is made of effective LAI computed for all the cases.
- However, when the LUT is restricted to the turbid medium cases, panel C) shows that the best estimation performances are achieved. This situation is close to the one targeted by the CYCLOPES algorithm (Baret et al. 2007) where a neural network is trained over a turbid medium learning data set and applied over all situations globally. As a result, effective LAI values are estimated.
- Panel G) shows that estimates of effective LAI based on a LUT made of effective LAI computed over all the cases is performing poorer as when using a LUT made of turbid medium cases only (Panel C). This may be explained by a larger variability in the solutions as compared to a LUT restricted to turbid medium cases.
- Nevertheless, panels D) and F) clearly demonstrate that a LUT based on effective LAI is not able to estimate the actual LAI as expected.
- Reciprocally, panels B) and H) shows that effective LAI values are (by construction of our data set) poor estimates of the actual LAI, with a clear underestimation of the actual LAI values depending on the leaf clumping in agreement with what was observed in Figure 11.

The results observed for the other view angles (30° or 57°) are very similar to what is observed at 0° zenith view angle as demonstrated in Figure 15 and Figure 16. However, a slight degradation of the estimation performances is observed for 57° view zenith angle, probably because of the saturation effect that may be observed for the larger LAI values.

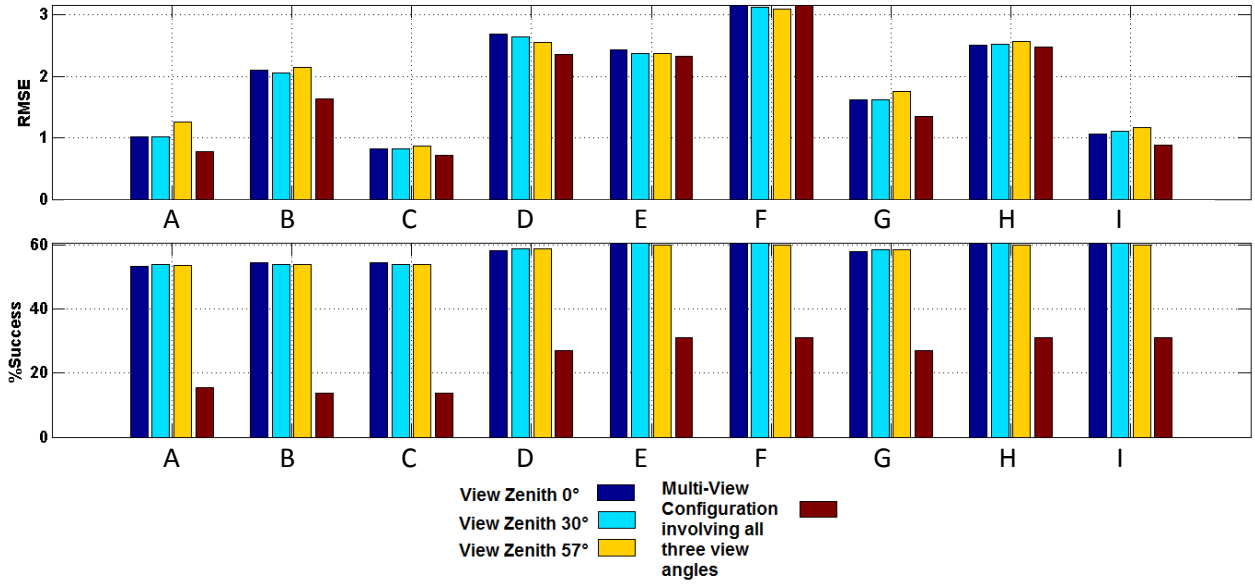

Figure 15. Synthesis of the retrieval performances and %success for CI = 1. The letters correspond to the panels in Figure 14.

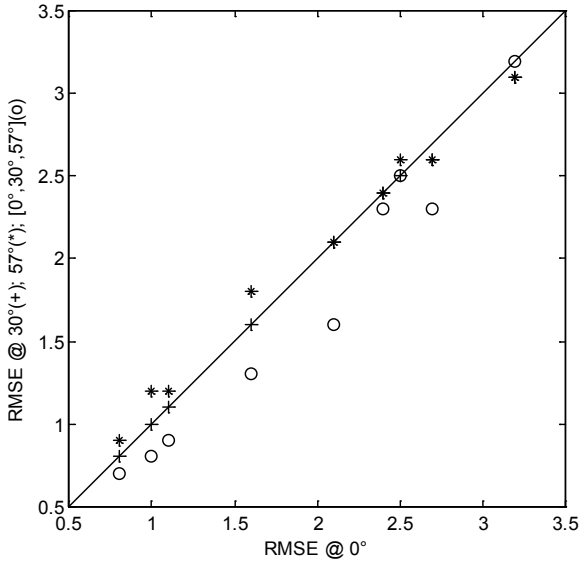

Figure 16. Comparison between the RMSE associated to LAI estimates for the 9 combinations of LUT and test considered (Table 3) for 30° (+), 57° (*) and multiangle [0°, 30°, 57°] as a function of the RMSE obtained for 0° view zenith angle.

When the 3 view directions are used concurrently in the LUT to estimate LAI, the same trends as using a single view direction are observed. However, the RMSE decreases significantly for most situations, and particularly those for which LAI_{eff} is targeted (Figure 15 and Figure 16). The LUT derived from the turbid medium cases and applied to the effective LAI from all the cases get the best performances (Figure 17 Panel C). The improved accuracy induced by the combination of view directions is explained by:

- The reduction of uncertainties associated to the radiometric mesurements because of the multiplication of configurations, or/and
- The use of additional information contained in the BRDF as already noticed by (Barnsley et al., 1997).

However, the %Success when using the three view directions concurrently is dramatically divided by a factor 2 or 3, (Figure 15) and reaching a minimum close to 17% when the turbid medium cases are considered since the reflectance hyperspace generated by the 9 SENTINEL 2 bands is sparsely populated. Alternative techniques should be thus designed to increase the local density of points around the test case, possibly using the interpolation techniques described earlier in section 2.2.2.

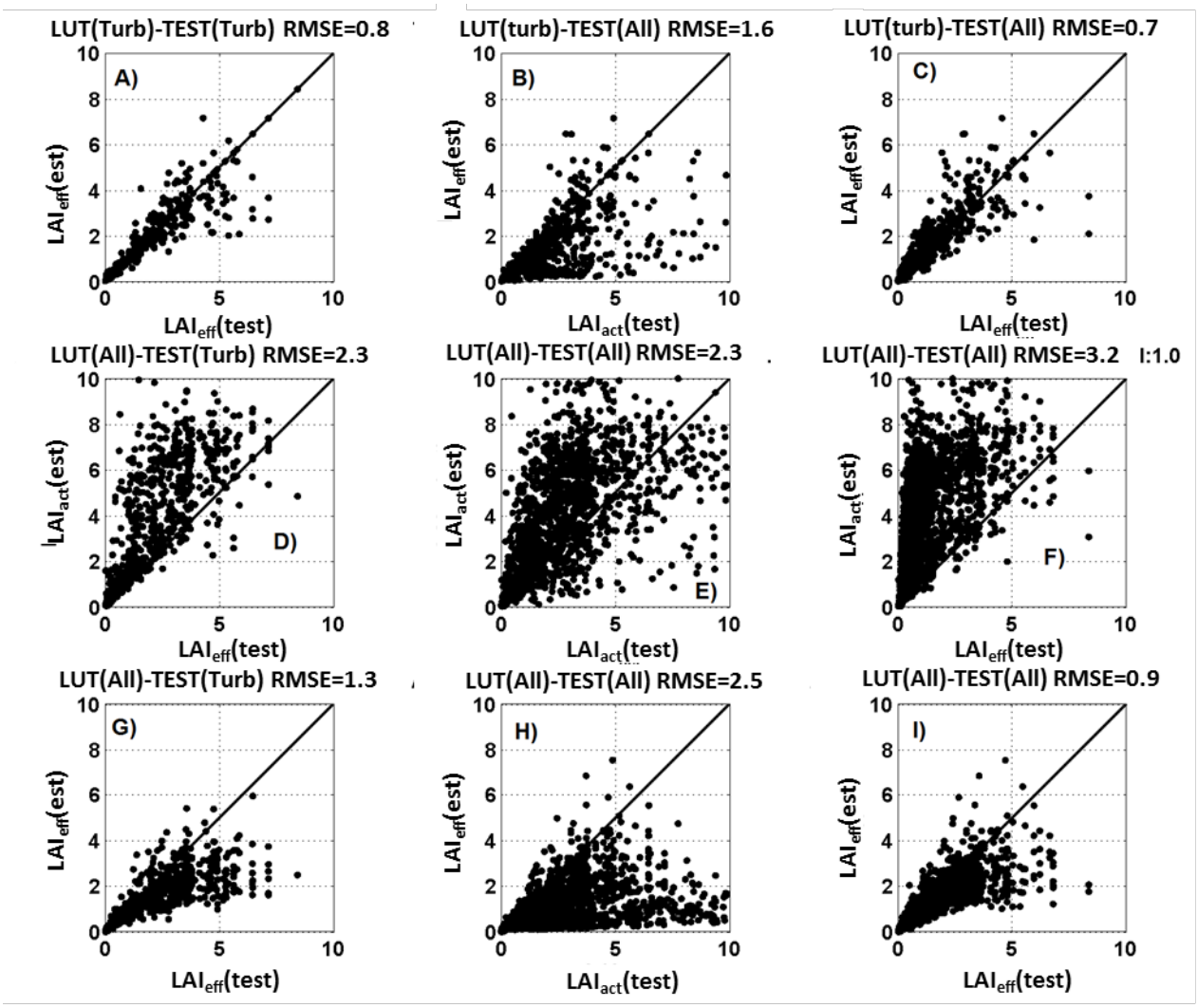

Figure 17. Estimated LAI (LAI_{app}) for Actual (LAI_{app}(act)) and Effective LAI (LAI_{app}(eff)) under the different canopy assumptions , Turbid (Tur) and both Turbid and non-Turbid (ALL), for a multi-view configuration and CI = 1

4 Conclusion

This study is based on a large set of simulations of 3D radiative transfer model which require intensive computation. It was achieved using an innovative approach to allow accurate interpolation in the spectral domain to get the canopy reflectance values for given leaf optical characteristics and soil reflectance in any spectral band from precomputed canopy reflectance values in a limited set of combinations of leaf reflectance and transmittance and soil reflectance. It is inspired from the spectral invariant theory, although simplified but probably more versatile than the usual theory. It may be ultilmately applied to 3D simulations of canopies represented by very detailed and realistic structure.

The approach used in this study to invert the radiative transfer model exploits Look Up Tables which are precomputed. Some refinements have been proposed to define the solution as a function of the confidence interval associated to the radiometric measurements. However, the procedure appears restrictive in a significant number of situations where no solution was found in the vicinity of the measurement because of the sparseness of the space of canopy realization. Efforts should thus be directed to improve both the criterion used for defining the solution as well as increasing the density of the space of canopy reéalisation which is possible using the previously developed interpolation techniques.

The results show that a LUT specialized over the cases targeted performs better than a more generic LUT, which was however expected, confirming the ill-posed nature of the inverse problem in remote sensing. However, the main output of this study is that reflectance measurements are more directly related to the effective LAI than to the actual LAI. The effective LAI was defined here as the LAI that provides the same directional green fraction as that of the actual canopy targeted. However, this study considers generic cases, without using any prior information on the target. This result may probably be different for specific situations where the space of canopy realization is concentrated in a restricted convex hull and where canopy variables have well known ranges of variability and where covariance between the variables are known as well. It should be noted that the variability on canopy structure considered in this study is restricted to the clumping at the stand scale, i.e. leaves are grouped within the tree crowns and the trees are randomly distributed in the scene. Therefore, clumping at the shoot or landscape scales have not been considered. However, the results obtained in this study should still be on value, since the shoot scale clumping may be treated through effective leaf optical properties (shoot scattering properties may be replaced by the equivalent lef reflectance and transmittance). Nevertheless the landscape clumping may still pose a problem for kilometric resolution observations while it could be addressed by extending our simulations to this additional factor of variability.

The results obtained in this study show also that similar performances were obtained for the 3 view directions considered (0°, 30° and 57° zenith angle) with a slight degradation of performances for the more oblique view angle because of the early saturation expected. It is possible that the information conveyed by multiple scattering effect plays a critical role in the relative stability of retrieval performances. The use of multidirectional observations improve significantly but only slightly the retrieval perfomances. This should be further investigated to better understand wether the improvement is mainly coming from the reduction of the effect of radiometric uncertainties by the multiplication of samples (observation directions) or if it is due to an increase of the radiometric information.

This study is based on radiative transfer silmulations and therefore includes no sources of uncertainties coming from the assumptions on canopy structure. The study should therefore be extended to actual observations where both actual and effective LAI values may be derived from ground measurements such as hemispherical photographs. This may be more easily achieved in the next future when actual Sentinel 2 observations will be available.

Acknowledgment: This research has received funding from the Geoland2 European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n°218795. A. Verger was funded by the VALi+d postdoctoral program (FUSAT, GV-20100270).

References

Anderson, M. C., Norman, J. M., Kustas, W. P., Houborg, R., Starks, P. J., and Agam, N.: A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales, Remote Sensing of Environment, 112, 4227-4241, 2008.

Bacour, C.: Contribtion a la determination des parametres biophysiques des couverts vegetaux par inversion de modeles de reflectance: analyse de sensibilitie et configurations optimales, PhD, Universite Paris 7 - Denis Diderot,, Paris (France), 2001.

Baret, F., and Guyot, G.: Potentials and limits of vegetation indices for LAI and APAR assessment, Remote Sensing of Environment, 35, 161-173, 1991.

Baret, F., Jacquemoud, S., and Hanocq, J. F.: About the soil line concept in remote sensing, Advances in Space Research, 13, 281-284, http://dx.doi.org/10.1016/0273-1177(93)90560-X, 1993a.

Baret, F., Jacquemoud, S., and Hanocq, J. F.: The soil line concept in remote sensing, Remote Sensing Reviews, 7, 65-82, 10.1080/02757259309532166, 1993b.

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

Baret, F., and Buis, S.: Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems, in: Advances in Land Remote Sensing, edited by: Liang, S., Springer Netherlands, 173-201, 2008.

Barnsley, M. J., Allison, D., and Lewis, P.: On the information content of multiple view angle (MVA) images, International Journal of Remote Sensing, 18, 1937-1960, 1997.

Barr, A. G., Black, T. A., Hogg, E. H., Kljun, N., Morgenstern, K., and Nesic, Z.: Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, Agricultural and Forest Meteorology, 126, 237-255, 2004.

Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C.: A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data, Remote Sensing of Environment, 114, 1312-1323, 2010.

Boulet, G., Chehbouni, A., Braud, I., Vauclin, M., Haverkamp, R., and Zammit, C.: A simple water and energy balance model designed for regionalization and remote sensing data utilization, Agricultural and Forest Meteorology, 105, 117-132, 2000.

Buermann, W., Dong, J., Zeng, X., Myneni, R. B., and Dickinson, R. E.: Evaluation of the utility of satellite based vegetation leaf area index data for climate simulations, Journal of Climate, 14, 2001.

Campbell, G. S.: Derivation of an angle density function for canopies with ellipsoidal leaf angle distributions, Agricultural and Forest Meteorology, 49, 173-176, http://dx.doi.org/10.1016/0168-1923(90)90030-A, 1990.

Chen, J. M., and Black, T. A.: Foliage area and architecture of plant canopies from sunfleck size distributions, Agricultural and Forest Meteorology, 60, 249-266, http://dx.doi.org/10.1016/0168-1923(92)90040-B, 1992a.

Chen, J. M., and Black, T. A.: Defining leaf area index for non-flat leaves, Plant, Cell & Environment, 15, 421-429, 10.1111/j.1365-3040.1992.tb00992.x, 1992b.

Chen, J. M., Liu, J., Leblanc, S. G., Lacaze, R., and Roujean, J.-L.: Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption, Remote Sensing of Environment, 84, 516-525, 2003.
Chen, J. M., Chen, X., Ju, W., and Geng, X.: Distributed hydrological model for mapping evapotranspiration using remote sensing inputs, Journal of Hydrology, 305, 15-39, 2005a.

Chen, J. M., Menges, C. H., and Leblanc, S. G.: Global mapping of foliage clumping index using multi-angular satellite data, Remote Sensing of Environment, 97, 447-457, 2005b.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., and Wang, L.: Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem, Remote Sensing of Environment, 84, 1-15, 2003.

Consoli, S., D'Urso, G., and Toscano, A.: Remote sensing to estimate ET-fluxes and the performance of an irrigation district in southern Italy, Agricultural Water Management, 81, 295-314, 2006.

De Pury, D. G. G., and Farquhar, G. D.: Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models, Plant, Cell & Environment, 20, 537-557, 10.1111/j.1365-3040.1997.00094.x, 1997.

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

Dong, H., Wenze, Y., Tan, B., Rautiainen, M., Ping, Z., Jiannan, H., Shabanov, N. V., Linder, S., Knyazikhin, Y., and Myneni, R. B.: The importance of measurement errors for deriving accurate reference leaf area index maps for validation of moderate-resolution satellite LAI products, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1866-1871, 10.1109/tgrs.2006.876025, 2006.

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

Feng, D., Chen, J. M., Plummer, S., Mingzhen, C., and Pisek, J.: Algorithm for global leaf area index retrieval using satellite imagery, Geoscience and Remote Sensing, IEEE Transactions on, 44, 2219-2229, 10.1109/tgrs.2006.872100, 2006.

Fernandes, R., and G. Leblanc, S.: 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, 303-316, 2005.

Ganguly, S., Schull, M. A., Samanta, A., Shabanov, N. V., Milesi, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Generating vegetation leaf area index earth system data record from multiple sensors. Part 1: Theory, Remote Sensing of Environment, 112, 4333-4343, 2008.

Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC: http://www.wmo.int/pages/prog/gcos/documents/GCOSIP-10_DRAFTv1.0_131109.pdf, 2010.

Huang, D., Knyazikhin, Y., Dickinson, R. E., Rautiainen, M., Stenberg, P., Disney, M., Lewis, P., Cescatti, A., Tian, Y., Verhoef, W., Martonchik, J. V., and Myneni, R. B.: Canopy spectral invariants for remote sensing and model applications, Remote Sensing of Environment, 106, 106-122, 2007.

Huemmrich, K. F.: The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance, Remote Sensing of Environment, 75, 423-431, 2001.

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

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

Kandasamy, S., Lopez-Lozano, R., Baret, F., and Rochdi, N.: The effective nature of LAI as measured from remote sensing observations, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, 789-792,

Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., and Lee, R.-Y.: Image masking for crop yield forecasting using AVHRR NDVI time series imagery, Remote Sensing of Environment, 99, 341-356, 2005.

Knyazikhin, Y., Schull, M. A., Xu, L., Myneni, R. B., and Samanta, A.: Canopy spectral invariants. Part 1: A new concept in remote sensing of vegetation, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 727-735, 2011.

Kobayashi, H., Suzuki, R., and Kobayashi, S.: Reflectance seasonality and its relation to the canopy leaf area index in an eastern Siberian larch forest: Multi-satellite data and radiative transfer analyses, Remote Sensing of Environment, 106, 238-252, 2007.

Lewis, P., and Disney, M.: Spectral invariants and scattering across multiple scales from within-leaf to canopy, Remote Sensing of Environment, 109, 196-206, http://dx.doi.org/10.1016/j.rse.2006.12.015, 2007.

Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., and Smith, B.: Climatic Control of the High-Latitude Vegetation Greening Trend and Pinatubo Effect, Science, 296, 1687-1689, 10.1126/science.1071828, 2002.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G., and Cook, R.: Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1804-1817, 10.1109/tgrs.2006.872529, 2006.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sensing of Environment, 83, 214-231, 2002.

North, P. R. J.: Three dimensional forest light interaction model using a monte carlo method, IEEE Transactions on Geoscience and Remote Sensing, 34, 956 1996.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C.: Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends in Ecology & Evolution, 20, 503-510, 10.1016/j.tree.2005.05.011, 2005.

Price, J. C.: Estimating leaf area index from satellite data, Geoscience and Remote Sensing, IEEE Transactions on, 31, 727-734, 10.1109/36.225538, 1993.

Qin, Z., Su, G.-l., Zhang, J.-e., Ouyang, Y., Yu, Q., and Li, J.: Identification of important factors for water vapor flux and CO2 exchange in a cropland, Ecological Modelling, 221, 575-581, 10.1016/j.ecolmodel.2009.11.007, 2010.

Running, S. W., and Coughlan, J. C.: A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes, Ecological Modelling, 42, 125-154, 10.1016/0304-3800(88)90112-3, 1988.

Schubert, P., Eklundh, L., Lund, M., and Nilsson, M.: Estimating northern peatland CO2 exchange from MODIS time series data, Remote Sensing of Environment, 114, 1178-1189, 2010.

Schull, M. A., Knyazikhin, Y., Xu, L., Samanta, A., Carmona, P. L., Lepine, L., Jenkins, J. P., Ganguly, S., and Myneni, R. B.: Canopy spectral invariants, Part 2: Application to classification of forest types from hyperspectral data, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 736-750, 2011.

Steduto, P., and Hsiao, T. C.: Maize canopies under two soil water regimes: II. Seasonal trends of evapotranspiration, carbon dioxide assimilation and canopy conductance, and as related to leaf area index, Agricultural and Forest Meteorology, 89, 185-200, 10.1016/s0168-1923(97)00084-1, 1998.

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

Verger, A., Baret, F., and Weiss, M.: A multisensor fusion approach to improve LAI time series, Remote Sensing of Environment, 115, 2460-2470, 2011.

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

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, Remote Sensing of Environment, 115, 801-823, 2011.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., and Bing, Z.: Relating soil surface moisture to reflectance, Remote Sensing of Environment, 81, 246, 2002.

Weiss, M., Baret, F., Myneni, R., Pragnère, A., and Knyazikhin, Y.: Investigation of a model inversion technique for the estimation of crop charcteristics from spectral and directional reflectance data, Agronomie, 20, 3-22 2000.
Wiegand, C. L., Gerbermann, A. H., Gallo, K. P., Blad, B. L., and Dusek, D.: Multisite analyses of spectral-biophysical data for corn, Remote Sensing of Environment, 33, 1-16, 1990.

Wylie, B. K., Fosnight, E. A., Gilmanov, T. G., Frank, A. B., Morgan, J. A., Haferkamp, M. R., and Meyers, T. P.: Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains, Remote Sensing of Environment, 106, 399-413, 2007.

Yan, H., Wang, S. Q., Billesbach, D., Oechel, W., Zhang, J. H., Meyers, T., Martin, T. A., Matamala, R., Baldocchi, D., Bohrer, G., Dragoni, D., and Scott, R.: Global estimation of evapotranspiration using a leaf area index-based surface energy and water balance model, Remote Sensing of Environment, 124, 581-595, 10.1016/j.rse.2012.06.004, 2012.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F., and Moore Iii, B.: 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, 357-371, 2005.

3 Gap-filling and smoothing of MODIS LAI time Series data

These LAI retrievals are subject to the uncertainties and assumptions in the retrieval algorithms and the radiative transfer models (RTMs) (Chapter 2). These uncertainties along with the residual cloud contamination, imperfect atmospheric or directional corrections introduces significant noise in the LAI time-series estimates. Consequently, the current products are characterized by significant temporal noise, which is not a characteristic of the actual LAI values. Changes in the LAI estimates result as a consequence of the incremental processes in vegetation such as leaf development and senescence. Hence, these changes in time are expected to be relatively smooth. However, such smooth temporal evolution is not always observed in the current LAI time series products(Verger et al., 2008;Weiss et al., 2007a). The LAI time series products are also associated with significant amount of discontinuities due to cloud cover (Weiss et al., 2007a) as well as system failures. The MODIS C5 LAI products, data used in this study, has been recognized as one of the time series estimates characterized by significant noise and gaps (discontinuities), due to a combination of various factors such as short compositing window (8 days), the maximum value compositing algorithm used, instabilities of the retrieval algorithms, etc. However, the MODIS C5 products are an improvement over the earlier C4 MODIS LAI products(De Kauwe et al., 2011). Due to the significance of LAI and its time series estimates to understand vegetation dynamics improving the quality of these time series estimates has been the focus of many investigations (Yuan et al., 2011;Verger et al., 2011b;Jiang et al., 2010b;Borak and Jasinski, 2009;Feng et al., 2008).

Time-series processing, therefore, forms the first stage in many studies on environment and vegetation. The earliest methods used in remote sensing to fill gaps and smooth time-series estimates, often called compositing, operate over a local temporal window focusing on minimizing artifacts due to cloud or snow contamination, and atmospheric or directional corrections. Some of these compositing algorithms were reviewed by (Ql and Kerr, 1997), including the well-known Maximum Value Compositing (MVC) (Holben, 1986) and the Best Index Slope Extraction (BISE) (Viovy et al., 1992). More recently, (Hird and McDermid, 2009) made an empirical comparison of six time series processing methods, focusing on NDVI (Normalized Difference Vegetation Index) (Rouse et al., 1974), and demonstrated that the logistic functions(Beck et al., 2006) and Asymmetric Gaussian functions (Jönsson and Eklundh, 2004) to outperform other simple local filtering methods. (Jiang et al., 2010b), compared three statistical methods namely Seasonal Auto Regressive Integrated Moving Average (SARIMA) (Box et al., 2008a), Dynamic Harmonic Regression (DHR) (Young et al., 1999) and Seasonal Trend Decomposition (STL) (Cleveland et al., 1990), focusing on MODIS LAI time series estimates, to smooth and provide forecast over a season. Fourier transforms (Azzali and

Menenti, 1999) and wavelet decomposition (Martinez and Gilabert, 2009) have also been used to process the time series estimates. These studies have demonstrated the superiority of the temporally local methods over global methods (applied over the entire time series) (Hird and McDermid, 2009;Beck et al., 2006;Jonsson and Eklundh, 2002). Physically based corrects were also proposed in order to correct the known factors of variability resulting in the GIMMS data set (Tucker et al., 2005). However, EMD had been successfully employed for correcting orbital drifts and directional effects (Pinzón et al., 2005). Further, the choice of the time-series processing methods for filling gaps and smoothing time series may have a large impact on the accuracy of the phenology extracted from these series estimates (Atkinson et al., 2012;Hird and McDermid, 2009).

This brief review shows that wide-range of methods are available for processing the time-series estimates of biophysical variables from medium resolution satellite sensors. However, identifying the potentials and limitations of these methods is difficult in the absence of comprehensive evaluation of the methods. Comparative evaluations of the time-series processing methods that are available are often qualitative and when quantitative, are centered on a small sample of global conditions. Further, most of these comparisons were made on NDVI. Previous comparative evaluations also paid very little attention to structure of the gaps in the time series estimates. Finally, only a small fraction of the studies on time series processing employed methods that operates on the entire time series (temporally global methods) such as decomposition based methods.

This chapter evaluates the capacity of several methods (8 methods) to provide faithful reconstruction of time series in the presence of significant amount of discontinuities and noise. The chapter evaluates the methods based on their ability to process time series with variable lengths of missing estimates, fidelity of reconstruction and the smoothness of the reconstructed temporal profiles. The ability of the methods to capture phenology stages is also quantified in this study. The methods selected for the evaluation were well referenced and are based on local curve fitting and decomposition techniques. This evaluation is based on the MODIS C5 LAI products (Shabanov et al., 2005) at 1Km spatial and 8-day temporal resolution over a 9 year period.

Biogeosciences Discuss., 9, 17053–17097, 2012
www.biogeosciences-discuss.net/9/17053/2012/
doi:10.5194/bgd-9-17053-2012

Biogeosciences Discussions

A comparison of methods for smoothing and gap filling time series of remote sensing observations: application to MODIS LAI products

S. Kandasamy, F. Baret, A. Verger, P. Neveux, and M. Weiss

EMMAH-UMR 1114 – INRA UAPV, Domaine Saint-Paul, Site Agroparc,
84914 Avignon, France

Received: 20 October 2012 – Accepted: 11 November 2012 – Published: 4 December 2012

Correspondence to: F. Baret (baret@avignon.inra.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Moderate resolution satellite sensors including MODIS already provide more than 10 yr of observations well suited to describe and understand the dynamics of the Earth surface. However, these time series are incomplete because of cloud cover and associated with significant uncertainties. This study compares eight methods designed to improve the continuity by filling gaps and the consistency by smoothing the time course. It includes methods exploiting the time series as a whole (Iterative caterpillar singular spectrum analysis (ICSSA), empirical mode decomposition (EMD), low pass filtering (LPF) and Whittaker smoother (Whit)) as well as methods working on limited temporal windows of few weeks to few months (Adaptive Savitzky-Golay filter (SGF), temporal smoothing and gap filling (TSGF) and asymmetric Gaussian function (AGF)) in addition to the simple climatological LAI yearly profile (Clim). Methods were applied to MODIS leaf area index product for the period 2000–2008 over 25 sites showing a large range of seasonal patterns. Performances were discussed with emphasis on the balance achieved by each method between accuracy and roughness depending on the fraction of missing observations and the length of the gaps. Results demonstrate that EMD, LPF and AGF methods were failing in case of significant fraction of gaps (% Gap > 20 %), while ICSSA, Whit and SGF were always providing estimates for dates with missing data. TSGF (respectively Clim) was able to fill more than 50 % of the gaps for sites with more than 60 % (resp. 80 %) fraction of gaps. However, investigation of the accuracy of the reconstructed values shows that it degrades rapidly for sites with more than 20 % missing data, particularly for ICSSA, Whit and SGF. In these conditions, TSGF provides the best performances significantly better than the simple Clim for gaps shorter than about 100 days. The roughness of the reconstructed temporal profiles shows large differences between the several methods, with a decrease of the roughness with the fraction of missing data, except for ICSSA. TSGF provides the smoothest temporal profiles for sites with % Gap > 30 %. Conversely, ICSSA, LPF, Whit, AGF and Clim provide smoother profiles than TSGF for sites with % Gap < 30 %. Impact of the accuracy and

smoothness of the reconstructed time series were evaluated on the timing of phenological stages. The dates of start, maximum and end of the season are estimated with an accuracy of about 10 days for the sites with % Gap < 10 % and increases rapidly with % Gap. TSGF provides the more accurate estimates of phenological timing up to % Gap < 60 %.

1 Introduction

Leaf area index, LAI, is recognized as an essential climate variable (GCOS, 2006) since it plays a key role in vegetation functioning and exchanges of mass and energy between the atmosphere, the plant and the soil. LAI is defined as half the area of the green elements per unit horizontal surface (Chen and Black, 1992). Satellite observations in the reflective solar domain have been used intensively for more than a decade to monitor LAI dynamics over the globe using medium resolution sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer) (Myneni et al., 2002), VEGETATION (Deng et al., 2006; Baret et al., 2007), MERIS (MEdium Resolution Imaging Spectrometer) (Bacour et al., 2006a) or AVHRR (Advanced Very High Resolution Radiometer) (Ganguly et al., 2008). LAI was thus used in numerous investigations including climate change (Pettorelli et al., 2005; Kobayashi et al., 2007), global carbon fluxes (Wylie et al., 2007; Schubert et al., 2010) land cover (Jakubauskas et al., 2002; Boles et al., 2004; Thenkabail et al., 2005; Heiskanen and Kivinen, 2008) and land cover change (Hansen et al., 2002, 2008; Coops et al., 2009; Pouliot et al., 2009) or crop production forecasting (Kastens et al., 2005; Dente et al., 2008; Becker-Reshef et al., 2010). For all these applications, the availability of long and continuous LAI time series is essential as outlined in (GCOS, 2010).

The current products show significant discontinuities mainly due to cloud or snow cover (Weiss et al., 2007) as well as system failure. Further, because of residual cloud contamination, imperfect atmospheric or directional corrections as well as instability of the inversion process, products may be characterized by significant temporal noise not expected for the actual LAI values. LAI is the result of incremental processes in vegetation such as leaf development and senescence. Therefore, LAI should be relatively smooth with time. However, this temporal smoothness property is not always observed over current LAI products as demonstrated by several authors (Weiss et al., 2007; Verger et al., 2008; Camacho et al., 2012). The MODIS LAI product was recognized as one of the less smooth mainly because of a combination of factors including a short (8 days) compositing window, the maximum value compositing algorithm used and instability in the retrieval algorithm, though MODIS collection 5 products are an improvement over the collection 4 products (De Kauwe et al., 2011). Several investigations have been focusing on the improvement of the MODIS products using specific filtering techniques to get smoother temporal series (Gao et al., 2007; Borak and Jasinski, 2009; Jiang et al., 2010; Verger et al., 2011; Yuan et al., 2011).

Time Series processing is thus an important ingredient of a biophysical algorithm in order to get the expected continuous and smooth dynamics required by many applications. The earliest methods used in remote sensing, often called compositing were reviewed by Qi and Kerr (1974). They mostly operate over a local temporal window, focusing on minimizing artifacts due to cloud or snow contamination, atmospheric or directional residual effects. They included the well-known MVC (Maximum Value Compositing) (Holben, 1986) and BISE (Best Index Slope Extraction from) (Viovy et al., 1992). More recently, (Hird and McDermid, 2009) reviewed the abundant literature on time series processing in remote sensing, mostly focusing on NDVI (Normalized Difference Vegetation Index) (Rouse et al., 1974). They further compared six methods operating over a restricted temporal window and demonstrated that under their conditions, logistic or (Beck et al., 2006) Asymmetric Gaussian Function (Jönsson and Eklundh, 2004) curve fitting methods were outperforming more simple local filtering methods. Jiang et al. (2010) compared three statistical methods both to smooth the time series and to provide forecast over a season. These methods are based on the decomposition of the time series into noise, seasonal variability and trend and require therefore relatively long time series of observations. Fourier transforms (Azzali and

Menenti, 1999), or wavelet decomposition (Martinez and Gilabert, 2009) have also been used to characterize the phenology of vegetation from medium resolution observations. However, several studies have demonstrated the superiority of local methods, i.e. based on a restricted temporal window, as compared to Fourier transform methods applied to the whole time series (Jönsson and Eklundh, 2002; Beck et al., 2006; Ma and Veroustraete, 2006; Hird and McDermid, 2009). Physically based corrections were also proposed in order to correct for the known factors of variability, resulting in the GIMMS data set (Tucker et al., 2005). However orbital drift and directionality were rather corrected using Empirical Mode Decomposition techniques (Pinzon et al., 2005). More recently, the Long Term Data Record (LTDR) series derived from AVHRR sensors (Vermote et al., 2009) and CYCLOPES (Baret et al., 2007) and GEOV1 (Meroni et al., 2012) derived from VEGETATION proposed also global time series based on physical principles.Alcaraz-Segura et al. (2010) and Meroni et al. (2012) showed that significant differences were observed between these several NDVI time series, making the identification of anomalies and trends more complex. The choice of the smoothing gap filling or compositing method may have a large impact on the accuracy of the phenology extracted from the reconstructed time series (Hird and McDermid, 2009; Atkinson et al., 2012).

This brief review of studies focusing on satellite time series from medium resolution sensors shows that a number of methods are available. It is however still difficult to identify the potentials and limitations associated since no comprehensive evaluation is available. Comparison is often qualitative, or when quantitative, it is mostly centered on a small sample of global conditions. Most of them have been applied to NDVI rather than on true biophysical variable such as LAI. Further, very little attention was paid to the missing data structure: as a matter of fact satellite observations present missing data mostly because of cloud masking which creates irregular time steps between actual observations of the surface. Gap filling is therefore an important aspect of the processing. Finally, only a small fraction of the studies were employing methods capable of processing the time series as a whole such as the decomposition methods.

The objective of this study is to evaluate the capacity of several methods to provide faithful reconstruction of time series in presence of significant amount of missing observations as well as observations contaminated by uncertainties. The methods will therefore be compared using several criterions including the ability to run over periods without observations of variable length, the fidelity of reconstructed values with the actual ones and the smoothness of the reconstructed temporal profiles. In addition, consequences on the ability to capture phenology stages will be also quantified since this is a usual application of the time series

Eight methods were selected because they were well referenced while being based either on local curve fitting techniques, or decomposition techniques working on the time series as a whole. Of the eight methods, except ICSSA and EMD, all the other methods are familiar methods for processing biophysical time series data. ICSSA and EMD, commonly used in other subject areas, were considered in this study due to their potential ability to recover seasonal trends and reduce noise. The study is based on MODIS LAI collection 5 product (Shabanov et al., 2005) at 1km spatial and 8 days temporal sampling over a 9 yr period. The MODIS LAI products were demonstrated to get relatively good accuracy (closeness to actual ground observations) but were suffering from lack of precision (temporal or spatial consistency), resulting into shaky temporal profiles as stated earlier. Processing of such time series is therefore expected to result in significant improvement of its consistency. A sample of sites selected to represent the range of variability expected over the globe was considered.

2 Approach, data and methods

The MODIS LAI products will first be described. Then the 8 methods selected will be briefly presented. Finally, the approach for evaluating the methods and the associated metrics used will be described.

2.1 The MODIS data and preprocessing

The data used in this study are the MODIS Collection 5 LAI products (MOD15A2) derived from TERRA and AQUA platforms. The products were downloaded from the land processes distributed active archive center (https://lpdaac.usgs.gov/products/modis_products_table/mod15a2) for the 2000–2008 period. They correspond to 1 km spatial sampling interval using a sinusoidal projection system. The temporal sampling is 8 days based on a daily composition: all observations available in the 8-day compositing window are accumulated, and the one getting the maximum FAPAR value is selected. The main MODIS LAI retrieval algorithm relies on the inversion of a 3-D radiative transfer model using the red and near infrared bidirectional reflectance factor values, associated uncertainties, the view-illumination geometry and biome type (within eight types based on MOD12Q1 land cover map) as inputs (Myneni et al., 2002; Shabanov et al., 2005). If the main algorithm fails, a back-up procedure is triggered to estimate LAI from biome specific NDVI based relationships. However, the LAI estimates using the back-up algorithm are of lower quality mostly due to residual clouds and poor atmospheric correction (Yang et al., 2006a, b). Hence, these estimates are not used in this study and are considered as missing observations. Although MODIS LAI product has been extensively validated (e.g. De Kauwe et al., 2011; Ganguly et al., 2008), high level of noise was inducing shaky temporal profiles and unrealistic seasonality (Kobayashi et al., 2010), which justifies the interest of using MODIS LAI products for smoothing and gap-filling investigations.

A first preprocessing step was applied to remove unexpected abrupt variations in the time series: values that are substantially different from both their left- and right-hand neighbors and from the median in a 72 days length local window are considered as missing values as proposed (Jönsson and Eklundh, 2004) in the TIMESAT toolbox. Further, for Evergreen Broadleaf forests presenting reduced seasonality and high level of variability in the time series because of frequent occurrence of residual cloud contamination, any value lower than the first decile are eliminated since these usually low LAI values are not expected in this canopy type.

2.2 The methods investigated

Eight candidate methods (Table 1) were selected, including both decomposition techniques generally applied to the whole time series and curve fitting methods working on a limited temporal window. Decomposition methods split the signal into additive components. The time series are then reconstructed using only the components of interest, usually removing the high frequency components considered as noise. Decomposition methods should capture the seasonality and the trend signals observed over the whole time series, which may be exploited in the reconstruction phase to replace missing values. Curve fitting techniques adjust the parameters of a functional by minimizing a cost function that is usually the sum of quadratic differences between observations and simulations. Because the adjustment is operated over a limited temporal window, only a limited amount of information is used when filling gaps.

2.2.1 Iterative Caterpillar Singular Spectrum Analysis Method (ICSSA)

This is a modification of the CSSA (Golyandina and Osipov, 2007) method developed to describe time series and fill missing data by decomposing the time series into empirical orthogonal functions (EOF). This modified version was proposed by Kandasamy et al. (2012) to correct for overestimation of seasonal valleys and better fitting to the peaks as compared to the original CSSA formulation. The method requires 2 parameters: the window length and the number of eigenvectors (orthogonal functions) used for the reconstruction. Better reconstruction can be obtained for large number of eigenvectors but at the cost of a decrease of the smoothness. After trial and errors, the number of eigenvector was set to 1, and the window length was set to 40 days. This method allows filling gaps and forecasting data at the extremities of the time series.

2.2.2 Empirical Mode Decomposition method (EMD)

This method proposed by Huang et al. (1998) consists in decomposing the time series into a small number of Intrinsic Mode Frequencies (IMFs) derived directly from the time series itself using an adaptive iterative process where the data are represented by intrinsic mode functions, to which the Hilbert transform can be applied. The method requires setting 2 parameters: the threshold for convergence and the maximum number of IMFs. The threshold for convergence is set to 0.3 according to (Huang et al., 1998) and the maximum number of IMFs was restricted to 10 after trial and errors. The first IMF, mostly affected by noise, was smoothed using a uniform mean kernel to remove the high frequency fluctuations at the expense of a loss in the amplitude (Demir and Erturk, 2008). Note that the EMD method requires the time series to be continuous. To allow the application of EMD to MODIS time series, the missing data within 128 days were filled by linear interpolation as proposed by Verger et al. (2011). However, when the time series contains gaps longer than 128 days, the whole series was not reconstructed. As a matter of fact, linear interpolation provides generally poor performances in case of long periods without observations.

2.2.3 Low Pass Filtering (LPF)

This method originally proposed by Thoning et al. (1989) was adapted by Bacour et al. (2006b) for better retrieving the seasonality from AVHRR time series. A time dependent function with 10 terms (2 polynomial and 8 harmonic terms) is first adjusted to the data. Then, the residuals of this first fitting are filtered with a low pass filter using two cut-off frequencies defined to separate the intra-annual and inter-annual variations. The final reconstruction is obtained by summing the polynomial and harmonic terms with the filtered residuals. Although this method may be considered as well as based on curve fitting, it applies over the time series as a whole. This method requires the data to be continuous. Hence, similarly to the EMD method, the gaps within 128 days were filled by linear interpolation. For gaps longer than 128 days, LPF was considered unsuccessful and, in this case, it results in missing reconstructed values for the whole time series.

2.2.4 Whittaker Smoother (Whit)

This method proposed by Whittaker (1923) is based on the minimization of a cost function describing the balance between fidelity expressed as the quadratic difference between estimates and actual observations and roughness expressed as the quadratic difference between successive estimates. This balance is controlled by a smoothing parameter. The higher this value, the smoother is the result but at the expense of fidelity. Finding an appropriate value of the smoothing parameter is difficult, as it depends on the data. After trial and errors the smoothing parameter was set to 100. The smoothness is also controlled by the order of differentiation which is fixed to 3 as proposed by Eilers (2003) for time series data sampled at regular intervals but with missing observations.

2.2.5 Adaptive Savitzky Golay Filter (SGF)

This method proposed by Chen et al. (2004) iteratively applies the Savitzky-Golay filter (Savitzky and Golay, 1964) to match the upper envelope of the time series. This specific adaptation was designed to minimize the effects of cloud and snow contamination that generally decreases the estimates of vegetation indices such as NDVI as well as biophysical variables such as LAI. The original Savitzky-Golay filter consists in a local polynomial fitting with two parameters: the length of the temporal window used and the order of the polynomial. As proposed by Chen et al. (2004), the values of these parameters were optimized for each case to get the best match between observations and reconstructed values. However, the range of variation of the window length was restricted to 72–112 days, and that of the polynomial order to 2–4. Missing data in the time series were filled by linear interpolation independently of the size of the gaps as proposed in the original version.

2.2.6 Temporal Smoothing and Gap Filling (TSGF)

This method is another adaptation of the Savitzky-Golay filter where the polynomial degree was fixed to 2 but the temporal window may be asymmetric and variable in length. It was designed by Verger et al. (2011) to better handle time series with missing observations. The temporal window corresponding to a nominal date is adjusted to include at least 3 observations within a maximum 64 days period on each side of the nominal date. If less than 6 observations are available within the maximum ±64 days temporal window, the polynomial fitting is not applied. The gaps in the reconstructed data are filled by an iterative (2 iterations) linear interpolation within 128 days window. For periods with missing data longer than 128 days, the interpolation is not applied which results in missing data. The possible flattening of the reconstructed time series observed over peaks was further corrected by scaling the smoothed series to the observations in a local window (Verger et al., 2011).

2.2.7 Asymmetric Gaussian Fitting (AGF)

This method has been proposed by Jönsson and Eklundh (2002, 2004) within the TIMESAT toolbox. A Gaussian function is adjusted locally over the growing and senescing part of each season. The functions are finally merged to get a smooth transition from one season to another. This method can handle small gaps. The original TIMESAT implementation of this method included 3 conditions preventing the near constant and noisy data from being processed. Two of these conditions (minimum seasonality in the data and maximum fraction of missing data of 25 %) were not considered here to enlarge its domain of validity in case of missing data. This thus allows more rigorous comparison with the other methods. However, the last condition was kept: the method was not applied over seasons with gaps longer than 72 days and, in these cases, reconstructions for the entire time series are not provided.

2.2.8 Climatology (Clim)

The climatology describes the typical yearly time course. It was included within the set of methods investigated since it may provide smooth and complete time series with however no changes from year to year. The climatology was computed every 8 days during a year by averaging the values available over a ±12 days window across all the years of the time series. The climatology was then corrected to provide more continuous and consistent temporal patterns as proposed in Baret et al. (2011). To provide smoother values, a simple Savitzky-Golay filter was applied (Savitzky and Golay, 1964). Note that the climatology may present missing values when no observations in the 24 days temporal window centered on a given date in the year were available across the whole years of the time series. In such situations, linear interpolation was used to fill gaps shorter than 128 days. Gaps longer than 128 days will result in missing data. Once the average yearly time course was computed, it was replicated across all the years considered to provide a reconstructed time series.

2.3 Evaluation approach

The approach proposed to evaluate the methods is based on two steps as sketched in Fig. 1. The first one is dedicated to the preparation of reference time series over a limited number of representative sites. As a result of this step, two time series will be output: (i) a time series with no gaps and small uncertainties considered as the reference (LAI_{ref}), and (ii) a time series with no gaps with realistic uncertainties (LAI_{comp}). In the second step, time series with variable occurrence of missing data will be simulated (LAI_{sim}) from the previously LAI_{comp} time series. Each method will be applied on this LAI_{sim} data to get the corresponding reconstructed time series (LAI_{rec}). Finally, the LAI_{rec} obtained with each of the 8 methods will be evaluated based on a range of metrics describing the fidelity, the roughness of time series and the accuracy of phenological stages that can be derived from the time series.

2.3.1 Generation of the reference and completed LAI time series

In the first step, few sites were selected with the objective to show a wide variability of seasonal patterns while having a minimal number of missing data. For this purpose, the 420 BELMANIP2 sites identified by Baret et al. (2006) to represent the variability of vegetation types and conditions around the world were considered. These 420 sites were first classified according to GLOBCOVER land cover map (Defourny et al., 2009) with the original classes aggregated into 5 main classes: Shrub/Savannah/bare area (SB), Grasslands and Crops (GC), Deciduous Broad Leaf Forests (DB), Evergreen Broad Leaf Forests (EB) and Needle Leaf Forests (NF). For EB and NF sites, most sites show significant fraction of missing data. The 5 sites showing the minimal gap fraction with a large variability in seasonal patterns were finally selected. The same process was applied to SB, GC and EB classes, resulting in a total of 25 sites (5 sites for each of the 5 biome classes) (Table 2)

The 8 methods presented earlier were applied to each of the 25 sites and show very good agreement. The median across all methods is a good approximation of the expected LAI product values (Fig. 2) with 4 out of the 8 methods investigated very close (RMSE (Root Mean Square Error) lower than 0.05) to the median across all methods. The time series made with the median across the 8 methods will therefore be considered as the reference values, LAI_{ref}, in the following. This LAI_{ref} does not show any missing data since the gaps in the original time series were filled by the reconstructed values of the 8 methods. LAI_{ref} constitutes a good reference with minimal uncertainties attached to the LAI values because of the temporal smoothing coming from each method and the computation of the median across the 8 methods. A second set of time series was generated to provide realistic LAI values: the original LAI values, LAI_{ori} were complemented at the location of missing data by LAI_{ref} values contaminated by a noise that was randomly drawn within the distribution of residuals ($LAI_{ref} - LAI_{ori}$) for each site. These realistic but continuous temporal profiles with no gaps (LAI_{comp}) will be used in the second step of the approach for simulating time series with gaps.

2.3.2 Simulation of time series with gaps

In the second step, emphasis was put on the occurrence of missing data (% Gap). The gap structure observed over each one of the 420 sites was applied to the completed time series (LAI_{comp}). This allows the gap structure to be more realistic as compared to other strategies that would consist in randomly simulating gaps. However, vegetation type and the associated climate experienced, hence the cloud occurrence and corresponding gap structure, are probably correlated. To account for such possible dependency, the gap structure applied to one of the 25 sites was selected within the gap sites belonging to the same vegetation class (Table 3). Note that the balance amongst vegetation classes in BELMANIP2 was preserved (Table 3) providing approximate representativeness of global scale conditions regarding the occurrence of missing data: SB and CG represent about 2/3 of the land area, associated with relatively low fraction of missing data (gap percentage, Fig. 3). Forests represent about 1/3 of the global land area with relatively high fraction of missing data. However, sites with less than 9 observations for the whole 9 yr period (i.e. less than one observation per year in average) were not considered since none of the methods will be able to provide a fair reconstruction of the LAI time course. A total of 384 sites were finally used to simulate the missing data (Table 3). Because each vegetation class is represented by 5 sites used for the LAI_{ref} and LAI_{comp} values, a total of 1920 cases (384 × 5) with realistic LAI uncertainties and gap structure were finally available.

2.4 Metrics used to quantify performances

The performances of the 8 methods considered in this study were evaluated based both on LAI values as well as phenology. Note that when the reconstructed LAI values (LAI_{rec}, Fig. 1) were outside the definition domain ($0 < LAI_{rec} < 7$), the reconstructed value was systematically set to the closest bound (0 or 7). Note also that in several situations, the methods may fail to reconstruct the whole time series due to long periods

of gaps. This will be quantified by the reconstruction fraction (% Reconstructions), i.e. the fraction of dates with reconstructed values in LAI_{rec} time series.

2.4.1 Metrics based on LAI values

The RMSE (Root Mean Square Error) computed over all cases quantifies the fidelity of the reconstruction of the time series:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\left(\mathrm{LAI}_{\mathrm{rec}}^{j}(t) - \mathrm{LAI}_{\mathrm{ref}}^{j}(t)\right)^2}{\sum_{j=1}^{N} n_j}} \quad (1)$$

where $LAI^{j}_{ref}(t)$ and $LAI^{j}_{rec}(t)$ are respectively the reference and the reconstructed values for date t and case j, n is the number of dates with observations for case j and N is the number of cases considered.

Finally, the metrics proposed by Whittaker (1923) called here roughness will be used to quantify the shaky nature of the reconstructed time series:

$$\mathrm{Roughness} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\left(\mathrm{LAI}_{\mathrm{rec}}^{j}(t) - \mathrm{LAI}_{\mathrm{rec}}^{j}(t-1)\right)^2}{\sum_{j=1}^{N} n_j}} \quad (2)$$

2.4.2 Metrics based on phenology

The 5 evergreen broadleaf forest sites were excluded from these metrics since the identification of seasonality was questionable at the single pixel scale considered in this study, and would result in large uncertainties in the timing of phenological stages if they exist (some sites do not show obvious seasonality). Three phenological events were considered: the Start of Season (SoS), Maximum of Season (MoS) and End of Season (EoS). SoS and EoS were defined similarly to Jönsson and Eklundh (2002) as the timing when LAI reaches 20 % of the whole LAI amplitude before (SoS) or after (EoS) the timing of maximum LAI (MoS). The reference dates of these three stages were derived by applying this phenology extraction method to the LAI_{ref} data (P_{ref}). Then the RMSE for the timing of SoS, MoS and EoS are computed:

$$\mathrm{RMSE\,(days)} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{s=1}^{m_j}\left(P_{\mathrm{rec}}^{j}(s) - P_{\mathrm{ref}}^{j}(s)\right)^2}{\sum_{j=1}^{N} m_j}} \quad (3)$$

where $P^{j}_{ref}(s)$ and $P^{j}_{rec}(s)$ are, respectively, the reference and the reconstructed dates for the phenological events and case j, m is the number of phenological events for case j(i.e. the number of seasons in the time series j) and N is the number of cases considered.

3 Results

The methods will first be evaluated with regards to fidelity and roughness of the reconstructed time series. Then, they will be evaluated with regard to their ability to describe the phenology. In both cases, the impact of the occurrence of missing data (called % Gap) will be analyzed.

3.1 Performances for LAI reconstruction

Before investigating quantitatively the performances through the several metrics envisioned, the main features of each method will be qualitatively assessed. Five cases of LAI_{rec} within the 1920's ones have been selected (Fig. 4) and their temporal profiles plotted against LAI_{ref}. They represent the 5 typical vegetation types under a medium occurrence of missing data. Visual inspection shows that:

– The climatology is often shifted from the reference temporal profile, highlighting the inter-annual fluctuations, particularly for non-forest vegetation types (SB and CG in Fig. 4).

– In presence of periods with long and continuous missing data, several methods were not able to reconstruct the time series over these periods, particularly TSGF and Clim, while AGF, EMD, LPF fail for the entire time series (SB in Fig. 4). However, the other methods (ICSSA, Whit, SGF) showing continuity in LAI_{rec} do not always provide realistic (as compared with LAI_{ref}) reconstructions in such cases.

– When observations show a significant level of temporal noise (the forest sites in Fig. 4: DB, EB and NF), significant differences are observed between the methods, both in terms of fidelity (closeness to LAI_{ref}) and roughness, particularly for SGF and EMD.

3.1.1 Capacity to reconstruct the temporal series in presence of missing data

All the methods were not able to fill the gaps, i.e. to provide an estimated value in gaps. This was quantified by the % Success, i.e. the fraction of gaps that were able to be filled. Whit, SGF, ICSSA allow to fill most of the gaps even if they are very long (Fig. 5a). Conversely, EMD, LPF and AGF show a rapid decrease of % Success with the length of the gaps, with no reconstructions for gaps longer than 85 (AGF) to 130 days (EMD, LPF). Even for small gaps, only 50 % of them were filled. This is due to the fact that a specific gap may be associated to other ones in a close vicinity in the time series. TSGF is able to fill gaps up to gap length of 128 days as expected by its definition. The climatology shows also a progressive decrease of % Success with gaps of 128 days length being filled in 80 % of cases because of the accumulation of observations over the 9 yr period.

The capacity to fill individual gaps has consequences on the reconstruction fraction (% Reconstructions), i.e. the fraction of dates with reconstructed LAI values, LAI_{rec}, relative to the total number of dates in the LAI_{comp} time series. Only three methods (Whit, SGF, ICSSA) were able to provide a continuous reconstructed time series over all the cases investigated (Fig. 5b) even for large occurrence of missing data in agreement with Fig. 5a. In contrast, AGF is characterized by the smallest % Success (Fig. 5a) and % Reconstruction (Fig. 5b): only 50 % of the dates are reconstructed for cases with more than 25 % of missing data in their time series (Fig. 5b). LPF and EMD that do not accept gaps longer than 128 days (Table 1) show also a similar drastic decrease of the reconstruction fraction with the occurrence of missing data in the cases considered (Fig. 5b). The TSGF method, although also not filling gaps longer than 128 days is more resilient to the occurrence of gaps: TSGF was able to reconstruct more than 50 % of the data for cases with more than 60 % of missing data. When a gap longer than 128 days appears in a time series, the remaining parts of the time series are reconstructed. This was not the case for LPF and EMD for which the whole time series was not reconstructed for cases having a gap longer than 128 days. Clim allows reconstructing most of the time series, even for cases with large amount of missing data, benefiting from the replications between years, cloudy days being not always the same day of the year.

To improve the robustness of the metrics used to characterize the performances on LAI and phenology reconstruction, they will be computed only when % Reconstructions > 50 % (Fig. 5b). As a consequence, all the methods will be compared for cases with less than % Gap < 20 %; TSGF, Clim, ICSSA, SGF and Whit will be compared for 20 % < % Gap < 60 %; and for % Gap > 60 %, only Clim, ICSSA, SGF and Whit will be compared (Fig. 5b).

3.1.2 Fidelity to LAI_{ref}

Fidelity is quantified by RMSE. To better highlight the reconstruction capacity of the methods, RMSE were computed by comparing LAI_{rec} with LAI_{ref} either over actual dates with observations or over dates with missing data in LAI_{sim}. Results show that, except for Clim and SGF, all the other methods show generally good fidelity with LAI_{ref} for the dates where observations are available (Fig. 6). These good performances

are observed almost independently from the occurrence of missing data (Fig. 6). The higher RMSE values observed for SGF is due to a positive bias induced by the fitting of the upper envelope of the observations. Clim shows a RMSE value close to that of SGF (Fig. 6) that mostly refers to the inter-annual variability of LAI seasonality.

Over dates of missing observations, the reconstruction capacity degrades rapidly as a function of the length of gaps for all methods except Clim that keeps a RMSE value around 0.35 independently from the gap length as expected (Fig. 7a). LPF and TSGF provide the best performances up to gap length around 100 days when Clim starts to be the best method. AGF, Whit, EMD and SGF show similar performances with RMSE lower than the climatology up to gap length around 70 days, while ICSSA performances rapidly degrade with the length of gaps. The fidelity of reconstructions in gaps as a function of the fraction of missing observations in the time series (Fig. 7b) derives logically from the reconstruction performances as a function of the gap length (Fig. 7a). The RMSE values computed over dates of missing observations are relatively low for all methods up to % Gap < 20 %. Then, SGF, Whit and ICSSA show a rapid increase of the RMSE with % Gap, with poorer performances as compared to Clim for % Gap > 30 % (Fig. 7b). These three methods show obvious artifacts when reconstructing long gaps (Fig. 4, non-forest sites 5 and 338). TSGF shows relatively low RMSE values up to 60 % gap (Figs. 6, 7b). Clim shows similar performances over dates with missing data (Fig. 7b) and dates with observations (Fig. 6) as expected since it is not dependent on the local observations.

3.1.3 Roughness

The roughness was computed over the whole reconstructed time series and is presented as a function of % Gap (Fig. 8). Results show that for % Gap < 30 %, EMD and SGF show the highest roughness values in agreement with the previous qualitative observations (Fig. 4). The behavior of SGF is controlled by its iterative nature that puts emphasis on fidelity relative to the upper envelope. For EMD, the 10 modes selected were showing variable patterns and it was difficult to find a better boxcompromise between smoothness and fidelity. ICSSA shows a roughness value close to that of Clim for % Gap < 20 %. However, the roughness of ICSSA strongly increases when % Gap > 20 %. This is partially due to inconsistencies observed in its temporal pattern with abrupt variations in the periods with high discontinuities in the data (Fig. 4, jumps observed between the lowest and the highest values when data are missing for non-forest sites 5 and 338). AGF and LPF show the smoothest temporal profiles however limited to cases with % Gap < 20 %. Whit provides always smooth reconstructed profiles, even for large amount of missing data. This is obviously controlled by the lambda parameter. Whit is just slightly rougher than LPF and AGF. TSGF shows a slightly higher roughness values than Whit for the % Gap < 30 %, with a significant decrease when % Gap increases. This is due to the linear interpolation used to fill the gaps that explains also the decrease of roughness for EMD, SGF and Clim when % Gap increases.

3.2 Performances for describing the phenology

The capacity of the several methods to identify the main phenological stages (SoS, MoS and EoS) was evaluated using the dates derived from LAI_{ref} as a reference. The performances (RMSE in days) were analyzed as a function of the occurrence of missing data (% Gap). Results show a general degradation of RMSE when % Gap increases for the three stages considered.

Closer inspection of performances in terms of RMSE for SoS estimates shows large differences between methods (Fig. 9a). For complete time series (% Gap = 0), RMSE values are between 3 days (LPF) and 15 days (AGF), with the exception of the climatology with a RMSE around 25 days, indicating a significant inter-annual variability in the timing of SoS. EMD, TSGF, Whit and SGF have RMSE around 1 days. For discontinuous time series, Whit, SGF and ICSSA show a continuous and steep increase of RMSE with % Gap. Conversely, the RMSE values of Clim and, in a lesser way, TSGF increase moderately with % Gap. Similar patterns are observed for MoS (Figure 9b) with however smaller differences between methods for % Gap < 20 % and a slightly lower rate

of increase of RMSE with % Gap except for Clim. The performances for EoS (Fig. 9c) appear to be very similar to what is observed for SoS (Fig. 9a). The Climatology (Clim) performs better than ICSSA, SGF and Whit for % Gap > 40 % for SoS and EoS, and for % Gap > 50 % for MoS. TSGF yields the smallest RMSE for % Gap > 20 % for SoS, MoS and EoS with however only small differences as compared to Clim for EoS (Fig. 9a–c).

4 Discussion and conclusion

This study compares 8 methods designed to improve the continuity and consistency of time series by filling gaps created by missing observations and smoothing the temporal profiles to reduce local uncertainties. However, the performances of the different methods for processing time series depend on their implementation. The selected methods were applied here as close as possible to their standard implementation including the original parameterization as proposed by their authors. The time series considered correspond to actual MODIS LAI products over a sample of sites that were selected to be both representative of the diversity of seasonal patterns and of the distribution of the missing observations. This approach was expected to improve the realism of the context of the analysis that accounts for the implicit links between the vegetation type and the distribution of missing observations. This may be critical for filling gaps or smoothing the time series. The approach allowed defining a set of reference time series used to quantify the accuracy of each of the 8 methods as a function of the fraction of missing observations.

Results clearly show that some methods including LPF, AGF and EMD were failing in about 50 % of the situations when the fraction of missing observations was larger than 20 % which represents about 60 % of the situations investigated here. This is partly due to the principles on which these methods are based, but also partly to their implementation. Consequently, great care should be taken with the implementation of such methods to improve their rate of applicability in case of significant periods with missing observations. Conversely, ICSSA, Whit and Clim methods were applicable in almost all situations while TSGF shows intermediate behavior.

For the methods resilient to periods of missing observations of significant length, their capacity to provide realistic interpolation between actual observations was challenged in cases corresponding to medium to high fraction of missing data. SGF, designed to fit the upper-envelope of observations, performs poorly (large RMSE and positive Bias) over MODIS LAI time series. Better filtering principles are thus required to reject outliers possibly contaminated by residual clouds. ICSSA and Whit show unreliable interpolated values in the medium (few weeks) to large (few months) periods of missing data although these methods are adjusted over the whole time series. The TSGF method appears to provide the most reliable interpolation capacity due to its adaptive temporal window, although limited to gaps smaller than 128 days. For longer periods without observations, the Clim method appears to be the best one provided that enough data are available over the time series of years used to build the climatology. Note that the reconstruction performances for the best methods and for gaps shorter than 100 days fulfills the GCOS criterion on LAI uncertainties (RMSE> 0.5) (GCOS, 2010) although the reconstruction uncertainty is only part of the error budget.

Each method is based explicitly (Whit) or implicitly (the other methods) on a balance between fidelity and smoothness. This is clearly demonstrated when plotting Roughness and RMSE performances for each of the 25 selected sites (Table 2) for a class of occurrence of missing data (Fig. 10). For each method, all the 25 sites are approximately organized around a line passing through the origin. The slope of this line indicates the balance between fidelity and roughness. For relatively continuous time series (0 % < % Gap < 15 %), TSGF, ICSSA and Whit focus more on fidelity than smoothness (Fig. 10, left). Conversely, LPF, EMD and AGF are focusing more on smoothness than fidelity. SGF constitutes a particular case because the fidelity is targeting the upper envelope of the points, resulting in larger RMSE values, while roughness is also quite important as described previously. Clim provides the steepest slope, with smooth temporal profiles but a loose match with observations. Note that the slope of Clim is in

between that of LAI_{comp} and LAI_{ref} (for which RMSE was replaced by the standard deviation between observations). For the larger occurrence of missing data (Fig. 10, right), the slopes increase significantly due to an increase of RMSE mostly due to inaccurate reconstructions in the gaps, and a decrease of roughness due to more simple patterns in the observations, except for ICSSA as noticed earlier.

The slope between RMSE and Roughness (Fig. 10) appears thus a good indicator of the balance between fidelity (RMSE) and smoothness of each method and its associated sets of parameters. The overall performances may be described by the distance to the ideal case (RMSE = Roughness = 0) in the [Roughness, RMSE] feature plane averaged over the 25 sites considered: the closer to the origin [0, 0], the smoother and the better match with LAI_{ref} (low RMSE). The behavior of each method as a function of the occurrence of missing data is well sketched in the [Performances, Slopes] feature plane (Fig. 11). For low amount of missing data, all the methods provide good performances except SGF and Clim for the reasons exposed previously. When the fraction of missing data increases, each method follows a particular pattern (the black arrow in Fig. 11) with a degradation of the performances and an increase in the slope indicating more emphasis on the smoothness of the temporal profiles. For medium and high occurrence of missing data, TSGF provides clearly the best overall performances although restricted to gaps smaller than 128 days, followed by Whit. SGF and ICSSA show poor performances.

The consequences of the application of the several time series processing methods on their capacity to describe phenology characteristics were finally evaluated. As expected, the methods providing the best accuracy on LAI estimation were also more accurate for dating specific phenological events such as start, maximum and end of season (Fig. 12).

The effect of gaps on the derivation of time series appears as a major limitation of the accuracy of the reconstructed temporal profiles. Techniques based on the processing of the time series as a whole (ICSSA, EMD, LPF, Whit and Clim) were not demonstrated to perform systematically better than techniques based on a limited temporal window (AGF, SGF, TSGF) although they were expected to fill long gaps with the "experience" gained across the several years available in the time series. Local methods were generally more faithful but were lacking capacity to fill long gaps. Most methods were performing poorer than Clim for gaps longer than about 100 days. Future works should therefore be dedicated to develop methods where the features derived from the exploitation of the several years available in the time series including the climatology, could be injected more explicitly as a background information for improving the reliability of methods working over a limited time window such as a season or part of it, with emphasis on the capacity to provide accurate phenological timing as proposed in Verger et al. (2012).

Acknowledgements. This research has received funding from the Geoland2 European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement #218795. A. Verger was funded by the VALi + d postdoctoral program (FUSAT, GV-20100270). Many thanks also to the MODIS teams who made available to the wide community the LAI data used in this study.

References

Alcaraz-Segura, D., Liras, E., Tabik, S., Paruelo, J., and Cabello, J.: Evaluating the Consistency of the 1982–1999 NDVI Trends in the Iberian Peninsula across Four Time-series Derived from the AVHRR Sensor: LTDR, GIMMS, FASIR, and PAL-II, Sensors, 10, 1291–1314, 2010.

Atkinson, P. M., Jeganathan, C., Dash, J., and Atzberger, C.: Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology, Remote Sens. Environ., 123, 400–417, 2012.

Azzali, S. and Menenti, M.: Mapping isogrowth zones on continental scale using temporal Fourier analysis of AVHRR-NDVI data, Int. J. Appl. Earth Observ. Geoinform., 1, 9–20, 1999.

Bacour, C., Baret, F., Béal, D., Weiss, M., and Pavageau, K.: Neural network estimation of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data: principles and validation, Remote Sens. Environ., 105, 313–325, 2006a.

Bacour, C., Bréon, F.-M., and Maignan, F.: Normalization of the directional effects in NOAA-AVHRR reflectance measurements for an improved monitoring of vegetation cycles, Remote Sens. Environ., 102, 402–413, 2006b.

Baret, F., Morissette, J., Fernandes, R., Champeaux, J. L., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigue, S., and Nickeson, J.: Evaluation of the representativeness of networks of sites for the global validation and inter-comparison of land biophysical products. Proposition of the CEOS-BELMANIP, IEEE Trans. Geosci. Remote, 44, 1794–1803, 2006.

Baret , F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Weiss, M., Samain, O., Roujean, J. L., and Leroy, M.: LAI, FAPAR and FCOVER CYCLOPES global products derived from VEGETATION, Part 1: Principles of the algorithm, Remote Sens. Environ., 110, 275–286, 2007.

Baret, F., Weiss, M., Verger, A., and Kandasamy, S.: BioPar Methods Compendium – LAI, FAPAR and FCOVER from LTDR AVHRR series. Report for EC contract FP-7-218795, available at http://www.geoland2.eu/portal/documents/CA80C881.html, INRA-EMMAH, Avignon, 46 pp., 2011.

Beck, P. S. A., Atzberger, C., Hogda, K. A., Johansen, B., and Skidmore, A. K.: Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI, Remote Sens. Environ., 100, 321–334, 2006.

Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C.: A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data, Remote Sens. Environ., 114, 1312–1323, 2010.

Boles, S. H., Xiao, X., Liu, J., Zhang, Q., Munkhtuya, S., Chen, S., and Ojima, D.: Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data, Remote Sens. Environ., 90, 477–489, 2004.

Borak, J. S. and Jasinski, M. F.: Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States, Agr. Forest Meteorol., 149, 320–332, 2009.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M.: GEOV1: LAI, FAPAR Essential Climate Variables and FCover global time series capitalizing over existing products, Part 2: Validation and inter-comparison with reference products, Remote Sens. Environ., in review, 2012.

Chen, J. M. and Black, T. A.: Defining leaf area index for non-flat leaves, Defining leaf area index for non-flat leaves, Plant Cell Environ., 15, 421–429, 1992.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L.: A simple method for reconstructing a high quality NDVI time series data set based on the Savitzky-Golay filter, Remote Sens. Environ., 91, 332–344, 2004.

Coops, N. C., Wulder, M. A., and Iwanicka, D.: Large area monitoring with a MODIS-based Disturbance Index (DI) sensitive to annual and seasonal variations, Remote Sens. Environ., 113, 1250–1261, 2009.

De Kauwe, M. G., Disney, M. I., Quaife, T., Lewis, P., and Williams, M.: An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, Remote Sens. Environ., 115, 767–780, 2011.

Defourny, P., Bicheron, P., Brockmann, C., Bontemps, S., Van Bogaert, E., Vancutsem, C., Pekel, J. F., Huc, M., Henry, C., Ranera, F., Achard, F., di Gregorio, A., Herold, M., Leroy, M., and Arino, O.: The first 300 m global land cover map for 2005 using ENVISAT MERIS time series: a product of the GlobCover system, Proceedings of the 33rd International Symposium on Rememote Sensing of Environment, Stresa, Italy, 1–4, 2009.

Demir, B. and Erturk, S.: Empirical mode decomposition pre-process for higher accuracy hyperspectral image classification, IGRARSS, IEEE, Boston, USA, 939–941, 2008.

Deng, F., Chen, J. M., Chen, M., and Pisek, J.: Algorithm for global leaf area index retrieval using satellite imagery, IEEE Trans. Geosci. Remote, 44, 2219–2229, 2006.

Dente, L., Satalino, G., Mattia, F., and Rinaldi, M.: Assimilation of leaf area index derived from ASAR and MERIS data into CERES-W heat model to map wheat yield, Remote Sens. Environ., 112, 1395–1407, 2008.

Eilers, P. H. C.: A Perfect Smoother, Anal. Chem., 75, 3631 pp., 2003.

Ganguly, S., Schull, M. A., Samanta, A., Shabanov, N. V., Milesi, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Generating vegetation leaf area index earth system data record from multiple sensors, Part 1, Theory, Remote Sens. Environ., 112, 4333–4343, 2008.

Gao, F., Morisette, J., Wolfe, R. E., Ederer, G., Pedelty, J., Masuoka, E. J., Myneni, R., Tan, B., and Nightingale, J. M.: An algorithm to produce temporally and spatially continuous MODIS LAI time series, IEEE Geosci. Remote Sens. Lett., 5, 60–64, 2008.

GCOS: GCOS-107: Supplemental details to the satellite based component of the "implementation plan for the global observing system for climate in support of the UNFCCC", GCOS/WMO, Geneve, Switzerland, 103 pp., 2006.

GCOS: GCOS-13: Implementation plan for the global observing system for climate in support of the UNFCCC, GCOS-138, WMO, 186 pp., 2010.

Golyandina, N. and Osipov, E.: The "Caterpillar"-SSA method for analysis of time series with missing values, J. Stat. Plan. Infer., 137, 2642–2653, 2007.
Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., and Carroll, M.: Towards an operational MODIS continuous field percent tree cover algorithm: examples using AVHRR and MODIS data, Remote Sens. Environ., 83, 303–319, 2002.
Hansen, M. C., Shimabukuro, Y. E., Potapov, P., and Pittman, K.: Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover, Remote Sens. Environ., 112, 3784–3793, 2008.
Heiskanen, J. and Kivinen, S.: Assessment of multispectral, -temporal and -angular MODIS data for tree cover mapping in the tundra – taiga transition zone, Remote Sens. Environ., 112, 2367–2380, 2008.
Hird, J. N. and McDermid, G. J.: Noise reduction of NDVI time series: An empirical comparison of selected techniques, Remote Sens. Environ., 113, 248–258, 2009.
Holben, B. N.: Characteristics of maximum-value composite images from temporal AVHRR data, Int. J. Remote Sens., 7, 1417–1434, 1986.
Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London, Series A, Mathematical, Physical and Engineering Sciences, 454, 903–995, 1998.
Jakubauskas, M. E., Legates, D. R., and Kastens, J. H.: Crop identification using harmonic analysis of time series AVHRR NDVI data, Computers Electronics in Agriculture, 37, 127–139, 2002.
Jiang, B., Liang, S., Wang, J., and Xiao, Z.: Modeling MODIS LAI time series using three statistical methods, Remote Sens. Environ., 114, 1432–1444, 2010
Jönsson, P. and Eklundh, L.: Seasonality extraction by function-fitting to time series of satellite sensor data, IEEE Trans. Geosci. Remote, 40, 1824–1832, 2002.
Jönsson, P. and Eklundh, L.: TIMESAT – a program for analyzing time series of satellite sensor data, Computers and Geosciences, 30, 833–845, 2004.
Kandasamy, S., Neveux, P., Verger, A., Buis, S., Weiss, M., and Baret, F.: Improving the consistency and continuity of MODIS 8 day leaf area index products, International Journal of Electronics and Telecommunications, 58, 141–146, 2012.
Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., and Lee, R.-Y.: Image masking for crop yield forecasting using AVHRR NDVI time series imagery, Remote Sens. Environ., 99, 341–356, 2005.
Kobayashi, H., Suzuki, R., and Kobayashi, S.: Reflectance seasonality and its relation to the canopy leaf area index in an eastern Siberian larch forest: Multi-satellite data and radiative transfer analyses, Remote Sens. Environ., 106, 238–252, 2007.
Ma, M. and Veroustraete, F.: Reconstructing pathfinder AVHRR land NDVI timeseries data for the Northwest of China, Adv. Space Res., 37, 835–840, 2006.
Martinez, B. and Gilabert, M. A.: Vegetation dynamics from NDVI time series analysis using the wavelet transform, Remote Sens. Environ., 113, 1823–1842, 2009.
Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R., Eerens, H., and Leo, O.: A protocol for the evaluation of agreement between space remote sensing time series and application to SPOT-VEGETATION fAPAR products, International Journal of Applied Observations and Geoinformation, in press, 2012.
Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83, 214–231, 2002.
Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C.: Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends Ecol. Evol., 20, 503–510, 2005.
Pinzon, J. E., Brown, M. E., and Tucker, C. J.: Satellite time series correction of orbital drift artifacts using empirical mode decomposition, in: EMD and its Applications, edited by: Huang, N. E. and Shen, S. S. P., World Scientific Publishers, Singapore, 167–186, 2005.
Pouliot, D., Latifovic, R., Fernandes, R., and Olthof, I.: Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data, Remote Sens. Environ., 113, 1749–1759, 2009.
Qi, J. and Kerr, Y.: On current compositing algorithms, Remote Sens. Rev., 15, 235–256, 1997.
Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C.: Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFC, 371 pp., 1974.
Savitzky, A. and Golay, M. J. E.: Smoothing and differentiation of data by simplified least square procedures, Anal. Chem., 36, 1627–1639, 1964.

Schubert, P., Eklundh, L., Lund, M., and Nilsson, M.: Estimating northern peatland CO_2 exchange from MODIS time series data, Remote Sens. Environ., 114, 1178–1189, 2010.

Shabanov, N. V., Huang, D., Yang, W., Tan, B., Knyazikhin, Y., Myneni, R. B., Ahl, D. E., Gower, S. T., and Huete, A. R.: Analysis and Optimization of the MODIS Leaf Area Index Algorithm Retrievals Over Broadleaf Forests, IEEE Trans. Geosci. Remote, 43, 1855–1865, 2005.

Thenkabail, P. S., Schull, M., and Turral, H.: Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data, Remote Sens. Environ., 95, 317–341, 2005.

Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory, part 2: Analysis of the NOAA GMCC data, 1974–1985, J. Geophys. Res., 94, 8549–8565, 1989.

Tucker, C. J., Pinzón, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E., and El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, Int. J. Remote Sens., 26, 4485–4498, 2005.

Verger, A., Baret , F., and Weiss, M.: Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products, Remote Sens. Environ., 112, 2789–2803, 2008.

Verger, A., Baret , F., and Weiss, M.: A multisensor fusion approach to improve LAI time series, Remote Sens. Environ., 115, 2460–2470, 2011.

Verger, A., Baret, F., Weiss, M., Kandasamy, S., and Vermote, E.: The CACAO method for smoothing, gap filling and characterizing seasonal anomalies in satellite time series, IEEE Trans. Geosci. Remote, in press, 2012.

Vermote, E., Justice, C., Csiszar, I., Eidenshink, J., Myneni, R., Baret, F., Masuoka, E., and Wolfe, R.: A Terrestrial Surface Climate Data Record for Global Change Studies, American Geophysical Union Fall Meeting, San-Francisco, USA, 2009.

Viovy, N., Arino, O., and Belward, A.: The Best INdex Slope Extraction (BISE): a method for reducing noise in NDVI timeseries, Int. J. Remote Sens., 13, 1585–1590, 1992.

Weiss, M., Baret , F., Garrigues, S., Lacaze, R., and Bicheron, P.: LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION, part 2: Validation and comparison with MODIS Collection 4 products, Remote Sens. Environ., 110, 317–331, 2007.

Whittaker, E. T.: On a new method of graduation, Proc. Edinburgh Math. Soc., 41, 63–75, 1923.

Wylie, B. K., Fosnight, E. A., Gilmanov, T. G., Frank, A. B., Morgan, J. A., Haferkamp, M. R., and Meyers, T. P.: Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains, Remote Sens. Environ., 106, 399–413, 2007.

Yang, W., Shabanov, N. V., Huang, D., Wang, W., Dickinson, R. E., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Analysis of leaf area index products from combination of MODIS Terra and Aqua data, Remote Sens. Environ., 104, 297–312, 2006a.

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D. E., Gower, S. T., Nemani, R. R., Knyazikhin, Y., and Myneni, R.: MODIS leaf area index products: from validation to algorithm improvement, IEEE Trans. Geosci. Remote, 44, 1885–1898, 2006b.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sens. Environ., 115, 1171–1187, 2011.

Table 1. List of the methods investigated. Length of processing window (Whole means that the processing window is the whole time series) and maximum gap length tolerated are indicated.

Abrev.	Method	Principles	Processing Window length (days)	Maximum gap length (days)	Reference
ICSSA	Iterative Caterpilar Singular Spectrum Analysis	Decomposition into EOF's using Eigen value decomposition	Whole	–	Golyandina and Osipov (2007)
EMD	Empirical Mode Decomposition	Decomposition into IMF's by "sifting"	Whole	128	Huang et al. (1998)
LPF	Low Pass Filtering	Fitting a harmonic curve to the series, followed by 2-pass filtering of the residuals	Whole	128	Bacour et al. (2006b)
Whit	Whitaker	Penalized Least Square Regression – Smoothness is governed by a parameter value	Whole	–	Eilers (2003)
SGF	Adaptive Savitzky-Golay Filter	Savitzky-Golay filter with iterations to fit the upper envelope of the series	72–112	–	Chen et al. (2004)
TSGF	Temporal Smoothing and Gap Filling	Savitzky-Golay filtering with flexible window and linear interpolation for gaps	48–128	128	Verger et al. (2011)
AGF	Asymmetric Gaussian Filter	Fitting Asymmetric Gaussian Function to seasons	60–300	72	Jönsson and Eklundh (2002, 2004)
Clim	Climatology	Inter-annual median for each 8-day period	Whole	128*	Baret et al. (2011)

* The maximum gap length applies on the yearly reconstructed time series, not on the original time series.

Table 2. Sites selected for the study for the 5 biomes. The site number in BELMANIP2 ensemble of 420 sites, latitude and longitude and fraction of missing data (% Gap) are indicated.

Biome	Site #	Lat.(°)	Lon. (°)	% Gap
Shrub Savannah Bare (SB)	5	−34.02	−65.63	1
	136	−18.46	44.41	3
	176	12.49	36.34	2
	186	10.70	39.41	1
	293	−21.54	143.83	1
Crop Grassland (GC)	69	38.63	−98.91	5
	127	−27.61	27.95	1
	225	35.09	−1.00	1
	280	−31.38	116.87	1
	338	25.99	68.52	0
Deciduous Broad leaf forests (DB)	131	−11.99	16.43	11
	146	−5.45	31.74	9
	162	4.86	28.80	8
	165	5.98	31.18	1
	296	−16.45	142.62	1
Evergreen Broadleaf Forests (EB)	19	−11.75	−53.35	41
	30	−2.68	−63.65	48
	50	17.59	−89.78	45
	142	−4.60	23.44	38
	320	24.54	121.25	41
Needle Leaf Forests (NF)	54	28.39	−108.25	11
	55	26.53	−106.68	13
	62	39.49	−120.83	18
	65	30.28	−83.85	16
	244	43.86	−1.10	20

Table 3. Number of sites per vegetation class in BELMANIP2 set of sites, and number of cases considered in the gap simulation experiment.

Vegetation class	SB	CG	DB	EB	NF	Total
Nb. sites in BELMANIP2	144	123	35	36	46	384
Nb. cases (time series) simulated	720	615	175	180	230	1920

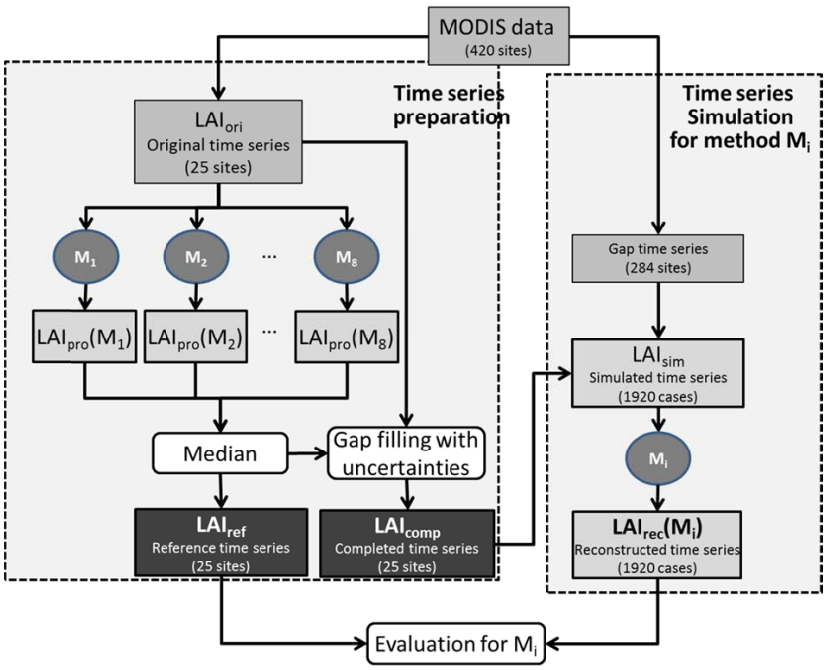

Fig. 1. Flow chart describing the approach used. M_i corresponds to method i within the 8 investigated. LAI_{rec} (M_i) corresponds to the reconstructed time series based on method M_i.

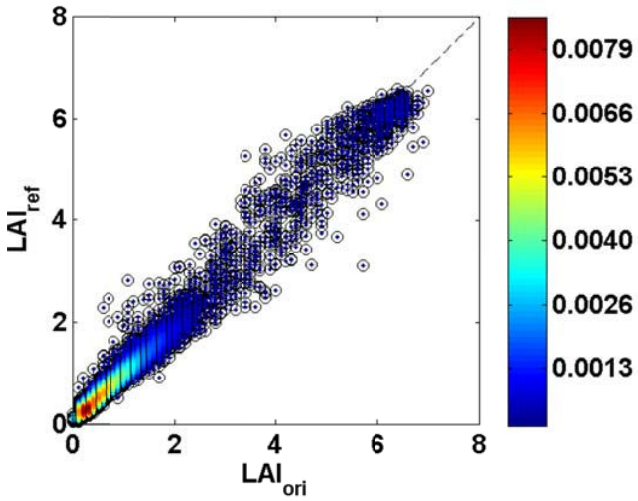

Fig. 2. Comparison between the original LAI values (LAI_{ori}) and the median of the reconstructed values (LAI_{ref}) based on the 8 methods considered over the 25 selected sites (Table 2) ($N = 7561$; $R^2 = 0.90$; RMSE = 0.231; Bias = −0.008). The colors of this density plot correspond to the frequency of data in each of the 0.1 × 0.1 cells LAI values.

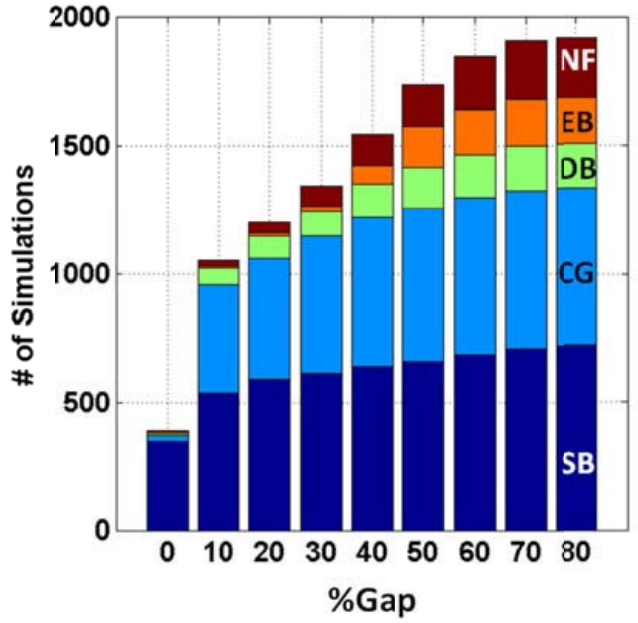

Fig. 3. Cumulated distribution of the fraction of missing data (% Gap) in the simulated time series (LAI_{sim}) for each of the 5 vegetation classes (SB, CG, DB, EB, NF).

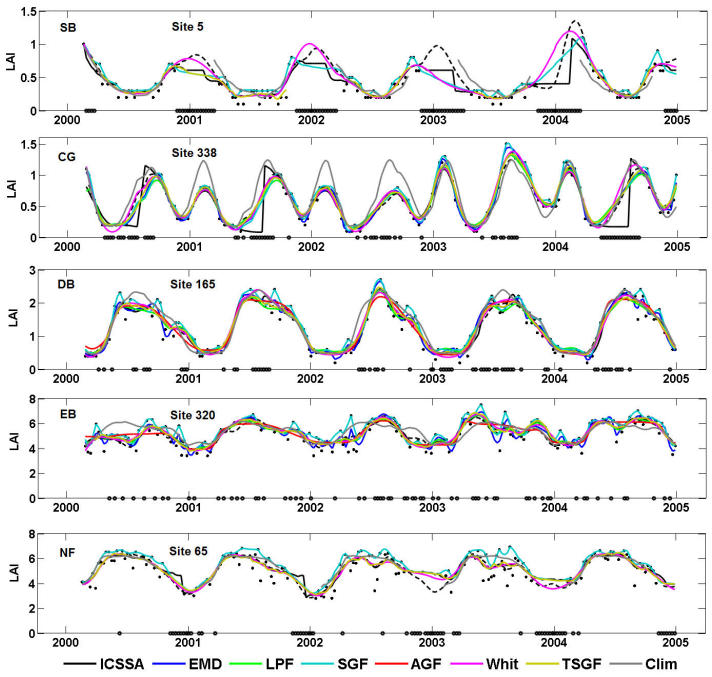

Fig. 4. Time series (LAI_{rec}) reconstructed by the several methods in presence of medium occurrence of missing data (25 % < % Gap < 35 %). Black dots correspond to LAI_{comp} at the location of observations. Empty circles on the x-axis correspond to the dates of missing data. The dashed black curve corresponds to LAI_{ref}. Note that because of the structure of missing data, EMD, LPF and AGF were not reconstructed for sites 5 and 65, as well as AGF for site 338.

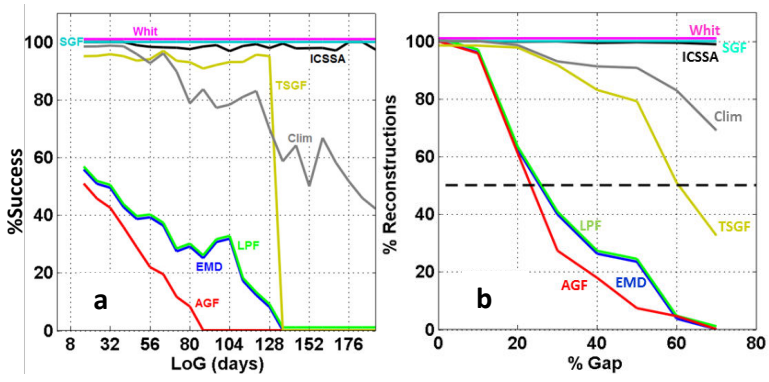

Fig. 5. (a) Fraction of gaps reconstructed (% Success) as a function of the length of the gaps (LoG). **(b)** Fraction of missing data reconstructed (% Reconstructions) as a function of the % Gap. The horizontal dashed line represents the 50 % threshold of % Reconstructions. The several methods are represented by different colors. Some values were slightly shifted vertically to ease the reading when curves were overlapping.

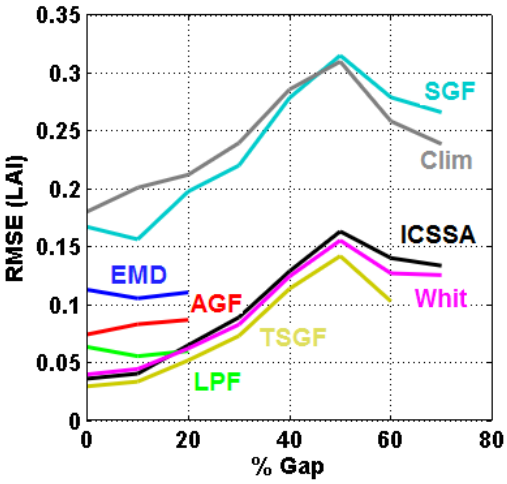

Fig. 6. RMSE as a function of % Gap. The RMSE is computed between LAI_{ref} and the reconstructed LAI_{rec} time series over dates with actual observations in LAI_{sim}. The values were slightly shifted vertically to ease the reading when curves were overlapping.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

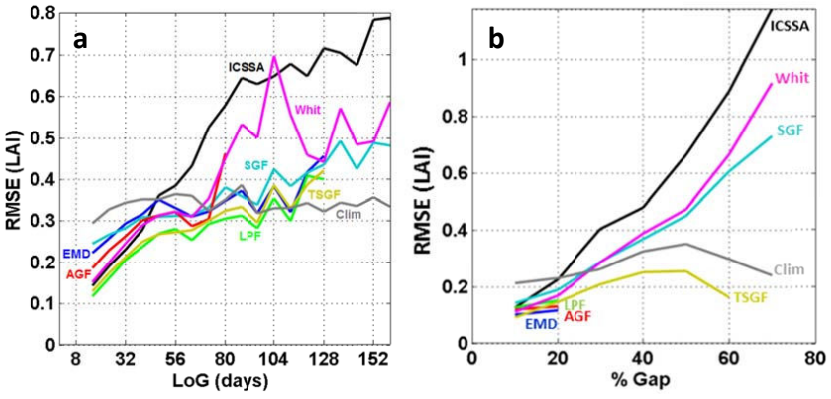

Fig. 7. RMSE as a function of the length of gaps **(a)** and fraction of missing observations (% Gap) **(b)**. The RMSE is computed between LAI_{ref} and the reconstructed LAI_{rec} time series over dates with missing observations in LAI_{sim}.

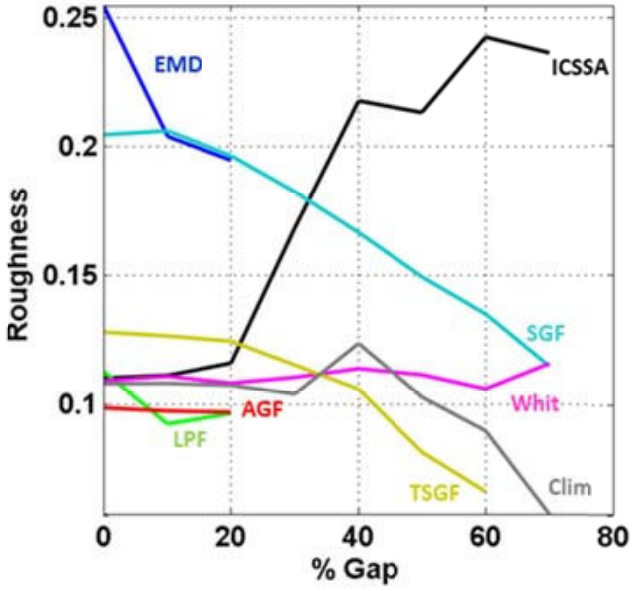

Fig. 8. Roughness of LAI_{rec} as a function of % Gap in LAI_{sim}.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

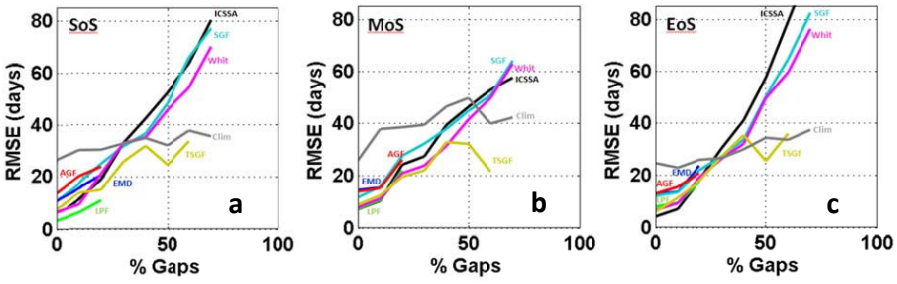

Fig. 9. RMSE relative to the timing (in days) of the start of season **(a)**, maximum of season **(b)** and end of season **(c)**. The RMSE is evaluated between the phenological dates computed with LAI_{ref} and those derived from the reconstructed LAI_{rec} time series using the several methods investigated.

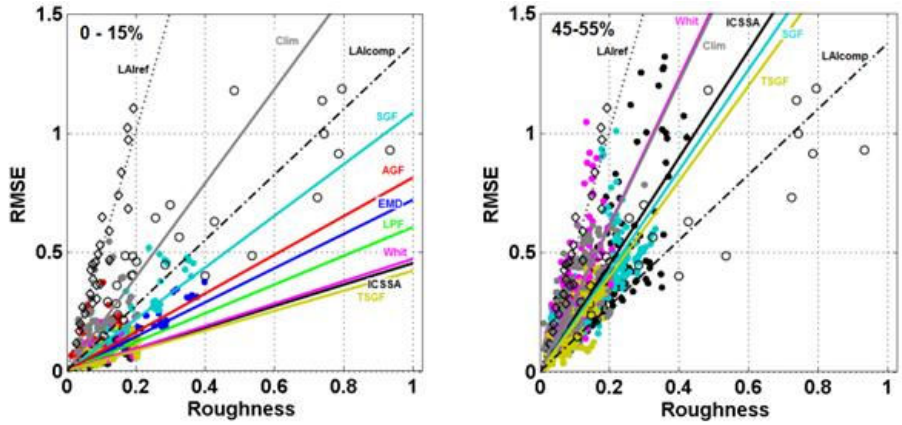

Fig. 10. RMSE as a function of Roughness observed over the reconstructed times series using the several methods. Color dots correspond to the values for the different methods over the 25 sites for cases with % Gap in the selected classes of occurrence of missing data: 0–15 % (left) and 45–55 % (right). Note that EMD, LPF and AGF are only displayed for the lowest occurrence of missing data (0 % < % Gap < 15 %). Black circles correspond to LAI_{comp} and black diamonds to LAI_{ref}. Lines correspond to the zero-offset linear regressions.

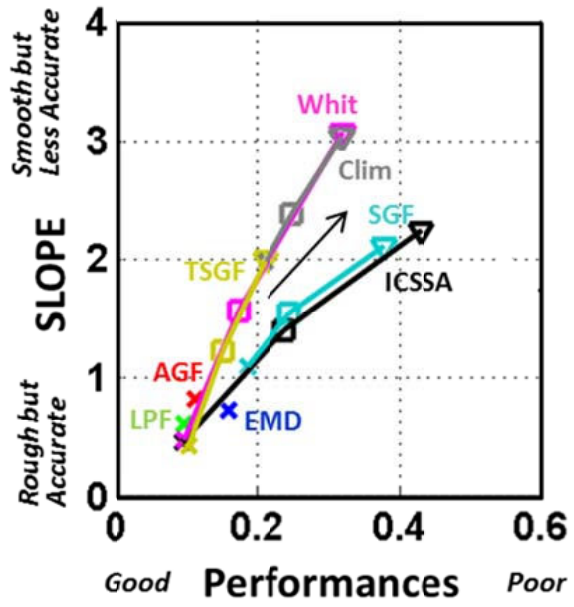

Fig. 11. Performances (distance to the origin) and slopes in the [Roughness, RMSE] feature plane (Fig. 10) associated to each method for 0 % < % Gap < 15 % (X), 25 % < % Gap < 35 % (□) and 35 % < % Gap < 45 % (▽). Note that EMD, LPF and AGF are only displayed for the lowest % Gap class. The black arrow indicates the effect of an increase of the fraction of missing data.

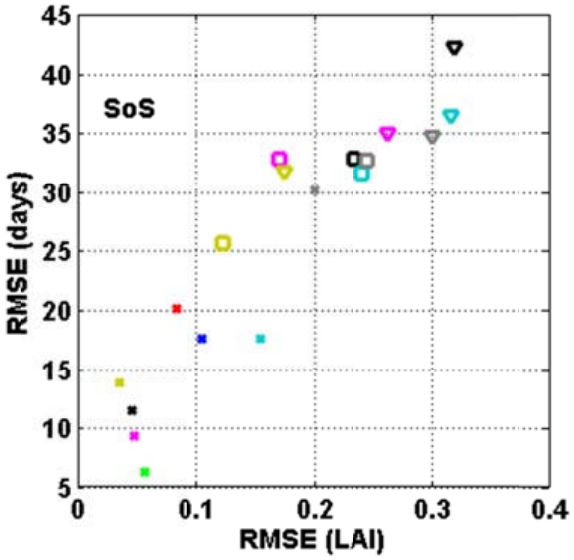

Fig. 12. Accuracy of the start of season retrieval expressed in RMSE (days) as a function of the accuracy of LAI estimated expressed in RMSE. The same colors (corresponding to methods) and markers (corresponding to classes of % Gap) as in Fig. 11 are used.

4 Near-Real Time Estimation of LAI from daily AVHRR LAI estimates

The 'BioPar' core mapping services of the GEOLAND 2, aims at producing continuous time series estimates in near real-time time, every 10-days (Lacaze, 2009). In the previous chapter, some of the potential methods for processing the time series estimates of LAI are explored for their ability to fill gaps (i.e. estimate missing values) and smooth the LAI time series estimates under varying geographic and gap conditions. Some of the methods studied in the previous chapter are capable of predicting near real-time LAI values given the past values in case of absence of radiometric observations from the sensors. In addition to the gap-filling/smoothing methods explored in the previous chapter, other methods do exist to provide future estimates. However, not many studies are available that help the user to choose an appropriate method, from a set of competing methods, for processing LAI time series data on a global scale to produce near-future estimates based on past estimates. In this context, it is of interest to GEOLAND 2 to find an appropriate method to produce near real-time (NRT) estimates of LAI by systematically and quantitatively evaluating the potential methods on a global scale under different gap conditions: amount and length of gaps.

This study was conducted using the AVHRR daily estimates of LAI. From the study on the gap-filling/smoothing methods explored in the previous chapter, the Whittaker (Whit) method (Eilers, 2003) was identified as one of the potential methods for producing near real-time estimates of LAI. The Time Series Smoothing and Gap-filling (TSGF) method (Verger et al., 2011b), explored in the previous chapter, had been improved to be included in the NRT processing chain of the LAI estimates from the VEGETATION sensor (GEOV2). Hence, it is of interest from the perspective of GEOLAND 2 to evaluate its (GEOV2 NRT algorithm that is based on TSGF and CACAO(Verger et al., 2012)) performance for AVHRR daily estimates, which is one of the sensors of interest to GEOLAND 2. One additional method, Gaussian Process Model (GPM) regression, is also identified as a potential method to make near-future predictions from past estimates. In this chapter, these three methods were systematically and quantitatively evaluated on their ability to produce near real-time estimates under varying amount of gaps and under varying length of periods without observations. In this study, the LAI time series estimates are based on the BELMANIP 2.1 sites, an ensemble of 445 sites distributed globally selected to represent global vegetation. Following the methodology applied in Chapter 3, this study uses the same reference sites but excludes sites south of the tropic (latitude -23°), to avoid mixing of the gap-structures observed in the northern and southern hemisphere.. This study validated the suitability of the GEOV 2 algorithm for processing time series estimates from AVHRR sensors.

NEAR REAL TIME ESTIMATES OF LEAF AREA INDEX FROM AVHRR TIME SERIES DATA

S. Kandasamy, A. Verger, F. Baret, M. Weiss
INRA UMR114, EMMAH, Domaine Saint-Paul, Site Agroparc, F-84914 Avignon, France

ABSTRACT
Near real time (NRT) estimation of Leaf Area Index (LAI) products is essential for monitoring rapid surface process changes within operational systems. This study assesses the performance of three time series processing methods, the Whittaker method (WM), Gaussian Process Model (GPM) and the GEOV2 algorithm, for the NRT estimation of LAI. The methods are evaluated using time series of daily LAI derived from AVHRR observations over a selection of BELMANIP2 sites representative of seasonal patterns of global biome vegetated areas and under varying level of noise and missing data. A simulation experiment was designed for the evaluation of method predicting capabilities with special emphasis on the global and local structure of missing data in the time series. Results show that the different methods achieve similar performances when the fraction of missing data over the whole time series is lower than 65% (RMSE<0.2) or the length of gaps is smaller than 10 days (RMSE values lower than 0.35). Conversely, for fraction of gaps higher than 65% or periods of gaps longer than 10 days GEOV2 is found to be more robust than the WM and GPM methods since it benefits from the use of a climatology derived from the data. Our findings support the operational use of GEOV2 algorithm for NRT estimation of biophysical products at global scale.

Keywords— time series, LAI, vegetation, missing data, near-real time

1. INTRODUCTION

Leaf Area Index (LAI), defined as the total green leaf area per unit horizontal ground area (Chen and Black 1992), is key variable for studying the earth surface processes and for monitoring surface vegetation (Yuan et al. 2011)-(David S 1995). It controls the size of the surface-atmosphere interface and hence influences the mass-energy exchanges between the surface ecosystems and the atmosphere. LAI has been identified as one of the essential climate variables(GCOS 2010). LAI also plays a key role in the studies on soil water content (Zhang and Wegehenkel 2006), vegetation dynamics (Jonsson and Eklundh 2002), or crop yield estimation (Jia et al. 2011). In the last decades, global LAI estimates have been produced from medium spatial resolution satellite sensors including MODIS(Knyazikhin et al. 1999; Knyazikhin et al. 1998), AVHRR (Baret et al. 2011), VEGETATION (Baret et al. 2007; Weiss et al. 2007) and MERIS (Bacour et al. 2006a). Nevertheless these LAI products are affected by noise and show some discontinuous mainly due to cloud cover which limits their use.

Many methods have been used for removing noise and smoothing the time series including median smoothing (Reed et al. 1994), curve fitting (Bradley et al. 2007), low-pass filtering (Bacour et al. 2006b), Savitzky-Golay filtering (Chen et al. 2004; Verger et al. 2011), double logistic and asymmetric Gaussian function (Jonsson and Eklundh 2002; Jönsson and Eklundh 2004), wavelet decomposition (Sakamoto et al. 2005), penalized least square regression (Eilers 2003) and singular spectrum analysis (Kandasamy et al. 2012). Similarly, many attempts have been made to fill the gaps in the LAI time series estimates based on data fusion (Verger et al. 2011) and spatio-temporal filling methods (Fang et al. 2008) including statistical regression methods (Eilers 2003; Falge et al. 2001), singular spectrum analysis (Kandasamy et al. 2012) and climatology based approaches (Verger et al., 2012).

A part from the continuity and smoothness of LAI time series suited by most of users, several applications including operational systems for crop management or short term weather prediction require near real time (NRT) estimates of LAI. Due to the discontinuities in the satellite time series, methods able to predict the short-term LAI values based on the past (historical) estimates are required. NRT estimate is challenging as compared to gap

filling because the former uses only the past estimates for making predictions while the latter benefits from the availability of data on either side (past and future) of the gaps. (Jiang et al. 2010) evaluated three methods: Dynamic Harmonic Regression (DHR), Seasonal Trend Decomposition based on Loess (STL) and Seasonal Auto Regressive Integrated Moving Average (SARIMA) able to characterize the non-stationary time-series of MODIS LAI and predict future values. Their results indicate that the SARIMA model gives the best prediction, DHR produces the smoothest curve, and STL is more sensitive to noise in the data. Zhiqiang et al. (2011) proposed a real-time inversion method based on the use of Ensemble Kalman Filter and the statistical SARIMA to recursively update LAI as new MODIS observations are delivered with improvements over the MODIS LAI product in terms of smoothness and accuracy as compared to field measurements. However very little attention was paid to the presence of missing data (gaps) in the satellite time series, mostly because of cloud masking which creates irregular time steps between actual observations of the surface.

This study compares the performances of three methods to make NRT estimations of LAI: the Whittaker Method (WM) (Eilers 2003), the Gaussian Process Model (GPM) (Rasmussen and Williams 2005) and the GEOV2 (Baret et al. 2012) algorithm. This study is conducted under the premises of GEOLAND2 European program which aims to develop an operational system for environmental monitoring. According to users specifications LAI is required to be provided every 10-days at NRT with an overall accuracy of 0.5 LAI (Lacaze 2009). This study uses LAI time series estimates derived from AVHRR daily observations in GEOLAND2 (Baret et al. 2011). AVHRR time series present significant amount of noise and discontinuities that allow a thorough evaluation of the methods. Further AVHRR provides long time series of data from 1981 and its continuity is ensured through the METOP global monitoring program. Special emphasis was put on the effect of missing data on the performance of NRT estimates with attention paid to the structure of occurrence of missing values, i.e. the fraction of gaps in the time series and the length of gaps preceding the date of prediction.

2. EXPERIMENT, DATA AND METHODS

2.1. Data

The daily LAI dataset used here is derived from the AVHRR Long Term Data Record (LTDR, Version 3) (WWW2). LTDR provides global coverage at 0.05° (5.6 km at equator) sampling interval in a latitude/longitude climate modeling grid and at a daily temporal step for the 1981 – 2000 period (Vermote et al. 2009a; Vermote et al. 2009b). The LAI estimation is based on the previous work of (Verger et al. 2008) demonstrating that neural networks could be trained to consistently estimate a given product from the reflectance measured by another sensor providing that a strong link exists between the inputs (radiometric signal) and outputs (the products). This principle is applied here to mimic GEOV1/VGT LAI product (Baret et al. 2010) from historical AVHRR archive data. Further details on the generation of the LAI dataset are provided in (Baret et al. 2011; Verger et al. 2012).

2.2. Methods

2.2.1 Whittaker Method (WM)

This method proposed by Whittaker (1923) is based on a very interesting principle that consists in minimizing a cost function describing the balance between fidelity expressed as the quadratic difference between estimates and actual observations and roughness expressed as the quadratic difference between successive estimates. This balance is controlled by a smoothing parameter. The higher this value, the more is the emphasis on the smoothing but at the cost of fidelity. Finding an appropriate value of the smoothing parameter is difficult, as it depends on the data. After trial and errors the smoothing parameter was set to 35000. The smoothness is also controlled by the order of differentiation which is fixed to 2 as proposed by Eilers (2003) for time series data sampled at regular intervals but with missing observations.

2.2.2 Gaussian Process Model (GPM)

Gaussian Process is a Bayesian machine learning regression approach based on a particularly effective kernel based method for placing a prior distribution over the space of functions (Rasmussen and Williams 2005). As other machine learning algorithms, the goals of GPM are primarily to make predictions as accurately as possible and to understand the behavior of the learning algorithms. As compared to neural networks which are based on adaptive basis functions for learning hidden features, in the GPM the prediction is based on fixed basis functions without loss of generalization capacity. This gives advantages with respect to the interpretation of model

predictions and provides a well-founded framework for learning and model selection. The predicted mean which is a linear combination of observations has an associated predicted variance which depends on the distance of observation to the Gaussian kernel functions. The kernel functions can be very sophisticated because all hyper-parameters can be learned effectively by maximizing the marginal likelihood in the training set. For further details on the GPM, the interested reader is referred to Rasmussen and Williams (2005) book and the associated Matlab code (http://gaussianprocess.org/gpml/) here used. GPM has recently been demonstrated as a powerful regression tool in other remote sensing predictions problems (Pasolli et al. 2010; Verrelst et al. 2012).

2.2.3 GEOV2 NRT algorithm

The GEOV2 (Baret et al. 2012) algorithm has been recently proposed for the near real time estimation of the second version of Geoland2 biophysical products. The method is based on dedicated time series processing techniques that combine an innovative climatology fitting based approach (Verger et al. 2012) to ensure robustness and continuity in the estimates with an adaptive Savitzky-Golay filter (Verger et al. 2011) to better adapt to the local variations in the data. The former is based on a consistent climatology adjustment of the climatology to actual observations (CACAO) in each sub-season. The latter is a temporal smoothing gap filling (TSGF) approach (Verger et al. 2011) based on an adaptative Savitzky-Golay second degree polynomial fitting with a flexible and asymmetric temporal window which is defined by the existence of n=6 observations within a semi-period between 30 and 60 days at each side of the date being processed. The dissymmetry in the observations for the near real time case (observations are only available for the past period) is solved using the CACAO values. Similarly when less than 6 observations exist in the past 60 days period, the CACAO values are used for gap filling. The climatology used in CACAO is derived as the inter-annual median of daily AVHRR LAI data in the 1981-1999 period within a 30-day composition window at dekadal temporal step and at the pixel level (Verger et al. 2012).

2.3. Experiment

As in the previous study (Kandasamy et al. 2012) the approach proposed to evaluate the methods is based on two steps as sketched in Figure 1. The first one is dedicated to the preparation of reference time series over a limited number of representative sites. As a result of this step, two time series will be output: (i) a time series with no gaps and small uncertainties considered as the reference (LAI_{ref}), and (ii) a time series with no gaps with realistic uncertainties (LAI_{comp}). In the second step, time series with variable occurrence of missing data will be simulated (LAI_{sim}) from the previously LAI_{comp} time series. For the last year of the time series (year 2000) each method will be applied on this LAI_{sim} data by considering only the past observations at each date to get the corresponding NRT estimates (LAI_{NRT}). Finally, the LAI_{NRT} obtained with each of the 3 methods will be evaluated based on a range of metrics describing the accuracy of the NRT estimates as compared to LAI_{ref} as a function of the fraction of gaps (%Gap) and length of gaps (LoG).

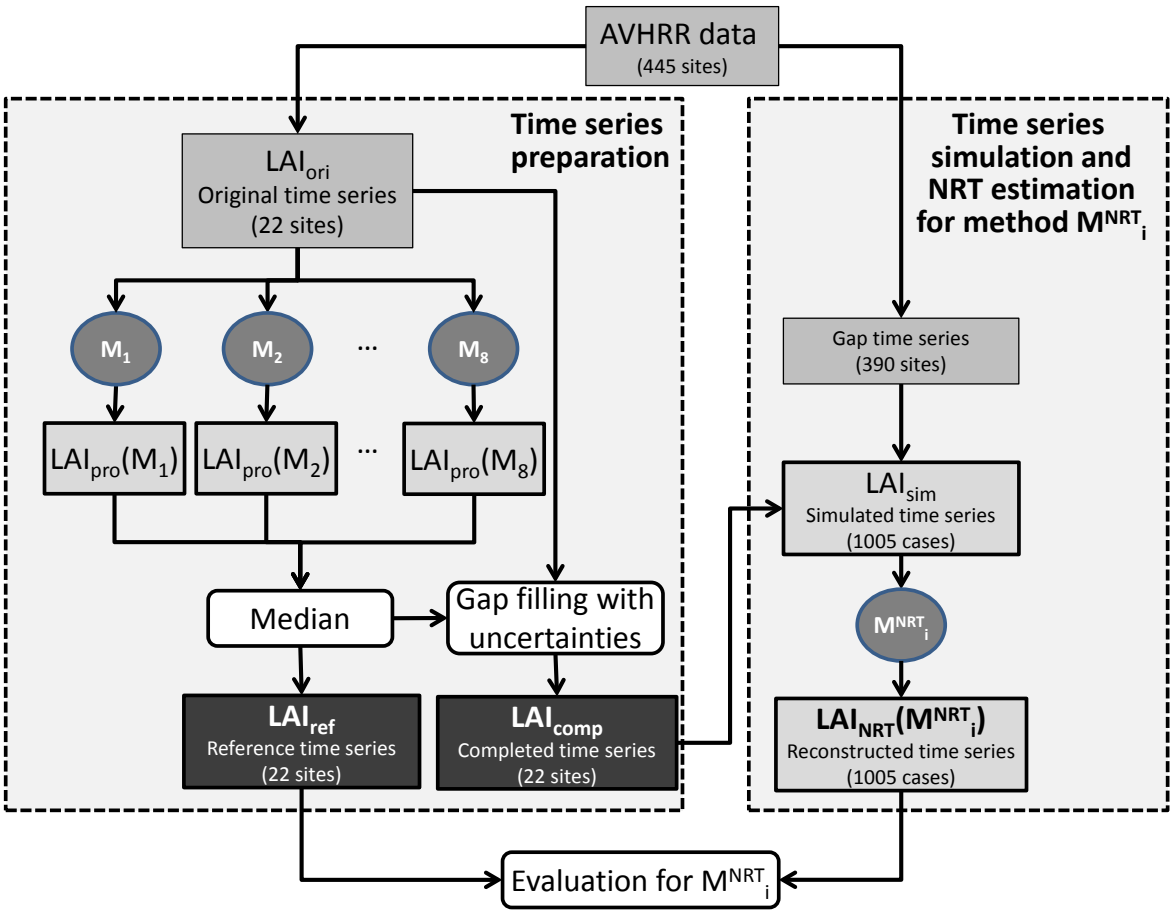

Figure 1. Flow chart describing the experiment. $M_{1\text{-}8}$ correspond to the 8 methods used to create the reference LAI_{ref} data and M^{NRT}_i to the NRT method i within the 3 investigated. $LAI_{NRT}(M^{NRT}_i)$ corresponds to the NRT estimates based on method M^{NRT}_i.

2.3.1. Generation of the reference and completed LAI time series

This study was conducted over the 445 BELMANIP2 sites identified by Baret et al. (2006) to represent the variability of vegetation types and conditions around the world. To deal with the seasonal asymmetry between northern (334 sites) and southern (111) sites and avoid simulating unrealistic gaps, only the sites having latitude higher than -23º latitude were selected which results in 412 sites. These sites were first classified according to GLOBCOVER land cover map (Defourny et al. 2009) with the original classes aggregated into 5 main classes: Shrub/Savannah/bare area (SB), Grasslands and Crops (GC), Deciduous Broad Leaf Forests (DB), Evergreen Broad Leaf Forests (EB) and Needle Leaf Forests (NF). For each of these classes, 5 BELMANIP2 sites were selected by (Kandasamy et al. 2012) in an attempt to show a large variability in seasonality while having minimum number of missing observations (Table 1). From these 25 sites, the 22 sites located northern of the -23º latitude were considered in this study. Note that forest sites have still significant occurrence of missing data since it was difficult to find sites with more clear sky conditions.

Table 1. Sites selected for the study for the 5 biomes. The site number in BELMANIP2 ensemble of 445 sites, latitude and longitude, fraction of missing data (%Gap) and number of LAI_{sim} simulated cases are indicated.

Site	Biome	Lat.(°)	Lon.(°)	%Gap	# simulated cases
136	Shrub Savannah Bare (SB)	-18.46	44.41	68	12
176		12.49	36.34	63	137
186		10.7	39.41	67	137
293		-21.54	143.83	55	12
69	Crop Grassland (GC)	38.63	-98.91	63	92
225		35.09	-1.00	56	92
338		25.997	68.52	43	92
131	Deciduous Broad leaf forests (DBF)	-11.99	16.43	83	11
146		-5.45	31.74	85	11
162		4.86	28.76	87	22
165		5.98	31.18	80	22
296		-16.45	142.62	63	11
19	Evergreen Broadleaf Forests (EBF)	-11.75	-53.34	83	27
30		-2.68	-63.65	96	27
50		17.6	-89.78	89	24
142		-4.59	23.44	86	27
320		25.54	121.25	95	24
54	Needle Leaf Forests (NLF)	28.39	-108.25	68	45
55		26.53	-106.68	64	45
62		39.49	-120.83	67	45
65		30.28	-83.85	80	45
244		43.86	-1.1	87	45

To create the reference LAI_{ref} time series, the 3 methods presented earlier and 5 additional methods described in (Kandasamy et al. 2012): Iterative Caterpillar Singular Spectrum Analysis (ICSSA) (Kandasamy et al. 2012), Empirical Mode Decomposition (EMD) (Huang et al. 1998), Low Pass Filtering (LPF) (Bacour et al. 2006b), Adaptive Savitzky-Golay Filter (SGF) (Chen et al. 2004) and Asymmetric Gaussian Function (AGF) (Jonsson and Eklundh 2002; Jönsson and Eklundh 2004) were applied to each of the 22 sites. Figure 2 illustrates the main features of the time series reconstructed by the several methods over the site 176 (Table 1). The performances of the reconstructions are very similar (within 0.05 LAI) for 6 out of the 8 methods (Table 2). Further, 4 of the 8 methods investigated are very close to the median across all methods with a RMSE (Root Mean Square Error) lower than 0.1. The median appears as a robust estimate of the expected LAI that mitigates the differences between the several methods and their possible limitations (e.g. higher RMSE for EMD, Table 2). This way the median across methods shows no bias as compared to the original data and overall performances around 0.4 in terms of RMSE (Table 2). Figure 3 shows that the median across all methods is a good approximation of the expected LAI product values. The time series made with the median across the 8 methods will therefore be considered as the reference values called LAI_{ref} in the following. This LAI_{ref} does not show any missing data since the gaps in the original time series were filled by the reconstructed values of the 8 methods. LAI_{ref} constitutes a good reference with minimal uncertainties attached to the LAI values because of the temporal smoothing coming from each method and the computation of the median across the 8 methods. A second set of time series was generated to provide realistic LAI values: the LAI_{ori} were complemented at the location of missing data by LAI_{ref} values contaminated by a noise that was randomly drawn within the distribution of residuals (LAI_{ref}-LAI_{ori}) for each site. These realistic but continuous temporal profiles with no gaps (LAI_{comp}) will be used in the second step of the approach for simulating time series with gaps.

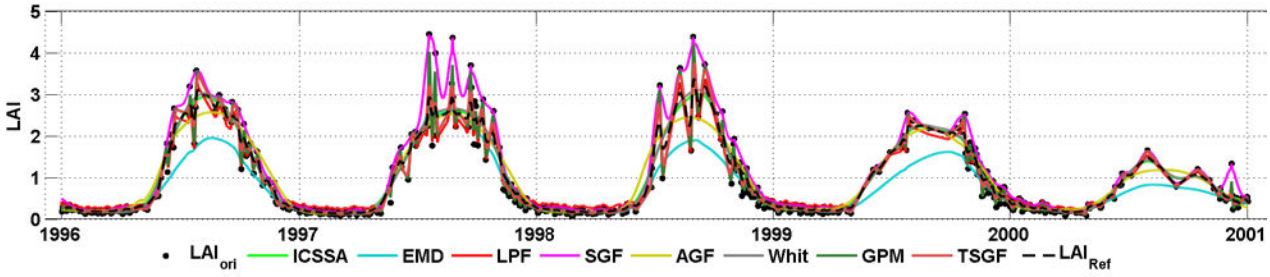

Figure 2. Illustration of the reconstruction by the 8 different methods: ICSSA (green), EMD (cyan), LPF (red), SGF (magenta), AGF (yellow), WM (gray), GPM (dark green) and TSGF (light red) over the site 176 (Table 1). Black dots correspond to LAI_{ori} at the location of available observations. The dashed black curve corresponds to LAI_{ref}.

Table 2. Performances in terms of RMSE and BIAS associated to the different methods used for generating LAI_{ref} as compared to LAI_{ori}. Statistics for LAI_{ref} resulting from the median of different methods are also indicated.

Method	RMSE	Bias
ICSSA	0.45	-0.02
EMD	0.97	0.02
LPF	0.11	0.02
SGF	0.56	0.31
AGF	0.51	-0.02
Whit	0.44	-0.02
GPM	0.18	$4x10^{-4}$
TSGF	0.32	-0.001
Median LAI_{ref}	0.40	$8x10^{-4}$

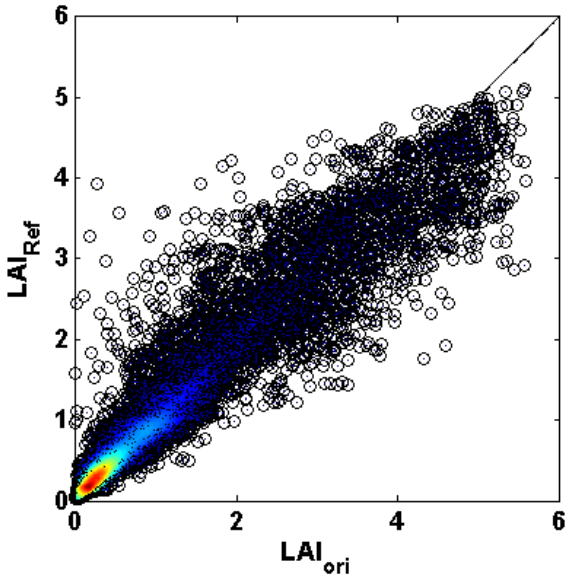

Figure 3. Comparison between the original LAI values (LAI_{ori}) and the reference LAI computed as the median of the reconstructed values (LAI_{ref}) based on the 8 methods considered over the 22 selected sites (Table 1) (N=10457; R^2=0.94; RMSE=0.40; Bias=$8x10^{-4}$).

2.3.2. Simulation of time series with gaps

In the second step, emphasis was put on the occurrence of missing data (gap). The gap structure observed over each one of the 412 sites used for the simulations was applied to the completed time series (LAI_{comp}). This allows keeping the gap structure more realistic as compared to other strategies that would consist in randomly drawing gaps. However, vegetation type and the associated climate experienced, hence the cloud occurrence and corresponding gap structure, are probably correlated. To account for such possible dependency, the gap structure applied to one of the 22 sites was selected within the gap sites belonging to the same vegetation class and hemisphere. A total of 1005 time series with realistic LAI uncertainties and gap structure were finally available (Table 1). The cases with long gaps preceding the date being estimated correspond to the time series with a higher fraction of missing data (%Gap) as shown in Fig 4. The %Gap increases exponentially with the LoG for LoG<10 days and asymptotically reaches a constant value around 85% for LoG>10 days. The distribution of LoG shows the opposite tendency with an exponential decay of frequencies with the LoG (black dotted line in Fig. 4).

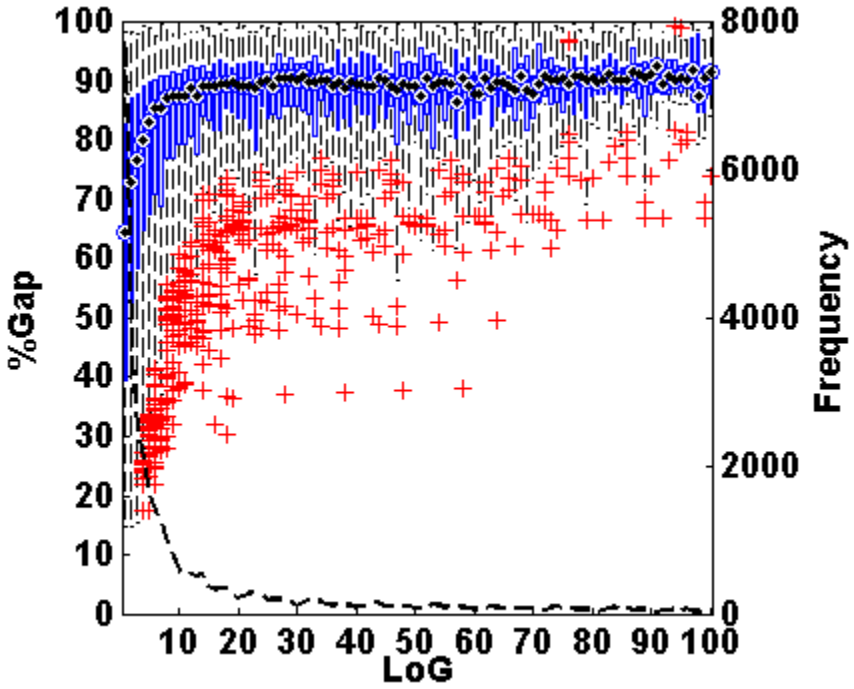

Figure 4. Box plot of the fraction of gaps (%Gap) in the AVHRR LAI time series from 1996 to the estimation date as a function of the length of gap (LoG) in days preceding the estimation date. Circles correspond to the median values. Bars indicate the 25 and 75 percentiles. Outlier values are marked by plus symbols. The dashed line corresponds to the frequency of LoG.

2.3.3. Metrics used to quantify the performances

Finally the three NRT evaluated methods will be applied to LAI_{sim} data every 10-days starting from 1-Jan-2000 by considering only the data before the date of prediction. The performances of Whit, GPM and GEOV2 methods to make NRT estimates were evaluated based on the RMSE between the LAI_{NRT} estimates and the reference LAI_{ref} values:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}\sum_{t=1}^{n_j}\left(LAI_{NRT}^{j}(t)-LAI_{ref}^{j}(t)\right)^2}{\sum_{j=1}^{N} n_j}}$$

where $LAI_{ref}^{j}(t)$ and $LAI_{NRT}^{j}(t)$ are respectively the reference and the NRT estimates for the dekadal date *t* and case *j*, *n* is the number of evaluated dates for the case *j* and *N* the number of cases considered. The performances of the three evaluated methods will be assessed as a function of %Gap and LoG.

4. RESULTS

The performance of the methods to produce NRT estimates is evaluated both globally as a function of the fraction of missing data in the time series and locally with the missing data structure in a local window preceding the estimation date.

4.1. Global assessment of NRT predictions with the fraction of missing data in the time series

The temporal evolution of NRT estimates is first qualitatively assessed and then a quantitative evaluation of their performances is conducted.

4.1.1. Temporal profiles

The temporal profiles of NRT estimates shown in Figures 5 and 6 illustrate the main features of the three evaluated methods under different levels of missing data. Visual inspection shows that:

- From low to moderate occurrence of missing data (Figures 5-6 A-B) the estimates of the three methods show similar temporal evolutions with a good agreement with the reference LAI_{ref} data.
- For very high levels of missing data (Figures 5-6 C) Whit and GPM show some limitations in terms of fidelity (closeness to LAI_{ref}). Whit provides unrealistic retrievals when most of observations are missing in Figure 5C. GPM is not able to retrieve the second peak for the double season crop shown in Figure 6C. Conversely, GEOV2 algorithm benefits from the use of the climatology filling approach and provides more robust estimates in cases with very high amount of gaps.

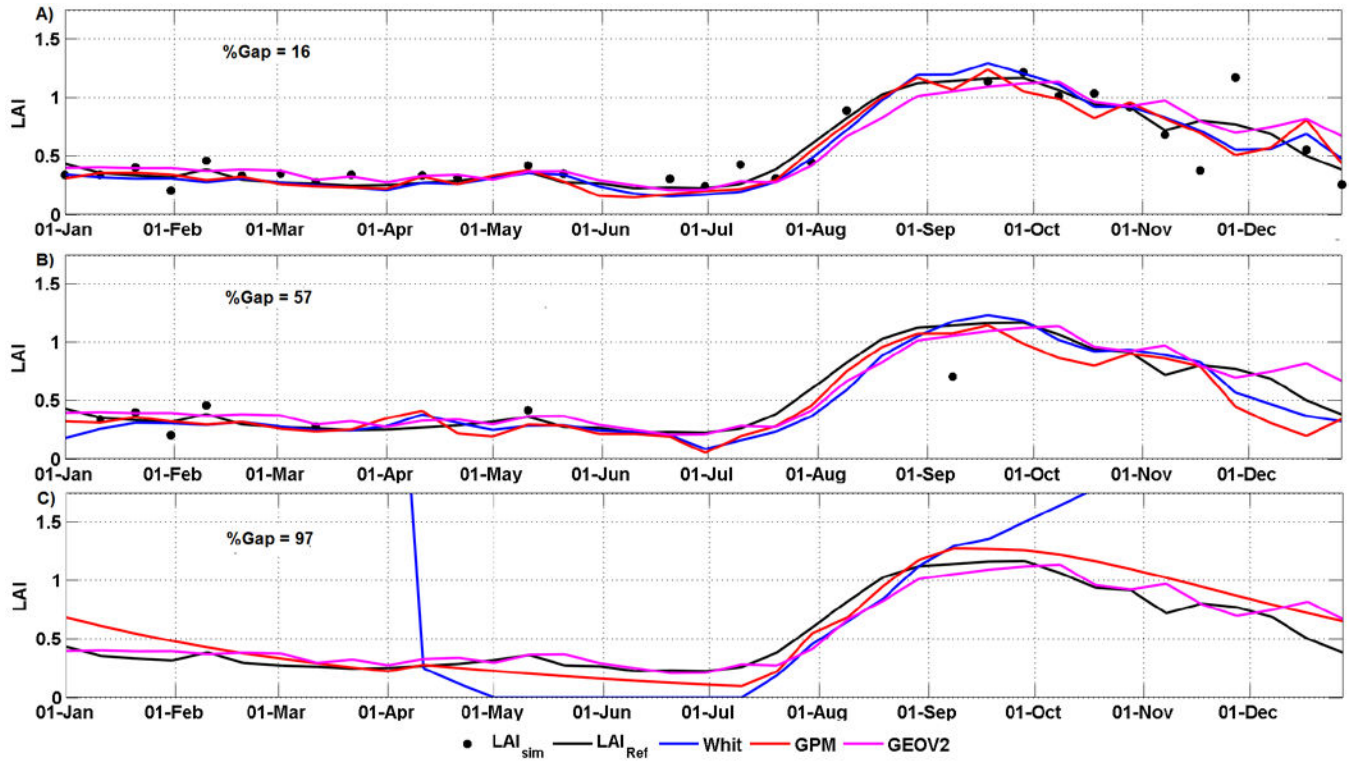

Figure 5. NRT estimates of the three methods for the shrubland site 176 (Table 1) for three levels of missing data: a) %Gap=16%, b) %Gap=57% and c) %Gap=97%. The LAI_{sim} and LAI_{ref} used respectively as input and reference data are also shown.

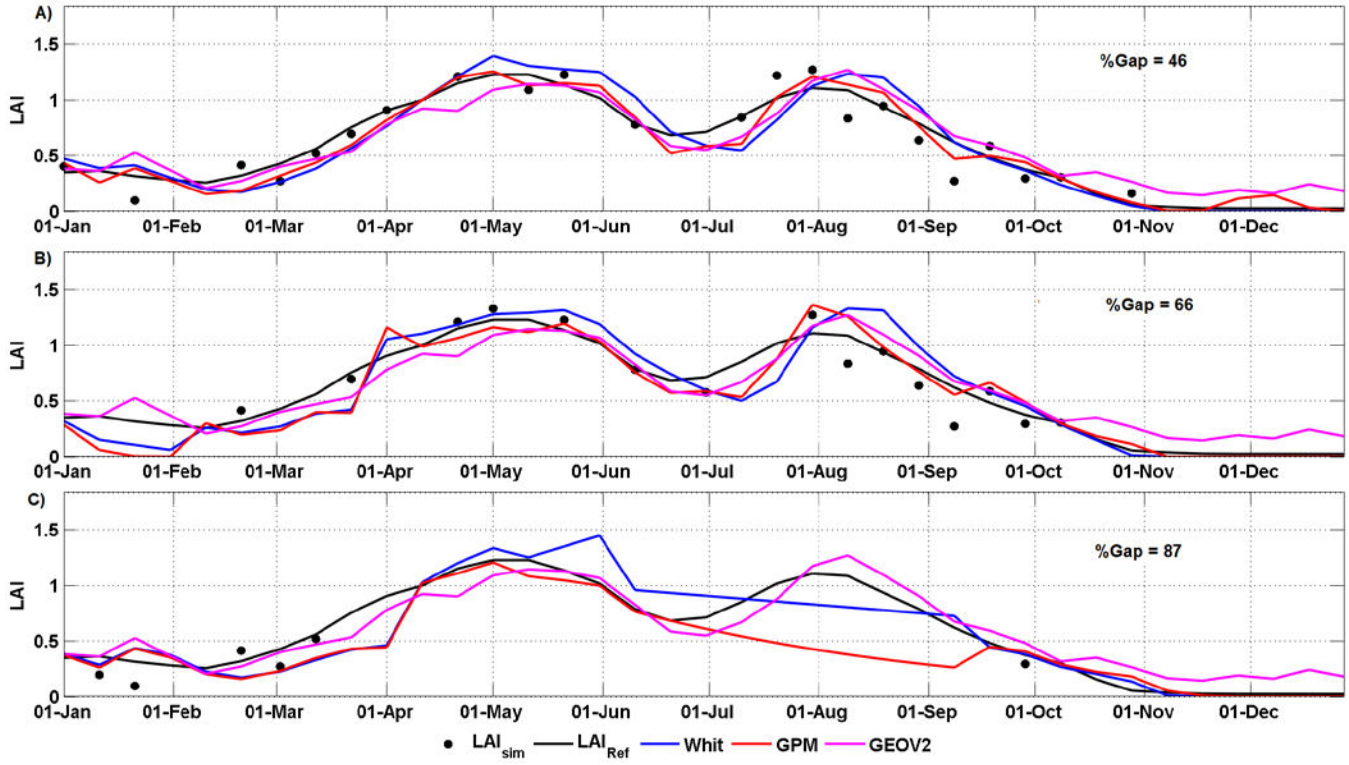

Figure 6. NRT estimates of the three methods for the cropland site 69 (Table 1) at varying fractions of missing data: a) %Gap=46, b) %Gap=66 and c) %Gap=87. The LAI_{sim} and LAI_{ref} used respectively as input and reference data are also shown.

4.1.2. Performance of NRT estimates with missing data

A quantitative assessment of the performances of the methods for NRT estimation of LAI as a function of the %Gaps is shown in Figure 7. As expected the performances evaluated through the RMSE as compared to LAI_{ref} decrease as the %Gaps increases. For %Gaps lower than 65%, all the methods perform similarly with RMSE lower than 0.2. For higher %Gaps levels an exponential increase of RMSE is observed. Whit and GPM show obvious limitations for NRT estimation when the time series show very high levels of missing data which correspond to a large number of cases as the frequency distribution indicates (green line). This confirms the previous findings in Figure 5-6C. Conversely, GEOV2 shows relatively low RMSE values within 0.4 up to %Gap of 95%.

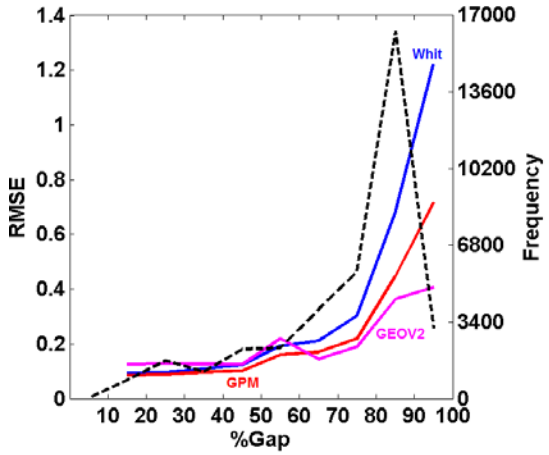

Figure 7. RMSE as a function of %Gap for the three methods. The RMSE values are computed for %Gap intervals of 10% width. The dashed line corresponds to the frequency of occurrence of missing data.

4.2. Local assessment of NRT estimates with missing data structure

A closer inspection of the influence of the local structure of missing data is here conducted. First, the influence of the LoG is evaluated. Then the effect of the number of available observations in the prediction accuracy is investigated.

4.2.1. Length of gaps

The performance of the different methods degrades exponentially with the LoG (Figure 8). For LoG lower than 10 days the three methods provide identical results with RMSE lower than 0.3. Whit and GPM perform similarly for LoG lower than 40 days. For longer gaps the RMSE of Whit continues to increase with the LoG while GPM reaches asymptotically a constant maximum value around 0.7. GEOV2 is the method that performs the best with RMSE values lower than 0.5 up to 60 days. Note that 60 days corresponds to the maximum compositing window period of TSGF (section 2.2.3). For LoG>60 days CACAO climatological values are used to fill the gaps. For LoG>80 days the RMSE of GEOV2 is similar to the RMSE of GPM and none of the methods provide reliable estimates in these extreme conditions. The noisy nature of the RMSE curve for the high LoG values show the instabilities of the NRT estimates and it could also be marginally attributed to local accidents in the dataset due to the lower number of simulation cases in these conditions (green line).

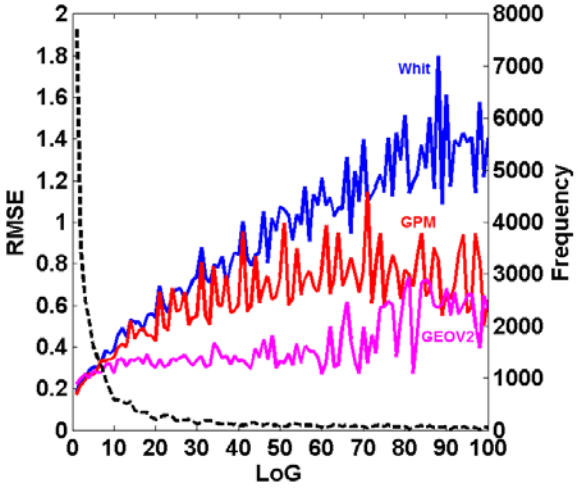

Figure 8. RMSE of the three methods as a function of LoG. The dashed line corresponds to the frequency of occurrence of LoG.

4.2.2. Number of observations in a local window

Finally, to complement the previous analysis, the performances of the methods were investigated as a function of the amount of available data in the 60-day period preceding the dekadal dates of prediction (Figure 9). A 60-day local window was selected for the assessment because it corresponds to the composition window of TSGF in GEOV2 algorithm while the other two methods (Whit and GPM) use all the available data prior to the date of the estimates. The uncertainty of the three methods decrease with the number of available observations with an exponential decay of the RMSE (Figure 9). GEOV2 is found to perform better than the other two methods when the number of observations in the 60-day window is lower than 10. When more than 10 observations are available the three methods perform similarly and reasonably well with RMSE≈ 0.2 for 20 out of the 60 potential observations. For the extreme case when any observation is available in the 60-day window, Whitt clearly performs the worst (RMSE ≈ 1.8) while GEOV2 performs the best within the expected accuracy levels required for most of LAI applications (RMSE ≈ 0.5). GPM constitutes an intermediate solution with RMSE around 0.5 when the number of observations is higher than 5.

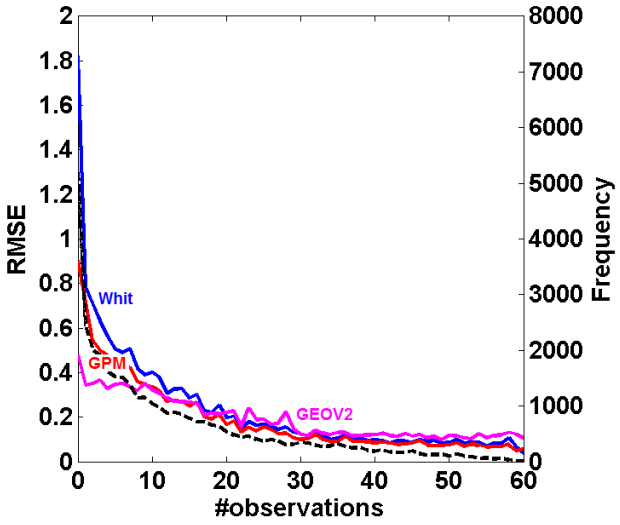

Figure 9. RMSE as a function of the number of observations in a 60-day local window preceding the dates for the NRT estimation. The dashed line corresponds to the frequency of occurrence of missing data.

5. DISCUSSION AND CONCLUSION

This study compares three methods, Whit, GPM and GEOV2, for the near real time estimation of LAI. Whit and the GPM use all the past observations in the time series to make a NRT prediction but without any auxiliary information to constrain the temporal evolution. GEOV2, on the other hand, is a combination of temporal strategies based on Savitzky-Golay and climatological fitting approaches allowing to adapt to local variations in the data when available and allowing at the same time to produce robust estimates in case of missing observations.

Methods were tested with actual AVHRR daily LAI products characterized by a significant level of noise and missing data. A simulation experiment was designed to apply the methods over a representative sample of seasonal patterns and distribution of missing observations. Special attention was paid on the influence of global and local structure of missing data on the performances of NRT estimates. This methodology was expected to improve the realism of the NRT estimation assessment.

Results show that different methods perform similarly and reasonably well (RMSE lower than 0.2) when the fraction of missing data is lower than 65%. For higher amounts of missing data in the time series, Whit and GPM show some artifacts in their NRT estimates while GEOV2 provides robust estimates with an overall performance better than 0.4 in terms of RMSE which fulfills user requirements. The local structure of missing data appears also to be a critical factor on the overall performance of NRT estimates. For periods of missing data lower than 10 days, which correspond to the most common situations, all the methods seem to perform similarly with RMSE values lower than 0.4. The performance of GPM and Whit degrade rapidly with the length of gaps while GEOV2 appears to be more robust with RMSE values lower than 0.5 up to 60 days with missing data because the used climatology plays a major regularization role. Similarly for the situations with less than 10 observations in a local window of 60 days the GEOV2 appears more reliable than Whit and GPM approaches demonstrating the potential of using a climatology fitting approach in cases of significant amount of gaps.

Global and local structures of gaps are linked: time series having long gaps mostly correspond to high percentage of missing data (LoG>20 days correspond to %Gap>65%, Figure 10). To better differentiate the effect of local and global structures of gaps on the performances of the methods, the impact of %Gap was analyzed for three different LoG scenarios: time series with LoG varying from 0-20, 20-40 and 40-60 days (Figure 10). The overall performances for a particular method are determined both by LoG and %Gap. However, LoG appears to be the dominant factor in determining the performances. For LoG≈10 days, no significant differences exist between the three methods with a gradual degradation in their performances as a function of %Gaps (RMSE varying from 0.1 to 0.4). For the time series having LoG>20 days, GEOV2 seems to more robust than Whit and GPM which show a

sharp degradation with %Gap with RMSE up to 0.9 and 1.1 for GPM and Whit, respectively, as compared to RMSE<0.4 for GEOV2.

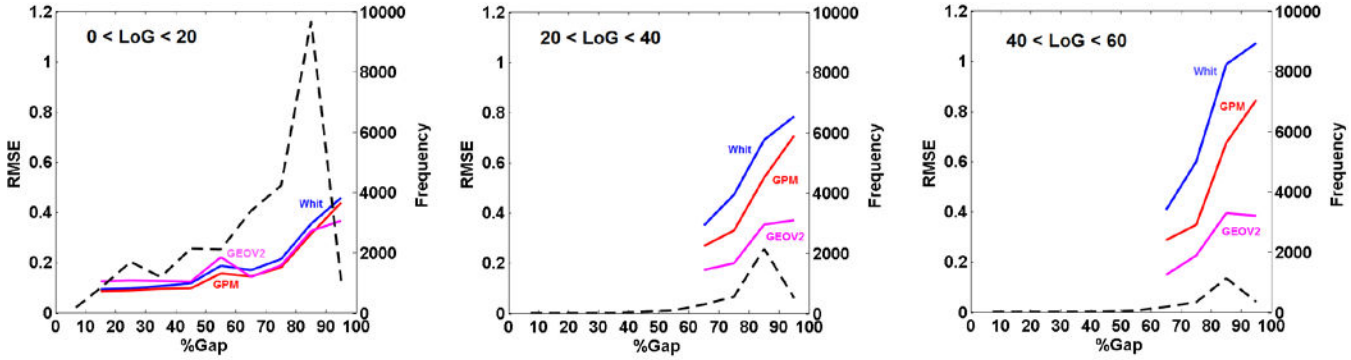

Figure 10. Performances of the methods as a function of %Gap for different LoG. The 'dashed' line corresponds to the frequency of occurrence of missing data.

In terms of the required computer resources, GEOV2 is 5 times more demanding than Whit but 200 times faster than GPM. GEOV2 will be implemented operationally for the generation of near real time biophysical products at the global scale every 10 days from VEGETATION data. GEOV2 algorithm will be further adapted to other sensors including AVHRR/METOP, PROBA-V, Sentinel-3 for time series continuity.

Acknowledgment: This research has received funding from the Geoland2 European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n°218795. A. Verger was funded by the VALi+d postdoctoral program (FUSAT, GV-20100270). The authors would like to thank E. Vermote for making the LTDR dataset available.

REFERENCES

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

Bacour, C., Bréon, F.-M., & Maignan, F. (2006b). Normalization of the directional effects in NOAA–AVHRR reflectance measurements for an improved monitoring of vegetation cycles. *Remote Sensing of Environment, 102*, 402-413

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

Baret, F., Morissette, J., Fernandes, R., Champeaux, J.L., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigue, S., & Nickeson, J. (2006). Evaluation of the representativeness of networks of sites for the global validation and inter-comparison of land biophysical products. Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing, 44*, 1794-1803

Baret, F., Weiss, M., Verger, A., & Kandasamy, S. (2011). BioPar Methods Compendium - LAI, FAPAR and FCOVER from LTDR AVHRR series. In, *Report for EC contract FP-7-218795. Available at* http://www.geoland2.eu/portal/documents/CA80C881.html (p. 46). Avignon: INRA-EMMAH

Baret, F., Weiss, M., Verger, A., & KANDASAMY, S. (2012). BioPar Methods Compendium: LAI, FAPAR and FCOVER from VEGETATION P products series. In (pp. 1-47): INRA

Bradley, B.A., Jacob, R.W., Hermance, J.F., & Mustard, J.F. (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment, 106*, 137-145

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment, 91*, 332-344

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

David S, S. (1995). Terrestrial biogeochemical cycles: Global estimates with remote sensing. *Remote Sensing of Environment, 51*, 49-56

Defourny, P., Bicheron, P., Brockmann, C., Bontemps, S., Van Bogaert, E., Vancutsem, C., Pekel, J.F., Huc, M., Henry, C., Ranera, F., Achard, F., di Gregorio, A., Herold, M., Leroy, M., & Arino, O. (2009). The first 300 m global land cover map for 2005 using ENVISAT MERIS time series: a product of the GlobCover system,. In, *Proceedings of the 33rd International Symposium on Rememote Sensing of Environment*. Stresa (Italy)

Eilers, P.H.C. (2003). A Perfect Smoother. *Analytical Chemistry, 75*, 3631

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., & Wofsy, S. (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology, 107*, 43-69

Fang, H., Liang, S., Townshend, J.R., & Dickinson, R.E. (2008). Spatially and temporally continuous LAI data sets based on an integrated filtering method: Examples from North America. *Remote Sensing of Environment, 112*, 75

GCOS (2010). Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. In (pp. This draft 2010 Update of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC2011 was compiled based on an expert meeting in early February 2009, and subsequent iterations by a task team under leadership of Paul Mason (former Chair of the GCOS Steering Committee) with participation by GCOS panel chairs and GCOS, GOOS and GTOS secretariat staff.): World Meteorological Organization

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., & Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454*, 903-995

Jia, K., Wu, B., Tian, Y., Li, Q., & Du, X. (2011). An effective biophysical indicator for opium yield estimation. *Computers and Electronics in Agriculture, 75*, 272-277

Jiang, B., Liang, S., Wang, J., & Xiao, Z. (2010). Modeling MODIS LAI time series using three statistical methods. *Remote Sensing of Environment, 114*, 1432-1444

Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on, 40*, 1824-1832

Jönsson, P., & Eklundh, L. (2004). TIMESAT: a program for analyzing time-series of satellite sensor data. *Computers & geosciences, 30*, 833-845

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2012). A comparison of methods for smoothing and gap filling time series of remote sensing observations. Application to MODIS LAI products. *Submitted to Biogeosciences*
Knyazikhin, Y., Glassy, J., Privette, J.L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morissette, J.T., Votava, P., Myneni, R.B., Nemani, R.R., & Running, S.W. (1999). MODIS Leaf area index (LAi) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document. In (p. http://eospso.gsfc.nasa.gov/atbd/modistables.html)

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., & Running, S.W. (1998). Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research, 103*, 32257 – 32275

Lacaze, R. (2009). GEOLAND 2: Towards an Operational GMES Land Monitoring Core Service - BioPar User Requirements. In (pp. 1-23): GEOLAND 2

Pasolli, L., Melgani, F., & Blanzieri, E. (2010). Gaussian Process Regression for Estimating Chlorophyll Concentration in Subsurface Waters From Remote Sensing Data. *Geoscience and Remote Sensing Letters, IEEE, 7*, 464-468

Rasmussen, C., & Williams, C. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press

Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., & Ohlen, D.O. (1994). Measuring Phenological Variability from Satellite Imagery. *Journal of Vegetation Science, 5*, 703-714

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment, 96*, 366-374

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

Verger, A., Baret , F., & Weiss, M. (2011). A multisensor fusion approach to improve LAI time series. *Remote Sensing of Enviroment, 115*, 2460-2470

Verger, A., Baret , F., Weiss, M., Kandasamy, S., & Vermote, E. (2012). The CACAO method for smoothing, gap filling and characterizing seasonal anomalies in satellite time series. *IEEE transactions on Geoscience and Remote Sensing,* DOI 10.1109/TGRS.2012.2228653

Vermote, E., Justice, C., & Bréon, F.M. (2009a). Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 898-908

Vermote, E., Justice, C., Csiszar, I., Eidenshink, J., Myneni, R., Baret, F., Masuoka, E., & Wolfe, R. (2009b). A Terrestrial Surface Climate Data Record for Global Change Studies. In, *American Geophysical Union Fall Meeting*. San-Francisco (USA)

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

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*, 317-331

Whittaker, E.T. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc., 41*, 63-75

Yuan, H., Dai, Y., Xiao, Z., Ji, D., & Shangguan, W. (2011). Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. *Remote Sensing of Environment, 115*, 1171

Zhang, Y., & Wegehenkel, M. (2006). Integration of MODIS data into a simple model for the spatial distributed simulation of soil water content and evapotranspiration. *Remote Sensing of Environment, 104*, 393-408

5 Conclusions and Possible Future Work

This thesis was conducted under the premises of GEOLAND 2, which aims to produce operational biophysical variable products in near-real time through its 'BioPar' services. This thesis focuses on the nature of the LAI retrievals estimated through remote sensing observations, followed by improving the quality (continuity and consistency) of the LAI time series estimates and the estimation of LAI in near-real time from past estimates, in case of unavailable observations. Towards this objective, the thesis is organized into three chapters each dealing with an area of focus. In the process, the thesis had utilized different Canopy Radiative Transfer Models, LAI time series gap-filling/smoothing methods and methods for near-real time estimations of LAI. The studies on the LAI time series processing had been conducted using the LAI time series estimates from MODIS and AVHRR sensors. This chapter summarizes the observations and findings made in the three chapters that makes this thesis and identifies potential future work in the subject

5.1.1 Effective Nature of LAI Estimations from Remote Sensing observations

The nature of retrieval and the dependency of the retrieval process on the uncertainty envelope over the set the potential solutions for an observed reflectance was studied. Towards this objective a LUT based approached is adopted, using 3 different radiative transfer models. The RTMs, namely SAIL, GEOSAIL and FLIGHT models, differ in their assumptions of the canopies and in their complexities. The SAIL model assumes a random foliage distribution while GEOSAIL and FLIGHT allows clumping with increasing details on the placement and shape of the crowns. The FLIGHT is based on monte-carlo simulations and hence is computationally expensive. This issue is partly addressed by exploiting the spectral-invariance property of the radiative transfer equation (RTE). A Gaussian noise model with multiplicative and additive noise levels of 2% and 1% respectively was used. The respective noise levels were used for both the band dependent and independent noise.

From the study, irrespective of the RTM used, optimal retrieval performances were obtained when the uncertainty envelope was close to the variation introduced by the noise in the observed reflectances, i.e., for a confidence interval of 1. The retrieval performance is generally found to be better when the canopy assumptions used in the LUT is consistent with that of the test cases. Also, the retrieval performances degrade with the increase in the complexity of the models, which may be explained as due to the ambiguities associated with the additional parameters in the complex models. The estimated LAI from the radiometric observations were the closest to the effective LAI, as derived from gap fraction measurements, but with some noisy estimates (under- and

over- estimations), particularly for higher LAI values. It is also found that an effective LAI estimated using an LUT based on canopies with random foliage distribution is better estimated than with the LUT representing clumped canopies. The effective LAI, however, usually underestimates the actual LAI of the clumped canopies. This study, also found that the multi-view observations are unlikely to greatly increase the accuracy and precision of the estimates. However, use of multi-view configurations may require a denser LUT for acceptable rate of successful LAI estimations.

This study was based on the Gaussian noise/uncertainty model for studying the dependency of the LAI retrieval performance on the uncertainty envelope over potential solutions. Possible relation between the characteristics of the noise and the uncertainty model, used in the LAI retrieval performance, may exist and may affect the retrieval performances. Similarly, the average leaf angle distribution and the leaf inclination distribution function is likely to influence the estimation performance of a given view direction. However, these aspects of the retrieval process are not conducted as part of this study and would be the focus of future studies.

5.1.2 Improving the quality of MODIS time series estimates

High-quality, continuous and consistent, LAI products are vital for the retrieval of phenological parameters, study land cover changes, global carbon fluxes, etc. However,the LAI retrievals from the satellite observations are usually characterized by high levels of discontinuities and noise, mainly due to cloud- or snow- cover, instabilities and assumptions in the retrieval algorithms, incorrect atmospheric or directional corrections and sensor failures. Such inconsistencies and discontinuities greatly limit the usefulness of the LAI products in many applications. Consequently many studies have been devoted to the smoothing for the time series estimates of LAI and to estimate the missing estimates. However many of these studies were made on NDVI rather than the LAI estimates, themselves. Further in the absence of a thorough evaluation of the potentials and limitations of the different competing methods on a global scale, it becomes necessary to make a systematic and thorough evaluation of the competing methods.

In this study, eight different time-series processing methods were compared using the MODIS 8-day LAI products. The eights methods are: Iterative Singular Spectrum Analysis (ICSSA), Empirical mode decomposition (EMD), low-pass filtering (LPF) based on Fourier decomposition, Whittaker smoother (Whit), Adaptive Savitzky-Golay filter (SGF), temporal smoothing and gap-filling method (TSGF), asymmetric Gaussian fitting (AGF) and climatology (clim). From this study, the EMD, AGF and LPF methods are found to provide estimates in less than 50% of situations, when the amount of gaps increases beyond 20%. ICSSA, Whit and SGF methods were able to provide estimations in most of the situations for all amounts of gaps. However, an examination of the RMSE of the estimates reveals that their estimates beyond 50% of gaps may not be

reliable. The TSGF, however, is more balanced. It is successful in estimating the missing observation in more than 50% of the situations for up to 60% of gaps. It also provides the smoothest temporal profiles for cases with more 30% of gaps. The TSGF is also found to better preserve the phenology parameters up to 60% of gaps.

TSGF method is a hybrid method using a polynomial filter and linear interpolation in a temporally adaptive window. The AGF and SGF, though are local methods, lack the flexibility of the TSGF method resulting in lower performances. The global methods are particularly found to be not suitable for processing LAI time estimates with many gaps like that of MODIS. Hence, it could be concluded that temporal methods, adapted for the characteristics of the data being processed are better suitable than general methods. The TSGF methods developed for the GEOLAND 2 project is chosen as the best among the selected methods.

It may be possible to improve the performance of the ICSSA and Whit methods by adapting a hybrid approach similar to that of the TSGF method (Verger et al., 2012), which may equally be possible with other methods as well. It would be interesting to evolve mechanisms that could automatically choose the appropriate processing method based on the characteristics of the data, which may be of use in many other applications in addition to remote sensing. Realization of computationally efficient fusion methodologies, based on multi-resolution approaches like EMD, for the estimation of long time data records from multiple sensors is expected to be an important aspect of the future studies. Finally, the LAI temporal evolution as a function of other climate and environmental factors in a multi-resolution approach may provide a better understanding of the earth processes and their interactions at different spatial scales.

5.1.3 Near Real-Time Estimation of LAI from AVHRR daily estimates of LAI

Near Real-Time estimation (NRT) (10 days) of LAI is of interest to GEOLAND 2 for operational purposes. In this chapter, the potentials and limitations of three competing methods were studied to provide near real-time estimates of LAI from past estimates. The methods identified for the study were, the Gaussian process model (GPM) regression, the Whittaker method (Whit) and the GEOV2 algorithm for processing VEGETATION time series estimates (GEOV2). In this study, the ability of the different methods to faithfully provide LAI estimates every 10-day as a function of the amount of available estimates and the temporal gap between the date of prediction and the last available estimate it studied. The GEOV2 method is a combination of TSGF and CACAO method, developed for the GEOLAND 2 VEGETATION NRT products.

The study found the different methods to have similar performance for gap fraction up to 65%. After which, the performance of the Whit method is found to degrade rapidly with increase in the amount of gaps followed by the Gaussian Process model (GPM)

regression and the GEOV2 algorithm. The GEOV2 is found to have a RMSE of less than 0.4 for gap fraction up to 95% making it an ideal choice for processing the AVHRR data characterized by significant noise and discontinuities. Similarly, for gap lengths of less than 10 days between the last available estimate and the next date of prediction, all the methods are found to have similar performances. However, for gap lengths greater than 20 days, the GEOV2 algorithm is found to better perform than the GPM and the Whit methods, which may be attributed to the use of climatology as background information in this method. GEOV2 is a local method that processes the data in a temporal window of 60 days. The RMSE of the estimates is found to exponentially decrease with increase in the number of observations in this window. Again, the methods are found to have similar performances when more than 5 observations are available in this window, with the GEOV2 algorithm clearly outperforming the other two methods for lower number of observations.

Temporal compositing is used in many algorithms to improve the quality of the time series estimates (Ex. MODIS C5 LAI Products). Temporal compositing is also used in the GEOV algorithms to produce high-quality estimates of biophysical variables. GEOV1 employs compositing at the level of reflectance while in GEOV2 it is applied at the level of products (as a final stage). In GEOV2, compositing was used in the final stage to avoid loss of information due to compositing. However, the effect of this loss of information due to compositing on the near-future predictions has not been quantified / evaluated. Further, the information change in the time series estimates may also vary based on the compositing algorithm used. These evaluations are expected to form the focus of future studies in this subject.

References

Anderson, M. C., Norman, J. M., Kustas, W. P., Houborg, R., Starks, P. J., and Agam, N.: A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales, Remote Sensing of Environment, 112, 4227-4241, 2008.

Atkinson, P. M., Jeganathan, C., Dash, J., and Atzberger, C.: Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology, Remote Sensing of Environment, 123, 400-417, 10.1016/j.rse.2012.04.001, 2012.
Azzali, S., and Menenti, M.: Mapping isogrowth zones on continental scale using temporal Fourier analysis of AVHRR-NDVI data, International Journal of Applied Earth Observation and Geoinformation, 1, 9-20, 1999.

Bacour, C.: Contribtion a la determination des parametres biophysiques des couverts vegetaux par inversion de modeles de reflectance: analyse de sensibilitie et configurations optimales, PhD, Universite Paris 7 - Denis Diderot,, Paris (France), 2001.

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

Bacour, C., Bréon, F.-M., and Maignan, F.: Normalization of the directional effects in NOAA–AVHRR reflectance measurements for an improved monitoring of vegetation cycles, Remote Sensing of Environment, 102, 402-413, 2006b.

Baret, F., and Guyot, G.: Potentials and limits of vegetation indices for LAI and APAR assessment, Remote Sensing of Environment, 35, 161-173, 1991.

Baret, F., Morissette, J. T., Fernandes, R. A., Champeaux, J. L., Myneni, R. B., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., and Nickeson, J. E.: Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: proposition of the CEOS-BELMANIP, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1794-1803, 2006.

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

Baret, F., and Buis, S.: Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems, in: Advances in Land Remote Sensing, edited by: Liang, S., Springer Netherlands, 173-201, 2008.

Baret, F., Verger, A., Weiss, M., and Kandasamy, S.: Deriving GEOV1 LAI, FAPAR and FCOVER products from AVHRR LTDR. Algorithm Theoretical Based Document, INRA, Avignon, Draft Version, 30, 2011a.

BioPar Methods Compendium: LAI, FAPAR and FCOVER from LTDR AVHRR Series: http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP038-ATBD_VegetationVariables_AVHRRLTDR_INRA-I1.20.pdf, 2011b.

Barr, A. G., Black, T. A., Hogg, E. H., Kljun, N., Morgenstern, K., and Nesic, Z.: Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, Agricultural and Forest Meteorology, 126, 237-255, 2004.

Beck, P. S. A., Atzberger, C., Høgda, K. A., Johansen, B., and Skidmore, A. K.: Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI, Remote Sensing of Environment, 100, 321-334, 2006.

Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C.: A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data, Remote Sensing of Environment, 114, 1312-1323, 2010.

Bicheron, P., and Leroy, M.: A Method of Biophysical Parameter Retrieval at Global Scale by Inversion of a Vegetation Reflectance Model, Remote Sensing of Environment, 67, 251-266, 1999.

Boles, S. H., Xiao, X., Liu, J., Zhang, Q., Munkhtuya, S., Chen, S., and Ojima, D.: Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data, Remote Sensing of Environment, 90, 477-489, 2004.

Borak, J. S., and Jasinski, M. F.: Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States, Agricultural and Forest Meteorology, 149, 332, 2009.

Botkin, D. B.: Remote Sensing of the biosphere, National Research Council, New York, 1986.

Boulet, G., Chehbouni, A., Braud, I., Vauclin, M., Haverkamp, R., and Zammit, C.: A simple water and energy balance model designed for regionalization and remote sensing data utilization, Agricultural and Forest Meteorology, 105, 117-132, 2000.

Box, G., Jenkins, G., and Reinsel, G.: Time Series Analysis: Forecasting & Control (4th Edition), Prentice Hall, 2008.

Bradley, B. A., Jacob, R. W., Hermance, J. F., and Mustard, J. F.: A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data, Remote Sensing of Environment, 106, 137-145, 2007.

Buermann, W., Dong, J., Zeng, X., Myneni, R. B., and Dickinson, R. E.: Evaluation of the utility of satellite based vegetation leaf area index data for climate simulations, Journal of Climate, 14, 2001.

Camacho, F., Baret, F., Cernicharo, J., Lacaze, R., and Weiss, M.: Quality assessment of the first version of Geoland-2 biophysical variables produced at global scale, Third International Symposium on Recent Advances in Quantitative Remote Sensing, Torrent (Spain), 2010.

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L.: A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter, Remote Sensing of Environment, 91, 332-344, 2004.

Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., and Jin, S.: A simple and effective method for filling gaps in Landsat ETM+ SLC-off images, Remote Sensing of Environment, 115, 1053-1064, 2011.

Chen, J. M., and Black, T. A.: Defining leaf area index for non-flat leaves, Plant, Cell & Environment, 15, 421-429, 10.1111/j.1365-3040.1992.tb00992.x, 1992.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymov, J. A., Swift, E., Van der Sanden, J., and Pellika, P. K. E.: Derivation and validation of Canada wide coarse resolution leaf area index maps using high resolution satellite imagery and ground masurements, Remote Sensing of Environment, 80, 184 %! Derivation and validation of Canada wide coarse resolution leaf area index maps using high resolution satellite imagery and ground masurements, 2002a.

Chen, J. M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S. G., White, H. P., Hall, R. J., Peddle, D. R., King, D. J., Trofymow, J. A., Swift, E., Van der Sanden, J., and Pellikka, P. K. E.: 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, 165-184, 2002b.

Chen, J. M., Liu, J., Leblanc, S. G., Lacaze, R., and Roujean, J.-L.: Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption, Remote Sensing of Environment, 84, 516-525, 2003.

Chen, J. M., Chen, X., Ju, W., and Geng, X.: Distributed hydrological model for mapping evapotranspiration using remote sensing inputs, Journal of Hydrology, 305, 15-39, 2005.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion), Journal of Official Statistics, 6, 3--73, 1990.

Combal, B., Baret, F., and Weiss, M.: Improving canopy variables estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies, Agronomie, 22, 205-215, 2002.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., and Wang, L.: Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem, Remote Sensing of Environment, 84, 1-15, 2003.

Consoli, S., D'Urso, G., and Toscano, A.: Remote sensing to estimate ET-fluxes and the performance of an irrigation district in southern Italy, Agricultural Water Management, 81, 295-314, 2006.

Coops, N. C., Wulder, M. A., and Iwanicka, D.: Large area monitoring with a MODIS-based Disturbance Index (DI) sensitive to annual and seasonal variations, Remote Sensing of Environment, 113, 1261 %! Large area monitoring with a MODIS-based Disturbance Index (DI) sensitive to annual and seasonal variations, 2009.

De Kauwe, M. G., Disney, M. I., Quaife, T., Lewis, P., and Williams, M.: An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest, Remote Sensing of Environment, 115, 767, 2011.

De Pury, D. G. G., and Farquhar, G. D.: Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models, Plant, Cell & Environment, 20, 537-557, 10.1111/j.1365-3040.1997.00094.x, 1997.

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

Dong, H., Wenze, Y., Tan, B., Rautiainen, M., Ping, Z., Jiannan, H., Shabanov, N. V., Linder, S., Knyazikhin, Y., and Myneni, R. B.: The importance of measurement errors for deriving accurate reference leaf area index maps for validation of moderate-resolution satellite LAI products, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1866-1871, 10.1109/tgrs.2006.876025, 2006.

Eilers, P. H. C.: A Perfect Smoother, Analytical Chemistry, 75, 3636, 2003.

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

Feng, D., Chen, J. M., Plummer, S., Mingzhen, C., and Pisek, J.: Algorithm for global leaf area index retrieval using satellite imagery, Geoscience and Remote Sensing, IEEE Transactions on, 44, 2219-2229, 10.1109/tgrs.2006.872100, 2006.

Feng, G., Morisette, J. T., Wolfe, R. E., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Bin, T., and Nightingale, J.: An Algorithm to Produce Temporally and Spatially Continuous MODIS-LAI Time Series, Geoscience and Remote Sensing Letters, IEEE, 5, 60-64, 10.1109/lgrs.2007.907971, 2008.

Fernandes, R., and G. Leblanc, S.: 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, 303-316, 2005.

Ganguly, S., Samanta, A., Schull, M. A., Shabanov, N. V., Milesi, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Generating vegetation leaf area index Earth system data record from multiple sensors. Part 2: Implementation, analysis and validation, Remote Sensing of Environment, 112, 4318-4332, 2008a.

Ganguly, S., Schull, M. A., Samanta, A., Shabanov, N. V., Milesi, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Generating vegetation leaf area index earth system data record from multiple sensors. Part 1: Theory, Remote Sensing of Environment, 112, 4333-4343, 2008b.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R., and Yang, W.: Validation and Intercomparison of Global Leaf Area Index Products Derived From Remote Sensing Data, Journal of Geophysical Research, 113, doi:10.1029/2007JG000635, 2008.

Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC: http://www.wmo.int/pages/prog/gcos/documents/GCOSIP-10_DRAFTv1.0_131109.pdf, 2010.

About GCOS: http://www.wmo.int/pages/prog/gcos/index.php?name=AboutGCOS, 2012.

About GEO http://www.earthobservations.org/about_geo.shtml, access: 1-Oct, 2012.

GEOLAND 2: Project Background: http://www.gmes-geoland.info/project-background.html, 2011a.

GEOLAND 2: Project Structure: http://www.gmes-geoland.info/project-background/project-structure.html, access: 1-oct, 2011b.

Gitelson, A. A., Vina, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B.: Remote estimation of leaf area index and green leaf biomass in maize canopies, Geophys. Res. Lett., 30, 1248, 10.1029/2002gl016450, 2003.

About GMES: http://ec.europa.eu/enterprise/policies/space/gmes/, 2012.

Gobron, N., Pinty, B., Verstraete, M. M., and Widlowski, J. L.: Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications,

Geoscience and Remote Sensing, IEEE Transactions on, 38, 2489-2505, 10.1109/36.885197, 2000.

Goel, N. S., and Deering, D. W.: Evaluation of a Canopy Reflectance Model for LAI Estimation through Its Inversion, Geoscience and Remote Sensing, IEEE Transactions on, GE-23, 674-684, 10.1109/tgrs.1985.289386, 1985.

González-Sanpedro, M. C., Le Toan, T., Moreno, J., Kergoat, L., and Rubio, E.: Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data, Remote Sensing of Environment, 112, 810-824, 2008.

Global Terrestrial Observing System: Activities: http://www.fao.org/gtos/activities.html, 2007.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., and Carroll, M.: Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data, Remote Sensing of Environment, 83, 303, 2002a.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., and Carroll, M.: Towards an operational MODIS continuous field percent tree cover algorithm: examples using AVHRR and MODIS data, Remote Sensing of Enviroment, 83, 319 %! Towards an operational MODIS continuous field percent tree cover algorithm: examples using AVHRR and MODIS data, 2002b.

Heiskanen, J., and Kivinen, S.: Assessment of multispectral, -temporal and -angular MODIS data for tree cover mapping in the tundra–taiga transition zone, Remote Sensing of Environment, 112, 2367-2380, 2008.

Hermance, J. F., Jacob, R. W., Bradley, B. A., and Mustard, J. F.: Extracting Phenological Signals From Multiyear AVHRR NDVI Time Series: Framework for Applying High-Order Annual Splines With Roughness Damping, Geoscience and Remote Sensing, IEEE Transactions on, 45, 3264-3276, 2007.

Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., Masek, J. G., and Stenhouse, G.: Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model, Remote Sensing of Environment, 113, 1988-1999, 10.1016/j.rse.2009.05.011, 2009.

Hird, J. N., and McDermid, G. J.: Noise reduction of NDVI time series: An empirical comparison of selected techniques, Remote Sensing of Environment, 113, 248-258, 10.1016/j.rse.2008.09.003, 2009.

Holben, B. N.: Characteristics of maximum-value composite images from temporal AVHRR data, International Journal of Remote Sensing, 7, 1417-1434, 10.1080/01431168608948945, 1986.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454, 903-995, 10.1098/rspa.1998.0193, 1998.

Huemmrich, K. F.: The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance, Remote Sensing of Environment, 75, 423-431, 2001.

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

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

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F. M., and 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 Sensing of Environment, 52, 163-172, 1995.

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

Jakubauskas, M. E., Legates, D. R., and Kastens, J. H.: Crop identification using harmonic analysis of time-series AVHRR NDVI data, Computers and Electronics in Agriculture, 37, 127-139, 2002.

Jiang, B., Liang, S., Wang, J., and Xiao, Z.: Modeling MODIS LAI time series using three statistical methods, Remote Sensing of Environment, 114, 1432, 2010a.

Jiang, B., Liang, S., Wang, J., and Xiao, Z.: Modeling MODIS LAI time series using three statistical methods, Remote Sensing of Environment, 114, 1432-1444, 2010b.

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

Jonsson, P., and Eklundh, L.: Seasonality extraction by function fitting to time-series of satellite sensor data, Geoscience and Remote Sensing, IEEE Transactions on, 40, 1824-1832, 10.1109/tgrs.2002.802519, 2002.

Jönsson, P., and Eklundh, L.: TIMESAT: a program for analyzing time-series of satellite sensor data, 30, 833-845, 2004.

Kandasamy, S., Lopez-Lozano, R., Baret, F., and Rochdi, N.: The effective nature of LAI as measured from remote sensing observations, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010, 789-792.

Kandasamy, S., Neveux, P., Verger, A., Buis, S., Weiss, M., and Baret, F.: Improving the consistency and continuity of MODIS 8 days Leaf Area Index Product, Signal Processing Symposium, Jachranka Village, Warsaw, Poland, 2011.

Kandasamy, S., Baret, F., Verger, A., Neveux, P., and Weiss, M.: A COMPARISON OF METHODS FOR SMOOTHING AND GAP FILLING TIME SERIES OF REMOTE SENSING OBSERVATIONS. APPLICATION TO MODIS LAI PRODUCTS, Biogeosciences Discussions, 9, 17053-17097, 10.5194/bgd-9-17053-2012, 2012a.

Kandasamy, S., Neveux, P., Verger, A., Buis, S., Weiss, M., and Baret, F.: Improving the Consistency and Continuity of MODIS 8 Day Leaf Area Index Products, International Journal of Electronics and Telecommunications, 58, 141-146, 10.2478/v10177-012-0020-8, 2012b.

Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., and Lee, R.-Y.: Image masking for crop yield forecasting using AVHRR NDVI time series imagery, Remote Sensing of Environment, 99, 341-356, 2005.

MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod15.pdf, access: 3-Oct, 1999.

Kobayashi, H., and Dye, D. G.: Atmospheric conditions for monitoring the long-term vegetation dynamics in the Amazon using normalized difference vegetation index, Remote Sensing of Environment, 97, 519-525, 2005.

Kobayashi, H., Suzuki, R., and Kobayashi, S.: Reflectance seasonality and its relation to the canopy leaf area index in an eastern Siberian larch forest: Multi-satellite data and radiative transfer analyses, Remote Sensing of Environment, 106, 238-252, 10.1016/j.rse.2006.08.011, 2007.

Lang, A. R. G., and McMurtrie, R. E.: Total leaf areas of single trees of Eucalyptus grandis estimated from transmittances of the sun's beam, Agricultural and Forest Meteorology, 58, 79-92, 10.1016/0168-1923(92)90112-h, 1992.

Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., and Smith, B.: Climatic Control of the High-Latitude Vegetation Greening Trend and Pinatubo Effect, Science, 296, 1687-1689, 10.1126/science.1071828, 2002.

Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., and Wu, S.: Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield, Mathematical and Computer Modelling, 10.1016/j.mcm.2011.10.038.

Maass, J., Vose, J. M., Swank, W. T., and Martinez-Yrizar, A.: Seasonal changes of leaf area index (LAI) in a tropical deciduous forest in west Mexico, Forest Ecology and Management, 74, 171-180, 10.1016/0378-1127(94)03485-f, 1995.

Martinez, B., and Gilabert, M. A.: Vegetation dynamics from NDVI time series analysis using the wavelet transform, Remote Sensing of Environment, 113, 1823-1842, 10.1016/j.rse.2009.04.016, 2009.

Moody, A., and Johnson, D. M.: Land-Surface Phenologies from AVHRR Using the Discrete Fourier Transform, Remote Sensing of Environment, 75, 305-323, 2001.

Morisette, J. T., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J. E., Garrigues, S., Shabanov, N. V., Weiss, M., Fernandes, R. A., Leblanc, S. G., Kalacska, M., Sanchez-Azofeifa, G. A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A. N., Lewis, T. E., Iiames, J. S., Colombo, R., Meroni, M., Busetto, L., Cohen, W. B., Turner, D. P., Warner, E. D., Petersen, G. W., Seufert, G., and Cook, R.: Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1804-1817, 10.1109/tgrs.2006.872529, 2006.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sensing of Environment, 83, 214-231, 2002.

North, P. R. J.: Three-dimensional forest light interaction model using a Monte Carlo method, Geoscience and Remote Sensing, IEEE Transactions on, 34, 946-956, 10.1109/36.508411, 1996.

Pandya, M. R., Singh, R. P., Chaudhari, K. N., Bairagi, G. D., Sharma, R., Dadhwal, V. K., and Parihar, J. S.: Leaf area index retrieval using IRS LISS-III sensor data and validation of the MODIS LAI product over central India, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1858-1865, 10.1109/tgrs.2006.876028, 2006.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C.: Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends in Ecology & Evolution, 20, 503-510, 10.1016/j.tree.2005.05.011, 2005.

Piao, S., Friedlingstein, P., Ciais, P., Zhou, L., and Chen, A.: Effect of climate and CO2 changes on the greening of the Northern Hemisphere over the past two decades, Geophys. Res. Lett., 33, L23402, 10.1029/2006gl028205, 2006.

Pinzón, J. E., Brown, M. E., and Tucker, C. J.: EMD correction of orbital drift artifacts in satellite data stream, in: HILBERT-HUANG TRANSFORM AND ITS APPLICATIONS, edited by: Huang, N. E., and Shen, S. S. P., World Scientific, 167-183, 2005.

Pouliot, D., Latifovic, R., Fernandes, R., and Olthof, I.: Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data, Remote Sensing of Environment, 113, 1749-1759, 2009.

Price, J. C.: Estimating Leaf Area Index From Satellite Data, IEEE Transactions on Geoscience and Remote Sensing, 31, 727-734, 1993.

Qin, Z., Su, G.-l., Zhang, J.-e., Ouyang, Y., Yu, Q., and Li, J.: Identification of important factors for water vapor flux and CO2 exchange in a cropland, Ecological Modelling, 221, 575-581, 10.1016/j.ecolmodel.2009.11.007, 2010.

Ql, J., and Kerr, Y.: On current compositing algorithms, Remote Sensing Reviews, 15, 235-256, 10.1080/02757259709532340, 1997.

Rasmussen, C., and Williams, C.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series), The MIT Press, 2005.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O.: Measuring Phenological Variability from Satellite Imagery, Journal of Vegetation Science, 5, 714, 1994.

Richardson, A. J., Wiegand, C. L., Wanjura, D. F., Dusek, D., and Steiner, J. L.: Multisite analyses of spectral-biophysical data for Sorghum, Remote Sensing of Environment, 41, 71-82, 1992.

Roerink, G. J., Menenti, M., and Verhoef, W.: Reconstructing cloudfree NDVI composites using Fourier analysis of time series, International Journal of Remote Sensing, 21, 1911-1917, 10.1080/014311600209814, 2000.

Roujean, J.-L., and Lacaze, R.: Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: A pragmatic method and its validation, J. Geophys. Res., 107, 4150, 10.1029/2001jd000751, 2002.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C.: Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFC, 1974.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., and Lindquist, E.: Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data, Remote Sensing of Environment, 112, 3112-3130, 2008.

Running, S. W., and Coughlan, J. C.: A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary

production processes, Ecological Modelling, 42, 125-154, 10.1016/0304-3800(88)90112-3, 1988.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., and Ohno, H.: A crop phenology detection method using time-series MODIS data, Remote Sensing of Environment, 96, 366-374, 2005.

Sakamoto, T., Wardlow, B. D., Gitelson, A. A., Verma, S. B., Suyker, A. E., and Arkebauer, T. J.: A Two-Step Filtering approach for detecting maize and soybean phenology with time-series MODIS data, Remote Sensing of Environment, 114, 2146-2159, DOI: 10.1016/j.rse.2010.04.019, 2010.

Schubert, P., Eklundh, L., Lund, M., and Nilsson, M.: Estimating northern peatland CO2 exchange from MODIS time series data, Remote Sensing of Environment, 114, 1178-1189, 2010.

Shabanov, N. V., Dong, H., Wenze, Y., Tan, B., Knyazikhin, Y., Myneni, R. B., Ahl, D. E., Gower, S. T., Huete, A. R., Aragao, L. E. O. C., and Shimabukuro, Y. E.: Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests, Geoscience and Remote Sensing, IEEE Transactions on, 43, 1855-1865, 10.1109/tgrs.2005.852477, 2005.

Steduto, P., and Hsiao, T. C.: Maize canopies under two soil water regimes: II. Seasonal trends of evapotranspiration, carbon dioxide assimilation and canopy conductance, and as related to leaf area index, Agricultural and Forest Meteorology, 89, 185-200, 10.1016/s0168-1923(97)00084-1, 1998.

Thenkabail, P. S., Schull, M., and Turral, H.: Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data, Remote Sensing of Environment, 95, 317-341, 2005.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, International Journal of Remote Sensing, 26, 4485-4498, 10.1080/01431160500168686, 2005.
Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, Remote Sensing of Environment, 114, 106-115, 2010.

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

Verger, A., Baret, F., and Camacho, F.: 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, 415-426, 2011a.

Verger, A., Baret, F., and Weiss, M.: A multisensor fusion approach to improve LAI time series, Remote Sensing of Environment, 115, 2460-2470, 2011b.

Verger, A., Baret, F., Weiss, M., Kandasamy, S., and Vermote, E.: Quantification of LAI interannual anomalies by adjusting climatological patterns, Analysis of Multi-temporal Remote Sensing Images (Multi-Temp), 2011 6th International Workshop on the, 2011c, 113-116.

Verger, A., Baret, F., Weiss, M., Kandasamy, S., and Vermote, E.: The CACAO method for smoothing, gap filling and characterizing anomalies in satellite time series, IEEE Transactions on Geoscience and Remote Sensing,, in press, 10.1109/TGRS.2012.2228653, 2012.

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

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, Remote Sensing of Environment, 115, 801-823, 2011.

Viovy, N., Arino, O., and Belward, A. S.: The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series, International Journal of Remote Sensing, 13, 1585-1590, 10.1080/01431169208904212, 1992.

Weiss, M., Baret, F., Myneni, R., Pragnère, A., and Knyazikhin, Y.: Investigation of a model inversion technique for the estimation of crop charcteristics from spectral and directional reflectance data, Agronomie, 20, 3-22, 2000.

Weiss, M., Baret, F., Smith, G. J., Jonckheere, I., and Coppin, P.: Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling, Agricultural and Forest Meteorology, 121, 37-53, 2004.

Weiss, M., Baret, F., Garrigues, S., and Lacaze, R.: LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products, Remote Sensing of Environment, 110, 317-331, 2007a.

Weiss, M., Baret , F., Garrigues, S., Lacaze, R., and Bicheron, P.: LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. part 2: Validation and comparison with MODIS Collection 4 products., Remote sensing of Environment, 110, 317-331, 2007b.

Wiegand, C. L., Gerbermann, A. H., Gallo, K. P., Blad, B. L., and Dusek, D.: Multisite analyses of spectral-biophysical data for corn, Remote Sensing of Environment, 33, 1-16, 1990.

Wiegand, C. L., Maas, S. J., Aase, J. K., Hatfield, J. L., Pinter Jr, P. J., Jackson, R. D., Kanemasu, E. T., and Lapitan, R. L.: Multisite analyses of spectral-biophysical data for wheat, Remote Sensing of Environment, 42, 1-21, 1992.

Wylie, B. K., Fosnight, E. A., Gilmanov, T. G., Frank, A. B., Morgan, J. A., Haferkamp, M. R., and Meyers, T. P.: Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains, Remote Sensing of Environment, 106, 399-413, 2007.

Yan, H., Wang, S. Q., Billesbach, D., Oechel, W., Zhang, J. H., Meyers, T., Martin, T. A., Matamala, R., Baldocchi, D., Bohrer, G., Dragoni, D., and Scott, R.: Global estimation of evapotranspiration using a leaf area index-based surface energy and water balance model, Remote Sensing of Environment, 124, 581-595, 10.1016/j.rse.2012.06.004, 2012.

Young, P. C., Pedregal, D. J., and Tych, W.: Dynamic harmonic regression, Journal of Forecasting, 18, 369-394, 10.1002/(sici)1099-131x(199911)18:6<369::aid-for748>3.0.co;2-k, 1999.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sensing of Environment, 115, 1171, 2011.

Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F., and Moore Iii, B.: 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, 357, 2005.

Zhiqiang, X., Shunlin, L., Jindi, W., and Xiyan, W.: Use of an ensemble Kalman filter for real-time inversion of leaf area index from MODIS time series data, Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009, 2009, IV-73-IV-76.

Zhu, X., Chen, J., Gao, F., Chen, X., and Masek, J. G.: An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions, Remote Sensing of Environment, 114, 2610-2623, 2010.

Summary

Monitoring biophysical variables at a global scale over long time periods is vital to address the climate change and food security challenges. Leaf Area Index (LAI) is a structure variable giving a measure of the canopy surface for radiation interception and canopy-atmosphere interactions. LAI is an important variable in many ecosystem models and it has been recognized as an Essential Climate Variable. This thesis aims to provide global and continuous estimates of LAI from satellite observations in near-real time according to user requirements to be used for diagnostic and prognostic evaluations of vegetation state and functioning. There are already some available LAI products which show however some important discrepancies in terms of magnitude and some limitations in terms of continuity and consistency. This thesis addresses these important issues. First, the nature of the LAI estimated from these satellite observations was investigated to address the existing differences in the definition of products. Then, different temporal smoothing and gap filling methods were analyzed to reduce noise and discontinuities in the time series mainly due to cloud cover. Finally, different methods for near real time estimation of LAI were evaluated. Such comparison assessment as a function of the level of noise and gaps were lacking for LAI.

Results achieved within the first part of the thesis show that the effective LAI is more accurately retrieved from satellite data than the actual LAI due to leaf clumping in the canopies. Further, the study has demonstrated that multi-view observations provide only marginal improvements on LAI retrieval. The study also found that for optimal retrievals the size of the uncertainty envelope over a set of possible solutions to be approximately equal to that in the reflectance measurements. The results achieved in the second part of the thesis found the method with locally adaptive temporal window, depending on amount of available observations and Climatology as background estimation to be more robust to noise and missing data for smoothing, gap-filling and near real time estimations with satellite time series.

Keywords: Leaf Area Index, radiative transfer model inversion, satellite, reflectance, time series processing, near real-time estimation

Résumé

Le suivi des variables biophysiques à l'échelle globale sur de longues périodes de temps est essentielle pour répondre aux nouveaux enjeux que constituent le changement climatique et la sécurité alimentaire. L'indice foliaire (LAI) est une variable de structure définissant la surface d'interception du rayonnement incident et et d'échanges gazeux avec l'atmosphère. Le LAI est donc une variable importante des modèles d'écosystèmes et a d'ailleurs été reconnue comme variable climatique essentielle (ECV). Cette thèse a pour objectif de fournir des estimations globales et continues de LAI à partir d'observations satellitaires en temps quasi-réel en réponse aux besoins des utilisateurs pour fournir des diagnostiques et pronostiques de l'état et du fonctionnement de la végétation. Quelques produits LAI sont déjà disponibles mais montrent des désaccords et des limitations en termes de cohérence et de continuité. Cette thèse a pour objectif de lever ces limitations. Dans un premier temps, on essaiera de mieux définir la nature des estimations de LAI à partir d'observations satellitaires. Puis, différentes méthodes de lissage te bouchage des séries temporelles ont été analysées pour réduire le bruit et les discontinuités principalement liées à la couverture nuageuse. Finalement quelques méthodes d'estimation temps quasi réel ont été évaluées en considérant le niveau de bruit et les données manquantes.

Les résultats obtenus dans la première partie de cette thèse montrent que la LAI effectif et bien mieux estimé que la valeur réelle de LAI du fait de l'agrégation des feuilles observée au niveau du couvert. L'utilisation d'observations multidirectionnelles n'améliore que marginalement les performances d'estimation. L'étude montre également que les performances d'estimation optimales sont obtenues quand les solutions sont recherchées à l'intérieur d'une enveloppe définie par l'incertitude associée aux mesures radiométriques. Dans la deuxième partie consacrée à l'amélioration de la continuité et la cohérence des séries temporelles, les méthodes basées sur un fenêtre temporelle locale mais de largeur dépendant du nombre d'observations présentes, et utilisant la climatologie comme information a priori s'avèrent les plus intéressantes autorisant également l'estimation en temps quasi réel.

Mots-clés: Indice foliaire, inversion, modèle de transfert radiatif, satellite, séries temporelle, réflectance, estimation en temps quasi-réel.