DIAL
Digital access to libraries

"Satellite remote sensing priorities for better assimilation in crop growth models : winter wheat LAI and grassland mowing dates case studies"

Curnel, Yannick

Abstract

In a context of markets globalization, early, reliable and timely estimations of crop yields at regional to global scale are clearly needed for managing large agricultural lands, determining food pricing and trading policies but also for early warning of harvest shortfalls. Crop growth models are often used to estimate yields within the growing season. The uncertainties associated with these models contribute to the uncertainty of crop yield estimations and forecasts. Satellite remote sensing, through its ability to provide synoptic information on growth conditions over large geographic extents and in near real-time, is a key data source used to reduce uncertainties in biophysical models through data assimilation methods. This thesis aims at assessing possible improvements for the assimilation of remotely-sensed biophysical variables in crop growth models and to estimate their related errors reduction on modelled yield estimates. Assimilated observations are winter wheat leaf area in...

Document type : *Thèse (Dissertation)*

Référence bibliographique

Curnel, Yannick. *Satellite remote sensing priorities for better assimilation in crop growth models : winter wheat LAI and grassland mowing dates case studies.* Prom. : Defourny, Pierre ; Oger, Robert

Available at:
http://hdl.handle.net/2078.1/160968

[Downloaded 2019/04/19 at 05:25:37]

Université catholique de Louvain
Faculty of Bioscience Engineering
Earth and Life Institute – Environmental sciences

SATELLITE REMOTE SENSING PRIORITIES FOR BETTER ASSIMILATION IN CROP GROWTH MODELS: WINTER WHEAT LAI AND GRASSLAND MOWING DATES CASE STUDIES

Yannick Curnel

Avril 2015

Thèse présentée en vue de l'obtention
du grade de docteur en sciences agronomiques
et ingénierie biologique

Promoteurs : Pr. Pierre Defourny
Dr Robert Oger

Jury Composition:

President:
Pr. DELVAUX Bruno (UCL, Belgium)

Promoters:

Pr. DEFOURNY Pierre (UCL, Belgium)
Dr Ir OGER Robert (CRA-W, Belgium)

Readers:

Dr DERAUW Dominique (CSL, Belgium)
Dr De WIT Allard (WUR, Netherlands)
Pr. DRAYE Xavier (UCL, Belgium)
Pr. TYCHON Bernard (ULg, Belgium)

This research has been financed by the Walloon Agricultural Research Centre (CRA-W) in the framework of MIMOSA project and realised in close collaboration with the Earth and Life Institute (Environmental sciences, ELIE) of UCL (Université catholique de Louvain)

WALLOON AGRICULTURAL RESEARCH CENTRE (CRA-W)

Agriculture and natural environment Department (D3)
Farming systems, territories and information technology Unit (U11)
Gembloux – Belgium

UNIVERSITE CATHOLIQUE DE LOUVAIN

Earth and Life Institute
Environmental sciences
Louvain-la-Neuve - Belgium

ACKNOWLEDGEMENTS

To paraphrase a famous quote of the marathon runner Emil Zatopek “If you want to run, run a mile. If you want to experience a different life, runs a marathon”, I would be tempted to say “If you want to be a scientist, work in a research centre or in a university. If you want to experiment a different scientific life, make a PhD”.

The analogies between a PhD and a marathon are numerous. Starting my scientific career, I was thinking that I was not yet able to take up the challenge. Progressively, the experience increasing, one day you decide that you are ready.

You have then to choose your destination, to define your finish line. The choice of my thesis subject was the result of my personal interest (and experience) for mathematical models and remote sensing but also the encounter in 2008 in a taxi on the way to a meeting in Stresa with the person that will be one of my promoters, Mr Defourny.

Many people think that a PhD, as a marathon, is a long and lonesome effort but it is not true. You have first to find your coaches. They will not give you a piggyback but they will provide you advices will share their experience and will help you to face the difficulties. In this context, I would like to thank first of all warmly my two promoters. First, I would like to thank Mr Oger Robert that has been my director for 10 years. He has been always available, notably for giving me good scientific advices (even if I have not realised it immediately or if I have given the impression not to hear them). I hope that he will enjoy his retirement. I would like also to thank my second promoter, Mr Defourny Pierre. Even If he is always very busy and that it is therefore not always easy to meet him, his global vision of the problem and his encouragements were for me very useful. I am really grateful to all members of my Jury, especially Mr De wit Allard who helped me a lot all along this PhD (especially on PyWofost) and welcomed me for a 10 days training in Wageningen UR premises.

When you start a PhD, as when you start running after the departure signal, motivation and energy is at the maximum. Unfortunately, soon or later, you are likely to “hit the wall”. Fortunately, at this moment, you can find people that refuel your batteries and help you to face the difficulties. I would like to express my gratitude to all the “supporters” along the road. I think first to my family and more especially to my wife Cindy (who had to face sometimes my moody character) and my light in the dark, my sweet daughter Lucie born during my PhD experience. I would like to thank my colleagues at CRA-W. It would be too long to name all of them. I would just cite Quentin Leroy, with who have shared my office and my taste for rock and heavy metal music and Viviane Planchon for listening and supporting me in my numerous periods of doubts. My gratitude goes also to the CRA-W as institution for the funding of this thesis through the Moerman projects.

No real progression is really possible without sparing partners. In this context, I especially think to all the PhD students involved in the GLOBAM project (who have shown me the way by already crossing the finish line): Emilie Bériaux, Aline Léonard, Louis Kouadio and Grégory Duveiller with who have had and I still have long

scientific discussions on remote sensing and crop growth modelling. I won't be fair not to cite the ENGE team. Thank you very much for welcoming me.

During this PhD, I have also had the opportunity and the chance to share my experience with Thibault Delvaux and Christophe Bocquet in the frame of their master theses. I really hope that I have been a good coach and that they have learned something during this period. From my side, I would like to thank them (as well as Aline Léonard) for the processing of SAR data and their collaboration to this thesis. A part of it belongs to them.

To conclude, I would say that this PhD, even if long was the road and hard was the way, was a rewarding experience where you learn a lot about yourself, your flaws and your qualities.

I have started with an Emil Zatopek quote, I will finally conclude by another one "What has passed is already finished with. What I find more interesting is what is still to come".

Gembloux, April 2015

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AE	Assimilation Efficiency
AgRISTARS	Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing
AMIS	Agricultural Market Information System
APES	Agricultural Production and Externalities Simulator
AVHRR	Advanced Very High Resolution Radiometer
AWIFS	Advanced WIde Field Sensor
B-CGMS	Belgian Crop Growth Monitoring System
BRDF	Bidirectional Reflectance Distribution Function
CAP	Common Agricultural Policy
CCWS	China CropWatch System
CEOS	Committee on Earth Observation Satellites
CERES	Crop Environment REsource Synthesis
CGMS	Crop Growth Monitoring System
CLAIR	*Clevers* Leaf Area Index by Reflectance
COSW	Carte d'occupation du sol de Wallonie
CRA-W	Walloon agricultural research Centre
CSDM	Canopy Structural Dynamic Model
CV	Coefficient of Variation
DFF	*Dépression Fagne Famenne*
DN	Digital Number
DOY	Day Of Year
DSSAT	Decision Support System for Agrotechnology Transfer
EMS	Electro-Magnetic Spectrum
ENL	Equivalent Number of Looks
EO	Earth Observation
ESA	European Space Agency
EU	European Union
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
fAPAR	fraction of Absorbed Photosynthetically Active Radiation
FASAL	Forecasting Agricultural output using Space, Agrometeorological and Land based observations
G	Grazing (parcels)
GAI	Green Area Index

GEO	Group on Earth Observation
GEOGLAM	Global Agricultural Monitoring initiative of GEO
GIEWS	Global Information and early Warning System
GMFS	Global Monitoring for Food Security
GLOBAM	GLOBal Agricultural Monitoring systems by integration of Earth observation and modeling techniques (STEREO II project)
HPAC	*Haut Plateau de l'Ardenne Centrale*
IFOV	Instantaneous Field Of View
IR	Infra Red
JECAM	Joint Experiment for Crop Assessment and Monitoring
JM	Just Mowed (parcels)
LAI	Leaf Area Index
LACIE	Large Area Crop Inventory Experiment
LEAFMOD	Leaf Experimental Absorptivity Feasibility MODdel
LIBERTY	Leaf Incorporating Biochemistry Exhibiting Reflectance and Transmittance Yields
LINGRA	LINTUL GRAssland
LINTUL	Light INTerception and UtiLization simulator
LPIS	Land Parcel Identification System
M	Mowing (parcels)
MAE	Minimal Assimilation Efficiency
MARS	Monitoring Agriculture by Remote Sensing
MCYFS	MARS Crop Yield Forecasting System
MERIS	MEdium Resolution Imaging Spectrometer instrument
MIR	Medium Infra Red
MODIS	MOderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-InfraRed
NM	Not mowed (parcels)
NNT	Neural NeTworks
NUTS	Nomenclature des unités territoriales statistiques
OECD	Organisation for Economic Co-operation and Development
OSSE	Observing System Simulation Experiment
PRI	PRecision Image
PROBA-V	PRoject for On-Board Autonomy-Vegetation
PROSAIL	model coupling PROspect and SAIL

PROSPECT	leaf optical PROperties SPECTral model
PSF	Point Spread Function
RADAR	RAdio Detection And Ranging
RMAE	Relative Mean Absolute Error
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
RS	Remote Sensing
RTM	Radiative Transfer Model
RUE	Radiation Use Efficiency
SAIL	Scattering by Arbitrarily Inclined Leaves(model)
SAVI	Soil Adjusted Vegetation Index
SCEM-UA	Shuffled Complex Evolution Metropolis parameter optimization algorithm
SIGEC	*Système Intégré de Gestion et de Contrôle*
SLOP	Stochastic model for Leaf Optical Properties
SNR	Signal-to-Noise Ratio
SODA	Simultaneous parameter Optimization and Data Assimilation
SPAN	Life span of leaves growing at 35°C
SPOT	*Satellite Probatoire pour l'Observation de laTerre*
SRTM	Shuttle Radar Topography Mission
STEREO	Support to the Exploitation and Research in Earth Observation data
STICS	*Simulateur mulTIdisciplinaire pour les Cultures Standard*
SUCROS	Simple and Universal CROp growth Simulator
SWAP	Soil Water Plant Atmosphere
SWIR	Short-Wave InfraRed
TDWI	Initial total crop dry weight
TM	Thematic Mapper
TOA	Top Of Atmosphere
TOC	Top Of Canopy
TRIM	Three-dimensional Radiation Interaction Model
TSUM1	Temperature sum from emergence to anthesis
TUE	Transpiration Use Efficiency
TDWI	Initial total crop dry weight
USDA	United State Department of Agriculture
UV	UltraViolet
VI	Vegetation Index

VZA	View Zenith Angle
WCM	Water Cloud Model
WMAE	Weighted Mean Absolute Error
WOFOST	WOrld FOod Studies (crop growth model)

LIST OF SYMBOLS

A	Forecasted matrices of ensemble states (EnKf)
AB	Aboveground grassland Biomass [kg/ha]
A_i	Forecasted LAI for ensemble member *i* [m²/m²]
A^a	Analysed matrices of ensemble states (EnKf)
A_i^a	Analysed LAI for ensemble member *i* [m²/m²]
AE	Assimilation efficiency [%]
$AE(j)$	Assimilation efficiency of iteration *j* [%]
$BS_{mod,i}$	True backscatter value at time *i* [dB]
$BS_{RS,i}$	Synthetic backscatter value at time *i* [dB]
c	Velocity of light (299,792 km/s)
CV	Coefficient of variation [%]
D_i	Perturbed LAI value used to update ensemble members
$FGYA_{ij}$	*F*inal grain yield after assimilation for truth *i* and iteration *j* [kg/ha]
$FGYO$	Final grain yield from the ”control” run [kg/ha]
$FGYT_i$	Final grain yield of the truth *i* [kg/ha]
GAI_o	MODIS-observed Green Area Index [m²/m²]
GAI_m	Modeled Green Area Index [m²/m²]
H	Measurement operator (EnKf)
m	Number of winsorized values [-]
v	Frequency [s^{-1}]
λ	Wavelength [m]
σ^0	Sigma nought or radar backscatterring coefficient [dB]
$LAI_{mod,i}$	True leaf area index at time *i* [m²/m²]
$LAI_{RS,i}$	Synthetic leaf area index at time *i* [m²/m²]
P_e	Ensemble variance/covariance matrices (EnKf)
Rain_3d	Sum of rainfall over the 72h preceding SAR acquistions (mm)
R_e	Observation variance/covariance matrices (EnKf)
T_{base}	Base effective temperature [°c]
TBA_{ijk}	Total biomass yield after assimilation for truth *i*, iteration *j* and repetition *k*[kg/ha]
TBO	Total grassland biomass yield from the ”control” run [kg/ha]
TBT_i	Total grassland biomass yield of truth *i* [kg/ha]
T_e	Effective temperature [°c]
$T_{max,e}$	Maximum effective temperature [°c]

INTRODUCTION

Context of the world food insecurity

According to the UN Food and Agriculture Organisation (FAO), one and a half times the amount of food needed to provide everyone in the world with an adequate and nutritious diet is already produced, yet one in seven people is suffering from hunger. People are hungry because they cannot afford food, not because the population is growing so fast that food is becoming scarce.

Indeed, though prices of food commodities on world markets (adjusted for inflation) declined substantially from the early 1960s to the early 2000s when they reached a historic low level (Figure 1), a slow increase of these prices have been observed from 2003 to 2006 before a sudden rise from 2006.

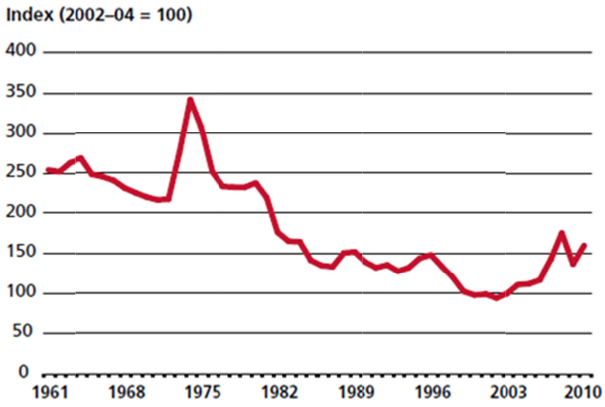


Figure 1.- FAO food price index evolution for 1961-2010 period (FAO, 2011a).

All the commodity group price indices composing the global food price index have followed the same trend with various intensities (figure 2).

The surge in food prices in the last years, which took many by surprise, has been the most marked of the past century in its magnitude, duration and the number of commodity groups whose prices have increased. The ensuing crisis has resulted in a 50-200% increase in selected commodity prices, driven 110 million people into poverty and added 44 million more to the undernourished. High food prices have had dramatic impacts on the lives and livelihoods, including increased infant and child mortality, of those already undernourished or living in poverty and spending 70-80% of their daily income on food (Nellemann *et al.*, 2009). Large short-term price change can have indeed long-term impacts on development by reducing the children's consumption of key nutrients during first days of life which leads to a permanent reduction of future earning capacity and therefore increasing the likelihood of future poverty and thus slowing the economic development process (FAO, 2011a).

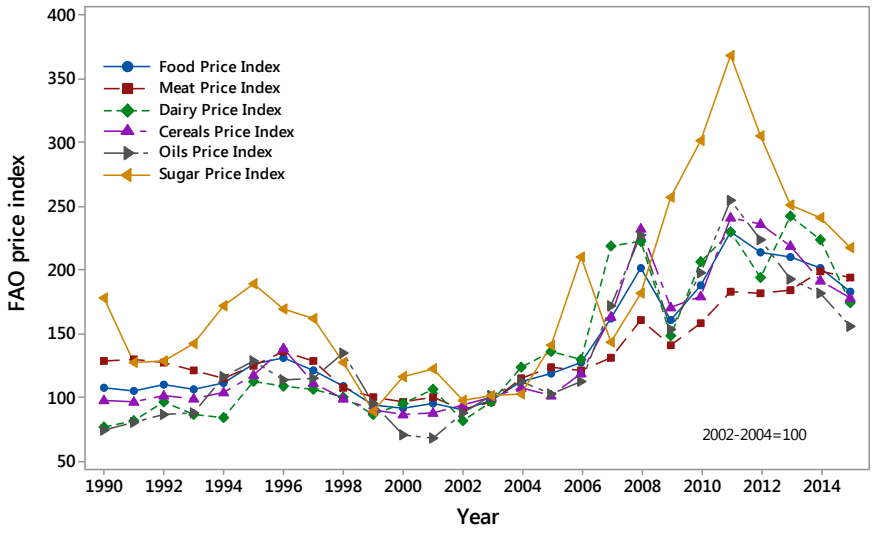


Figure 2.- FAO price indices evolution on 1990-2015 period (source: www.fao.org)

Concerning the evolution of the situation, many medium to long-term projection models seems to suggest that food commodity prices will remain relatively high over the next decade or so. The Organisation for Economic Co-operation and Development (OECD) and FAO (OECD/FAO, 2011) project that world prices for rice, wheat, maize and oilseeds in the five years from 2015/16 to 2019/2020 will be higher in real terms by 40, 27, 48 and 36 percent respectively than in the five years from 1998/99 to 2002/2003 (FAO, 2011a).

Although various observers attach different degrees of importance to assorted factors, there is a relatively strong consensus that multiple factors had a role in the price increases.

First of these factors is the weather shocks, such as drought events observed in Australia in 2005-2007 or in Russia during summer 2010, inducing lower production and trade. For example, the droughts in Russia during the summer 2010 incited the Russian authorities to announce a ban on wheat exports due to the expected production shortfall. The result was a new spike of the global wheat price which prompted fears of a similar food crisis as in 2007-2008. Though it didn't happened mainly due a relatively good replenishment of the grain global stocks by two previous years of good harvests, the FAO cereals price index rose above record of 2008 (Figure 2). Main explanations lie in the floods observed in the important grain-producing region of Queensland and the particularly dry summer in the American "corn belt" (Duveiller, 2011). If the frequency of extreme weather events increases as expected (IPCC, 2012 ; Olesen *et al.*, 2011), such production shocks will be more frequent which will tend to increase also the prices volatility and may lead to

unpredictable changes in prices, especially if stocks are low to begin with (FAO, 2011b).

Biofuel policies have also created new linkages between the price of oil and the price of food commodities. When oil prices increase, demand for biofuels will increase and will therefore compete with food production. For example, the corn equivalent of the energy used for a few minutes' drive could feed a person for a day while a full tank of ethanol in a large 4-wheel drive suburban utility vehicle could almost feed one person for a year. The current policies promoting Biofuel, such as the Renewable Energy Directive (RED) implemented by the European Union stating that the share of renewable energy sources (including non-liquids) should increase to 10% of total transport fuel use by 2020, increase the pressure on food markets (OECD/FAO, 2011). Because world oil prices have historically been more volatile than food prices, world food markets may also be subject to increased volatility.

The increase of food prices can be also explained by the growth in food demand and need resulting from the combined effects of world population growth to over 9 billion by 2050, rising incomes and dietary changes towards higher meat intake. Meat production is particularly demanding in terms of energy, cereal and water.

The largest population increase is projected to occur in emerging countries (Figure 3), mainly in Asia and particularly in China, India and Southeast Asia. These emerging economies have long term economic growth putting upward pressure on prices for petroleum and fertilizer because of the resource-intensive nature of their economic growth and leading to increased demand for meat, and hence animal feed, as diets diversified.

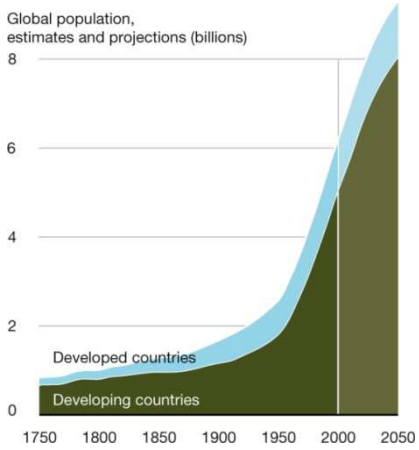


Figure 3.- Human population growth in developed and emerging countries (source: Nellemann *et al.*, 2009)

The higher prices for petroleum and fertilizer, besides the impact on biofuel demand, raise indirectly food production, transport and distribution costs and subsequently the food prices.

Dietary shifts towards more meat will induce a much larger share of cropland for grazing and feed production for the meat industry. As a part of cereal production is used for animal feeding, it hampers the cereal availability for human consumption which raises food prices.

Slower growth of cereal yields (and production), especially those of rice and wheat, during the past 20 years as a result of low investment over the previous three decades, the increased demand on commodity future markets as a result of both speculation and portfolio diversification and the low levels of stocks, are among the other factors explaining the increase of food prices.

Consequences of higher food prices are different according to the type of population. Consequences are also different at short and long term. In the short term, the benefits of high prices go primarily to wealthy farmers with a large marketed surplus. As the poorest people usually buy more food than they sell, high food prices tend to worsen poverty, food insecurity and malnutrition. High food prices can have benefits however by representing an opportunity to spur long-term investment in agriculture which should contribute to sustainable food security in the longer run.

In addition to the impact of high or low prices, the variability and the unpredictability in food prices can have also important effects and repercussions even if average food prices remain constant. When prices fluctuate substantially, even if they are tolerable on average, the short-term shocks make both smallholder farmers and poor consumers vulnerable to long-term poverty traps. Prices volatility tends to limit the private farm-level investments leading for example to a lower fertilizer use and subsequently to a lower productivity. On a macroeconomic scale, prices volatility reduces the ability to function as signals guiding the resource allocation. Investments are therefore not necessary directed to optimal sectors of the economy which reduces the economy growth. Moreover, the poverty and the famine possibly induced by the volatile food prices can induce food riots and social problems damaging the investment climate and the subsequent use of subsidies limiting investment in public goods such as agricultural research, education, health and roads for example (FAO, 2011a).

There are therefore clear needs for price stability and for systems and methods providing accurate, efficient and affordable information on crop yield outlook and estimated cultivated areas estimates in near-real time at the scale of provinces and countries as basis for regional crop production estimates. This information is of interest for public and private organizations working at international and national levels, especially if such information relates to the major centres of production, or to regions with food security problems or regions with high climatic risk of crop failures (de Wit *et al*., 2008). Information on the current situation and outlook for global agriculture shapes expectations about future prices and allow markets to function more efficiently. Better information and analysis of global and local markets as well as an improved transparency could reduce the incidence and magnitude of panic-driven price surges (FAO, 2011a). Timely information about potential and observed harvest shortfalls can also hasten early identification of problems areas and help to organise and optimise food supplies at a regional scale (Duveiller *et al*., 2011a).

A first major policy response to price volatility in food and agricultural markets has been recently given by the establishment, at the request of the G20 Agriculture Ministers in 2011, of the Agricultural Market Information System (AMIS) and the GEOGLAM initiative. AMIS is an inter-agency platform seeking to strengthen collaboration and dialogue among main producing, exporting and importing countries with a view to enhance food market transparency and to encourage policy action in response to market uncertainty (www.amis-outlook.org). GEOGLAM contributes to the AMIS bulletin through its crop monitoring platform sharing satellite-derived information to the different operational agricultural monitoring systems.

Agricultural monitoring at global and local scale

Accurate yield predictions are of great economic importance. To be of use, the crop production data has to be delivered to the authorities in a timely manner.

Economic globalization as well as the increasing consideration of environmental concerns in national, European and even world policies has also gradually but firmly changed the farming world. Gradually the emphasis has been set on farm profitability, environmental-friendly production modes and food safety. On a more local level, this evolution has constrained farmers to be able to meet more and more precise technical and economical requirements. Needed information to support their decisions should be made available as quick as possible. The quantitative and qualitative monitoring of agricultural productions (arable crops and grasslands) throughout the growing season is a part of the necessary information to meet cross compliance at region, farm or parcel level.

Assessment of crop production implies to distinguish, identify and measure the crops areas and to estimate their yields. The primary information is to determine before harvest time what a farmer is growing, in order to predict yields, anticipate the market, control possible expenditure (e.g. subsidies) and estimate the support needed. There is a major time constraint as the relevant information has to be gathered in the period between crop emergence and crop harvesting, which is just a few months (ESA, 2012).

At large scales (regional, national, continental or global), the accurate determination of crop production using traditional ground sampling strategies is not possible. This is due to the large number of factors affecting crop growth and their high spatio-temporal variability. For example, soil fertility and water availability, temperature and radiation regime, irrigation and fertilisation practices, pests and disease control, amongst others, affect crop growth to varying degrees.

In the same way, for Defourny *et al*. (2007), historical time series of cultivated areas and yields can serve to predict areas and yields of subsequent years but these predictions can be rather inaccurate due to a year-to-year variability of both areas and yields as observed in Europe. Additional information about the current year is therefore commonly needed. Inter-annual variability of croplands is usually limited contrary to yields varying more largely from one year to another and widely from one field to another, or even within the same field (Justice and Becker-Reshef, 2007).

According to Ray *et al*. (2015), climate variation accounts roughly for a third of global crop yield variability. While some areas show no significant influence, in substantial areas of the global breadbaskets, more than 60% of the yield variability can be explained by climate variability.

To face the aforementioned limitations of these methods, different strategies have been developed for monitoring crop growth and predicting crop yields and productions. First approach is based on regression models linking meteorological or vegetation indices to crop yields or productions (Tucker *et al.*, 1985; Doraiswamy *et al*., 2003; Becker-Reshef *et al*., 2010; Lobell and Burke, 2010; Mkhabela *et al*., 2011). Lobell and Field (2007) have for example shown that at global scale simple indicators of growing season temperatures and precipitation explained 30% or more of year-to-year variations in average yields for the 6 most grown crops. Changes in weather conditions do not, however, solely determine the extent of yield variability. Crops respond variously to the timing of weather events such as drought, heavy rains, low and high temperatures depending on crop and phenotypic stability of the cultivar grown and management practices (Peltonen-Sainio *et al.*, 2010). Moreover, most of these statistical models are not global and one of their main drawbacks is that they are only applicable in the region and the range of conditions for which they are developed (Becker-Reshef *et al*., 2010). Statistical models have also other shortcomings as problems of co-linearity between predictor variables (e.g. temperatures and precipitation), assumptions of stationarity (e.g. that past relationships will hold in the future, even if management events evolve) and low signal-to-noise ratios in yields (Lobell and Burke, 2010).

A second approach, getting round the limitations of the first approach based on statistical models, is based on simulation models. By the end of the 1960s, computational power was no longer a restriction (Passioura, 1996) allowing and stimulating the first attempts to synthesize detailed knowledge on plant physiological processes in order to explain the functioning of crops as a whole. Insights into various processes were expressed using mathematical equations and integrated in so-called simulation models. Though first objective of these models was to increase the understanding of crop behavior by explaining crop growth and development in terms of the underlying physiological mechanisms, their use has extended spurred by various research questions. In addition to their explanatory function, the applicability of well-tested models for extrapolation and prediction was quickly recognized and more application-oriented models were developed. For example, demands for advisory systems for farmers and scenario studies for policy makers resulted in the evolution of models geared respectively towards tactical and strategic decision support (Bouman *et al.*, 1996a).

These process-based models are typically developed and tested using experimental trials and thus offer the distinct advantage of leveraging decades of research on crop physiology and reproduction, agronomy and soil sciences among other disciplines. Yet these models also require extensive input data on cultivar, management and soil conditions that are unavailable in many parts of the world. More significantly, even in the presence of such data these models can be very difficult to calibrate because of

uncertain parameters (Lobell and Burke, 2010). A model is a simplified representation of a system, and a system is a limited part of reality that contains interrelated elements (de Wit, 1982).

The parallel development of agroecosystem models and remote sensing techniques led to an early fusion of these fields and to the development of synergic applications (Dorigo *et al.,* 2007). The combined use of remote sensing derived biophysical/biochemical state variables and agroecosystem models is expected to improve their predictive performance, especially at regional scale (Launay and Guérif, 2005). The two technologies present indeed obvious complementarities: for example, whereas the daily time step simulation capabilities of cropping systems models are excellent for crop growth analyses in the temporal domain, remote sensing images offer great opportunity to understand spatial crop growth patterns. Conversely, whereas model inputs requirements have limited the use of cropping systems models for spatial crop growth analyses, several practical problems, including cloud cover and satellite revisit time, have limited the reliability of remote sensing (especially in the optical domain) as temporal crop analysis tool. With the integration of these technologies, the problems associated with one can be offset by the benefits of the other (Thorp *et al., 2010*).

Agriculture is a major user of satellite remote sensing data (Moulin *et al.*, 1998). Preliminary research and development on satellite monitoring of agriculture started with the ETRS sensor (Landsat system) in the early 1970's (Justice and Becker-Reshef, 2007). Few years after the launch of the first civilian satellite for Earth Observation in 1972 (Landsat-1), the North-American Large Area Crop Inventory Experiment (LACIE) has demonstrated that improved accuracy in USDA predictions of wheat production can be achieved by the use of satellite imagery. The basic approach used in LACIE was to combine estimates of the land area planted in wheat with estimates of yield per unit area (Erickson, 1984). One of the main successes of the LACIE experiment was to be able to predict with great accuracy the 30% shortfall in Soviet spring wheat six weeks before the harvest (Liang, 2004). The Agriculture and resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) program extended the LACIE methodology to other crops and regions and also used remote sensing data to assess crop condition along the season and relating it to yield (Boatwright and Whitebread, 1986). For Europe, the council of Ministers of the European Union (EU) decided in 1988 to set up a project to improve the provision of agricultural statistics which are necessary to manage the large budgets involved in the European Common Agricultural Policy (CAP). This project, known as the MARS project (Monitoring Agriculture by Remote Sensing), comprised different activities such as regional crop inventories, satellite-based area estimates, assessment of foreign agricultural production and an agricultural information system (Council of the European Community, 1988). Several other global crop programs such as the USDA Foreign Agricultural service (FAS), the UN-FAO Food Security Global Information and Early Warning System (GIEWS), the USAID Famine Early Warning System (FEWS) and the EU Global Monitoring of Food Security (GMFS) make use of satellite observations for regional to global scale agricultural monitoring (Justice and Becker-Reshef, 2007).

Several countries have also established systems for monitoring national and foreign crop production as e.g. with the Chinese CropWatch system (Bingfang, 2006; Wu *et al.*, 2014), the FASAL system in India or the B-CGMS system in Belgium (Tychon *et al.*, 1999). Examples of the major regional to global agricultural monitoring systems are presented in Justice and Becker-Reshef (2007).

Current trend, with the GEOGLAM initiative, is geared toward the enhancement and the strengthening of the different agricultural monitoring programs through international networking, operationally focused research as well as data / method sharing. The task of GEOGLAM, endorsed in the G20 Heads of states' declaration (Cannes, November 2011), is to "coordinate satellite monitoring observation systems in different regions of the world in order to enhance crop production projections and weather forecasting data". More specifically, GEOGLAM aims to provide AMIS with an international and transparent multi-source, consensus assessment of crop growing conditions, status, and agro-climatic conditions, likely to impact global production. Assessments are produced operationally since September 2013 and published in the AMIS Market Monitor Bulletin.

Most of the current operational monitoring systems use however a very limited area of the potentialities of remote sensing. Remote sensing imagery is most of the time used for descriptive analyses, crop mapping or to derive growth indicators used as explanatory variables in statistical models sometimes in conjunction with other growth indicators derived from agroecosystem models. The several techniques for merging remote sensing data with model simulations developed in a next paragraph, known as "data assimilation" procedures and initially suggested by Wiegand *et al*. (1986) have been widely studied in the two last decades at parcel, local up to regional scale (De Wit and Van Diepen, 2007) but no operational use of these methods in monitoring systems is on the agenda.

Satellite Earth observation for agriculture monitoring

Nowadays, the number of instruments orbiting and observing the earth is large. These instruments can be divided in two broad categories: passive and active (A comprehensive review of remote sensing principles is available in Curnel, 2014)

Whatever the category of sensors (passive or active), distinction can be also performed on *spatial resolution* and *temporal resolution*.

The *spatial resolution* of a remote sensing system is the smallest possible feature that can be detected by this system. Spatial resolution notion should not be confused with the pixel size which is the smallest units of an image. Spatial resolution of current satellite sensors varies from tens of centimetre up to 1 km or more. Distinction is commonly made between (very) high (or fine), medium and low (or coarse) spatial resolutions. Limits for these categories are not really fixed as technological improvements constantly move these limits by providing new types of instrument. For example, MERIS (Medium Resolution Imaging Spectrometer) and

MODIS (MOderate Resolution Imaging Spectroradiometer) or more recently PROBA-V (PRoject for On-Board Autonomy-Vegetation) instruments attempting to replace AVHRR and VEGETATION by providing information on pixel grids of 250-300m were first classified as medium resolution satellites but are now largely considered as having a coarse spatial resolution (Duveiller, 2011). Current trends are to set upper limit at 4 and 30 m respectively for high and medium resolution images.

The *absolute temporal resolution* of a remote sensing system is equivalent to the revisit period which refers to the length of time it takes for a satellite to complete one entire cycle and therefore to image the same area at the same viewing angle a second time (http://www.nrcan.gc.ca/). Due to some degrees of overlap in the imaging swaths or adjacent orbits for most satellites and the increase in this overlap with increasing latitude, some areas of the earth tend however to be re-imaged more frequently. Also, some satellite systems are able to point their sensors to image the same area between different satellite passes. The *actual temporal resolution* of a sensor depends therefore on a variety of factors, including the satellite / sensor capabilities, the swath overlap and latitude. For example, SPOT sensors are configured to permit off-nadir viewing which can reduce revisit time from 26 days (i.e. the satellites' orbit cycle) to 2-3 days for HRVIR (Visible and Infrared High-Resolution) sensor and one day for the VEGETATION sensor (Khorram *et al*., 2012).

Up to recently, temporal and spatial resolutions were inversely proportional. Most coarse spatial resolution sensors such as AVHRR, VEGETATION or MODIS are nearly available on a daily basis while a medium resolution sensor such as Landsat 8 have a minimum revisit period of 16 days. High spatial resolutions instrument with daily revisit capabilities have recently appeared (e.g. the Canadian RapidEye and the Taiwanese FORMOSAT- 2 sensors) but they can only provide such observation frequency over a limited geographic extend. It can be noted that another recent acquisition strategy for high spatial resolution is to increase the geographic extent that can be covered at a given date instead of the revisit frequency. It is the option followed with the Indian AWIFS instrument and the Disaster Monitoring Constellation (DMC) with respective swaths of 740 and 660 km (Duveiller, 2011).

An other step has been done by the ESA Earth observation program for COPERNICUS and the associated Sentinels missions. COPERNICUS aims at providing reliable and up-to-date environmental information that are crucial notably to understand how our planet and its climate are changing, the role played by human activities in these changes and how these will influence our daily lives in order to allow decision makers, businesses and citizens to take the right action.

Each Sentinel mission (6 missions are scheduled) is based on a constellation of two satellites to fulfill revisit and coverage requirements, providing robust datasets for Copernicus Services (table 1). These EO data are complemented with *in-situ* sensors (ground-based stations, airborne and maritime sensors). ESA uses its multi-mission ground systems to acquire, process, archive and distribute data from other satellites - so called Third Party Missions. Data from these 3rd party missions are also included within the scope of COPERNICUS and are distributed under specific agreements with the owners or operators of those missions (table 2). 3rd party missions

complement Sentinels missions by providing higher resolution imagery and/or augmenting Sentinels revisit capabilities.

The first Sentinel satellite (Sentinel-1A, a C-band sensor), successfully launched in April 2014, paves the way to a new era in earth observation. Sentinels products will not only combine a good spatial resolution with a good revisit time but will be provided on a free and open basis. A particular attention has been also put on compatibility between sensors from the program but also between EO programs developed in other part of the world. Scientists from Sentinel-2 and Landsat-8 have for example been working together to make their data compatible and to develop joint archives. Satellite data products could be significantly improved if these were not limited to individual sensors but would combine complementary platforms across space agencies and sensor types (Butler, 2014).

Let's note that with passive sensors, even if the revisit time is short, the number of valid observations is sometimes limited due to cloud conditions. For example, the tandem of satellites of the forthcoming Sentinel-2 mission should ensure a 5 days revisit time (1-3 days in case of emergencies such as floods and earthquakes by rolling and tilting the satellite). However, the maximum effective coverage time in summer considering cloud cover can reach up to 85 days in Europe (Figure 4).

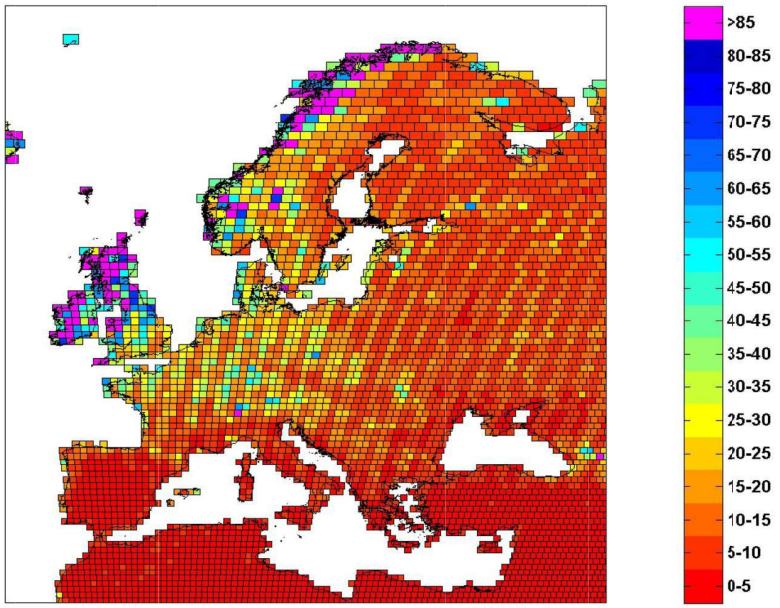


Figure 4.- Maximum effective coverage time (number of days) for Sentinel-2 tandem of satellites (<15% cloud cover, 68% confidence) (Martimort, 2009)

Active sensors, due to their ability to penetrate clouds and obtain imagery regardless of cloud conditions (Moran *et al.*, 1997), are not affected by this restriction of valid observations.

In the frame of the assimilation of remotely-sensed information in crop growth models the required spatial and temporal resolution, independently from the accuracy of this information, will vary according to different parameters.

Most of the models are crop specific and therefore, in order to retrieve crop specific information from the remotely-sensed information, high resolution images are required to avoid problems of mixed pixels. The maximum spatial resolution will depend of the structure of the agricultural landscape and the size of fields for the considered crop. In case of very fragmented landscape with small fields, the use of high/medium resolution images are compulsory.

The requested temporal resolution will be different according to the selected assimilation strategy and according to the considered assimilated variable. The forcing assimilation strategy requires for example remote sensing information available at the same time step than the model or at least with a frequency allowing interpolation. Besides the frequency, the period of availability of remotely-sensed data has an influence on assimilation performances. Dente *et al.* (2008), assimilating LAI observations derived from MERIS in CERES-wheat model, have for example showed that the lack of one or two remotely-sensed LAI observations doesn't affect significantly the assimilation process as far as the remaining observations cover well the most important wheat phenological period i.e. when LAI reaches the maximum value. In other words, the lack of information (LAI observations) at the moment of maximum development of the wheat crop (i.e. stem elongation and heading stage) significantly increases the errors on yields estimation.

Table 1.- Main Sentinel missions /instruments characteristics (adapted from Berger *et al*., 2012)

	Sentinel-1	Sentinel-2	Sentinel-3					Sentinel-5p	Sentinel-4	Sentinel-5
Launch A-unit/B-unit	2014/18 months after A-unit	2015/18 months after A-unit	2015/18 months after A-unit					?	2019 — embarked on MTG-S1 and MTG-S2	2020 + — embarked on Metop-SG
Design lifetime per unit	7.25 yrs (consumables for 12 yrs)	7.25 yrs (consumables for 12 yrs)	7.5 yrs (consumables for 12 yrs)					7 yrs		
Orbit	Sun-sync, 693 km/incl. 98.18/LTAN 18:00	Sun-sync, 786 km/LTDN: 10:30	Sun-sync, 814.5 km, LTDN: 10:00					Sun-sync, 824 km, LTDN: 13:30	GEO	Sun-sync — under definition
Instrument	C-band SAR	MSI (multi-spectral-instrument)	OLCI (ocean and land colour instrument)	SLSTR (sea and land surface temperature radiometer)	SRAL (Sentinel-3 Ku/C radar altimeter)	MWR (microwave radiometer)	POD (precise orbit determination)	TROPOMI (UV–VIS–NIR–SWIR (UVNS) pushbroom spectrometer — in tandem with NPOESS-VIIRS for cloud screening)	UVN + utilisation of the infrared sounder IRS on MTG-S and imager data from the MTG-I platforms for cloud screening	UVNS + utilisation of TIR data from IR sounder and imagery data for cloud screening on Metop-SG
Coverage	Global/20 min per orbit	All land surfaces and coastal waters + full med. sea between: − 56 and + 84° latitude, 40 min imaging per orbit	Global	Global	Global	Global	Global	Global	Europe	Global
Revisit	12 days (6 days for A- and B-units)	10 days (5 days for A- and B-units)	< 4 days (< 2 days for A- and B-units)	< 4 days (< 2 days for A- and B-units)	27 days			Daily	60 min (goal: 30 min)	
Spatial resolution/swath width	Strip mode: 5 × 5m/80 km interferometric wide-swath mode: 5 × 20m/250 km (standard mode) extra-wide-swath mode: 20 × 40 m/400 km wave mode: 5 × 5m/20 × 20 km	Depending on spectral band 10–20–60 m/290 km	300 m/1270 km (with 5 westward tilted cameras to avoid sun-glint, fully within SLSTR nadir and oblique swath)	500 m (VIS, SWIR) 1 km (MWIR, TIR)/1675 km (nadir)/750 km (backwards), nadir and 55° backwards viewing	> 2 km (centred nadir and within OLCI and SLST swath)	20 km (centred nadir and fully collocated with SRAL)		7 km/~ 2100 km	8 km/N/S FOV: 3.65°	
Spectral coverage/resolution	5.405 GHz — VV + VH, HH + HV	13 spectral bands: 443 nm–2190 nm (incl. 3 bands at 60 m for atmos. corr.)	21 spectralbands: 400 nm– 1020 nm	9 bands: (0.55–12 µm — new bands at 1.3 and 2.2 µm)	Ku/C-band altimeter	23.8/36.5 GHz	GPS, LRR and DORIS	270–495 nm 710–775 nm 2314–2382 nm/0.25 nm to 1.1 nm (depending on the band)	305–400 nm 400–500 nm 750–775 nm/0.12 nm to 0.5 nm (depending on the band)	
Radiometric resolution/accuracy	1 dB (3 s)	12 bit/ < 5%	2% absolute	TIR NEdT: 50 mK	Total range error: 3 cm	3 K absolute (0.6 K relative)	3 cm final accuracy	/2%		

Table 2.- Main operational and future SAR and optical Copernicus contributing missions

SAR platforms / sensors								
System	**Country**	**Bands**	**Incidence**	**polarisation**	**Spatial resolution (m)**	**Swath width (km)**	**Temporal resolution (d)**	**Operationnality**
RADARSAT-2	Canada	C	20-50°	HH, VV, HV, VH	3-100	20-500	2-3	2007-now
TerraSAR-X	Germany	X	20-60°	HH, VV, HV, VH	1-16	5-150	2.5-11	2007-now
COSMO-SKYMED	Italy	X	20-50°	HH, VV, HV, VH	1-100	3040	1-16	2007-now
PAZ	Spain	X	20-60°	HH, VV, HV, VH	1-18	10-100	15+2/11	Scheduled in 2015
Optical platforms / sensors								
System	**Country**	**Bands**			**Spatial resolution (m)**	**Swath width (km)**	**Temporal resolution (d)**	**Operationnality**
PROBA-V	Belgium	VIS, NIR, SWIR			100-333	500	1	2013-now
RapidEye	Canada	VIS, Red edge, NIR			6.5	77	1-5.5	2009-now
Landsat 8	USA	VIS, NIR, SWIR, (PAN)			15-100	185	16	2013-now
DMC	Internation. consortium	VIS, NIR			30-40	600	14	2003-now
MODIS	USA	VIS, NIR, MIR, TIR			250-1000	2330	1-2	1999-now

Remotely-sensed biophysical variables

The signal that is recorded by remote sensing instruments is driven by physical processes governing the radiative transfer within the soil, canopy and the atmosphere. The interaction of radiation with canopies and soils depends on the optical, thermal or dielectric properties of the elements as well as on their number, area, orientation and position in space. Therefore remote sensing allows deriving directly only canopy or soil primary biophysical variables that can be merged with model simulations (Baret, 2000).

With a variable degree of success and at different spatial and temporal resolutions, passive or active remote sensing has been indeed used to estimate soil and crop characteristics such as leaf area index (Myneni *et al*., 2002; Haboudane *et al*., 2004; Bacour *et al*., 2006 ; Baret *et al*., 2007 ; Wu *et al*., 2007, Verger *et al*., 2008; Nguy-Robertson *et al.*, 2014; Delegido *et al.*, 2015), biomass (Di Bella *et al.*, 2005 ; Beeri *et al.*, 2007; Liu *et al*., 2010; Claverie et al., 2012) , chlorophyll contents (Haboudane *et al*., 2002 ; Tilling *et al*., 2007, Zhang et al., 2008), evapotranspiration (Gómez *et al.*, 2005; Mutiga *et al*., 2010) or soil moisture (Ahmad *et al*., 2010, Gherboudj *et al*., 2010; Fang and Lakshmi, 2014).

Each spectral domain is sensitive to particular canopy or soil characteristics. The reflective optical domain (400-2500 nm) will provide estimates of canopy structural variables and biochemical composition (pigment, water...). Thermal infrared domain and passive µ-wave will depend on the surface temperature and canopy structural variables. The active microwave domain will provide information on soil roughness, moisture, as well as canopy structure and water content. Therefore, the combined use of several spectral domains is likely to be the only way to extract the maximum amount of information on canopy or soil biophysical characteristics (Jensen, 1983; Baret, 2000). The main primary and secondary variables (according to Baret, 2000) are listed in table 3.

LAI is probably one the most commonly derived variable. The Leaf Area Index (LAI) of a plant canopy is a quantitative measure of the area of green leaf material present in the canopy per unit ground surface (Gobron and Verstraete, 2009). More specifically, it is defined as one-half of the total green leaf area per unit of horizontal ground surface area (Chen and Black, 1992). The LAI is an important vegetation biophysical variable widely used for crop growth monitoring and yields estimation or land surface process simulation (Clevers and van Leeuwen, 1996; Xiao *et al*., 2011) but also for parameterization of climate models (Buermann *et al*., 2001; Gobron, 2008). LAI is a good indicator of crop status and is closely linked to several other crop and soil variables such as biomass, yield, crop nitrogen uptake, nutrition status and water stress occurrence (Casa *et al*., 2012).

LAI is operationally estimated from remotely-sensed optical imagery at global scale for daily to monthly periods in the frame of several initiatives such as MODIS from NASA and in the framework of the global land component of the COPERNICUS Initial Operations (GIO) providing a series of bio-geophysical products on the status

and evolution of land surface at global scale. Currently operationally available on a daily basis from SPOT-VGT at 1 km resolution, a GIO LAI product (derived from PROBA-V) with a higher spatial resolution (1/3 km) is expected for the beginning of 2015. Several global datasets of LAI and fAPAR were produced using NOAA/AVHRR (Sellers *et al.*, 1994; Mynemi *et al*., 1997; Los *et al.*, 2000, Masson *et al.*, 2003), SPOT/VEGETATION (Deng *et al*., 2006; Baret *et al*., 2007), ENVISAT/MERIS (Bacour *et al.*, 2006) or TERRA+AQUA/MODIS (Mynemi et al., 2002; Yang *et al*., 2006). Validation and inter-comparison of such products reveals that they do not necessarily agree with respect to the spatial heterogeneity and inter-annual variability (Garrigues *et al.*, 2008).

These global scale LAI products are unfortunately not crop specific as their spatial resolution, selected to favour the observation frequency, is too coarse which limits their use for agricultural monitoring. In this context, a current trend aims at the establishment of frameworks to retrieve spatially and temporally consistent crop specific information from satellite imagery that can be coupled with crop specific growth models at regional scale (e.g. Duveiller, 2011; Bériaux, 2011).

Remote sensing can be also used to retrieve phenological (Xin *et al*., 2002; Curnel and Oger, 2006b; Jin and Eklundh, 2014) and management information such as mowing dates (Herold *et al.*, 2000;Courault*et al.* 2010, Hadj Said *et al*., 2011; Voormansik *et al*., 2013; Dusseux *et al*., 2014). Knowledge of plant phenology is essential for most agroecosystem models since it governs partitioning of assimilates (Dorigo *et al*., 2007). Therefore a precise knowledge of the phenological status of the plants will greatly improve the results obtained by agroecosystem models (Delécolle *et al.,* 1992, Jégo *et al.*, 2012).

Various algorithms have been developed to retrieve biophysical and biochemical variables from reflectance data or backscattering coefficient (see Dorigo *et al*. (2007) or Baret and Buis (2008) for a review).

Roughly, these algorithms can be subdivided in two main categories: *statistical* and *physical approaches.* To face some limitations of both approaches, a kind of intermediate approach has also emerged with the development of semi-empirical models.

Table 3.- Estimation of canopy, leaf or soil biophysical variables as a function of the spectral domain used. The level of accuracy and robustness of the estimation is indicated by the "+" ("++++" accurate and robust; "-" no estimates possible). Secondary biophysical variables are also indicated (Baret, 2000)

	Biophysical Variables	Spectral domains				
		Visible Near Infrared	Near Infrared Short Wave Infrared	Thermal Infrared	Active µ-wave (radar)	Passives µ-wave
Canopy structure	*LAI*	+++	+++	+	++	+
	Leaf orientation	+++	+++	+	+	+
	Leaf size and shape	+	+	+	+	+
	Canopy height	-	-	-	++	-
	Canopy water mass				+++	+++
Leaf characteristics	Chlorophyll content	+++	-	-	-	-
	Water content	-	+++	-	+++	+++
	Temperature	-	-	++++	-	++
Soil characteristics	Surface soil moisture	-	+	+	+++	+++
	Roughness	+	+	-	++	+
	residues	+++	++	-		-
	Organic matter	++	++	-	-	-
	Soil type	++	++	+		
Secondary variables	*fCover*	++++	++++	++	++	+
	fAPAR	++++	++++			
	albedo	++++	+++			
	Long wave flux	-	-	++++	-	-

Statistical approaches

Statistical approaches search for a consistent relationship between the spectral signature of an object, in general the leaf or canopy reflectance, or its dielectric behaviour and the biophysical or biochemical variable of interest. To establish such relationships, spectral, radar, biophysical or biochemical measurements have to be taken under varying field or laboratory conditions and for different plant species or cultivars and phenological stages. The accuracy of the measurements and the range of conditions considered for the development of a relationship determine to a larger extent the validity and the portability of the relationships.

The traditional way of linking remotely-sensed information is by simple (Clevers, 1989; Svoray and Shoshany, 2002; Dente *et al.*, 2008) or multiple regression (Bsaibes *et al.*, 2009) techniques but recently more sophisticated statistical approaches such as partial least square (Hansen and Schjoerring, 2003; Darvishzadeh *et al.*, 2008) and artificial neural networks (Verger *et al.*, 2008; Bsaibes *et al.*, 2009) have been introduced.

The spectral information (reflectance, transmittance or absorptance) is rarely used directly to construct the response function to biophysical and biochemical variables. Manipulations are frequently used to enhance subtle spectral features and to reduce undesired effects caused by variations in soil reflectance, sun and view geometry, atmospheric composition, and other leaf or canopy properties. As specified by Dorigo

et al. (2007) standard manipulations includes normalization (Chappelle *et al.*, 1992), logarithmic transformation (Jacquemoud *et al*., 1995 ; Yoder and Pettigrew-Crosby, 1995), continuum removal (Kokaly and Clark, 1999; Mutanga *et al.*, 2004) and the calculation of first or second derivatives (Huang *et al*., 2004).

However, the most widespread method used to reduce background effects and enhance spectral features is to express spectral reflectance in a combination of a limited number of (transformed) spectral bands, to create what is known as a vegetation index (VI). Most VIs concentrate on the red-edge region, which is the region between 680 and 800 nm that is characterized by a sharp decrease of chlorophyll absorption from maximum absorption around 680 nm to almost zero absorption at 800 nm. This makes this wavelength range very well suited to study vegetation characteristics (Baret *et al.,* 1992).

In the agricultural monitoring context, VIs have been widely used as proxy characterizing the growing conditions and have been therefore linked to crop yield (Tucker *et al*., 1980; Wiegand *et al.*, 1991; Ren *et al*., 2008; Vescovo and Gianelle, 2008; schut *et al*., 2009; Becker-Reshef *et al*., 2010; Cicek *et al.*, 2010).

Though the statistical approaches are rather easy to implement, a good validated relationship between the VIs and the variables of interest is compulsory which implies expensive and time-consuming measurements if this relationship wants to be valid for a wide range of species, canopy conditions, and view constellations. These limitations have induced the development of *physical / mechanistic approaches.*

Mechanistic approaches

Remote sensing data result from radiative transfer processes within canopies that depend on canopy variables, and observational configuration (wavelength, view and illumination conditions). Radiative transfer models summarize our knowledge on the physical processes involved in the photon transport within the vegetation canopy or atmosphere, and simulate the radiation field reflected or emitted by the surface for given observational configuration, once the vegetation and the background soil as well as possibly the atmosphere are specified. Retrieving canopy characteristics from the radiation field as sampled by the satellite sensor needs to “invert” the radiative transfer model (Baret and Buis, 2008). The mechanistic / *physical approach* consists therefore of inverting a radiative transfer model for the estimation of leaf and canopy properties. Canopy variables include the variables of interest for the agricultural monitoring such as *LAI*, and the other variables that are not of direct use for the agricultural monitoring but that influence the radiative transfer, such as soil background properties. The causal relationship between the variables of interest and remote sensing data corresponds to the forward (or direct) problem (figure 5). Conversely, retrieving the variables of interest from remote sensing measurements corresponds to the inverse problem, i.e., developing algorithms to estimate the variables of interest from remote sensing data as observed in a given configuration. Prior information on the type of surface and on the distribution of the variables of interest can also be included in the retrieval process to improve the performances (Baret and Buis, 2008).

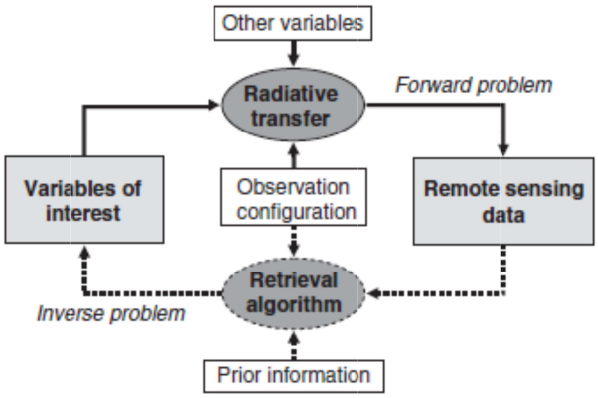


Figure 5.- Forward (solid lines) and inverse (dashed lines) problems in remote sensing (Baret and Buis, 2008)

Radiative transfer models can be envisaged at two different levels: leave and canopy transfer models. More details concerning radiative transfer models can be found in Curnel (2014).

Semi-empirical models

Semi-empirical models are based on theoretical principles and statistical relationships but are extremely simplified. They use indeed only a few parameters and can be easily inverted to retrieve biophysical variables.

An example of semi-empirical model used with optical remote sensing systems is the Monteith's efficiency model (Monteith, 1972; Kumar and Monteith, 1981). Monteith's efficiency model uses measured data of global radiation and reflectance measurements for derivation of *f*APAR to model dry matter production. An other example is the CLAIR model (Clevers 1988, 1989). In this model the WDVI (weighted difference vegetation Index) is ascertained as a weighted difference between measured near-infrared (NIR) and red reflectances, assuming that the ratio of NIR and red reflectances of bare soil is constant. The WDVI is subsequently used for estimating LAI according to the inverse of an exponential function.

For radar remote sensing systems, a commonly used model (Bouman 1991, 1992; Prévot *et al.*, 1993; Inoue *et al.*, 2002, Svoray and Shoshany, 2002; Lucau-Danila, 2008) is the Water Cloud Model (Attema and Ulaby, 1978).

Estimation of uncertainties on remotely-sensed biophysical variables

The establishment of a statistical relationship or the development of an operational algorithm for retrieving a biophysical parameter should take into account the uncertainties involved in the modelling phase as well as the uncertainties in the input data.

Accounting for these different sources of uncertainties is rather complicated. For this reason, the uncertainty on remotely-sensed biophysical variables is most of time assessed through the comparison with measurements of these biophysical variables on field.

This comparison with field measurement data is relatively straightforward when biophysical variables are derived from (very) high resolution data, i.e. when pixel size is lower than the size of the fields where the *in situ* measurements are realised. When biophysical variables are derived from medium resolution data (such as MODIS data), the comparison and subsequently the uncertainty assessment become more challenging. To face this problem Yao *et al.* (2008) suggest, with a view to validate LAI products derived from MODIS at 1 km resolution, to use high resolution imagery as scaling transfer bridge. They used a Landsat-TM imagery to retrieve 30 m resolution LAI maps that they have validated with field measurement data. The 1 km coarse resolution LAI maps were then generated by degrading the Landsat-TM LAI maps through the nearest neighbour re-sampling and finally compared with the LAI retrieved from MODIS data.

When comparing remotely-sensed products with field measurements, it is important to be sure that the biophysical variables derived from remote sensing are strictly comparable with biophysical variables. Duveiller (2011), whose initial objectives were to retrieve a crop specific (winter wheat) LAI from 250m MODIS data on the basis of the inversion of a radiative transfer model, found that the biophysical variable retrieved was closer a green area index (GAI) than from a Leaf area index.

The knowledge of the level of uncertainty on a remotely-sensed biophysical is very important as it provides the degree of confidence that we can have on these data. In the frame of assimilation, the level of uncertainty of assimilated observations is often used by different assimilation techniques (such as e.g. ensemble Kalman filter).

To take the example of Leaf Area Index, one of the biophysical variables considered in this thesis, Yi *et al.* (2008) have for example retrieved wheat LAI from 500m reflectances MODIS products through the inversion of one-dimensional (1-D) radiative transfer model. A poor correlation was observed between modelled and measured LAI (varying between 1 and 6) when daily C4 MODIS data are used. The observed RMSE in this situation is equal to 1.1. The use of daily C5 MODIS data, presenting an improved atmospheric correction algorithm compared to the daily C4 MODIS data, allows improving the estimation of LAI. The RMSE between the modelled and observed LAI data decreases up to 0.7. For both C4 and C5 collections, the LAI tends to be overestimated when the sensor is operated with a large view zenith angle in the backscattering direction. In the same way, Duveiller (2011) has observed a RMSE of 0.58 and 0.63 between remotely-sensed (based on SPOT images) and observed GAI respectively when the remotely-sensed GAI is estimated on the basis of an empirical model or on the basis of a physical approach from a neural network inversion of PROSAIL radiative transfer model. He has also generated smoothed GAI estimations from 250m MODIS reflectance and has compared these estimations with upscaled punctual SPOT GAI estimations. This comparison showed the performance is related to the date of acquisition. According

to Duveiller (2011) the GAI is harder to estimate when it reaches its peak value (RMSE close to 1). However, the dispersion of the estimation (coefficient of variation) was observed as relatively stable all along the season suggesting that the overall performance of the MODIS GAI is stable. Purity of pixels has also an impact on the observed uncertainty: mixed pixels tend to present higher RMSE than pure pixels.

Crop growth models

By the end of the 1960s, computers had evolved sufficiently to support the first attempts to synthesize detailed knowledge on plant physiological processes, in order to explain the functioning of crops as a whole. These first (heuristic) models were meant to increase the understanding of crop behaviour by explaining crop growth and development. Simulation models are powerful tools for testing our understanding of crop performance by comparing simulation results and experimental observations, thus making explicit gaps in our knowledge. Experiments can then be designed to fill these gaps. Well-tested crop growth models can be used to explore, in a quantitative way, the relative importance of crop characteristics, such as physiological and morphological traits as well as environmental characteristics, in a manner that would not be possible in field experimentation.

Over the years, new insights and different research questions motivated the further development of simulation models (Bouman *et al*., 1996a). Progressively, more application-oriented models were developed to meet various demands and objectives formulated by farmers, scientists and policy makers (Murthy, 2004).

Crop growth models meet for example demands of farmers for advisory systems. Crop growth models have been indeed used in numerous studies to help farmers in day-to-day, i.e. tactical decision making. They have been used to investigate the effects of management options such as sowing time, plant population density, irrigation timing and frequency and fertilizer applications in different environmental conditions on long-term mean yield and yield probability (e.g. Aggarwal *et al.,* 1994; Aggarwal and Kalra, 1994). More recently, results of crop growth models have been applied to tactical decision making using knowledge based such as expert systems and decision models. For example, the CERES models have been integrated in the Decision Support System for Agrotechnology Transfer (DSSAT) which allow users to combine the technical knowledge contained in the crop growth model with economic considerations and environmental impact evaluations with a view to facilitate economic analysis and risk assessment of farming enterprises (Jame and Cutforth, 1996, Jones *et al*., 2003). An other example is the GrazeGro herbage growth model included in the GrazeMore decision support system developed to provide to dairy farmers, consultants and advisors a tool for improving management and utilization of pasture in European dairy production systems (Barrett *et al*., 2005).

Demands of policy makers for policy management tools have also geared the evolution of the models towards tactical and strategic support systems. The issues range from global (a recent concern is for example the study of the impact of climate changes on crops (Kang *et al*., 2009; Guo *et al*., 2010; Supit *et al*., 2010; White *et al*.,

2011; Supit *et al*., 2012)) to field level (effect of crop rotation on soil quality…) issues (Murthy, 2004). The main use of crop growth models as management tools concerns the yields predictions (see 'agricultural Monitoring at global and local scale' section).

Most of the crop growth simulation models are mechanistic i.e. that they attempt to explain not only the relationship between parameters and simulated variables, but also the mechanism of the described processes. Most of the models have been initially developed and evaluated at field scale but it has become a common practice to apply them from the field to a (supra-) national scale (Palosuo *et al.*, 2011).

Currently, four main process-oriented approaches are used to simulate crop growth: (1) Radiation use efficiency (RUE) with e.g. CERES family (Uehara and Tsuji, 1993) and STICS (Brisson *et al*., 2003), (2) Transpiration use efficiency (TUE) with e.g. GLAM (Challinor *et al*., 2004), (3) Combination of RUE and TUE with e.g. CropSyst (Stöckle *et al*., 2003) and (4) Photosynthesis e.g. CGMS/WOFOST and all the SUCROS family models (van Ittersum *et al.*, 2003).

Palosuo *et al.* (2011) grouped 8 widely used, easily accessible and well-documented crop growth simulation models according to 5 major crop growth processes namely leaf area development and light interception, light utilization, crop phenology, soil moisture dynamics and nitrogen balance. These 8 crop simulation models included in the comparison are APES (Donatelli *et al.*, 2010), CROPSYST (Stöckle *et al.*, 2003), DAISY (Abrahamsen and Hansen, 2000; Hansen *et al.*, 1990; Hansen, 2000), DSSAT (Ritchie and Otter, 1985; Hoogenboom *et al.*, 2003; Jones *et al.*, 2003), FASSET (Olesen *et al.*, 2002a,b; Berntsen *et al.*, 2003), HERMES (Kersebaum, 1995; Kersebaum and Beblik, 2001; Kersebaum (2007), STICS (Brisson *et al.*, 1998, Brisson *et al.*, 2003) and WOFOST (van Diepen *et al.*, 1989 ; Supit *et al.*, 1994 ; Boogaard *et al.*,1998).

Uncertainty in crop growth models

Over the past decades, modelling biophysical processes in general has benefited from significant developments, including dramatic growths in computational power, ever increasing availability of distributed observations, and improved understanding of the physics and dynamics of the natural systems. This led to the building of higher levels of complexity into models, and an advance from lumped, conceptual models toward semi-distributed and distributed physics-based models. Paradoxically, while these advances reflect our growing understanding, they have also increased the need for concrete methods to deal with the increasing uncertainty associated with the models themselves, and with the observations required for driving and evaluating the models.

It is now being broadly recognized that proper consideration of uncertainty in simulations and predictions is essential for both purposes of research and operational modelling. The value of a prediction to relevant decision making processes is limited if reasonable estimates of the corresponding simulation / prediction uncertainty are not provided (Liu and Gupta, 2007).

There are several alternative taxonomies of uncertainty sources in distributed models (van Oijen and Ewert, 1999; Katz, 2002; Walker *et al.*, 2003; Rubarenzya *et al.,* 2007). According to Katz (2002), three major sources of uncertainty can be identified: *(1)* model structure, *(2)* measurement error and *(3)* natural variability.

Model structure errors result from inaccurate treatment of modelled processes, inexact numerical schemes and inadequate resolutions (Bert *et al*., 2007). This source of uncertainty reflects the modeler's limited understanding of the model. The uncertainty related to model structure is the most difficult source of uncertainty to quantify (Chatfield, 1995).

Measurements errors arise when attempting to measure an unknown physical variable. These errors include both those of a random nature whose magnitude reflects the precision on which the measurements are based and those due to systematic error (Katz, 2002).

Natural variability is a major source of uncertainty for most environmental variables (state variables / initial conditions) as they exhibit systematic differences over space and time, as well as inherent randomness. Most variables relevant to crop models (e.g. soil nitrogen or water content at sowing) have high spatial and temporal variability which makes the accurate estimation of these quantities difficult and costly. Values for a quantity of interest may be available only for a limited number of sites or conditions. Consequently, many inputs to crop models are often derived from indirect measurements, estimations from other variables (e.g. solar radiation) or measurements or estimations at sites other than the simulation site must be used to fill data gaps, or be considered as representative. The way in which these derived variables are computed may introduce substantial amounts of uncertainty. Both measurement uncertainty and natural variability uncertainty are associated with incomplete or imperfect knowledge of model inputs such as empirical quantities, initial conditions and boundary conditions (Bert *et al*., 2007). According to Palosuo *et al.* (2011), a judicious use of crop models implies calibration for their most important parameters before being able to apply them with confidence and a minimal calibration for phenological dates is not sufficient to generate robust crop cultivar-specific yield estimates for different environments.

Assimilation of remotely-sensed variables, such as for example LAI, aims mainly at correcting consequence of natural variability on modelling.

A good knowledge of the different uncertainty sources and of their respective importance turns out to be very important in order to be able to improve models capabilities. An accurate knowledge of the uncertainty level of the most sensitive model inputs or of the model state variables is for example a key information in probabilistic assimilation methods such as the Ensemble Kalman filter (EnKf). The information is also important in calibration strategy related method to identify parameters or initial conditions to calibrate/adjust and subsequently to identify their range of possible variation.

Relative importance of the different sources of errors is also function of the model and of the model application level.

Paluoso *et al.* (2011) comparing performances of eight crop growth models (previously mentioned in 'crop growth models' paragraph) for winter wheat during 49 growing seasons across Europe observed that some models performed better than others in estimating grain yield and other crop variables, but none could unequivocally be termed robust and accurate in terms of yield prediction across different environments and for different crop cultivars. Good prediction of crop yield for some models came at the cost of overestimating or underestimating harvest index or total biomass. Other models showed a distinct bias towards under- or overestimating yields. Paluoso *et al.* (2011) also observed that mean model predictions are in relatively good agreement with observed yields. This result was also observed for the kind of study applied on winter barley (Rötter *et al*., 2012). The authors of both studies support therefore both the use of multi-model ensembles rather than relying on single models.

De Wit and van Diepen (2007) stated, on the basis of the results of Pellenq and Boulet (2004), that parameter uncertainty represents likely the largest source of uncertainty in crop modelling for point application where meteorological conditions are relatively well known. On the other hand, several studies (Easterling *et al.*, 1998; Mearns *et al*., 2001) also demonstrated that when aggregating model output to (higher) regional scales, uncertainty in the input forcings become dominant over uncertainty in parameters (soil parameters in these studies). This was also confirmed by Lobell *et al.* (2007).

Several studies analyzed the effects of using spatially aggregated climate data on simulated yields by crop models (van Bussel *et al.*, 2011). It has been shown (Easterling *et al*., 1998) that agreement between simulated and observed yields for wheat and maize improves considerably when climate data were disaggregated from 2.8° x 2.8° to approximately 1° x 1° resolution, with disaggregation below 0.5° showing no further improvement. Disaggregation of soil data showed little effect on model results (Easterling *et al*., 1998). Olesen *et al.* (2000) and De Wit *et al.* (2005) reported as well that a grid cell size of 0.5° x 0.5° is an appropriate size to simulate yields at regional scale. However, other studies showed (e.g. Baron *et al.*, 2005; Wassenaar *et al.*, 1999) that especially in regions where water-limitations play a role, model results for wheat and millet yields were sensitive to aggregation of precipitation and soil data. Van Bussel *et al.* (2011) have shown that the aggregation of available phenological information improves the spatial data coverage of a region, but reduces spatial heterogeneity. Importantly, the use of aggregated weather data and emergence dates for simulations of crop phenology has little effect on the aggregated predicted phenological events. Results of this study suggest that, for the model used (AFRCWHEAT2), spatially aggregated weather data and emergence dates to simulate the length of the growing season for winter wheat is justified for grid cells with a maximum area of 100 km.

For Hansen and Jones (2000), the availability of input data of adequate quality and spatial coverage is perhaps the most serious practical constraints to application of crop models at regional or larger scale.

Assimilation of remotely-sensed biophysical variables

Reducing uncertainty in biophysical models can be mainly reached either by the acquisition of more informative and higher quality data (e.g. through the development of improved measurements and observation networks), by the development of improved models through better representation of physical processes and/or the use of better mathematical techniques or by the development of efficient and effective techniques allowing a better extraction and assimilation of all the available data.

The last approach presents obviously room for improvement. There is indeed clearly a need for robust techniques that effectively and efficiently use information from observation into models to produce improved estimations and predictions. Remote sensing, through its ability to provide synoptic information on growth conditions over large geographic extent and in near real-time, is one of the most frequent data source used for assimilation purposes. The union of remote sensing and cropping systems modelling has been investigated for a diverse set of agricultural applications including regional crop monitoring and yield prediction (Prévot *et al.*, 2003; Doraiswamy *et al.,* 2004; Launay and guérif, 2005; De Wit and van Diepen, 2008; Dente *et al*., 2008; Courault *et al.*, 2010), precision crop management (Jones and Barnes, 2000; Seidl *et al*., 2004) and evapotranspiration mapping (Olioso *et al.*, 2005).

Data assimilation can be defined as procedures aiming to produce physically consistent representations or estimates of the dynamical behaviour of a system by merging the information present in imperfect models and uncertain data in an optimal way to achieve uncertainty quantification and reduction (Liu and Gupta, 2007). Observation and model estimates provide different kinds of information and in different time and spatial scales; therefore, when used together, they can provide a level of accuracy that cannot be obtained when used individually (Alavi *et al*., 2009). It is supposed that a better estimation of model state or input variables can improve the estimations or predictions. Various assimilation strategies have been developed, with various levels of complexity within each strategy. The different ways to combine a crop model with radiometric observations (ground measurements or satellite data) were initially described by Maas (1988) and their classification was revisited by Delécolle *et al*. (1992).

These different assimilation techniques have been recently summarized in Dorigo *et al.* (2007). Roughly, three main categories of data assimilation strategy can be identified namely the *forcing strategy*, the *calibration strategy* and the *updating strategy*.

These three strategies are schematically summarized in figure 6.

The *forcing* strategy aims at replacing a model state variable, initial condition or input variable with data derived from remote sensing. It notably supposes that the model is for example not able to simulate correctly the state variable and implies the assimilated remote sensing data to be available at the same time step than the

model one or at least with a frequency allowing interpolation (Thorp *et al*., 2010). Model initial conditions often used in forcing strategies are emergence (Casa *et al*., 2012) or mowing (Courault *et al.*, 2010) dates.

With the *calibration* strategy, model parameters or initial conditions are adjusted to obtain an optimal agreement between the simulated and the observed state variables, which characterise the system behaviour. The sensitive and uncertain model parameters or initial conditions are calibrated either manually or automatically by running the model with various combinations of parameter values within realistic ranges (Dorigo *et al*., 2007). This strategy is certainly the most frequently used (Maas, 1988; Bouman, 1995; Doraiswamy *et al*., 2004; Dente *et al*., 2008; Jarlan *et al*., 2008; Duveiller *et al*., 2010).

The *updating* strategy consists of the continuously updating of model state variables, whenever an observation is available. This method is based on the assumption that a better simulated state variable at day *t* will also improve the accuracy of the simulated state variable at succeeding days (Dorigo *et al*., 2007). Crop models are still generally deterministic in contrast to models within the fields of oceanography, meteorology and (more recently) land surface hydrology, where probabilistic approaches are generally accepted and advanced algorithms for sequential data assimilation such as nudging, variational methods and (Ensemble) Kalman filtering have been developed. As a result, crop models provide no information on uncertainties of the model states during simulation which is crucial for a successful application of most sequential data assimilation algorithms. In principle, probabilistic methods and data assimilation can be applied to crop growth models as well. Particularly the use of the Ensemble Kalman filter (Evensen, 2007) is interesting for crop models because it combines a probabilistic approach with sequential data assimilation (de Wit, 2007). Ensemble Kalman Filter (EnKF), allowing to account for errors in both simulated and observed state variables, has been used in several recent studies (Pellenq and Boulet, 2004; de Wit and van Diepen*,* 2007; Wu *et al.*, 2012).

Each assimilation strategy has its strengths and its drawbacks. The propagation of errors can be problematic in forcing strategy where the model forgets its own information to follow the observed state variables or initial conditions, including the observations errors. The calibration and updating strategies have more flexibility in assimilating remotely-sensed state variables / initial conditions and their associated errors in the model. If the number of observations is sufficient and their errors relatively small and considering that the physical description of underlying process is an acceptable representation of the natural system, the calibration strategy is expected to give more representative input parameters and improve model's predictions. Processing/computation time is however critical for calibration strategy due to the iterative optimization procedure. The number of recalibrated parameters has to be small compared to the available observations of the considered state variable and to the uncertainty on these observations in order to avoid overfitting problems (Dumont *et al*., 2012). The two other methods need only one run. However, for some updating methods (ensemble Kalman Filter), as some measure of

uncertainty in the model state variables must be propagated through the system, it could be also computationally expensive if the model is complex or the model spatial domain is large. The updating methods also imply most of the time more important modification of source codes as it is necessary to adjust state variables during model run (Dorigo *et al.*, 2007).

Whatever the considered assimilation strategy, improvement of simulation results due to data assimilation is notably influenced by (1) the uncertainties in the model and (2) the frequency and accuracy of the remote sensing observations and of the state variables/initial conditions derived from these remotely-sensed observations.

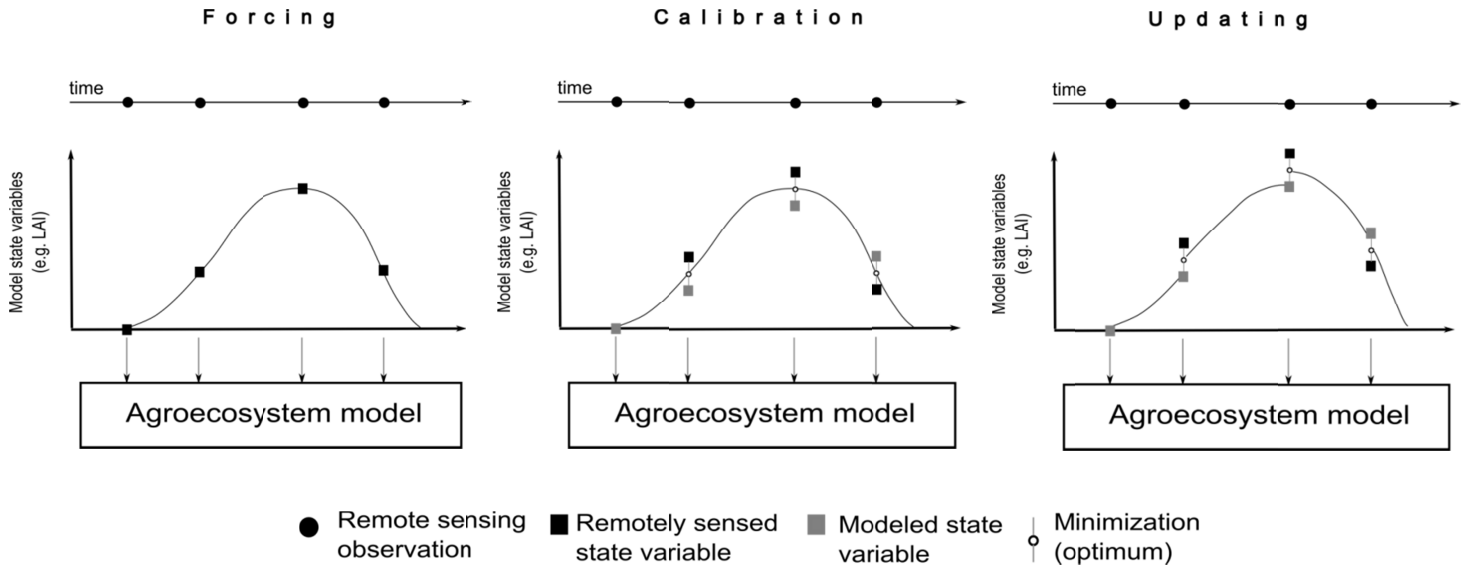


Figure 6.-Schematic representation of different methods for the assimilation of remotely sensed variables in agroecosystem models (adapted from Delécolle *et al.*, 1992)

Scope and objectives

Data assimilation techniques are frequently used to reduce uncertainty in biophysical models. The performance of data assimilation strongly depends on the assimilation protocol and the uncertainty specification (Pellenq and Boulet, 2004). The assimilation strategy, the level of uncertainty on both modelled and observed variables, the availability of the observations assimilated in the model or the objectives pursued are indeed some of the factors to consider.

In this context, the main objective of this thesis is to assess possible improvements for the assimilation of remotely-sensed biophysical variables in crop growth models and to estimate their related errors reduction on modelled yield estimates.

More specifically, the improvements of assimilation of biophysical variables derived from satellite data are studied from different points of view (corresponding to 3 thesis objectives): ***(1)*** the interplay between the accuracy and the frequency of remotely-sensed data, ***(2)*** the assimilation strategies and their implementation, and ***(3)*** the different sources contributing to the uncertainty on remote sensing data and biophysical variables.

(1) Interplay between the accuracy and the frequency of remotely-sensed data

The first objective focuses on the requirements in terms of accuracy and frequency of remotely-sensed data in order to reach a given level of errors reduction on yields estimates thanks to data assimilation.

The assessment of these requirements on observations would imply to have, over the same area of interest, time series of remote sensing images with different temporal resolution and different levels of accuracy. Such an idealistic situation is not realistic from a practical point of view. To cope with this limitation and to systematically investigate the assimilation performances, approaches based on synthetic data called “Observing System Simulation Experiment” (OSS Experiment) are selected.

OSS Experiments consist of identical twin experiments, a widely used technique to assess the impact of data assimilation. This setup consists in considering data simulated with one model run as observations and assimilating them afterwards into another run of the same model with different initial conditions. The convergence of the second run towards the first, defined as the “truth”, can be measured to quantify the data assimilation effectiveness in driving the model with “wrong” initial conditions towards the truth (Raicich and Rampazzo, 2003).

OSS Experiments have been initially developed (Arnold and Dey, 1986, Atlas, 1997) and mainly used (Raicich and Rampazzo, 2003; Reichle *et al*., 2008b; Wei and Malanotte-Rizzoli, 2010) in meteorology and oceanography. Applications of OSSE in agronomy are currently limited (Pellenq and Boulet, 2004, Pauwels *et al*., 2007).

The proposed OSS Experiments aim at assessing, first of all, the improvements on modelled yields estimations that can currently be expected from assimilation with the

available and forthcoming (e.g. COPERNICUS) sensors but can also provide guidelines for future satellite missions.

It is also important to note that OSS experiments, though designed for a specific objective, have also an exploratory finality and hence allow studying the limits of a system.

(2) Assimilation strategies and their implementation

In a second step, the thesis aims at investigating the role of assimilation strategy in the assimilation improvement. Many assimilation strategies have been developed historically for meteorology and later on for oceanography before being extended to other domains such as agronomy. The thesis studies therefore the potentialities and limitations of these assimilation strategies in the specific context of the assimilation in crop growth models. Conclusions of this study are drawn from the OSS Experiments but also more empirically on the basis of experiments based on field observations.

(3) Main sources of uncertainty on remote sensing data and biophysical variables

The third objective, based on field experiments, is to identify the main sources of uncertainties and to assess separately their impact on assimilation performances. Uncertainties on satellite remote sensing data and subsequently on biophysical variables derived from these data, considered as a whole in OSS Experiments, arise indeed from different sources related to sensors' characteristics, atmospheric and meteorological conditions and characteristics of the monitored target.

All the analyses contribute to assess the conditions of success of satellite remote sensing data assimilation in crop growth models.

In order to broaden the scope of the study, both a state variable and an initial condition respectively derived from optical and SAR imagery, are assimilated using three main categories of data assimilation strategy (forcing, calibration, updating).

Factors limiting the efficiency of the assimilation of satellite remote sensing are specific to type of the sensors (e.g. optical, SAR), the characteristics of the considered sensor and of the monitored crop, the assimilated variables but also the assimilation strategy.

Considering all the combinations of these different factors is not conceivable within this PhD. For this reason, only some of them are studied. Two analysis scenarios have been considered (table 4).

Table 4.- Assimilation scenarios considered within the framework of the thesis

Assimilated observation	**Assimilation strategy**	**Crop**	**RS data**
State variable (LAI)	Calibration Updating	Winter wheat	Optical
Initial conditions (Mowing dates)	Forcing	Grassland	SAR

As presented in table 4, the thesis is focused on winter wheat and grasslands as these 2 crops occupy a major place in European and world agriculture.

Wheat is grown on more than 240 million ha, larger than for any other crop, and world trade is greater than for all other crops combined (Curtis, 2002). Winter wheat is the most represented crop In the EU-27 (EUROSTAT, 2010).

Worldwide, grasslands represent indeed 3500 million ha i.e. roughly 26 percent of the world land area and 70 percent of the world agricultural area, and contain about 20 percent of the world's soil carbon stocks (Schlesinger, 1977; Ramankutty e*t al*., 2008; FAOSTAT, 2009).

Assimilated observations are different according to the target crop: Leaf Area Index (derived from optical data) for winter wheat and the mowing calendar derived from SAR data for grassland.

In the thesis, LAI is assimilated according to two assimilation strategies, namely the updating and calibration strategies. These two strategies, well-tailored and frequently used in the frame of LAI assimilation in crop growth models (e.g. Doraiswamy *et al*., 2004; Dente *et al*., 2008; de Wit and van Diepen, 2007; Wu *et al*., 2012), use contrasting approaches. The updating strategy uses indeed a sequential assimilation scheme, correcting model state variables whenever an observation is available, while the calibration strategy uses a variational approach, adjusting model parameters or initial states with a view to obtain an optimal agreement (usually the minimization of a cost function) between the simulated and the observed state variables time series.

Mowing calendar is an important component of grassland management. Mowing date as well as mowing frequency conditions the quantity as well as the quality of the harvested grass (Hermann *et al.,* 2005). As stated by Dusseux *et al.* (2013), practices such as mowings (but also grazing) and their intensity of use have also different environmental impact concerning biodiversity, soil degradation and water pollution. In addition, grassland management practices are spread out within the growing season according to farmers' needs but also climate which induces subsequently a huge uncertainty on the start of modelling. As several mowings can be observed for a same parcel during a growing season, it can induce a huge uncertainty on the total yield at the end of the season. Identification of grassland management is therefore of high importance, especially if we also consider the legislative framework. For example, according to the Decision n° 529/2013/EU of EU legislation (European Commission, 2013), a mandatory accounting on greenhouse gas emissions and removals resulting from activities related to grazing land management is to be phased in between 2013 and 2021. From 2016 to 2018, Member States will be required to report to the Commission on the inventory systems in place or being developed to estimate emissions and removals from grazing land management (Barrett *et al.*, 2014).

As mowings are punctual events characterized by temporal differences due to climate and farm management, the use of multi-temporal image series ideally with a high temporal frequency is therefore required. To face the limited use of optical images in Belgium arising from a frequent cloud cover, mowing calendar is estimated

in the thesis from SAR data which are almost independent of weather and illumination conditions and assimilated using the forcing assimilation strategy.

Researches activities have been divided in 4 main analyses. As aforementioned, two of these analyses are based on synthetic data sets (OSS experiments), the two remaining analyses being based on field campaign measurements synchronously completed with satellite observations.

Outline of this thesis

The core of this thesis (chapters 1 to 4) is based on scientific papers which are at different levels of progress towards being published in peer-reviewed journals. Some have already been published or accepted for publication, while others still need to be submitted.

Chapter 1 presents the results of an Observing System Simulation Experiment aiming to assess the efficiency of 2 assimilation techniques based on recalibration and on Ensemble Kalman filter (EnKf).Efficiency is assessed on the basis of the reduction of the errors on simulated yields by WOFOST model (winter wheat). The assimilated observation is the leaf area index (LAI). The study aims also to assess the uncertainty level and the temporal availability needed on remotely-sensed LAI to reach two given objectives (25 and 50%) in terms of errors on yields reduction.

Chapter 2 presents the results of the assimilation of remotely-sensed LAI derived from MODIS sensor in WOFOST model. One of the objectives is to possibly validate some of the results acquired in chapter 1 but also to stress some of the uncertainty sources on remote sensing data. The study assesses notably the influence of the landscape neighbouring the assimilated pixels on LAI time series and subsequently on assimilation.

Chapter 3 presents the results of the second OSS Experiment focused on grassland. Assimilated variable is the mowing calendar assessed on SAR time series. As in chapter 1, the main objective is to assess the efficiency of the considered assimilation procedure (forcing strategy) for different levels of accuracy and temporal availability on radar data.

Chapter 4 assesses the detection of mowing in grassland on the basis of SAR backscattering coefficients time series including different factors driving the backscattering variability. The analysis also aims at assessing the impact of these sources on backscattering coefficients and *de facto* on assimilation performances.

The main findings and future perspectives are summarized and discussed in the conclusion of this document.

CHAPTER 1 : Assessment of remotely-sensed LAI assimilation potentialities based on an OSS Experiment[1]

1. Introduction

Crop models simulating crop growth under different environmental and management conditions by taking various limiting factors (e.g. soil, weather, water, nitrogen) into account in a dynamic way are good tools for diagnosing crop growth conditions or for predicting yields over large areas (Launay and Guérif, 2005). The success of these crop growth monitoring systems strongly depends on the crop simulation model's ability to quantify the influence of weather, soil and management conditions on crop yields and on the system's ability to properly integrate model simulation results over a range of spatial scales (de Wit and van Diepen, 2008).

Unfortunately, crop growth monitoring systems applied over large areas and relying on a spatially distributed crop growth model are typically confronted with large uncertainty in the spatial distribution of soil properties and initial soil conditions, crop parameters, meteorological forcings (Hansen and Jones, 2000) as well as management practices. Within the crop growth model, this uncertainty influences the simulation of two important physiological processes i.e. the simulation of crop canopy development which determines light interception and the potential of photosynthesis and the simulation of moisture content in the soil determining the actual evapotranspiration and reduction of photosynthesis as a result of drought stress (de Wit and van Diepen, 2007).

Assimilation of remotely-sensed biophysical variables is one of the possible options to reduce this uncertainty in biophysical models (Prévot *et al*., 2003; Launay and Guérif, 2005, de Wit and van Diepen, 2008, Jarlan *et al*., 2008; Quaife *et al*., 2008).

Retrieval of vegetation canopy biophysical properties or soil characteristics from remote sensing is gradually moving from research to operational contexts. The availability of remotely-sensed information at increasingly high spatial and temporal resolution opens up the possibility of developing several applications at a range of scales, from globe to field (Casa *et al.*, 2012).

The estimation of the soil and crop biophysical variables derived from satellite imagery is uncertain and this for several reasons. First of all, the measurement device is not perfect inducing small measurement errors to which is added a most significant noise due to atmospheric conditions. Secondly, the observations models relating satellite observations to the soil or crop biophysical variables (e.g. empirical or radiative transfer models) are not perfect. Third, the spatial and temporal scale on

[1]Adapted from Curnel *et al.*, 2011

which the measurement are made is rarely the one on which the observation is required. Observation models and interpolation of these observations implies hypotheses or requires ancillary information that induces new uncertainties (Pellenq and Boulet, 2004).

Despite these uncertainties, crop variables derived from satellite imagery represents an additional source of information on crop growth that could be favourably incorporated in crop growth models. Indeed, if simulation or observation by itself is not able to provide an accurate description of the considered system, combining them should improve both predictive and retrospective models capabilities. Ground data can be also used in this context but it can hardly compete with the synoptic spatial and temporal coverage provided by remote sensing imagery. Ground data are also tainted with considerable measurement errors and their collection is expensive as well as time consuming.

Combining both model and observations is the task of data assimilation methods (Reichle, 2008; Lahoz *et al*., 2010). These methods aim at minimising uncertainties in the estimation of a given modeled state (as e.g. LAI) in an optimal way, i.e. based on statistical criteria. All of these methods aim at reducing the discrepancy between the measured and the simulated observations by adjusting either uncertain parameters and/or initial conditions (Prévot *et al*., 2003; Launay and Guérif, 2005; Dente *et al*., 2008) or the model state variables (De Wit and van Diepen, 2007; Reichle *et al.*, 2008a).

The performance of data assimilation strongly depends on the assimilation protocol and the uncertainty specification (Pellenq and Boulet, 2004). A comprehensive assessment of the performance of an assimilation scheme based on real observations is unfortunately not conceivable as it requires too many observations. In order to systematically investigate the assimilation performance, an approach based on synthetic data called “Observing System Simulation Experiment” (OSSE) has been developed.

This chapter presents the results of an OSS Experiment assessing the efficiency of the assimilation of leaf area index (LAI) derived from satellite in the World Food Studies (WOFOST) crop growth model.

Two assimilation methods, based on model parameters recalibration and on Ensemble Kalman Filter (EnKF), are considered. It is expected that these assimilation methods improve the estimation of final grain yields. The magnitude of the final grain yield estimation improvement is supposed to be, among other things, in connection with the uncertainty on the LAI derived from remote sensing in addition to the uncertainty of the model itself. In this context, besides the comparison of both assimilation approaches, the OSS Experiment also assesses the accuracy and temporal availability required on the retrieved LAI to reach a given objective in terms of errors reduction on final grain yields estimation.

2. Material and Methods

2.1. Crop growth model

We used the WOFOST (WOrld FOod STudies) crop simulation model as a basis for our work (Diepen *et al.*, 1989; van Ittersum, 2003). WOFOST is a mechanistic crop growth model that describes plant growth by using light interception and CO_2 assimilation as growth driving processes and by using crop phenological development as growth controlling process. The model can be applied in two different ways: (1) a potential mode, where crop growth is purely driven by temperature and solar radiation and no growth limiting factors are taken into account; (2) a water-limited mode, where crop growth is limited by the availability of water. The difference in yield between the potential and water-limited mode can be interpreted as the effect of drought. Currently, no other yield-limiting factors (nutrients, pests, weeds, farm management) are taken into account.

Interpolated weather data (rain, temperature, global radiation or wind speed) are combined with crop (management) and soil (hydraulic properties) data to simulate on a daily time step the crop life cycle from sowing or emergence to maturity or harvest. The WOFOST model provides daily estimates of biomass, grain yields and different state variables, such as LAI and soil moisture, for different crop types over a range of agro-environmental conditions.

The analysis described in this paper was carried with a particular implementation of WOFOST called PyWOFOST. In this implementation WOFOST compiled routines were linked with the Python interpreter which greatly increases the flexibility of the system allowing more sophisticated analyses such as probabilistic ensemble modelling and parameter optimization. The model implementation can be downloaded from http://www.wofost.wur.nl.

2.2. Area of Interest (AOI)

Although the OSS Experiment is mainly based on synthetic data and that subsequently no field data are needed, it is still worthwhile to have a first guess of the uncertainty on model parameters and initial states in order to avoid the use of “non-realistic synthetic data” which could lead to erroneous conclusions. In our experiment, the ADAM data set (Baret *et al.*, 2001) is used to ground the experiment to realistic conditions and later on to possibly validate some of the conclusions of the OSS Experiment. Indeed, recent studies (Duveiller *et al.*, 2011a; Duveiller *et al.*, 2011b) provide a quantitative assessment of the retrieval performance of LAI from remote sensing imagery on the basis of the ADAM data set.

The ADAM (Assimilation of spatial Data within Agronomic Models) experiment was a unique comprehensive experiment focusing on wheat crops and located in Funduléa (30 km East from Bucharest, Romania). The objective of this project was to develop and evaluate methods able to exploit high spatial resolution satellite observations to optimize cultural practices, estimate the production and evaluate environmental impacts.

The site (44°27'38.43''N; 26°37'14.34''E) is about 20×20km², flat, and includes mainly large agriculture fields. An intensive experimental campaign was conducted in 2000-2001 over wheat fields using 42 sampling units (20x20m²), providing a unique set of information on crop growth and growing conditions. Concurrently, meteorological data as well as 39 SPOT-HRV images and 15 SAR mages (ERS and Radarsat) were acquired.

2.3. General description of the OSS Experiment

OSS Experiments consist of identical twin experiments, a widely used technique to assess the impact of data assimilation. This setup consists in considering data simulated with one model run as observations and assimilating them afterwards into another run of the same model with different initial conditions. The convergence of the second run towards the first, defined as the "truth", can be measured to quantify the data assimilation effectiveness in driving the model with "wrong" initial conditions towards the truth (Raicich and Rampazzo, 2003).

The OSS experiment, focused on winter wheat, has been conducted similarly for the two considered assimilation methods (EnKF- and calibration-based assimilation methods). Figure 1.1 presents an overview of the OSS Experiment.

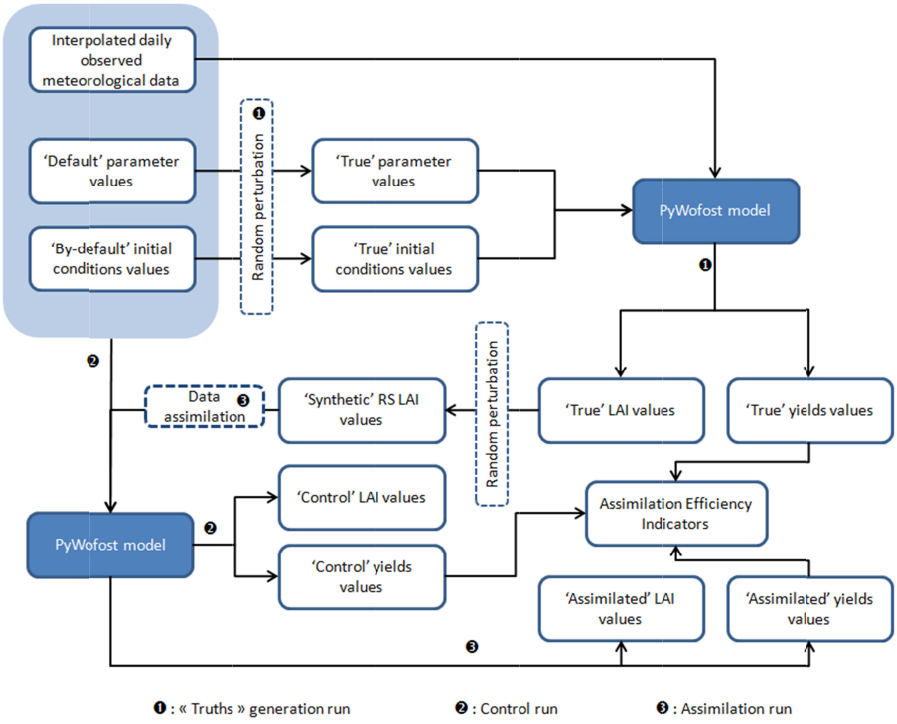


Figure 1.1.- Overview of the Observing System Simulation Experiment (OSSE)

The OSS Experiment is divided in 3 main steps.

First step: "truths" generation run (1)

In the truths generation run, a set of "synthetic observations" used afterwards in the assimilation step is computed. To reach this objective, a so-called "true" set of initial conditions and parameters is defined on the basis of the "by-default" set of initial conditions and parameters. "Default" values are the ones used operationally and can be considered as the best estimations of these initial conditions and parameters for a given area without any additional observations and therefore without assimilation.
Due to errors on measurements devices, heterogeneity or lack of information, the true initial conditions or parameter values are not necessarily the same than the ones used "by-default" in the model. In our analysis, it is assumed that the "true" initial conditions or model parameters values are possible states of a normal distribution centred on the "default" values. "True" model initial conditions and parameters considered in the OSS Experiments are random elements picked up within this normal distribution.

The generated set of "true" model initial conditions and parameters are then used to run the model in order to generate "true" state variable (LAI) and yields values. A stochastic noise term is added to the "true" LAI values in order to reflect observations errors as well as possible errors due to the (radiative transfer) model and inversion errors, thereby generating a "synthetic" remotely-sensed LAI.

Second step: control run (2)

The control run (often called "open loop" run) represents the best model estimate of the truth without the benefit of data assimilation. WOFOST model is run with its "default" initial conditions and parameters. The 'control' LAI and yield values generated during this run will be used to evaluate the improvement or alteration of the simulations due to the assimilation of the "synthetic" LAI values.

Third step: assimilation run (3)

In the last step of the OSS Experiment, the "synthetic" remotely-sensed LAI values defined in the first step are assimilated back in the WOFOST model. LAI and yield estimations generated after this assimilation phase will be compared with the values generated in the "truths" generation run which allow assessing the performance of the considered assimilation method.

2.4. Practical implementation of the OSS Experiment

2.4.1. Uncertain model parameters and initial states

In WOFOST, the number of parameters and subsequently the number of "candidates" for defining the model error is relatively high. In the frame of the OSS Experiment, two model parameters (SPAN and TSUM1 parameters) and one initial state (crop emergence date), influencing directly the leaves growth and subsequently the estimation of yields, have been considered in the "truths generation" run.

The SPAN parameter represents the life span (in days) of leaves growing at 35°C. SPAN parameter determines the rate and point of senescence of LAI. This parameter is also influenced by nutrients (a lack of nitrogen leads to early leaves browning) as well as by pests and diseases which are aspects that are not modelled in WOFOST. Influence of SPAN on LAI and yields have already been stressed in previous studies (Curnel and Oger, 2006a; Martin *et al.,* 2006).

TSUM1 is a variety parameter defined as the sum of effective temperatures from emergence to anthesis (°C d). In their sensitivity analysis of the CGMS model, Ndacyayisenga and Terres (1998) have identified this parameter as one of the most important for grain and biomass yields estimation in WOFOST model. Their analysis highlights the importance of rigorously calibrating crop growth cycles and therefore refining the estimation of emergence, anthesis and maturity dates. We could assume that the values currently available in WOFOST database are correctly calibrated and that the uncertainty on these parameters is low. However, a study (Curnel and Oger, 2004) in Belgium tends to contradict this assumption. In this study, TSUM1 have been estimated for 13 municipalities during the 1998-2003 period on the basis of phenological observations. On average, the observed TSUM1 was very close to the default calibrated value but the uncertainty around this mean was rather large (with a coefficient of variation equal to 7.5%) generating subsequently a significant uncertainty on final grain yield.

Local management and climate also affect LAI. Sowing date and subsequently crop emergence date is mainly influenced by climatic conditions and therefore quite variable from year to year. Since crop emergence is the moment at which the simulation starts in WOFOST, this parameter has an influence on LAI development.

Table 1.1 presents the 'default' values for these parameters and initial condition as well as the uncertainty level considered on them in the frame of the OSS Experiment.

Table 1.1.- 'Default' value and defined uncertainty level for the model parameters and initial condition considered in the OSS Experiment

	Units	**'Default' value**	**Uncertainty level (s.d.)**
SPAN	days	35	4
TSUM1	°C d	1504	75
Emergence date	-	27/10	7 days

As we considered only a random noise (no biais), a random perturbation was added to TSUM1, SPAN and crop emergence date follows a normal distribution centred on "default" values and that no correlation between parameters exists. The standard deviations for emergence date and TSUM1 considered here correspond to the standard deviations computed on the basis of the ADAM field data. As no information on SPAN parameter is available in the ADAM data set, the standard deviation has been arbitrarily fixed to 4 days such as the SPAN values randomly selected within the normal distribution centred on "default" value (35 days) are included within the lower (17 days) and upper (50 days) limits defined for this parameter in WOFOST.

The levels of uncertainty considered on SPAN, TSUM1 and crop emergence induce a significant spread on LAI (Figure 1.2) and subsequently on final grain yields estimation (table 1.2).

According to the considered parameter or initial state, the evolution of the induced LAI spread is different. We can notice, for example, that the spread on LAI induced by the considered uncertainty on SPAN parameter is only expressed in the decreasing phase of LAI.

The standard deviation observed subsequently on final grain yields estimation is close on 900 kg/ha, corresponding to a coefficient of variation around 9.4% (table 1.2). Uncertainty on final grain yields is lower when the uncertainty on model parameters / initial condition is considered separately.

Table 1.2.- Descriptive statistics for final grain yields (in kg/ha) when considering individually and all together an uncertainty on emergence date, TSUM1 and SPAN parameters (200 realizations, year 2001).

	Mean	**s.d.**	**min**	**max**
TSUM1	8462	541	6660	9087
SPAN	8599	645	5540	9631
Emergence date	8520	366	7560	8945
All together	8314	888	4717	9794

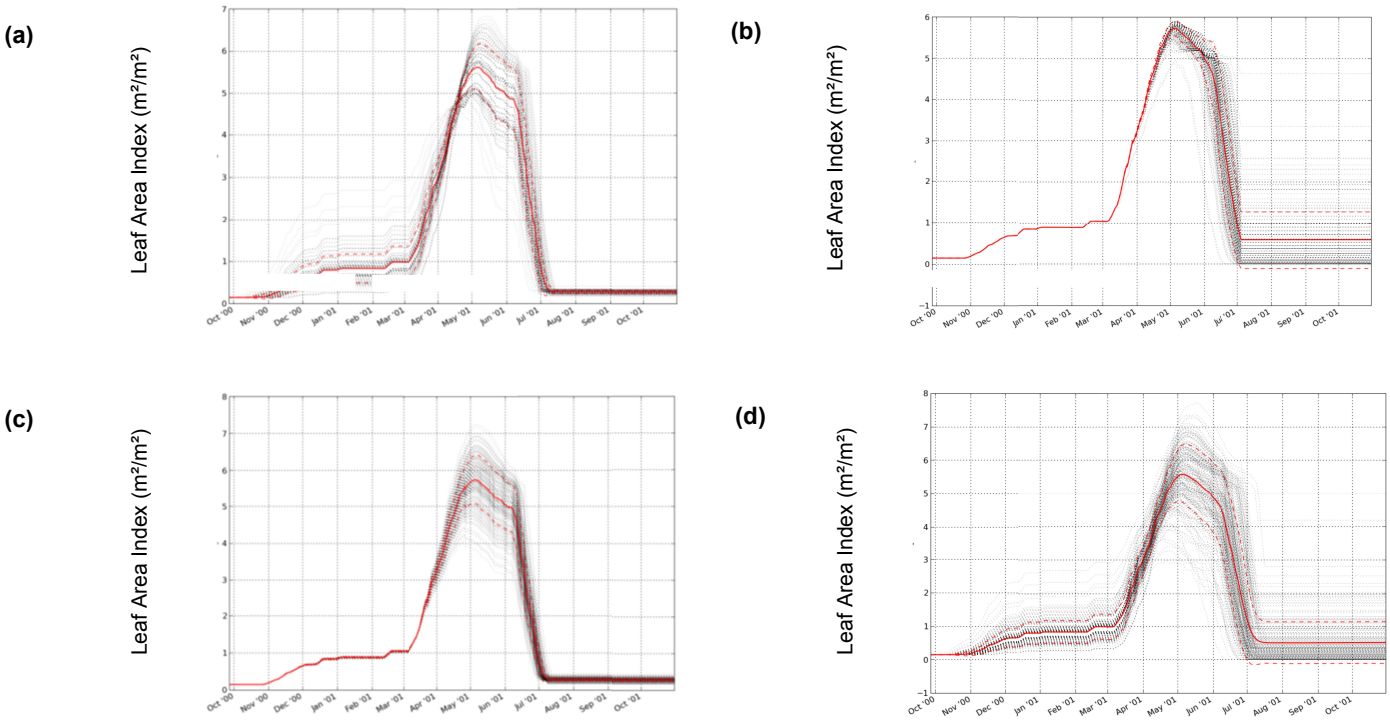


Figure 1.2.- Graphical representation of the spread on modeled LAI resulting from an uncertainty (standard deviation) of 7 days on crop emergence date **(a)**, 4 days on SPAN parameter **(b)** and 75°C on TSUM1 parameters **(c)** and on the 3 aforementioned parameters considered all together **(d)** - (200 realizations).

2.4.2. Assimilated LAI

As previously mentioned (paragraph 2.3), synthetic LAI time series (considered to be comparable to those derived from remote sensing and which will be assimilated back in the model) are generated from the true LAI time series by adding a stochastic noise term to the "true" LAI values. For each truth:

$$LAI_{RS,i} = LAI_{mod,i} + \varepsilon_i (Eq.1)$$

$$with \varepsilon_i \sim N(0, \sigma_i)$$

Where $LAI_{RS,i}$ is the synthetic leaf area index at time *i*, $LAI_{mod,i}$ the true leaf area index at time *i* and ε_i the added stochastic noise at time *i*.

The stochastic noise added to the true leaf area index value follows a normal distribution. The standard deviation σ_i associated to the distribution will be considered in our experiment as proportional to the true LAI values:

$$\sigma_i = CV * LAI_{mod,i} \qquad (Eq.2)$$

Where CV is the considered coefficient of variation. The coefficient of variation has been considered as constant whatever the time *i*. Three different values (5, 10 and 15%) will be considered (see paragraph 2.4.6. "Definition of OSSE scenarios").

2.4.3. Assimilation methods

For both assimilation methods, considered uncertain model parameters and initial conditions are the same than the ones used in the truths generation run.

Ensemble Kalman Filter (EnKF)

The Ensemble Kalman Filter (EnKF) is a Monte Carlo variant of the Kalman filter (Evensen, 2003) able to incorporate available observations sequentially in time. The probability distribution of a model state (including both model parameters and model responses) is represented empirically by an ensemble of realizations. The EnKF performs in turn a model forecast where the model responses (state variables) are propagated forward in time based on the model dynamics and a filter update in which the ensemble of the model state is adjusted by incorporating available observations. EnKF addresses two important factors in traditional inverse problems: the uncertainty of the model state variable and the sensitivity of the model state variable to the model parameters (Chen *et al*., 2009).

The basic filter update step in an EnKF for each ensemble member can be defined as

$$A^a = A + P_e * H^T (H P_e H^T + R_e)^{-1} * (D - HA) \ (Eq.3)$$

where A and A^a are the forecasted and analysed matrices of ensemble states, P_e and R_e are the ensemble and observation variance/covariance matrices, H is the measurement operator and (D-HA) are the innovation vectors.

Considering that assimilation only concern one state variable (LAI) and that the "synthetic" LAI is directly assimilated in the model, P_e and R_e are in this case

variances and the measurement operator is an identity matrix. Equation 3 can be reduced to:

$$A_i^a = A_i + P_e/(P_e + R_e) * (D_i - A_i)\ (Eq.4)$$

Where A_i^a and A_i are the analysed and forecasted LAI for ensemble member *i*, P_e and R_e are respectively the variances on the modeled LAI and the "synthetic" LAI and D_i is the perturbed LAI value used to update ensemble members *i* (Burgers *et al.*, 1998).

Like in the study of De Wit and Van Diepen (2007), an ensemble of 50 members was used since these authors found that this ensemble size was a reasonable trade-off between computation time and accuracy of ensemble mean and variance.

Variances on modelled LAI (P_e) are derived from these 50 members. As aforementioned, the ensemble is generated by considering only an uncertainty on 3 model parameters. No uncertainty has been considered on the model (considered as perfect, model errors Q = 0) and on meteorological forcings.

Recalibration-based assimilation method

The second assimilation method aims at recalibrating / reinitializing the WOFOST model. Sensitive and uncertain model parameters/initial values are tuned until the temporal behaviour of model LAI (LAI_{mod}) reaches the best agreement with the multi-temporal remotely-sensed ("synthetic") LAI (LAI_{RS}).

A variational assimilation algorithm is applied in order to seek the optimal input parameters minimizing the difference between LAI_{mod} and LAI_{RS}. The maximum likelihood solution to the problem is obtained simply by minimizing the squared errors:

$$J(p) = \sum_{i=1}^{N} \left(LAI_{RS,i} - LAI_{mod,i}\right)^2\ (Eq.5)$$

Where N is the number of assimilated LAI observations, $LAI_{RS,i}$ and $LAI_{mod,i}$ are respectively the ("synthetic") observed and modelled LAI at acquisition day *i* and *p* is the set of model parameters/initial states considered. The number of model parameters and initial states considered in the method will be at the most equal to 2 in order to avoid over-parameterization, especially when the number of observations is low.

Optimization has been done using the unconstrained Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963).

2.4.4. Assessment of assimilation efficiency

The efficiency assessment for the two assimilation methods as well as the estimation of the accuracy and temporal availability needed on observed LAI will be mainly estimated on the basis of the assimilation efficiency (AE) indicator. This indicator is computed from Relative Mean Absolute Error (*RMAE*) values estimated for the situation with and without assimilation (respectively $RMAE_{TA(j)}$ and $RMAE_{TO}$ indicators). The number of truths considered (N) is equal to 20. In order to have more

stable and reliable results, random perturbation to generate the “synthetic” remotely-sensed LAI values has been repeated several times (*k* iterations). For practical reasons (computation time), the number of considered iterations has been also fixed to 20.

The Relative Mean Absolute Errors are computed for each iteration *j* (j=1,…, K)as follow:

$$RMAE_{TO} = \frac{1}{N}\sum_{i=1}^{N} |FGYT_i - FGYO|/|FGYT_i| \ (Eq.6a)$$

$$RMAE_{TA(j)} = \frac{1}{N}\sum_{i=1}^{N} \left|FGYT_i - FGYA_{ij}\right|/|FGYT_i| \ (Eq.6b)$$

Where $FGYT_i$ is the final grain yield of the truth *i*, *FGYO* the final grain yield from the ”control” run i.e. the reference grain yield without assimilation and $FGYA_{ij}$ the final grain yield after assimilation for truth *i* and iteration *j*. Differences and ratios of final grain yields have all been considered in absolute values.

For EnKf-based assimilation technique, the final grain yields after assimilation ($FGYA_{ij}$) considered in eq. 6b is the mean of the ensemble final grain yields.

The same 20 truths are considered for the 20 different iterations.

The Assimilation efficiency AE indicator, derived from these Relative Mean Absolute errors, can be computed (for each iteration *j*) as follow:

$$AE(j) = 100 * \left(1 - RMAE_{TA(j)}/RMAE_{TO}\right) \ (Eq.7)$$

In the OSS Experiment the truths have been generated considering an uncertainty on 3 elements (2 parameters and one initial state). Considering that, as aforementioned, the maximum number of recalibrated parameters and/or initial state has been set to 2 in recalibration-based assimilation method, recalibrating on one or two parameters/initial state implies that the other selected parameters or initial state remains at their ‘default’ value in the recalibration phase. It therefore seems highly probable that the recalibrated values of the parameters and/or initial state on which the recalibration is performed have to be different from the true values since the recalibration tries to compensate artificially the deviation from the LAI time series obtained with the ‘default’ parameters values. This situation is met more often than not when the true values of the parameters or initial state not concerned by the recalibration are different from the ‘default’ parameters values. As the Levenberg-Marquardt algorithm is unconstrained, recalibrated parameters and initial state values can possibly take extreme values that could be considered as unrealistic leading in some situations to unrealistic final grain yields estimations. The phenomenon can particularly be stressed when the true parameters and initial state values have been sampled at the edges of the distribution.

Instead of discarding the unrealistic final grain yields estimation which could lead to an optimistic overestimation of the Assimilation Efficiency AE(j) value for some iterations but also with a view to have the same basis of comparison, a robust approach has been selected for the estimation of the mean Assimilation Efficiency value by considering the Winsorized mean instead of the simple arithmetic average.

The Winsorized mean is computed as the arithmetic mean after the replacement of the *m* smallest observations by the (*m*+1)st smallest observation and the replacement of the *k* largest observations by the (*m*+1)st largest observation. In other words, the observations are Winsorized at each end. The mean Assimilation Efficiency AE, used as reference to assess and compare the performances of the two assimilation methods, can be computed as follow:

$$AE = \frac{1}{k}\left[(m+1)\left(AE(j)_{(m+1)} + AE(j)_{(k-m)}\right) + \sum_{l=m+2}^{k-m-1} AE(j)_{(l)}\right] \quad (Eq.8)$$

With $AE(j)_{(l)}$ the ordered AE(j) values (l=1,…,k), *m* is the number of winsorized AE(j) values and has been fixed to 2 which represents 10% of the sample. *k* is the number of iterations.

Winsorized means have been computed with SAS/STAT® 9.1.3 statistical software.

Assimilation Efficiency AE indicator represents an estimation of the average reduction of errors on final grain yields estimation. A positive value for AE means therefore that the considered assimilation technique allows some improvement in final grain yields estimation. On the contrary, a negative value for AE means that the errors on final grain yields estimation are on average increased after assimilation and that in this situation, it is better to not assimilate and to use the model parameters and initial state 'default' values.

2.4.5. Assessment of the accuracy and temporal availability needed on remotely-sensed LAI

The second objective of the OSS Experiment aims at the assessment of the accuracy needed on remotely-sensed LAI in order to reach a given objective in terms of errors reduction on final grain yields estimation. In this frame, two objectives of 25% and 50% of errors reduction have been defined. To provide an order of magnitude, these 2 objectives correspond respectively to an errors reduction on final grain yields estimation of around 2.3 and 4.6 Qt / ha respectively.

Considering the probabilistic approaches used in the OSS Experiment, the assessment of the accuracy and of the temporal availability needed on remotely-sensed LAI can't be done considering exclusively the (Winsorized) Assimilation Efficiency value but has also to consider the variability observed between the different (k) iterations.

We have therefore decided to consider in our analysis the lower confidence level (α=5%) of the Winsorized Assimilation Efficiency value provided by SAS/STAT® 9.1.3 software (as far as this value is positive). The "new" indicator defined by this way has been called "Minimal Assimilation Efficiency" (MAE) and represent the lowest Assimilation Efficiency value that can be expected for the considered assimilation method.

The minimum accuracy and temporal availability needed on remotely-sensed LAI to reach one of the two defined objectives in terms of errors reduction on final grain

yields estimation will be assessed on this Minimal Assimilation Efficiency (MAE) indicator: each scenario presenting a Minimal Assimilation Efficiency value higher than the defined objectives (25 or 50%) is considered as suitable.

2.4.6. Definition of OSSE scenarios

The OSS Experiment described above has been performed for different "scenarios" in order to estimate the accuracy needed on Leaf Area Index (LAI) derived from remote sensing to reach a given objective in terms of yields estimation accuracy.

On the basis of a defined objective of Relative Mean Absolute Errors reduction, expressed by the (Minimal) Assimilation Efficiency indicator, it will be possible to identify the scenarios which allow this objective to be reached. Besides the uncertainty on LAI derived from remote sensing, the acquisition frequency and the moment when these observations are available have to be taken into account as well. The required level of precision on final grain yield could indeed be reached either with accurate remotely-sensed LAI observations with low revisit frequency or with less accurate but more frequent remotely-sensed observations.

Three levels of uncertainty on remotely-sensed LAI observations (coefficient of variation of 5, 10 and 15%) and four acquisition frequencies (every 3 days, every week, every 2 weeks and every month) have been defined. Three 'periods of assimilation' have been also defined: an 'early' period between the onset of greenness and the maximum of LAI (1st of March until the 15th of April), a 'late' period including also some observations in the senescent phase from the 1st of March up to the 15th of June) and finally a 'full' period including all the growing season from the 1st of March until July (see Figure 1.2).

Each combination of these 3 factors (the level of uncertainty on the remotely-sensed LAI observations, the acquisition frequency and the assimilation period) represents a given scenario. 36 scenarios can be therefore potentially defined. However, some of them can be seen as unrealistic and have been discarded. For example, it appears unrealistic to recalibrate the SPAN parameter in the early period as the uncertainty on SPAN parameter has no impact on modelled LAI during this period. For each of these considered scenarios, the indicators assessing the assimilation performance ($RMAE_{TO}$, $RMAE_{TA(j)}$, AE, MAE) have been computed.

3. Results

3.1. Results for EnKF-based assimilation technique

Assimilation Efficiency values for EnKF-based recalibration technique are presented in table 1.3.

Table 1.3.- Assimilation Efficiency (AE), standard error of the Assimilation Efficiency and Minimal Assimilation Efficiency (MAE) for EnKF-based assimilation strategy

Observations availability (time step in days)	*CV (%)*	***Early period***			***Late period***			***Full period***		
		AE	*s.e. AE*	*MAE*	*AE*	*s.e. AE*	*MAE*	*AE*	*s.e. AE*	*MAE*
3	*5*	5.00	1.44	1.59	-47.94	2.12	-52.96	-57.80	2.54	-63.81
	10	6.21	1.08	3.66	-46.81	1.84	-51.16	-53.98	2.27	-59.36
	15	5.37	1.02	2.95	-42.38	1.84	-46.72	-48.64	0.94	-50.86
7	*5*	5.97	0.75	4.19	-38.42	2.05	-43.27	-50.49	1.97	-55.14
	10	5.00	1.11	2.39	-35.04	1.03	-37.48	-48.54	2.60	-54.68
	15	5.53	0.75	3.76	-34.35	0.94	-36.57	-48.77	1.85	-53.15
14	*5*	3.77	0.96	1.50	-36.44	1.33	-39.58	-39.27	1.12	-41.91
	10	3.95	1.34	0.79	-36.24	1.26	-39.22	-38.77	1.53	-42.38
	15	6.47	0.50	5.27	-32.64	0.74	-34.40	-36.19	1.53	-39.81
30	*5*	6.36	0.63	4.87	-33.20	0.30	-33.92	-33.35	0.62	-34.82
	10	7.65	0.74	5.89	-30.48	0.52	-31.71	-30.58	0.56	-31.90
	15	8.96	0.60	7.54	-27.85	0.46	-28.95	-26.25	0.50	-27.44

Results for EnKF-based assimilation method are not very good: with the exception of the early period, Assimilation Efficiency (AE) values are always negative. Instead of reducing the average error on final grain yields estimation as expected, EnKF-based assimilation tend to increase this error. The increase is at best around 25-30% but can reach more than 50%. In other words, it is preferable in most situations to run the model with the 'default' parameters values rather than assimilating remotely-sensed LAI with Ensemble Kalman Filter.

These results appear counter-intuitive. Indeed, AE values tend to increase with the decrease of the number of observations and in a lower extent with the increase of the uncertainty on remotely-sensed LAI. Whatever the assimilation period considered, higher Assimilation Efficiency values are observed when assimilated remotely-sensed LAI observations are available every month and when the uncertainty on remotely-sensed LAI reaches 15% (CV). The standard error on Assimilation Efficiency values follows globally the same pattern: a reduction when the time step between assimilated remotely-sensed LAI observations and/or the uncertainty on these observations increases.

The hypothesis for explaining these counter-intuitive results lies in the method itself and corresponds to what we call the "phenological shift". Indeed with the Ensemble Kalman Filter method, for each observation/assimilation date, the value of each realization/each member of the ensemble is updated in function of the uncertainty of modelled and observed LAI. At each assimilation step, due to uncertainty considered on emergence date and on TSUM1 parameter step, all the members of the ensemble are not at the same phenological stage. Different modelled LAI values at different development stages are therefore updated on the basis of an observation at a defined development stage which logically introduces some "perturbations". The fact that in WOFOST the value of some parameters are linked to the phenological stage further aggravates the situation.

A simplified illustration of the "phenological shift" is presented in figure 1.3.

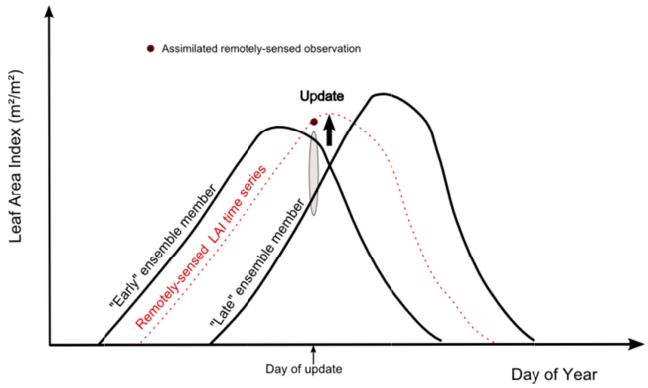


Figure 1.3.- Schematic representation of « phenological shift »

Two ensemble members are represented in figure 1.3, one with an early crop emergence date ("Early ensemble member") and one with a late crop emergence

date (“Late ensemble member”). The red dotted line represents a remotely-sensed LAI time series. At the day of update (let’s suppose that no assimilation was done earlier), assimilated remotely-sensed LAI observations is in our example close to the peak of LAI while the model LAI values of the “Early” and “Late” ensemble members are at this moment respectively in the decreasing and increasing phase. The updated LAI values of two ensemble members could be (according to the uncertainty level on the assimilated remotely-sensed observation) updated towards higher values while, for example, the LAI values of the early ensemble member is in a decreasing phase. In the same way, LAI value for the late ensemble corresponding to a value at mid-term of the increasing phase will suddenly raise up, possibly up to a value that could possibly correspond, if the uncertainty on observed assimilated LAI is very small for example, to a value close to LAI peak.

Illustration in figure 1.3 overstates the “phenological shift” but illustrates clearly this concept and its possible consequences.

The “phenological shift” hypothesis can explain why the AE values are higher when the uncertainty on assimilated remotely-sensed is higher. Indeed, when the uncertainty on assimilated observations is high, the contribution of these observations in the corresponding updated modelled LAI values (A_i^a in equation 4) is reduced. In other words, the observations have less weight in the updated values computation which reduces the perturbation introduced by the phenological shift. It can also explain why the Assimilation Efficiency values are a bit higher, even slightly positive in the early period, when the number of assimilated observations is lower. Less observation mean less perturbation induced by the phenological bias.

The existence of a phenological shift is not necessarily synonym of negative assimilation efficiency. Figure 1.4 presents the Assimilation Efficiency values computed for 50 truths and for different levels of uncertainty considered on TSUM1 parameter and crop emergence (responsible for the phenological shift). In this example, the coefficient of variation (uncertainty) on remotely-sensed LAI, available every week, is equal to 10%. These levels of uncertainty have been chosen separately to avoid any possible compensation of the effects between the 2 factors, uncertainty on the SPAN parameter remaining at the level defined in the OSS Experiment. When an uncertainty is considered on TSUM1 parameter, uncertainty on crop emergence is set to zero and conversely.

Assimilation Efficiency decreases with the increase of the considered level of uncertainty and therefore with the supposed increase of the impact of the “phenological shift” both for TSUM1 and crop emergence. Maximum AE value (equals to 34%) is indeed observed when no uncertainty is considered on TSUM1 and on crop emergence i.e. in the absence of phenological shift.

AE is only negative when the uncertainty level on crop emergence is equal to 6 days. These observations reinforce our hypothesis on the “phenological shift” and its effects.

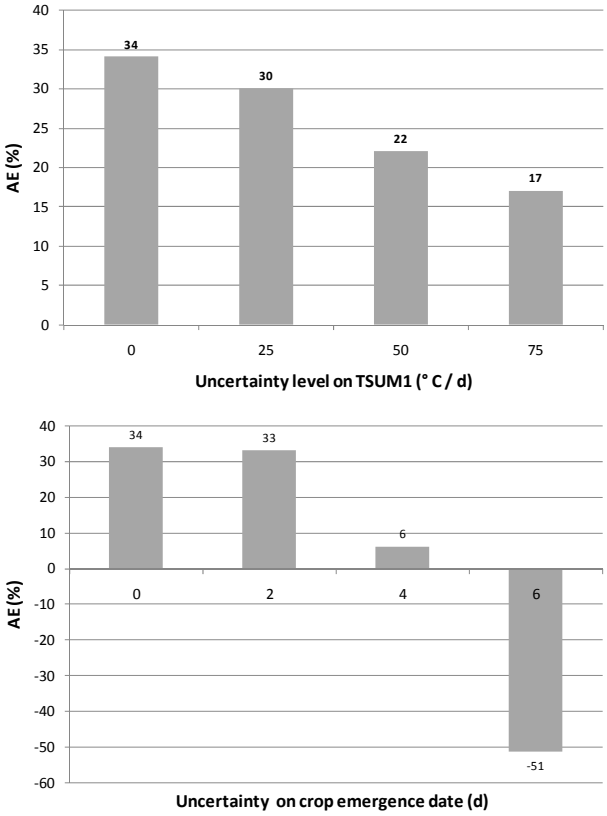


Figure 1.4.- Assimilation Efficiency (%) with EnKf assimilation technique for different levels of uncertainty on TSUM1 parameter (upper) and crop emergence date (bottom) – CV on RS-LAI=10%, observations every week, late period

We can therefore conclude that assimilating LAI with the EnKF-based assimilation method as implemented in our OSS Experiment is not the most suitable approach.

Considering these very poor results, the study aiming to estimate the accuracy and the temporal availability needed on assimilated remotely-sensed LAI has not been performed for EnKF-based assimilation strategy.

3.2. Results for recalibration-based technique

Table 1.4 presents the results of the OSS Experiment for the intermediate uncertainty level considered on the assimilated remotely-sensed LAI (10%). Results follow globally the same pattern for the two other considered uncertainty levels on remotely-sensed LAI.

Unlike the results for the EnKF-based assimilation technique, AE values are generally positive which means that globally recalibrating uncertain parameters improves final grain yields estimation. While the improvement appears limited in some situations, it can sometimes reach very interesting values up to 50-60%. In a

crop yields forecasting perspective, it is particularly interesting to notice that significant improvements can be already reached early in the season. Improvements reach indeed already quite often 20% in the early period. Performing the recalibration later in the season allows reaching better results. In the late season, a reduction of errors on final grain yields estimation around 40% can be observed in many situations. At the end of the season (full period), therefore more in a yields estimation than prediction context, a recalibration on both TSUM1 and SPAN parameters allows to reach up to 60% of errors reduction.

Increasing the number of assimilated data tends therefore to improve the AE values. The fact that the decrease of the AE values with the increase of the time step between assimilated remotely-sensed data is not systematically observed can certainly be imputed to the stochastic approach used in the OSS Experiment coupled with a possible insufficient number of iterations.

When comparing single-parameter recalibrations with recalibrations based on a combination of two parameters for a given assimilation period, it is not so surprising to notice that the best AE values are often observed for a recalibration on two parameters. Some exceptions can be however found in the early period. While during this period, the best recalibration type tends to be a joint recalibration on both TSUM1 and emergence date when remotely-sensed LAI observations are available every 3 days -1 week, it is not the case anymore when remotely-sensed LAI observations are available only every 2 weeks or every month, especially if the comparison is made on the MAE Indicator.

In a limited number of situations, negative AE values can be observed. They often involve the SPAN parameter, especially when this parameter is considered alone. In the late period, i.e. the period including normally all the increasing and a part of the decreasing phase of LAI, AE is at best equal to -80.75%. In other words, recalibrating SPAN alone during this period increases the average error on final grain yield estimation by at least 80% compared to the situation without assimilation. The origin of this extreme situation observed for SPAN parameter is multiple.

Table 1.4.- Winsorized Assimilation efficiencies (AE), standard error of the Assimilation Efficiency and Minimal Assimilation Efficiency (MAE) for recalibration-based assimilation strategy and an uncertainty level on assimilated remotely-sensed LAI set to 10%

		AE				s.e. AE				MAE			
	Observations availability (time step in days)	***3***	***7***	***14***	***30***	***3***	***7***	***14***	***30***	***3***	***7***	***14***	***30***
Early period	TSUM1	-5.55	-1.21	2.65	-5.23	0.95	1.35	2.21	1.78	-7.59	-4.08	-2.05	-9.02
	SPAN[2]	-	-	-	-	-	-	-	-	-	-	-	-
	Emergence date	24.00	21.44	20.45	17.61	0.64	0.83	0.87	0.85	22.64	19.66	18.60	15.81
	TSUM1-SPAN	-4.93	-1.10	-1.04	-1.34	1.12	1.22	1.79	2.11	-7.32	-3.70	-4.86	-5.84
	TSUM1-Emergence date	38.25	28.54	15.54	14.60	3.14	4.09	5.30	3.64	31.56	19.83	4.25	6.85
	SPAN-Emergence date	24.10	23.12	20.79	19.74	0.64	0.40	0.88	1.05	22.73	22.26	18.90	17.50
Late period	TSUM1	40.07	39.46	37.22	39.34	0.82	1.41	1.25	2.05	38.32	36.45	34.56	34.97
	SPAN	-80.75	-89.77	-84.88	-88.09	0.82	1.10	1.92	1.87	-82.50	-92.11	-88.96	-92.07
	Emergence date	32.33	33.14	24.64	23.34	0.41	1.26	2.60	3.50	31.45	30.45	19.10	15.88
	TSUM1-SPAN	53.30	24.65	28.18	9.49	1.47	3.13	3.51	2.49	50.16	17.97	20.70	4.18
	TSUM1-Emergence date	46.35	46.88	45.45	41.58	0.77	1.07	1.55	2.38	44.70	44.61	42.16	36.50
	SPAN-Emergence date	6.05	-17.32	-19.80	-37.86	2.11	3.31	4.66	2.95	1.56	-24.38	-29.72	-44.15
Full period	TSUM1	25.60	27.44	25.38	22.18	0.64	0.61	1.44	1.48	24.24	26.14	22.32	19.02
	SPAN	-42.15	-41.00	-41.19	-56.57	0.38	0.69	0.92	0.95	-42.96	-42.46	-43.14	-58.60
	Emergence date	20.69	22.16	21.58	20.52	0.60	0.73	0.83	1.35	19.41	20.61	19.81	17.64
	TSUM1-SPAN	65.24	64.15	55.24	34.63	1.08	1.78	2.43	4.26	62.93	60.36	50.06	25.55
	TSUM1-Emergence date	27.38	26.48	26.63	24.05	0.41	0.71	1.21	1.81	26.50	24.97	24.06	20.20
	SPAN-Emergence date	43.66	45.14	39.18	5.88	3.28	3.81	4.24	3.32	36.68	37.02	30.15	-1.19

[2]No data are displayed as SPAN parameter as this parameter has no influence in the early period

First of all, as previously mentioned, SPAN parameter only influences the decreasing phase of LAI and its value is notably conditioned by the time when LAI peak occurs. In the model, the date of the LAI peak is mainly dependant on the start of the simulation (i.e. the crop emergence date) and on the time it takes from this point onwards to reach the LAI peak (represented by the TSUM1 parameter). The date of the LAI peak is therefore mainly dependant on the two other factors considered as uncertain in our OSS Experiment. We can easily imagine that recalibrated SPAN values have to take values different from the true values and most probably unrealistic values as an uncertainty on the moment of LAI peak exists. On the other hand, an accurate estimation of SPAN implies having observations on the whole decreasing phase of LAI and not only a part of this phase as in the late period. This can explain why the AE values increase between the late and the full period including all or nearly all the decreasing phase of LAI.

The same explanations apply for TSUM1 parameter in the early period for which negative or at best slightly positive Assimilation Efficiency values can be observed. Indeed, TSUM1 parameter has ideally to be computed from emergence to anthesis, which roughly represents all the increasing phase of LAI. This phase is not fully covered in the early period. Furthermore, the crop emergence is considered as fixed when recalibration is performed on TSUM1 only, even though this model initial state has been considered to be uncertain in the “truths” generation run in our OSS Experiment. With respect to this, we can observe that the joint recalibration of crop emergence and TSUM1 improves significantly the AE value during the early period. Let’s note that the very similar values observed between recalibrations on TSUM1 and on TSUM1-SPAN on one hand, and between recalibrations on emergence date and on SPAN–Emergence date on the other, can be easily explained by the fact that SPAN parameter has no influence on the increasing phase of LAI.

The best recalibration types per scenario, i.e. the recalibration type with the lowest MAE values, are presented in table 1.5.

Clear patterns can be identified for each assimilation period.

In the early period, except when the uncertainty on remotely-sensed LAI is equal to 5%, best MAE values are observed when recalibrations are performed only on crop emergence date. Acquisition of very accurate observations (CV equals to 5%) provides significant improvement. From a minimum of 15-20%, reduction of errors on final grain yields estimations reaches in this situation up to 45-50%. It is furthermore nearly the only possibility to meet the lowest objective in terms of errors reduction (25%) in the early period.

Table 1.5.- Best recalibration type and corresponding Minimal Assimilation Efficiency per scenario (EM: emergence date)

Observations availability (time step in days)	*CV (%)*	**Early period**		**Late period**		**Full period**	
		Best recal. type	*MAE*	*Best recal. type*	*MAE*	*Best recal. type*	*MAE*
3	*5*	EM-TSUM1	47.58	TSUM1-SPAN	54.04	TSUM1-SPAN	64.02
	10	EM-TSUM1	31.56	TSUM1-SPAN	50.16	TSUM1-SPAN	62.93
	15	EM	21.31	EM-TSUM1	44.21	TSUM1-SPAN	63.25
7	*5*	EM-TSUM1	46.38	EM-TSUM1	46.27	TSUM1-SPAN	64.15
	10	EM	19.66	EM-TSUM1	44.61	TSUM1-SPAN	60.36
	15	EM	17.32	EM-TSUM1	39.26	TSUM1-SPAN	49.00
14	*5*	EM-TSUM1	44.21	EM-TSUM1	44.65	TSUM1-SPAN	64.60
	10	EM	18.60	EM-TSUM1	42.16	TSUM1-SPAN	50.06
	15	EM	18.80	TSUM1	33.41	TSUM1-SPAN	32.79
30	*5*	EM-TSUM1	30.61	EM-TSUM1	46.91	TSUM1-SPAN	51.31
	10	EM	15.81	EM-TSUM1	36.50	TSUM1-SPAN	25.55
	15	EM	15.95	EM-TSUM1	25.75	EM-TSUM1	26.46

On the contrary, this first objective is systematically reached in the late period. This is not the case for the second objective (50% of errors reduction), which is only reached when assimilated LAI observations are available every 3 days with an uncertainty lower than 15%. These results show despite everything that, considering the potentialities offered by the current available sensors and our working hypotheses, an improvement around 30-40% can be already expected in a crop yield forecasting approach. Best results are usually obtained by a joint recalibration on crop emergence and TSUM1 parameter.

The full period presents the highest MAE values, usually obtained through a joint recalibration on TSUM1 and SPAN parameters. During this period, the second objective in terms of errors reduction (50%) is reached in a majority of situations.

Let's note that when the best recalibration type (e.g. a joint recalibration on crop emergence date and TSUM1 in the full period when observations are available every month and present an uncertainty level of 15%) is not the one predominantly observed in the corresponding assimilation period (a joint recalibration on TSUM1 and SPAN parameters in our example), MAE values for these situations are generally very close to the MAE value of the predominantly observed recalibration type.

From a general point of view, reducing the uncertainty on LAI observations for a given observation availability scheme allows better improvements of the MAE values than reducing the time step between LAI observations for a given uncertainty on LAI observations.

4. Discussions

One of the major findings of the OSS Experiment is the poor efficiency observed for the EnKF-based assimilation technique. In our experiment, EnKF-based Assimilation of LAI into the WOFOST model produces generally results which are the opposite of those normally expected from an assimilation technique: in most cases the average error on final grain yields estimation is increased after assimilation. These unexpected results have been explained by the existence of a phenological shift induced by the uncertainty considered on TSUM1 and crop emergence.

In real conditions, uncertainty on phenological stage (phenological shift) will systematically exist whatever the considered area of interest. It seems, for example, difficult to avoid uncertainty regarding crop variety or sowing date and subsequently on emergence date as sowings never occur at the same time for all fields within a given region and vary at all events from year to year due to the variability of climatic conditions.

These results don't necessarily mean that the method can't provide satisfying results: distinctly positive AE values have been observed in our experiment when lower uncertainty levels on TSUM1 and crop emergence are considered i.e. when the phenological shift is reduced. Positive Assimilation Efficiency values may also be observed when assimilation is based on variables that are less sensitive to such

phenological shift. For example, de Wit and van Diepen (2007) have observed that assimilating coarse resolution satellite soil moisture estimates derived from microwave sensor in order to correct errors in the WOFOST water balance with Ensemble Kalman Filter (EnKF) clearly improves the relationship between modelled yields and official yield statistics for winter wheat. Nevertheless, there is a risk that the phenological shift will systematically hamper the performance of this assimilation technique (at least as implemented in our experiment) compared to other assimilation techniques as the recalibration-based approach used in our analysis.

Assimilation techniques based on the Ensemble Kalman Filter have been initially developed and successfully used in oceanography and meteorology i.e. in the frame of continuous processes for which there is no parallel controlling process like phenology. These EnKf-based assimilation techniques aim at correcting sequentially the uncertainty observed on the modelled variable without correcting the causes of this uncertainty. This is not the case for the second assimilation method considered in our experiment which attempts to "recalibrate" some of the uncertain model parameters and/or initial state. This explains why, in most situations, the estimation of final grain yields is improved in average, and sometimes in a significant way with a reduction of more than 50% of the average error on final grain yields estimation compared to the situation without assimilation.

An objective of 25% errors reduction on final grain yields estimation can be reached in a quite high number of assimilated LAI observations availability schemes and uncertainty levels. However, when this objective is doubled, the possibilities are significantly restricted and generally imply having observations all along the growing season, frequent and accurate assimilated LAI observations with an uncertainty level (CV) equal to 10% (or ideally lower). From a global point of view, reducing the uncertainty level on the observations seems to be more efficient to reach this errors reduction objective. These results suggest that the remote sensing community should focus efforts on improving LAI retrieval algorithms rather than improving the temporal availability of the observations.

It is important to note that the scenarios defined in our experiment are ideal situations. The regular time step considered for observation is unrealistic in practice with optical remote sensing instruments due to high variability in cloud cover and other atmospheric conditions. In order to attain valid observations every 3 or 7 days, either the revisit frequency or the field of view of the satellites needs to be increased substantially. Different sensors can already provide images almost on a daily basis but it comes often at the expense of smaller geographical coverage associated with consequent acquisition costs (e.g. SPOT, RapidEye, Venμs) limiting greatly the application or of a coarser spatial sampling i.e. coarser pixels (e.g. MODIS/MERIS providing daily reflectances at 250-300 m) resulting in LAI obtained from potentially more heterogeneous surfaces which induces scaling errors (Garrigues *et al*., 2006). Although alternative solutions exist, such as for example the Disaster Monitoring Constellation (DMC) which is able to provide up to daily HiRes images (32 m spatial resolution) with a relatively good spatial coverage (340 km swath), much is expected from the coming Sentinel-2 mission which will survey the world with 20m spatial

resolution every 5 days under cloud-free conditions (every 15-30 days considering the presence of clouds) and will be distributed at no cost.

The representation of the accuracy of LAI retrieved from remote sensing using the CV is also idealized. We have indeed considered that the level of uncertainty on remotely-sensed LAI observations was constant all over the growing season and that this uncertainty follows a normal distribution. It is in reality harder to accurately retrieve reliable values when LAI is small because in such cases it can become hard to discriminate the green vegetation from the background. Similarly, when LAI is high the remote sensing signal can become saturated leading to an asymmetric distribution of errors. Further discrepancies can come from the fact that, when looking at canopies such as wheat, the variable retrieved from remote sensing is closer to Green Area Index (GAI) rather than LAI (Duveiller *et al.,* 2011a). Finally, the actual retrieval performances are generally lower than those considered here. For instance, Duveiller *et al.* (2011a) have observed RRMSE values on Green Area Index (GAI) derived from a time series of SPOT/HRV and SPOT/HRVIR for winter wheat ranging between at best a little bit lower than 25 % in the green-up phase up to more than 100% in the late senescence phase.

Many other hypotheses have been set in this experiment, notably normal distributions for the uncertain model parameters and initial state and the fact that their default values represent the most probable values (no bias), the perfection of the model or the fact that remotely-sensed LAI represents a non-biased vision of the reality.

The obvious outperformance of recalibration-based technique has also to be relativized. The arbitrary decision to assume that all modelling errors are on model parameters tends to favour this technique which is able to tackle the problem directly (i.e. updating the parameters that are explicitly degraded in the synthetic experiment set-up), contrary to a state-based assimilation technique such as EnKf. Lower performances are probably to be expected by the introduction of an incorrect model structure or random errors on meteorological forcings for example.

All these results therefore need imperatively to be confirmed on a real case study. However, our experiment has the merit to show that improvements are possible with a simple recalibration-based assimilation strategy. If we work on the assumption that LAI observations can be available every 2 weeks with an uncertainty level (CV) equal to 15%, a reduction of errors on final grain yield estimation of 30-35 % can already be reached in a crop yields forecasting perspective (in the late period). Moreover, if the frequency and the quality of remote sensing data can be improved with a view to reach these objectives even early in the season, the algorithms used to derive biophysical variables from the remote sensing information or used in the assimilation phase can be in the same way improved. The recalibration-based assimilation algorithm used in our OSS Experiment could be improved by using a weighted approach when minimizing the errors between observed and modelled LAI values by providing less weight to LAI observations when the uncertainty on these estimations is considered as the strongest and/or when the number of available observations is low. We could also imagine performing a smoothing operation on the observed

remotely-sensed LAI observations in addition to the application of exclusion rules (for example defining that the LAI can't exceed a given value at a given period) prior to assimilation or using hybrid approaches, combining the merits of the two assimilation strategies, by using a prior recalibration of emergence date followed by an assimilation based on EnKF strategy during the rest of the season.

5. Conclusions

From this Observing System Simulation Experiment, considering the selected uncertainty levels on the three considered parameters, we can first of all conclude that the EnKF-based assimilation technique is not suitable in the frame of the assimilation of LAI observations in the WOFOST model. In most situations, the average error on final grain yields estimation is indeed increased compared to the situation without assimilation which means that it is more accurate to use simply the default parameters values. The poor behaviour of this assimilation technique is mainly due to the existence of a phenological shift induced in the frame of our OSS Experiment by the joint uncertainties on TSUM1 and crop emergence date.

Contrasting with the EnKF-based assimilation technique, the recalibration-based assimilation technique allows significant improvement in the final grain yields estimation. In the most ideal situations, i.e. when the uncertainty on remotely-sensed observations is very low (5% CV) and/or when these observations are available all along the growing season, our recalibration-based assimilation approach allows a reduction of the average error on final grain yields estimation of more than 45-50%. The moment of assimilation, the uncertainty level on remotely-sensed data and the time step between the assimilated LAI observations logically have an influence on assimilation performance. For each assimilation period, it is possible to identify the parameter or the combination of parameters providing the best results. Best recalibration type is globally a single recalibration on crop emergence date when the LAI observations are available before the peak of LAI (early period). When LAI observations are also available in the decreasing phase of LAI, best recalibration types are a joint recalibration on crop emergence date and TSUM1 and a joint recalibration on TSUM1 and SPAN parameters respectively when LAI observations are also available in a part (late period) or all along (full period) the decreasing phase of LAI. The identification of a particular and different recalibration type per assimilation period can be explained by the existing interdependence between the considered uncertain parameters and initial state and the necessity to have a good correspondence between the period on which the parameters have to be computed and the period on which the LAI observations are available. A good recalibration of TSUM1 implies a good recalibration of crop emergence date and a good recalibration of SPAN parameter implies a good recalibration of crop emergence and TSUM1 parameter.

Our experiment has identified, considering our initial hypotheses, how frequently and how accurately LAI has to be retrieved from remote sensing data to reach either 25 or 50% of error reduction on final grain yields estimation. Remotely-sensed LAI observations have to be available at least on a period including the maximum of LAI

(late period) to reach the most realizable objective. Although the objective of 25% error reduction on final grain yields estimation can be reached in a relatively high number of assimilated LAI observations availability and uncertainty levels, the possibility to reach the second objective of reducing the error by 50 % is much more restricted as it requires frequent and accurate assimilated LAI observations all along the growing season, most of time with an uncertainty level (CV) of 10% or lower.

None of the present-day sensors can really pretend meeting operationally the identified accuracies and availabilities on remotely-sensed LAI observations. However, forthcoming missions such as Sentinel-2, coupled with improved assimilation techniques, could make a significant step towards the fulfilment of these requirements in terms of accuracy and availability.

CHAPTER 2 : Influence of pixels' purity on assimilation performance

1. INTRODUCTION

Uncertainty on remote sensing variables is an important factor to take into account in the specific context of data assimilation. Chapter 1 has stressed that a reduction of the errors on yields estimation is preferably reached by a reduction of the uncertainties on the assimilated remotely-sensed observation.

The level of uncertainty on remotely sensed observation is notably function of the sensors characteristics, of the corrections applied on raw remotely-sensed data and of the methods (statistical relationships, inversion of radiative transfer models…) used to transform the satellite signal in the target biophysical variable.

Multitemporal data assimilation applications are for example dependent on the accurate registration of the data into a common spatial framework (Roy, 2000). All the images composing the time series need to be perfectly coregistered so that image pixels are different (temporal) views of the same object or multitemporal pixel entity. Such a coregistration is achieved by mapping to a common system of grid cells, all the geolocated observations of each acquisition being allocated into a predefined output grid cell (Gómez-Chova *et al*., 2011). The preprocessing step aiming at allocating geolocated observations of each acquisition into a predefined output grid cell is known as gridding and is an important phase as it could introduce gridding artifacts, undesirable effects induced by differences in dimensions and orientation between observations and grid cells (Thorp *et al*., 2010). The gridding artifacts come from two effects, the "pixel shift" corresponding to the mismatch between location on the ground of observations and the predefined grid cells for storing the observations and the "geolocation error", considered as modest by Wolfe *et al.* (2002), corresponding to errors in assigning geolocation coordinates to observations.

According to Tan *et al*. (2006), these gridding artifacts strongly influence the local spatial properties of MODIS images. The sensor observation in any grid cell is only partially derived from the location of the cell, with the average overlap between observations and their grid cells being less than 30%. The observation cover (obscov) included in the MODIS data reports for each cell the actual overlap with the current observation.

Typically, the grid cell dimensions are set equal to the nadir observation dimensions (Wolfe *et al*., 1998; Tan *et al*., 2006). A common misconception is indeed to believe that the observational footprint is the geometric projection of a rectangular pixel onto the Earth's surface (Cracknell, 1998). In practice, observations have elliptical shapes as the surface area convolved with the system point spread function (PSF) defines the area that is physically sensed (Schowengerdt, 2007). Gridding effects are mainly observed for whiskbroom sensors designed for providing a high revisit time (usually

one to three days) while conserving a medium spatial resolution (around 250-300 m). Wide field-of-view whiskbroom sensors, such as MODIS, present progressively overlapping observations further from the nadir, producing different pixel overlaps depending on the view zenith angle and inducing therefore a mismatch between observations and grid cells for any acquisition date. The large swath of these sensors ensuring the up to daily revisit time induces also a mismatch between observations of different dates as revisit time doesn't match with the repeat cycle. For example, MODIS sensor with its 2330 km swath width allowing achieving a near-daily global coverage has only a repeat cycle of 16 days. The daily acquisitions within this time window are therefore acquired from a different orbit and subsequently from different observation geometry. Differences in the footprint of each pixel among the different acquisitions are therefore expected. These differences are also present between images acquired from the same orbit but at different dates due to the unavoidable small shifts in satellite position and in the acquisition time (Thorp *et al.,* 2010).

The variability of footprint within an image and between the different acquisitions of a satellite time series represents a limited problem in case of homogeneous areas but is much more problematic in case of heterogeneous areas combining different land covers. This problem of mixed pixels is logically frequently met with medium resolution images. Even for pushbroom sensors like SPOT-Vegetation (1 km) and ENVISAT-MERIS (300 m) allowing a most content footprint across track, fragmented landscapes remain a challenge.

As crop growth models provide crop specific information, assimilated observations in these models have to be crop specific as well. In order to make use medium resolution data for crop growth monitoring, it is required to select pixels whose observational footprints fall (at least to a certain extent) within the target crop specific fields. To face gridding effects and complex acquisition geometry problems, Duveiller and Defourny (2010) have developed a method for the MODIS sensor whose result is a pixel purity map providing for a given landscape an estimate of the percentage of the information obtained from surfaces that are covered with the target crop (winter wheat) for every cell in the MODIS grid. The approach consisted in convolving a kernel modelling the MODIS spatial response over detailed winter wheat masks constructed from detailed crop acreage maps. This methodology was used by De Wit *et al*. (2012) to generate winter wheat GAI time series with a view to assimilate them in WOFOST growth model (Diepen *et al*., 1989; Van Ittersum *et al.*, 2003). Assimilation was done, for MODIS grid cells presenting at least a pixel purity of 75%, by optimizing two important model parameters (TDWI and SPAN parameters) in order to minimize the difference between the simulated and observed GAI time series for each individual pixel and year.

Building on these experiences, the main objective of this study is to assess the impact of pixels purity (and gridding effects) on assimilation. This impact has been more specifically assessed on the MODIS data set and the recalibrated parameters (TDWI and SPAN) considered in De Wit *et al.* (2012) for winter wheat. Although the assimilation approach considered in De Wit *et al.* (2012) allow improving significantly the relationship between simulation results and reported yields at regional level,

relatively low assimilation efficiencies are sometimes observed. This study aims at assessing if and how the incomplete pixel purity influences the results of the assimilation and subsequently at understanding and identifying some of the elements that could improve the assimilation approach developed by De Wit *et al.* (2012). Our study extends therefore this analysis by investigating the spatial (and temporal) distribution of recalibrated optimized model parameters and subsequently of GAI time series taking into consideration the influence of an incomplete pixel purity but also in a lower extent gridding effects. The analysis has been performed at model and MODIS grid cells level by considering the proportion of the main land use classes within respectively these grids.

2. MATERIAL AND METHODS

2.1. Area of Interest

Considering the availability of ancillary data such as detailed yearly crop maps at field level, this work focuses primarily on the Walloon territory in Belgium and more precisely on two agricultural regions, the 'Limoneuse' (loamy) and 'Condroz' agricultural regions (Fig 2.1). Agricultural regions are defined on the basis on soil and climate criteria.

a. Agronomical characterization of the 2 agricultural regions

Figure 2.1 presents the main land use per municipality in Walloon region (year 2009). Obvious differences between the two agricultural regions can be noticed.

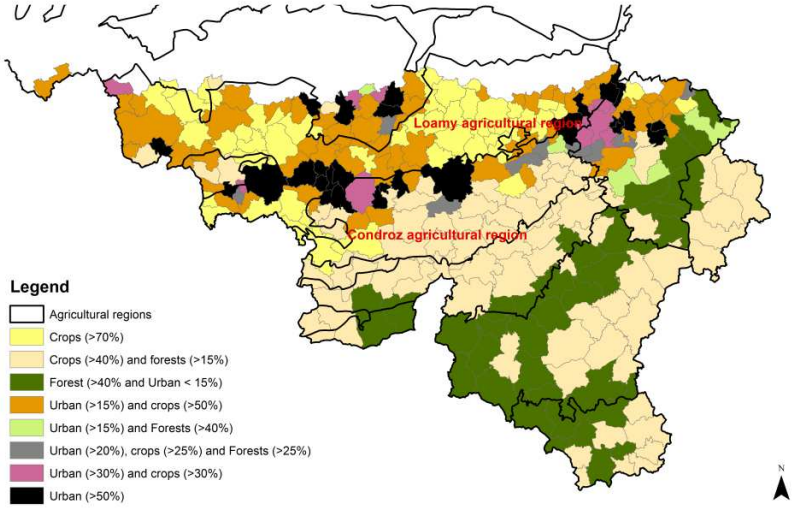


Figure 2.1.- Typology per municipality (only Walloon region is represented) – (source : based on Fontaine *et al.*, 2011)

The ‘Limoneuse’ (Loamy) agricultural region is the largest and the most fertile agricultural region in Walloon region due notably to sometimes relatively important proportion of loess soils. “Crops” is the main land use and represents most of time more than 70% of the surface. Winter wheat is the dominant crop (Figure 2.2). Nearly 50% of the national cultivated area in winter wheat can be found in this region. At Walloon region level, this proportion reaches 72.5% (table 2.1). The 3 other main crops in terms of importance are sugar beet, fodder maize and potatoes (all summer crops). Compared to Condroz agricultural region, the proportion of forests and grasslands is relatively low (Figure 2.2).

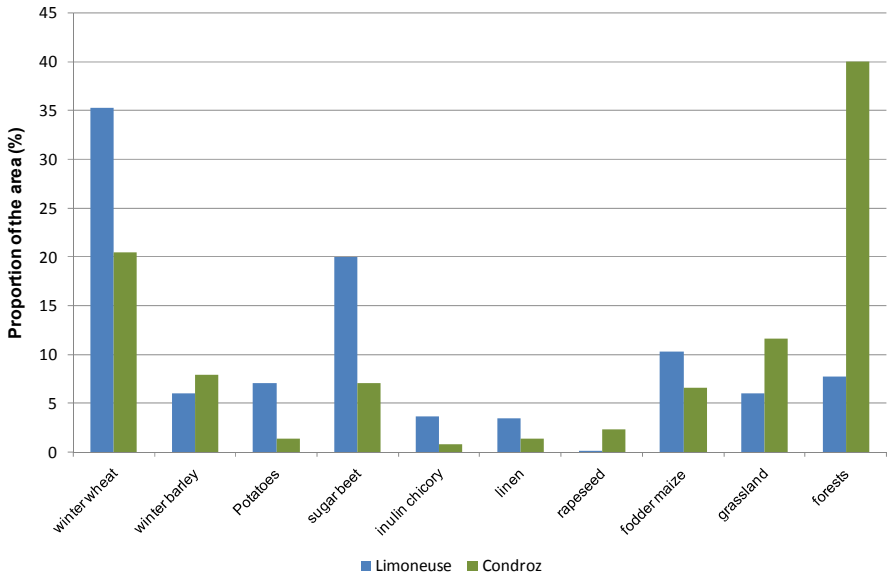


Figure 2.2.- Proportion of the area covered by the most representative crops and by forests according to the sum of area in Limoneuse (Walloon region part) and Condroz agricultural region (year 2002) – source : http://statbel.fgov.be/

The Condroz is characterized by a hilly landscape with clay outcrops on limestone in the valleys favourable to grasslands (representing more or less 40% of the Condroz cultivated area) and loamy beds on sandstone allowing the cultivation of cereals, sugar beet and oilseed. Condroz agricultural region is mainly covered by “crops” (roughly 60%) but also by a significant proportion of forests (25%). As far as “crops” are concerned, a distinction has to be made between arable crops and grasslands. Grasslands and arable crops share this territory. Grasslands area is roughly the half of the winter wheat area. Condroz is the second agricultural region for winter wheat production in Walloon region as it represents 23.0% of its cultivated area. In Walloon region, Limoneuse and Condroz agricultural regions represent therefore both 95.5 % of cultivated area in winter wheat.

Table 2.1.- Average cultivated area in winter wheat (period 2000-2009)

	Agricultural region		Walloon region	Belgium
	Limoneuse	*Condroz*		
winter wheat area in ha	92,277	29,438	127,341	194,593
Prop. cultivated area (% Belgium)	47.4	15.1	65.4	100.0
Prop. cultivated area (% Walloon region)	72.5	23.0	100.0	

Unlike the Limoneuse agricultural region Condroz, despite a relative homogeneity in terms of growing conditions, is subdivided in two sub-regions (Figure 2.3).

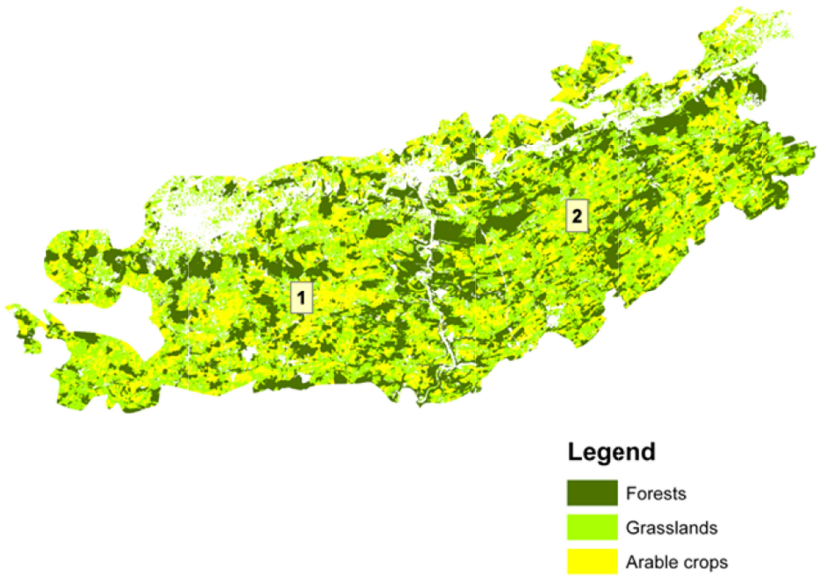


Figure 2.3.- Land use (main classes) of the Condroz agricultural region - source: COSW (version 2.07, 2011) – 1: western part of Condroz, 2: Eastern part of Condroz

Western part of Condroz is characterized by large productive terraces, with low stressed relief. Arable crops are predominant in this area. Eastern part of Condroz is characterized by an alternation between arable crops and grasslands, with punctual forests and shrubs elements. Forests/woody elements and grasslands are more present in this part of the Condroz.

These land use differences between and within agricultural regions are of great interest to study the impact of pixels purity and gridding effects. The hypothesis to

test is that the (nearly) “crop specific” GAI time series are also influenced by the signal not coming from the target crop (winter wheat) and that this influence depends on the land use type.

a. Climatic conditions

From a meteorological point of view, slight differences exist between the two agricultural regions. Table 2.2 presents the sum of effective temperatures from the 1st January to the 30th of May and from the 1st of June to the 30th of August on the MODIS acquisition period (2000-09). Effective temperatures (Te) are computed considering a base temperature (Tbase) below which no development occurs. Above a certain maximum effective temperature ($T_{max,e}$), T_e remains constant.

$$\begin{cases} T_e = 0 & T \leq T_{base} \\ T_e = T - T_{base} & T_{base} < T < T_{max,e} \\ T_e = T_{max,e} - T_{base} & T \geq T_{max,e} \end{cases}$$

Base and maximum effective temperatures for winter wheat have been fixed to 0° and 30°C respectively. End of May/beginning of June is the period when heading of winter wheat occurs in Belgium and August is usually the harvest month.

Temperature is a primary factor affecting crop development. The total effective temperature (also called accumulated thermal time) to which a plant has been exposed is commonly used by modelers to quantitatively assess the crop phonological development (Hodges, 1990).

Whatever the year, the sums of effective temperatures are higher for the Loamy region, the average daily difference of effective temperature between the two regions being equal roughly to 0.5°C d. October to December period covers the sowing and part of the season which is not modelled (simulations in WOFOST start for winter crops the 1st of January).

Table 2.2.- Average sum (standard deviation) of effective temperatures on 2000-09 period by agricultural region (in °C d)for October to December, January to May , June to August and from January to August periods

Agric. region	Jan.-May	June-Aug.	Jan.-Aug.	Oct.-Dec.
Condroz	*1130 (134)*	*1045 (72)*	*2175 (131)*	*670 (85)*
Loamy	*1212 (131)*	*1068 (76)*	*2280 (125)*	*717 (80)*

2.2. MODIS GAI time series

2.2.1. Terra/Aqua MODIS observations

Launched in 2000, the MODIS instrument that is aboard the Terra platform began a new era in remote sensing by providing an interesting balance of spatial, temporal and spectral resolutions. MODIS provides observational information in the following

36 spectral bands: 29 bands at 1-km spatial resolution at nadir, 5 bands at a resolution of 500 m and 2 bands at a resolution of 250 m. The MODIS swath enables the instrument to scan the Earth over a range of 2340 km, thereby providing global coverage every 1–2 days. In 2002, a second MODIS instrument was launched on the Aqua platform. The advantages of the MODIS data are that they are pre-processed, free of charge and readily available to the scientific community. As the agricultural landscape in Belgium is heavily fragmented, only the 2 bands at 250-m resolution, corresponding to reflectance in the red and near infrared spectral bands, are used in this paper. These data are part of the MOD09 collection 5 products and are distributed using the L2G grid, which has squared pixels in a sinusoidal projection.

2.2.2. Selection of suitable MODIS grid cells

Given the complex acquisition geometry of the MODIS observations (Tan *et al.*, 2006; Wolfe *et al.*, 1998) and given that the spatial sampling requirements to monitor the Belgian landscape are similar to the nominal MODIS spatial resolution, extra care needs to be applied when relating SIGEC (IACS) database with the MODIS L2G grid. SIGEC database was created by the government of the Walloon region and indicates what crop has been sown in every field in a given year In order to ensure grid-target adequacy, the method proposed by Duveiller and Defourny (2010) was used. This approach consists in convolving a kernel modelling the MODIS spatial response over the detailed wheat masks constructed from the SIGEC database. The result is a pixel purity map that provides an estimate of the percentage of the information obtained from surfaces that are covered with the target crop winter wheat for every cell in the MODIS grid. The cells in the grid with a purity of at least 75% are retained; for each of these valid locations, a time series is constructed by gathering the pixels in all of the available daily MODIS images. These time series are further filtered by screening out any pixels that do not correspond to the highest quality standards based on the MODIS flags and to some thresholds in the view zenith angle (*VZA*) and the Observation coverage (*Obscov*).

2.2.3. Generation of crop specific GAI time series from MODIS

GAI values have been retrieved over Walloon region from the multispectral reflectance using an algorithm developed for CYCLOPES (Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites) products (Baret *et al.*, 2007) for the years 2000 through 2009. This approach was successfully used by Duveiller *et al.* (2011b) to retrieve crop specific GAI at high spatial resolution along the entire growing season and scaled up to MODIS data by Duveiller *et al.* (2011a).

With a view to discard MODIS GAI estimated time series deemed as unsuitable for assimilation in WOFOST model, De Wit *et al.* (2012) have applied a filtering procedure based on the Canopy Structural Development Model (CSDM) (Koetz *et al.*, 2005). The filtering procedure assesses the GAI time series using six criteria designed on the basis on the a priori knowledge of the agricultural characteristics of winter wheat cultivation in the area and on an exploratory analysis of satellite GAI estimates.

The selection procedure has severely restricted the initial dataset. The percentage of remaining GAI observed time series ranged from 11.4% (in 2000) and 56.7% (in 2002), the average being equals to 22.8%.

2.3. Spatially distributed crop growth model

2.3.1. Crop growth model

We used the WOFOST (WOrld FOod STudies) crop simulation model (Diepen *et al.*, 1989; van Ittersum *et al.*, 2003). WOFOST is a mechanistic crop growth model that describes plant growth using light interception and CO2 assimilation as growth driving processes and crop phenological development as a growth controlling process.

The leaf photosynthetic active area in WOFOST is calculated by multiplying the Specific Leaf Area (SLA, in ha kg−1) by the living leaf biomass (in kg/ha) for each leaf age class. Together with leaf biomass, the living stem and pod biomass can contribute to the photosynthetic active area and is calculated by multiplying stem and pod biomass by the Specific Stem Area (SSA) and Specific Pod Area (SPA). The sum of the leaf area, pod area and stem area is stored in WOFOST as the variable 'Leaf Area Index'. With respect to photosynthesis, stem area and pod area are treated similarly to leaf area. In practice, however, the contribution of stem and pod biomass to the leaf area index is zero, as the values of the SPA and SSA parameters are zero in the European model setup. Therefore, a possible inconsistency exists between the WOFOST modelled Leaf Area Index and the satellite estimated green area index (GAI) (De wit *et al*., 2012).

However, the contribution of living stem and pod is probably negligible and considering moreover that the yellowing of pods is not calculated in WOFOST, it could induce inconsistencies at the end of the growing season. In this study we have subsequently chosen to match the WOFOST modelled LAI directly with the satellite GAI and therefore not to adjust the SPA and SSA parameters in the WOFOST model as the values of SSA and SPA are not well known. In the remainder of this paper, the term 'GAI' will be used.

2.3.2. Spatial implementation of crop growth simulations

The crop simulation model was implemented spatially by building upon the framework provided by the Crop Growth Monitoring System (CGMS). CGMS allows the regional application of WOFOST by providing a database framework that handles model input (e.g. weather, soil and crop parameters), model output (crop indicators such as total biomass and leaf area index), aggregation to statistical regions and yield forecasting (Vossen and Rijks, 1995; Genovese, 1998; Boogaard *et al*., 2002; de Wit *et al*., 2009). CGMS is part of the MARS (Monitoring Agricultural Resources) crop yield forecasting system that was developed by the Agri4Cast Action of the Joint Research Centre in Ispra, Italy. Since 1994, CGMS has monitored crop growth in Europe, Anatolia and the Maghreb with a spatial resolution of 25 km × 25 km and a temporal resolution of one day. The primary purposes of the CGMS are to provide information regarding weather indicators and crop status during the growing season

and crop yield forecasts at the level of EU member states early in the crop growth season. The principal user of the CGMS is the European Commission's Directorate General for Agriculture (De Wit *et al*., 2012).

2.4. Assimilation approach

Assimilation approach is based on an optimization method aiming to minimize a given objective function expressing the deviation between simulated and remotely-sensed estimation of GAI through the adjustment of selected model parameters. The two selected parameters are the TDWI and SPAN parameters.

The TDWI parameter strongly influences the initial growth rate and has a high degree of uncertainty due to two reasons. First, because the WOFOST model begins on January 1st, the growth of wheat in autumn is not taken into consideration (contrary to PyWOFOST used in chapter 1 (which explains why this parameter is used instead of crop emergence). Second, the effects of winter conditions on standing crop biomass are not taken into consideration. Variability of the TDWI parameters strongly influences the rate of increase of the crop GAI and also affects the maximum GAI that can be reached during the growing season. The importance of calibrating this parameter in the WOFOST model was demonstrated for winter wheat by Yuping *et al.* (2008) for the North China Plain.

The SPAN parameter determines the rate at which green leaves will turn brown and therefore determined the rate of crop senescence. Although this parameter is a characteristic of the variety of wheat, the senescence rate is also influenced by drought, fertilisation (nitrogen shortage decreases SPAN) and pest/disease infestation. Therefore, variability in crop management influences the SPAN parameter.

Together, the TDWI and SPAN parameters make it possible to model the increasing and decreasing branches of the GAI trajectory and the GAI maximum. Because these parameters operate separately on the increasing and decreasing branches, little correlation between the two parameters can be expected while still providing large flexibility in adapting the shape of the GAI trajectory. These two parameters are however not able to correct the phenology and notably the moment of the GAI peak. Figure 2.4 presents the evolution of simulated GAI for 5 different values regularly distributed within the respective valid ranges of the parameters (considered separately, the non-considered parameter remaining at its default value).

The optimization in De Wit *et al.* (2012) was implemented by minimizing the Weighted Mean Absolute Error (WMAE) between the modeled GAI (GAI_m) and the MODIS-observed GAI (GAI_o):

$$WMAE = \frac{\sum_1^n w.|GAI_o - GAI_m|}{\sum_1^n w}$$

n is the number of observations within the considered MODIS-observed GAI time series and *w* is a weighting factor used with a view to face the unevenly distribution of GAI MODIS observations that can be occasionally observed. Indeed, if the

increasing branch of MODIS GAI includes more observations than the decreasing branch, using equal weights would induce to give more weight to this increasing branch (and vice-versa) which is not desirable. The relative weight associated to each observation is therefore inversely proportional to the number of observations found in a temporal window of 11 days around the considered observation. In the optimisation procedure GAI values below 0.5 m²/m² were discarded.

The optimisation was implemented using the COBYLA (Constrained Optimisation BY Linear Approximations) algorithm in the NLopt Python library (Johnson, 2011; Powell, 1998). COBYLA was constrained to search for parameter solutions over valid ranges of the SPAN (20 ≤ SPAN ≤ 50 days) and TDWI (50 ≤ TDWI ≤ 500) parameters, and the combination of SPAN and TDWI that minimised the objective function was used. The default values for SPAN and TDWI (35 days and 210 kg/ha, respectively) were used as the starting point for the COBYLA algorithm.

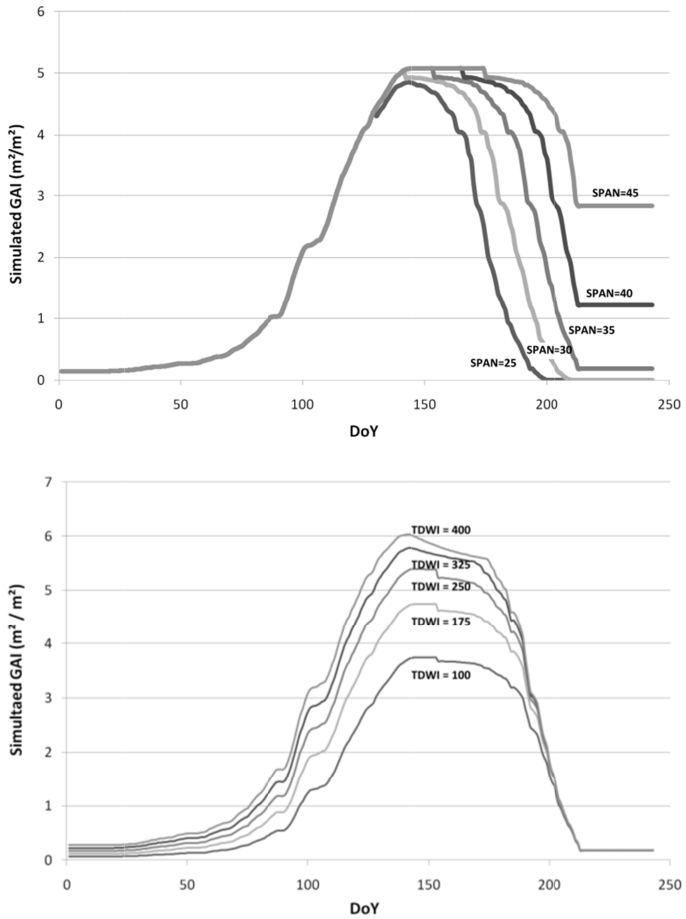


Figure 2.4.- Evolution of simulated GAI according to SPAN (above) and TDWI (above) values

Simulations presented in Figure 2.4 have been generated for a grid located in the Loamy region in 2002. Variations of recalibrated parameters values induced variations in simulated yields (Figure 2.5). The contour plot has been generated by considering a step of 2 days for SPAN parameter within the 20-50 days range and a step of 25 kg/ha for TDWI parameter within the 50-500 kg/ha range.

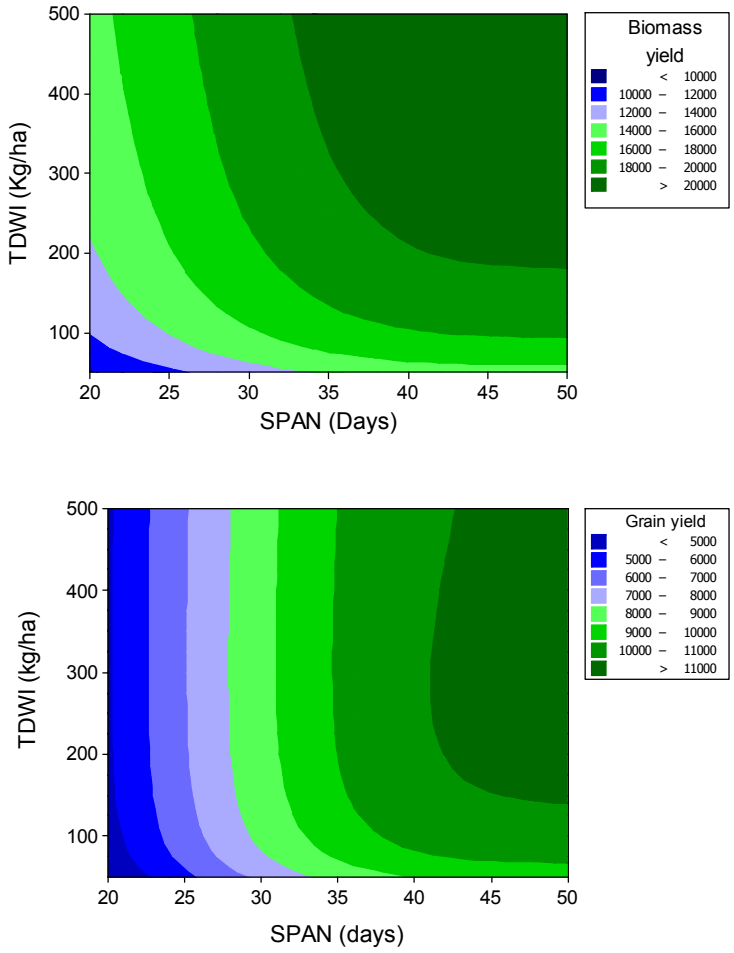


Figure 2.5.- Evolution of simulated biomass in kg/ha (above) and grain (below) yields according to SPAN and TDWI (below) values

Simulated grain yields are more sensitive to SPAN parameter compared to TDWI which influences mainly simulated biomass yield. For grain yield, we can observe that for a given SPAN value TDWI has nearly no influence on simulated yields and that for a given TDWI value the effect of a variation of SPAN value on grain yield tends to be less important for higher SPAN values.

3. Methods

The assimilation procedure has been applied to all of the MODIS grid cells with suitable GAI time series over all of the available years. For each MODIS grid cell and each year, the WOFOST model has been executed using the soil, weather and crop information corresponding to the 10 x 10 km model grid at which the observations were obtained. The GAI observations for each of these MODIS Grid cells have been also retrieved, and the optimisation algorithm executed to obtain the minimum value of the objective function. The approach yielded a combination of SPAN and TDWI as well as the WOFOST-simulated GAI and biomass values for the years 2000 to 2009.

Considering the difference of land use between the 2 considered agricultural regions, a first exploratory analysis has been conducted to assess if the spatial variability of the recalibrated parameter values can be associated to this difference of land use. In order to reach this objective, SPAN and TDWI recalibrated values have been averaged at the 10 x 10 km model grid and mapped. Only model grids holding at least 10 MODIS grid cells have been considered. The spatial patterns were afterwards visually explored to assess if a difference in recalibrated parameter values can ne also observed between the two agricultural regions.

A more detailed study has been afterwards performed to assess the effects of the incomplete purity (up to 25% of the information for a MODIS grid is not specific to winter wheat) and of the possible remaining gridding effects on assimilation results. The contribution of the two effects can't unfortunately be dissociated. By sake of simplicity, we will use the term of "contamination of the MODIS grid cells" for these two effects considered as a whole. Analyses have been performed at two different scales. Analyses were first performed at model grids level and afterwards at a more detailed scale i.e. at MODIS grid cells. The analyses have been performed for year 2002 as this year presents the highest number of suitable MODIS grid cells for both agricultural regions.

In the frame of the analyses at model grids level, the proportion of the different land cover within each 10 x 10 km model grid has been computed on the basis of the land use map of Wallonia (COSW version 2.07, 2011) for forests and the yearly Belgian LPIS (Land Parcels Identification System) maps for crops and grasslands.

MODIS grid cells contamination has been assessed through the relationships between the recalibrated parameters values and the corresponding proportion of land cover at simulation grid level. Only land cover expected to have an important contribution on MODIS grid cells contamination were considered. These considered land covers are summer crops, grasslands and forests. For summer crops, only the 3 main crops of this category have been considered namely maize, potatoes and sugar beet.

A proportion of the different land covers has been also computed for the analysis at MODIS grid cells level. Contrary to the previous analysis, this proportion has been computed within buffers circumscribing the MODIS pixels (at nadir) considered as proxies for the PSF.

In order to separate at best the effect of the contamination of a MODIS grid cell by a given land cover (e.g. summer crops) from the effect coming from another land cover (e.g. grasslands), a selection (Figure 2.6) of 100 MODIS grid cells has been performed for each of the consider land cover. The selection has been done on the maximization of the sum of the percentages of the considered land cover and of winter wheat as well as the percentage of the considered land cover (within the buffer).

For each day of observation, the average GAI values by category of land covers were then computed. The resulting GAI time series for winter wheat was visually compared to the resulting GAI time series for the 3 other land covers. Only day of observations presenting at least 40% (40 observations) in the two compared categories of land covers have been displayed.

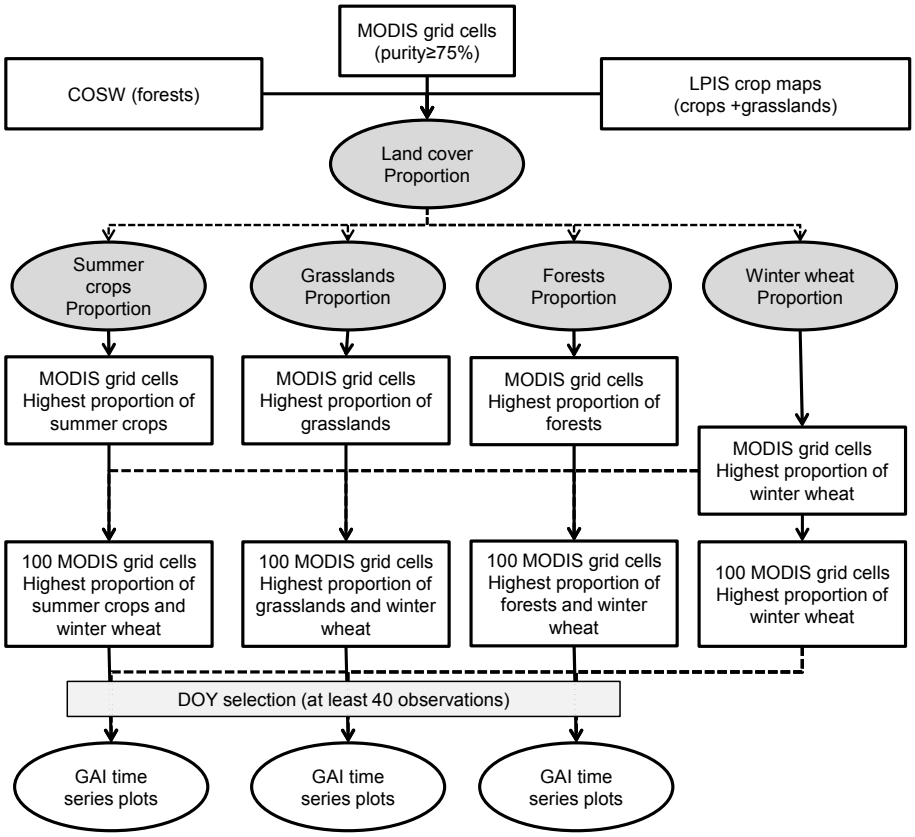


Figure 2.6.-MODIS grid cells selection procedure (MODIS grid cells contamination study)

An analysis of the codistribution of parameters values has been finally performed.

4. Results

4.1. Spatial distribution of the recalibrated parameters

Maps of recalibrated parameters values tend to present a spatial pattern (Figure 2.7, year 2002 displayed). Differences of recalibrated parameters values, averaged at the 10x10 km model grid, can be observed between agricultural regions but also within agricultural regions. The distinction between the two agricultural regions can be clearly seen. Within the Condroz agricultural region, a distinction seems also to be observed between its eastern and western parts.

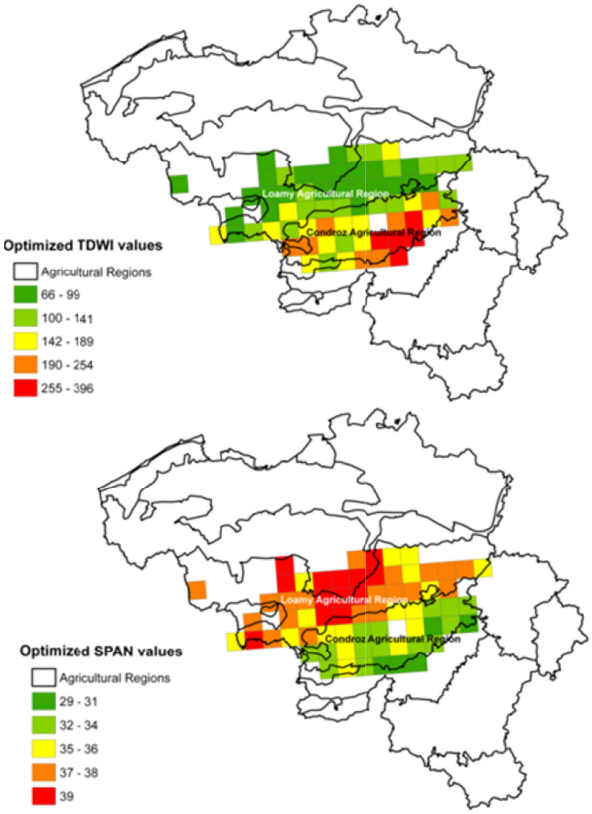


Figure 2.7.- Average TDWI and SPAN values per 10 x 10 km model grids for year 2002 (only model grids with at least 10 MODIS grid cells are considered)

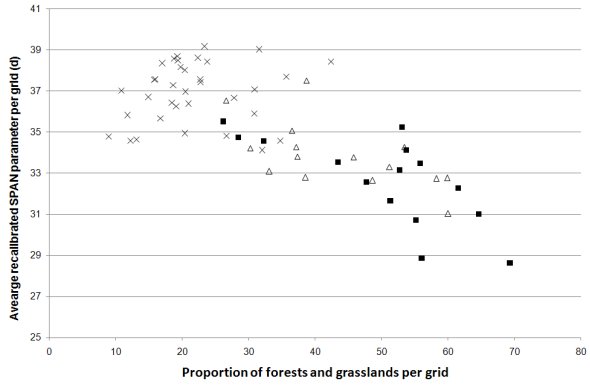
Explaining this spatial variability is of prime importance. The main question is to know if the variability observed corresponds to the field reality or not. The observed variability seems to be too important to be only imputed to the "natural" variability (i.e. the variability really observed on the field). The incomplete purity and gridding effects could be some of the factors contributing to this "higher than expected" variability. This statement is assessed in the next paragraphs.

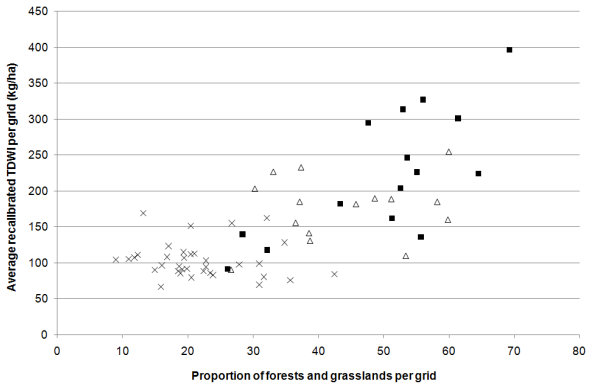
4.2. MODIS grid cells contamination (model grids level)

Observed spatial patterns of recalibrated parameters in figure 2.7 seem to match, at least visually, with spatial patterns observed in the land use maps presented in figures 2.1 and 2.3. In order to test the possible influence of landscape, supposed to be linked to the incomplete purity and to a remaining gridding effect, on the MODIS observations and subsequently on recalibrated parameters the relationships between the average recalibrated parameters values and the percentage of forests and grasslands per model grid have been assessed (figure 2.8). Proportion of forests and grasslands has been considered as these 2 classes represent the two main classes besides the arable crops class. It also represents evergreen elements that could "influence" the crop specific signal all along the growing season.

The relationship between the average date of MODIS GAI peak and the percentage of forests and grasslands was also assessed. Date of maximum GAI values for the different MODIS GAI time series were estimated on the basis of the CSDM model used in the filtering procedure.

Figure 2.8 shows (for year 2002) the influence of landscape on recalibrated parameters and also on the moment of GAI peak. Lower SPAN values (higher TDWI values) are observed when the proportion of grasslands and forests is high.





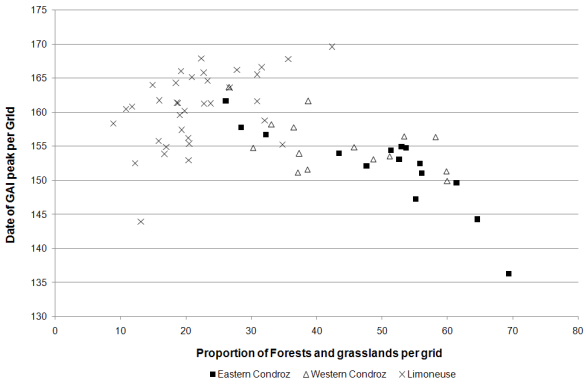


Figure 2.8.- Relationship between the average recalibrated SPAN and TDWI values and the average date of GAI peak with the proportion of forests and grasslands per model grid (year 2002)

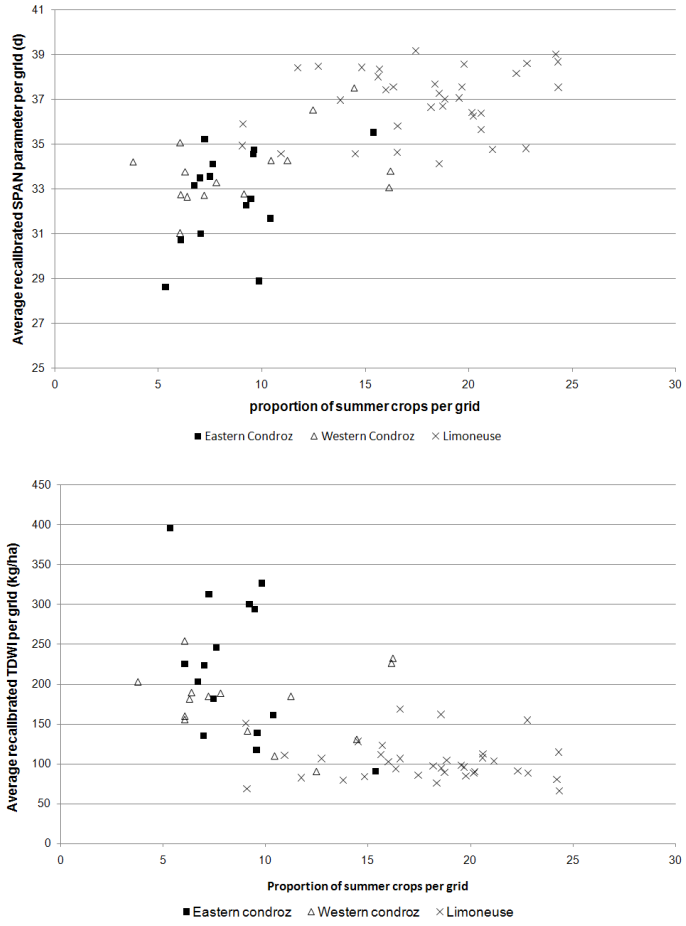
Let's note that the maximum GAI value is, whatever the year, reached significantly earlier in Condroz agricultural region compared to Limoneuse agricultural region which is contradictory according to reality. Indeed, phenological data collected in variety trials of the Walloon agricultural research centre (CRA-W) in Gembloux (Limoneuse agricultural region) and Ciney (Condroz Agricultural region) for 15 years between 1987 and 2004 show no differences of heading date (phenological moment closes to the moment of maximum LAI). The same varieties have been used in both sites for the same year. The average heading date for Ciney and Gembloux is respectively equals to DOY 152 and 153. The average and the standard deviation of the difference of average heading date per year (Ciney minus Gembloux) is respectively equal to -0.5 and 3 days.

No agronomic (sowing dates…) or meteorological information (tables 2.2 and 2.3) can explain this contradictory observation. The WOFOST model runs on 2000-09 period with the default parameters provides an average difference of 4 days between the peaks of GAI of Limoneuse and Condroz agricultural regions, the peak of GAI in Limoneuse agricultural region occurring earlier. The fact that up to 25% of the information does not come from winter wheat crop is a hypothesis for explaining this situation.

The range of dates of GAI peak observed in figure 2.8 is also too wide to be associated to the natural variability of this phenological parameter all the more since it is average values (per model grid) that are displayed and is aforementioned in contradiction with field observation. The hypothesis of a "contamination" of MODIS GAI time series by land use other than winter wheat, in this case grasslands and forests, could be therefore a plausible explanation for these relationships. We can however also observe that the average date of GAI peak provided by WOFOST without assimilation is equal to 140 and 146 respectively for Limoneuse and Condroz agricultural regions in 2002. The corresponding dates estimated on the basis of the MODIS GAI time series are in average equal to 161 and 154, the gap is particularly important especially for Limoneuse region (140 vs. 161). Other factors have therefore

to play a part. Considering the importance of summer crops in Loamy region (Figure 2.3), a possible 'contamination' of this category of crops is also to consider.

If we plot the average per model grid of the recalibrated SPAN and TDWI values and of the dates of GAI peak (estimated by CSDM) against the proportion of summer crops, inverse relationships than the ones obtained with grasslands and forests can be observed (Figure 2.9).



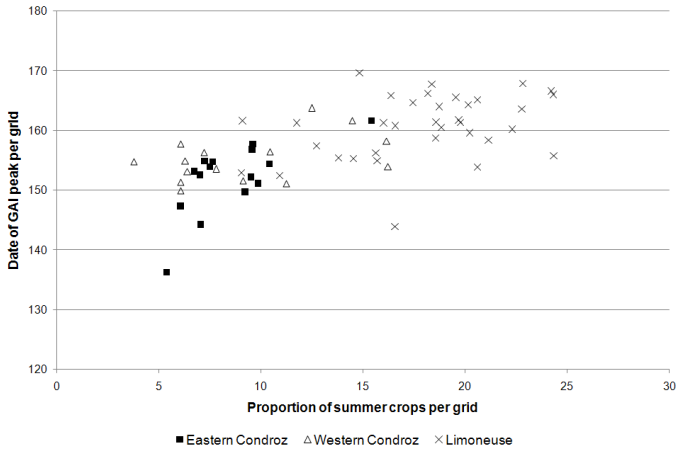


Figure 2.9.- Relationship between the average recalibrated SPAN and TDWI values and the average date of GAI peak with the proportion of summer crops per model grid (year 2002)

From these different relationships, we can obviously conclude that, despite the pre-selection of MODIS grid cells notably according to the purity in terms of winter wheat occupation, the surrounding landscape seems to have an influence on observed GAI values and subsequently on recalibrated parameters. It is however not possible at this level to know exactly which is exactly the (most) influencing landscapes but also how the different landscapes / land uses influence the observed GAI time series.

4.3. MODIS grid cells contamination (MODIS grid cells level)

The selection procedure of MODIS grid cells for winter wheat (figure 2.6) led to a final set of 100 pure grid cells: 99.9% of the buffers area is covered by the crop (standard deviation of 0.4%).

These MODIS grid cells were compared to selections of 100 MODIS grid cells presenting respectively a high proportion of summer crops, grasslands and forests. The comparison is concerned with the time series of the average GAI values.

4.3.1. Summer crops

The average proportion of summer crops in the selected 100 MODIS grid cells is equal to 18.1% (standard deviation of 5.8%), the average proportion of winter wheat being equal to 77.3% (standard deviation of 7.1%).

Figure 2.10 presents the time series of the average GAI values for pure and summer crops “contaminated“ MODIS grid cells.

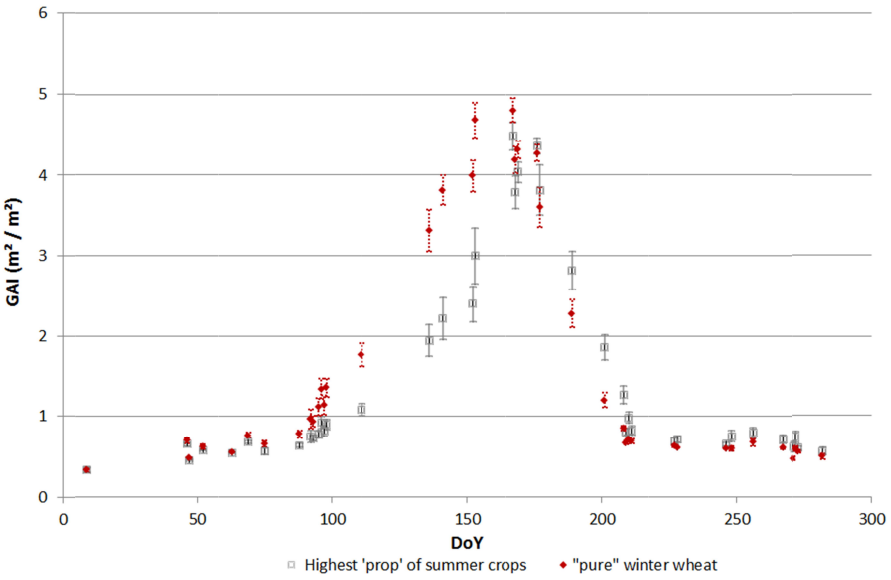


Figure 2.10.- Time series of average observed GAI values (and 95% confidence intervals of the mean) for pure winter wheat MODIS grid cells and for MODIS grid cells presenting the 10% highest proportion of summer crops (year 2002)

The average GAI values are lower in the increasing phase of GAI for MODIS grid cells presenting a high proportion of summer crops but these values tend to be higher in the decreasing phase. The maximum GAI value tends to occur earlier in case of pure winter wheat MODIS grid cells.

The explanation for the difference of GAI profiles between the situation with pure winter wheat MODIS grid cells and the situation with MODIS grid cells presenting a higher proportion of summer crops could rely in the difference of phenological development between winter and summer crops. Start of growing season and therefore of GAI evolution for summer crops occurs at best in April. Before and just after this period, for the MODIS grid cells presenting a high proportion of summer crops, a part of the surface within these MODIS grid cells does not contribute (or only a few) to global GAI value and it can explain why the GAI values for MODIS grid cells presenting a high proportion of summer crops are lower than the GAI values for "pure winter wheat" MODIS grid cells before heading. Maximum of GAI values for summer crops occurs also later, after the winter wheat heading (heading being roughly the moment when winter wheat reaches its maximum GAI values). Therefore after the winter wheat heading, when for "pure winter wheat" MODIS grid cells the GAI values start to decrease, a part of the surface observed by MODIS grid cells presenting a high proportion of summer crops contribute to an increase of GAI values which could explain first of all the slight discrepancies observed near the maximum GAI values and also the highest GAI values observed for MODIS grid cells presenting an higher proportion of summer crops in the decreasing phase of GAI.

The time shift of the GAI peak occurring in case of MODIS grid cells contamination by summer crops could also explain why the average GAI peak is observed later in Limoneuse agricultural region compared to the Condroz agricultural region while it is the contrary which is expected. If we look at the proportion of MODIS grid cells including at least 5% of summer crops in the 2 agricultural regions, we can observe that these MODIS grid cells represent roughly 26% of the total number of considered MODIS grid cells in Loamy agricultural region and only 12% in Condroz agricultural region. As the other sources of MODIS grid cells contamination do not particularly tend to change the moment of the observed GAI peak, it is not too surprising in these conditions to observe a later date of GAI peak for Limoneuse agricultural regions.

4.3.2. Grasslands

The MODIS grid cells presenting a higher proportion of grasslands (average proportion of grasslands: 13.4% - standard deviation : 4.9%) tend to present a very similar pattern (Figure 2.11) to the one observed for pure winter wheat MODIS grid cells which is not too surprising as the 2 land cover tends to present rather similar signature for most of the season.

Looking more into details, small shifts can be observed. The MODIS grid cells presenting a higher proportion of grasslands tends to exhibit higher average GAI values at the beginning and the end of the growing season and lower average GAI values in-between. This behaviour seems to be coherent with what could be expected. Grasslands represent indeed evergreen elements. At the beginning of the growing season most of the grasslands, whatever their type of management (mowing or grazing), looks like winter wheat except that the soil can be considered as fully covered by this type of vegetation. It is therefore not too surprising to observe at this moment slightly higher values for MODIS grid cells presenting an higher proportion of grasslands. As the winter wheat starts growing, its GAI increases as well and become probably higher in average than those observed in grassland notably due to the grazing and mowing operations which tends to decrease GAI value. At a certain moment in the decreasing phase of winter wheat GAI, the GAI for MODIS grid cells presenting an higher proportion of grasslands tends to become higher than the GAI for “pure” winter wheat MODIS grid cells probably due to the fact that grasslands are still growing at this period. Considering the lack of precise information on grasslands management, it is unfortunately not possible to validate this hypothesis.

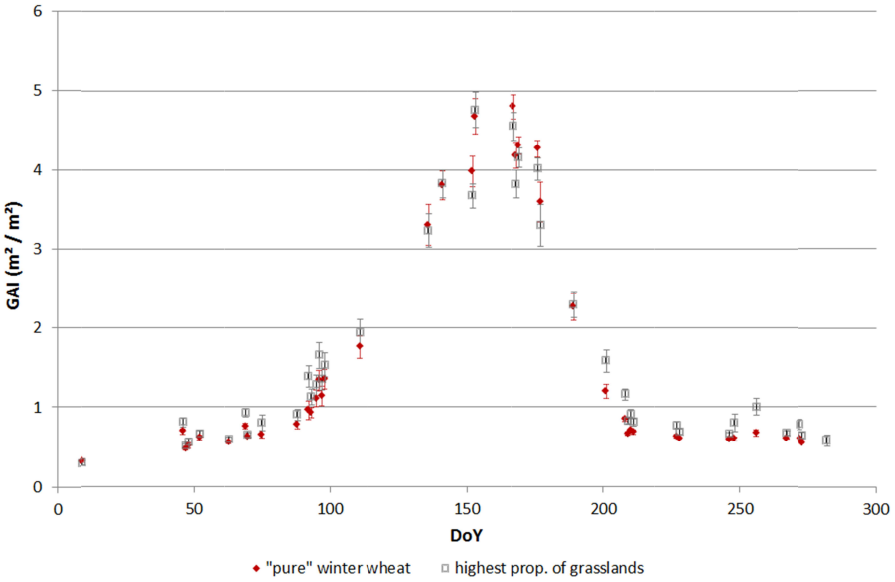


Figure 2.11.- Time series of average observed GAI values (and 95% confidence intervals of the mean) for pure winter wheat MODIS grid cells and for MODIS grid cells presenting the 10% highest proportion of grasslands (year 2002)

4.3.3. Forests

The average proportion of forests for the MODIS grid cells presenting a higher proportion of forests is equal to 8.5% (1st quartile: 2.9 % - 3rdquartile: 10.9%). The proportion of winter wheat for the MODIS grid cells presenting a higher proportion of forests is rather high with an average of 87.4%.

Evolution of average GAI values for MODIS grid cells presenting a higher proportion of forests (figure 2.12) is similar to the one observed for MODIS grid cells presenting an higher proportion of summer crops i.e. lower average GAI values compared to situations of pure winter wheat MODIS grid cells in the increasing phase of GAI and in the peak region and higher GAI values in the decreasing phase. The differences of GAI values are however smaller. The similarity with the summer crops can be explained by the fact that broadleaves trees represent the majority of tree species (at least 80%, CPDT (2011)) as budbreak of broadleaves trees (and subsequently the start of the contribution to GAI) occurs in spring. The lower differences of GAI values can probably be explained by the lower proportion of forests within the buffers.

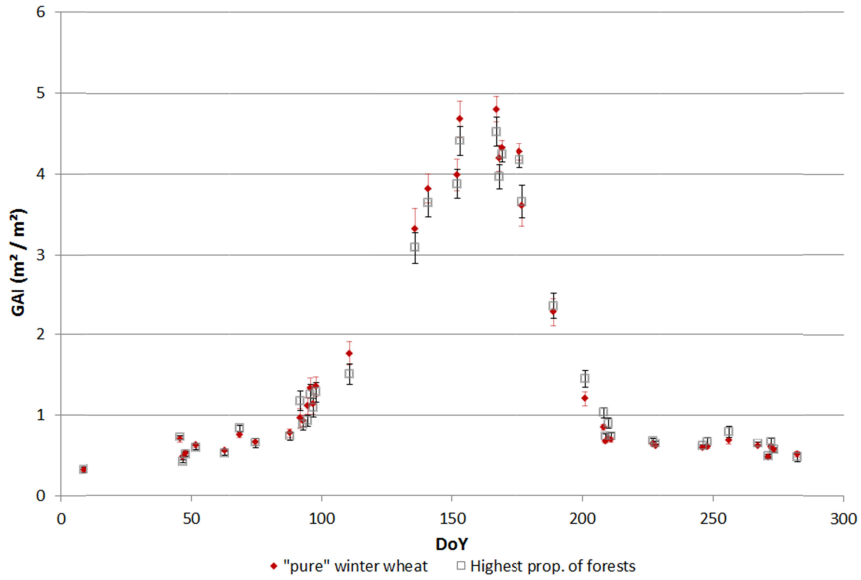


Figure 2.12.- Time series of average observed GAI values for pure winter wheat MODIS grid cells and for MODIS grid cells presenting the 10% highest proportion of forests (year 2002)

4.3.4. Conclusions

The MODIS grids contamination tends to have a significant impact on observed GAI time series, especially the contamination by summer crops which tends to induce a phenological shift of the GAI time series. The MODIS grids contamination by summer crops tends indeed to shift the observed GAI time series later in the season. The observed GAI peaks are for example observed later in the season for MODIS grids presenting a higher proportion of summer crops and later than the expected range of dates for GAI peak in winter wheat. This phenological shift has an impact on assimilation procedure as seen in paragraph 4.1. The time shift of the GAI peak occurring in case of MODIS grids contamination by summer crops could also explain why the average GAI peak is observed later in loamy agricultural region compared to the Condroz agricultural region while it is the contrary which is expected. If we look at the proportion of MODIS grids including at least 5% of summer crops in the 2 agricultural regions, we can observe that these MODIS grids represent roughly 26% of the total number of considered MODIS grids in loamy region and only 12% in Condroz region. As the other sources of MODIS grids contamination do not particularly tend to change the moment of the observed GAI peak, it is not too surprising in these conditions to observe a later date of GAI peak for Limoneuse agricultural regions.

4.4. Analysis of recalibrated parameters co-distribution

Recalibrated parameters values per agricultural region and per year are presented in table 2.3. Both recalibrated TDWI and SPAN values present a spatial and temporal variability.

Table 2.3.- Average and standard deviation of recalibrated TDWI and SPAN values per year and per agricultural region (L: Limoneuse agricultural region – C :Condroz agricultural region)

			TDWI				**SPAN**			
Year	*Nb obs*		*Average*		*St. dev.*		*Average*		*St. dev.*	
	L	C	L	C	L	C	L	C	L	C
2000	429	25	135	199	70	66	32.5	31.8	5.0	4.9
2001	384	168	107	105	55	48	30.7	28.1	3.6	4.3
2002	1647	607	107	192	52	94	36.7	33.4	2.8	3.3
2003	1054	282	125	159	48	65	32.3	31.2	2.8	2.4
2004	942	308	213	271	118	125	31.9	30.3	3.1	3.5
2005	420	282	108	136	64	93	26.1	24.2	3.3	4.0
2006	258	139	165	242	103	134	25.1	24.7	3.3	3.2
2007	416	105	141	195	62	98	34.8	33.2	3.3	4.1
2008	855	466	116	214	45	97	35.3	32.3	2.5	2.7
2009	402	149	61	64	33	40	25.3	27.0	2.4	5.0

Recalibrated TDWI values, and therefore the initial amount of biomass at simulation start (the 1st January) are systematically lower for loamy agricultural region. TDWI is a parameter summarizing, at least partly, the crop growth in autumn and effects of winter conditions on crop biomass. Higher TDWI values for Condroz region seem therefore to show that this region is earlier concerning winter wheat growth development than loamy agricultural region. This observation is in contradiction with the field reality, notably with the observations available in table 2.2.

As far as SPAN parameter is concerned, contrary to the TDWI parameter, SPAN values are roughly systematically higher for loamy region than for Condroz. No field data are unfortunately available to confirm or contradict this observation.

Standard deviations tend to be more important, whatever the recalibrated parameter, in Condroz compared to the loamy region.

It is also important to mention that in average the recalibrated parameters values are most of time lower than the default parameters values (equals to 35 days and 210 kg/ha respectively for SPAN and TDWI parameters).

Despite the initial hypothesis of a relative independence between the two recalibrated parameters, the computation of correlation coefficients shows that very highly significant and negative relationships exist between the TDWI and SPAN parameters (table 2.4). The coefficients of correlation are most of time higher than 0.5 (in

absolute value). Let's note however that the coefficients of correlation are slightly biased by the constraints put on parameters recalibrated values (ranges of possible values defined for both parameters).

Table 2.4.- Coefficients of correlation between TDWI and SPAN parameters per year

Year	Loamy Reg.	Condroz Reg.
2000	-0.73***	-0.62***
2001	-0.45***	-0.65***
2002	-0.56***	-0.52***
2003	-0.62***	-0.33***
2004	-0.32***	-0.14*
2005	-0.49***	-0.64***
2006	-0.52***	-0.52***
2007	-0.35***	-0.39***
2008	-0.58***	-0.45***
2009	-0.09ns	-0.17*

As the theoretical relative independence between the two parameters can't be really questioned, explanations for such systematic relationships can be probably found in the recalibration procedure. One of the main consequences is therefore that both parameters loss their original meaning. In other words, it means that the recalibrated values for the two parameters can't be interpreted which could explain at least partly the contradictory results observed above.

In WOFOST model, maximum of GAI value is reached at simulated heading. Simulated GAI remains at its peak value for some days, usually for less than one week. An exploratory analysis of simulated GAI time series shows sometimes the existence of a particularly long plateau of simulated maximum GAI value, especially for years presenting higher SPAN values (lower TDWI values) as for example year 2002 and within a year for Loamy agricultural region.

Figure 2.13 illustrates this plateau for a selected MODIS grid cell in 2002 a simulated GAI time series resulting from SPAN/TDWI recalibration and the corresponding GAI observations used for this calibration. The existence of a plateau at maximum simulated GAI value is clearly noticeable.

Figure 2.13 seems also to show a temporal discrepancy between observed and simulated GAI values, especially in the increasing phase of GAI. The maximum simulated GAI value is also lower than the maximum MODIS-estimated GAI values.

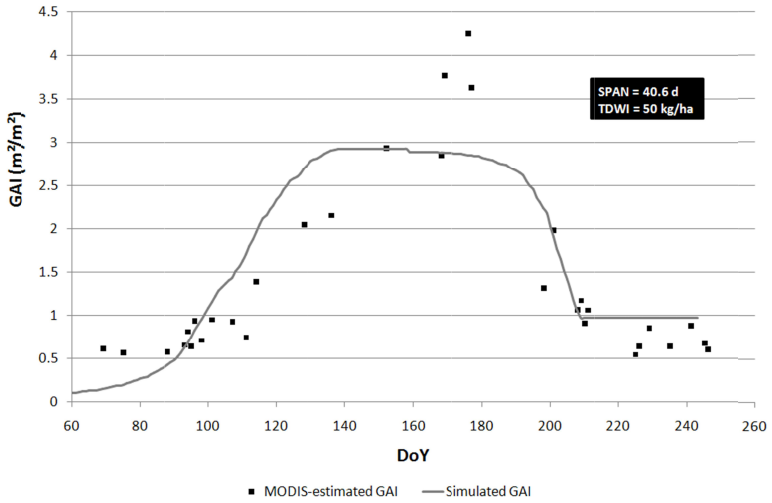


Figure 2.13.- MODIS-estimated GAI observations and simulated GAI time series after SPAN/TDWI recalibration for a MODIS grid in 2002

Situation presented in figure 2.13 is not an isolated case. For 2002, the average number of days to observe a GAI decrease of 0.1 m²/m² from the peak of GAI is equal to 25 and 13 days respectively for loamy and Condroz regions. The correlation of this number of days with SPAN parameter is strong, especially for loamy agricultural region: 0.83*** and 0.59*** respectively for loamy and Condroz agricultural regions

In order to have an order of magnitude of the time shift between (remotely-sensed) observations and simulations time series, we have estimated the day of maximum GAI values for the different MODIS-estimated GAI time series on the basis of the CSDM model used in the filtering procedure. For each MODIS grid in 2002, the difference between the days of maximum GAI value derived from MODIS and simulated GAI time series has been computed. The estimation based on CSDM model has to be considered carefully as the moment of the peak as well as the GAI value of this peak is dependent on the particular fitting between the CSDM sigmoid increasing and the decreasing exponential curves. It can however provide a first estimation of the maximum GAI timing.

The difference is in average equal to 19 and 9 days respectively for loamy and Condroz agricultural region (Figure 2.14). GAI peaks estimated from MODIS tend therefore to occur later than the GAI peaks estimated by simulation.

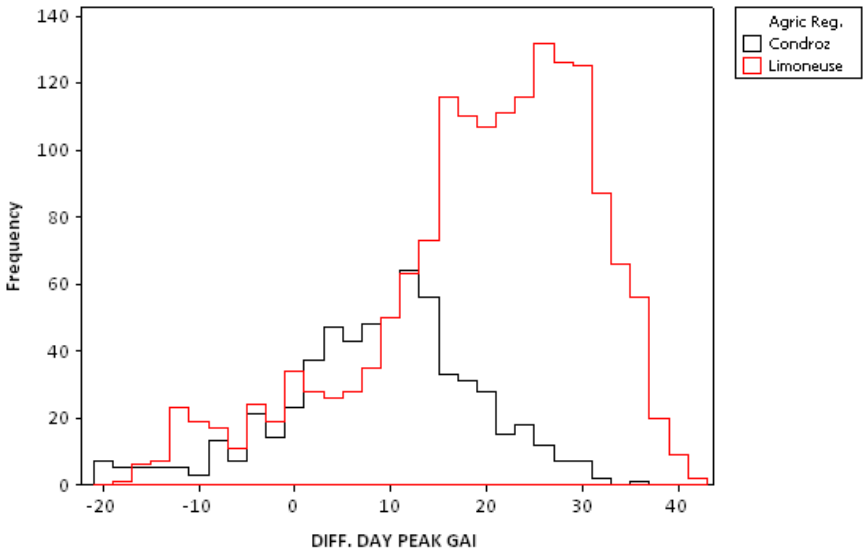


Figure 2.14.- Difference between (CSDM) estimated and simulated days of maximum GAI values for Limoneuse and Condroz agricultural regions in 2002

Let's note also that even if only pure winter wheat MODIS grid cells are selected, a phenological discrepancy between remotely-sensed and simulated GAI time series remains. In our study, maximum GAI value for the purest winter wheat MODIS grid cells in 2002 is observed at DoY 167. Considering that remotely-sensed observations are not available on a daily time step, no certitude exists on the fact that this date corresponds to the day of the GAI peak but it seems that this moment can be estimated (on average) between DoY 153 and 167. The average date of maximum GAI value estimated with the CSDM is equal to DoY 157 (values for the 1st and 3rd quartiles are respectively equal to 152 and 166) for these purest winter wheat MODIS grid cells. As far as WOFOST model is concerned, the peak of GAI is observed at last at DoY 146 (for Condroz agricultural region). The interquartile range (equal to 14 days) can be also considered as relatively wide. Possible explanations for such a range are first of all the variability in the sowing dates and the existence of different varieties (genotypes) but also the uncertainty existing on the estimation of the date of GAI peak by the CSDM fit which is dependent on the temporal availability and on the uncertainty of the GAI observations. Considering the medium resolution of MODIS sensor, information recorded for a given MODIS grid cells is likely to come from different parcels at different phenological stages. This probably influences GAI values, with a variable intensity according to the day of observation due to the existence of the gridding effects, and subsequently the global shape of the observed GAI time series (and therefore the estimation of the date of the GAI peak by the CSDM fit).

Phenological observations collected in 2002 by the Walloon Agricultural Research Centre in two varieties trials both located in Limoneuse agricultural region identify in

average the moment of the GAI peak (heading phenological stage) at DoY 151 (earlier observation: DoY 147, later observation: DoY 155) and DoY 154 (earlier observation: DoY 149, later observation: DoY 158) which is quite close to the CSDM estimation of GAI peak moment. Though it has to be validated on a larger data set, this could mean that phenology derived from remote sensing match with field reality. This assumption is strengthened by the results of Sakamoto *et al.*, 2010 who have observed, comparing the date of 4 key phenological stages derived from MODIS data time series with ground-based observations, a RMSE ranging between 3 and 7 days for maize and soybean.

The joint recalibration on SPAN and TDWI is unfortunately not able to correct phenology. It means that for a given simulation grid and a given year, the simulated day of maximum GAI value is not affected by the SPAN/TDWI recalibration. SPAN parameter influences mainly the decreasing phase of GAI. As this decreasing phase starts just after the maximum GAI value, the existence of a significant shift between simulated and observed GAI peaks (that we will call “phenological discrepancy” in the rest of the study), has therefore logically an influence on SPAN parameter recalibration. Indeed, an observed situation with a quick leaves fading after the peak could be seen as a favourable situation (high SPAN value meaning that the leaves stay green for a long time) if the modeled peak of GAI is significantly earlier than the observed one. In this situation, the simulated yield will be consequently overestimated (considering that the TDWI is unchanged).

Recalibrated TDWI value is logically also influenced by the phenological discrepancy. When the phenological discrepancy is important, recalibration of the model is therefore constrained to increase artificially the SPAN parameter value which constraints subsequently the recalibrated TDWI parameter value. This can probably explain why the maximum GAI value observed in figure 2.13 is low compared to observations and why the phenological discrepancy is essentially observed in the increasing phase of GAI.

The relationship between the difference of observed and simulated date of GAI peak and recalibrated SPAN parameter at agricultural region level on 2000-09 period is globally positive (figure 2.15). Higher SPAN values are observed for years where observed dates of GAI peak occur later than the simulated ones (and conversely). As the speed of leaves fading (characterized by the SPAN parameter) is supposed to be independent of the GAI peak date, this obvious relationship has low probability to be a matter of chance and illustrates once again the impact of phenological discrepancy.

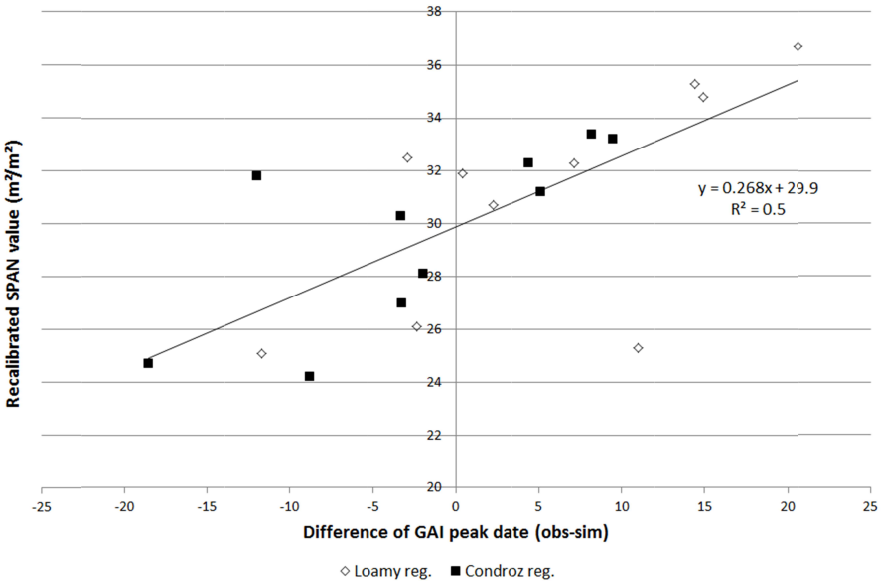


Figure 2.15.- Relationship between the recalibrated SPAN parameter and the difference of observed and simulated GAI peak dates (observed –simulated)

One of the main conclusions that can be drawn is that recalibrating jointly on SPAN and TDWI parameters is not recommended as far as no phenological adjustment is performed prior to this recalibration. It confirms some of the results of a previous study based on a synthetic data set (Curnel *et al*., 2011) where it was observed that recalibrating on SPAN parameter without adjusting the moment of the GAI peak (through a recalibration of TSUM1 or emergence date) was found inefficient.

5. Discussions and conclusions

This study aimed at assessing if and how the incomplete pixel purity influences the results of the assimilation and subsequently at understanding and identifying some of the elements that could improve the assimilation approach developed by De Wit *et al.* (2012).

The lack of purity is a consequence of the medium spatial resolution of MODIS sensors. Some of characteristics of the MODIS sensors, such as the PSF or gridding effects, can be also associated to the problems linked to the lack of purity. If a significant proportion of summer crops is observed within selected MODIS grid cells, it tends to (phenologically) shift the observed GAI time series, to decrease observed GAI values in the increasing phase of GAI and to increase these values in the decreasing phase of GAI. These effects that can be clearly identified despite a relatively low proportion of summer crops are mainly due a difference of crop calendars.

The assimilation procedure based on SPAN and TDWI parameters used in the analysis is not able to correct this phenological shift. The consequences of this phenological discrepancy uncorrected by the SPAN/TDWI recalibration are first of all that recalibrated parameters do not reflect the observed situations. For example, a situation of quick leaves fading could be seen as a normal or even an ideal situation (high SPAN value) if the phenological shift between the observed and simulated GAI time series is important. The phenological shift tends to provide higher SPAN (lower TDWI) values when the observed GAI peak occurs after the simulated GAI peak (situation of year 2002 for example) and tends to provide lower SPAN (higher TDWI) values when the observed GAI peak occurs before the simulated GAI peak. The importance of correctly simulating the moment of the GAI peak (and also its values) with a view to obtain a correct estimation of the SPAN parameter was already highlighted in Curnel *et al.* (2011) and in Kouadio (2012). It also raises the issue of equifinality of data assimilation and the related risk to misinterpretation of the recalibrated parameter values.

From a practical point of view, the direct repercussion on the assimilation procedure is that the rules used in the filtering procedure of GAI time series has to be strengthened. One step in this direction could be to select only MODIS grid cells selected within a pure or nearly pure winter wheat environment, i.e. increasing the purity level threshold currently fixed to 75% up to 90-95% for example. A compromise has however to be find so as not to lose the natural variability (due to differences of cultivars, of fertilization scheme…) that can be observed for a given year in a defined region. Being very strict on purity level will drastically decrease the number of selected MODIS grid cells especially if the landscape is highly fragmented. For example in 2002, selecting MODIS grid cells with a purity higher than 90% reduce the number of observations from 1647 in Limoneuse agricultural region to 572 (reduction of 67%) and from 607 in Condroz agricultural region to 194 (reduction of 68%). The number decreases to 289 and 105 respectively for Limoneuse and Condroz agricultural regions if the purity threshold level is set to 95%. A question is to know if to possibly losing some of the natural variability by a more severe selection on purity level is more interesting in terms of assimilation efficiency than possibly having a better representation of growing conditions but introducing some noise in the time series. As we work in this study on regions that are defined on the basis of growing conditions criteria such as meteorological and soil conditions (in other words, growing conditions are relatively similar within a given agricultural region), it is not unrealistic to believe that the natural variability within a given agricultural region is relatively low and therefore that selecting the purest winter wheat MODIS grid cells could improve assimilation efficiency.

The temporal shift between the simulated and remotely-sensed GAI peaks, even when pure pixels are considered, is questioning. It can arise from a wrong estimation of model parameter(s) influencing the simulated date of maximum GAI value. In WOFOST, for winter crops, this date is mainly driven by TSUM1 parameter, i.e. the sum of effective temperatures from the 1st January up to heading. TSUM1 is normally a variety (genotype) parameter. In the version of WOFOST used in this study, TSUM1 has been recently recalibrated after a spatial clustering in two zones of equal

"variety" (Djaby *et al*, 2009). Each variety is defined by a single set of crop parameters which are kept constant within a zone. Concept of variety as considered in Djaby *et al*, 2009 is purely a model concept completely different from varieties (genotypes) in plant breeding. The zones were defined according climatic gradients as well as topographic features. Despite this recalibration, an average RMSE of 7.6 days was observed between simulated and observed day of anthesis. This remaining error is mainly due to the non-consideration of genotypes in simulation. Considering TSUM1 parameter instead of TDWI could be an alternative to allow a phenological adjustment but unfortunately it couldn't probably be able to adjust correctly the GAI values (as TDWI can do). TSUM1 and possibly TSUM2 could be computed for each MODIS grid based on GAI observations or more exactly on CSDM smoothed time series. These TSUM1/TSUM2 values could be then simply forced in the model prior to a recalibration on TDWI and SPAN. Let's note that considering the dependency observed between the recalibrated parameters, an other possibility could be to adjust first TDWI on the increasing phase of GAI (still considering a prior forcing with TSUM1) before any recalibration of SPAN parameter. Whatever the proposed method, it implies however to have full confidence in the GAI observations as representation of the field (at least on the phenological point of view) and in the ability of the CSDM to smooth observed GAI time series. Although some crop growth stage observations collected in field trials tend to support the hypothesis that phenology derived from remote sensing match with field reality it has to be validated on a larger data set. Differences in sowing dates can also explain partly the differences between simulations and observations. In WOFOST for winter crops, simulation starts at pseudo emergence date fixed at the 1st January and it is therefore unfortunately not possible to directly recalibrate the sowing (or emergence) date to match the observations.

More generally, the differences between the simulated and remotely-sensed dates of maximum GAI stresses once again the recurrent problem of the usually low size of observations data sets used to calibrate models.

Despite that and the fact that the results of this analysis has to be confirmed for other years and regions, this study has however the merit to lay the foundations for further improvements of assimilation techniques of remote sensing data in crop growth models.

CHAPTER 3 : Potential performances of SAR-Estimated grassland mowing calendar assimilation in LINGRA model based on an OSS Experiment

1. Introduction

Monitoring grassland resources at regional level is an integral part of a forage production system. Grassland resources should ideally be analyzed concomitantly with other aspects such as land or livestock resources. The monitoring is usually done to quantify available resources, understand the interrelationships between the different components and appraise development options or estimate livestock support capacity (Harris, 2000).

Monitoring grassland resources through field observations is difficult, expensive and time consuming. The modeling approach can be in this context a solution to face these difficulties.

Various models have been developed to simulate grass growth at regional and local level (Sheehy and Johnson, 1988; Thornley, 1991; Schapendonk *et al.*, 1998; Stilmant *et al.*, 2001; Woodward, 2001; Barret *et al.*, 2005; Ruget *et al.*, 2006). These models generally use meteorological inputs considering that the prevailing climatic conditions, notably air temperature, light and rainfall are the main determining factors for grass growth (Barret *et al.*, 2005) in addition to soil water supply. Besides the climatic conditions, the level of N fertilizer applied is also important and, given that other factors being equal, is often the main limiting factor for herbage production (Whitehead, 1995). In the grassland models, nitrogen fertilization can be either specified (Stilmant *et al.*, 2001; Ruget *et al.*, 2006) or considered as optimal (Bouman *et al*., 1996a).

As observed in chapters 1 and 2, synchronizing modeled phenology is an important factor to consider with a view to reach a good accuracy in yields estimation and *a fortiori* in yields predictions. In grasslands, phenology is partly driven by farmers' management, mainly mowings. In Belgium up to 3 (and sometimes 4) mowings can be observed within a growing season. These mowings are spread out all along the growing season according climatic conditions, farmers' needs and objectives (Dusseux *et al.*, 2013). A huge uncertainty on mowing dates used in grassland growth models is therefore observed.

As aforementioned, uncertainty in biophysical models can be reduced by assimilation of remotely-sensed data. Mowings are punctual events. In order to monitor these events, multi-temporal image series with a high temporal frequency are required. This requirement can be hardly met in Belgium with optical sensors considering the frequent cloud cover. On the contrary, SAR systems, due to their ability to penetrate

cloud and haze layers (Attema *et al*., 1998) and their independence to sun illumination, offer a high data acquisition frequency even in bad weather conditions which represent a serious asset.

With a variable degree of success and at different spatial resolutions, SAR data were used to monitor and estimate grassland characteristics such as grass height (Hill *et al.*, 1999), biomass (Svoray and Shoshany, 2002; Moreau and Le Toan, 2003, Wang *et al*., 2013), leaf area index (Dente *et al.*, 2008).The methodologies used to link these biophysical variables with SAR data include "traditional" approaches such as those based on statistical relationships, e.g. simple (Stolz and Mauser, 1997; Svoray and Shoshany, 2002; Moreau and Le Toan, 2003) or multiple regressions (Hill *et al*., 1999), or based on the inversion of a semi-empirical model such as the water cloud model (Svoray and Shoshany, 2002 ; Graham and Harris, 2003; Dabrowska-Zielinska *et al*., 2007) or the Oh model (Oh *et al*., 1992). More complicated models have been also developed. An extensive review of these models can be found in the Kornelsen and Coulibaly (2013).

Grassland management can also be followed through SAR data. Herold *et al.* (2000) have for example shown that a discrimination of the management mode (mowing or grazing) with SAR images and mainly with the L band (the other bands tested were the C and X bands) was possible. Dusseux *et al.* (2014) have showed that SAR-derived (RADARSAT-2) variables (polarization ratio and polarimetric decomposition) allow a very good discrimination of grasslands from crops in Central-north Brittany (classification accuracy of 0.98) but unfortunately that the HH/VV ratio, commonly used for land monitoring in agricultural areas, was characterized by high variance for each date which do not allow to discriminate grassland management practices. These findings are in accordance with those from Voormansik et al. (2013) who have also noted with TERRASAR-X that it was not possible to distinguish tall grass from short grass on the basis of HH/VV ratio. The same authors have however assessed that it was however easy to detect areas with freshly cut grass lying horizontally on ground on the basis of dual polarimetric dominant scattering alpha angle. Let's note that a current trend, exacerbated by the forthcoming sentinel missions, relies in the combination of SAR with optical data for grassland management (Finnigan, 2013; Wang et al., 2013; Dusseux *et al*., 2014; Hong *et al*., 2014).

Backscattering measurement is also tainted by errors. Errors arise notably as a result of instrument noise, calibration errors, aliasing or statistical uncertainties due to speckle. Even accurately measured and stable in time SAR data display significant speckle (Bally and Fellah, 1995). Each SAR image pixel includes a high number of individual scatterers. When a coherent electromagnetic radiation interacts with a rough surface, the scatters generate return signals with random phases, which interact with each other. Because of constructive and destructive interferences in the radar returns from independent point scatterers within a resolution cell, the intensity in the radar image of homogeneous target is not constant. The interferences are the cause of the grainy appearance of the generated image and lead to random changes in the pixel's brightness (Tupin *et al*., 2014). As a consequence, the image of a homogeneous surface shows pixels values with a high dispersion (González

Sanpedro, 2008; Giustarini *et al*., 2015). Speckle limits the ability to correctly interpret SAR images, restricts image classification and introduces uncertainty in ground surface parametric inversion (Huang and Liu, 2007).

Though several techniques (mainly filters) were developed to reduce speckle noise in SAR imagery (Durand *et al*., 1987; Huang and van Genderen, 1996; Qiu *et al*., 2004; Huang and Liu, 2007; de Leeuw and Tavares de Carvalho, 2009; De Keyser *et al*., 2012; Buemi *et al*., 2014; Torres *et al*., 2014), the speckle noise cannot be fully eliminated. The efficiency of these filtering techniques is limited by the difficulty of modeling SAR speckle as several field and sensors characteristics (e.g. field size and geometry, moisture, dielectric constant, wavelength, polarization or view angle) can generate different speckle noises.

The impact of the SAR backscattering uncertainty on assimilation should be therefore assessed. In order to reach this objective a synthetic analysis, similar to the one used in chapter one, has been performed. More specifically, this chapter presents the results of an OSS Experiment assessing the efficiency of the assimilation of mowing dates derived from SAR data in the LINGRA grass growth model (Bouman *et al*., 1996b; Schapendonk *et al*., 1998) and aiming to identify the accuracy and the temporal availability required on SAR observations to reach a given objective in terms of errors reduction on final total biomass estimation.

2. Material and Methods

2.1. Grassland growth model

We used the LINGRA (LINtul GRAss) grassland growth model as a basis for our work (Bouman *et al*., 1996b Schapendonk *et al*., 1998).

LINGRA, derived from the LINTUL (Light INTerception and UtiLisation simulator) concepts as proposed by Spitters (*In* Bouman *et al*., 1996b), is a sink/source growth model initially developed to predict the productivity of *Lolium perenne* grasslands across the member states of the EU at the level of potential and water-limited productions. The model has been also used to quantify land-use evaluation and study the effects of climate change on grass growth (Rodriguez *et al*., 1999), used as basis for the development of new models (Barrett *et al*., 2005) and adapted for Timothy (*Phleum pratense*) swards (Höglind *et al*., 2001).

Simulated key processes are light utilization, leaf formation, leaf elongation, tillering, and carbon partitioning to storage, shoots and roots. The sward is assumed to remain in a vegetative stage. The integration level is kept high and the number of processes has been restricted to key parameters, and only a small number of processes involving these parameters are dynamically simulated. On the other hand, parameters that have relatively little impact on crop growth, or which knowledge is scarce, have been treated using a static approach.

Two levels of production, potential (depending only on intercepted solar radiation and temperature) and water-limited, are simulated. Soil nutrients are considered as at an optimal level and there is no simulation of mineral and manure nutrition. In contrast to

arable crops, the grassland plants are frequently defoliated due to grazing or mowing. LINGRA is mainly suitable for simulating mowings. In the model, the mowing dates are optional, either read from an external file or when the grass reaches a certain amount of biomass, fixed by the user (Schapendonk *et al*., 1998).

2.2. Area of Interest

Although the OSS Experiment is mainly based on synthetic data and that subsequently no field data are really compulsory, it is still useful to have a first guess of the uncertainty on model parameters and/or initial states in order to avoid the use of “non-realistic synthetic data” which could lead to erroneous conclusions.

Simulations were performed for year 2008 and for a 50 x 50 km grid located in Wallonia where the proportion of grasslands area is important (>25 % of the total area).

In our experiment, data collected in the vicinity of Libramont municipality (49° 55′ N, 5° 22′ E) within the frame of the on-going MIMOSA project (Buffet *et al*., 2010) was used to ground the experiment to realistic conditions and later on to possibly validate some of the conclusions of the OSS Experiment.

Data, collected during the first mowing period in 2008 and during the first and second mowing period in 2010, includes grass height and biomass as well as management surveys (mowing and grazing dates).

The 2-year survey has notably allowed stressing that even in a limited area grassland management practices are spread out during all the vegetation period. As an example in 2008 for the first mowing period, roughly 6 weeks separate the first (last dekad of May) and last (first dekad of July) mowings. This large range of mowing dates can be explained by weather conditions (temporal distribution of rainfall events), farmers objectives in terms of grassland quantity and quality but also by some regulations (as e.g. agri-environmental measures) granting subsidies for late mowed parcels.

2.3. Description of the OSS Experiment

As previously explained in chapter 1, OSS Experiments consist of identical twin experiments, a widely used technique to assess the impact of data assimilation, in which data selected from one model run is assimilated into another run of the same model with different initial conditions. The convergence of the second run towards the first, defined as the “truth”, can be measured to quantify the data assimilation effectiveness in driving the model with “wrong” initial conditions towards the truth (Raicich and Rampazzo, 2003).

Figure 3.1.presents an overview of the OSS Experiment.

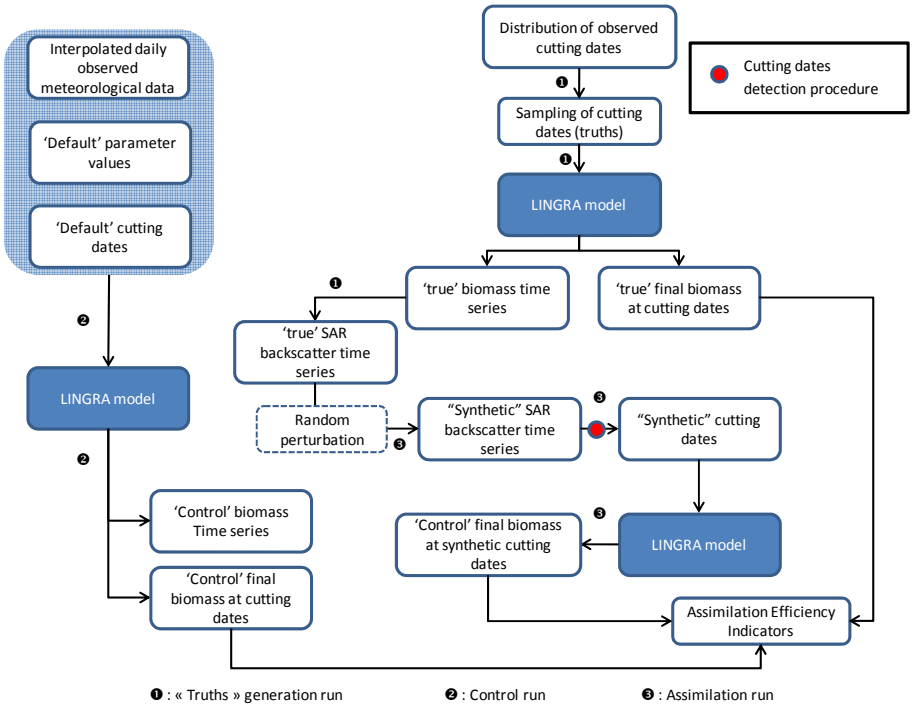


Figure 3.1.- Overview of the OSS Experiment

The OSS Experiment is divided in 3 main steps.

First step: "truths" generation run (1)

In the truths generation run, a set of "synthetic observations" used afterwards in the assimilation step is computed. To reach this objective, a so-called "true" set of mowing calendars is defined on the basis of the default mowing calendar considered in the LINGRA model.

"Default" mowing calendar is the one used operationally and can be considered as the best estimation of this calendar for a given area without any additional observations and therefore without assimilation. In our study, a 3 mowing per year scheme has been considered. Default mowing dates have been set to DOY 166 (15th of June), 215 (3th of August) and 264 (21th of September). The default mowing calendar is in accordance with observations made in the area of interest.

Considering that mowing operations are mainly dependent on meteorological conditions and farmers management (dairy farmers tend for example to mow earlier to favour grass quality), the true mowing calendars are not necessarily the same as the ones used "by-default" in the model.

True mowing calendars have been defined by considering 10 days before and after the default mowing dates in addition to the default mowing dates themselves. 3 mowing dates per mowing period (e.g. DOY 156, 166 and 176 for the first mowing

period) were therefore considered and consequently 27 (3^3) mowing calendars (27 "truths") have been defined.

The generated set of "true" mowing calendar is then used to run the model in order to generate "true" biomass time series. True biomass time series have been afterwards converted in backscattering time series by the way of a statistical relationship (Equation 1) presented in Svoray and Shoshany (2002):

$$\sigma^0 = -1.88 * \ln(AB) - 16.24 \qquad (Eq. 1)$$

With σ^0 the SAR backscattering (in dB) and AB the aboveground biomass (in kg/m² of dry matter).

This empirical relationship, simple and straightforward, established using ERS-2 SAR data, presents a negative shape which seems to be coherent with other studies (Taconet *et al.*, 1994, Stolz and Mauser, 1997).

A stochastic noise term is finally added to the "true" backscattering values in order to account for the uncertainty on SAR backscattering (resulting from instrument noise, speckle…) generating by this way "synthetic" backscattering values.

$$BS_{RS,i} = BS_{mod,i} + \varepsilon_i$$

$$with\ \varepsilon_i \sim N(0, \sigma)$$

With $BS_{RS,i}$ the synthetic backscattering value at time *i*, $BS_{mod,i}$ the true backscattering value at time *i* and ε_i the added stochastic noise at time *i*.

The stochastic noise added to the true backscattering values is assumed to follow a normal distribution. Different levels for the standard deviation σ associated to the distribution were considered (0.5, 1 and 1.5 dB – see paragraph 2.4. "Definition of OSSE scenarios")

An example of true and synthetic time series is presented in figure 3.2.

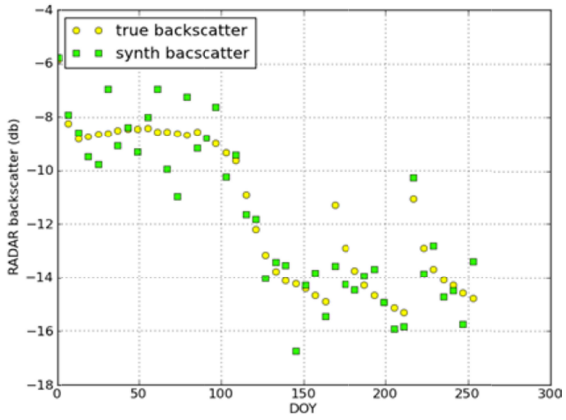


Figure 3.2.- Example of true (yellow dots) and synthetic (green squares) radar backscattering values (observations available every 6 days with an uncertainty of 1 dB on backscattering values)

Second step: control run (2)

The control run (often called “open loop” run) represents the best model estimate of the truth without the benefit of data assimilation. LINGRA model is run with its aforementioned “default” mowing calendar. The ‘control’ yield values generated during this run are used to evaluate the improvement or alteration of the simulations due to the assimilation of the “synthetic” mowing calendars.

Third step: assimilation run (3)

In the last step of the OSS Experiment, a procedure is used to estimate “synthetic” mowing calendars on the basis of the “synthetic” backscattering time series defined in the first step.

Mowing dates will be estimated for each synthetic backscattering time series based on the following procedure (also explained in figure 3.3):

1. For each mowing period, a time frame within which we search for the mowing date is defined. The time frame goes from 20 days before the default mowing date for the considered mowing period up to 20 days after this default mowing date.
2. Within the defined time frame, the mowing window (time interval between two observations where the mowing is supposed to occur) corresponds to the interval with the biggest difference between two successive observations as far as this difference is positive (considering that backscattering value when the biomass is high, i.e. before the mowing, is lower than after the mowing).
3. To be considered as valid mowing window, the two backscattering values of the interval should be on either side of a threshold value arbitrarily set in our study to 0.2 kg/m² converted in backscattering value with equation 1.
4. Considered mowing dates are the average of the two dates at both ends of the mowing windows.

If no mowing date is identified for a given mowing period, the mowing date used to run LINGRA is the default mowing for the considered mowing period.

Figure 3.3 illustrates the mowing date detection procedure (for one mowing period).

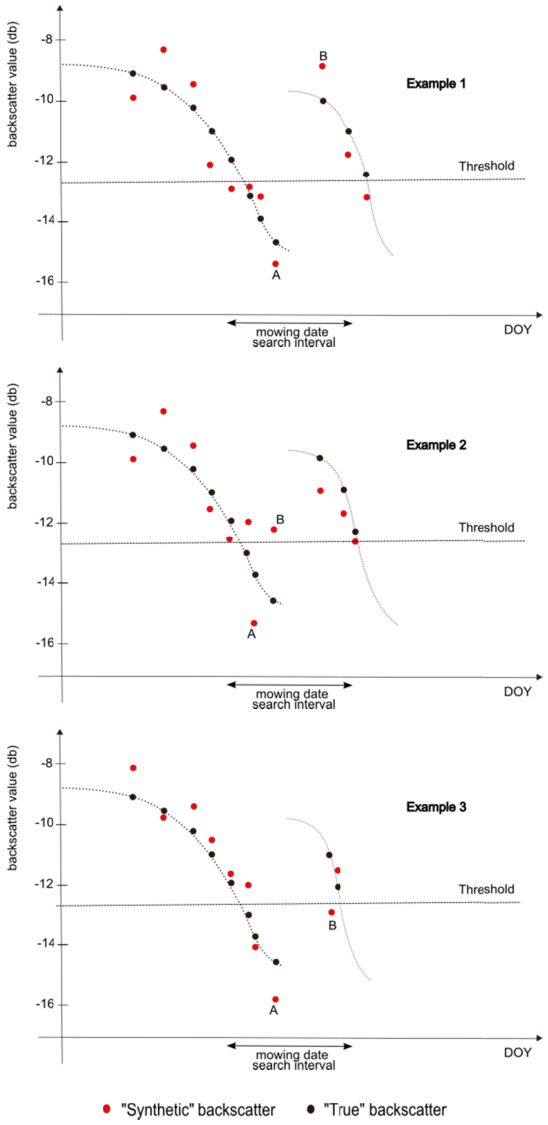


Figure 3.3.- Possible cases mowing dates detection procedure

In the different examples, the biggest differences between two successive backscattering values are observed between observations 'A' and 'B'. In example 1, these two observations are on both side of the defined backscattering threshold value and surround the true mowing. In example 2, the two observations are also on both side of the backscattering threshold value but in this case the mowing window is not well identified. In the last example, the two observations presenting the biggest difference of backscattering values are on the same side of the threshold. In this case, as doubts exist, no mowing date is estimated and the default mowing date is used instead.

LINGRA is run considering these synthetic mowing calendars and the resulting yields values are compared with the yields generated in the “truths” generation run which allow assessing the performance of the considered assimilation method.

2.4. Definition of OSSE scenarios

As the study aims at identifying the accuracy and the temporal availability needed on SAR observations to reach a given objective in terms of errors reduction on final grain yields estimation, different scenarios of uncertainty on SAR observations and time step between SAR observations have been defined.

As mentioned in paragraph 2.4.2., three different standard deviations σ associated to the stochastic noise added to the true backscattering values have been considered: 0.5, 1 and 1.5 db. To provide an order of magnitude, an uncertainty of 1.5 dB is roughly what can be expected with ERS SAR .PRI and .GEC products. Let’s note that the lower uncertainty levels considered in this analysis can for example be reached with these products by pixel averaging. Magnitude of the accuracy improvement in this case is function of the parcel size concerned by the averaging technique (Bally and Fellah, 1995).

Two regular time steps between observations were considered: 6 and 12 days. These time steps correspond to the expected time steps with the forthcoming ESA Sentinel-1 SAR mission, 6 days being for the two-satellite constellation (Snoeij *et al.*, 2008).

2.5. Assessment of assimilation efficiency

As for the winter wheat OSS Experiment (chapter 1), the estimation of the accuracy and temporal availability needed on SAR observations is mainly estimated on the basis of an indicator computed from Relative Mean Absolute Error (RMAE) values estimated for the situation with and without assimilation, namely the assimilation efficiency (AE) indicator.

As aforementioned, the number of truths considered is equal to 27. In order to have more stable and reliable results, random perturbations to generate the “synthetic” backscattering values have been repeated several times. For the sake of computation time, the number of iterations considered was fixed to 40.

As the day at which we start to consider the regular time step between observations can influence the results of the mowing date estimation, six different start dates (repetitions) was considered. When the regular time step is equal to 6, start has been set from DOY 1 to 6. When the regular time step is equal to 12, starts have been randomly selected within 2-days successive interval between DOY 1 and 12 (one start randomly selected between DOY 1 and 2, one start between DOY 3 and 4...etc.).

The Relative Mean Absolute Errors are computed as follows:

$$RMAE_{TO} = \frac{1}{N1}\sum_{i=1}^{27}\frac{|TBT_i - TBO|}{|TBT_i|} \qquad (Eq.2)$$

$$RMAE_{TA} = \frac{1}{N2}\sum_{i=1}^{27}\sum_{j=1}^{40}\sum_{k=1}^{6}\frac{\left|TBT_i - TBA_{ijk}\right|}{|TBT_i|} \qquad (Eq.3)$$

Where TBT_i is the total biomass yield of truth *i*, *TBO* the total biomass yield from the "control" run, i.e. the reference total biomass without assimilation and TBA_{ijk} the total biomass yield after assimilation for truth *i* (i=1,...,27), iteration *j* (j=1,..., 40) and repetition *k* (k=1,..., 6). Differences and ratios have all been considered in absolute values. N1 and N2 represents the total number of observations.

$RMAE_{TO}$ and $RMAE_{TA}$ have been computed for each mowing period but also for the whole mowing season (i.e. considering the 3 mowing periods as a whole). Total biomasses considered in equations 2 and 3 are therefore either the biomass mowed at the considered mowing period or the sum of the biomass for the 3 considered mowings.

The Assimilation efficiency AE indicator, derived from these Relative Mean Absolute errors, can be computed as follows:

$$AE = 100 * \left(1 - \frac{RMAE_{TA}}{RMAE_{TO}}\right) \qquad (Eq.4)$$

Assimilation Efficiency (AE) indicator represents an estimation of the average reduction of errors on total biomass yields estimation. A positive value for AE means therefore that the assimilation of the estimated mowing dates globally improves the total biomass yields estimation. At the opposite, a negative value for AE means that the errors on total biomass yields estimation are on average increased after assimilation and that in this situation, it is better to not assimilate and to use the model parameters and initial states 'default' values.

2.6. Assessment of the accuracy and temporal availability needed on SAR observations

The main objective of the OSS Experiment aims at the assessment of the accuracy needed on remotely-sensed SAR observations in order to reach a given objective in terms of errors reduction on final grass yields estimation. In this frame, two objectives of 25% and 50% of errors reduction have been defined.

Each scenario presenting an Assimilation Efficiency value higher than the defined objectives (25 or 50%) is deemed as suitable.

3. Results

Table 3.1 presents the results for the mowing dates estimation. Globally, the estimation procedure allows the identification of a mowing date in more than three-quarter of the situations. The percentage of identification is logically higher when the

uncertainty level on backscattering values is low and/or the time step between satellite observations is short.

The average differences between the estimated and the true mowing dates are systematically negative. This bias increases with the level of uncertainty on backscattering and/or the time step between observations.

The standard deviation follows logically the same trend towards an improvement according to the level of uncertainty on backscattering values and the time step between observations. The estimation of mowing date based on SAR data allows to divide the standard deviation up to more than by 4 compared to the situation without assimilation i.e. when simply using the default values: from 8.2 days when the default mowing dates are used, the standard deviation falls up to 1.7-1.8 days when observations are available every 6 days and uncertainty level on backscattering values is equal to 0.5 db. It seems also more interesting to improve accuracy on observations than the time step between these observations.

Computation of AE confirms the significant potential offered by the detection of mowing dates (table 3. 2).

Positive Assimilation Efficiency values are observed nearly for all our considered scenarios meaning that the detection of mowing dates based on SAR data and the use of the information in LINGRA growth model always improve the estimation of total biomass produced over the 3 mowing periods. Assimilation efficiency values are slightly higher when the mowing periods are considered separately. Though the improvements are sometimes rather limited, especially when a time step of 12 days is observed between SAR observations, a reduction around 75 % of the errors on total biomass can be overall expected when SAR observations are available every 6 days with an uncertainty of 0.5 db.

Relative Mean absolute Error without assimilation ($RMAE_{TO}$) is low (3.4%) when the sum of mowed biomass is considered. This low value is linked to the specification (value, distribution) of the uncertainty applied on mowing dates but also is due to the fact that our experiment has been performed only for one year; consequently the inter-annual variability is not taken into consideration. $RMAE_{TO}$ is higher when the mowing periods are analysed separately and increases logically from the 1st to the 3rd mowing period. From 6.0% for the first mowing period, it reaches 15.6% at the 3rd mowing period.

Roughly, the first objective aiming to an errors reduction on total biomass of 50% can be only reached when time step between SAR observations is equal to 6 days and the level of uncertainty is equal or lower than 1 db.

The second and less demanding objective aiming to an errors reduction on total biomass of 25% can be logically reached in more situations. This objective is obtained when the time step between observations is equal to 6, whatever the level of uncertainty. When the time step is doubled, the objective is reached when level of uncertainty is equal to 0.5 dB and hardly reached when the level of uncertainty is equal to 1db.

Reduction of the assimilation efficiency is relatively limited, for a given time step, when the uncertainty level rises from 0.5 to 1 dB but seems to be more important when the uncertainty level reaches 1.5 db. Reduction of the assimilation efficiency on the other hand tends to be more important when, for a given uncertainty level, the time step goes from 6 to 12 days. These results tend to show that for improving yield estimations reducing the time between two observations is comparatively more efficient than reducing the uncertainty level on these observations.

Table 3.1.- Percentage of identified mowing dates, mean and standard deviation of the difference between the estimated and the true mowing dates per mowing period (with / without replacement by default value)

		% of identified mowing dates			*Mean difference between estimated and true mowing dates*			*Std. of the difference between estimated and true mowing dates*		
Timestep	***σ***	**1st mowing period**	**2nd mowing period**	**3rd mowing period**	**1st mowing period**	**2nd mowing period**	**3rd mowing period**	**1st mowing period**	**2nd mowing period**	**3rd mowing period**
6	*0.5*	99.9	99.7	100.0	-0.5 / -0.5	-0.5 / -0.5	-0.5 / -0.5	1.7 / 1.7	1.8 /1.7	1.7 / 1.7
6	*1*	95.1	93.4	98.5	-1.0 / -0.9	-0.6 / -0.6	-0.6 / -0.6	4.0 / 3.7	3.7 / 3.2	3.4 / 3.3
6	*1.5*	90.0	88.0	95.6	-1.9 / -1.8	-1.0 / -1.0	-0.7 / -0.7	6.7 / 6.7	6.4 / 6.1	6.3 / 6.3
12	*0.5*	88.9	86.3	98.0	-1.4 / -0.9	-0.9 / -1.1	-0.9 / -0.8	4.0 / 3.3	4.3 / 3.2	3.6 / 3.4
12	*1*	83.6	79.8	95.1	-2.6 / -2.4	-1.2 / -1.4	-1.1 / -1.0	5.7 / 5.5	5.3 / 4.2	4.8 / 4.5
12	*1.5*	79.3	76.0	93.0	-3.5 / -3.7	-1.8 / -2.4	-1.8 / -1.9	8.1 / 8.2	7.1 / 6.6	6.8 / 6.7
With default values		-	-	-	0.0	0.0	0.0	8.2	8.2	8.2

Table 3.2.- $RMAE_{TA}$, $RMAE_{TO}$ and AE values per mowing period considered separately and all together (sum of mowed biomass)

Time step	*σ*	1st mowing period			2nd mowing period			3rd mowing period			All mowing periods		
		$RMAE_{TA}$	$RMAE_{To}$	AE(%)	$RMAE_{TA}$	$RMAE_{To}$	AE(%)	$RMAE_{TA}$	$RMAE_{To}$	AE(%)	$RMAE_{TA}$	$RMAE_{To}$	AE(%)
6	*0.5*	1.5	6.0	75	2.7	10.8	75	3.7	15.6	76	1.2	3.4	67
6	*1*	2.4	6.0	60	4.4	10.8	60	5.4	15.6	66	1.6	3.4	54
6	*1.5*	4.4	6.0	31	7.5	10.8	31	9.4	15.6	40	2.6	3.4	25
12	*0.5*	3.4	6.0	48	5.6	10.8	48	7.2	15.6	54	2.2	3.4	36
12	*1*	4.9	6.0	18	6.6	10.8	39	8.3	15.6	47	2.6	3.4	24
12	*1.5*	7.1	6.0	-18	8.9	10.8	18	10.7	15.6	31	3.3	3.4	4

4. Discussions and conclusions

Our Observing System Simulation Experiment (OSSE) has allowed highlighting and quantifying the errors reduction to be expected on total biomass by the detection of the mowing dates based on SAR observations and the assimilation of these mowing dates in LINGRA model.

Reduction of errors in estimating total biomass can reach up to 67% in optimal conditions i.e. when SAR observations are available every 6 days with an uncertainty level equal to 0.5 dB which is roughly what can be expected from the forthcoming Sentinel-1 mission with its two satellites.

It is important to note that the scenarios defined in our experiment are ideal situations. The regular time step between SAR observations is probably an optimist vision of the reality as well as the constant level of uncertainty all along the growing season. Many other hypotheses have been set in this experiment as the normal distribution considered for the errors on SAR observations and the constant uncertainty over the growing season, the perfection of the LINGRA model, an optimal fertilization or the fact that the remotely-sensed observations represent a non-biased measurement of field biomass. The simulated improvements are probably the best to be expected but even divided by two the improvements remain significant.

Results are also function of the methodology used to detect the mowing dates. Our results tend to demonstrate that our detection procedure provides satisfactory results despite a systematic negative bias. The reason explaining that the average differences between the estimated and the true mowing dates are systematically negative has probably different origins. In case of mowing period incorrectly identified, the probability to identify a mowing period earlier to the right mowing period seems to be higher.

For the first mowing period, where the biases are the most important, a part of the explanation lies also in a higher proportion of non-identification for true mowing dates occurring 10 days after the default mowing dates (i.e. mowing dates occurring at DOY 176). A later mowing at this period means slightly higher daily temperatures and subsequently a faster re-growth. The probability to have the two observations defining identified mowing window on the same side of our defined threshold is higher (and the subsequently the probability to use the default value is higher).This procedure is probably only suitable for mowed parcels. Biomass variations along the growing season in grazed parcels are indeed certainly too small to be detected but the procedure could be possibly used for land use mapping.

The methodology also starts from the assumption that mowing corresponds to an increase of backscattering. This assumption is probably only valid for a specific range of conditions. Besides the vegetation biomass, the backscattering is indeed influenced by several factors such as, among others, soil moisture content and roughness, incidence angle, wavelength or canopy structure. The impact of these different factors will be studied more closely in chapter 4.

All these results need of course to be confirmed by empirical study but our experiment tend however to show that, despite the aforementioned hypotheses and approximations, the detection of cutting dates on the basis of SAR observations and the subsequent use of this information in LINGRA model allows improvements of total biomass estimation. Our results can also provide objectives for future missions dedicated to crop growth monitoring.

Whatever the scenario of temporal availability and of accuracy on SAR observations, an improvement is observed. If an objective of 25% errors reduction on total biomass can be reached in nearly all situations (the only situation where this 25% objective is not reached is when SAR observations present an uncertainty of 1.5 dB and are available every 12 days), the second objective aiming at an errors reduction of 50% is only observed when SAR observations are available on a 6 days basis and with an uncertainty lower or equal to 1 db.

A better estimation of grassland production, through a better estimation of the mowing calendar, is important for grassland management both at farm and regional level but also in the frame of the use of this information in decision support systems. A better estimation of biomass means also a better estimation of the grass quality as an inverse relationship links grassland productivity and grassland energetic value.

An accurate estimation of mowing dates could also be used for regional monitoring and control purposes. For example in Wallonia in the frame of agri-environmental measures, subsidies are granted to farmers accepting to delay their mowing operations (a late mowing favouring the fauna and flora biodiversity). Grassland parcels concerned by the agri-environmental measure can't be mowed before a certain date (the 20th of June or the 1st of July according to the region).This study shows also that SAR observations, as far as the temporal resolution is good enough, could be therefore used to check if the conditions for granting the subsidies are met or at least to orientate the controls of subsidised grassland parcels. Assessing mowings calendar is also an opportunity to assess land-use intensity in grasslands which has a significant impact on biodiversity and on landscape habitat. Intensive mowings, involving frequent destruction of above-ground plant organs or a generally low ability to resprout, result in a rapid decline of sensitive species and enrichment of more disturbance-tolerant ruderal taxa.

CHAPTER 4 : Estimation of mowing dates on the basis of SAR (ERS-2) observations

1. Introduction

The Observing System Simulation Experiment implemented in chapter 3 has stressed the importance of correctly estimating the mowing dates for accurate yields estimation. The method used in chapter 3, most probably only valid for a specific range of conditions, starts from the assumption that mowing corresponds to an increase of backscattering. Other factors can in addition influence the backscattering and subsequently the mowing detection. Therefore, before considering the assimilation in model like Lingra, the priority is to assess whether a mowing date can actually be detected using SAR time series.

The total backscattering is a complex sum of the backscattering from vegetation and soil. As the microwave frequency can penetrate both canopy and soil up to a difficult-to-determine depth, assessing if the signal is dominated by the soils or crop conditions is difficult (Moran *et al.*, 2012). Penetration capabilities of a SAR signal within a canopy is function of the sensor configuration including frequency, incidence angle and polarization but is also dependent on water content, vegetation and soil characteristics (Gherboudj *et al*., 2011).

Shorter wavelength such as X-band and C-band interacts mainly with the top canopy layers while longer wavelength such as L-band penetrates deeper within the canopy leading to higher scattering contribution from the soil (Ulaby *et al*., 1984 *In* Jiao *et al*., 2010; Henderson and Lewis, 1998). C-band is usually considered as rather suitable frequency for agricultural purposes as it has a wavelength of about 5 cm which is comparable to the size of the leaves and the stems of cereal crops (Wooding *et al.*, 1995; Skriver *et al*., 1999).The sensitivity of C-band SAR to crop conditions is however dependent upon the crop type. Ferrazzoli *et al.* (1997), correlating crop biomass with backscattering at C-HV at an incidence angle of 35° for various crops, have observed a strong relationship (r^2=0.75) for crops such as wheat, colza and alfalfa. A lower relationship (r^2=0.31) was unfortunately observed for crops such as corn, sunflower and sorghum due to a signal saturation. For Ferrazolli et al. (1992) this saturation is observed at a LAI of 2-3 m^2/m^2 for C-VV and C-HH backscattering. The same saturation value is reported for VV/VH ratio by Blaes *et al.* (2006). According to McNairn and Brisco (2004), this saturation is reached in corn when the plants reach approximately 1 m height.

The vegetation contribution to total backscattering is comparatively more important when the incidence angle increases. Indeed, the pathlength within the vegetation increases and therefore enhances the response to crop conditions. Steep incident angles (<30°) are more interesting in a context of soil moisture estimation due to the decreased effects of soil roughness and vegetation attenuation (Dabrowska-Zielinska *et al.*, 2007 ; Balenzano *et al*., 2011; Moran *et al.*, 2012; Gherboudj *et al*., 2011).

The penetration of SAR signal may be finally also dependent on polarization if the orientation of crops coincides with the orientation of polarization. A stronger backscattering from vertically orientated crops (such as wheat, rice or grass) is for example usually observed with VV polarization.

In the same way, directions of planting row can have an impact on total SAR backscattering. Picoli *et al.* (2013) have for example observed in Sugarcane that backscattering values from fields with rows perpendicular to the range direction of satellite were higher (~ 1.2 dB) than those from fields with rows horizontal to this direction. According to Moran *et al.*, 2012, this impact can induced differences of 5-10 dB up to 20 dB in extremes situation. Wind can change the orientation of the crop and increase or decrease by this way significantly the backscattering. A variation between 2.5 and 7dB has been reported (Voormansik *et al*., 2013).

Many studies are currently exploiting the potentialities offered by fully polarimetric sensor (McNairn and Brisco, 2004). Good relationships can be observed for agricultural crops between biomass or LAI and polarizations differences (Brown *et al.*, 2003) or ratio (De Roo *et al*., 2001; Mattia *et al*., 2003, Della Vecchia *et al.*, 2008).

The vegetation canopy typically results in volume scattering and the attenuation of radar signal due to vegetation varies according to its dielectric properties (i.e. its water content) and its physical structure. When biomass increases, the backscattering becomes decreasingly sensitive to soil properties (Dabrowska-Zielinska *et al.*, 2007). Skriver *at al.* (1999) have observed that at the end of the growing season the C-band backscattering from crop vegetation was dominated by volume scattering. Luckman (1998) and Hill *et al.* (2005) have also shown, based on a polarimetric model-based decomposition (Freeman and Durden, 1998), that volume scattering was dominant in grassland at C-band (the secondary scattering being the surface scattering). For Dabrowska-Zielinska *et al.* (2007), the volume scattering of vegetation becomes the dominant scattering type when the LAI is higher than 3 for L-band. Joseph *et al.* (2010) have however assessed for corn that even at peak of biomass and large incidence angles the backscattering remains sensitive to soil moisture.

Many studies (Allen and Ulaby, 1984; Sofko *et al*., 1989; wood *et al*., 2002; Riedel *et al.*, 2002; Riedel and Schmullius, 2003; Saich and Borgeaud, 2000) have stressed the increase of the backscattering coefficient due to the presence of water on crops (after a rainfall or dew formation). This increase is significant: Sofko *et al.* (1989) have for example observed a 2-4 dB increase of backscattered signal of canopy just after a rainfall and Wood *et al.* (2002) an increase of 1.7-2.5 dB for crops on which dew was formed. Concerning grasslands, Stolz and Mauser (1997) show that for a same grass height, backscattering coefficients corresponding to data acquired in dry conditions tend to be lower than the ones acquired in rainy conditions but also that the higher dielectric constant due to the droplets on the plants increases the signal significantly only for higher canopy. According to Riedel *et al.* (2002), intercepted rainfall and dew influence on backscattering is different especially at C-band. For

these authors, the backscattering variations at C-band after a rain event strongly depend on the vegetation structure and the growth stage while the impact due to dew is independent from vegetation type.

As far as the contribution of the soil in the total backscattering is concerned, it is mainly influenced by soil moisture content and surface roughness (Benallegue *et al*., 1995; Zribi and Dechambre, 2002; Löw *et al*., 2005). The dielectric constant (water content) of the soil is the main factor influencing the intensity of the backscattering (Álvarez-Mozos *et al*., 2005). If other factors are ignored, the backscattering coefficient increases with an increase of soil moisture until the moisture content reaches roughly 35% in volume. At this moment, the radar signal becomes insensitive to soil moisture (Bériaux *et al*., 2011; Kornelsen & Coulibaly, 2013).The granulometry and the texture have also an influence but this influence is much lower (Fellah, 1997).

The main objective of this study is to assess conditions for success of grassland mowings detection based on SAR data considering the numerous factors influencing the backscattering.

Two preliminary investigations support this assessment. The first one aims at assessing the backscattering coefficient sensibility to the main sources of variability such as water content and biomass. As grasslands can be either managed by mowing or grazing, a second preliminary study assesses the potentialities of SAR data to classify the grassland parcels according their land use or management mode.

Results of this chapter have been initiated on the basis of master theses of Thibault Delvaux and Christophe Bocquet.

2. Material and Methods

2.1. Field campaigns

Two field campaigns were completed in 2008 and 2010 in two ecological territories (Figure 4.1) located in Wallonia (Belgium), namely the 'Dépression Fagne-Famenne' (DFF) and the "haut plateau de l'Ardenne centrale" (HPAC) ecological territories.

For both years, a set of grassland parcels has been monitored. In 2008, the monitoring has been only done in HPAC ecological territory. For each of these parcels, the type of grassland management (mowing or grazing) and the current status of the parcels (i.e. for example if cows are currently grazing or if the mowing parcel has been cut or not) have been recorded. The GPS coordinates have been also collected at the borders of the parcels along roads

For practical reasons, these parcels were selected in a limited area within the ecological territories: in the vicinity of Libramont-Chevigny and Beauraing towns (represented by red triangles in figure 4.1) respectively for HPAC and DFF ecological territories.

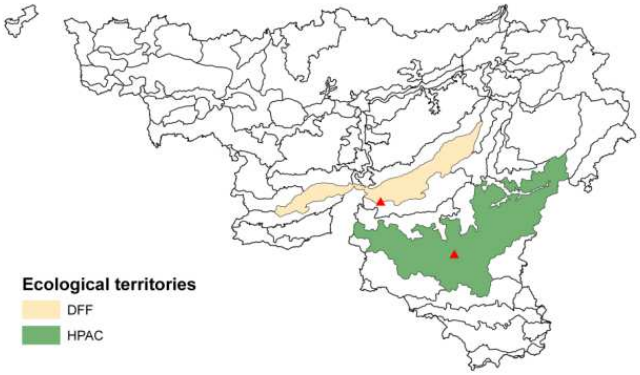


Figure 4.1.- Ecological territories including the monitored sites during the 2008 and 2010 field campaigns (DFF: "depression fagne famenne" ; HPAC: "hauts plateaux de l'ardenne Centrale")

61 grasslands parcels were monitored in 2008, 108 in 2010. Their distribution per management type and ecological territory is reported in table 4.1.

Table 4-1.- Number of observed parcels per year, ecological territory and management mode

Year	*Ecological territory*	Mowing	Grazing
2008	HPAC	28	33
2010	HPAC	14	33
	DFF	34	27

In 2008, 6 surveys were realized more or less on a weekly basis from the 28th of May up to the 7th of July. In 2010, management of grassland parcels was recorded for both ecological territories from the 25th of May up to the 28th of June 2010 at the same visit frequency than in 2008. Timing of field data collection is available in figures 4.2 to 4.4.

2.2. SAR data

A set of 15 ERS-2 images (Precision image (PRI) products) have been acquired for both field campaigns. ERS-2 sensor operates in C-band (5.3 GHz) and presents an incidence angle of 23° in VV polarization (table 4.2).

ERS-2 acquisitions (n=5) for 2008 are limited to HPAC ecological territory while acquisitions for 2010 (n=10) covered both territories. For both years of field campaign, taking into the cost of field campaigns, only the first mowing period was studied.

Table 4-2.- General information on acquired ERS-2 PRI images

Field Campaign	Date of acquisition	Orbit	Track	Pass	Spatial coverage[1]	
					HPAC	*DFF*
2008	18/05	68370	380	Descending	x	
	03/06	68599	108	Descending	x	
	16/06	68792	381	Ascending	x	
	22/06	68871	380	Descending	x	
	05/07	69064	72	Ascending	x	
2010	04/05	78619	108	Descending	x	
	17/05	78812	301	Ascending	x	x
	23/05	78891	380	Descending	x	x
	05/06	79084	72	Ascending	x	x
	08/06	79120	108	Descending	x	
	11/06	79163	151	Descending		x
	21/06	79313	301	Ascending	x	x
	27/06	79392	380	Descending		x
	10/07	79585	72	Ascending	x	
	16/07	79664	151	Descending		x

[1] 'x' in the HPAC or DFF columns means that the ERS-2 images covered the monitored area within respectively the HPAC and DFF ecological territories.

Due to programming constraints and the necessity to have frequent images for monitoring the grasslands management, it was not always possible to acquire images at the same mode (ascending or descending). Moreover, it was not always possible to cover the 2 ecological territories with a single image.

Ascending and descending images of ERS-2 are acquired respectively around 21:45 and 10:30. The risk of dew, and subsequently to observe increase of backscattering coefficient values, at both acquisition time is therefore plausible.

2.3. Meteorological conditions

The 2 selected ecological territories present slightly different growing conditions. If higher temperatures are recorded in the DFF ecological territory (synonymous with a faster grasslands growth and probably earlier mowings), rainfall is slightly higher in HPAC ecological territory.

In 2008, between April and July (4 months), 270 mm and 327 mm of rainfall were observed in 70 days (more than 1 day over 2) respectively for DFF and HPAC ecological territories. In 2010, on the same period, the number of rain days as the amount of rainfall was lower: rain occurred only 52 days and generated 186 and 201 mm respectively for DFF and HPAC.

Figures 4.2 to 4.4 present for both ecological territories the daily rainfall on the periods of grassland parcels monitoring. These figures show clearly that the number of rain days, and subsequently the potential number of days where backscattering can be affected, is relatively important whatever the considered ecological territory and the considered year.

Considering the highest frequency of rainfall events in 2008, it is not too surprising to see that proportionally more ERS-2 SAR images were acquired during a rain day and within or close to a rain period. The probability to record more noise in the backscattering in 2008 is therefore more important.

Meteorological data used in the frame of this study are hourly rainfall for two meteorological stations belonging to the PAMESEB network (www.pameseb.be) close (< 20 km as the crow flies) to the monitored parcels.

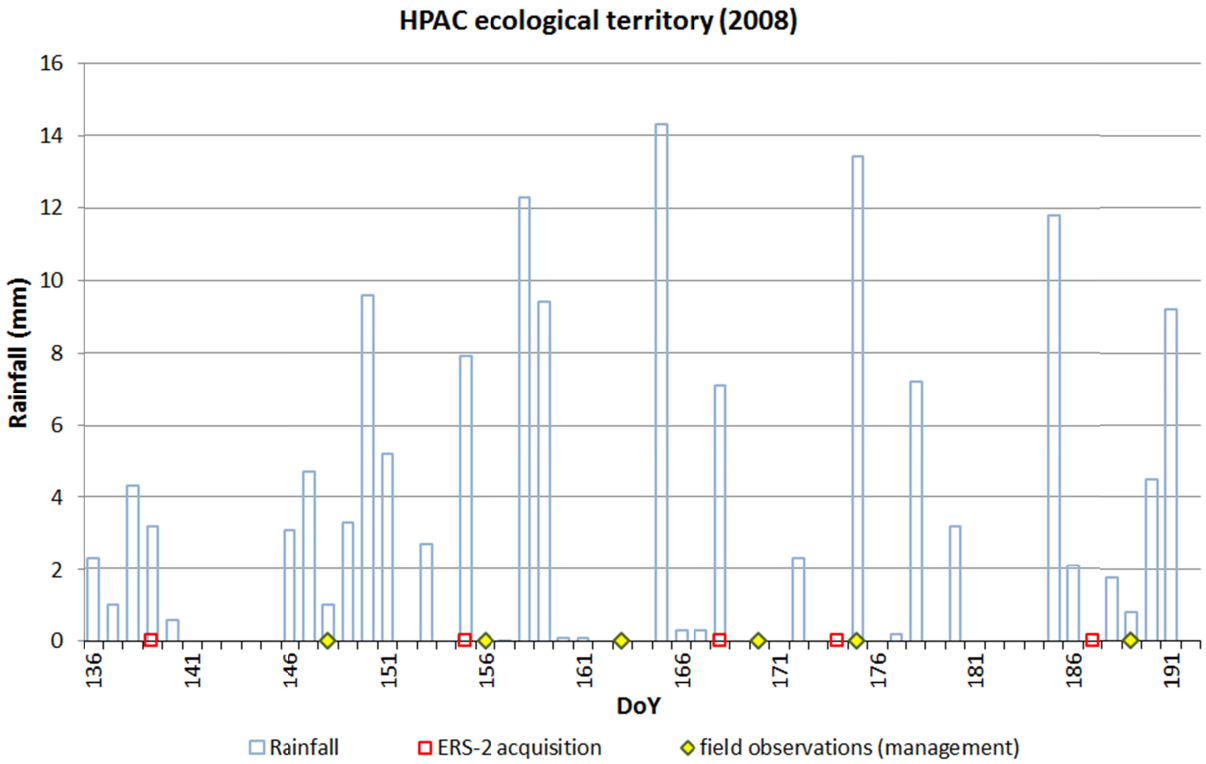


Figure 4.2.- Amount of rainfall (mm) and distribution of ERS-2 images and field observations (management) in HPAC ecological territory (year 2008)

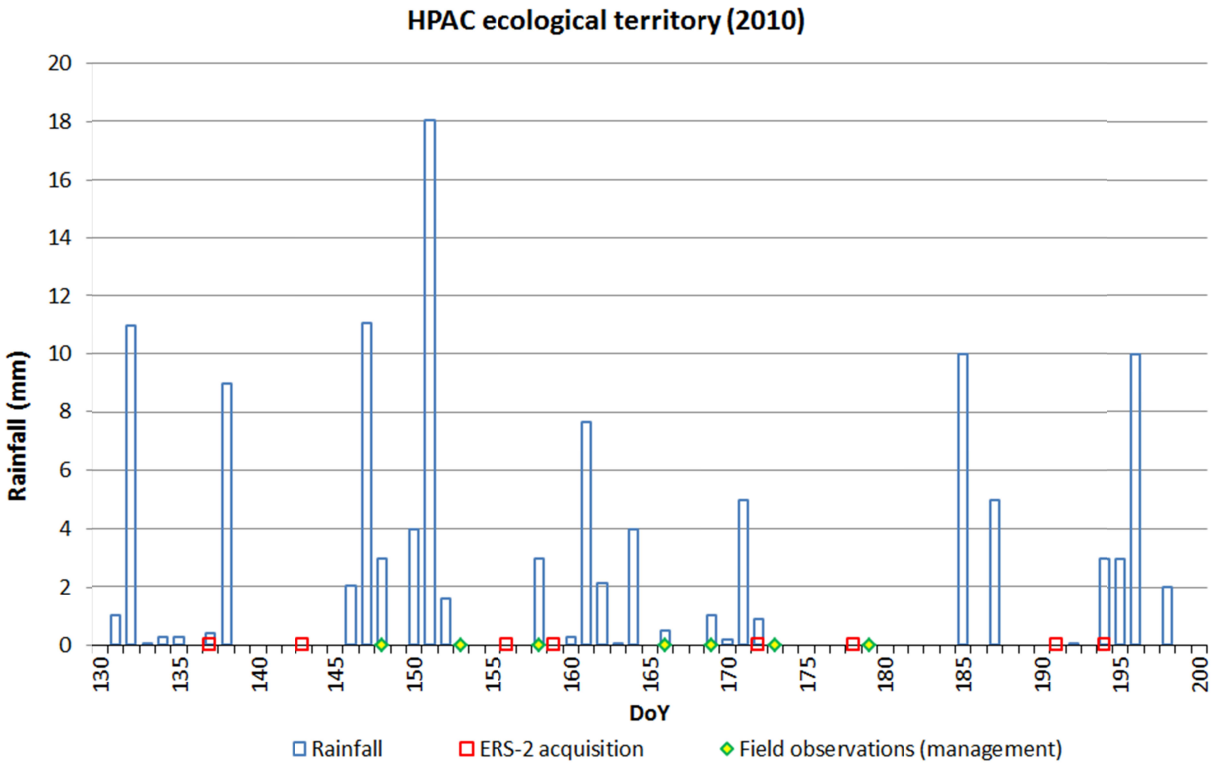


Figure 4.3.- Amount of rainfall (mm) and distribution ofERS-2 images and field observations (management) in HPAC ecological territory (year 2010)

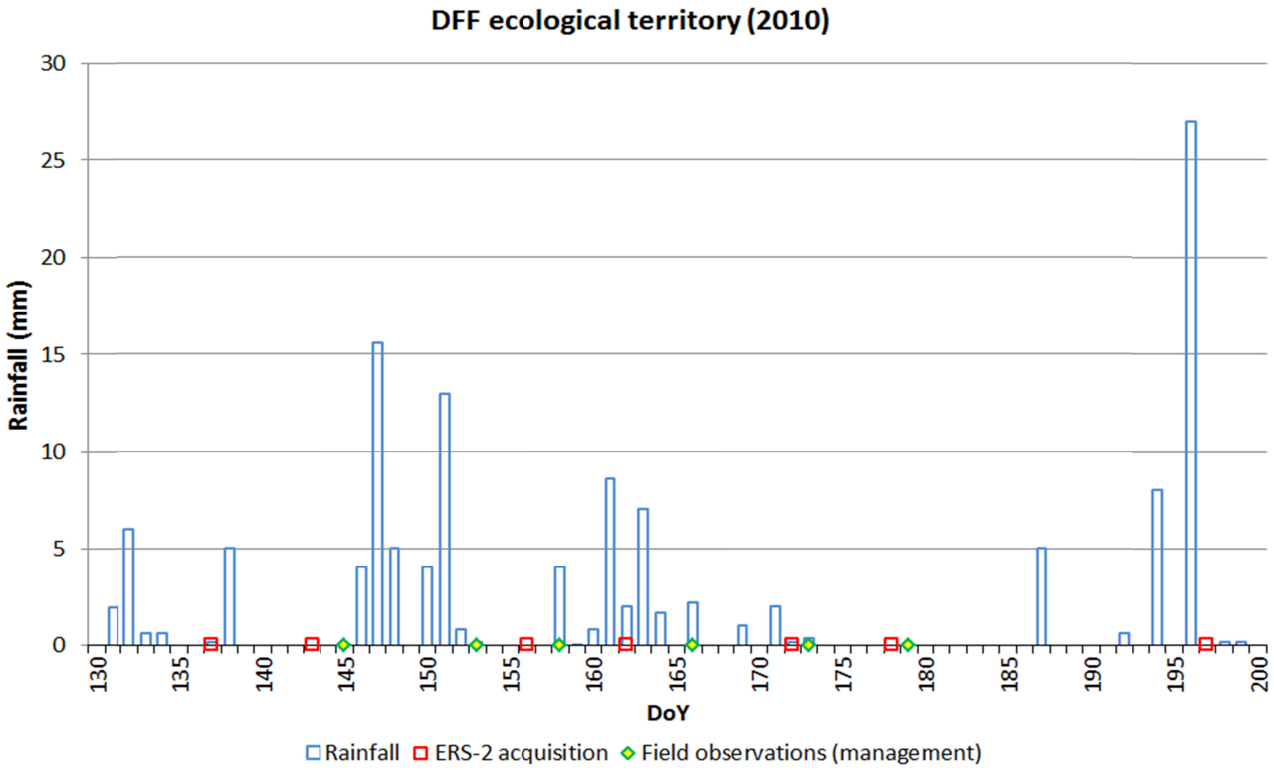


Figure 4.4.- Amount of rainfall (mm) and distribution ofERS-2 images and field observations (management) in DFF ecological territory (year 2010)

2.4. Methods

2.4.1. SAR data preprocessing

All the ERS-2 images have been preprocessed with SARscape module of ENVI software (*Exelis Visual Information Solutions, Boulder, Colorado)*. The preprocessing includes the images coregistration as well as geometric and radiometric corrections taking into account the local incidence angle of the SAR beam. A digital elevation model sampled at 90m (SRTM 3 sec) was used as ancillary data.

Images in ascending and descending modes have been preprocessed separately.

2.4.2. Computation of backscattering coefficients and selection of management parcels

The GPS coordinates of each parcel were identified on the basis of the annual Land Parcel Identification System (LPIS) maps.

For each parcel, a 30m perimeter buffer was delineated. Pixels in this buffer have been discarded to limit border effects and the backscattering coefficients were averaged.

In order to reduce speckle noise, only parcels including a certain number of pixels was considered. A threshold level of 100 pixels is recommended by Bally and Fellah (1995), roughly 75% and 97% of confidence level respectively for radiometric resolution bounds of +/- 0.5 and 1.0 dB. However, as the number of remaining parcels was too low (especially in HPAC ecological territory), the threshold was consequently reduced to 50 pixels. With this number of pixels, the confidence drops to 59% and 89% respectively for radiometric resolution bounds of +/- 0.5 and 1.0 dB.

Moreover, only parcels presenting at least 3 dates of observation have been considered.

The number of mowing parcels after this selection is equal to 10 and 24 respectively for HPAC and DFF ecological territories. As far as the grazing parcels are concerned, the number is equal to 29 and 23 respectively for HAPC and DFF ecological territories.

2.4.3. Land use detection

First, the parcel signatures per management mode were studied and the impact of biomass and water content on backscattering assessed through empirical relationships.

Water content is probably one of the most important sources of uncertainty on backscattering coefficients but no synchronous field measurement is available for these grassland parcels. To take this factor into account, the cumulated rainfall over the 72h preceding the SAR acquisitions (rain_3d) were computed as a proxy thanks to a very dense station network.

Based on the results of this exploratory analysis, a discrimination rule identifying the land use, i.e. the management type, was defined while the land cover is already known.

The data set has been split in a calibration data set composed of data for HPAC ecological territory (including therefore the 2 years, 2008 and 2010) and a validation data set composed of data for DFF ecological territory (year 2010) to assess the robustness of the proposed criterion.

2.4.4. Mowings detection (exploratory analysis)

This study aims mainly at studying the effect of mowings on backscattering and at assessing if the mowings or more exactly the mowing windows (time interval between two SAR images where the mowing has occurred) can be unambiguously identified.

Building on the land use correctly identified for all the parcels, two different approaches were defined, the first one dealing with the spatial variability while the second focuses on the temporal variability.

The first approach works in relative terms by comparing per acquisition date the backscattering coefficients of mowing parcels with the median backscattering coefficient of grazing parcels (G parcels). For mowing parcels, a distinction is made between parcels that are not yet mowed (NM parcels) and parcels that have just been mowed (JM parcels). This approach relies on the assumption that the fluctuation of biomass in grazing parcels is quite limited all along the growing season, at the contrary of mowing parcels with biomass in NM and JM parcels respectively higher and lower compared to the G parcels.

The second approach works also in relative terms but the comparison is made from an acquisition date to another. The approach aims at comparing the differences of backscattering between consecutive SAR acquisitions for NM and JM parcels.

For both approaches, decision rules to detect mowings were defined.

The dates of field campaign survey do not match necessarily with the SAR acquisition dates (figure 4.5). The time interval (Δt_2) between the moment the parcel is observed as mowed (tf_2) and the next SAR acquisition (ts_2) can be sometimes important. It varies between 3 and 18 days (7 days in average). The time interval (Δt_1) between the moment the parcels are observed for the last time as not mowed (tf_1) and the next SAR acquisition (ts_2), varies between 6 and 24 days (12 days in average). Mowing dates are therefore uncertain and included between tf_1 and tf_2.

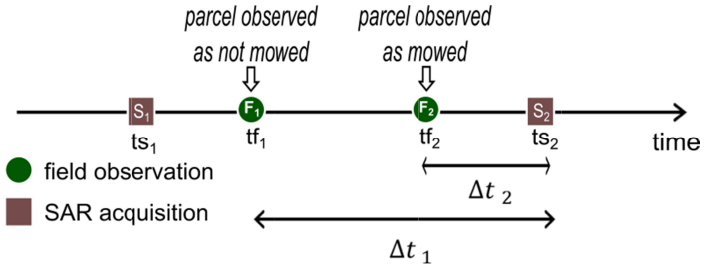


Figure 4.5.- Schematic representation of SAR acquisitions and field observations temporal distributions

In some parcels, grass has therefore regrown or has possibly been managed differently (for grazing) which limits the possibility to study the effect of mowing on backscattering.

To take partly into account this issue an additional selection was made for mowing parcels (for both approaches) based on the average of tf_1 and tf_2. Without any additional information, this average date represent the best estimator of the mowing dates. After a defoliation induced by the mowing, new leaves are formed using carbohydrates stored in the stubble of the plants. Respiration, especially of roots, increases after defoliation which limits the amount of carbon translocated to the growing tissues in the first 5-10 days after defoliation (Pearson & Ison, 1997). Regrowth is subsequently limited during this period. For this reason, parcels presenting a difference between the average of tf_1 and tf_2 and the next SAR acquisition (ts2) higher than 10 days have been discarded.

In the frame of the second approach, the time step between the SAR acquisition before tf_1 and the SAR acquisition following tf_2 can be sometimes very long for JM parcels. The difference is indeed not necessarily computed on two successive acquisitions. If a SAR image is acquired between the 2 field surveys F_1 and F_2, the status of the parcel at that moment is unknown. In such a situation, the difference computed between SAR acquisition S_2 and S_1 can be relatively important, only due to grassland growth. All the parcels with time step between S1 and S2 higher than 20 days have been discarded. The number of parcels remaining after this selection procedure is equal to 29.

3. Results

3.1. Field data analysis (mowing dates)

In 2008, most of the parcels have been observed as mowed (Figure 4.6) after day of year (DoY) 170 (18th of June). Some drier periods can be indeed observed in figure 4.2 in this period (after DoY 168 and 180).

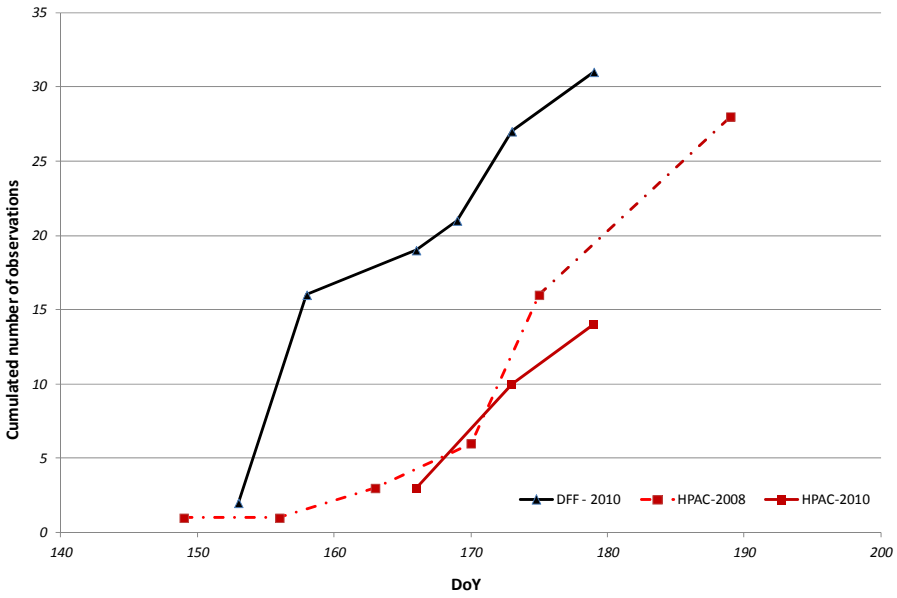


Figure 4.6.- Cumulated number of management parcels observed as mowed at the given day of observation for DFF ecological territory in 2010 and HPAC ecological territory in 2008 and 2010.

In 2010, as far as the HPAC ecological territory is concerned, first mowing operations in management parcels were observed the 15th of June (DoY 166) and the last ones the 28th of June (DoY 179). No parcels were observed as mowed during the survey at DoY 169 which is not too surprising considering the meteorological conditions at that moment. In the DFF ecological, roughly 50% of the monitored management parcels (16 over 31) have been mowed before the 7th of June (DoY 158), mainly after the 2nd of June (DoY 153) when a window of 4-5 dry days can be observed (Figure 4.4). The following mowings are more distributed in time.

In grassland growth models, only one mowing date is indeed usually used for a given region and a given mowing period. For a given ecological territory and a given year, we can observed that the range of mowing dates (for the 1st mowing period) is large (sometimes more than 50 days between the first and the last observed dates of mowing) which justifies furthermore our attempt to estimate these mowing dates by remote sensing.

3.2. Parcels signature

The distribution of backscattering coefficients for grasslands in C-VV ranges from -6.5 to -15 dB (Figure 4.7). These values are coherent with values provided for grassland with VV polarization in C-band (Stolz and Mauser, 1997; Hill et *al.*, 1999; Saich and Borgeaud, 2000,) but also in X-band (Bargiel *et al.*, 2010).

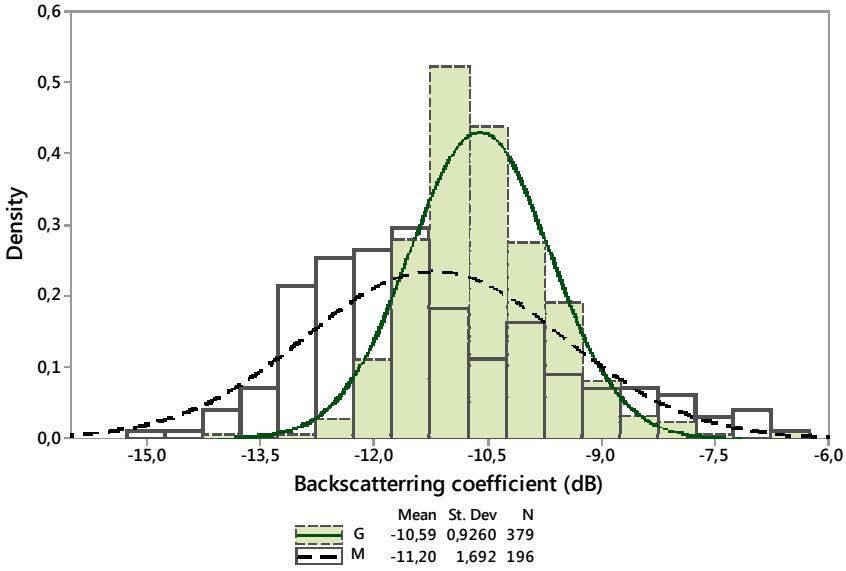


Figure 4.7.- Distribution of backscattering coefficient per management mode (G: grazing; M: mowing) and adjustments (plain and dashed lines) under the normal distribution assumption

Obvious differences can be noticed between grazing (G) and mowing (M) parcels: grazing parcels tend to present higher backscattering coefficients and a lower dispersion around the mean (lower standard deviation).

These differences are observed whatever the year and/or the ecological territory (results not shown).

These contrasted distributions can be explained by a difference of biomass. In grazing parcels, grass is periodically ingested by herds. We can therefore consider that the fluctuation of biomass in grazing parcels is quite limited all along the growing season, at the contrary of mowing parcels. The comparison of backscattering coefficients of mowing parcels before and just after a mowing reinforce this assumption: the average backscattering coefficient for mowing parcels before the mowing (NM parcels) is equal to -11.69 dB (σ=1.33 dB, n=151) while mowing parcels just after the mowing (JM parcels) present a backscattering coefficient average of -9.50 dB (σ=1.75dB, n=44). When biomass increases, the contribution of volume scattering to (total) backscattering increases as well which tends to attenuate the signal and to reduce the backscattering coefficient values.

Biomass can probably not explain alone the distribution of backscattering coefficients observed per management mode. As aforementioned, other factors influence the backscattering.

If we plot the backscattering coefficient average of mowing parcels (before the mowing) against the average coefficient of grazing parcels (figure 4.8) for each acquisition date, we can notice that conditions prevailing at the moment of images acquisition affect similarly the two management modes: lower backscattering coefficient values for grazing parcels tends indeed to correspond to lower backscattering coefficient values for mowing parcels and vice-versa.

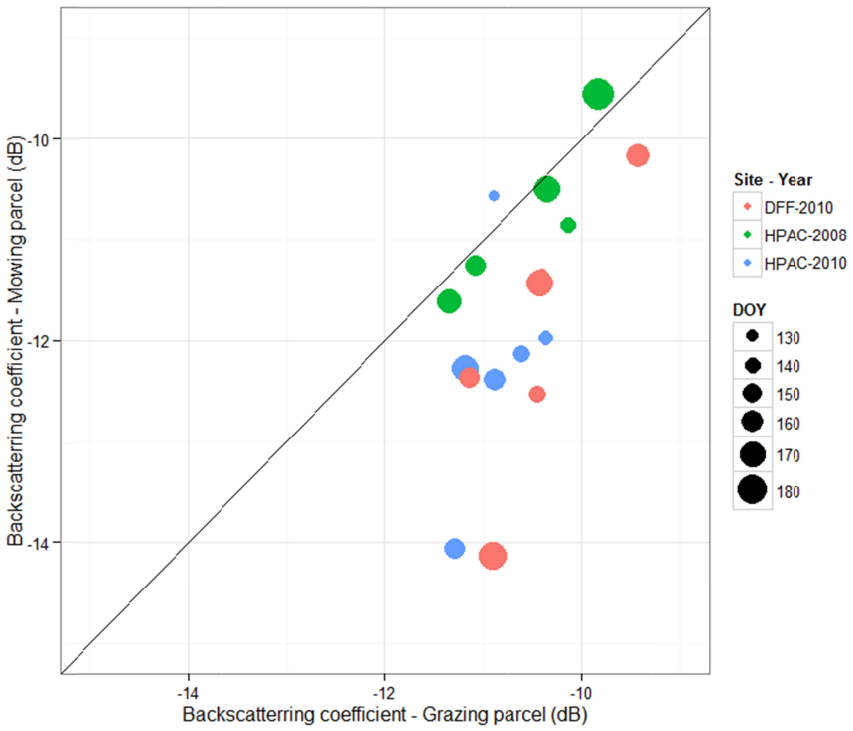


Figure 4.8.- Relationship between the average backscattering coefficient per acquisition date for grazing parcels and the average backscattering coefficient for mowing parcels (diagonal line represents the 1:1 line) – the points size is proportional to the day of year (DoY)

A deviation in regards to 1:1 line is observed, especially when backscattering coefficient values for NM parcels become lower than -12 dB. This deviation can certainly be imputed to the effect of biomass on backscattering.

3.3. *Effect of water content on backscattering*

Water content is probably the most significant factor influencing the backscattering.

The relationship between average backscattering coefficients per ecological territory and per acquisition date and the sum of rainfall over 72 hours (3 days) before this date (rain_3d) shows, as expected, an increase of backscattering coefficient with the increase of rainfall (Figure 4.9).

Coefficient of determination is equal to 0.4. No significant impact of type of orbit (ascending or descending) is observed. We however observe that in absence of rainfall (at meteorological station level), descending ERS-2 images (acquired around 10h30) tend to present higher backscattering coefficient values than images acquired in ascending mode (around 21h45). This difference possibly means that dew is more present in the morning when images are acquired in descending mode.

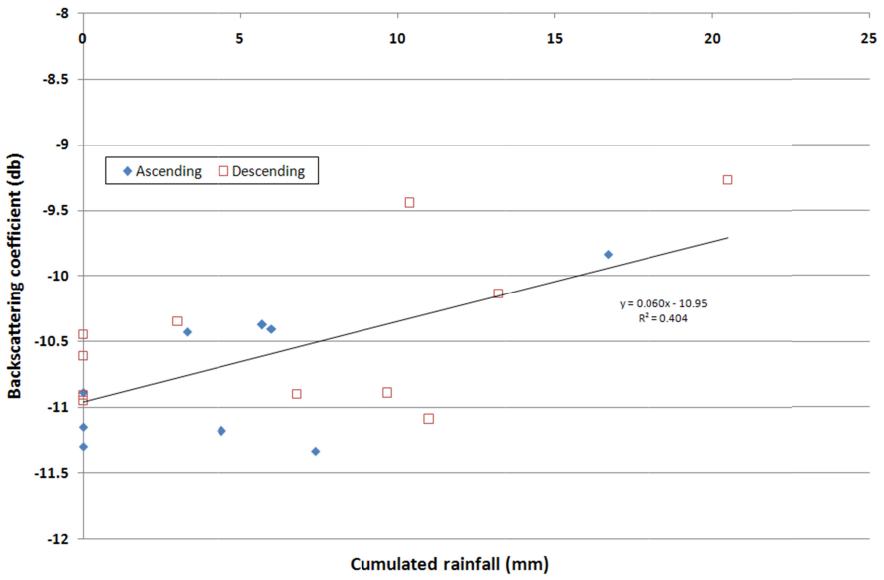


Figure 4.9.- Evolution of the average backscattering coefficient per acquisition and per ecological territory for grazing parcels according to the cumulated rainfall recorded at meteorological stations level over 72 h before the ERS-2 acquisitions

The relationship observed in figure 4.9 can explain partly the variability observed for backscattering coefficients of grazing parcels in figure 4.6 and has direct consequences on land use detection. Observations in the tails of the distributions for grazing and possibly mowing parcels (figure 4.7) are probably not outliers but correspond to backscattering in particularly dry or wet conditions.

In order to assess the influence of water content on biomass backscattering, the evolution of the differences between average backscattering coefficients per

acquisition date and per ecological territory of G and NM parcels and rain_3d has been reported (figure 4.10).

The differences of backscattering are higher in dry condition. When the soil is dry, the contribution of biomass water content on backscattering is predominant. When the soil moisture increases, its contribution on backscattering increases and then reaches a level where contributions of both soil and plants to the SAR response tend to be similar and thus confused. This behavior was also observed by Auquière (2001), Dabrowska-Zielinska et al. (2007) and Beriaux (2011). Therefore, G and NM parcels can't be distinguished and hamper the detection of land use. When the soil moisture further increases, the backscattering from the soil masks the contribution from the biomass water content.

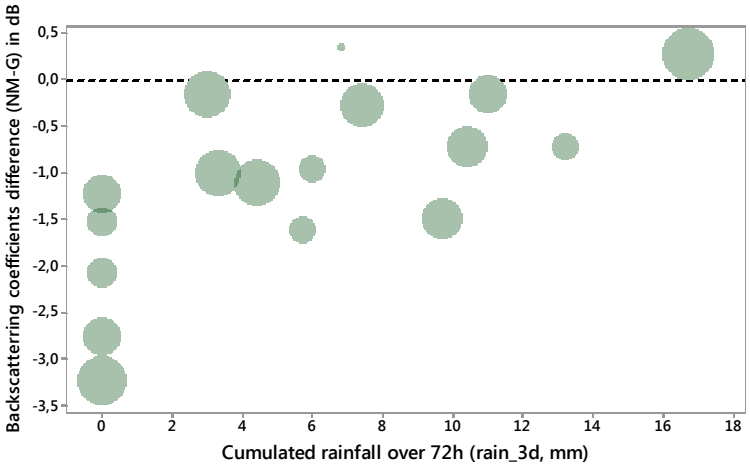


Figure 4.10.- Differences between average backscattering coefficients of NM and G parcels for each acquisition as a function of the cumulated rainfall over 3 days before the SAR acquisition (rain_3d) - the points size is proportional to the day of year (DoY)

3.4. Management detection

Based on these results, a procedure building on the difference in backscattering coefficients distributions (figure 4.7) and the effect of water content on backscattering coefficient (figure 4.9) is proposed to classify the land use.

More specifically, the procedure is based on the distribution of backscattering coefficients of grazing parcels. If an observation for a given parcel steps outside this distribution, this parcel is not considered as a grazing parcel and is subsequently labelled as a mowing parcel. As shown in figure 4.9, water content has to be taken into account to define the distribution of grazing parcels for a given situation (a given year and/or a given area for example).

The procedure is schematically presented in figure 4.11.

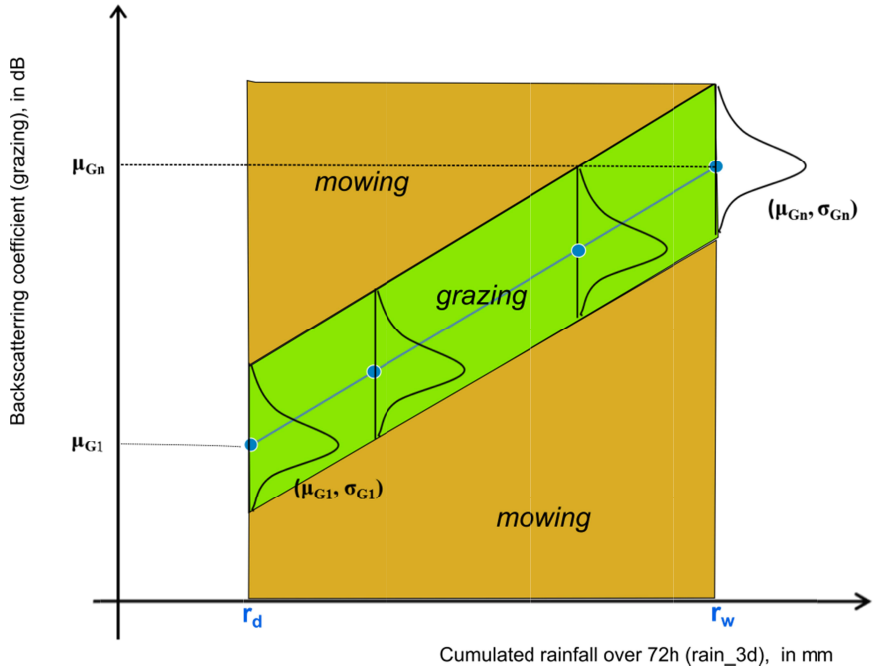


Figure 4.11.- Schematic representation of the procedure to classify grassland parcels according to management type (mowing or grazing)

The limits of the distribution for grazing parcels for a given day (a given value of rain_3D) are set both on the relationship between the average backscattering of grazing parcel and the cumulated rainfall over 72h (rain_3d) and on the assumption of Gaussian distribution of backscattering coefficients for grazing parcels for a given acquisition date. If all the backscattering values available for a given parcel (and a given mowing period) are included within the limits of the probability (fixed at 99%) interval of this distribution then this parcel is labelled as grazing.

The standard deviation values for the distributions of backscattering coefficients for grazing parcels per acquisition date ($\sigma_{g1},\ldots,\sigma_{Gn}$ in figure 4.11) have been calibrated based on HPAC data set.

μ_{G1} and μ_{Gn} are the average backscattering coefficients respectively for the driest (r_d) and wettest (r_w) dates of observations.

Figure 4.12 presents the distribution of these standard deviations for all the acquisition dates.

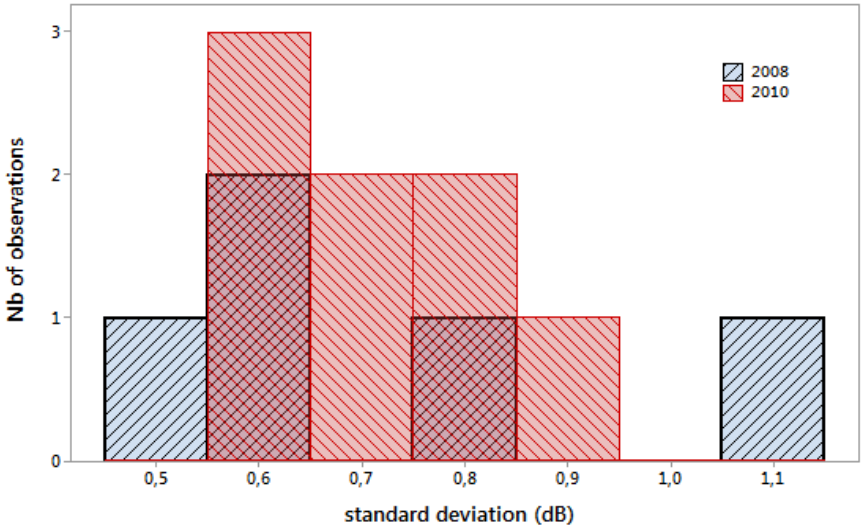


Figure 4.12.- Distribution of standard deviations of backscattering coefficients for grazing parcels per acquisition date

For the sake of simplicity and considering that no significant relationship has been found between these standard deviations and the rain_3d (results not shown), the median value (0.66 dB) of the standard deviations has been selected to set the distribution interval allowing to make the distinction between the grazing and the mowing parcels.

Application of the procedure on the calibration data set provides as expected very promising classification results (table 4.3). Producer's accuracy for grazing and mowing parcels are respectively equal to 79.3% and 90.0%. The overall classification accuracy is equal to 82.1%.

Table 4.3.- Confusion matrix (calibration data set)

Reference →		Grazing	Mowing	*total*	**User's accuracy**
classification	Grazing	23	1	*24*	**95.8%**
	Mowing	6	9	*15*	**60.0%**
total		*29*	*10*	*39*	
Producer's accuracy		**79.3%**	**90.0%**		

The classification results according to the years tends to present a lower classification results for mowing parcels in 2008 (table 4.4).

Most of the SAR data in 2008 were acquired in wet conditions (high rain_3d values). In these conditions, as observed in figure 4.9, backscattering coefficients of mowing parcels tend to be similar to those observed for grazing parcels which can explain the confusion between the 2 land uses.

Table 4.4.- Confusion matrix per year (calibration data set)

2008					
Reference →		Grazing	Mowing	*total*	**User's accuracy**
classification	Grazing	5	1	*6*	**83.3%**
	Mowing	1	2	*3*	**66.7%**
total		*6*	*3*		
Producer's accuracy		**83.3%**	**66.7%**		
2010					
Reference →		Grazing	Mowing	*total*	**User's accuracy**
classification	Grazing	18		*18*	**100.0%**
	Mowing	5	7	*12*	**58.3%**
total		*23*	*7*		
Producer's accuracy		**78.2%**	**100.0%**		

Applied on validation data set, the procedure provides as well quite good classification results (table 4.5). The overall classification accuracy is equal to 78.7%. The producer's accuracy is also good and reaches 82.6% and 75.0% for respectively grazing and mowing parcels.

Table 4.5.- Confusion matrix (validation data set)

Reference →		Grazing	Mowing	*total*	**User's accuracy**
classification	Grazing	19	6	*25*	**76.0%**
	Mowing	4	18	*22*	**81.8%**
total		*23*	*24*	*47*	
Producer's accuracy		**82.6%**	**75.0%**		

3.5. Mowings detection

As mentioned in paragraph 3.2, the backscattering ofgrassland parcels just after mowing (JM parcels) tends to be lower than before the mowing but also lower than backscattering of grazing parcels. The distribution of backscattering coefficient (figure 4.13) is also wider (σ=1.75 dB).

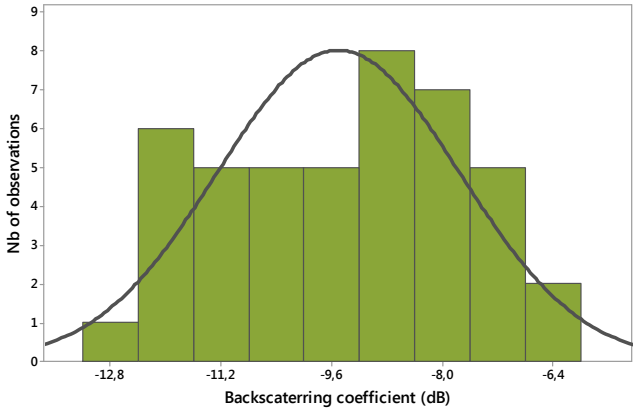


Figure 4.13.- Distribution of backscattering coefficients of JM parcels

As for grazing parcels, the backscattering increases with the sum of rainfall over the 72h preceding SAR acquisitions (results not shown).

3.5.1. Approach based on the comparison with the median backscattering of G parcels

The differences for each acquisition date between the backscattering coefficient of JM parcels and the median of backscattering coefficient of G parcels is most of the time positive (figure 4.14). The difference average equals 0.52 dB (median = 0.82 dB). These results were expected as biomass per hectare remaining after mowing tends to be lower than the one observed in grazing parcels.

Negative values, sometimes lower than -1dB, can be however observed that would mean that biomass per hectare is higher in these parcels. Let's note that grassland swaths have been observed in 4 of the 5 parcels with a difference of backscattering lower than -1 dB. These observations have been made 5-6 days before the SAR acquisition. A possible explanation for these lower differences of backscattering coefficient values could be the presence of these grass swaths at the day of ERS-2 acquisition. Let's also note that 6 over the 8 parcels for which swaths have been observed a couple of days before the next SAR acquisition present a lower difference of backscattering compared to median backscattering coefficient of G parcels. No evidence of the influence of grassland swath on backscattering coefficient has been found in literature. We can however notice that McNairn *et al.* (1997) have shown that, using C-VV SAR band, the backscattering coefficient values of winter barley residue plots were significantly lower from those of bare soils / no residue plots. The

plausibility of the hypothesis of an effect of remaining swath on the field on backscattering is a bit reinforced by the fact that at the next acquisition, the backscattering coefficient values for these parcels are a bit higher or similar than median backscattering of grazing parcels.

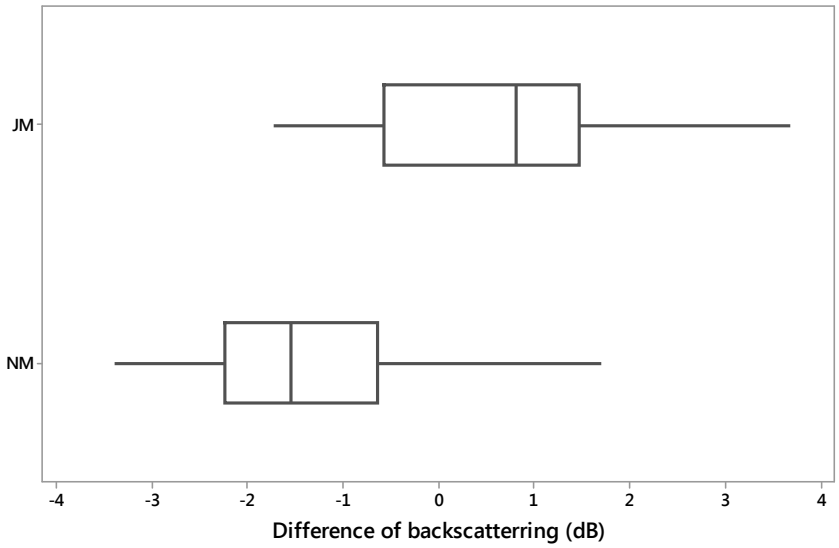


Figure 4.14.- Difference (in dB) of backscattering coefficient of NM and JM parcels with the median backscattering coefficient of G parcels

Unlike for the JM parcelsthe differences between the backscattering coefficient values of NM parcels with the median of backscattering coefficient values of G parcels are most of time negative. In average, this difference is equal to -1.34 dB (median of -1.54 dB).

The range is however quite wide for both categories: the standard deviation for NM and JM parcels (compared to the median backscattering of G parcels) is respectively equal to 1.17 dB and 1.35 dB. This dispersion is observed for each acquisition date. No pattern can be associated to the rainfall distribution (figure 4.15). This situation can hamper the detection of mowings (by inducing the probability of false detection of mowings).

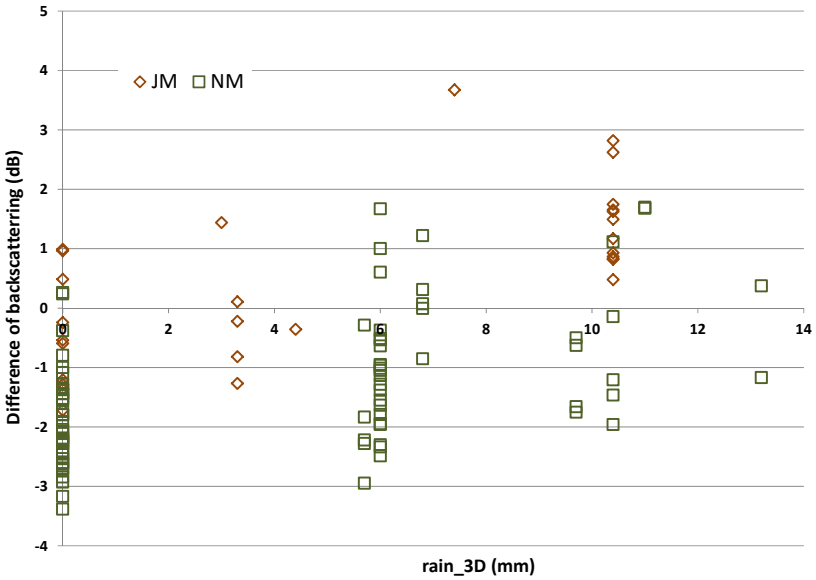


Figure 4.15.- Difference of backscattering (in dB) with the median backscattering of G parcels for NM and JM parcels according to rain_3D (in mm)

A threshold value for the difference with median backscattering of G parcels could be to set to detect mowings, all the values above this threshold being labeled as a mowing.

When the threshold is low (-0.5 dB), mowings are identified in 21 parcels over 29 but for 13 of these parcels are false detection of mowings are observed i.e. values above the threshold are observed for these parcels when they are not yet mowed. It means that mowings are unambiguously identified in 8 parcels (no false detection of mowings).

Increasing the threshold decreases the number of parcels where mowings are identified. The number of parcels with no false detection (equal to 8) remains however constant up to a threshold of 0.5 dB.

Table 4.6.- Number (%) of parcels where a mowing is detected and number of parcels with false mowings according to the threshold value fixed on the difference of backscattering with median backscattering of G parcels.

Threshold (dB)	Nb of parcels (mowings)	Nb of parcels with false mowings detection
-0.5	21 (72%)	13
0	18 (62%)	10
0.5	15 (52%)	7
1	9 (31%)	6

This relatively low efficiency of the mowings detection procedure based on relative difference in space is probably due to some variability / uncertainty factors not considered in the analysis.

The second approach could possibly better take into account these factors as far as they remain constant from an acquisition date to another.

3.5.2. Approach based on the temporal difference of backscattering

Figure 4.16 presents the backscattering difference (per parcel) between SAR acquisitions according to their corresponding difference of rain_3d for NM, G and JM parcels. The SAR acquisitions used to compute this difference are different according to the parcels category. For G parcels category, considering that the biomass variation in these parcels from an acquisition to another is supposed as marginal, the all the possible combination of backscattering differences have been computed. For NM parcels category, only the difference between two successive SAR acquisition dates is considered.

For JM parcels category, as unfortunately the field campaign dates do not match necessarily with the SAR acquisition dates, the considered SAR acquisitions are the SAR acquisition following the mowing and the SAR acquisition preceding the last field observation of the parcel as not mowed.

Whatever the parcels category, the differences are computed by subtracting the value for the earlier acquisition from the later. Positive values for the rain_3d and backscattering differences mean therefore respectively that the later SAR image has been acquired in wetter conditions and present a higher backscattering compared to the previous considered one.

Figure 4.16 allows stressing different aspects. The first one is that mowing induces most of time an increase of backscattering. A decrease is only observed when the presence of grassland swaths is suspected.

The variation of the difference of backscattering for G parcels is lower compared to the situation observed for JM. This variation is essentially due to the variation of water content as the G parcels biomass is considered as stable from an acquisition date to another. This assumption of biomass stability is reinforced in figure 4.16: when conditions are similar for both SAR images (difference of rain_3D=0), no difference of backscattering is observed.

As for JM and G parcels, a global trend towards an increase of the backscattering difference with the increase of the difference of rain_3D can be observed for NM parcels. The difference of backscattering is positive when the difference of rain_3D is positive meaning that the increase of water content has a stronger effect (increase) on backscattering than the increase of biomass.

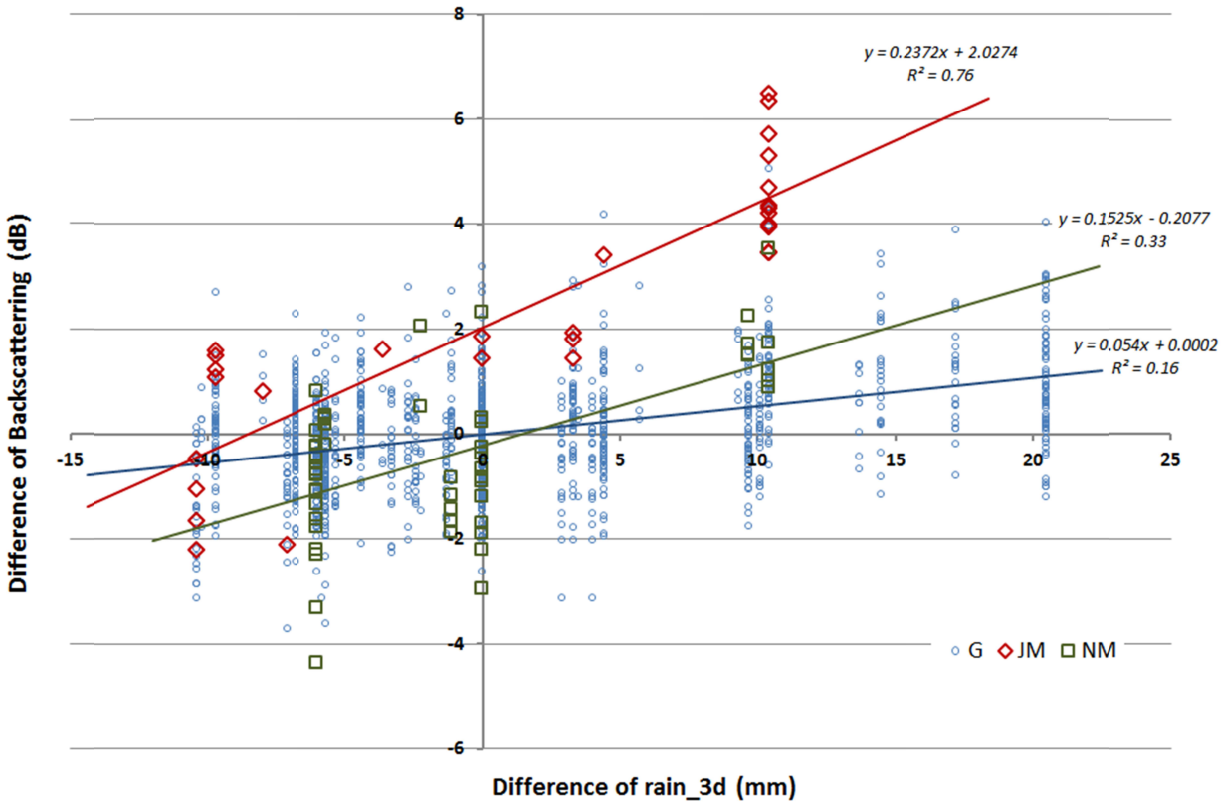


Figure 4.16.- Relationship between the backscattering difference (per parcel) between SAR acquisitions according to their corresponding difference of rain_3d for NM, G and JM parcels

We can also notice that regression line for NM crosses the regression line for G parcels when the difference of rain_3D is close to zero, the difference of backscattering being higher for NM parcels when the difference of rain_3D is positive. This behavior is probably due to the fact that the contribution of biomass on backscattering tends to be masked at higher water content (figure 4.10).

The difference between the regression lines for NM and JM parcels for a given difference of rain_3d varies in our conditions roughly between 2 and 3 dB. The standard errors of regression are quite high, equal to 1.23 and 1.20 dB respectively for NM and JM parcels. Figure 4.16 allows a much better understanding of the respective influence of biomass accumulation, mowing and moisture on the backscattering.

A decision rule can be defined to detect mowings based on a similar approach than the one used to detect land use (chapter 3.4), i.e. based on the prediction intervals of regressions for NM and JM parcels. These prediction intervals (at 99%) overlap (figure 4.17). There are therefore situations where a difference of backscattering can't be attributed to a mowing (area in blue in figure 4.15) without confusion with the increase of backscattering for NM parcels.

Mowing can be identified without ambiguity only when the difference of backscattering is higher than the higher limit of the confidence interval for NM parcels (area in yellow in figure 4.17). At 99% of confidence level, only 4 mowings can be detected (no false detection of mowings is observed) in these conditions. If the confidence interval is reduced to 95%, the number of identified mowings increases up to 14 (48% of mowings detected) but a false detection of mowings is observed for 2 parcels. These 2 parcels are different from the 14 where a mowing has been detected which means that the mowing interval is not correctly estimated for these 2 parcels. In a context of no optimal data set, these results show that this temporal approach is quite promising and relies on a relevant rationale.

Let's note also that the effect of water content on contribution of volume scattering on total backscattering (observed in figure 4.10) has not been considered yet and that the relationships for JM and NM parcels are specific to the (average) time step between successive SAR acquisitions considered to compute the differences. In our study, this time step for NM parcels varies between 6 and 13 days (average of 8 days). If this time step is smaller, the difference of backscattering will be smaller as well (and conversely). For JM parcels, the average is equal to 17 days. The relationship for JM parcels is also specific to the time step between the moment the parcels are observed as mowed and the next SAR acquisitions. Minimizing these 2 time steps should increase the difference of backscattering and ease the detection of mowings.

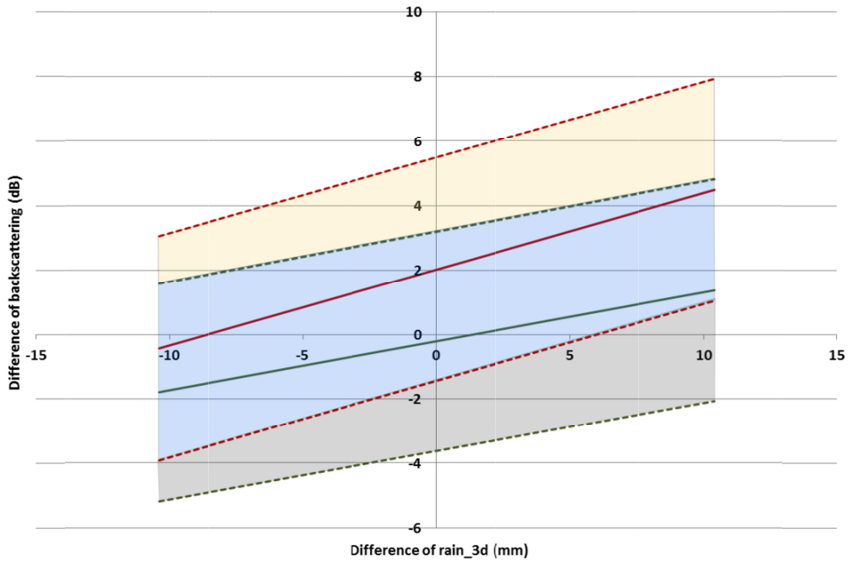


Figure 4.17.- Prediction intervals at 99% confidence level (dashed lines) for the regression lines (solid lines) between the difference of backscattering according to the difference of rain_3d for JM (in red) and NM (green) parcels

.

4. Conclusions and discussions

As expected, the conditions of acquisition have an impact of backscattering coefficient values.

An increase of backscattering has been observed with the increase of the proxy (rain_3d, cumulated sum of rainfall over 72 hours before the SAR images) used to represent the water content.

At the contrary, a decrease of backscattering with the increase of biomass tends to be observed at least in dry conditions (low values for rain_3d). The decrease of backscattering with the increase of biomass is attenuated in higher water content conditions considering that in these conditions the contributions of both soil and plants to the SAR response are tangled.

This confusion of contributions of both soil and plants to the SAR response could hamper the classification of grassland parcels according to their land use / management type (mowing vs grazing). Our study has however shown that good overall classification accuracy (around 80%) can be reached in validation thanks to the backscattering distribution of grazing parcels as far as some SAR images are acquired in dry conditions.

The significant effect of water content on backscattering implies to consider it with a view to detect mowings. In this context, two approaches have been considered. The first one works per acquisition date (spatial approach) and is based on the differences of backscattering with the median of grazing parcels. The second one uses a temporal approach and is based on the relationship between the differences of backscattering between 2 successive SAR images against the differences of the corresponding rainfall history (rain_3d).

Though both approaches allows identifying mowings, the second approach provides better results with roughly half of the mowings detected (against roughly on quarter with the first approach) with a limited number of false detections. These better results observed for the second approach are probably due to the fact that some factors of variability that were not considered in our analysis (such as maybe the local incident angle) are masked in the temporal approach.

The 2 methodologies could be certainly refined and possibly improved by considering a better proxy for water content. The current proxy does not take for the moment into account the distribution and the intensity of rainfall during the considered period of 72h. A possible solution could be therefore the use of a hydrological model such as for example the SWAP ('Soil Water Atmosphere Plant') model (Kroes *et al*., 2008; Van Dam *et al*., 2008) as shown by Bériaux (2011) for maize.

In the same way, additional analyses are needed to assess the impact of the errors made on land use classification on mowings detection.

The efficiency of the mowings detection can be however considered as satisfying considering the relatively low temporal availability of ERS-2 acquisitions and the reduction of the number of usable observations due to non-optimal organizations of

field campaigns (non-coincidence between the days of acquisition and the days of field campaigns).

This study has also showed that the field observations protocol has to be modified. Field observations should be done as much as possible on the same day/ same hour than SAR acquisitions but also in-between these SAR acquisitions (especially if the time step between SAR acquisitions is important). The frequency and the moment of field observations between 2 SAR images should be adapted according to *a priori* information on the probability of mowings. Mowings occurred mainly in dry conditions, field surveys could be organised in priority when these conditions are forecasted.

The study has also raised a possible effect on swaths at the moment of SAR acquisitions on backscattering. This effect is currently hypothetic and has to be assessed in the frame of new field campaigns but if this effect is validated, accounting it could also improve the efficiency of mowings detection. Such a revisit capability is very recently achieved by the recently achieved Sentinel-1 SAR sensor.

The efficiency of mowings detection with the second approach could be certainly improved if the time step between SAR acquisitions gets smaller: the differences of backscattering for NM parcels would remain limited (as the increase of biomass between the 2 acquisitions would be small) and the difference of backscattering for JM parcels at or close to its maximum.

CONCLUSIONS AND PERSPECTIVES

Conclusions and discussions

The main objective of this thesis was to assess possible improvements for the assimilation of remotely-sensed biophysical variables in growth models and to estimate their related errors reduction on modelled yield estimation.

The assimilation improvement of biophysical variables derived from satellite data has been studied from different points of view which will structure this synthesis.

(1) Interplay between the accuracy and the temporal resolution of remotely-sensed data

Two Observing System Simulation Experiments (OSSE) have been used to assess the requirements in terms of accuracy and temporal resolution of remotely-sensed data in order to reach a recommended level of error reduction on yield estimation after data assimilation (*chapters 1 & 3*). Different observations to assimilate (covering optical and microwave spectral domains) and crop growth models have been considered in these experiments.

Both experiments confirm that assimilation of remotely-sensed data derived from the currently available sensors in growth models represents already a significant potential to improve yields estimation but also show that these improvements are not the optimum and that the errors on yields estimations can still be reduced.

Improving both the accuracy and temporal resolution of remotely-sensed data provides logically the best reductions of errors on yields estimations. However, according to the assimilated observation, the component (accuracy or temporal resolution) to target in priority can be different.

In the first OSS experiment (*chapter 1*), focusing on the assimilation of LAI time series in WOFOST model, reduction of the errors on final grain yields estimation can reach up to 65% but require having frequent and accurate (with an uncertainty level (CV) equal to 10% or ideally lower) LAI observations all along the growing season.

Substantial improvements can be fortunately reached earlier in the season (ideally at least when the peak of LAI is reached) which is a prerequisite to set up a system able to provide accurate and timely information about potential and observed harvest shortfalls and able to tackle the food security challenge. Reduction of errors on final grain yields estimation of 30-35 % have been observed with LAI observations presenting an uncertainty level of 15% and available every 2 weeks from the onset of greenness up to the beginning of the LAI decreasing phase.

From a global point of view, the OSS experiment has also stressed that reducing the uncertainty level on the LAI estimate seems to be more efficient to reach high errors reduction (> 50%) suggesting that the remote sensing community should rather focus efforts on improving LAI retrieval algorithms rather than improving the temporal availability of the observations.

Currently, the uncertainty level on remotely-sensed LAI reached with the new generation of high-resolution optical sensors such as LANDSAT-8 or RapidEye is around 15-20% (Shang *et al.*, 2014; Kross *et al.*, 2015). The uncertainty level is function of the sensor characteristics but also of the retrieval algorithm (Rivera Caceido, 2014).In these conditions, the expected errors reduction on yields estimations varies, according to the temporal availability, between 15 and 65 % but is usually lower than 50%.Errors reduction should be more important with the forthcoming Sentinel-2 mission as the uncertainty level on LAI is expected at 10% (Drusch *et al.*, 2012).

The temporal resolution is more critical for punctual events such as mowing dates. The second OSS experiment (*chapter 3*) has allowed to highlight that mowing period (period between two SAR observations during which the mowing occurs) could be identified on the basis of ERS-2 synthetic data at worst with an accuracy of 76% as far as the uncertainty level on backscattering coefficient is lower or equal than 1.5 dB and the timestep between two synthetic ERS-2 images is equal or lower than 12 days. Forcing LINGRA grassland growth model with the estimated mowing dates allows once again, as expected, to reduce the errors on simulated total grassland biomass compared to the situation without assimilation by at least 25% (up to 65-75%) when SAR observations are available every 6 days.

(2) The assimilation strategies and their implementation

A ‘sequential’ (based on Ensemble Kalman Filter) and a ‘calibration’ assimilation strategy have been used and compared in *chapter 1*.

The substantial improvements on yields estimations were reached only with the second approach. The existence of a significant phenological (temporal) shift between simulated and observed LAI time series can indeed limit the efficiency of assimilation methods as far as these methods are not able to correct this phenological shift. An increase of the errors on final grain yields estimation (of at least 27-28%) has been indeed observed with Ensemble Kalman Filter (EnKf) assimilation method. The EnKf (sequential) assimilation method is not able to correct the phenology contrary to the recalibration based assimilation method as far as this method consider parameters (such as TSUM1 or crop emergence) allowing this correction. The poor performance observed of the EnKf cannot be however imputed to the method in itself but only on the existence of this phenological shift: In the situations where the phenological shift between simulated and observed was small, EnKf assimilation method was also able to reduce the errors on final yields estimation.

A phenological shift/discrepancy arises mainly from uncertainties on the parameters and initial states (e.g. dates of sowing or emergence) driving the phenological stages in the models but can also arise, in some situations, from a lack of pixel purity and possible gridding effects resulting from the use of medium resolution sensors (*Chapter 2*). The use, in this thesis, of a sensor such as MODIS has the advantage to

provide quantitative remote sensing information with a time step more compatible with an operational use for assimilation in crop growth models but its medium resolution and its associated gridding effects induced by this whiskbroom imply that the observational footprint is not fully covered by the targeted crop. Chapter 2 has shown that the part of the information of these mixed pixels that doesn't come from the target crop influences the global GAI time series. This influence is different according the contributing crop(s) that are not the target crop. A significant proportion of summer crops within selected MODIS grid cells tends for example to (phenologically) shift the observed GAI time series, to decrease observed GAI values in the increasing phase of GAI and to decrease these values in the decreasing phase of GAI.

Prior efforts to correct any possible phenological shift / synchronise in time simulations and observations have therefore to be undertaken before any other efforts of improvement. The utility of adjusting the decreasing phase of LAI, already highlighted in previous studies (Curnel and Oger, 2006a ; Martin *et al*., 2006), has been for example confirmed in chapter 1 but only if a readjustment of the increasing phase (through crop parameters such as TSUM1 or crop emergence) is achieved. In other words, it means that without adjustment of the moment of the peak of LAI reduction of errors on final grain yields estimation by recalibration of the SPAN parameter is at best limited or inexistent (with in this case an increase of the errors). The necessity of correctly adjusting the moment (and the value) of peak of LAI has been also observed by Kouadio (2012).

Recalibrating parameters that are not able to adjust remotely-sensed and simulated phenology (e.g. joint recalibration on SPAN and TDWI parameters in *chapter 2*) can partly falsifies the interpretation of recalibrated parameters. Selection of the parameters for the recalibration must therefore pay attention to the equifinality issue which may lead to the loss of their biophysical meaning.

(3) Sources contributing to the uncertainty on remote sensing data and biophysical variables.

In addition to the effect of pixels' purity inducing an uncertainty on LAI and subsequently reducing the assimilation performances, this thesis has also assessed the influence of different factors (water content, biomass) on backscattering whose the uncertainty can limit the mowings detection based on backscatterring.

Similarly to crop sowing or crop emergence, mowing dates are management events to consider in grassland growth models notably with a view to adjust observed and simulated phenology (*chapter 3*). These mowing dates can be estimated by remote sensing either with optical data (e.g. Courault *et al*., 2010; Hadj Said *et al*.; 2011, Lips (2011) or from SAR data as shown in this thesis (*Chapters 3 and 4*).

Different factors (e.g. incidence angle, surface roughness and biomass) influence SAR backscattering similarly for grazing and mowing parcels. The impact of water content on backscattering has been more specifically assessed (*chapter 4*) based on a proxy (cumulated rainfall over the 72h preceding the SAR acquisitions).

Chapter 4 has allowed to confirm, as expected in regards to literature review, that grassland backscattering increases with the increase of water content and to assess that it decreases with the biomass at least in dry conditions.

The comparison of backscatterings of mowing parcels with the median backscattering of grazing parcels (whose fluctuations of biomass are supposed constant all along the growing season) has allowed stressing that soil backscattering becomes predominant in the total SAR backscattering when water content increases masking progressively the contribution of biomass.

This confusion of contributions of both soil and plants to the SAR response could hamper the classification of grassland parcels according to their land use / management type (mowing vs grazing) and subsequently detection of mowings if SAR acquisitions are acquired exclusively in high water content conditions.

A good overall classification accuracy (around 80%) was however obtained for the validation data set using a methodology based on the backscattering distribution of grazing parcels and so, despite a sub-optimal organization of field campaigns (non-coincidence between the days of acquisition and the days of field campaigns) and a relatively low revisit of ERS-2 (C-band sensor) images.

Two approaches have been defined in *Chapter 4* to detect mowings respectively looking at relative difference either in space or in time. These 2 approaches have been designed considering the effect of water content on backscattering. At best, only half of these mowings have been correctly identified. Among the two approaches, the one using a temporal approach provides the best results as it probably allows masking some of these factors. These non-optimal results are mainly induced by the sub-optimal organization of field campaigns but also due to some factors of uncertainty that have not been accounted in the study. The considered approach can be therefore considered as promising as rooms for improvement exist.

Perspectives

In order to further improve the current achievements, several ways of investigation can be identified.

Data assimilation can be seen as the crossroads of two worlds, the world of crop modelling and the world of remote sensing. Data assimilation techniques bind these 2 worlds, the objectives being most of time to improve outputs of the models. Neither (noisy) data nor (inaccurate) models can give a complete description of a system, but their optimal combination through data assimilation technique can indeed provide a more complete and coherent picture of the state of it, which is essential to understand, monitor and more importantly predict its evolution.

Improvements can therefore concern the different parts of the systems, i.e. either the model or the assimilated observation but also the assimilation techniques.

Improvements based on assimilated observations (LAI)

Improvement of assimilated observations can be reached by different ways. These ways of improvement vary according to the observations.

For a state variable such as the LAI considered in this PhD (*Chapters 1 and 2*), it is important to be sure that the remotely-sensed and simulated state variable strictly represent the same biophysical variable. The biophysical variable retrieved from satellite imagery usually referred as LAI is more close to a plant area index (PAI) (Duveiller, 2011) and subsequently more related to photosynthetically active elements of the canopy. Commonly used crop growth models ordinarily supposed to simulate the LAI of the crops (in its whole, not only of its canopy). A perspective of improvement could be to ensure a perfect concordance between assimilated observations and simulations either by adapting the radiative transfer models or any techniques used to retrieve the biophysical variables from remote sensing or by modifying the models.

In the hypothesis where assimilated observations and simulated state variables are strictly identical, as far as the objective is to retrieve crop specific information (yields) at a regional level, the notion pixel purity (as illustrated in this thesis) is very important. The challenge of discriminating pixels that correspond to a particular crop type, a prerequisite for crop specific agricultural monitoring, remains daunting when the signal encoded in pixels stems from several land uses (mixed pixels), e.g., over heterogeneous landscapes where individual fields are often smaller than individual pixels (Löw and Duveiller, 2014). As things are now, the acquisition of images presenting a temporal availability compatible with agriculture monitoring comes unfortunately with a coarser spatial resolution and its inherent problems of pixels purity in fragmented agricultural regions. Before the announced advent of sensors such as Sentinel missions, combining theoretically high spatial resolution and high temporal resolution, a selection of pixels according to their purity in regards to the target crop has to be performed. Selecting pixels reduces the number of available information within the considered region at which the crop specific yields have to be estimated. A very strict selection on purity could lead to loose a part the regional / spatial variability which could influence the accuracy of yields estimations at regional level. A balance has therefore to be found between the pixels' purity and spatial variability. In this context, a study could done to assess the best compromise between these two aspects allowing to reach a given level of accuracy on yields estimations at regional level. Results of this study could also have an indirect implication on the use of cloudy optical images. If to characterise a (homogeneous) region, only a few observations rigorously selected are needed to ensure improvement of simulations, the probability to exploit cloudy optical images is increased as usually the cloud cover do not concern all the target region. It stressed however that the crop yield estimations at a regional level should be done for homogeneous region in terms of growing conditions and agricultural practices, such as agricultural regions and therefore not for administrative divisions such as NUTS.

The tricky compromise between spatial and temporal resolutions as well as spatial coverage, partly already solved with sensors as AWiFS is close to being solved with the forthcoming missions and more specifically with the current trend towards the definition of virtual satellite constellation. A virtual satellite constellation is defined the Committee on Earth Observation Satellites (CEOS) as a "set of space and ground segment capabilities that operate in a coordinated manner to meet a combined and common set of Earth Observation requirements". Existing satellite data products could be significantly improved if these were not limited to individual sensors but would combine complementary platforms across space agencies and sensor types (Wulder *et al*., 2014). NASA and USGS are currently collaborating with ESA to pursue cross-calibration of Landsat-8 and Sentinel-2 sensors, and to investigate approaches for normalizing differences between the respective data products (including view & solar angle BRDF, and spectral band passes). The wider swath of Sentinel-2 will provide increased temporal coverage (Drusch et al. 2012), with a single Sentinel-2 reducing revisit times to 8-days from the current 16-day revisit cycle of Landsat-8 (Roy et al. 2014). With both Sentinel-2a and b in orbit the expected 5 day revisit in combination with Landsat-8 will provide for an observation approximately every 3 days. Reducing revisit time is crucial in some regions of the world where the cloud cover is important.

Independently from the effect linked to the use of higher spatial resolution images, uncertainty on remotely-sensed biophysical parameters such as LAI can be reduced through the improvement of retrieval algorithms. Algorithms based on neural networks (e.g. CYCLOPES algorithm) were commonly used to retrieve biophysical parameters. Though the robustness of these algorithms has been proved in various operational processing chains, they present some limitations (Rivera Caicedo, 2014). LAI derived from CYCLOPES is for example less accurate at higher values due to saturation effect in the radiative transfer simulation and the neural network inversion algorithm (Bacour et al., 2006; Weiss et al., 2007). They also tend to underperform for situations not considered in their training phase. Besides, training neural networks involve tuning several parameters that may greatly impact the robustness of the model. These algorithms are currently being replaced by more advanced, simpler to train algorithms such as machine learning regression algorithms (Camps-valls and Bruzonne, 2009). These algorithms are able to cope with the saturation effect and, using a Bayesian approaches, are able to provide in addition to high accuracies uncertainty intervals for the predictions which make them more suitable for operational applications (Rivera Caicedo, 2014).

Improvements based on assimilated observations (mowing dates)

In the frame of the estimation of model initial conditions such as mowing dates or in a lower extend as crop emergence or sowing date (more difficult to estimate with precision), the use of SAR images represent probably an advantage over optical images. SAR is indeed an attractive source of data as these sensors can acquire imagery regardless of the presence of clouds or the lack of sunlight. If the revisit time is very small (every 1-3 days for examples), as mowing occurs during dry

periods where the probability to have usable optical images is higher, this advantages is probably slightly reduced. However, even in dry conditions, the probability to have a cloud-free image over all the area of interest is very low. A recent publication (Whitcraft *et al.*, 2014) has showed that the early and mid-agricultural growing season, which are important periods for crop type area identification and crop yield forecasting, are characterized by both frequent and pervasive cloud extent in many parts of the world and that during these periods but also during other portions of the agricultural growing season less than half of 8 day composites would be 70% clear. The authors suggest that in these areas/ time periods, optical, polar-orbiting imaging is not likely to be viable option for operational monitoring and that alternatives such as SAR ought to be considered. When the objective is to assess yields at a regional level with a grasslands growth model, it is also important to know the distribution of the mowing dates within the considered region. SAR allow to a complete picture of this regions which is not necessarily the case with optical systems.

The use of SAR data to estimate mowing dates could be also improved by dissociating the backscattering in its different components. It could be indeed useful to dissociate the backscattering from the soil and the backscattering coming from the vegetation. The semi-empirical water-cloud model(Attema and Ulaby, 1978), defining the radar radiation backscattered from the vegetation canopy as the incoherent sum of reflected radiation from vegetation and the underlying soil following its attenuation by the vegetation layer, could be used in this context (after calibration) to estimate grass biomass.

As mowing operations occurs during relatively dry periods, meteorological information could be used to refine the estimation if the temporal availability of the remote sensing is not short enough. With the approach considered in this thesis, we can only define the period where the mowing operation occurs and each date within this time interval has the same probability to be the mowing date. By using meteorological data (rainfall data), the objective is to provide a probability density function of mowing occurrence for all the days included within the time interval and possibly reduce the uncertainty on mowing dates. This approach could be also used to limit false detection of mowings.

The current trend towards an improvement of temporal resolution for SAR sensors can also open the door to other techniques allowing monitoring the grassland biomass and subsequently of mowing dates as for example the interferometry. Radar interferometry is a technique that extracts three-dimensional information of the observed surface by using the phase content derived from a couple of images. The interferometric phase measures the path length difference between the target and the two sensor sensors locations from which a three-dimensional position of the images resolution elements (i.e. height maps is derived). The interferometric correlation, or coherence, measures the variance of the interferometric phase estimate. It decreases with increasing system noise, volume scattering and temporal changes and therefore contains thematic information (Blaes and Defourny,

2003).Different studies have already demonstrated the increased information added by SAR coherence to SAR monitoring over agricultural areas. Blaes and Defourny (2003) have for example shown by using the 1-day interval SAR acquisition capabilities from 2 identical C-band instruments (ERS1/2) that a high sensitivity of the coherence to the plant height and the canopy cover exists. Strong relationships between the plan height and the coherence for the 4 monitored crops (winter wheat, sugar beet, potato and maize): Coefficient of determination ranged between 0.64 for maize to 0.92 for winter wheat. The more the height of the crops was high, the lower the coherence. These results are confirmed by Barrett *et al.* (2012) who have observed that for the majority of cases, especially in C-band, the change in vegetation tended to be the predominant source of decorrelation (and therefore to the loss of coherence). A mowing event, modifying the height of the crops, would significantly change (increase) the coherence and would therefore allow to detect mowing operations. Coherence images with one-day interval are also expected to contribute significantly to crop discrimination (Bériaux, 2011). COSMO SkyMED and Tandem-X sensors could be used for this purpose.

For McNairn and Brisco (2004), an ideal SAR system for any agricultural application would have in general multi-frequency, multi-temporal, and multi-polarization capabilities, and, in addition, very high spatial resolution compared to the parcel size. ERS-2 sensor used in the thesis is a single frequency (C-band) and single polarization (VV) sensor with an average incident angle of 23° and a repeat cycle of 35 days. The best conditions to monitor grassland with SAR data were therefore not really met. The use of a dual (e.g. TerraSAR-X, Sentinel-1) or quad (e.g. RADARSAT-2) polarization sensors would provide more discrimination of biomass through a separation of the different scattering mechanisms. Polarimetric SAR is indeed sensitive to the structural properties of a sensed object. The well-known decomposition theorem (entropy/anisotropy/alpha method) developed by Cloude and Pottier (1997) initially developed for fully polarimetric data but later on extended to encompass dual polarimetric SAR measurements (Cloude, 2007) could be used for this purpose. Voormansik *et al.* (2013) have tested different parameters derived from TerraSAR-X to detect grassland cutting practices. Though none of these parameters were found to be sensitive to grass height, the authors have however found that it was possible to detect areas with freshly cut grass lying horizontally on ground on the basis of dual polarimetric dominant scattering alpha angle. The use of shallower incident angles (>40°) is also expected to enhance interaction with vegetation. Sentinel-1 characteristics meet most of the requirements for a better monitoring of grasslands.

Improvements based on assimilation techniques

Concerning assimilation techniques, this thesis have shown that assimilation techniques that don't allow synchronizing in time simulated and observed LAI time series limit or hamper the efficiency of these techniques. Further investigations should be focused on the development of assimilation techniques allowing this synchronization in addition / prior to any possible update of the state variable(s)

and/or the recalibration of model parameters. According to the confidence that we have on models and/or on assimilated observations, uncertainties on models and/or on assimilated observations should be considered in these assimilation techniques. For example, calibration-based assimilation techniques only apply if, in addition to a sufficient number of observations, the observation error is small (Dorigo *et al.*, 2007). Moreover, one of the main weaknesses of these techniques is that all errors from input, output and model structures are attributed exclusively to model parameters (Chen *et al*., 2008). As sources of uncertainties probably exist both in models and observations, developed assimilation techniques should consider these uncertainties. Bayesian approaches such as those developed by Moradkhani *et al.* (2005), Chen *et al.* (2008) or Evensen (2009) aiming at a dual state-parameter estimation using EnKf or the Simultaneous parameter Optimization and Data Assimilation (SODA) method combining the strengths of the parameter search efficiency and explorative capabilities of the Shuffled Complex Evolution Metropolis (SCEM-UA) algorithm (Vrugt *et al.*, 2003) with the power and computational efficiency of EnKf (Vrugt *et al*., 2005) seem to be, in this context, an interesting way to investigate.

Other considerations

This thesis also allows wondering on the relevance to assimilate observations in crop growth models in regions, such as Belgium, where the variability of observed yields at regional level is limited. The question is particularly relevant for arable crops such as e.g. winter wheat. Assimilating a remotely-sensed variable such as LAI only in case of 'extreme' weather conditions for which the models tends to be less efficient could be the only situation for which assimilation could significantly improve the quality of yields estimations in regions where the variability of the observed yields is low. This assumption needs however to be further investigated. For grasslands, assimilation of mowings calendar has certainly more sense considering the high variability of the different mowing dates (linked to meteorological conditions but also to some management factors) and the subsequent highest variability of the annual yields, sum of yields for the different mowings which are all uncertain.

This thesis has furthermore stressed the added value that would represent the redaction of standardized protocols describing when data collections should be organized as well as the observation to record during these data collections would greatly improve the quality of further data collection and analyses based on these collected data. The redaction of a best practices protocol for further field campaigns in grasslands to support researches with SAR, based on the experience learned from this thesis is currently in preparation. This protocol is intended to support the JECAM (Joint Experiment for Crop Assessment and Monitoring) initiative whose the overarching goal is to reach a convergence of approaches, develop monitoring and reporting protocols and best practices for a variety of global agricultural systems (www.jecam.org).

Sentinels missions and more globally the ESA Earth observation program for COPERNICUS represent assuredly a new era for earth observation but is also a link in the hyper connectivity of our world. This increasing digital interconnection of

people and things, anytime and anywhere, will have profound social, political and economic consequences, and increasingly form part of our everyday lives, from the cars we drive and medicines we take, to the jobs we do and the governance systems we live in. According to the World Economic Forum, there will be 50 billion networked devices by 2020. The accuracy of the Copernicus products and their free and open access will enhance the development of applications and services, notably in agriculture. Development of communication and information technology and its democratization has induced higher expectations from farmers (and industries) for products and services able to help them in the daily and at short-term management of their farms. They have now access to data that were previously most of time only available to governments. As users, they expect this information to be rapidly, easily available and understandable but has also reliable. Crop yields estimations and predictions (at regional but also field level) are one of the products usually expected by farmers and industry. This thesis contributes to this current development of our society by its objective to improve the reliability of yield estimations and predictions and is part of the main context of resources (fertilizers, manure, etc.) optimisation, ecologically intensive agriculture and feed autonomy.

Development of the information and communication technology will also completely change the interactions between information users and producers. Exchanges of information are made more and more easy, through the development of crowdsourcing for example. It will markedly speed up the critical learning process for the remote sensing providers thanks to input and near real time feedbacks from the users. It will also reduce the costs related to field data collection. Increasing the reliability of EO products and services will probably induce a snowball effect enhancing the interaction between information users and producers.

Earth Observation is therefore moving more and more to an operational mode. Governments should take part to this evolution in order to fulfil its public service tasks by providing objective and devoid of commercial purpose information to farmers and more generally to the agricultural sector.

LIST OF PUBLICATIONS

Publications in peer-reviewed journals

Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and forest meteorology* 151(12), 1843-1855.

Curnel, Y., Duveiller, G., De Wit, A.J.W, Oger, R., Defourny, P. (in preparation). Influence of pixels purity on assimilation performance. To be submitted to *Agricultural and forest meteorology*.

Curnel, Y., Delvaux, T., Oger, R., Defourny, P. (In preparation). Estimation of mowing dates with ERS-2. To be submitted to *International Journal of remote sensing*.

Curnel, Y., Jacques, D., Gierlinger, N., Pâques, L.E., 2006. Variation in the decay resistance of larch to fungi. *Ann. For.Sci*. 65 (810), 8p.

Curnel Y., Jacques, D., Nanson, A., 2003. The first multisite clonal test of wild cherry (*Prunus avium L.)* in Belgium.*Sylvae Genetica* 52, 2, 45-52.

El Jarroudi, M., Kouadio, L., Bertrand, M., **Curnel, Y.**, Giraud, F., Delfosse, P., Hoffmann, L., Oger, R., Tychon, B., 2012.Integrating the impact of wheat fungal diseases in the Belgian crop yield forecasting system (B-CYFS).*Eur. J. Agron.*40**:** 8-17

Gao, Y., Wang, Y., Yin, S., **Curnel, Y.**, 2006.The application of world food study model (WOFOST) for yield prediction in Heilongjiang province.*Chinese Journal of agrometeorology* 27 (1), 27-30.

Gao, Y., Wang Y., Yin, S, Li, X., Curnel Y., 2006.Application of crop growth monitoring system for estimating crop yield (traduction from Chinese).*Journal of Northeast agricultural University*, 37 (5), 706-713.

Jacques, D., Marchal, M., **Curnel, Y**., 2004.Relative efficiency of alternative methods to evaluate wood stiffness in the frame of hybrid larch (Larix x Eurolepis Henry) clonal selection.*Annals of forest science*, 61 (1), 35-43.

Menzel A., Sparks T.H., Estrella N., Koch E., Aasa A., Ahas R., Alm-kübler K., Bissolli P., Braslavská O., Briede A., Chmielewski F.M., Crepinsek Z., **Curnel Y**., Dahl A., Defila C., Donnelly A., Filella Y., Jatczak K., Mage F., Mestre A., Nordli O., Peñuelas J., Pirinen P., Remisová, Scheifinger H., Striz M., Susnik A., Van Vliet A.J.H., Wielgolaski F.-E., Zach S., Zust A., 2006. European phenologicalresponse to climate change matches the warming pattern. *Global Change Biology* 12, 1969–1976.

Yu, C., Wang, Y., Gao, Y., **Curnel, Y**., 2005. Application of STATCAT, the software of a integrate statistic tool called CGMS. *Heilongjiang meteorology* 4, 1-3.

Book chapter

Douglas, G.C., Pliura, A., Dufour, J., Mertens, P., Jacques, D., Fernandez-Manjares, J., Buiteveld, J., Parnuta, G., Tudoroiu, M., **Curnel, Y**., Thomasset, M., Jensen, V., Knudsen, M. A., Foffová, E., Chandelier, A.; Steenackers, M., 2013. Common Ash (*Fraxinus excelsior* L.). Book chapter, Forest tree breeding in Europe : current state-of-the-art and perspectives. ed. / Luc E. Pâques.Vol. II Springer, 403-462 (Managing Forest Ecosystems, Vol. 25).

Publications in Conference proceedings

Curnel, Y., Oger, R., 2006. Agrophenology indicators from remote sensing: state of the art. ISPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates, Stresa (Italy), 31-38.

Duveiller, G., Kouadio, L., **Curnel, Y**., Djaby, B., de Wit, A., Tychon, B., and Defourny, P. 2010.Exploring the potential of crop specific green area index time series to improve yield estimation at regional scale. In Sobrino, J. A., editor, *Proceedings of the 3rd In- ternational Symposium on Recent Advances in Quantitative Remote Sensing (RAQRS'III), September 27th to October 1st, 2010, Torrent (Valencia), Spain*. Universitat de Valencia, Servicio de Publicaciones.

Ghysel, F., **Curnel, Y**., Stilmant, D., Oger R. Modéliser les performances techniques, économiques et environnementales des systèmes agraires afin d'explorer les voies d'évolution possibles "OptiMAE". Carrefour des Productions animales, 13 : 112-114.

Ghysel, F., Curnel, Y. , Hennart, S. , Oger, R. , Decruyenaere, V., Stilmant, D. 2009. *Modéliser les systèmes bovins pour analyser l'impact des mesures agri-environnementales sur leurs performances.*In: Rencontres Recherches Ruminants, 93-96.

Jarroudi, M., Kouadio, L., Martin, B., Giraud, F., Delfosse, P., Hoffmann, L., **Curnel, Y.**, Tychon, B., 2010.*Modelling plant diseases impact with the Belgian Crop Growth Monitoring System.*In: Proceedings of Agro2010 The XIth ESA Congress, Montpellier, France, August 29th - September 3rd 2010, 2 pp.

Picard, I., Eerens, H., Oger, R., **Curnel, Y**., Tychon, B., Ozer, P., Diepen, K. Van, Boogaard, H.L., Genovese, G., Nègre, T., Wang, Y., Li, X., 2005.Use of SPOT-VEGETATION in different versions of the European crop growth monitoring system. In: Proceedings of the 2nd international VEGETATION user conference; 1998-2004: 6 years of operational activities. - Luxembourg : EC, 2005 - ISBN 92-894-9004-7, 411 - 420.

Contribution to a collective work

Demarée G.R. and **Curnel Y.**, 2008.Plant phenology in Belgium. *In:* The history and current status of plant phenology in Europe –COST Action 725: Establishing a European Data Platform for Climatological Purposes, J. Nekovar, E. Koch, E. Kubin, P. Nejedlik, , T.H. Sparks, F.-E. Wielgolaski (eds), ISBN 978-951-40-2091-9, ISBN 978-951-40-2091-9,

REFERENCES

Abrahamsen, P., Hansen, S., 2000. Daisy: an open soil–crop–atmosphere system model. Environ. *Model.Softw*. 15, 313–330.

Aggarwal, P. K., Kalra, N., 1994. Analyzing the limitations set by climatic factors, genotype, and water and nitrogen availability on productivity of wheat II. Climatically potential yields and management strategies. *Field Crops Research*38, *93-103.*

Aggarwal, P. K., Kalra, N., Singh, A. K., Singha, S. K., 1994. Analyzing the limitations set by climatic factors, genotype, and water and nitrogen availability on productivity of wheat I. The model description, parameterization and validation.*Field Crops Research 38, 73-91.*

Ahmad, S., Kalra, A., Stephen, H., 2010.Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources* 33(1), 69-80.

Alavi, N., Warland, J.S., Berg, A., 2009. Assimilation of soil moisture and temperature data into land surface models: A survey. *In: data assimaltion for atmospheric, Oceanic and hydrologic applications,* Springer-verlag, Berlin Heidelberg, 429-448.

Allen, C.T., Ulaby, F.T., 1984. Characterization of the microwave extinction properties of vegetation canopies.Ann Arbor, MI 48109, University of Michigan, College of engineering, Radiation laboratory.

Álvarez-Mozos, J., Casali, J., González-Audicana, M., Verhoest, N.E.C., 2005. Correlation between ground measured soil moisture with RADARSAT-1 derived backscattering coefficient over an agricultural catchment of Navarre (North of Spain). *Biosystems Engineering* 92 (1), 119-133.

Arnold, C.P. Jr., Dey, C.H., 1986. Observing-system simulation experiments: Past, present, and future. *Bull. Amer. Meteor. Soc.* 67, 687-695.

Atlas, R., 1997. Atmospheric observation and experiments to assess their usefulness in data assimilation. *J.Meteor.Soc.Jpn.*75,111-130.

Attema, E. P. W. and Ulaby, F. T., 1978. Vegetation modeled as a water cloud. *Radio Science*, *13*(2), 357-374.

Attema, E., Duchossois, G., Kohlhammer, G., 1998. ERS-1/2 SAR land applications: Overview and mainresults. *IEEE – geoscience and remote sensing symposium, IGARRS,* pp. 1796-1798.

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sens. Environ.* 93, 53–67.

Auquière, E., 2001. SAR temporal series interpretation and backscattering modelling for maize growth monitoring, PhD thesis, 227 pp, Université Catholique de Louvain, UCL-geomatics, Louvain-la-Neuve, Belgium.

Bacour, C., Baret, F., Béal, D., Weiss, M., pavageau, K., 2006. Neural network estimation of *LAI, fAPAR, fCover* and *LAIxC$_{ab}$* from top of canopy MERIS reflectance data: Principles and validation. *Remote Sens. Environ.*105, 313-325.

Balenzano, A, Mattia, F., Satalino, G., Davidson, M.W.J., 2011. Dense temporal series of C-band and L-band SAR data for soil moisture retrieval over agricultural crops. *IEEE journal of selected topics in applied earth observations and remote sensing* 4(2), 439-450.

Bally, P., Fellah, K., 1995. Evaluation of the accuracy of the backscattering coefficient measurement in SAR data products. ESA, Earth Sciences Division, Technical note, July 1995, 36p.

Baret F., 2000. ReSeDa, Assimilation of multisensor and multitemporal remote sensing data to monitor soil & vegetation functioning. Final report, 59p.

Baret, F., Jacquemoud, S., Guyot, G., Leprieur, C., 1992. Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sens. Environ.*41 (2–3), 133–142.

Baret, F., Vintila, R., Lazar, C., Rochdi, N., Prévot, L., Favard, J.-C., de Boissezon, H., Lauvernet, C., Petcu, E., Petcu, G., Voicu, P., Denux, J.-P., Poenaru, V., Marloie, O., Simota, C., Radnea, C., Turnea, D., Cabot, F., Henry, P., 2001. The ADAM database and its potential to investigate high temporal sampling acquisition at high spatial resolution for the monitoring of agricultural crops. *Roman. Agric. Res.* 16, 69–80.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Nino, F., Weiss, M., Samain, O., Roujean, J. L., Leroy, M., 2007. LAI, fAPAR and fCOVER cyclopes global products derived from vegetation: Part 1: Principles of the algorithm. *Remote Sens. Environ.*110(3), 275–286.

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

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., Smets, B., 2013. GEOV1: LAI and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Principles of development and production. *Remote Sensing of Environment* 137, 299-309.

Baron, C., Sultan, B., Balme, M., Sarr, B., Traoré, S., Lebel, T., Janicot, S., Dingkuhn, M., 2005. From GCM grid cell to agricultural plot: scale issues

affecting modeling of climate impact. *Phil. Trans. R. Soc. Lond.*B 360, 2095–2108.

Barrett, P.D., Laidlaw, A.S., Mayne, C.S., 2005. GrazeGro: a European herbage growth model to predict pasture production in perennial ryegrass swards for decision support. *Europ. J. Agronomy* 23, 37–56

Barrett, B., Whelan, P., Dwyer, N., 2012. The use of C- and L-band repeat pass interferometric SAR coherence for soil moisture change detection in vegetated areas. *The open remote sensingJournal* 5, 37-53.

Barrett, B., Nitze, I., Green, S., Cawkwell, F., 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sensing of Environment* 152, 109-124.

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

Beeri, O., Phillips, R., Hendrickson, J., Frank, A.B., Kronberg, S., 2007. Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie. *Remote Sens. Environ.*110, 216-225.

Benallegue, M., Taconet, O., Vidal-Madjar, D., Normand, M., 1995. The use of radar backscattering signals for measuring soil moisture and surface roughness. *Remote Sens. Environ.*53, 61-68.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. *Remote sensing of Environment* 120, 84-90

Bériaux, E., 2011. Leaf Area Index retrieval from multi-annual and multi-polarization SAR time series for crop monitoring.PhD thesis, UCL, 225 p.

Bériaux, E., Lambot, S., Defourny, P., 2011. Estimating surface-soil moisture for retrieving maize leaf-area index from SAR data. *Journal canadien de télédétection*, 37(1), 136-150.

Berntsen, J., Petersen, B.M., Jacobsen, B.H., Olesen, J.E., Hutchings, N.J., 2003. Evaluating nitrogen taxation scenarios using the dynamic whole farm simulation model FASSET. *Agric. Syst.* 76, 817–839.

Bert, F.E., Laciana, C.E., podestà, G.P., Satorre, E.H., Menéndez, A.N., 2007. Sensitivity of CERES-maize simulated yields to uncertainty in soil properties and daily solar radiation. *Agricultural systems* 94, 141-150.

Bingfang, W., 2006.Introduction of China Crop Watch System with remote sensing. ISPRS Archives XXXVI-8/W48. Workshop proceedings: Remote sensing support to crop yield forecast and area estimates. 30th November-1st December 2006, Stresa (Italy), 15-18.

Blaes, X., Defourny, P., 2003. Retrieving crop parameters based on tandem ERS 1/2 interferometric coherence images. *Remote Sens. Environ.*88, 374-383.

Blaes, X., Defourny, P., Wegmüller, U., Della Vecchia, A., Guerriero, L., Ferrazzoli, P., 2006. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 4, 2006, p. 791-800.

Boatwright, G., Whitebread, V., 1986. Early warning and crop condition assessment Research. Geoscience and Remote Sensing, *IEEE transactions on Geoscience and Remote Sensing* (ISSN 0196-2892); GE-24, 54-64

Boogaard, H.L., van Diepen, C.A., Rötter, R.P., Cabrera, J.M.C.A., van Laar, H.H., 1998.WOFOST 7.1.User's guide for the Wofost 7.Crop Growth Simulation model and WOFOST control centre 1.5.52.DLO Winand staring centre, Wageningen, 142 p.

Boogaard, H.L., Eerens, H., Supit, I., van Diepen, C.A., Piccard, I., Kempeneers, P., 2002. Description of the MARS Crop Yield Forecasting System (MCYFS). Joint Research Centre, Study contract number 19226-2002-02-F1FED ISP NL.

Bouman, B. A., 1991.Crop parameter estimation from ground-based xband (3-cm wave) radar backscattering data. *Remote Sens. Environ. 37*(3), 193-205.

Bouman, B. A., 1992. Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. *International Journal of Remote Sensing* 13(14), 2564-2581.

Bouman, B.A.M., 1995. Crop modelling and remote-sensing for yield prediction. *Netherlands J. Agricult. Sci.* 43(2), 143-161.

Bouman, B.A.M., van Keulen H., van Laar, H.H., Rabbinge R., 1996a. The 'School of de Wit' crop growth simulations models: A pedigree and historical overview. *Agricultural Systems* 52 (2/3), 171-198.

Bouman, B.A.M., Schapendonk, A.H.C.M., Stol W., van Kraalingen, D.W.G, 1996b. Description of the grassland growth model LINGRA as implemented in CGMS. AB-DLO, 56 p.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* 18, 311–346.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillere, J.P., Hénault C., Maraux, F., Seguin B., Sinoquet H., 2003. An overview of the crop model STICS. *Europ. J. Agronomy* 18, 300-306.

Brown, S.C.M., Quegan, S., Morrison, k., Bennett, J.C., cookmartin, G., 2003.High-resolution measurements of scattering in wheat canopies-implications for crop parameter retrieval. *IEEE transactions on Geoscience and Remote sensing* 41(7), 1602-1610.

Bsaibes, A., Courault, D., Baret, F., Weiss, M., Olioso, A., Jacob, F., Hagolle, O., Marloie, O., Bertrand, N., Desfond, V., Kzemipour, F., 2009. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. *Remote Sens. Environ.* 113, 716-729.

Buermann, W., Dong, J., Zeng, X., Myneni, R.B., Dickinson, R.E., 2001. Evaluation of the utility of satellite-based vegetation Leaf Area Index for climate simulations. *Journal of Climate* 14, 3536-3550.

Buemi, M.E., Frery, A.C., Ramos, H.S., 2014. Speckle reduction with adaptative filters. *Pattern Recognition Letters* 36, 281-287.

Buffet, D., Curnel, Y., Goffart, J.P., Vanwindekens, F., Van Den Wyngaert, L., 2010. Analyse des méthodes d'intégration des techniques de modélisation et de l'information satellitaire multicapteurs dans les systèmes d'aide à la décision. Projet MIMOSA, rapport intermédiaire, septembre 2010, 112 p + annexes

Burgers, G., jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.* 126 (6), 1719–1724.

Butler, D., 2014. Earth observation enters next phase. *Nature* 508, 160-161. doi:10.1038/508160a

Camps-Valls, G., Bruzzone, L., 2009. Kernel methods for Remote Sensing Data Analysis.Wiley & Sons, UK, 434 p.

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., Baret, F., 2012. Forcing a wheat crop model with LAI data to access agronomic variables : Evaluation of the impact of model and LAI uncertainties and comparison with an empirical model. *Europ. J. Agronomy* 37, 1-10.

Challinor, A. J., Wheeler, T.R., Craufurd, P.Q., Slingo, J.M., Grimes, D.I.F., 2004.Design and optimisation of a large-area process-based model for annual crops. *Agricultural and Forest Meteorology*, 124, (1-2) 99-120.

Chappelle, E.W., Kim, M.S., McMurtrey, I., James, E., 1992. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. *Remote Sens. Environ*. 39 (3), 239–247.

Chatfield, C., 1995. Model uncertainty, data mining and statistical-inference.*J. Roy. Stat. Soc. A Sta.* 158, 419–466.

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

Chen, M., Liu, S., Tieszen, L.L., Hollinger, D.Y., 2008. An improved state-parameter analysis of ecosystem models using data assimilation. *Ecological modelling* 219, 317-326.

Chen, Y., Oliver, D.S., Zhiang, D., 2009. Data assimilation for nonlinear problems by Ensemble Kalman Filter with reparameterization. *J.Petrol. Sci. Eng.* 66, 1–14.

Cicek, H., Sunohara, M., Wilkes, G.A., McNairn, H., Pick, F., Topp, E., Lapen, D.R., 2010. Using vegetation indices from satellite remote sensing to assess corn and soybean response to controlled tile drainage. *Agricultural Water Management 98*(2), 261-270.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.-F., Huc., M., Keravec, P., Béziat, P., Fieuzal, R., Ceschia, E., Dedieu, G., 2012. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote sensing of Environment* 124, 844-857.

Clevers, J.G.P.W., 1988. The derivation of a simplified reflectance model for the estimation of Leaf Area Index. *Remote Sens. Environ.* 25, 53-69.

Clevers, J.G.P.W., 1989. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sens. Envir*. 29(1), 25-37.

Clevers, J.G.P.W., van Leeuwen, H.J.C., 1996. Combined use of optical and microwave remote sensing data for crop growth monitoring. *Remote sens. Environ*. 56, 42-51.

Clevers, J.G.P.W., 1997. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sens. Envir*.61 (2), 221-228.

Clevers, J.G.P.W. Jongschaap, R.E.E., 2001. Imaging spectrometry for agricultural applications.*In:* Imaging spectrometry: basic principles and prospective applications. F.D. van der Meer & S.M. de Jong. - Dordrecht : Kluwer Academic Publishers, pp. 157 - 199.

Cloude, S., Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing* 35(1), 68–78.

Cloude, S., 2007. The dual polarization entropy/alpha decomposition: a PALSAR case study. In *PolInSAR2007 Workshop*, Frascati, 2007. Council of the European Community, 1988. Council decision of 26 September 1988 adopting a pilot project on remote sensing applied to agricultural statistics. *Official Journal of the European Union* L273, 12-17.

Courault, D., Hadria, R., Ruget, F., Olioso, A., Duchemin, B., Hagolle, O., Dedieu, G., 2010. Combined use of FORMOSAT-2 images with a crop model for

biomass and water monitoring of permanent grassland in Mediterranean region. *Hydrol.Earth Syst. Sci*. 14, 1731-1744.

COSW (2011). Carte d'Occupation du Sol de Wallonie, Version 2.07.

CPDT, 2011. Diagnostic territorial de la Wallonie 2011. Service Public de Wallonie, 287 p.

Cracknell, A. 1998. Synergy in remote sensing–What's in a pixel? *International Journal of Remote Sensing* 19(11), 2025–2047.

Curnel, Y.; 2014. Basic principles of Remote sensing. Technical note – U11 (vol. 1) – 48 p. http://cra.wallonie.be/document/download/613

Curnel, Y. and Oger, R., 2004. Influence of phenological uncertainty on B-CGMS winter wheat yield predictions: results of a sensitivity analysis. Paper prepared in the frame of the 2nd European scientific meeting on the Crop Growth Monitoring System (CGMS), Prague, Czech Republic, 8th October 2004, 7pp.

Curnel, Y. and Oger, R., 2006a. Conception d'un système combiné de prévision des rendements et d'avertissements phytosanitaires en Région Wallonne pour la culture du blé: Développement d'un module « maladie » pour B-CGMS. Internal report, ULG 05/1224297 convention, 11pp.

Curnel, Y. and Oger, R., 2006b. Agrophenology indicators from remote sensing: state of the art. ISPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates, Stresa (Italy), 31-38.

Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and forest meteorology* 151(12), 1843-1855.

Curtis, 2002. Wheat in the world *In:* Bread wheat, improvement and production. FAO Plant Production and Protection Series 30, 565 p.

Dabrowska-Zielinska, K., Inoue, Y., Kowalik, W., Gruszczynska, M., 2007. Inferring the effect of plant and soil variables on C- and L-band SAR backscatter over agricultural fields, based on model analysis. *Space Research* 39, 139-148.

Darvishzadeh, R., Skidmore, A.K., Schlerf, M., Atzberger, C.G., Cho, M.A., 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS journal of photogrammetry and remote sensing* 63 (4), 409-426.

Defourny, P., Blaes, X., Bogaert, P., 2007. Respective contribution of yield and area estimates to the error in crop production forecasting, Proceedings of the ISPRS Archives XXXVI-8/W48 Workshop: Remote sensing support to crop yield forecast and area estimates.

De Keyser, E., Vernieuwe, H., Lievens, H., Álvarez-Mozos, J., De Baets, B., Verhoest, N.E.C., 2012. Assessment of SAR-retrieved soil moisture uncertainty induced by uncertainty on modeled soil surface roughness. *International Journal of Applied Earth Observation and Geoinformation* 18, 176-182.

Della vecchia,A., Ferrazzoli, P., Guerriero, L.,Ninivaggi, L., Strozi, T., Wegmüller,U., 2008.Observing and modeling multifrequency scattering of maize during the whole growth cycle. *IEEE Transactions on Geoscience and Remote Sensing* 6(11), 3709−3718.

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

Delegido, J., Verrelst, J., Rivera, J.P., Ruiz-Verdú, A., Moreno, J., 2015.Brown and green LAI mapping through spectral indices. *International Journal of Applied Earth Observation and Geoinformation* 35, 350-358.

De Leeuw, M.K., Tavares de Carvalho, L.M., 2009. Performance evaluation of several adaptive speckle filters for SAR imaging. Anais XIV simpósio Brasileiro de sensoriamento remoto, Natal (Brasil), INPE, 7299-7305.

Deng, F., Chen, J., Plummer, S., Chen, M., and Pisek, J. 2006. Algorithm for global Leaf Area Index Retrieval Using Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 44(8), 2219–2229.

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

De Roo, R. D., Du, Y., Ulaby, F. T., & Dobson, M. C. (2001). A semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing* 39(4), 864 872.

De Wit, C.T., 1982. Simulation of living systems. *In* Simulation of plant growth and crop production, eds F.W.T. Penning de Vries & H.H. van Laar. Simulation monographs, PUDOC, Wageningen, The Netherlands, 3-8.

De Wit, A.J.W., Boogaard, H.L., Van Diepen, C.A., 2005. Spatial resolution of precipitation and radiation: the effect on regional crop yield forecasts. *Agr.Forest Meteorol.* 135, 156–168.

De Wit, 2007. Regional crop yield forecasting using probabilistic crop growth modeling and remote sensing data assimilation. PhD thesis, Alterra scientific contributions 20, 154 p.

De Wit, A.J.W. and van Diepen, C.A., 2007. Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest meteorology* 146, 38-56.

De wit, A.J.W., Dabrowska-Zielinska, K., van Diepen, C.A., 2008. Preface of Modern Methods in Crop Yield Forecasting and Crop Area Estimation issue. *International Journal of Applied Earth Observation and Geoinformation* 10, 401-402.

De Wit, A., Van Diepen, K., 2008. Crop growth modeling and crop yield forecasting using satellite-derived. *International Journal of Applied Earth Observation and Geoinformation* 10, 414-425.

De Wit, A.J.W., van Diepen, C.A., Boogaard, H.L., 2009. Research activities in regional crop modelling and yield forecasting. *Agro Informatica* 22 (2), 25 - 27.

De wit, A.J.W., Duveiller, G., Defourny, P., 2012. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agricultural and Forest Meteorology* 164, 39 - 52.

Di Bella, C., Faivre, R., Ruget, F., Seguin, B., 2005. Using VEGETATION satellite data and the crop model STICS-prairie to estimate pasture production at the national level in France. *Physics and Chemistry of the Earth* 30, 3-9.

Diepen, C.A., van Wolf, J., van Keulen, H., 1989. WOFOST: a simulation model of crop production. *Soil Use Manage.* 5, 16-24.

Diepen van, K. and Boogaard, H., 2009. History of CGMS in the MARS project. *Agro-informatica* 22 (2), 11-14.

Djaby, B., de Wit, A., Tychon, B., 2009. Influence of regional distribution of crop parameters on the winter wheat yield prediction performance of a spatially distributed crop model. Stereo II project SR/00/101 – GLOBAM 2nd activity report, 213-235.

Donatelli, M., Russell, G., Rizzoli, A.E., Acutis, M., Adam, M., Athanasiadis, I.N., *et al.*, 2010.A component-based framework for simulating agricultural production and externalities. *In:* Brouwer, F.M., van Ittersum, M.K. (Eds.), Environmental and Agricultural Modeling-integrated Approaches for Policy Impact Assessment. Springer, pp. 63–108.

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

Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote sensing* 69, 665-674.

Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using landsat and MODIS. *Remote Sens. Envir.*92, 548-559.

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

Dumont, B., Vancutsem, F., Seutin, B., Bodson, B., Destain, J.-P., Destain, M.-F., 2012. Simulation de la croissance du blé à l'aide de modèles écophysiologiques: Synthèse bibliographique des methods, potentialities et limitations. *BASE* 16 (3), 382-392.

Durand, J.M., Gimonet, B.J., Perbos, J.R., 1987. Speckle in SAR images: an evaluation of filtering techniques. *Advances in Space* Research 7(11), 301-304.

Dusseux, P., Gong, X., Corpetti, T., Hubert-Moy, L., Corgne, S., 2013. Contribution of radar images for grassland management identification. Proc. Of SPIE vol. 8531, 853104-1-7.

Dusseux, P., Corpetti, T., Hubert-Moy, L., Corgne, S., 2014. Combined use of multi-temporal optical and radar satellite images for grassland monitoring. *Remote Sens.* 6, 6163-6182. Doi:10.3390/rs6076163.

Duveiller, G., 2011. Crop specific green area index retrieval from multi-scale remote sensing for agricultural monitoring.PhD thesis, Université catholique de Louvain (UCL), 181 p.

Duveiller, G. and Defourny, P., 2010. A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sens. Envir.*114, 2637-2650.

Duveiller, G., Kouadio, L., Curnel, Y., Djaby, B., de Wit, A., Tychon, B., Defourny, P., 2010. Exploring the potential of crop specific green area index time series to improve yield estimation at regional scale. In Sobrino, J. A., editor, *Proceedings of the 3rd International Symposium on Recent Advances in Quantitative Remote Sensing (RAQRS'III), September 27th to October 1st, 2010, Torrent (Valencia), Spain*. Universitat de Valencia, Servicio de Publicaciones.

Duveiller, G., Baret, F., Defourny, P., 2011a. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. *Remote Sens. Envir.*115, 2686-2701.

Duveiller, G., Weiss, M., Baret, F., Defourny, P., 2011b.Retrieving wheat Green Area Index during the growing season from optical time series measurements based on neural network radiative transfer inversion. *Remote Sens. Envir*. 115, 887-896.

Duveiller, G., Baret, F., Dfourny, P., 2012.Remotely sensed green area index for winter wheat crop monitoring: 10-year assessment at regional scale over a

fragmented landscape. *Agricultural and forest meteorology* 166-167, 156-168.

Duveiller, G., López-Lozano, R., Seguini, L., Bojanowski, J.S., Baruth, B., 2013. Optical remote sensing requirements for operational crop monitoring and yield forecasting in Europe. Proceedings 'Sentinel-3 OLCI/SLSTR and MERIS/(A)ATSR workshop', Frascati, Italy, 8p.

Easterling, W.E., Weiss, A., Hays, C.J., Mearns, L.O.,1998. Spatial scales of climate information for simulating wheat and maize productivity : the case of the US Great plains. *Agricultural and forest meteorology* 90, 51-63.

Erickson, J.D., 1984. The LACIE experiment in satellite aided monitoring of global crop production. In *The role of terrestrial vegetation in the Global Carbon Cycle: Measurements by remote sensing*, edited by G.M. Woodwell (Chichester : Wiley), 191-217.

ESA, 2012, agricultural monitoring, fact sheetBR-128/II**,** 60p.

European Commission (EC), 2013. Decision No 529/2013/EU of the European Parliament and of the Council of 21 May 2013 on accounting rules on greenhouse gas emissions and removals resulting from activities relating to land use, land-use change and forestry and on information concerning actions relating to those activities, Official Journal of the European Union (2013) 5620138097

EUROSTAT, 2010. Agricultural statistics, main results 2008-2009.Eurostat pocketbooks, ISBN 978-92-79-15246-7, 186p.

Evensen, G.E., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn*. 53, 343-367.

Evensen, G.E., 2007. Data assimilation, the Ensemble Kalman Filter. Springer, 279p.

Evensen, G.E., 2009. The Ensemble Kalman Filter for combined state and parameter estimation. *Control systems* 29(3), 83-104.

Fang, H., Wei, S., Liang, S., 2012a. Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sens. Envir.*119, 43–54

Fang, H., Wei, S., Jiang, C., Scipal, K., 2012b. Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sens. Envir.*124, 610-621.

Fang, B. and Lakshmi, V., 2014. Soil moisture at watershed scale: Remote sensing techniques. *Journal of hydrology* 516, 258-272.

FAO, 2011a. The state of food insecurity in the world – How does international price volatility affect domestic economies and food security. 55p.

FAO, 2011b.The state of the world's land and water resources for food and agriculture (SOLAW), managing systems at risk.Food and Agriculture Organization of the United Nations, Rome and Earth scan, London, 308 p.

FAOSTAT, 2009.Statistical Database 2007. Rome.

Ferrazzoli, P., Paloscia, S., Pampaloni, P., Schiavon, G., Solimini, D., Coppo, P. 1992. Sensitivity of microwave measurements to vegetation biomass and soil moisture content: a case study. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, pp. 750–756.

Ferrazzoli, P., Paloscia, S., Pampaloni, P., Schiavon, G., Sigismondi, S., Solimini, D. 1997. The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 35, pp. 5–17.

Finnigan, C., 2013. Developing a grassland biomass monitoring tool using a time series of dual polarimetric SAR and optical data. Thesis, University of Saskatchewan (Saskatoon), 115p.

Fontaine, K., Neri, P., Defourny, P., Hanin, Y., 2011. L'occupation du sol en Wallonie. *In :* Diagnostic territorial de la Wallonie 2011, Ed. Ghislain Géron, Conférence Permanente du Développement Territorial (RW). Dépôt légal : D/2011/11802/59, page 251-260

Freeman, A., Durden, S., 1998.A three-component scattering model for polarimetric SAR data.*IEEE transactions on geoscience and remote sensing* 36, 963-973.

Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sens. Environ.* 105 (4), 286–298.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T.,Weiss, M., Nickeson, J. E.,Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y., Yang, W., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research-biogeosciences* 113 (G2):G02028.

Genovese, G.P., 1998. The methodology, the results and the evaluation of the MARS crop yield forecasting system. In: Rijks, D., Terres, J.M., Vossen, P. (Eds.), Agrometeorological Applications for Regional Crop Monitoring and Production Assessment. Office for Official Publications of the EU, Luxembourg, 67–119, EUR 17735 EN.

Gherboudj, I., Magagi, R., Berg, A.A., Toth, B., 2010. Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. *Remote Sens. Envir.*115(1), 33-43.

Gherboudj, I., Magagi, R., Berg, A.A., Toth, B., 2011. Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. *Remote Sensing of Environment* 115 (1), 33-43.

Giustarini, L., Vernieuwe, H., Verwaeren, J., Chini, M., Hostache, R., Matgen, P., Verhoest, N.E.C., De Baets, B. 2015. Accounting for image uncertainty in SAR-based flood mapping. *International Journal of applied Earth Observation and Geoinformation* 34, 70-77.

Gobron, N., 2008. Leaf Area Index (LAI) *in* Terrestrial essential climate variables for climate change assessment, mitigation and adaptation, GTOS 52, FAO, 44 p.

Gobron, N. and Verstraete, M.M., 2009. Assessment of the status of the development of the standards for the terrestrial essential climate variables, Leaf Area Index (LAI), report version 10, 22 p.

Gómez, M., Olioso, A., Sobrino, J.A., Jacob, F., 2005. Retrieval of evapotranspiration over the Alpilles/ReSeDA experimental site using airborne POLDER sensor and a thermal camera. *Remote Sens. Envir.*96, 399-408.

Gómez-Chova, L., Zurita-Milla, R., Alonso, L., Amorós-López, J., Guanter, L., Camps-Valls, G., 2011. Gridding artifacts on medium-resolution satellite image time series: MERIS case study. *IEEE Transactions on Geoscience and Remote Sensing* 49, 2601-2611.

Graham, A.J. and Harris, R., 2003. Extracting biophysical parameters from remotely sensed radar data: a review of the water cloud model. *Progress in Physical Geography* 27(2), 217–229

Guo, R., Lin, Z., Mo, X., Yang, C., 2010. Responses of crop yield and water use efficiency to climate change in the North China Plain. *Agricultural Water Management*97, 1185-1194.

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

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

Hadj Said, M., Laghrouche, M., Malek, M., Hadria R., Olioso, A., Courault, D., Ameur, S., 2011. Detection supervisee des dates des opérations de fauche des prairies irriguées dans la plaine de la Crau, France. *Télédétections* 10 (2-3), 129-136.

Hajkowicz, S., Negra, C., Barnett, P., Clark, M., Harch, B., Keating, B., 2012. Food price volatility and hunger alleviation – Can Cannes work ? Agricultural and food security 1:8, 12p.

Hansen, S., 2000.Daisy, a Flexible Soil–Plant–Atmosphere System Model. Equation Section 1. The Royal Veterinary and Agricultural University, Copenhagen, p. 47.

Hansen J.W. and Jones J.W., 2000. Scaling-up crop models for climate variability applications. *Agric. Syst.* 65 (1), 43–72.

Hansen, J., Jensen, H.E., Nielsen, N.E., Svendsen, H., 1990. DAISY—A Soil Plant System Model. Danish Simulation Model for Transformation and Transport of Energy and Matter in the Soil–Plant–Atmosphere System. National Agency for Environmental Protection, Copenhagen.

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

Harris, P.S., 2000. Grassland resource for pastoral systems. FAO plant production and protection paper 162, 166p.,ISBN92-5-104537-2

Hoogenboom, G., Jones, J.W., Porter, C.H., Wilkens, P.W., Boote, K.J., Batchelor, W.D., Hunt, L.A., Tsuji, G.Y., 2003. Decision Support System for Agrotechnology Transfer Version 4.0. Volume 1: Overview. University of Hawaii, Honolulu, HI.

Henderson, F. M. and Lewis, A. J., 1998. Introduction, in *Principles & Applications of imaging radar.Manual of Remote Sensing. Third Edition*, edited by F. Henderson, M. and Lewis, A., J., pp. 1-7.

Hermann, A., Kelm, M., Kornher, A., Taube, F., 2005.Performance of grassland under different cutting regimes as affected by sward composition, nitrogen input, soil conditions and weather – a simulation study. *Europ. J. Agronomy* 22, 141-158.

Herold, M., Hajnsek, I., Schmullius, C., 2000. Multifrequency and Polarimetric Radar Remote Sensing of Grassland - geo-biophysical Parameter Retrieval with Airborne E-SAR Data, in Buchroithner, M. F., 2001. A Decade of Trans-European Remote Sensing Cooperation, Proceedings of the 20th EARSeL Sympsium Remote Sensing in the 22st Century: A Decade of Trans-European Remote Sensing Cooperation, 14 - 16 June 2000, Dresden, Germany, 95-102.

Hill, M.J., Donald, G.E., Vickery, P.J., 1999. Relating radar backscatter to biophysical properties of temperate perennial grassland. *Remote Sens. Environ.* 67,15-31.

Hill, M., Ticehurst, C.J., Lee, J.-S., Grunes, M.R., Donald, G.E., Henry, D.,2005. Integration of optical and radar classification for mapping pasture type in Western Autralia. *IEEE transactions on geoscience and remote sensing* 43, 1665-1681.

Hoges, T., 1990.Predicting crop phenology.CRC press, 248 p.

Höglind, M., Schapendonk, A.H.C.M., Van Oijen, M., 2001. Timothy growth in Scandinavia: combining quantitative information and simulation modelling. *New Phytol.*151, 355–367.

Hong, G., Zhang, A., Zhou, F., Brisco, B., 2014. Integration of optical and synthetic aperture radar (SAR) images to differentiate grassland and alfalfa in prairie area. *International Journal of Applied Earth observation and geoinformation* 28, 12-19.

Huang, S.Q. and Liu, D.Z., 2007. Some uncertain factor analysis and improvement in spaceborne synthetic aperture radar imaging. *Signal processing* 87, 3202-3217.

Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.* 93 (1–2), 18–29.

Huang, Y. and van Genderen, J.L., 1996. Evaluation of several speckle filtering techniques for ERS-1&2 imagery. *International archives of photogrammetry and remote sensing* vol. XXXI, part B2, 164-169.

Hutchinson, C.F., 1991. Use of satellite data for famine early warning in sub-Saharan Africa. *Int. J. Remote Sens.* 12, 1405-1421.

Inoue, Y., Kurosu, T., Maeno, H., Uratsuka, S., Kozu, T., Dabrowska- Zielinska, K., Qi, J., 2002. Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biologicalvariables, *Remote Sens. Environ*. *81*(2-3), 194-204.

ICTSD, 2013. Agriculture and food security group, proposals and analysis, Geneva, Switzerland, 88 p.

IPCC, 2012. Managing the risks of extreme events and disasters to advance climate change adaptation –Special report of the intergovernmental panel on climate change, 594p.

Jacquemoud, S., Verdebout, J., Schmuck, G., Andreoli, G., Hosgood, B., 1995. Investigation of leaf biochemistry by statistics. *Remote Sens. Environ*. 54 (3), 180–188.

Jame, Y.W., Cutforth, H.W., 1996. Crop growth models for decision support systems. *Can. J. Plant Sci.* 76, 9-19.

Jarlan, L., Mangiarotti, S., Mougin, E., Mazzega, P., Hiernaux, P., Le Dantec, V., 2008. Assimilation of SPOT/VEGETATION NDVI into a sahelian vegetation dynamics model. *Remote Sens. Environ*.112, 1381-1394.

Jégo, G., Pattey, E., Liu, J., 2012. Using Leaf Area Index, retrieved from optical imagery, in the STICS crop model for predicting yield and biomass of field crops. *Fields crop research*131, 63-74.

Jensen, J.R., 1983. Biophysical remote sensing. *Annals of the Association of American Geographers* 73 (1), 111-132.

Jiao, X., McNairn, H., Shang, J., Liu, J., 2010. The sensitivity of multi-frequency (X, C and L-band) radar backscatter signatures to bio-physical variables (LAI) over corn and soybean fields. *In.:* Wagner W., Székely, B. (eds.): ISPRS TC VII Symposium – 100 years ISPRS, Vienna, Austria, July 5-7, IAPRS, vol. XXXVIII, part 7B., 317-321.

Jin, H., Eklundh, L., 2014. A physically based vegetation index for improved monitoring of plant phenology. *Remote Sensing of Environment* 152, 512-525.

Johnson, S.G., 2011. The NLopt Nonlinear-optimization Package, http://ab-initio.mit.edu/nlopt.

Jones, D. and Barnes, E.M., 2000. Fuzzy composite programming to combine remote sensing and crop models for decision support in precision crop management. *Agricultural Systems* 65:137-158

Jones, J.W., Hoogenboom G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. *Europ. J. Agronomy* 18, 235-265.

Joseph, A.T., van de Velde, R., O'Neill, P.E., Lang, R., Gish, T., 2010. Effects of corn on C- and L-band radar backscatter : A correction method for soil moisture retrieval. *Remote sensing of Environment* 114, 2417-2430.

Justice, C., and Becker-Reshef, I.E., 2007.Report from the Workshop on developing a strategy for Global Agricultural Monitoring in the framework of Group on Earth Observations (GEO), *UN FAO Report.*

Kang, Y., Khan, S., Ma, X., 2009. Climate change impacts on crop yield, crop water productivity and food security – A review. *Progress in Natural Science* 19, 1665-1674.

Katz, R.W., 2002. Techniques for estimating uncertainty in climate change scenarios and impact studies. *Clim. Res*. 20, 167-185.

Kersebaum, K.C., 1995. Application of a simple management model to simulate water and nitrogen dynamics. *Ecol.Model.* 81, 145–156.

Kersebaum, K.C., 2007. Modelling nitrogen dynamics in soil–crop systems with HERMES. *Nutr. Cycl. Agroecosyst.* 77, 39–52.

Kersebaum, K.C. and Beblik, A.J., 2001. Performance of a nitrogen dynamics model applied to evaluate agricultural management practices. In: Shaffer, M.J., Ma, L., Hansen, S. (Eds.), Modeling Carbon and Nitrogen Dynamics for Soil Management. Lewis Publishers, Boca Raton, 549–569.

Khorram, S.,Koch, F.H.,van der Wiele, C.F.,Nelson, S.A.C, 2012. Remote sensing. Springer briefs in space development, Springer (Eds), 134p.

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

Kokaly, R.F. and Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67 (3), 267–287.

Kornelsen, K.C., Coulibaly, P., 2013. Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. *Journal of hydrology 476*, 460-489.

Kouadio, A.L., 2012. Prévisions des rendements du blé d'hiver à échelle régionale par modélisation de la courbe de chute de l'indice foliaire. PhD thesis, University of Liege, 187p.

Kroes, J. G., Van Dam, J. C. , Groenendijk, P., Hendricks, R. F. A., Jacobs, C. M. J., 2008. SWAP version 3.2. Theory description and user manual, *Alterra Report.*, Wageningen, 266 p.

Kross, A., McNairn, H., Lapen, D., Sunohara, M., Champagne, C., 2015. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Erath observation and geoinformation* 34, 235-248.

Kumar, M., Monteith, J. L., 1981.Remote sensing of crop growth. Plants and the Daylight Spectrum. Smith, H. New York, Academic Press**:** 133-144.

Lahoz, W., Khattatov, B., Ménard, R. (Eds.), 2010. Data assimilation : Making sense of observations. Springer, 732 p.

Launay, M. and Guérif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agricult.Ecosyst. Environ.* 111 (1-4), 321-339.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math*. 2 (2), 164–168.

Liang, S., 2004.Quantitative remote sensing of land surfaces. John Wiley & Sons, Inc., Hoboken, new Jersey.

Liu, Y. and Gupta, H.V., 2007. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour. Res.* 43, W07401, doi: 10.1029/2006WR005756, 18 p.

Liu, J., Pattey, E., Miller, J.R., McNairn, H., Smith, A., Hu, B., 2010. Estimating crop stress, above ground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. *Remote Sens. Environ*.114, 1167-1177.

Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agric. Ecosyst. Environ.* 94, 205–220.

Lobell, D.B. and Field, C.B., 2007. Global scale climate-crop yield relationships and the impacts of the recent warning. *Environ. Res. Lett.* 2, 1-7.

Lobell, D.B., Ortiz-Monasterio, J.I., Falcon, W.P., 2007. Yield uncertainty at the field scale evaluated with multi-year satellite data. *Agricultural Systems* 92, 76-90.

Lobell, D.B. and Burke, M.B., 2010. On the use of statistical models to predict yield responses to climate change. *Agricultural and Forest Meteorology* 150, 1443-1452.

Los, S.O., Pollack, N.H., Parris, M.T., Collatz, G.J., Tucker, C.J., Sellers, P.J., Malmstrom, C.M., DeFries, R.S., Bounoua, L., Dazlich, D.A., 2000. A global 9-yr biophysical land surface dataset from NOAA AVHRR data. *Journal of hydrometeorology* 1, 183-199.

Löw, A., Ludwig, R., Mauser, W., 2005. Use of microwave remote sensing data to monitor spatio temporal characteristics of surface soil moisture at local and regional scales. *Advances in geosciences* 5(49), 49-56.

Löw, F., Duveiller, G., 2014. Defining the Spatial resolution requirements for crop Identification Using Optical Remote Sensing. *Remote Sens. 6*, 9034-9063; doi:10.3390/rs6099034

Lucau-Danila, C., 2008. LAI retrieval from SAR remote sensing for crop monitoring at regional scale, PhD thesis, 142 pp, Université Catholique de Louvain, UCL-geomatics, Louvain-la-Neuve, Belgium.

Luckman, A., 1998. The effects of topography on mechanisms of radar backscatter from coniferous forest and upland pasture. *IEEE transactions on geoscience and remote sensing* 36, 1830-1834.

Maas, S.J., 1988. Using satellite data to improve model estimates of crop yield. *Agronomy Journal*. 80: 655-662.

Macdonald, R.B. and Hall, F.G., 1980. Global crop forecasting.*Science* 208 (4445), 670-679.

Maier, S. W., Lüdeker, W., Günther, K. P., 1999. SLOP: A Revised Version of the Stochastic Model for Leaf Optical Properties. *Remote Sens. Environ*. *68*(3), 273-280.

Mannschatz, T., Pflug, B., Borg, E., Feger, K.-H., Dietrich, P., 2014. Uncertainties of LAI estimation from satellite imaging due to atmospheric correction. *Remote sensing of environment* 153, 24-39.

Marquardt, D.W., 1963. An algorithm for the least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math*. 11 (2), 431–441.

Martin, B., Tychon, B., El Jarroudi, M., Curnel, Y., Oger, R., 2006. Green leaf area decline of wheat top three leaves in Belgium and G–D of Luxembourg from 2003 to 2006: the relationships with grain yield. Presentation at the III CGMS Experts meeting and GEOLAND training workshop, Arlon, Belgium, 23–25 October 2006, 13 slides.

Martimort, P., 2009. Sentinel-2, optical high resolution mission for GMES operational services. Presentation at the 'Lansat science meeting', June 2009, Rochester, NY, USA, 17 slides.

Masson, V., Champeaux, J.-L., Chauvin, F., Meriguet, C., Lacaze, R.2003. A Global database of Land Surface Parameters at 1-km Resolution inMeteorological and Climate Models. *Journal of Climate* 16(9), 1261–1282.

Mattia, F., Le Toan, T., picard, G., Posa, F.I., D'Alessio, A., Notarnicola, C., Gatti, A.M., Rinaldi, M., Satalino, G., Pasquariello, G., 2003. Multitemporal C-band radar measurement on wheat fields.*IEEE Transactions on Geoscience and Remote Sensing 41*(7), 1551-1560.

McNairn, H., Van der Sanden, J. J., Brown, R. J., Ellis, J., 2000.The potential of RADARSAT-2 for crop mapping and assessing crop condition, Proceedings of the Second International conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista, Florida, 10-12 January 2000.

McNairn, H., Brisco, B., 2004. The application of C-band polarimetric SAR for agriculture : a review. *Can. J. Remote Sensing 30(3)*, 525-542.

Mearns, L.O., Easterling, W., Hays, C., Marx, D., 2001. Comparison of agricultural impacts of climate change calculated from high and low resolution climate change scenarios: part I. The uncertainty due to spatial scale. *Climatic change* 51 (2), 131-172.

Mkhabela, M.S., Bullock, P., Raj, S., Wang, S., Yang, Y., 2011. Crop yield forecasting on the Canadian prairies using MODIS NDVI data. *Agricultural and Forest Meteorology* 151, 385-393.

Monteith, J.L., 1972.Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9, 747–766.

Moran, M.S., Inoue, Y., Barnes, E.M., 1997.Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ*. 62, 319-346.

Moran, S., Alonso, L., Moreno, J.F., Pilar Cendrero Mateo, M., Fernando de la Cruz, D., Montoro, A., 2012. A RADARSAT-2 quad-polarized time series for

monitoring crop and soil conditions in Barrax, Spain. *IEEE Transactions on Geoscience and Remote Sensing* 50(4), 1057-1070. DOI: 10.1109/TGRS.2011.2166080

Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in water resources* 28, 135-147.

Moreau, S. and Le Thoan, T., 2003. Biomass quantification of Andean wetland forages using ERS satelliteSAR data for optimizing livestock management. *Remote Sens. Environ*. 84(4), 477-492.

Moulin, S., Bondeau, A., Delecolle, R., 1998.Combining agricultural crop models and satellite observations: from field to regional scales. *Int. J. remote sensing* 19 (6), 1021-1036.

Murthy, V.R.K., 2004. Crop growth modeling and its applications in agricultural meteorology.Satellite remote sensing andGIS applications in agricultural meteorology.Proceedings of a Training Workshop held 7-11 July 2003 in Dehra Dun (India), AGM-8, WMO/TD-No. 1182, 423p.

Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sens. Environ.*89 (3), 393–408.

Mutiga, J.K., Su, Z., Woldai, T., 2010.Using satellite remote sensing to assess evapotranspiration: case study of the upper Ng'iro North Basin, Kenya. *International Journal of Applied Earth Observation and Geoinformation* 125, S100-S108.

Mynemi, R.B., Ramakrishna, R., Nemani, R., Running, S., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE transactions on geosciences and remote sensing* 35(6), 1380-1393.

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

Ndacyayisenga, N. and Terres, J.M., 1998. Analyse de sensibilité du modèle de suivi de croissance des cultures « CGMS ». In: Utilisation des techniques d'analyse de variance. European Commission, JRC, Ispra, Italy, 35 p.

Nellemann, C., MacDevette, M., Manders, T., Eickhout, B., Svihus, B., Prins, A. G., Kaltenborn, B. P. (Eds), 2009.*The environmental food crisis – The environment's role in averting future food crises.*A UNEP rapid response assessment.United Nations Environment Programme, GRID-Arendal, ISBN: 978-82-7701-054-0, 104p.

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

OECD/FAO, 2011.*OECD-FAO Agricultural Outlook 2011-2020*, OECD Publishing and FAO. http://dx.doi.org/10.1787/agr_outlook-2011-en

Oh, Y., Sarabandi, K., Ulaby, F.T., 1992. An empirical model and inversion technique for radar scattering from bare soil surfaces. *IEEE Trans. Geosci. Remote Sens.* 30, 370-381.

Olesen, J.E., Bocher, P.K., Jensen, T., 2000. Comparison of scales of climate and soil data for aggregating simulated yields of winter wheat in Denmark. *Agr.Ecosyst. Environ.* 82, 213–228.

Olesen, J.E., Berntsen, J., Hansen, E.M., Petersen, B.M., Petersen, J., 2002a. Crop nitrogen demand and canopy area expansion in winter wheat during vegetative growth. *Eur. J. Agron.* 16, 279–294.

Olesen, J.E., Petersen, B.M., Berntsen, J., Hansen, S., Jamieson, P.D., Thomsen, A.G., 2002b. Comparison of methods for simulating effects of nitrogen on green area index and dry matter growth in winter wheat. *Field Crops Res.* 74, 131–149

Olesen, J.E., Trnka, M., Kersebaum, K.C., Skjelvag, A.O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., Micale, F., 2011. Impacts and adaptation of European crop production systems to climate change. *Europ. J. Agronomy* 34, 96-112.

Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J.-P., Braud, I., Jacobs, F., Lecharpentier, P., Ottlé, C., Calvet, J.-C., Brisson, N., 2005. Future directions for advanced evapotranspiration modeling : Assimilation of remote sensing data into crop simulation models and SVAT models. *Irrigation and Drainage Systems* 19, 377-412.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Şaylan, L., Šiška, B., Rötter, R., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *Europ. J. Agronomy* 35, 103-114.

Passioura, J.B., 1996. Simulations models: science, snake oil, education, or engineering ? *Agron. J.* 88, 690-694.

Pauwels, V.R.N., Verhoest, N.E.C., De Lannoy, G.J.M., Guissard, V., Lucau, C., Defourny, P., 2007. Optimization of a coupled hydrology/crop growth model through the assimilation of observed soil moisture and LAI values using an Ensemble Kalman Filter. *Water Resources Research* 43(4), W04421, doi:10.1029/2006WR004942.

Pearson, C.J., Ison, R.L., 1997. Agronomy of grassland systems, 2nd edition. Cambridge university press, 222 p.

Pellenq, J., Boulet, G., 2004. A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. *Agronomie* 24 (2004), 197–204.

Peltonen-Sainio, P., Jauhiainen, L., Trnka, M., Olesen, J. E., Calanca, P., Eckersten, H., Eitzinger, J., Gobin, A., Kersebaum, K. C., Kozyra, J., Kumar, S., Marta, A. D., Micale, F., Schaap, B., Seguin, B., Skjelvåg, A. O., Orlandini, S., 2010. Coincidence of variation in yield and climate in Europe. *Agriculture, Ecosystems & Environment.*139, (4), 2010, 483-489.

Picoli, M. C. A., Lamparelli, R.A.C., Sano, E.E., Batista de Mello, J.R., Rocha, J.V. (2013). Effect of sugarcane-planting row directions on ALOS/PALSAR satellite images. *GIScience & Remote Sensing* 50(3), 349-357.

Powell, M.J.D., 1998. Direct search algorithms for optimization calculations. *Acta Numer.* 7, 287–336.

Prévot, L., I. Champion, and G. Guyot, 1993. Estimating surface soil moisture and leaf area index of a wheat canopy using a dual-frequency (C and X bands) scatterometer. *Remote Sens. Environ*.46(3), 331-339.

Prévot, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., Brisson, N., 2003. Assimilating optical and radar into the STICS crop model for wheat. *Agronomie* 23, 297-303.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B.E., Disney, M., Bowyer, P., 2008. Assimilating canopy reflectance into an ecosystem model with an Ensemble Kalman Filter. *Remote Sens. Environ*.112, 1347-1364.

Qiu, F., Berglund, J., Jensen, J.R., Thakkar, P., Ren, D., 2004. Speckle noise reduction in SAR imagery using a local adaptative median filter. *GIScience and Remote Sensing* 41, N°3, 244-266.

Raicich, F. and Rampazzo, A., 2003. Observing System Simulation Experiments for the assessment of temperature sampling strategies in the Mediterranean Sea. *Annales Geophysicae* 21, 151-165.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem.Cycles,* 22(1), GB1003.

Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. *Nature communications* | 6:5989 | DOI: 10.1038/ncomms6989.

Reichle, R.H., 2008. Data assimilation methods in the Earth sciences. *Advances in water Resources* 31, 1411-1418.

Reichle, R.H., Crow, W.T., Keppenne, C.L., 2008a. An adaptive Ensemble Kalman Filter for soil moisture data assimilation. *Water Resour. Res.* 44, W03423, doi:10.1029/2007WR006357.

Reichle, R.H., Cow, W.T., Koster, R.D., Sharif, H.O., Mahanama, S.P.P., 2008b. Contribution of soil moisture retrievals to land data assimilation products. *Geophys.Res.Lett.* 35, L01404, doi:10.1029/2007GL031986.

Ren, J., Chen, Z., Zhou, Q., Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation* 10(4), 403–413.

Riedel, T., Pathe, C., Thiel, C., Herold, M., Schmullius, C., 2002. Systematic investigation on the effect of dew and interception on multifrequency and multipolarimetric radar backscatter signals. Proceedings of the 3rd International Symposium on Retrieval of Bio- and Geophysical Parameters from SAR Data for Land Applications, 11th - 14th September 2001, SCEOS, Sheffield, UK, 99-104.

Riedel, T. and Schmullius, C.C., 2003. Effect of interception of the backscattering behaviour from crops.Proceedings IGARSS'03, 21–25 July 2003, Toulouse, France, on CD.

Ritchie, J.T. and Otter, S., 1985. Description and performance of CERES-Wheat: A User oriented Wheat Yield Model. ARS Wheat Yield Project ARS-38. Natl. Tech. Info. Serv., Springfield, MO, pp. 159–175.

Rivera Caicedo, J.P., 2014. Optimized and automated estimation of vegetation properties: Opportunities for Sentinel-2. PhD thesis, Universitat de Valencia, 171 p.

Rodriguez, D., Van Oijen, M., Schapendonk, A.H.M.C., 1999. LINGRA-CC: a sink–source model tosimulate the impact of climate change and management on grassland productivity. *New Phytol.* 144, 359–368.

Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takáč, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. *Field Crops Research* 133, 23-36.

Roy, D.P., 2000. The Impact of misregistration Upon Composited Wide Field of View Satellite Data and Implications for Change Detection. *IEEE transactions on geosciences and remote sensing* 38 (4), 2017-2032.

Roy, D.P.,Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., Mc Corkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne,

R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. *Remote sensing of Environment* 145, 154-172.

Rubarenzya, M.H., Willems, P., Berlamont, J., 2007. Identification of uncertainty in distributed hydrological modelling: case study of the Grote Nete catchment in Belgium. Available on website www.wrc.org.za

Ruget, F., Novak, S., Granger S., 2006. Du modèle STICS au système ISOP pour estimer la production fourragère. Adaptation à la prairie, application spatialisée. *Fourrages* 186, 241-256.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., Verma, S.B., Suyker, A.E., Arkebauer, T.J., 2010. A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data. *Remote Sensing of Environmen*t 114, 2146-2159.

Saich, O., Bourgeaud, M., 2000.Interpreting ERS SAR signatures of agricultural crops in Flevoland, 1993-1996. *IEEE T. Geoscience and Remote Sensing* 38(2), 651-657.

Shang, J., Liu, J., Huffman, T., Qian, B., Pattey, E., Wang, J., Zhao, T., Geng, X., Kroetsch, D., Dong, T., Lantz, N., 2014. Estimating plant area index for monitoring crop growth dynamic using Landsat-8 and RapidEye images. *Journal of applied remote sensing* vol. 8, 085196-1 to 085196-12.

Schapendonk, A.H.C.M., Stol, W., van Kraalingen, D.W.G., Bouman, B.A.M., 1998. LINGRA, a sink/source model to simulate grassland productivity in Europe. *European Journal of Agronomy* 9, 87-100.

Schlesinger, W.H., 1977. Carbon balance in terrestrial detritus. *Ann. Rev. Ecol. Syst*., 8: 51–81

Schowengerdt, R.A., 2007. *Remote sensing: models and methods for image processing*. San Diego: Academic Press., 3rd edition.

Schut, A., Stephens, D., Stovold, R., Adams, M., Craig, R., 2009. Improved wheat yield and production forecasting with a moisture stress index, AVHRR and MODIS data. *Crop. Pasture Sci.* 60(1), 60-70.

Sellers, P.J., Tucker, C.J., Collatz, G.J., Los, S.O., Justice, C.O., Dazlich, D.A., Randall, R.A.,1994. A global 1° by 1° NDVI data set for climate studies. Part 2: the generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing* 15(17), 3519-3545.

Seidl, M. S., Batchelor, W. D., Paz, J.O., 2004. Integrating remotely sensed images with a soybean modelto improve spatial yield simulation. *Transactions of the American Society of Agricultural Engineers* 47(6), 2081–2090.

Sheehy, J.E. and Johnson, I.R., 1988. Physiological models of grass growth. In: Jones, M.B., Lazenby,A. (Eds.), The grass crop: the physiological basis for production. Chapman & hall, London, 243-275.

Skriver, H., Svendsen, M.T., Thomsen, A.G., 1999. Multitemporal C- and L-band polarimetric signatures of crops. *IEEE transaction on geoscience and remote sensing* 37(5), 2413-2429.

Snoeij, P., Attema, E., Davidson, M., Floury, N., Levrini G., Rosich, B., Rommen B., 2008. Sentinel-1, the GMES Radar Mission. IEEE Radar Conference, Rome (Italy), 26-30th 13 of May 2008, 5p.

Sofko, G. J., Sloboshan, J., McKibben, M., Koehler, J., Brisco, B., 1989. Variation of microwave radar cross-section of wheat during the early hours of a rainfall. In: IGARSS '89/12th Canadian symposium on remote sensing, Vancouver, BC (pp. 1191–1194). Canadian Remote Sensing Society.

Soria-Ruiz, J.,McNairn, H., Fernandez-Ordonez, Y., Bugden-Storie, J., 2007. Corn monitoring and crop yield using optical and RADARSAT-2 images, in *Geoscience and Remote Sensing Symposium. IGARSS 2007. IEEE International*, edited, Barcelona.

Spitters, C.J.T., 1990. Crop growth models: their usefulness and limitations. *Acta Hortic.* 267, 349–368.

Stilmant, D., Rabier, F., Dufrasne, S., Oger, R., Buffet, D., 2001. Prévision des quantités et qualités de fourrages disponibles : aide à l'établissement d'une stratégie alimentaire à l'échelle de l'exploitation et de la région agricole. Rapport final Convention T4/42/67, Centre de Recherches Agronomiques, 86 p.

Stöckle,C. O., Donatelli, M., Nelson, R., 2003.CropSyst, a cropping systems simulation model. *Europe J. Agronomy* 18, 289-307.

Stolz, R. and Mauser, W., 1997. Evaluation of ERS data for biomass estimation of meadows.Proc.3rd ERS symp.On space at the service of our Environment, Florence (Italy), 17-21 March 1997, 203-207.

Supit, I., 2000. An exploratory study to improve the predictive capacity of the Crop Growth Monitoring System as applied by the European Commission.Treebook 4, ca 180 p, ISBN 90-804443-5-9 (paperback), Treemail Publishers, Heelsum.

Supit, I., Hooijer, A.A., van Diepen, C.A., 1994. System Description of the WOFOST 6.0 Crop Simulation Model Implemented in CGMS, Volume 1: Theory and Algorithms. EUR 15956 EN, Joint Research Center, Commission of the European Communities, Luxembourg.

Supit, I., van Diepen, C.A., de Wit, A.J.W., Kabat, P., Baruth, B., Ludwig, F., 2010. Recent changes in the climatic yield potential of various crops in Europe. *Agricultural systems* 103, 683-694.

Supit, I., van Diepen, C.A., de Wit, A.J.W., Wolf, J., Kabat, P., Baruth, B., Ludwig, F., 2012. Assessing climate change effects on European crop yields using the

Crop Growth Monitoring System and a weather generator. *Agricultural and Forest meteorology* 164, 96-111.

Svoray, T. and Shoshany, M., 2002. SAR-based estimation of areal aboveground biomass (AAB) of herbaceous vegetation in the semi-arid zone: a modification of the water-cloud model. *Int. j. remote sensing* vol. 23, no. 19, 4089–4100.

Taconet, O., Benallegue, M., Vidal-Madjar, D., Prevot, L., Dechambre M., Normand, M., 1994.Estimation of soil and crop parameters for wheat from airborne radar backscaterring data in C and X bands. *Remote Sens. Environ.*50, 287-294.

Tan, B., Woodcock, C. E., Hu, J., Zhang, P., Ozdogan, M., Huang, D., Yang, W., Knyazikhin, Y., Myneni, R.B., 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sens. Environ.*105, 98–114.

Thornley, J.H.M., 1991. A model of leaf growth tissue, acclimation and senescence. *Ann. Botany* 67, 219-228.

Thorp, K.R., Hunsaker, D.J., French, A.N., 2010. Assimilating leaf area index estimates from remote sensing into the simulations of a cropping systems model. *Transactions of the ASABE* 53(1), 251-262.

Tilling, A.K., O'Leary, G.J., Ferwerda, J.G., Jones, S.D., Fitzgerald, G.J., Rodriguez, D., Belford, R., 2007. Remote Sensing of nitrogen and water stress in wheat. *Field Crops Research* 104, 77-85.

Torres, L., Sant'Anna, S.J.S., da Costa Freitas, C., Frery, A.C., 2014. Speckle reduction in polarimetric SAR imagery with stochastic distances and non local means. *Pattern Recognition* 47, 141-157.

Tucker, C., Holben, B., and Elgin, J., 1980. Relationship of spectral data to grain yield variation. *Photogrammetric Engineering and Remote Sensing* 46, 657–666.

Tucker, C.J., Vanpraet, C.L., Sharman, M.J. and van Ittersun, G., 1985. Satellite remote sensing of total herbaceous production in the Senegalese Sahel 1980-1984. *Remote Sens. Environ.*17, 232-249.

Tupin, F., Nicolas, J.M., Inglada, J., 2014.Imagerie de télédétection. Editions Hermes-Lavoisier, *Traité IC2, série Signal et Image,* 369 p.

Tychon, B., Buffet, D., Dehem, D., Oger, R., Veroustraete, F., Wouters, K., 1999. Adaptation of the European crop growth monitoring system to the Belgian conditions.Proceedings of the International symposium on modeling cropping systems Lleida, 21-23 June 1999, 263-264.

Uehara, G. and Tsuji, G.Y., 1993. The IBSNAT project. *In* Systems approaches for sustainable agricultural development, eds F.W.T. Penning de Vries, P.

Teng, and K. Metselaar. Kluwer Academic Publisher, Dordrecht, The Netherlands, 505-513.

Ulaby, F.T., Allen, C.T., Eger, G., Kanemasu, E., 1984. Relating the microwave backscattering coefficient to Leaf Area Index. *Remote Sens. Environ.*14, 113-133.

Van Bussel, L.G.J, Ewert, F., Leffelaar, P.A., 2011. Effects of data aggregation on simulations of crop phenology. *Agriculture, Ecosystems and Environment* 142, 75-84.

Van Dam, J. C., Groenendijk, P., Hendricks, R. F. A., Kroes, J. G., 2008. Advances of modeling water flow in variably saturated soils with SWAP. *Vadose Zone Journal 7*(2), 640-653.

Van Ittersum, M.K., Leffelaar, P.A., Van Keulen, H., Krop, M.J., Bastiaans, L., Goudriaan, J., 2003. Onapproaches and applications of the Wageningen crop models. *Eur. J. Agron.* 18, 201–234.

Van Oijen, M., Ewert, F., 1999. The effects of climatic variation in Europe on the yield response of spring wheat cv. Minaret to elevated CO2 and O3: an analysis of opentop chamber experiments by means of two crop growth simulation models. *Eur. J. Agron.* 10, 249–264.

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

Vescovo, L. and Gianelle, D., 2008. Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). *Advances in Space Research* 41, 1764-1772.

Voormansik, K., Jagdhuber, T., Olesk, A., Hajnsek, I, Papathanassiou, K.P., 2013. Towards detection of grassland cutting practices with dual polarimetric TerraSAR-X data. *International Journal of Remote Sensing* 34, No. 22, 8081-8103.

Vossen, P. and Rijks, D., 1995. Early crop yield assessment of the E.U. countries: the system implemented by the Joint Research Centre. EUR 16318.Publication of the Office for Official Publications of the E.C., Luxembourg.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003.A shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.* 39(8), 1201, doi: 10.1029/2002WR001642.

Vrugt, J. A., Diks,C. G. H., Gupta,H. V., Bouten,W., Verstraten, J. M.,2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of

global optimization and data assimilation. *Water Resour. Res*. 41, W01017, doi:10.1029/2004WR003059

Walker, W.E., Harremoes, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Von Krauss, M.P.K., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr. Assess.* 4, 5–17.

Wang, X., Ge, L., Li, X., 2013. Pasture Monitoring using SAR with COSMO-SkyMed, ENVISAT ASAR, and ALOS PALSAR in Otway, Australia. *Remote Sens*. 5, 3611-3636 ; doi:10.3390/rs5073611.

Wassenaar, T., Lagacherie, P., Legros, J.-P., Rounsevell, M.D.A., 1999. Modelling wheat yield responses to soil and climate variability at the regional scale. *Climate Res.* 11, 209–220.

Wei, J. and Malanotte-Rizzoli, P., 2010. Validation and application of an Ensemble kalman Filter in the Selat Pauh of Singapore. *Ocean dynamics* 60, 395-401.

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

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 Sens. Environ.* 110, 317–331.

Whitcraft, A.K., Vermote, E.F., Becker-Reshef, I., Justice, C.O., 2014. Cloud cover throughout the agricultural growing season: impacts on passive optical earth observations. *Remote sensing of Environment*, article in press.

White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. *Field crops research* 124, 357-368.

Whitehead, D.C., 1995. Grassland Nitrogen. CAB International, London.

Wiegand, C.L., Richardson, A.J., Jackson, R.D., Pinter Jr., P.J., Aase, J.K., Smika, D.E., Lautenschlager, L.F., Mc Murtrey III, J.E., 1986. Development of agrometeorological crop model inputs from remotely sensed information. *IEEE transactions on Geoscience and Remote Sensing* 24, 90-97.

Wiegand, C.L., Richardson, A.J. Escobar, D.E., Gerbermann, A.H., 1991. Vegetation Indices in Crop Assessments. *Remote Sens. Environ. 35, 105-119.*

Wolfe, R.E., Roy, D.P., Vermote, E., 1998. MODIS Land Data Storage, Gridding, and Compositing Methodology: Level 2 Grid. *IEEE transactions on geosciences and remote sensing* 36(4), 1324-1338.

Wolfe, R., Nishihama, M., Fleig, A., Kuyper, J., Roy, D., Storey, J., Patt, F., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sens. Environ.* 83, 31-49.

Wood, D., McNairn, H., Brown, R.J., Dixon, R., 2002. The effect of dew on the use of RADARSAT-1 for crop monitoring choosing between ascending and descending orbits. *Remote Sens. Environ.*80, 241-247.

Woodward, S.J.R., 2001. Validating a model that predicts daily growth and feed quality of NewZealand dairy pastures. *Environ. Int.* 27, 133-137.

Wu, J., Wang, D., Bauer, M.E., 2007. Assessing broadband vegetation indices and QuickBird data in estimating leaf area index of corn and potato canopies. *Field Crops Research* 102, 33-42.

Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., 2012.Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield. Computer and computing technologies in agriculture V, IFIP advances in Information and communication technology 370, 410-418.

Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., 2014. Remote sensing-based crop monitoring experiences with China's CropWatch system. *International Journal of Digital Earth*, 7(2), 113-137.

Wulder, M.A., White, J.C., Hilker, T., Coops, N.c., Griffiths, P., Pflugmacher, D., Hostert, P., MAsek, J.G., 2014. Towards a satellite virtual constellation for land characterization; concepts and roles for sentinel-2 and Landsat. Abstract, Sentinel-2 for science workshop, 20-22 May 2014, ESA-ESRIN, Frascati (Rome), Italy.

Xiao, Z., Liang, S., Wang, J., Jiang, B., Li, X., 2011. Real-time retrieval of Leaf Area Index from MODIS time series data. *Remote Sens. Environ.*115, 97-106.

Yang, W., Huang, D., Tan, B., Stroeve, J., Shabanov, N., Knyazikhin, Y., Nemani, R., Myneni, R. 2006. Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the terra MODIS sensor: 2000-2005. *IEEE Transactions on Geoscience and Remote Sensing 44(7)*, 1829–1842.

Yao, Y., Liu, Q., Liu, Q., Li, X., 2008. LAI retrieval and uncertainty evaluations for typical row-planted crops at different growth stages. *Remote Sens. Environ.*112, 94-106.

Yi, Y., Yang, D., Huang, J., Chen, D., 2008. Evaluation of MODIS reflectance products for wheat leaf area index (LAI) retrieval. *ISPRS Journal of Photogrammetry & Remote Sensing* 63, 661-677.

Yoder, B.J., Pettigrew-Crosby, R.E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. *Remote Sens. Environ*. 53 (3), 199–211.

Yuping, M., Shili, W., Li, Z., Yingyu, H., Liwei, Z., Yanbo, H., Futang, W., 2008. Monitoring winter wheat growth in North China by combining a crop model

and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation* 10, 426–437.

Zhang, Y., Chen, J.M., Miller, J.R., Noland, T.L., 2008. Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. *Remote Sensing of Environment* 112, 3234-3247.

Zribi, M., Dechambre, M., 2002. A new empirical model to retrieve soil moisture and roughness from C-band radar data. *Remote sensing of environment* 84, 42-52.