Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Télédétection de la Biosphère Continentale - Ecologie

Présentée et soutenue par :
Amanda Gabriela Maia VELOSO

le : lundi 23 juin 2014

Titre :

Modélisation spatialisée de la production, des flux et des bilans de carbone et d'eau des cultures de blé à l'aide de données de télédétection : application au sud-ouest de la France.

Ecole doctorale :
Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E)

Unité de recherche :
Centre d'Études Spatiales des la Biosphère (CESBIO UMR 5126)

Directeur(s) de Thèse :

Eric CESCHIA, Valérie DEMAREZ, Richard ESCADAFAL

Rapporteurs :
A. BEGUE, Directrice de Recherche, CIRAD, Montpellier
R. CASA, Professeur, Université de la Tuscia, Viterbo, Italie
N. VIOVY, Ingénieur-Chercheur CEA, LSCE, Saclay

Membre(s) du jury :

J. C. CALVET, Directeur de recherche, CNRM-GAME, Toulouse
D. COURAULT, Chargée de recherche, INRA, Avignon
J. P. GASTELLU-ETCHEGORRY, Professeur, Université de Toulouse

Acknowledgments

Je tiens tout d'abord à remercier sincèrement Yann Kerr, le directeur du CESBIO, ainsi que toutes les personnes du CESBIO qui m'ont accueillies dans leur laboratoire et ont permis que cette thèse se déroule dans les meilleures conditions.

J'adresse de très chaleureux remerciements à Éric Ceschia, mon directeur de thèse, qui m'a encadré au cours de ces trois années. Merci pour ta confiance. Merci de m'avoir fait découvrir et aimer un métier et un domaine de recherche passionnants. Je tiens à exprimer ma gratitude aussi à Valérie Demarez, ma co-directrice de thèse. Merci de ton aide et de tes conseils éclairés. A vous deux, merci de m'avoir accompagné au long de ce « voyage ».

Je remercie les rapporteurs de ma thèse Agnès Bégué, Raffaele Casa et Nicolas Viovy, d'avoir consacré du temps à la lecture de mon manuscrit et de leurs précieux commentaires. Je suis également reconnaissante aux autres examinateurs du jury Jean-Christophe Calvet, Dominique Courault et Jean-Philippe Gastellu-Etchegorry pour l'intérêt porté à mon travail. Merci à tous d'avoir joyeusement participé à mon jury de soutenance. Ce fût un moment mémorable pour moi. Merci également à Richard Escadafal, qui a accepté d'être mon directeur de thèse administratif.

Un grand merci à l'ensemble du personnel technique pour leur participation à l'acquisition des données expérimentales, ainsi qu'à toutes les personnes ayant contribué au traitement, au recueil et à la compréhension des diverses données utilisées lors de ma thèse. Merci encore à tous les collègues avec qui j'ai eu beaucoup de plaisir à discuter et qui m'ont fait bénéficier de leur expérience en télédétection, agronomie, écologie, physique, informatique, mathématique… Je remercie également le personnel administratif et informatique qui ont toujours promptement répondu à mes sollicitations.

Un grand merci à tous mes collègues et amis du CESBIO avec qui je partage toujours de très bons moments. Merci pour les pauses, barbecues, piscines, soirées, CESBIO plage, CESBIO Noël, …

Je remercie de tout cœur tous mes amis (je ne vais pas vous nommer… vous vous reconnaitrez) d'avoir toujours été là pour moi (de près ou de loin), assurant le bon déroulement de cette thèse. Merci pour les moments de détente, les rigolades, la bonne humeur.

Um muito obrigada à toda a minha família. Dedico esta tese à vocês, que foram essenciais para o sucesso deste trabalho. Obrigada aos meu pais, pela confiança, por terem acreditado em mim, sempre! Obrigada pelo apoio incondicional, pelo amor. Obrigada aos meus irmãos, tão importantes na minha vida. Quen!

Un grand merci aussi à ma belle-famille pour leurs encouragements, les repas (!) et leur affection.

Pour finir, un énorme merci à mon amoureux qui m'a rejoint au cours de cette péripétie et qui a su m'accompagner, me guider, me supporter (pas toujours évident !) et me soutenir pendant ces années. Merci pour tout. Notre belle aventure ne fait que commencer =)

A tous ceux qui ont contribué et participé, directement ou indirectement, à l'accomplissement de ce travail… MERCI !

INDEX

Abstract

The agricultural lands that occupy more than one third of Earth's terrestrial surface contribute to climate change and are also impacted by those changes, since their production is conditioned by the climatic conditions and water resources. The main objective of this thesis is therefore to quantify and analyze the production and also the main components of the carbon and water biogeochemical cycles for crop ecosystems in contrasted climatic years, focusing specifically on the winter wheat crop, in order to identify the best strategies for maintaining crop production and reducing environmental impacts. The study area is located in southwest France.

For this purpose, we propose a regional modeling approach that combines: i) high spatial and temporal resolutions (HSTR) optical remote sensing data, ii) simple crop models and iii) an extensive set of *in-situ* measurements for models' calibration and validation. The combined use of these three 'tools' opens new perspectives for advanced agro-ecosystems modeling and monitoring at regional or global scales.

Remote sensing data from the Formosat-2 and SPOT satellites were used to produce green area index (GAI) dynamic maps, since GAI has a key role in soil-plant-atmosphere interactions. Two methods (empirical and physical) were tested for producing GAI maps and the physical method, based on the inversion of a radiative transfer model using artificial neural networks, was kept for the modeling approach since it presented good performances, with no specific calibration.

The dynamic GAI maps were used for calibrating the investigated crop models. The semi-empirical SAFY model, based on Monteith's light-use efficiency theory and adapted for remote sensing coupling, was then calibrated and evaluated in terms of GAI, biomass and yield estimates. Next, this model was modified in order to simulate the components of the net carbon fluxes: gross primary production (GPP), ecosystem respiration (Reco) and net ecosystem exchange (NEE). The resulting SAFY-CO_2 model was then coupled with a water budget module, based on the FAO-56 method. The final SAFYE-CO_2 model is therefore able to estimate the components of the crop water cycle, principally evapotranspiration and soil water content.

At last, it was possible to compute carbon budgets (NECB) and water budgets as well as some water use efficiency (WUE) indices that allowed evaluating the winter wheat crop ecosystems (locally or regionally), in terms of environmental and agronomical aspects.

The three investigated models (SAFY, SAFY-CO_2 and SAFYE-CO_2) reproduced correctly the biomass production (with relative error about 25%) and yield (relative

error ranging from 22% up to 36%, depending on the version of the model and the yield function estimates) for years with contrasted climatic conditions. The net carbon flux components estimated with the models were overall in agreement with the flux measurements performed on our two experimental sites (Auradé and Lamasquère), presenting very good correlations (R^2 about 0.9 for GPP, 0.75 for Reco and 0.84 for NEE). However, the performances of the models could still be improved by considering weeds development or re-growths events after harvest or by estimating GAI by means of radar remote sensing data for years with strong vegetation development.

The NECB calculated for different climatic, environmental and management conditions allowed identifying the main factors by which they are influenced and the potential levers for improving carbon storage in the crop soils and thus reducing the impacts of crop production on climate change. Regarding the evapotranspiration simulated by SAFYE-CO_2, they presented a good correlation (R^2 about 0.73) and satisfactory errors (RMSE about 0.47 $mm.day^{-1}$) when compared with the measurements. In terms of annual crop water consumption, the model estimates and measurements presented absolute differences between 20 $mm.year^{-1}$ and 77 $mm.year^{-1}$. Finally, the agronomical and environmental WUE indicators were calculated at regional scale. They will allow to identify sustainable strategies for preserving the environmental resources and maintaining crop production.

Résumé

Les terres agricoles, qui occupent plus d'un tiers de la surface continentale de la Terre, contribuent au changement climatique et sont aussi affectées par ces changements puisque leur production est contrainte par les conditions climatiques et les ressources en eau. L'objectif principal de cette thèse est donc de quantifier et d'analyser la production et aussi les principales composantes des cycles biogéochimiques du carbone et de l'eau des agrosystèmes, pour des années climatiques contrastées, afin d'identifier les meilleures stratégies pour maintenir la production et réduire les impacts environnementaux. Ce travail a été focalisé sur les cultures de blé du sud-ouest de la France.

Pour répondre à cet objectif nous proposons une approche de modélisation spatialisée qui combine : i) des données de télédétection optique à hautes résolutions spatiale et temporelle, ii) des modèles de culture semi-empiriques et iii) un ample dispositif de mesures *in-situ* pour la calibration et la validation des modèles. L'utilisation combinée de ces trois outils offre de nouvelles perspectives pour la modélisation et le suivi des agrosystèmes à l'échelle régionale et globale.

Les données satellites à hautes résolutions spatiale et temporelle issues des capteurs Formosat-2 et SPOT ont été utilisées pour produire des cartes dynamiques d'indice de surfaces vertes des plantes (Green Area Index, GAI) qui jouent un rôle fondamental dans les interactions sol-plante-atmosphère. Deux méthodes (empirique et physique) ont été testées pour produire ces cartes de GAI et l'approche physique basée sur l'inversion d'un modèle de transfert radiatif en utilisant des réseaux de neurones artificiels a été conservée pour la suite de ce travail car elle présente de bonnes performances, sans besoin de calibration.

Les séries temporelles de cartes de GAI ont ensuite été utilisées pour la calibration des modèles de culture étudiés. Le modèle semi-empirique SAFY basé sur la théorie de l'efficience de l'utilisation de la lumière de Monteith et adapté au couplage avec les données de télédétection a ainsi été calibré et évalué en termes d'estimations de GAI, de biomasse et de rendement. Par la suite, SAFY a été modifié afin de permettre la simulation des composantes des flux nets de CO_2: la production primaire brute (GPP), la respiration de l'écosystème (Reco) et les flux nets de CO_2 (NEE). Le modèle SAFY-CO_2 a ainsi été crée puis couplé à un module de bilan hydrique, inspiré de la méthode FAO-56 permettant au modèle résultant, nommé SAFYE-CO_2, d'estimer en plus des variables précédentes les composantes du cycle de l'eau des cultures, notamment l'évapotranspiration et la teneur en eau du sol. Au

final, des bilans de carbone (NECB) et d'eau ainsi que des efficiences de l'utilisation de l'eau (WUE) ont été calculés à partir des simulations de flux net de CO_2, d'eau et de rendement. Ces indicateurs ont permis d'évaluer localement et régionalement la performance des cultures de blé par rapport à des aspects agronomiques et environnementaux.

Les trois modèles étudiés (SAFY, SAFY-CO_2 et SAFYE-CO_2) reproduisent correctement la production de biomasse (avec des erreurs relatives d'environ 25 %) et le rendement (erreurs relatives entre 22% et 36 %, selon la version du modèle) pour des années avec des conditions climatiques contrastées. Les composantes du flux net de CO_2 estimées par les modèles sont globalement en accord avec les mesures de flux effectués sur les deux sites expérimentaux d'Auradé et de Lamasquère, présentant de très bonnes corrélations (R^2 environ 0,9 pour la GPP, 0,75 pour Reco et 0,84 pour la NEE). Cependant, les performances des modèles pourraient encore être améliorées en prenant en compte le développement des repousses et des mauvaises herbes, ou encore en estimant le GAI à partir de données de télédétection radar pour des fortes valeurs d'indice de végétation.

Les bilans de carbone calculés pour des conditions climatiques, environnementales et de gestion contrastées ont permis d'identifier des leviers d'actions pour améliorer le stockage du carbone dans les sols de culture et donc de réduire les impacts de la production agricole sur le changement climatique. Finalement, l'évapotranspiration simulée par SAFYE-CO_2 a présenté une bonne corrélation (R^2 autour de 0.73) et des erreurs correctes (RMSE environ 0,47 mm.j^{-1}) par rapport aux mesures. En termes de consommation annuelle en eau des cultures, les estimations et les mesures ont présenté des différences absolues entre 20 mm.an^{-1} et 77 mm.an^{-1}. Enfin, les indicateurs agronomiques et environnementaux de l'efficience d'utilisation de l'eau ont été calculés à l'échelle régionale. Ils permettront d'établir des stratégies à adopter pour la préservation des ressources environnementales et le maintien de la production agricole.

General Introduction

1. General & Specific Backgrounds

Climate and land use changes are major issues affecting the whole world society and all life on Earth. More specifically, according to the recent reports of the Intergovernmental Panel on Climate Change (IPCC, 2014), “climate change refers to a change in the state of the climate that can be identified by changes in the mean and/or the variability of its properties, and that persists for an extended period, typically decades or longer. Climate change may be due to natural internal processes or external forcing such as modulations of the solar cycles, volcanic eruptions, and persistent anthropogenic changes in the composition of the atmosphere or in land use”. The scientists have been studying this subject for years and there is now sufficient evidence that human influence on the climate system is clear. In recent decades, changes in climate and land use have caused impacts on natural and human systems on all continents and across the oceans. Continued emissions of greenhouse gases will cause further warming and changes in all components of the climate system (IPCC, 2014) and will have impacts on natural and cultivated vegetation distribution. Therefore climate change is a global problem and its limitation relies upon significant and continuous reductions of greenhouse gas emissions. In this context, the IPCC was created in 1988 by the World Meteorological Organization (WMO) and United Nations Environment Program (UNEP) to provide policymakers with regular assessments of the scientific basis of climate change, its impacts and future risks, and realistic response strategies for adaptation and mitigation [http://www.ipcc.ch/]. The IPCC has published four Assessment Reports, and before the end of this year (2014) the final version of the Fifth Assessment should be available. These reports underline the importance of climate and land use changes as challenges requiring international cooperation. They led, for example, to the creation of the United Nations Framework Convention on Climate Change (UNFCCC) and also supported the adoption of the Kyoto Protocol in 1997. The Fourth Assessment, published in 2007, has pointed out that most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gases concentrations. The latest (in progress) Fifth Assessment Report (AR5) has a more comprehensive assessment of scientific knowledge on climate change since 2007, focusing on the socio-economic aspects of climate change and its implications for sustainable development.

The atmospheric concentrations of the greenhouse gases (carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O)) have all increased since 1750 due to human activities. The CO_2 concentrations have increased by 40% since pre-industrial times (IPCC (2014);see Figure 1), primarily from fossil fuel emissions and secondarily from net land use change emissions. The ocean has absorbed about 30% of the emitted anthropogenic carbon dioxide, causing ocean acidification.

Atmospheric CO_2 represents the most active atmospheric phase of the global carbon cycle. The global carbon cycle can be viewed as a series of reservoirs of carbon in the Earth System, which are connected by exchange fluxes of carbon. Figure 2 illustrates the perturbation of the global carbon cycle caused by anthropogenic activities, averaged globally for the decade 2003–2012.

In addition to the carbon biogeochemical cycle, the water cycle (but also the N, P, *etc*. cycles) is also affected by climate change. According to the studies analyzed by the IPCC contributors, there is robust evidence that the freshwater-related risks of climate change increase significantly with increasing greenhouse gas concentrations. There is a high agreement that climate change over the 21st century is projected to reduce renewable surface water and groundwater resources significantly in most dry subtropical regions, probably intensifying competition for water among sectors. In many regions, change in precipitation or melting snow and ice may alter hydrological systems, affecting water resources in terms of quantity and quality (IPCC, 2014).

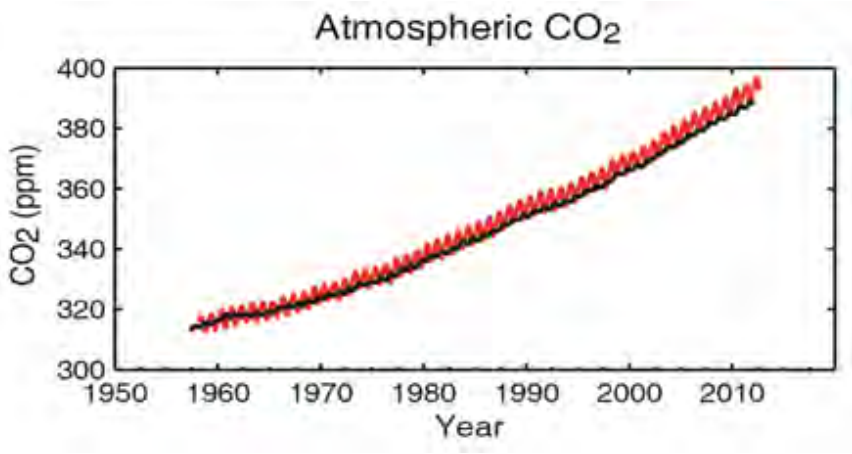


Figure 1- Atmospheric concentrations of carbon dioxide (CO_2) from Mauna Loa (19°32′N, 155°34′W – red) and South Pole (89°59′S, 24°48′W – black) since 1958, extracted from (Ciais et al., 2013).

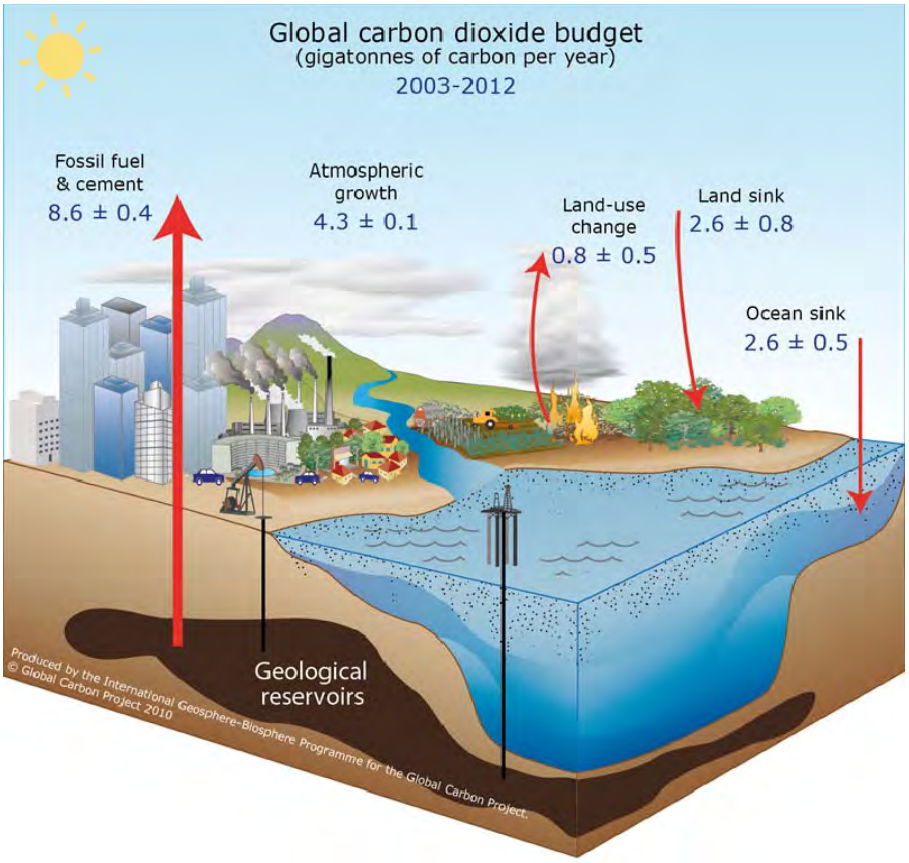


Figure 2- Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities, averaged globally for the decade 2003–2012. The arrows represent emission from fossil fuel burning and cement production; emissions from deforestation and other land-use change; and the carbon sinks from the atmosphere to the ocean and land reservoirs. The annual growth of carbon dioxide in the atmosphere is also shown. All fluxes are in units of PgC yr^{-1}. Source: Le Quéré et al., (2013); CDIAC Data; NOAA/ESRL Data; Global Carbon Project 2013.

In order to evaluate the sensitivity of food production to weather and climate, a number of models have widely been used to assess the response of crop yield to temperature. According to IPCC WGII AR5 Chapter 7 (IPCC, 2014), model results confirm the importance of known key physiological processes, such as the shortening of the time to maturity of a crop with increasing mean temperature (Iqbal and Arif, 2010), decline in grain set when high temperatures occur during flowering (Moriondo et al., 2011), and increased water stress at high temperatures throughout the growing cycle (Lobell et al., 2013). Based on many studies, covering a wide range of regions and crops, there is high confidence that the negative impacts of climate change on crop yields have been more common than positive impacts. Climate change has negatively affected wheat and maize yields for many regions (with medium confidence, IPCC (2014)). For the major crops (wheat, rice, and maize) in tropical and temperate regions, climate change without adaptation is projected to

negatively impact production for local temperature increases of 2°C or more above late-20th-century levels, although individual locations may benefit from it. Climate change is thus projected to progressively increase inter-annual variability of crop yields in many regions. These projected impacts will occur in the context of rapidly rising crop demand.

Global demand for agricultural crops is increasing, and may continue to do so for decades, driven by global population increase and greater per capita incomes anticipated through midcentury (Tilman et al., 2011). Both land clearing and more intensive use of existing croplands could partly contribute to the increase in crop production needed to meet such demand, but the environmental impacts and tradeoffs of these alternative paths of agricultural expansion might be devastating. Therefore, it is crucial to understand the environmental impacts of global crop production and to determine how to achieve greater yields with lower impacts to meet the world's future food security and sustainability needs.

The CLIMATOR project (Brisson and Levrault, 2010) was a pluridisciplinary research project carried out from 2007 to 2010 that aimed at evaluating the potential effects of climate changes over the French croplands. According to the project's results a series of positive and negative changes are likely to happen in the agricultural system in France. As positive effects, related to the temperature increase trend, we can cite: the possibilities of development of new croplands in the north of France; a quickening of the phenological cycles, thus partial avoidance of water stress and extreme temperatures for the summer crops; and a reduction of the autumn frost episodes for the winter crops. On the other hand, amongst the main negative effects of climate change, we list: an extension of the intercropping period, increasing the risks of soils leaching, carbon losses and erosion; reduction of the productivity (yield) in case of accentuated water stress for the non-irrigated crops; increase of the crops water requirements, amongst others. Maintaining crop production will require, for most crop species, some adaptations in the management practices (earlier sowing of winter crops, irrigation to ensure a good emergence…).

In France, like in most countries, agriculture is a major user of water resources (representing more than 50% of the total consumed water, up to 80% during summer). At present, the irrigated surface areas cover approximately 2Mha (against 0.8Mha in 1970) and represent 6% of the agricultural croplands. Besides, because some regions are dedicated to the cultivation of certain crop types, an amount of 50% of the irrigated surfaces is concentrated within three sole regions (Aquitaine, Centre and Midi-Pyrénées). As according to the CLIMATOR project, significant changes on the precipitation dynamics are predicted to happen in the southwest France (a 25% precipitation increase during winter and 15% precipitation decrease

during summer toward 2050), the need for irrigation in those three regions will probably increase.

2. Scope and Objectives

Given all those alarming findings, a number of questions are raised over the response of the agricultural systems to all those changes and challenges. “What are the main processes controlling the agro-ecosystems water and carbon cycles? What crop types/ rotations should be favored given the climate changes predictions? What can be done for enhancing the water use efficiency? How to reduce the croplands CO_2 and other GHG emissions? What strategies should be adopted for an optimal management of the lands and of the water resources at local, regional and global scales?” And so on.

In short, some of the great challenges of our time are to maintain an optimal (and sufficient) level of crop productivity, considering the specific characteristics of each region, to take into account the socio-economic conditions and, at the same, to reduce the environmental impacts of the agricultural ecosystems.

In this context, the main objectives of this thesis are 1) to analyze the production and also the carbon and water biogeochemical processes for crop ecosystems in contrasted climatic years, focusing specifically on the winter wheat crop, 2) to model, quantify and evaluate the components of the net carbon and water fluxes of these agro-ecosystems, 3) to compute the water and carbon budgets from the field up to the regional scale; to achieve a deep understanding of the interaction of the climatic and environmental conditions over the crops development and production; and finally, to identity potential solutions for a better management of the agro-ecosystems resources (essentially, in terms of carbon and water contributions) at the local and regional scales.

For these purposes, models that are able to adequately describe the response of agro-ecosystems to different types of management strategies and diverse climatological and environmental conditions are an indispensable tool. The so-called *crop models* should be able to describe the main processes that occur during plant growth and crop production: phenology, photosynthesis, respiration, dry matter production, water consumption, biomass partitioning, solar radiation absorption etc. In this frame, remote sensing may give access to the spatial variability and the time variation of information related either to crop development or crop water status, which may be used for driving crop models (Droogers and Bastiaanssen, 2002; Duchemin et al., 2002).

Remote sensing can significantly contribute to provide a timely and accurate picture of the agricultural production, as it is very suitable for gathering information over large areas with high revisit frequency. For a better understanding of the croplands carbon and water cycles, a great effort has been made for building new models and complete databases that combine satellite data and *in-situ* measurements. The combined use of these three 'tools' opens new perspectives for advanced agro-ecosystems modeling and monitoring over wide surfaces. Generally, the models provide valuable information of the crop characteristics, functioning and development at field scale. Therefore, when combined with the satellite observations, the modeling capacities might be extended to larger scales. Nonetheless, the *in-situ* data (measurements of biomass, yield, leaf area index, CO_2 and water fluxes …) remain essential for further understanding of the plant processes but also for calibrating and validating the models.

In this thesis we propose a regional modeling approach that combines: i) high spatial and temporal resolutions remote sensing data, ii) a simple crop growth model and iii) an extensive set of *in-situ* measurements (for models' calibration and validation), in order to deepen our understanding of the wheat crop ecosystems in southwest France in the perspective of a sustainable agricultural development system, that, in the future, could be transposed to other regions in the world. Indeed, the new upcoming space missions (such as SENTINEL satellites) provide good perspectives for attaining these purposes, since their increased swath width along with the short revisit time and number of available spectral bands would allow monitoring rapid changes in land surfaces (such as vegetation during the growing season) at fine spatial resolution and over large scales. Finally, with our approach we expect to provide an efficient tool that benefits of the whole potential offered by the future satellite observations for investigating and monitoring the crops development and for quantifying and evaluating the production, water and carbon budgets of these croplands at regional/global scales.

3. Outline of the thesis

The core of this thesis is composed of six main chapters. Each chapter is introduced below by stating its main objectives and by outlining its link with the others.

The first chapter defines the key processes related to the carbon and water biogeochemical cycles over croplands and then cites some of the methods used for estimating the components of the carbon and water cycles. Amongst these methods, a special focus is put on the crop modeling coupled to remote sensing products. In

the sequence, the basic principles about optical remote sensing and its potential for analyzing the vegetation dynamics are introduced.

Chapter 2 describes firstly the general characteristics of the study area. Next, the remote sensing data set and its main features are presented. Then, the ensemble of *in-situ* data, together with the protocols used for collecting those data, are described. The data presented in this chapter are used in the further sections for driving, calibrating and/or validating the developed approaches.

Chapter 3 presents the methods used for retrieving the biophysical variable green area index (GAI) from satellite images. The two investigated approaches (empirical and physical-based) are presented, followed by their validation by means of the presented *in-situ* data sets. The GAI maps obtained through this work are further employed for the regional crop modeling approaches.

Chapter 4 contains the description of the crop models used and developed during the thesis and also the approaches adopted for calibrating these models. At first, the original crop SAFY model is described, which is capable of simulating the plant growth dynamics (GAI) and production (biomass and yield). Then, the newly developed SAFY-CO_2 model is presented. In addition to GAI and crop production, SAFY-CO_2 also simulates the main components of the ecosystem carbon balance. Going further, the chapter is concluded by the presentation of the SAFYE-CO_2 model, which couples SAFY-CO_2 with a simplified water module, based on the FAO-56 method, allowing to also estimate the main components of the water cycle and a better estimate of soil respiration as well. By using a simplified modeling approach as support, this chapter demonstrates how a number of variables can be differently estimated depending on the available inputs and on the researchers' major objectives.

Chapter 5 contains the main results relative to the three studied models. Results of simulated GAI dynamics, biomass estimates, yield production, net carbon fluxes components (photosynthesis, ecosystem respiration), net ecosystem carbon budgets, crop evapotranspiration and soil water contents are compared against *in-situ* data over the study area. To conclude, an inter-comparison of the models results is performed and the main conclusions of the chapter are drawn.

Finally, Chapter 6 concludes this thesis by summarizing and discussing the main results from the previous chapters, by raising new questions and suggesting potential improvements for future researches in relation to our approach.

Introduction Générale

1. Contextes général et scientifique

Le climat et les changements d'affectation des terres sont des problèmes majeurs qui affectent la société mondiale et la vie toute entière sur terre. D'après les récents rapports du Groupe d'experts Intergouvernemental sur l'évolution du Climat (GIEC, IPCC en anglais), «le changement climatique fait référence à une modification de l'état du climat, pouvant être identifiée par des changements de la moyenne et/ou de la variabilité de ses propriétés, et qui persiste dans le temps, au moins des décennies. Le changement climatique peut être dû à des processus naturels internes ou au forçage radiatif externe tel que les variations des cycles solaires, les éruptions volcaniques, et les changements anthropogéniques pérennes de la composition de l'atmosphère ou de l'utilisation des sols ». Les scientifiques ont étudié le sujet durant des années et ont démontré l'influence humaine sur le système climatique. Dans les dernières décennies, les changements liés au climat et à l'affectation des terres ont touché l'humanité et l'environnement sur tous les continents et dans tous les océans. Les émissions continues de gaz à effet de serre engendreront toujours plus de changements dans toutes les composantes du système climatique (IPCC, 2014) et redistribueront spatialement la végétation naturelle et cultivée. Le changement climatique est donc un problème global et sa prévention dépend de réductions conséquentes et continues des émissions de gaz à effet de serre. L'IPCC a été créée dans ce contexte en 1988 par l'Organisation météorologique mondiale (WMO en anglais) et le Programme environnemental des Nations Unies (UNEP en anglais) afin de fournir aux décideurs des évaluations régulières sur des bases scientifiques du changement climatique, ses impacts et risques futurs, et des stratégies réalistes pour s'y adapter et le juguler [http://www.ipcc.ch/]. L'IPCC a publié quatre rapports d'évaluation, la version finale du cinquième devant être disponible avant la fin de l'année 2014. Ces rapports soulignent que l'importance du climat et des changements d'utilisation des sols nécessitent une coopération internationale. Ils ont par exemple mené à la création à la Convention-cadre des Nations-Unies sur le changement climatique (UNFCCC en anglais) et ont encouragé l'adoption du protocole de Kyoto en 1997. Les quatre rapports d'évaluation publiés en 2007 ont mis en exergue le lien étroit entre l'augmentation des températures moyennes globales observées depuis la deuxième moitié du XXe siècle et les concentrations de

gaz à effet de serre anthropogéniques. Le cinquième rapport de l'IPCC (AR5) offre une évaluation plus complète de la connaissance scientifique sur le changement climatique depuis 2007, et pointe les aspects socio-économiques du changement climatique et ses implications pour le développement durable.

Les concentrations atmosphériques en gaz à effet de serre (dioxyde de carbone (CO_2), méthane (CH_4) et oxyde d'azote (N_2O)) ont augmentées depuis 1750 à cause des activités humaines. Les concentrations de CO_2 ont augmentées de 40% depuis l'ère préindustrielle (IPCC, 2014a, voir Figure 1), d'abord à cause des émissions fossiles, puis des émissions nettes dues aux changements d'affectation des terres. L'océan a absorbé environ 30% des émissions anthropogéniques de dioxyde de carbone, entrainant l'acidification des océans.

Le dioxyde de carbone anthropogénique représente la composante la plus dynamique du cycle global du carbone. Le cycle global du carbone peut être vu comme une série de réservoirs de carbone dans le système de la planète, connectés par des échanges de flux de carbone. La figure 2 illustre les perturbations du cycle global du carbone causées par l'homme, moyennée sur la décennie 2003-2012.

En plus du cycle biogéochimique du carbone, le cycle de l'eau (mais aussi les cycles de l'azote, du phosphore, etc) est aussi affecté par le changement climatique. D'après les études analysées par les membres de l'IPCC, les risques liés à la disponibilité de l'eau augmentent de manière significative avec les concentrations des gaz à effet de serre. Il est communément admis que le changement climatique participera grandement, au XXIe siècle, à la réduction des ressources en eaux de surface renouvelables et en eaux souterraines dans la plupart des régions arides subtropicales, alimentant la compétition pour l'accès à l'eau. Dans beaucoup de régions, les changements de précipitation et la fonte des neiges et des glaces peuvent modifier les systèmes hydrologiques, entrainant une baisse de qualité et de quantité des ressources en eau (IPCC, 2014a).

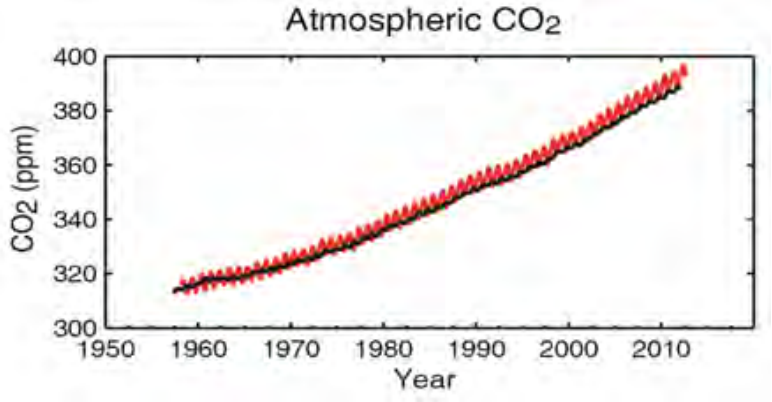


Figure 1- Concentration atmosphérique du dioxyde de carbone (CO_2), issu de Mauna Loa (19°32′N, 155°34′W – rouge) et du Pôle Sud (89°59′S, 24°48′W – noir) depuis 1958, extrait de (Ciais et al., 2013).

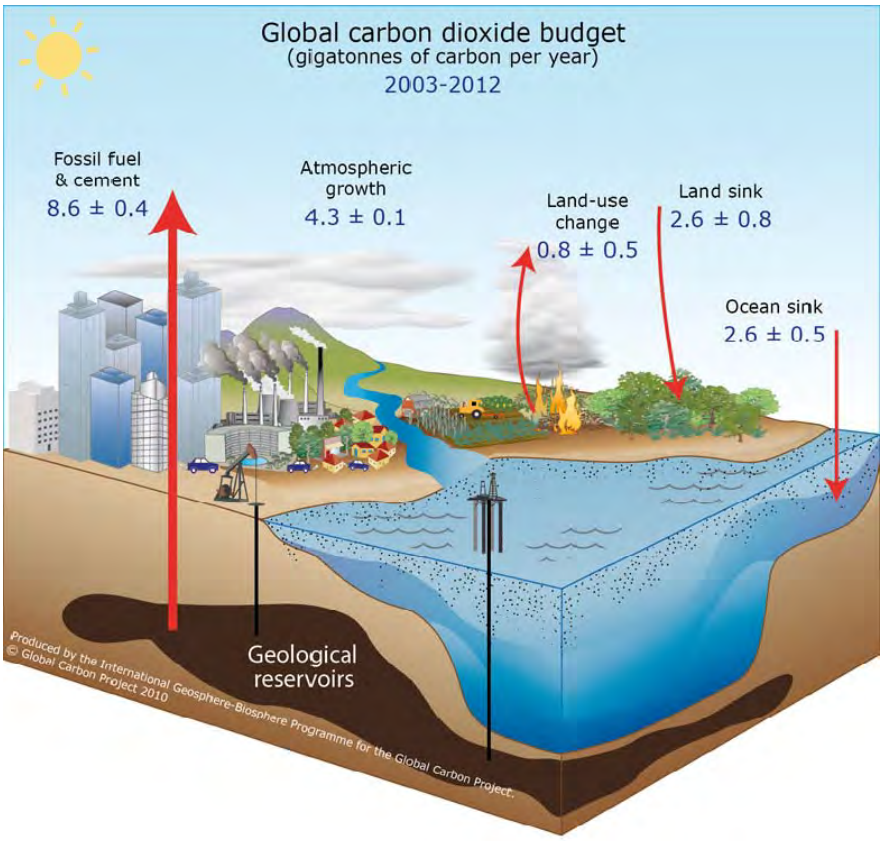


Figure 2- Représentation schématique des perturbations du cycle global du carbone causées par les activités humaines, moyennée sur la décennie 2003-2012. Les flèches représentent les émissions dues à la consommation des énergies fossiles et à la production de ciment ; émissions dues à la déforestation et autres changements d'affectation des sols ; et les puits de carbone de l'atmosphère aux océans et réservoirs terrestres. L'augmentation annuelle du dioxyde de carbone dans l'atmosphère est aussi soulignée. Toutes les unités de flux sont exprimées en $PgC.an^{-1}$. Source: Le Quéré et al., (2013); CDIAC Data; NOAA/ESRL Data; Global Carbon Project 2013.

Le secteur agricole subit directement les conséquences du changement climatique, car la production est conditionnée par les conditions climatiques et les ressources en eau. De plus, les pratiques agricoles participent à l'augmentation des émissions anthropogéniques des gaz à effet de serre. D'après la Food and Agriculture Organisation (FAO), des Nations Unies, les cultures agricoles couvrent 1.53 milliards d'hectares (environ 12% de la terre libre de glace) et les pâturages occupent 3.38 milliards d'hectares (environ 26% de la terre libre de glace). Au total, l'agriculture monopolise 38% de la surface terrestre de la planète (Foley et al., 2011). De plus, le rapport de l'IPCC affirme que les secteurs de l'agriculture, des forêts et les changements de l'occupation des sols comptent pour environ un quart des émissions nettes de gaz à effet de serre anthropogéniques (environ 10-12 Gt $CO_2eq.an^{-1}$), principalement à cause de la déforestation, des émissions agricoles provenant du

sol, des engrais et du bétail. (IPCC, 2014b). Pourtant, les incertitudes liées aux émissions de gaz à effet de serre du secteur agricole sont élevées et le potentiel de stockage de carbone de l'atmosphère dans les terres agricoles sous la forme de matière organique pourrait être un levier puissant pour juguler le changement climatique.

Afin d'évaluer la sensibilité de la production de nourriture par rapport au temps et au climat, de nombreux modèles ont été utilisés pour évaluer la réponse du rendement agricole aux températures. D'après l'IPCC WGII AR5, les résultats de modélisation confirment l'importance des processus physiologiques clé connus, tel que la réduction du temps de maturité avec l'accroissement de la température moyenne (Iqbal and Arif, 2010), la diminution de la quantité de grains lors d'épisodes de températures élevées pendant la floraison (Moriondo et al., 2011), et l'augmentation du stress hydrique durant le cycle de croissance quand les températures sont hautes (Lobell et al., 2013). D'après de nombreuses études effectuées sur un large panel de zones géographiques et de cultures agricoles, il est fort probable que le changement climatique a eu sur les champs agricoles des effets plus néfastes que positifs. Le changement climatique a eu probablement un impact négatif sur les rendements de blé et de maïs (IPCC, 2014a). Sans aucune adaptation des cultures majeures (blé, riz et maïs) des zones tropicales et tempérées, le changement climatique impactera négativement la production, pour une augmentation des températures de 2^{o}C ou plus par rapport aux niveaux de températures de la fin du XXe siècle. Le changement climatique risque aussi d'augmenter progressivement la variabilité interannuelle des rendements des cultures dans de nombreuses régions. Ces impacts prévus arriveront dans le contexte d'une demande alimentaire toujours croissante.

La demande globale en production agricole augmente maintenant et continuera sûrement à augmenter pendant des décennies, à cause d'une population globale et des revenus par tête croissants jusqu'au milieu du siècle (Tilman et al., 2011). Le défrichement des terres et une utilisation plus intensive des cultures existantes pourraient compenser le besoin d'une production toujours plus élevée, mais les impacts environnementaux et les compromis de cette expansion seraient dévastateurs pour l'environnement. Il est donc crucial de comprendre les impacts environnementaux de la production globale des cultures et de trouver comment atteindre des rendements plus importants tout en minimisant ces impacts afin d'obtenir la sécurité alimentaire future durable recherchée.

Le projet CLIMATOR (Brisson et Levrault, 2010), était entre 2007 et 2010 un projet de recherche pluridisciplinaire dont le but était d'évaluer les effets potentiels du changement climatique sur les cultures françaises. D'après les résultats de ce projet, une série de changements positifs et négatifs risque de toucher le système agricole

en France. On peut citer, comme effets positifs liés à la tendance croissante des températures : les possibilités de développement de nouvelles cultures dans le nord de la France, des cycles phénologiques plus rapides, donc un évitement plus probable de stress hydrique et de températures extrêmes pour les cultures d'été ; et une réduction des épisodes de gel d'automne pour les cultures d'hiver. D'un autre côté, nous pouvons citer les principaux effets négatifs du changement climatique : une extension de la période inter-culturale, un accroissement des risques de lessivage des sols, de pertes de carbone et d'érosion ; une réduction de la productivité en cas de stress hydrique important pour les cultures non irriguées ; une augmentation des besoins en eau, parmi d'autres impacts. Le maintien de la production agricole nécessitera pour la plupart des espèces cultivées des adaptations des pratiques agricoles (semer plus tôt les cultures d'hiver, une meilleure irrigation…).

En France comme dans la plupart des pays, l'agriculture a des besoins en eau considérables (plus de 50% de la totalité de l'eau consommée, environ 80% en été). A présent, les zones irriguées couvrent 2Mha (contre 0.8 Mha en 1970) et représentent 6% des cultures agricoles. De plus, 50% des surfaces irriguées sont concentrées dans trois régions seulement (Aquitaine, France et Midi-Pyrénées) puisque ces régions sont consacrées à certaines cultures en particulier. D'après le projet CLIMATOR, d'importants changements des dynamiques de précipitations sont prévus dans le Sud-Ouest français (une augmentation des précipitations de 25% en hiver et une baisse de 15% en été d'ici 2050). Le besoin en irrigation dans ces régions augmentera probablement.

2. Objectifs

Étant donné tous ces facteurs alarmants, de nombreuses questions se posent quant à l'adaptation des systèmes agricoles face à ces défis. « Quels sont les processus principaux qui contrôlent les cycles de l'eau et du carbone des agro-systèmes ? Quelles rotations de cultures devraient être favorisées au vu des prédictions de changement climatique ? Comment peut-on améliorer l'efficacité de l'utilisation de l'eau ? Comment réduire les émissions de CO_2 des cultures et les autres émissions de gaz à effet de serre ? Quelles stratégies devraient être adoptées pour une gestion optimale des terres et des ressources en eau aux niveaux local, régional et global ? ».

Pour résumer, les grands défis de notre temps sont : 1) de maintenir un niveau optimal et suffisant de la productivité des cultures, selon les caractéristiques

spécifiques de chaque région, 2) de prendre en compte les conditions socio-économiques et, en même temps, 3) de réduire les impacts environnementaux sur les écosystèmes agricoles.

Dans ce contexte, les objectifs principaux de cette thèse sont d'analyser la production et les cycles biogéochimiques du carbone et de l'eau pour les écosystèmes agricoles pendant des années de climat contrasté, avec une attention particulière sur la culture hivernale du blé ; de modéliser, quantifier et évaluer les composantes des flux nets de carbone et d'eau de ces écosystèmes ; de calculer les bilans de carbone et d'eau de l'échelle de la parcelle à l'échelle régionale ; d'apporter une compréhension profonde de l'interaction entre les conditions environnementales et le climat dans un contexte de développement et de production des cultures ; et finalement, d'identifier des solutions potentielles pour une meilleure gestion des ressources des agro-écosystèmes (essentiellement pour le carbone et l'eau) à l'échelle locale et régionale.

Pour ce faire, les modèles capables de décrire de façon adéquate la réponse des agro-écosystèmes à différentes stratégies de gestion et diverses conditions environnementales et climatologiques, sont un outil indispensable. Ces modèles devraient pouvoir décrire les principaux processus lors de la croissance et la production agricole : phénologie, photosynthèse, respiration, consommation d'eau, la production de la biomasse, l'absorption des radiations solaires etc. Dans ce cadre, la télédétection peut offrir des informations concernant la variabilité spatiale et temporelle de l'information liée au développement des cultures comme à l'état hydrique de la culture, pouvant être utilisées en assimilation dans les modèles de cultures (Droogers and Bastiaanssen, 2002; Duchemin et al., 2002).

La télédétection, capable de fournir des informations à large échelle avec une fréquence importante de revisite, peut grandement contribuer à décrire avec précision et dans le temps l'état de production des cultures. Afin de mieux comprendre les cycles du carbone et de l'eau des cultures un effort considérable a été fourni pour développer des nouveaux modèles et constituer des bases de données complètes combinant des données satellitaires et des mesures *in-situ*. L'utilisation conjointe de ces trois outils ouvre de nouvelles perspectives pour la surveillance et la modélisation des systèmes agricoles sur de grandes surfaces. En général, les modèles fournissent des informations utilises sur le développement, le fonctionnement et les caractéristiques des cultures à l'échelle de la parcelle. Ainsi, les capacités des modèles peuvent être élargies à grande échelle en utilisant conjointement des données satellitaires. Néanmoins, les données *in-situ* (mesures de biomasse, de rendement, d'indice de surface foliaire, de flux de CO_2 et d'eau) restent essentielles pour une meilleure compréhension des processus de la plante et du sol mais aussi pour la calibration et la validation des modèles.

Dans cette thèse, nous proposons une approche de modélisation régionale qui combine : i) des données de télédétection à résolutions temporelle et spatiale élevées, ii) un modèle de croissance simple des cultures, et iii) un grand jeu de mesures *in-situ* (pour la calibration et la validation des modèles), dans le but d'améliorer notre compréhension des cultures de blé dans le sud-ouest de la France, pour le développement durable d'un système agricole, qui, dans le futur, pourra se transposer à d'autres régions du monde. En effet, les nouvelles missions spatiales à venir (comme les satellites SENTINEL) proposent des perspectives intéressantes pour atteindre ces objectifs, comme leur fauchée élargie associée à une courte période de revisite et de nombreuses bandes spectrales permettra le suivi des changements terrestres rapides (comme la végétation durant la période de croissance) à une résolution spatiale fine et à grande échelle. Finalement, nous espérons fournir avec cette approche un outil efficace bénéficiant de tout le potentiel offert par les futurs satellites pour le suivi des cultures et pour l'évaluation de la production, des bilans de carbone et d'eau de ces cultures à l'échelle régionale et globale.

3. Plan de la thèse

Cette thèse s'articule autour de six chapitres principaux. Chaque chapitre est introduit ci-dessous en précisant ses objectifs principaux et en soulignant ses liens avec les autres chapitres.

Le premier chapitre définit les processus clé liés au cycle biogéochimique du carbone et de l'eau sur les cultures et aborde ensuite certaines méthodes utilisées pour l'estimation des composantes des cycles du carbone et de l'eau. Parmi ces méthodes, l'accent est mis sur les modèles de cultures couplés aux produits de télédétection. Les principes de base de la télédétection optique et son potentiel pour l'analyse des dynamiques de la végétation sont introduits.

Le deuxième chapitre décrit d'abord de manière générale de la zone d'étude. Puis, les données de télédétection et leurs principales caractéristiques. Ensuite, l'ensemble des données *in-situ* ainsi que les protocoles mis en place pour les collecter sont présentés. Les données décrites dans ce chapitre sont utilisées dans les chapitres suivants, pour le développement, la calibration et/ou la validation des approches.

Le chapitre 3 présente les méthodes utilisées pour l'estimation de l'indice de surface verte (GAI en anglais) à partir des images satellites. Les deux approches étudiées, physique et empirique, sont expliquées et validées au moyen des données *in-situ*. Les cartes de GAI obtenues sont ensuite employées pour la modélisation spatialisée des cultures.

Le chapitre 4 aborde la description des modèles de culture utilisés et développés durant cette thèse, ainsi que les approches retenues pour la calibration de ces modèles. D'abord, le modèle SAFY, capable de simuler les dynamiques de croissance de la plante (GAI) et la production (biomasse et rendement), est décrit. Ensuite, le nouveau modèle SAFY-CO_2 développé durant cette thèse est présenté. En plus du GAI et de la production, le modèle SAFY-CO_2 simule les principales composantes du bilan de carbone. Afin d'aller plus loin, le chapitre est conclut par la présentation du modèle SAFYE-CO_2, résultant du couplage de SAFY-CO_2 avec un module d'eau simplifié, basé sur la méthode FAO-56. Ce modèle permet aussi une meilleure simulation de la respiration du sol et une estimation des principales composantes du cycle de l'eau. En utilisant une approche de modélisation simplifiée comme support, ce chapitre démontre comment certaines variables peuvent être différemment estimées en fonction des entrées du modèle et des objectifs principaux du chercheur.

Le chapitre 5 contient les résultats principaux, relatifs aux trois modèles étudiés. Les résultats de simulations, donc les dynamiques de GAI, les estimations de biomasse, le rendement, les composantes du flux net de CO_2 (photosynthèse, respiration), les bilans nets de carbone, l'évapotranspiration de la culture et la teneur en eau du sol sont comparées aux données *in-situ* sur la zone d'étude. Pour conclure, une inter-comparaison des résultats des modèles est effectuée, et les principales conclusions sont énoncées.

Finalement, le chapitre 6 conclut cette thèse en résumant et discutant les résultats principaux des chapitres précédents. De nouvelles questions se posent alors, et des améliorations potentielles pour les recherches futures liées à notre approche sont proposées.

1. Study of the carbon and water biogeochemical cycles over croplands

This chapter presents the definition of the key processes related to the carbon and water biogeochemical cycles for crops. The main existing methods and approaches for measuring/estimating the components of these cycles are described, including *in-situ* devices, modeling and modeling coupled with remote sensing products. In a second part, a brief introduction to the conceptual basis of optical remote sensing and its link with characteristics of vegetation properties are presented. The methods for retrieving biophysical variables from the remote sensing observations are introduced and the link between these variables and the canopy intrinsic processes is outlined, by the means of the crop ecosystem models.

1.1. The water and carbon biogeochemical cycles: definitions, measurements methods and modeling

1.1.1. The water cycle in agriculture

Global population is increasing and so the water needs in different sectors (agriculture, industry, domestic) rise as well, requiring a wise use of this limited resource. Agricultural crops use a lot of water, mostly when irrigated, which makes it crucial to improve the agricultural water management.

An efficient agricultural water management requires reliable estimation of crop water requirement (evapotranspiration). The plant roots extract water from the soil to live and develop. The main part of this water does not remain in the plant, but escapes to the atmosphere through the plant's leaves and stems. This process is called transpiration. Some water also evaporates from soil surface, which is denominated as evaporation (Allen et al., 1998). Thus evapotranspiration (ETR) can be defined as the transfer of water from the soil surface (evaporation) and plants (transpiration) to the atmosphere. ETR is a critical component of the water balance at plot, field, farm, catchment, basin or global level (Zeleke and Wade, 2012). The crop water requirements mostly depend upon the climate (radiation, temperature), the soil type, the crop type (*e.g.* maize needs more water than wheat) and the growth stage (well developed crops need more water than crops that have just been planted).

The water budget is calculated in order to estimate the evolution of the soil water content over a given period of time. In the croplands context, the water budget is basically the difference between the effective amount of rainfall and irrigation received by the crop and the amounts of water lost by the crop and soil due to evaporation, transpiration, runoff and deep infiltration. A scheme of the crop water cycle is presented in Figure 1.3.

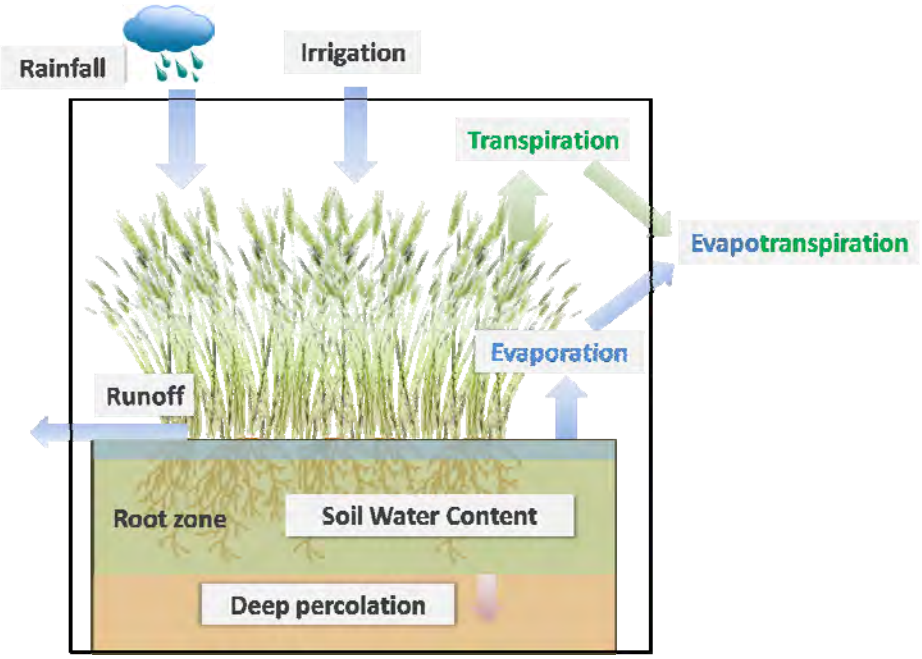


Figure 1.3- Main water cycle components at the crop field scale.

1.1.2. Approaches for estimating evapotranspiration

There are different methods for estimating ETR. A number of techniques may be used for measuring, separately or jointly, the components E and TR of an ecosystem. In agronomy, the lysimeters devices have been used for a long time (Qiu et al., 2008; Steiner and Hatfield, 2008) for measuring E or ETR. However, they may disturb significantly the soil when they are installed, making its behavior different inside and outside of these devices. The sap flow methods can be used to estimate the total transpiration of whole plants (Granier et al., 1996; Roupsard et al., 2006), including crop plants (Bethenod et al., 2000). Flux measurements by Eddy-Covariance (EC) are commonly used for a range of ecosystems for measuring ETR at the field scale. In some studies, isotopic measurement approaches were combined with EC methods for estimating the components E and TR separately over forest canopy (Williams et al., 2004). In other studies, two different levels of EC measurements were used to calculate the TR and the water use efficiency of the forest canopy regardless the underbrush beneath the canopy (Jarosz et al., 2008; Lamaud et al., 1996; Roupsard et al., 2006). However, this technique is not appropriate for croplands where the crop height is low. Béziat et al., (2013) have developed a simple statistical approach for partitioning those two components for crops based on EC and micro-meteorological measurements. Furthermore, TR can be also estimated from gas

exchanges measurement at the leaf level (Medrano et al., 2009; Steduto and Albrizio, 2005; Steduto et al., 1997).

In addition to the experimental methods, different modeling approaches are frequently used for simulating ETR and its components. Some of these approaches are characterized as empirical (Li et al., 2008; Ritchie, 1972), others are more complex, being designated as mechanistic, which are often based upon the so-called SVAT (Soil Vegetation for Atmosphere Transfer) models. The SVAT models may represent the land surface as a unique layer (soil + plant together, (1996; Noilhan and Mahfouf, 1996; Noilhan and Planton, 1989)), as two-layers (soil and plant separately, (Hu et al., 2009; Sellers et al., 1997; Shuttleworth and Wallace, 1985)) , as three-layers (bare soil, shaded soil and plant,(Boulet et al., 1999)) or even as multi-layer system (Ogée et al., 2003). The two or more layers models have a more realistic representation of the energy/water balance and can describe, separately, the contributions of the soil and the vegetation to the ETR. These models often require an important number of input parameters, which makes them more difficult to calibrate.

1.1.2.1. ETR estimation: modeling and remote sensing

The main methods classically used to measure ETR (above cited) are available at the field scale, but do not allow estimating the fluxes when dealing with large spatial scales (Courault et al., 2005).

Different methods have been developed for benefiting of the (spatial and temporal) information provided by remote sensing products for estimating ETR at regional scales. For this purpose, satellite optical and thermal infrared imagery can be a useful tool (Courault et al., 2005; Kite and Droogers, 2000; Van Niel and McVicar, 2004). The ecosystem models including water transfer in the soil–vegetation–atmosphere system, when linked with remote sensing data, may provide estimates of the water and energy balances on continuous regular space grids (Duchemin et al., 2006; Moulin et al., 1998; Olioso et al., 2005). This approach is often based on the combination of inversion, calibration and assimilation techniques, making them complex and difficult to be used and validated in an operational context (Combal et al., 2003; Olioso et al., 1999; Verhoef and Bach, 2003). At present, the FAO Penman-Monteith approach is considered as a standard method for ETR estimation in agriculture (Allen et al., 1998), being widely used for operational applications, especially by water resource managers. It is based on the concepts of reference evapotranspiration (ET_0) and crop coefficients (K_C), which have been introduced to separate the climatic demand from the plant response (Allen et al., 1998). According to this method, the reference surface is a well-watered green grass of uniform height and infinite extent, which is actively growing and completely

shading the ground (FAO-56 method). Therefore, to avoid ambiguous methods, the FAO recommends to estimate the reference evapotranspiration ET_0 using the Penman–Monteith equation (Monteith, 1985; Penman, 1948), with the parameterization elaborated by Allen et al., (1998). This method was adopted for this thesis; the ET_0 formula is described in Eq. (1.1). The equation uses standard climatic records of solar radiation (sunshine), air temperature, humidity and wind speed. Thereby, ET_0 is a climatic parameter expressing the evaporative demand of the atmosphere independent of crop type, crop development and management practices (Allen et al., 1998).

$$ET_0 = \frac{0.408\,\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34 u_2)} \quad (1.1)$$

where ET_0: reference evapotranspiration [mm day^{-1}],

R_n : net radiation at the crop surface [MJ m^{-2} day^{-1}],

G : soil heat flux density [MJ m^{-2}.day^{-1}],

T : mean daily air temperature at 2 m height [°C],

u_2 : wind speed at 2 m height [m s^{-1}],

e_s : saturation vapour pressure [kPa],

e_a : actual vapour pressure [kPa],

e_s-e_a : saturation vapour pressure deficit [kPa],

γ : slope of the vapour pressure curve [kPa °C^{-1}],

The evapotranspiration from a given crop grown and managed under standard conditions is called potential crop evapotranspiration (Etc). Standard condition is a disease-free, well fertilized crops, grown in large fields, under optimum soil water conditions, and achieving full production under the given climatic conditions (Zeleke and Wade, 2012). ETc represents the climatic “demand” for water by a given crop. The crop coefficient approach relates evapotranspiration from a reference crop surface (ET_0) to evapotranspiration from a given crop (ETc) through a crop coefficient Kc [Eq. (1.2)].

The actual crop evapotranspiration (ETR) corresponds to the real water consumption according to weather parameters, crops factors, management and environmental conditions (Courault et al., 2005). Nevertheless, several other crop and surface characteristics should also be considered, such as crop type and variety, crop stage development and root system development.

Thereby, crop evapotranspiration is calculated by multiplying ET_0 by a crop coefficient, expressing the difference in evapotranspiration between the actual cropped and reference grass surfaces. This coefficient is expected to comprise all differences in the physio-morphology between the crop in its actual status and the ideal reference grass (Duchemin et al., 2006). The differences can be combined into one single coefficient, or it can be split into two factors describing separately the differences in evaporation and transpiration between both surfaces. The choice of the approach depends on the user's objectives, the accuracy required and the climatic data available.

$$ETc = Kc \times ET_0 \tag{1.2}$$

$$ETR = Ks \times Kc \times ET_0 \tag{1.3}$$

$$ETR = \underbrace{Ke \times ET_0}_{E} + \underbrace{Kcb \times ET_0}_{T} \tag{1.4}$$

In the single crop coefficient approach, the effect of crop transpiration and soil evaporation are combined into a single coefficient. This coefficient includes a crop coefficient (K_c) and a stress coefficient (K_s, Eq. (1.3)). In the dual crop coefficient approach, the crop transpiration and soil evaporation are determined separately. Two coefficients are used: the basal crop coefficient (K_{cb}) to describe plant transpiration, and the soil water evaporation coefficient (K_e) to describe evaporation from the soil surface. The single K_c coefficient is thus replaced by the sum of K_{cb} and K_e [Eq. (1.4)]. The basal crop coefficient, K_{cb}, is defined as the ratio of ETR to ET_0 when the soil surface layer is dry but when the average soil water content of the root zone is adequate to sustain full plant transpiration. The soil evaporation coefficient, K_e, describes the evaporation component from the soil surface and depends upon the top-layer soil water content and the fraction of vegetation cover (Fcover) (Allen et al., 1998). Thus, for this approach the estimation of the coefficients K_{cb} and K_e is directly linked to two biophysical variables: the green plant area index (GAI) and the Fcover.

The use of remotely-sensed vegetation indices, such as the normalized difference vegetation index (NDVI; Rouse Jr et al., (1974)) or the soil-adjusted vegetation index (SAVI; Huete, (1988)), has been tested to predict crop coefficients at field and regional scales (Bausch and Neale, 1987; Choudhury et al., 1994; Duchemin et al., 2006). In most of the cases, the vegetation index derived from satellite data are used for providing estimates of GAI and FCover, which are afterward used for estimating crop coefficients K_{cb} and K_e. The methods for retrieving GAI from satellite data will be described in Chapter 3 and the relationships linking GAI and FCover, and then

between them and the coefficients K_{cb} and K_e will be presented in section 4.3.1 with the SAFYE-CO_2 model description.

1.1.3. The carbon cycle in agriculture

The fixation of atmospheric carbon dioxide (CO_2) by plant photosynthesis is a key process for the functioning of the crop ecosystems. Photosynthesis process absorbs energy from the sun, releases oxygen to the atmosphere and absorbs CO_2. The rate at which this last process occurs is termed as gross primary production (GPP). Subtracting the plant's day and night respiratory losses (autotrophic respiration) yields to net primary production (NPP), or the net flux at which carbon is stored as biomass. It can also be referred as the plant productivity. The NPP measures accumulation of biomass in any part of the plant, including root systems. Integration over the ecosystem and subtraction of additional respiratory losses (heterotrophic respiration) yields to the net ecosystem exchange (NEE). The NEE is therefore the net uptake or release of carbon influenced by climatic and by non-climatic factors, like the plant water supply, leaf area index, soil carbon dynamics, but it also depends upon crop type and management practices. When the NEE term is negative, the ecosystem is considered as an atmospheric CO_2 sink, and if NEE is positive, the ecosystem is considered as a CO_2 source.

The cited processes constitute the *natural* components of the carbon cycle. For the carbon budget assessment at the crop ecosystem level, we also consider the anthropogenic contribution, corresponding to carbon import (mainly as organic fertilization and seeds) and export (yield exported out of the field) terms. As it is illustrated in Figure 1.4, these are the key elements of the carbon budget at the crop field level. For estimating the total cropland carbon budget other carbon gains and losses processes should be taken into account, such as erosion, emission of volatile organic compounds and microbially-produced methane (CH_4; see (Smith et al., 2010) for a description of all the components of the croplands carbon budget and the methods used to estimate them). Nevertheless, these additional terms will not be considered by our (simplified) approach.

Furthermore, the storage or destocking of organic carbon in agricultural soils (soil organic carbon, SOC) is an important issue related to climate and land use changes management. Soil organic matter is one of the largest carbon reservoirs that is in rapid exchange with atmospheric CO_2, and it is thus important as a potential greenhouse gases source or sink. The loss of organic matter content may have negative effects over the physicochemical properties of the soil, which could result in a decrease in fertility, biodiversity, greater susceptibility to erosion and soil compaction... Indeed, Soltner (2006) has stated that, from an agronomical point of

view, its reduction would cause an inevitable degradation in soil fertility, even if optimal mineral fertilization was accomplished.

Several studies have been conducted to assess the potential storage of atmospheric carbon in agricultural soils. In France, the maximum potential for additional carbon storage in the soil could be of the order of 1 to 3 million tonnes per year according to a research conducted by INRA, in 2002. This storage would counterbalance up to 3-4 % of the annual GHG emissions in France, which is not negligible. Changes in land use and changes in agricultural practices (reduction of soil labor and tillage, etc.) would allow increasing the content of the soil carbon reservoirs and also reducing the contribution of agriculture to global warming. In this context, it can be noted that it is essential defining efficient strategies for maintaining (or increasing) the content of organic carbon in the soil reservoirs, since they play an important role on both climate and agricultural (crops production) issues.

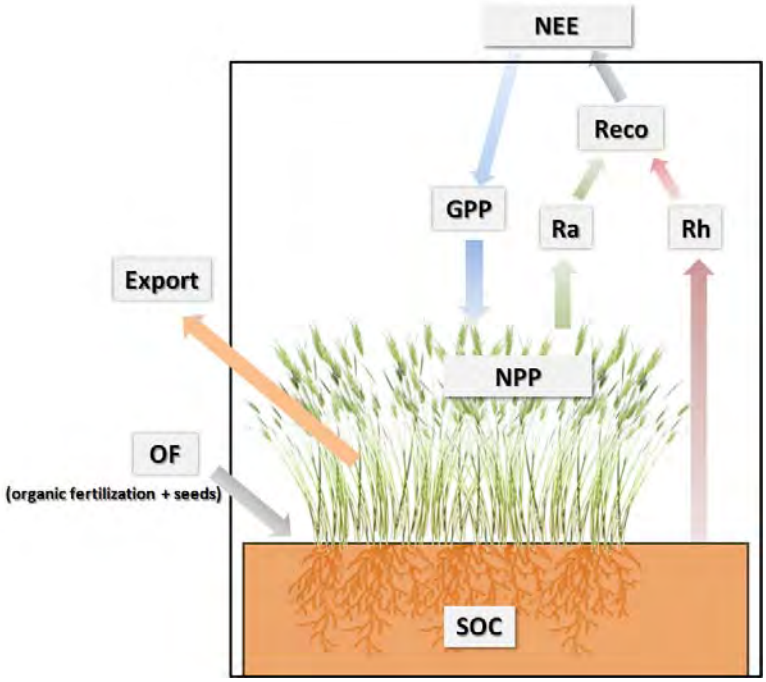


Figure 1.4- Main carbon cycle components at the crop field scale.

1.1.4. Approaches for estimating the components of the net CO_2 fluxes

1.1.4.1. *In-situ* measurements of the NEE components

Carbon fluxes studies based upon *in-situ* measurements are of great importance since they allow quantifying and understanding the impact of natural and anthropogenic processes over its biogeochemical cycle.

Different methods exist for measuring the components of the net CO_2 fluxes. Flux chambers can be used for quantifying CO_2 fluxes between soil or vegetation and atmosphere. These chambers are usually connected to infrared analyzers for measuring the plants respiration, in either manual (Damesin et al., 2002; Maier et al., 1998; Stockfors and Linder, 1998) or automatic (Ceschia et al., 2002) systems. In the same way, manual or automatic measurements of soil respiration were performed over a number of ecosystems in order to assess the total respiration (Rayment and Jarvis, 2000) or its autotrophic and heterotrophic components (Epron et al., 2001). Moreover, some chambers can be used to measure the leaf or branches photosynthesis (Dufrêne et al., 1993; Hoyaux et al., 2008; Wallin et al., 2001). These systems provide information about how the plant/soil elements respond to different environmental conditions (climate, nutrition, CO_2 concentration ...) and are very useful for the parameterization/validation of ecophysiological models.

Another valuable method for evaluating the net ecosystem CO_2 flux is the eddy covariance (EC) measurement technique. The need of measuring the carbon, water and energy exchanges between the soil-canopy and the atmosphere over a wide range of ecosystems has resulted in the establishment, since the late 90s, of EC flux measurements networks at regional (*e.g.* FLUXPYR), continental (Asiaflux, CarboAfrica, Asiaflux, Euroflux ...) and international (FLUXNET, Baldocchi et al.,(2001)) levels.

One of the main objectives of these networks is to create a homogenous database, for facilitating the analysis of the terrestrial ecosystems functioning in response to climate and anthropogenic events. Moreover, they aim at using the flux measurement data to develop, test and validate ecosystems models. Then, these models could be used to test the effect of different climatic and management changes over the ecosystems. These flux databases could also be used to assess regional and continental flux estimates, when coupled with modeling and remote sensing devices.

In general, the study of fluxes by EC over croplands is rather recent. The first studies were essentially focused on the seasonal cycles of net CO_2 and water fluxes and carbon budgets for maize/soybean crop rotations (Bernacchi et al., 2005;

Hollinger et al., 2005; Pattey et al., 2002; Suyker and Verma, 2009). Later, rice (Saito et al., 2005), sugar beet (Moureaux et al., 2006), wheat and triticale (Ammann et al., 1996; Baldocchi and Collineau, 1994) were also studied. More recently, various works resulting from the CarboEurope-IP program were published, enlarging the range of cultures and management systems investigated (Aubinet et al., 2009; Ceschia et al., 2010; Eugster et al., 2010; Lehuger et al., 2010; Moureaux, 2008; Sus et al., 2010). Different approaches can be applied to partition the NEE flux measurements into photosynthesis (GPP) and ecosystem respiration (R_{eco}) components. The most commonly used is the method proposed by Reichstein et al., (2005) that was modified by Béziat et al., (2009) to take into account crop specificities in terms of fast development and discontinuous land surfaces properties.

The CESBIO runs since 2005 two experimental sites (Auradé and Lamasquère) where Eddy-Covariance measurements are performed over crop fields with contrasted managements. These data will be described in Chapter 2, and will be used for validating our CO_2 and water flux modeling approaches (Chapter 5).

1.1.4.2. Components of the net CO_2 fluxes: modeling and remote sensing

The models used to study the net fluxes of CO_2 and the carbon dynamics of croplands are either: i) general biogeochemical models (such as *SPA* (Williams et al., 1996), *Ecosys* (Grant et al., 2007), *ISBA-Ags* (Calvet et al., 1998), *ORCHIDEE* (Krinner et al., 2005), *ORCHIDEE-STICS (Gervois et al., 2008)*) that, in some cases, constitute the land surfaces module of atmospheric models in large-scale studies. It is generally difficult for these models to take into account the specificities and complexity of the of agro-ecosystems, in particular the effects of management practices; either ii) agronomic models (*CERES* (Gabrielle et al., 1998), *STICS* (Brisson et al., 1998), *SAFY* (Duchemin et al., 2008)) that are primarily designed for simulating crop production (*NPP,* yield), but that can be adapted for also simulating the crop net CO_2 fluxes. This kind of model is generally based upon simplified modeling of the crop processes (*e.g.* direct conversion of fraction of absorbed light into biomass). In general, crop models have been widely used to simulate the crops development and production and were validated using biomass and/or yield data (Zhang et al., 2002) at the field scale. However, these models have rarely been evaluated in terms of daily (or half-hourly) net CO_2 fluxes (Lehuger et al., 2010). Adiku et al., (2006), which developed a CO_2 flux simulation model for barley crops, were the first to use *in-situ* flux data for the validation of a crop model. Nowadays, the flux measurements are more commonly used for the development and validation of the agronomic/ecosystem models. We can also cite mechanistic agronomic complex models like SUCROS (De Wit and

Goudriaan, 1978), AFRCWHEAT (Weir et al., 1984) and WOFOST (Supit et al., 1994), for which photosynthesis, maintenance and growth respiration processes are simulated explicitly.

The different ways to combine a crop model with remote sensing observations were initially described by Maas (1988a) and their classification was revisited by Delécolle et al., (1992). According to Moulin et al., (1998), four methods of satellite data integration have been identified: i) the direct use of a driving variable estimated from remote sensing information in the model; ii) the updating of a state variable of the model (*e.g.* LAI) derived from remote sensing; iii) the re-initialization of the model, *i.e*., the adjustment of an initial condition to obtain a simulation in agreement with the remotely-sensed derived observations; iv) the re-calibration of the model, *i.e*. the adjustment of model parameters to obtain a simulation in agreement with the satellite observations.

Maas was one of the firsts to use satellite data (Landsat/MSS) to derive LAI time course for sorghum (1988b) and winter wheat crops (1991) by re-initializing/re-parameterizing the semi-empirical GRAMI crop model (based upon the Monteith's efficiency model (1977)). The technique helped estimating yields, which is one component of the carbon budget, for different fertilization and irrigation treatments. The modeling approach adopted through this thesis is based upon these concepts: using biophysical variables derived from satellite data to parameterize at first a simple crop model. For this purpose, we use the SAFY model (Duchemin et al., 2008), which presents essentially the same formalisms as the GRAMI model: semi-empirical crop growth models, based upon the Monteith light-use efficiency theory, driven exclusively by metereological data (mean daily air temperature and incident daily radiation) and completely adapted to the use of remote sensing data. This model was then modified to simulate the different components of NEE.

Given that the crop carbon and water budgets are mainly driven by: 1) the climatic and environmental conditions (such as temperature, soil and air humidity...) that may affect, in different ways, the components of the net fluxes (*e.g* the photosynthesis, autotrophic and heterotrophic respirations); and 2) the agricultural practices (such as tillage, organic fertilizer, straw management and irrigation), crop models that are able to take into account both of these drivers (1 and 2) are certainly the most promising for estimating carbon and water fluxes/budgets (Grant et al., 2007; Huang et al., 2009; Wattenbach et al., 2010). In this context, through this thesis we develop an approach that modifies the original SAFY crop model in order to simulate the dynamics of carbon and water fluxes for the wheat croplands, and still benefiting of the remote sensing observations for calibrating the model and estimating those fluxes at regional scale. The basic principles of the optical remote

sensing and link with biophysical variables (assimilated by the crop model) are described in section 1.2.

1.1.5. Environmental budgets and indicators

Since croplands can play an important role in climate change mitigation, it is essential to achieve a comprehensive understanding of the processes driving the current patterns of cropland carbon and water dynamics (Wattenbach et al., 2010). For this purpose, the development of remotely-sensed biophysical products, which characterize the temporal and/or spatial heterogeneity of croplands, combined with agro-ecosystem models, which reproduce the main crop physiological processes, constitutes a valuable research tool to achieve those objectives.

At last, given the presented approaches and evolution of the SAFY model towards the SAFY-CO_2 and SAFYE-CO_2 versions, it is possible to compute budgets and indicators that allow evaluating the crop ecosystems (locally or regionally), in terms of environmental and agronomical aspects. This kind of information is fundamental for identifying the processes and management practices that should be privileged for reducing the environmental impacts of crops or for improving their functioning, and also for maintaining yield production.

1.1.5.1. The net ecosystem carbon budgets (NECB)

In order to calculate the carbon footprint of agro-ecosystems (Net Ecosystem Carbon Budget, NECB), in addition to the atmospheric net exchanges, *i.e.* the NEE, we should also consider the carbon losses corresponding to the biomass (NPP) fraction that is exported at harvest (C_{exp}), the carbon input of seeds (or tubers) at sowing and, in some cases, the input of carbon as organic fertilizers (OF) (see Figure 1.4). The difference between the net carbon inputs and the carbon loss at harvest reflects the evolution of the organic carbon stocks in the ecosystem at short term, which can be interpreted as the variation of soil organic carbon matter (ΔSOM, Eq. (1.5)) when the carbon budget integrates a longer period (typically a whole crop rotation). The sign of the annual NECB indicates if the ecosystem is a carbon sink (NECB negative) or a carbon source (NECB positive).

$$NECB = NEE - OF + C_{\exp} \approx \Delta SOM \tag{1.5}$$

Note that NEE is negative if the GPP term (photosynthesis) is greater than the Reco term (respiration).

1.1.5.2. The water use efficiency (WUE)

In the context of climate change and agricultural water resource limitations, the ability to identify drivers for improving agro-ecosystem water use efficiency (WUE) is

essential. Improving agricultural water resource management, and also the stocks of organic carbon in the soils, has become a major challenge worldwide. To respond to the increasing food demand, crop management must be attentive to climate variability and adapt their practices according to given conditions, for example, by using crops that produce more biomass per amount of water used at the plant and the ecosystem levels (Tallec et al., 2013). Therefore, quantifying and understanding the water use efficiency (WUE) of croplands is essential for facing agro and environmental challenges and for developing sustainable options regarding the agricultural practices.

Many definitions of WUE have been proposed depending on the objectives of the study and on the considered time and spatial scales (Katerji et al., 2008; Ritchie and Basso, 2008). For crop ecosystems, three main WUE approaches have been developed. The most commonly used WUE definition is based on an agronomical approach (WUE_{agro}) that considers biomass production per amount of water used at the field scale [Eq. (1.6)]. Biomass production can be defined either as the total aboveground biomass produced during the cropping season or as exported biomass (grain + straw if this last is exported) (Katerji et al., 2008). The first definition is more relevant for agronomical and environmental approaches, and the second is more relevant for production and economical issues. The water consumed can be defined either as evapotranspiration (ETR), or as precipitation plus irrigation when ETR cannot be estimated. This agronomical WUE has most often been calculated during the growing season, so the impacts of crop rotation, intercrop periods (re-growth or bare soil) and rain distribution on WUE on the annual time scale have not usually been tackled.

$$WUE_{agro} = \frac{\text{biomass production or exported biomass}}{\text{ETR or rain + irrigation}} \quad (1.6)$$

$$WUE_{eco} = \frac{\text{NEE}}{\text{ETR}} \quad (1.7)$$

$$WUE_{NECB} = \frac{\text{NECB}}{\text{ETR}} \quad (1.8)$$

$$WUE_{plt} = \frac{\text{GPP}}{\text{TR}} \quad (1.9)$$

To specifically address environmental issues associated with the water cycle in the agro-ecosystem (soil + plants), an ecosystem specific WUE (WUE_{eco}) has been defined (Sinclair et al., 1984; Steduto and Albrizio, 2005; Steduto et al., 1997). WUE_{eco} is the ratio between the net CO_2 fixation of the ecosystem (NEE) and water loss through ETR [Eq.(1.7)].

To account for carbon input through organic fertilization and output through biomass export at harvest, carbon budget WUE (WUE_{NECB}) was defined (Béziat et al., 2009) [Eq.(1.8)]. This indicator allows evaluating the crop management practices that yield to an optimal tradeoff between carbon budgets and water consumption (Kleemann and Gill, 2008).

Finally, studies assessing crop varietal performance, stress tolerance and plant adaptation to climate changes (CO_2 concentration or temperature increase) usually define a specific WUE for plants (WUE_{plt}) (Tambussi et al., 2007), corresponding to the ability of the canopy (or individual photosynthetic organs) to fix carbon via photosynthesis for a given amount of water loss through transpiration [Eq. (1.9)].

Given the objectives of this thesis, we will focus our analysis on the WUE_{NECB} and WUE_{agro} indicators (Chapter 5). The estimates of our developed modeling approach (SAFYE-CO_2, see section 5.3) will allow computing these indicators. Therefore, given the actual climate change context, in particular in southwest France (CLIMATOR project), it is important to have a general approach (like ours), including the water, carbon and production aspects, for establishing sustainable strategies for preserving the environmental resources.

1.2. Remote sensing principles

The key processes and methods related to the carbon and water biogeochemical cycles over croplands were introduced in the previous section, with a special interest on the methods using crop modeling coupled to remote sensing products. To complete this chapter, the present section introduces the basic principles about optical remote sensing and its link with vegetation monitoring.

The term *remote sensing* has been defined many times. "Remote sensing is the science, or art, of acquiring information about the Earth's surface without actually being in contact with it" (Fischer et al., 1975). In another words, "remote sensing is the science of deriving information about the Earth's land and water areas from images acquired at distance. It usually relies upon measurement of electromagnetic energy reflected or emitted from the features of interest" (Campbell, 2002).

Electromagnetic energy refers to the energy that comes from electromagnetic radiation. This radiation is composed of waves that move at the speed of light. Electromagnetic radiation can be described in terms of its wavelength (λ), which is the distance between the crests of the waves; or in terms of its frequency, which is the number of crests that pass by a fixed point during a fixed time interval. In the

remote sensing field, the convention is to describe the electromagnetic radiation using wavelength.

The electromagnetic spectrum can be defined as the continuum of energy that ranges from shorter wavelengths (including gamma and x-rays) to the longer wavelengths (including microwaves and broadcast radio waves)(Figure 1.5). It is divided into three main regions: the visible, infrared and microwave, which are further subdivided into bands, such as the blue, green and red bands of the visible region.

For this thesis we use only passive optical remote sensing. The optical sensors capture the radiances coming from the system Earth-Atmosphere illuminated by the sun, on the spectral regions of visible, infrared and thermal infrared. Passive optical sensors do not provide their own source of energy and depend upon solar radiation. Therefore, they detect natural radiation that is emitted or reflected by the scene being observed; *i.e.* the sensors record the returning radiation that is naturally radiated by the Earth's surface or that is emitted by the sun and reflected by the surface. The system sun-target-sensor is illustrated in Figure 1.6 for a wheat field. In this context, the signal measured by the sensor depends upon the characteristics of the environment (clouds or not, …), of the vegetation (crops species, phenology,…) and of the satellite system.

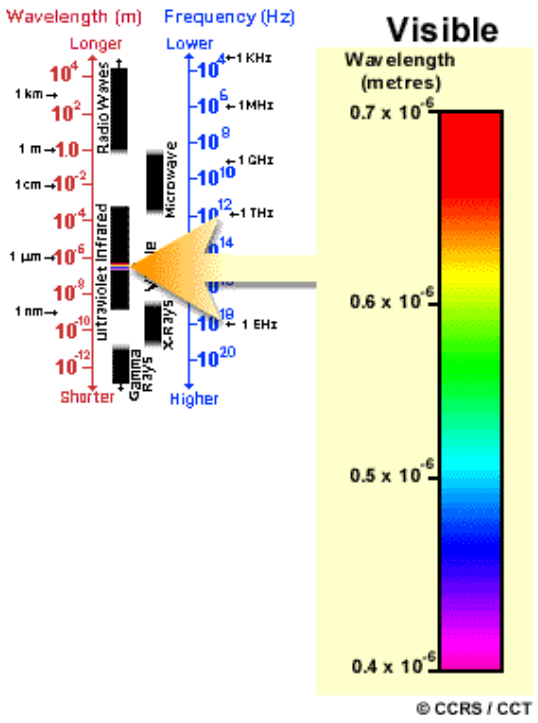


Figure 1.5- The electromagnectic spectrum, with a focus on the visible spectrum. The visible wavelengths cover a range from approximately 0.4 to 0.7 µm [extracted from (CCRS, 2014).

Concerning the environmental aspects, the recorded signal is affected by the condition of sun illumination, such as season and day of the year (intensity), and also by the position (latitude and longitude) of the scene relative to the sun. Besides, the radiation can be attenuated by atmospheric properties, such as gases concentration, temperature, water content and aerosols scattering and absorption. The physical characteristics of the vegetation, such as shape, structure, geometry and composition, will also change the properties of the reflected and emitted radiation.

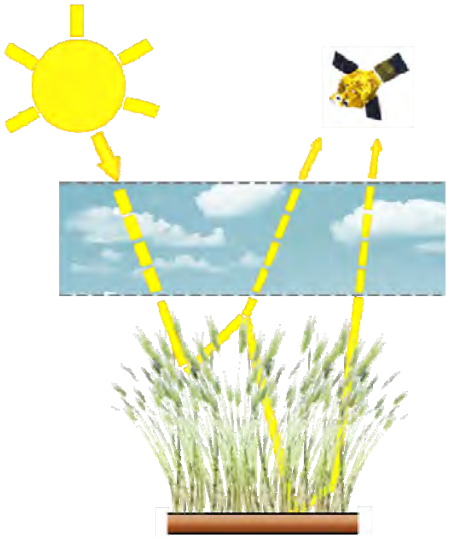


Figure 1.6- Scheme of the passive remote sensing approach. The satellite sensor records the radiation that is naturally radiated by the wheat field and/or that is emitted by the sun and reflected by the canopy. The recorded radiation is impacted by atmospheric effects, here represented by the blue cloudy sky.

Concerning the sensor properties, the altitude and geometry of observation of the satellite will directly impact the measured radiation. The positions of the sensor and of the sun relatively to the observed landscape are conventionally represented by three angles known as the zenith viewing angle (θ_V), the solar zenith angle (θ_S) and the relative azimuth angle ( ) (Figure 1.7). Moreover, each sensor has its own technical characteristics in terms of temporal, spatial, spectral, and radiometric resolution, that also determines the measured signal.

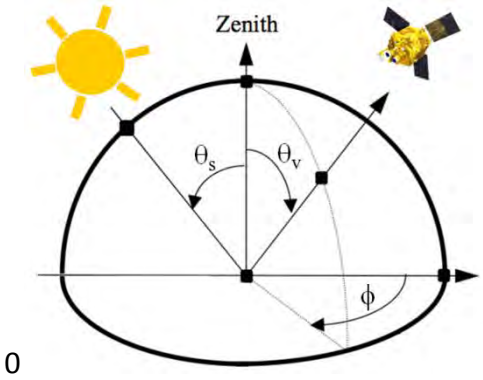


Figure 1.7- Observation and illumination geometries. The zenith viewing angle (θ_V), the solar zenith angle (θ_S) and the relative azimuth angle ( ) are indicated.

The radiometric resolution refers to the sensitivity of the sensor to the magnitude of the collected electromagnetic energy. The finer the radiometric resolution of a sensor is, the more sensitive it is to detecting slight differences in (reflected or emitted) energy. The spectral resolution describes the ability of the sensor to define fine wavelength intervals. The finer the spectral resolution is, the narrower the wavelength range for a given spectral band will be.

The spatial resolution refers to the size of the smallest element that can be resolved on the ground. In a digital image, the resolution is limited by the pixel size, *i.e.* the smallest area that can be separately recorded as an entity corresponds to the pixel size. The remote sensing observations are also characterized by their temporal resolution (or revisit period), which refers to the length of time it takes for a satellite to image the exact same area a second time. The temporal resolution depends mostly upon the length of time for the satellite to complete one entire orbit cycle, but also upon a variety of factors, including the sensor swath overlap and latitude.

In remote sensing, there is often a resolution trade-off. The different radiometric, spectral, spatial and temporal resolutions are often limiting factors for the utilization of the satellite image data for different applications. Because of technical constraints, the different resolutions categories must be balanced according to the objectives and capabilities of the sensor.

In the context of crop monitoring, having high spatial and temporal resolutions seems essential (West et al., 2010). Given the rapid changes of the vegetation through the growth cycle and likely clouds cover, a high temporal revisit is needed; and, for studying the processes at fine scales and being able to accurately discriminate the lands use and fields' borders, the high spatial resolution is thus required. For a long time, no satellite offered these characteristics. The recent availability of the Formosat-2 and RapidEye satellites, combining both high spatial and temporal resolutions (but over a limited geographic extent), and of the upcoming missions such as Sentinel-2 and Venus, provide great perspectives of

progress in crop monitoring/modeling research. In addition, the integrated application of multi-sensor and multi-temporal remote sensing data is also a recent trend, and consists in a useful approach to solve the problem of observation gaps.

1.2.1. Reflectance definition

In the remote sensing domain, it is important to define the concept of reflectance. The reflectance refers to the ratio of the total amount of radiation reflected by a surface to the total amount of radiation incident on the surface (Figure 1.8). Reflectance is therefore unitless and has a range limited to [0-1]. The reflectance with a specified geometry of incident and reflected electromagnetic radiation is called directional reflectance. Thus when incident and reflected radiation are both directional, the term bidirectional reflectance is applied. The concept of bidirectional reflectance is used in the design of satellite sensors. Hereafter, for the sake of simplicity, the term 'reflectance' will be used instead of bidirectional reflectance.

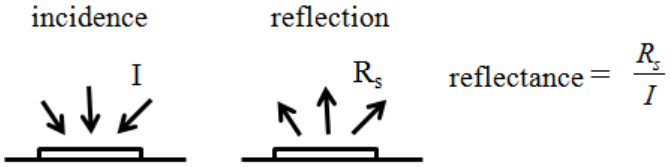


Figure 1.8- Illustration of the reflectance characterization.

The reflectance spectrum of an element is a plot of the fraction of radiation reflected as a function of the incident radiation wavelength and serves as a unique signature for the given element. In principle, an element can be identified from its spectral reflectance signature if the sensing system has sufficient spectral resolution to distinguish its spectrum from those of other materials. This idea provides the basis for multispectral remote sensing. For example, water has almost no reflectance in the infrared region. The vegetation has a typical spectral signature which enables it to be distinguished readily from other types of land cover in an optical/near-infrared image. The reflectance is low in both the blue and red regions of the spectrum, due to the light absorption by chlorophyll and other pigments for photosynthesis (Figure 1.9). It has a peak at the green region which makes vegetation presents green color to the human eye. In the near infrared region, the reflectance is much higher than that in the visible band due to the cellular structure of the leaves, making this region very useful for vegetation monitoring. These particular spectral properties of green vegetation, which are directly linked to photosynthesis, facilitate the retrieval of information on plant canopies from the electromagnetic signal measured by satellite remote sensing instruments (Tucker and Sellers, 1986).

Through this thesis, we aim at using the optical remote sensing–derived reflectances of the crop vegetation canopy for characterizing its biophysical properties (such as leaf are index), and then at applying those extracted information into crop modeling approaches.

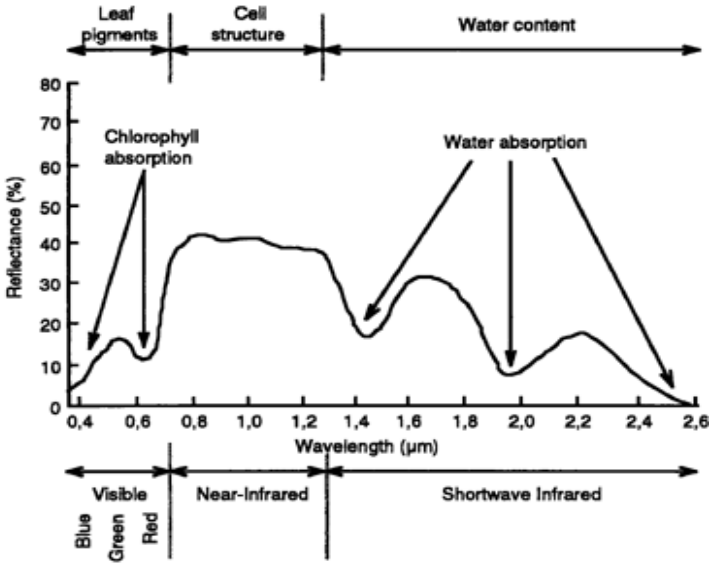


Figure 1.9- Typical spectral response characteristics of green vegetation showing the spectral effects of leaf pigments, cell structure, and water content (Hoffer, 1978).

1.2.2. From satellite-derived biophysical variables up to crop growth models

Remote sensing data have been widely used to estimate the canopy biophysical variables, such as leaf area index (LAI), daily fraction of photosynthetically active radiation absorbed by the vegetation(fAPAR), chlorophyll content (C_{ab}) and cover fraction (fCover), which are applied to crop monitoring, water and carbon cycle simulation, climatic modeling and global change researches, amongst others. Therefore, it is very important to precisely estimate these variables from remote sensing data at the local, regional or global scale (Combal et al., 2002; Xiao et al., 2008).

Nowadays, there are many methods for estimating biophysical variables from remote sensing data (Baret and Buis, 2008; Liang, 2005; Weiss and Baret, 1999). These methods can be divided into two main classes: the empirical and the physical-based approaches. Generally, empirical relationships have been established between spectral vegetation indices (which combine certain spectral bands, such as the NDVI (Tucker, 1979)) and ground measurements of the biophysical variables. The physical-

based methods often rely upon radiative transfer models, describing the relationship between canopy characteristics and reflectance, which are used in the inverse model to estimate the canopy biophysical variables from remote sensing data (Jacquemoud and Baret, 1993; Kuusk, 1991).

The leaf area index (LAI) is a particularly interesting variable, since it characterizes the main interface between atmosphere and vegetation, thereby governing the exchanges of energy, water and carbon. LAI can be defined as half the total developed area of green leaves per unit of ground horizontal surface area (Chen and Black, 1992). However, in the remote sensing field, because of the difficulty in distinguishing the different green elements of the plant (like leaves, stems,…), the alternative concept of Green Area Index (GAI) is commonly used. The GAI refers to the photosynthetically (green) active plant area, with no distinctions between leaves, stems, branches, and reproductive organs.

The GAI is a state variable that is involved in important physical and/or physiological processes of the crop canopy, such as photosynthesis, respiration and transpiration. Thereby, one of the great interests in retrieving GAI from remote sensing observations is to couple it with crop growth models that simulate these processes.

2. Study area and data description

This chapter will describe our study area and the data sets used through this thesis. The first part presents the general characteristics of the study site. Next, the remote sensing data set characteristics and images pre-processing chain are presented. The following section consists in the *in-situ* data set description. The *in-situ* data can be divided into two main groups regarding the origin and protocols that were used for collecting those data. The data groups are designated: the 'experimental sites' data and 'intensive field campaigns' data. Here, a particular focus on the 2011 field campaign, performed specifically for this thesis. Then the data obtained from collaborations with farmers' networks and/or associative corporations are summarized, followed by the general description of the soil maps compilation. The chapter is concluded by the presentation of the data considered as 'regional data', which includes the regional land use database (RPG), the SAFRAN meteorological data and the yield statistics at a regional (departmental) scale.

2.1. Study Area

The study area is located in southwest France, within a 24 km x 24 km area (1°10'E, 43°27'N), as showed in Figure 2.1. The site is situated next to Toulouse and is mostly part of the Haute Garonne department, but with a small part within the Gers department. The region is mainly covered by arable lands (around 60%). Major cultivated crops are wheat (*Triticum aestivum L.*), barley (*Hordeum vulgare L.*), rapeseed (*Brassica napus L.*), maize (*Zea mays L.),* soybean (*Glycine max)* and sunflower (*Helianthus annuus L.*).

The climate is temperate mild, characterized by warm and dry summers, sunny autumns, soft winters and by rainy springs. Annual precipitation is about 656 mm and annual mean temperature is 13°C. The Figure 2.2 shows the monthly average temperature and precipitation, calculated from 1980 to 2011 for the Toulouse-Blagnac station.

The Garonne River crosses the east side of the study area, and soil textures are dominated by clay and loam.

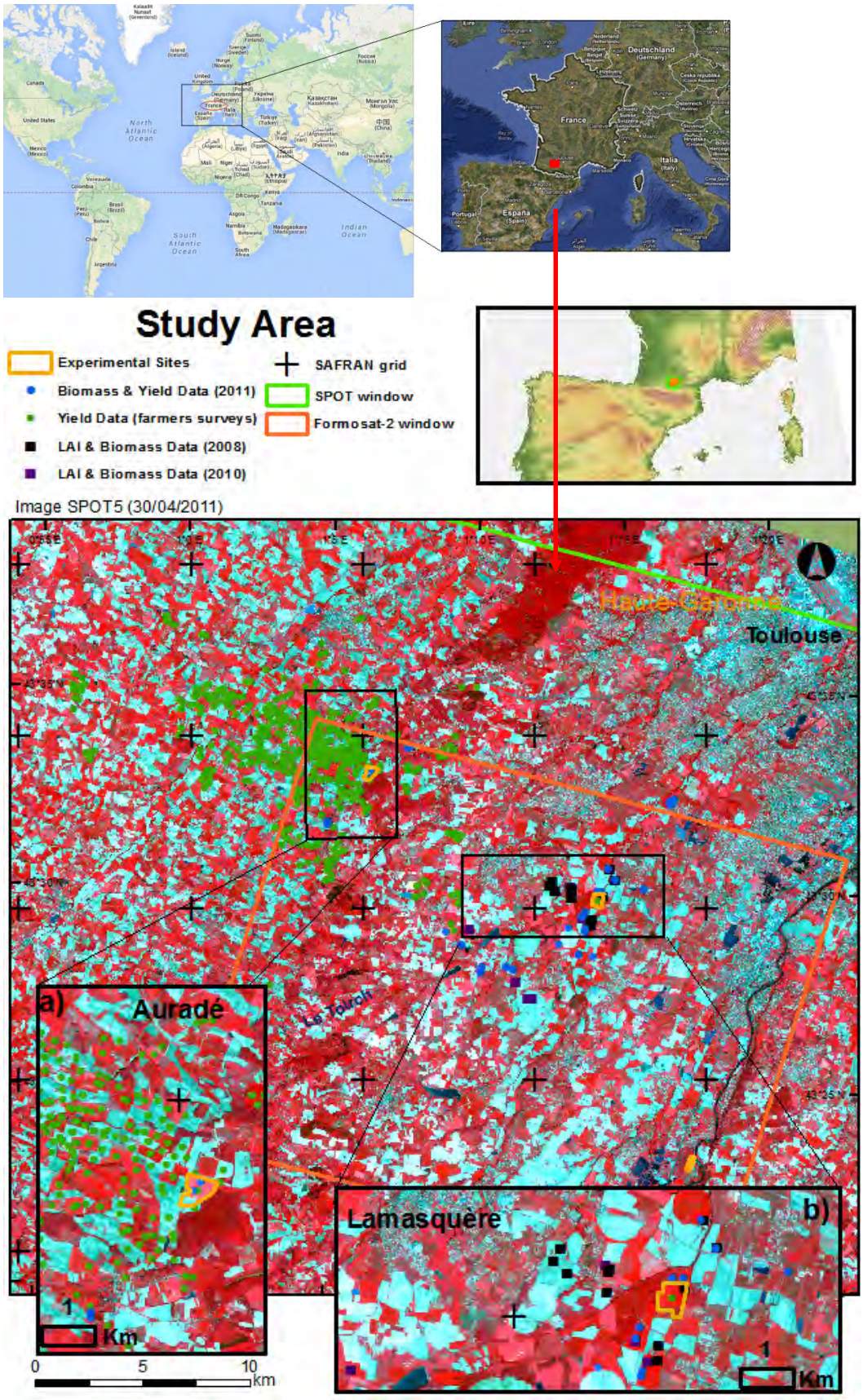


Figure 2.1- General map of the study area and the location of the experimental sites, intensive field campaigns, SAFRAN meteorological grid and extent of the satellite images.

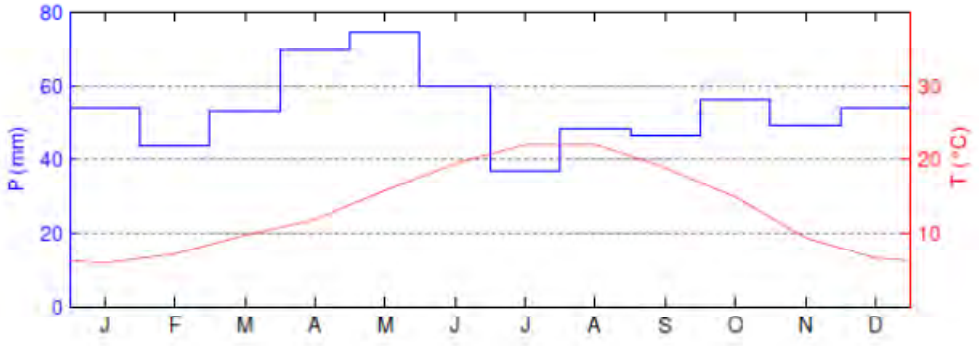


Figure 2.2- Ombro-thermal diagram created with the data registered by the Toulouse-Blagnac meteorological station (source: Météo-France) over the years 1980-2011. It shows the monthly average of temperatures (right axis) and precipitation (left axis). [Figure extracted from Claverie (2012)].

2.2. Remote sensing data

The need of continuous time series of high resolution images for a detailed monitoring of crop development led to the combined use of data from Formosat-2 and Spot satellites. Only high spatial resolution and gap-free satellite time series can provide enough information for efficient crop monitoring applications. The potential of remote sensing data is often limited by cloud cover and/or gaps in observations. Thus for this work we used data from both satellites since they have rather similar characteristics in terms of spectral bands and spatial resolution. This paragraph presents a description of both systems and the pre-processing chain applied to the satellite images.

2.2.1. Formosat-2

Formosat-2 (F2) is a Taiwanese satellite, launched in 2004. It has a sun-synchronous orbit at altitude of 891 kilometers and it takes approximately 103 minutes to orbit the earth. Formosat-2 provides images with spatial resolution of 8m in four reflective bands (Blue, Green, Red and Near-infrared). Two acquisition modes are available: the panchromatic mode and the multispectral one. Only the last one was used for this thesis. The sensor has a footprint of 24 km × 24 km (represented by the orange frame in Figure 2.1) and a revisit frequency of one day. Images for a given site can be always taken from the same position, which means they are taken at near constant viewing angles. Table 2.1 synthesizes the main characteristics of this satellite. A set of 131 images is available from 2006 to 2011 over the study area.

Table 2.1- Formosat-2 satellite key features

Formosat-2	
Spectral bands	0.45~0.52µm（Blue） 0.52~0.60µm（Green） 0.63~0.69µm（Red） 0.76~0.90µm（Near Infrared）
Spatial Resolution	8 meters
Image Swatch	24 kilometers
Viewing angle	+/- 45°
Revisit frequency	1 day

2.2.2. SPOT

The SPOT (Satellite Pour l'Observation de la Terre) satellites are part of CNES's Earth observation strategy. For this thesis, we use images taken by the SPOT 2, 4 and 5 satellites. The SPOT orbit is polar, circular, sun-synchronous, and phased. Each system has two viewing instruments and each SPOT scene covers a geographical area of 60 x 60 km (represented by the green frame in Figure 2.1). Two alternative modes of imaging are possible: panchromatic and multispectral modes. Data have spatial resolutions of 20m (SPOT 2 and 4) or 10m (SPOT 5) in the green, red and near infrared spectral bands for SPOT 2 plus short-wave infrared band for SPOT 4 and 5. The viewing angles varied between +/- 27°. Table 2.2 synthesizes the main characteristics of the SPOT satellites. A total of 69 SPOT images were acquired over the 6 years period for our study area.

Table 2.2- SPOT satellite key features

SPOT	
Spectral bands	0.52~0.60µm（Green） 0.63~0.69µm（Red） 0.76~0.90µm（Near InfraRed） 0.45~0.52µm（Short-Wave InfraRed）， for SPOT4 and 5
Spatial Resolution	20 meters, SPOT2 and 4 10 meters, SPOT5
Image Swatch	60 kilometers
Viewing angle	Cross-track : +/- 27°
Revisit frequency	2 to 3 days 1 day with full constellation

Figure 2.3 shows the calendar of image acquisitions for F2 and SPOT satellites over our study area, from 2006 until 2011. The F2 satellite presents more regular acquisitions than the SPOT ones, but still the use of SPOT series remains essential in order to fill the gaps of the F2 series and also in case of cloudy days, they might be helpful. Despite the numerous images data set, the presence of clouds and/or shadows in a number of images reduces the amount of exploitable data. For the year 2008, we can observe that from February until June, there is just one SPOT acquisition on the 26th of April. Unluckily this image is quite cloudy, constituting a gap in the sequence. Consequently we might not have enough information for monitoring the winter wheat development this year. Practically the same happens for 2009. From March until June, there is only one cloudy F2 acquisition on the 5th of May, what makes the series not fully usable.

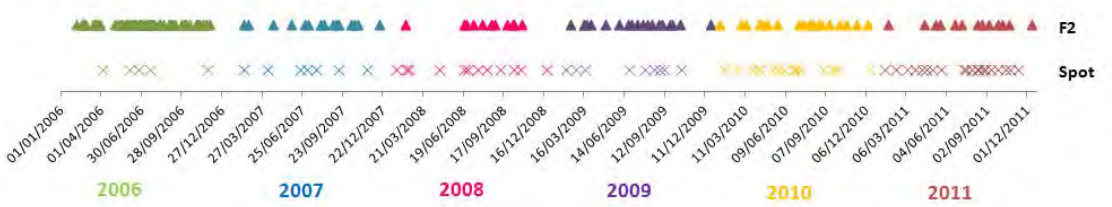


Figure 2.3- Calendar of image acquisitions from 2006 until 2011, for Formosat-2 (up) and SPOT 2,4 and 5 (bottom) satellites.

2.2.3. Images pre-processing

Both remote sensing datasets (F2 and SPOT) were processed with the KALIDEOS processing chain [http://kalideos.cnes.fr]. It provides valuable atmospheric, radiometric and geometric corrections, required for accurate time series studies.

The pre-processing starts with the ortho-rectification step. The geometric distortions are mainly due to the relative movement of the sensor, of the platform where the sensor is mounted and of the earth (rotation). The topographic characteristics of the observed areas can also contribute to these distortions. The geometrical corrections are therefore performed, in order to palliate these effects, providing images projected in a same cartographic system and having geographic correspondence between them (Figure 2.4).

The chosen approach consists in taking a reference image (already geometrically corrected) and a digital elevation model (DEM), then choosing remarkable points, identifying the agreement of the position of selected points in the reference and the processed images, and finally correcting the present image, so it matches precisely the reference image. The reference images were selected from the SPOT5 acquisitions, having a 2.5m resolution, and presenting viewing angle close to nadir

and no or small presence of clouds. The mean geometric correction accuracy obtained is on the order of 0.2 pixels (detailed information on LaFrance et al., (2012)).

The ortho-rectified images are then submitted to the radiometric corrections. Initially, the ortho-images have no physical sense. The raw radiometric values are therefore converted into top of atmosphere (TOA) reflectances. Next step is to obtain top of canopy (TOC) reflectance images, which are needed for our applications.

The atmospheric correction is one of the key steps to obtain valuable surface reflectances from satellite instruments operating in the visible and near-infrared domains (as F2 and SPOT). The main difficulty of this processing is the correction of the effects of atmospheric aerosols, because their abundance and nature is highly variable in time and space (Hagolle et al., 2008). In the visible domain, the top of atmosphere (TOA) reflectance above dark targets (dense vegetation cover for example) may change by more than 100% when comparing a hazy day to a clear day. To perform accurate atmospheric corrections, a good knowledge of the aerosol optical properties is necessary (Hagolle et al., 2008).

For this purpose, we simulate ground and TOA reflectances with the 6S model (Vermote et al., 1997), for a given range of surface reflectance values. After Look-Up Tables (LUT), relating ground and TOA reflectances, are created for each processed image (1 LUT per spectral band). Then, for each measured TOA reflectance pixel, the correspondent ground reflectance (TOC) value is found in the earlier mentioned LUT.

The quality of this inversion of surface reflectance depends mainly on the estimation of the atmospheric conditions at the moment of image acquisition. The 6S model requires as inputs: gaseous contents, aerosol optical depth, water vapor content, and atmospheric pressure, among others. The necessary data characterizing the atmospheric conditions are obtained from the AERONET network [*aeronet.gsfc.nasa.gov*] or from the NCEP (National Center for Environment Prediction, *www.ncep.noaa.gov*).

The TOC reflectance images are then obtained. Depending on the topography of the observed area, the DEM can be taken into account through the radiometric corrections. A valuable option is also to perform topographic corrections on the reflectance (because of sun or shade oriented slopes), when studying hilly areas. Figure 2.5 summarizes the main steps of the described pre-processing approach.

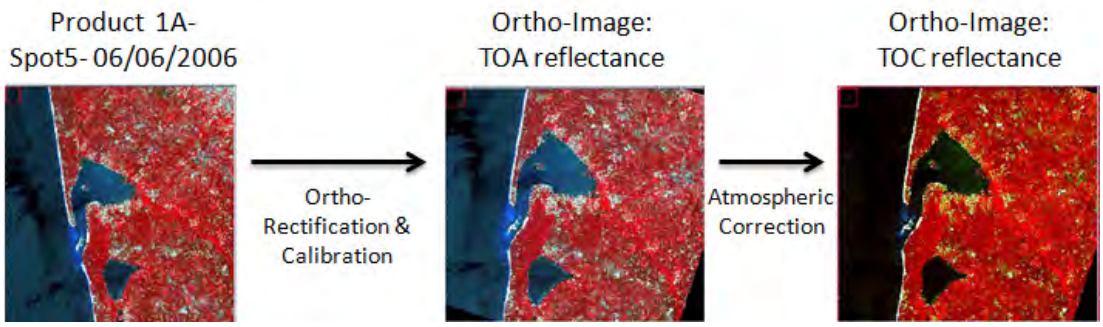


Figure 2.4- Representation of the processing sequence from Product 1A to TOC reflectances ortho-images. (Source: LaFrance et al, 2012).

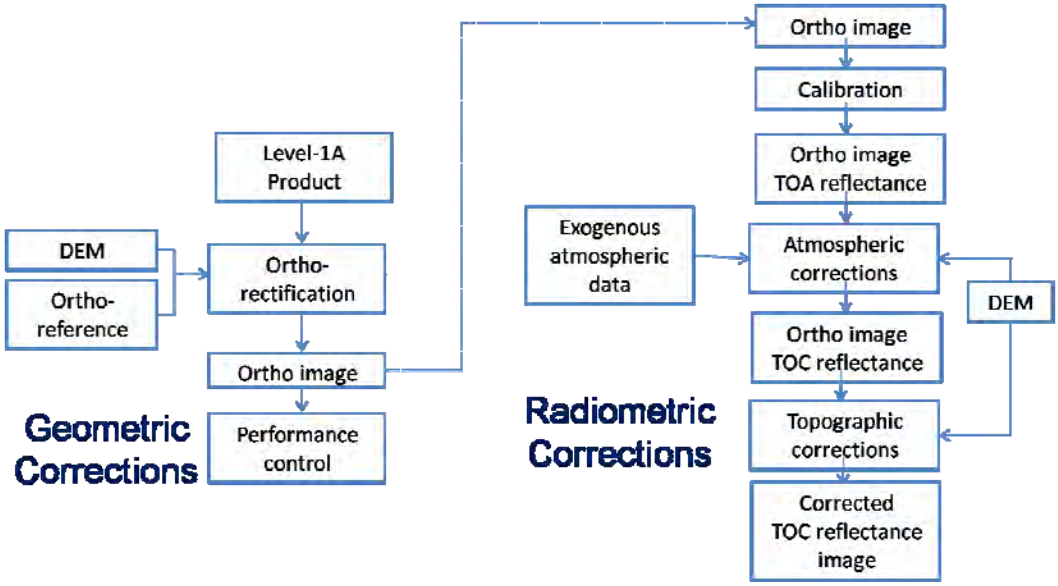


Figure 2.5- Flow diagram of the geometric and radiometric corrections performed over the satellite images. (Source: LaFrance et al., 2012).

2.3. *In-situ* data

2.3.1. The experimental sites

Two agricultural sites were instrumented by CESBIO in 2005 for performing micrometeorological and meteorological measurements. Hereafter, they are called 'experimental sites'. The *Auradé* and *Lamasquère* sites are located in southwest of Toulouse and are separated by 12km. They are part of the global network FLUXNET [http://fluxnet.ornl.gov] and the ICOS network [http://www.icos-infrastructure.eu/] for observation of surface fluxes (CO_2, latent and sensible heat). Both sites present similar climatic conditions but they have different soil properties, topography and management practices. Table 2.3 presents the main characteristics of the two sites.

Table 2.3- Site characteristics and general information on soil and meteorology. Climatic means were measured with each site meteorological station. Climatic normals correspond to means calculated over 7 years (2005-2011).

		Sites	
		Auradé	**Lamasquère**
Site characteristics	Coordinates	43°55'37''N, 01°10'61''E	43°49'65''N, 01°24'19''E
	Plot area [ha]	23.5	32.3
	Elevation [m]	245	180
	Slope [%]	2	0
	Exposure	ENE	
	Fetch in main wind directions [m]	260 W, 270 ESE	200 W, 140 ESE
Soil	Class of soil texture	clay loam	clay
	Particle analysis [% sand; % loam; % clay]	20.6 ; 47.1 ; 32.3	12.0 ; 33.7 ; 54.3
Climatic normals	Mean annual temperature [°C]	12.8	12.8
	Annual Precipitation [mm]	639.4	649.3

The Auradé plot is located on a hillside area near the Garonne river terraces. The plot is characterized by a rapeseed/winter wheat/sunflower/winter wheat rotation.

The Lamasquère plot is part of an experimental farm (milk and chicken production) owned by the EIP (Ecole d'Ingénieurs de Purpan). The instrumented site borders the "Touch" river and is characterized by a maize/winter wheat rotation. At the Auradé site only grain is exported and it receives just mineral fertilizers. At the Lamsquère site most of the aboveground biomass is exported for feeding the animals of the farm or it serves as litter. For more information on the sites management practices, see Béziat et al., (2009).

Several measurements were performed over the experimental sites since 2005. In addition to micro and meteorological measurements, destructive measurements of plant leaf area index and biomass were performed as well, in order to characterize the crops development during the vegetative period. Next paragraphs present a description of this set of ground data.

2.3.1.1. Destructive measurements: GAI and Biomass

For estimating leaf area index both destructive (direct) and non-destructive (indirect) measurements were carried out. Data were collected from the sowing until the maturity phase, once per month, approximately. The estimated variables were:

destructive green leaf area index (LAI), destructive green plant area index (GAI) and the effective green plant area index (GAIeff), estimated from the hemispherical photographs.

The protocol for collecting field data has changed through the years (from 2006 until 2011). In 2006 and in 2007, plants were collected on the two diagonals of the fields, monthly during slow vegetation development periods and every two weeks during fast vegetation development periods. In 2008 a single transect was done crossing the entire field, but not always on the same direction for each year. Since 2010 two transects are done in such a way that the vegetation samples are better covering the footprint of the flux measurements. Figure 2.6 shows how the collecting itinerary changed in the Auradé site, for the years when winter wheat was cultivated (2006, 2008 and 2010). Each different colour represents a different day of measurement during the plant development.

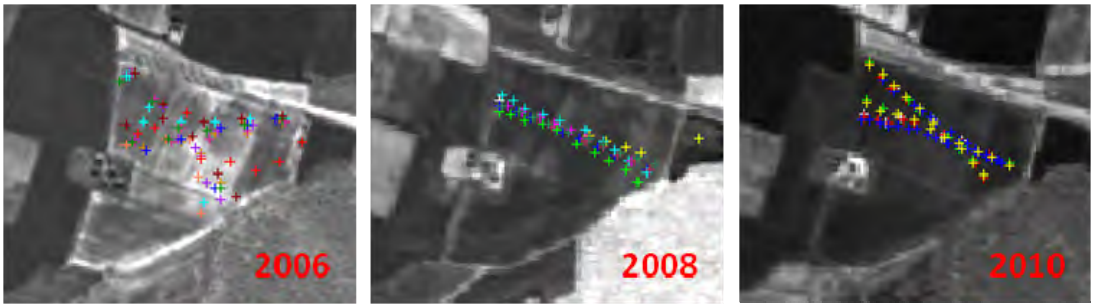


Figure 2.6- Illustration of the dynamics of the itinerary for collecting ground data over the Auradé site, in 2006, 2008 and 2010. For these years, winter wheat was cultivated. The 'crosses' represent the registered GPS points where measurements were performed. Each different color represents a date of measurement over the vegetative cycle.

For rapeseed, maize and sunflower, around 30 plants were collected at each date. For winter wheat, ten to twenty 1.5 meter long rows were collected at each sampling date, according to the collecting practices described earlier.

The organs of the collected plants were separated into (green and yellow) leaves, stem, flowers and fruits. Then the leaf and plant green surface areas were estimated by means of a LiCor planimeter (LI3100, LiCor, Lincoln, NE, USA). Afterward the destructive LAI and GAI variables were deducted.

Next step consisted in drying all the samples in an oven at 65°C during 48 hours minimum, then in weighting each of the samples. The dry aboveground biomass (DAM) was then estimated for each collected plant. Afterwards we estimated the mean DAM (and associated standard deviation) for each site-sampling date.

Table 2.4 summarizes the amount of *in-situ* measurements performed from 2006 until 2011 for winter wheat (information for summer crops are not detailed here).

At those two sites, the yield and description of technical operations were given by the farmers.

Table 2.4- Summary of the winter wheat ground measurements performed from 2006 until 2011 over the Auradé and Lamasquère sites (displayed numbers correspond to the amount of sampled dates per site-year) and through the intensive measurements campaigns (the numbers represent the quantity of sampled ESUs). The observed variables are effective GAI (DHP, from indirect measurements), green LAI, GAI, biomass and yield (from destructive measurements).

Year	Site / ESU	GAI (DHP)	LAI/GAI (destructive)	Biomass	Final Grain Yield
2006	Auradé	9	10	10	-
	Field Campaign	41	-	-	-
2007	Lamasquère	7	9	9	-
2008	Auradé	9	10	10	-
2009	Lamasquère	6	8	8	-
2010	Auradé	4	5	5	-
2011	Lamasquère	4	5	5	-
	Field Campaign	-	-	23*	23

*final biomass.

2.3.1.2. Indirect measurements: GAI estimated from hemispherical photographs

Non destructive measurements of green plant area index (GAI) were also performed, in order to analyze the GAI dynamics obtained by a different method. This indirect method demands less manual labour than the destructive one, and is less time-consuming. It allows collecting data more often during the vegetative season, and thus obtaining a regular monitoring of the vegetation, and over larger surfaces.

For our work, we chose to take digital hemispherical photographs (hereafter called as DHPs) as an indirect method for estimating GAI. There are several solutions for processing the hemispherical photographs in order to estimate the biophysical variable of interest. We can cite: Winscanopy (Regent Instruments, Quebec, Canada), GLA (Frazer et al., 1999), CIMES (Walter, 1994), Hemiview (Delta-T Device) and CAN-

EYE [www4.paca.inra.fr/CAN-EYE]. For this work we chose the CAN-EYE software. This software was developed by the EMMAH team, from the INRA Avignon. One of its main interests is that CAN-EYE processes a set of photographs at once, which reduces the processing time considerably.

CAN-EYE's approach is based on the estimation of gap fraction, which can be defined as the probability that an incident light beam reaches the ground, without being intercepted by the canopy. Therefore, measuring gap fraction is equivalent to measuring transmittance at ground level. Different methods have been developed to link the gap fraction with the canopy structure. Indeed, the variable that can be actually related to gap fraction is the Plant Area Index (PAI) rather than LAI, since gap fraction is mainly sensitive to any green or non-green vegetation elements within the canopy volume, including leaves but also stems and ears (Baret et al., 2010). However, given the fact the plant only photosynthesize when its organs are green, it seems more appropriate to use Green Area Index (GAI) as our variable of interest, referring only to the green (active) elements of the canopy. Besides, the GAI variable has been commonly used in light use efficiency-based crop models (like SAFY), since it is more closely related with the fraction of absorbed photosynthetically active radiation (fAPAR).

In CAN-EYE the gap fraction is calculated from the RGB images through a supervised classification (Demarez et al., 2008). Several classification options are available. It is possible to classify the soil (or gaps) pixels only, and then non-selected pixels are considered as green vegetation. Another possibility is to select the green vegetation pixels only (so non-selected pixels are considered as gaps). Still an option is to select both vegetation and soil pixels or even else, to consider more than two classes, in order to differentiate flowers or senescent vegetation, for example. A default classification based on predefined color segmentation is then proposed. If it is accepted, CAN-EYE applies an automatic color segmentation in which the total number of distinctive colors is reduced to 324. This reduction of the radiometric resolution simplifies the subsequent manual classification operation. The classification can be then interactively and iteratively refined if judged necessary (Demarez et al., 2008). A mask can be applied in order to remove parts of the photographs containing other elements than vegetation (like the photographer, a backpack, the pole...). Figure 2.7 shows an example of DHP taken downwards a wheat plot, the DHP when the undesired elements are removed and the final green vegetation-soil classification image.

Once the mean gap fraction is estimated for each set of DHPs, GAI can be calculated by applying a relation that links both variables (gap fraction and GAI). The Poisson law is commonly used in indirect methods in order to determine the GAI; its formula is shown in Eq. (2.1). This model assumes that leaves are uniformly and

randomly distributed, which is not always valid for different canopies, especially for those with aggregative patterns (Baldocchi and Collineau, 1994). To allow the use of the Poisson law, the concept of *effective* GAI is proposed (Chen and Black, 1991; Chen, 1996; Chen and Cihlar, 1995), which corresponds to the product of a clumping index with the *true* GAI estimates. Thus when random turbid medium assumptions of Poisson model are not satisfied, a modified Poisson model expression is applied, integrating the leaves aggregation into the estimation of the gap fraction [Eq. (2.2)]. The *true* GAI is generally more closely related with the destructive measurements of GAI. *Effective* and *true* GAI, both outputs of the CAN-EYE software, are then associated by the clumping index λ [Eq.(2.3)].

$$P_0(\theta,\varphi) = e^{\left(\frac{-G(\theta,\varphi)GAI_{eff}}{\cos\theta}\right)} \quad (2.1)$$

$$P_0(\theta,\varphi) = e^{\left(\frac{-\lambda(\theta)G(\theta,\varphi)GAI_{true}}{\cos\theta}\right)} \quad (2.2)$$

$$GAI_{eff} = \lambda(\theta) \times GAI_{true} \quad (2.3)$$

where θ and ϕ are the zenith and azimuth angles, respectively, of the direction of propagation of the incident beam, G(θ, ϕ) is the mean projection of a leaf area unit in a plane perpendicular to direction (θ, ϕ).

The clumping index λ(θ) is superior to 1 if the canopy presents a regular structure, it is inferior to 1 if the canopy has an aggregative pattern and λ(θ)=1 if the canopy is uniform.

Next, we describe succinctly the main steps for the DHPs processing, which allows to estimate the mono-directional gap fraction from each set of taken photographs.

Initially a classification is performed, classing the pixels as a 'green vegetation' or as a 'bare soil' pixel. Next step consists in calculating the gap fraction $P_0(\theta,\phi)$ for each of the DHPs series. Then the mono-directional gap fraction ($P_0(\theta)$) is computed, by integrating the gap fraction over all the zenithal angles. As a final step we estimate the biophysical variables (effective and true GAI, average leaf area (ALA), clumping index, etc..), by minimizing the root mean square error between the measured gap fraction (from classification) and the gap fraction simulated by the modified Poisson law.

DHPs acquisition

The DHPs were taken downwards at a position above the canopy, which allows controlling the distance between the camera and the canopy. It also allowed to increase the sampling surface by using a pole to take photographs at higher height and thus increase the spatial representativeness of the measurements (Demarez et al., 2008). For the experimental sites, the photos were taken at each sampled date, next to the position where the destructive measurements were performed. Thus, each set of DHPs per date is supposed to represent the variability of the field canopy.

The photographs were taken with a Nikon Coolpix 8400 camera having FC-8 fisheye lens. The camera was calibrated using the method described in [*http://www.avignon.inra.fr/can_eye]* to compute the optical centre of the camera–fisheye lens system. The distance between the objective and the top canopy was set to 1 m to avoid having individual leaves too close from the sensor. For tall canopies, the use of a pole was necessary. For more detailed information on this subject, see Demarez et al., (2008), Duthoit (2006) and [http://www6.paca.inra.fr/CAN-EYE/Documentation-Publications/ Documentation].

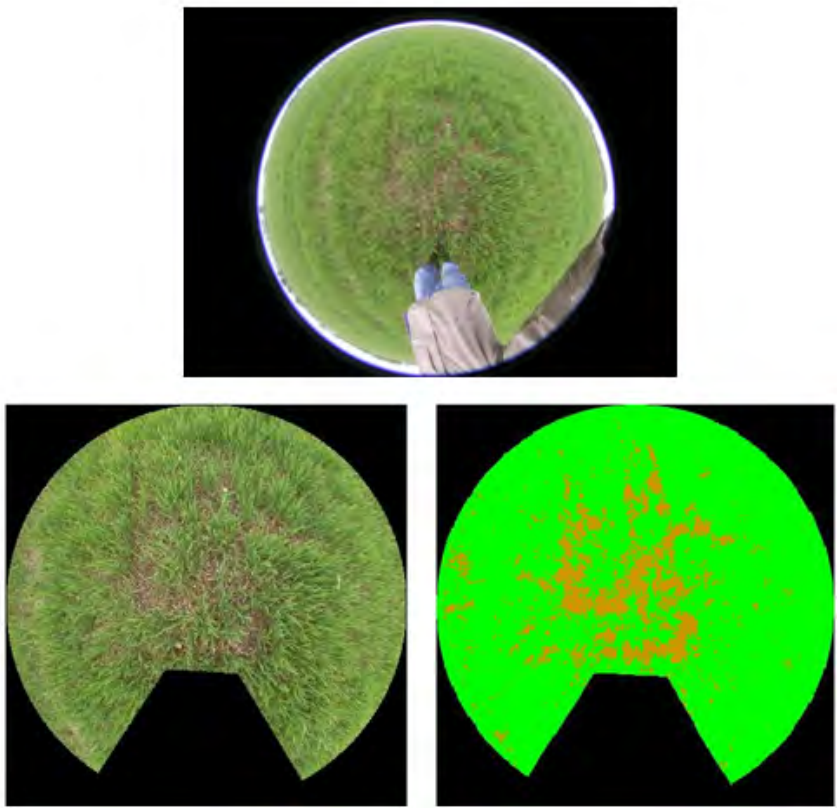


Figure 2.7- Example of a digital hemispherical photograph taken in April 2006 over a winter wheat plot. On top: the original photograph; on left bottom: the mask for removing undesired elements (here the photographer) is applied; right bottom: the result of the classification step, for which the pixels were classed as vegetation (green) or soil (brown).

2.3.1.3. Flux and meteorological measurements

Both experimental sites were instrumented in order to measure different climatic variables. Masts were installed in the middle of each experimental field. Turbulent fluxes of CO_2, water vapor (evapotranspiration and latent heat), sensible heat and momentum have been measured continuously by the Eddy-Covariance (EC) method (Aubinet et al., 1999; Baldocchi, 2003; Moncrieff et al., 1997) since March 2005. The EC devices were mounted at heights of 2.8 and 3.65 m at Auradé and Lamasquère, respectively. Instrument heights were chosen to be at least 1 m higher than crops at their maximum development (Béziat et al., 2009). The EC system is made of a three-dimensional sonic anemometer (CSAT 3, Campbell Scientific Inc, Logan, UT, USA) and an open-path infrared gas analyzer (LI7500, LiCor, Lincoln, NE, USA).

Then EdiRe software (Robert Clement, © 1999, University of Edinburgh, UK) was used to calculate fluxes. The net ecosystem exchange (NEE) was then calculated as the sum of turbulent fluxes of CO_2 (Fct) and changes in CO_2 storage (Fcs). Flux filtering, quality controls and gap filling were performed following the CarboEurope-IP recommendations. NEE was partitioned into gross primary production (GPP) and ecosystem respiration (R_{eco}) components (Béziat et al., 2009).

The evapotranspiration (ETR) measurements were partitioned into evaporation (E) and transpiration (TR) by means of a statistic method based on marginal distribution sampling (MDS) (Béziat et al., 2009; Béziat et al., 2013).

Concomitantly to the flux measurements, standard meteorological variables were recorded on each site to analyze and calculate turbulent fluxes. Different radiation components were measured: incoming and outgoing short-wave and long-wave radiation with a CNR1 (Kipp & Zonen, Delft, NL); net radiation with a NR-lite (Kipp & Zonen, Delft, NL); incoming photosynthetic photon flux density (PPFD) with a PAR-lite (Kipp & Zonen, Delft, NL) and direct and diffuse PPFD with a BF2 (Delta-T, Cambridge, UK) (Béziat, 2009).

Air temperature and relative humidity were measured with a Vaisala probe (HMP35A, Vaisala, Helsinki, Finland). Precipitation was recorded with a ARG100 rain-gauge (Environmental Measurements Ltd., Sunderland, UK), atmospheric pressure with a BS4 sensor (BS4, Delta-T, Cambridge, UK), wind speed and direction with a 014A wind speed sensor and a 024A wind direction until 2009 and with a windsonic (Campbell) thereafter.

Soil temperature, moisture and heat flux profiles were measured with TP107 (Campbell Scientific Inc, Logan, UT, USA), CS616 (Campbell Scientific Inc, Logan, UT, USA) and HFP01 (Hukseflux, Delft, NL) probes, respectively. Three measurement repetitions were performed on each site at depths of 0.05, 0.10, 0.30 m and a single measurement was taken at depths of 0.60 m and 1 m at Auradé and Lamasquère,

respectively (Béziat, 2009). In recent years, some of the sensors were replaced by other models in order to match the ICOS criteria.

2.3.1.4. Spatial representativeness of the flux measurements and remote sensing-based flux estimates

The spatial representativeness issues cannot be neglected when comparing/validating remote sensing based modeling approach with flux towers data. The fluxes measured by the Eddy-Covariance (EC) towers represent the integrated fluxes within the tower footprint area. The EC measurements are based on the hypothesis that the contributing area of the fluxes is topographically flat and the vegetation is uniformly distributed (Baldocchi, 2008; Baldocchi, 2003; Chen et al., 2012). However, in fact, it is rare to find ideal sites respecting these hypotheses. EC data are usually influenced by the sites heterogeneity; thus the analysis of the spatial representativeness of the measurements might be useful for better understanding and using the EC flux data. Footprint models have been developed to estimate the probability of fluxes originated from a particular place surrounding the tower.

A footprint study was performed over the Auradé site. We used the Horst & Weil footprint model (Horst, 1999; Horst and Weil, 1994) for estimating the contribution, per unit surface flux, of each upwind surface area to a measured vertical flux. Therefore the flux footprint can be defined as the relative weight given to each elemental surface flux (Horst and Weil, 1994). The EC flux tower's footprint generally varies depending on wind speed, wind direction, atmospheric stability, and canopy and tower heights (Chen et al., 2012). For this study, as outputs of the Horst & Weil model, the footprint is represented over a pre-defined matrix, with 8m-resolution pixels (to be compatible with the Formosat-2 images resolution).

We evaluated then the footprint climatology of the Auradé site in 2006, during the wheat crop season. Figure 2.8 shows the estimated average monthly climatology of the footprint, from January until September 2006. These maps illustrate the average value of contribution of each pixel to the flux measurement. We observe that the monthly footprint climatology is asymmetrically distributed around the tower and that it varies from one month to another. These changes are mainly due to the variation of the atmospheric conditions and also to the changes in canopy height. We note that in the periods of strong wheat development (thus increasing crop height), here from April until June, the footprint is situated closer to the tower. In June, for example, when the winter wheat has reached its maximal development, the maximal contribution of a single pixel close to tower can be up to 2%. Besides, we see that the areas that contribute the most to the flux measurements correspond to those within the dominant winds direction. Even if in these maps, the footprint is represented completely inside the Auradé field, some pixels situated outside the

field can also contribute to the flux measurements. We quantified then the proportion of the measurement coming from the exterior. This study demonstrated that the outside pixels have an extremely weak individual contribution of approximately 10^{-3}% and that their total contribution in a year basis is about 9 to 12%, in average. It means that generally around 90% of the measured fluxes come from inside the field, which is encouraging.

For this thesis, we worked at the field scale; thus we compared the fluxes simulated by the crop models for each crop field (Auradé and Lamasquère) with the fluxes measured by the flux towers. Good agreement was found between observed and estimated fluxes (see sections 5.3 and 5.4), which allowed validating our modeling approach. Nevertheless, for a more precise validation of the models, we recommend future works to run the crop models (SAFY-CO_2 and SAFYE-CO_2) at the pixel-scale and to combine the models outputs with a footprint modeling, in order to account for the heterogeneity over the EC flux footprint area. Therefore to avoid any spatial scale mismatching it would be valuable to use a footprint weighting approach to match satellite pixel-scale and the EC measurements scale.

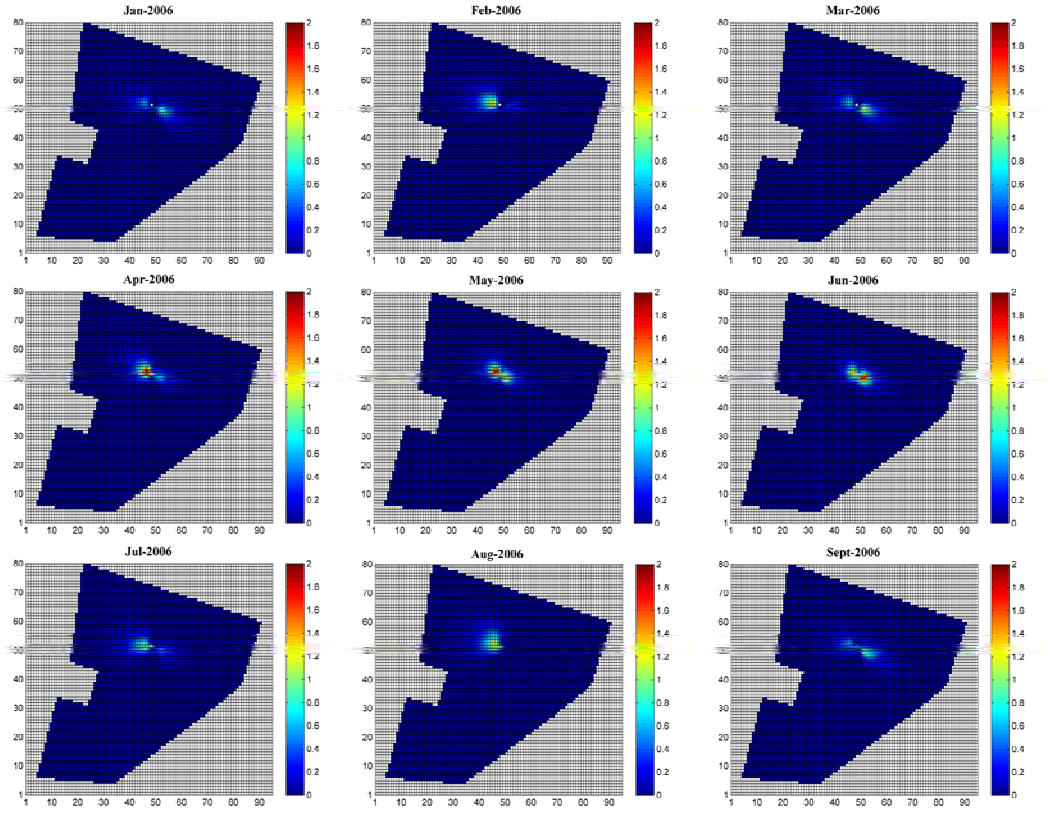


Figure 2.8- Average monthly footprint climatology for the Auradé site, during the 2006 wheat crop season. The position of the EC tower is indicated by a white dot. The climatology of the footprints from January to September are displayed. The x and y-axis represent the number of pixels of the represented scene. Each pixel has 8m resolution. The colorbar indicates the individual contribution (in percentage) of each pixel.

2.3.2. Intensive field campaigns

Over the years, some intensive field campaigns were accomplished by the CESBIO team, in order to have specific field data or to complete the actual set of field data from the experimental sites. Given that our objective is to study crop development and associated processes at regional scale, it is important having the maximum available field data, distributed over the whole study area. Four main campaigns were performed in order to measure LAI and biomass. One in 2008, one in 2009 and another in 2010, focusing on summer crops; and the most recent one in 2011, in the scope of this thesis, focusing on winter crops. Additionally, we have exploited some hemispherical photographs taken over wheat fields around Lamasquère in 2006.

We introduce here the concept of `ESU', standing for E*lementary Sampling Unit*. We call hereafter ESU the plots where biomass or/and green leaf/plant area index measurements were performed. The measurements over each ESU were done according to the VALERI sampling protocol [http://w3.avignon.inra.fr/valeri/], which states that the samples must be taken over a square elementary homogeneous surface. For our study, each ESU has 20m x 20m square area, corresponding to the dimensions of a pixel (SPOT). The coordinates of the center of the ESU must be recorded with a GPS, so the ESUs can be well situated in the satellite images, in order to link the *in-situ* data with the remote sensing data.

The 2008, 2009 and 2010 campaigns focused on summer crops: sunflower, maize and soybean. During the campaign, each ESU was sampled with 13 DHPs. Figure 2.9 illustrates the location where DHPs were taken within the ESU. The photographs were taken 7 to 10 times during the growing season. In 2008, the biomass was estimated from 10 plants collected near the ESUs through the vegetative development. In 2010 a total of 25 plants par ESU were collected around the ESU center (for more details see Claverie, (2012)).

2.3.2.1. The 2011 Field Campaign

After the field campaigns focusing on summer crops, we decided to improve the set of *in-situ* data for winter crops. Having a special interest on the yield information, in order to improve the validation of crop models yield estimates for winter crops, and given that yield destructive *in-situ* data were inexistent so far, we organized a campaign for measuring crops biomass at maturity stage (just before harvest) and final grain yield as well. The exceptional dry character of the season was an additional incentive for performing this campaign in the 2011 year. The ground data collected in a dry year would allow testing the robustness of our models to respond to extreme climatic conditions.

We selected 22 fields over our study area: 16 plots cultivated with winter wheat, 1 plot with barley and 5 plots containing rapeseed. The fields were chosen in a way to obtain a dataset covering a large range of yield/biomass values. For some fields, where remarkable heterogeneity was observed based upon NDVI analysis in spring time with F2 and SPOT images, two ESUs were assigned. Thus, for winter wheat 7 fields (of the 16) were sampled twice, counting a total of 23 sampled ESUs. The location of the sampled fields over the 24km^2 window (from F2 images) can be seen in Figure 2.10. For this thesis, we used only the data collected over wheat fields.

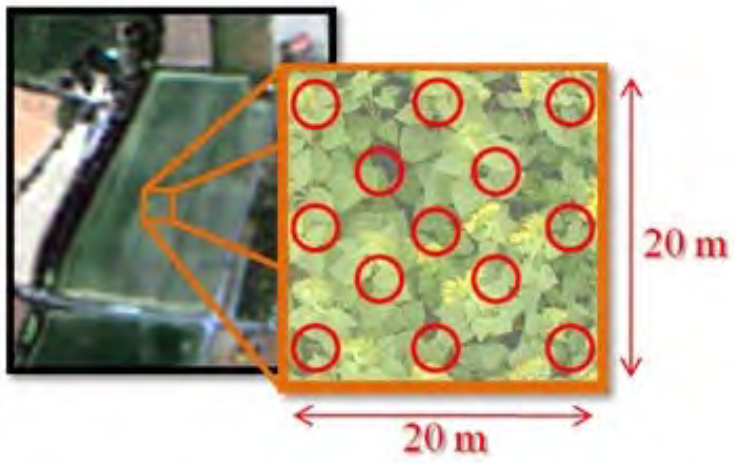


Figure 2.9- ESU sampling protocol: distribution of the 13 DHPs.

2.3.2.1.1. Measurements protocols

Next we describe briefly the protocols adopted for this campaign for collecting field data, according to each crop type.

a) **Wheat and barley:** For each ESU, five plots were sampled in order to have a good spatial representativeness. One plot was located in the center of the ESU and the four others at each ESU's corner. In each plot we collected 4 rows over 50cm length. The distance inter-rows was approximately 13cm, which results in 0.25m² sampling surface for each plot. The data of all five plots were then averaged, resulting in one mean measurement per ESU.

b) **Rapeseed:** As for wheat and barley, 5 plots were sampled for each ESU. In each plot we collected 1 row over 100cm length. The distance inter-rows was approximately 50cm, which results in 0.5m² sampling surface for each plot.

As it was described for the destructive measurements performed over the experimental sites, the *in-situ* data processing was essentially the same for this campaign: the organs of the collected plants were separated into leaves, straw, stems and ears or siliqua (for rapeseed) and dried in an oven during 48 hours. The differences are that i) no leaf area index measurements were done, given that the samples were collected once, at the end of the senescent phase (no green leaves), and ii) that destructive grain yield was measured after threshing, information that was not available for the experimental sites nor the previous intensive campaigns.

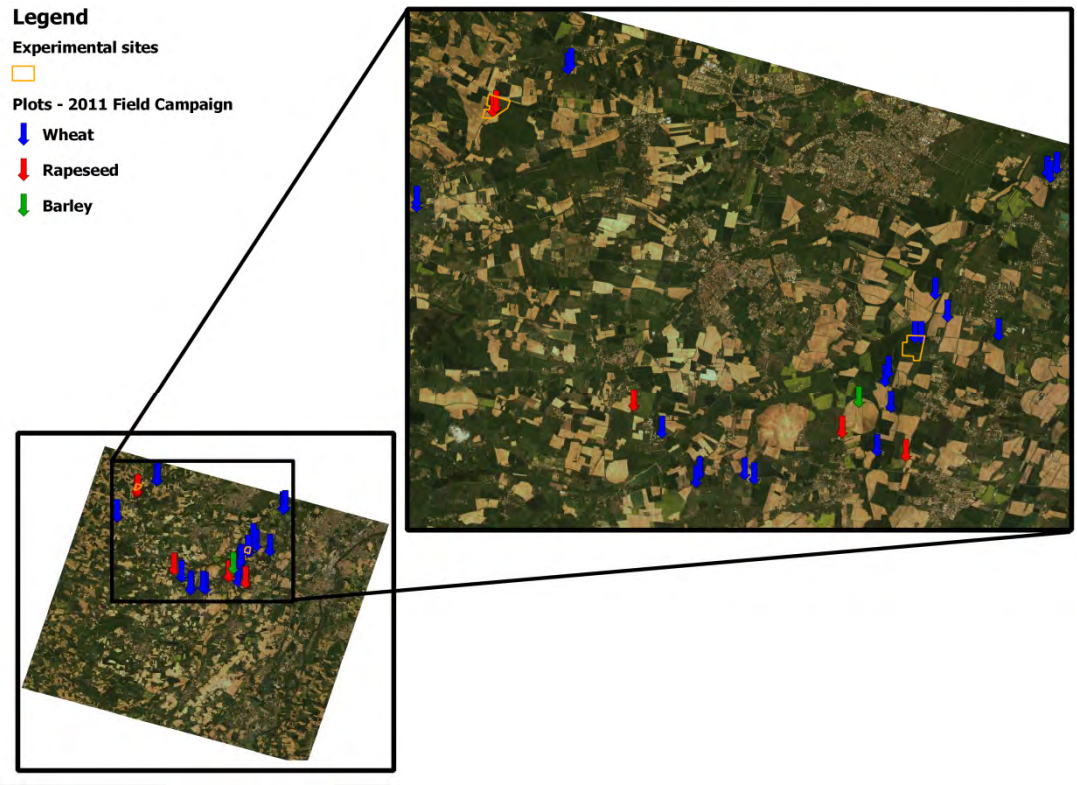


Figure 2.10- Location of the fields where biomass and yield destructive measurements were performed during the 2011 field campaign. Wheat plots are indicated by the blue arrows, rapeseed by the red ones and barley by the green one. A Formosat-2 image is displayed as background.

2.3.2.1.2. Analysis of biomass and yield collected data

The analysis of the biomass and yield measurements over the 23 wheat ESUs indicates large diversity. For final biomass, as observed in Figure 2.11, the values can vary from 500 $g.m^{-2}$ up to 1700 $g.m^{-2}$. The obtained mean biomass (and standard

deviation) is 1092 ± 344 g.m^{-2}. This behavior is also observed when considering the grain yield measurements, with values ranging from 1.87 t.ha^{-1} up to 8.39 t.ha^{-1}, with mean yield of 4.93 ± 1.86 t.ha^{-1} (Figure 2.12).

It is interesting to note that the heterogeneous aspect (visual) of the fields that were selected to be sampled with two ESUs is confirmed by the destructive measurements. Figure 2.13 shows the measured grain yield for each pair of ESUs within a same field. For example, for the first case, the ESU CA presented a yield of 8.4 t.ha^{-1} while the ESU CB (in the same field) presented a yield of only 6.1 t.ha^{-1}. Figure 2.14 shows the observed heterogeneity within-fields based on a false-color composition of a SPOT5 image for three paired plots. The SPOT image was taken on 21/05/2011, which corresponds to maximum development and beginning of senescence. The intra-field differences observed at this stage were reflected on the final grain yield and biomass, as it is illustrated in Figure 2.12.

From this data set we calculated harvest index for the winter wheat plots. The harvest index (HI) is defined as the ratio of final grain yield over the total final biomass. The HI was calculated for each of the 23 plots and a mean HI equal to 0.45±0.05 was found (Figure 2.15).

Besides, we observe the HI values have lower standard variation than those obtained for biomass or yield values. We performed a traditional regression as well, by plotting the biomass against yield measurements, where the curve's slope gives the HI value (Figure 2.16). The value found for the winter wheat harvest index will be useful for evaluating the performance of our crop models for estimating the final grain (in Chapter 4).

The causes of the observed intra-field heterogeneities might be various: as differences in soil depth, texture or moisture, or even in ground slope orientation. This topic, however, is not part of the main objectives of this thesis and will be investigated in depth by a recently ongoing PhD thesis (Enguerrand Burel, Cesbio, Toulouse, France).

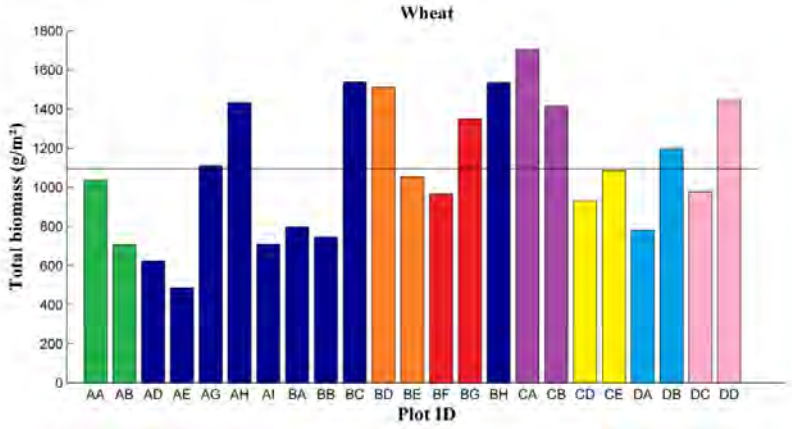


Figure 2.11- Final biomass measured for the 23 winter wheat plots collected on the 2011 field campaign. The black horizontal line indicates the mean final biomass for these plots. The 7 fields that were sampled with two ESUs each are colored by the colored bars in green, orange, red, purple, yellow, light blue and pink. The others fields are displayed in dark blue.

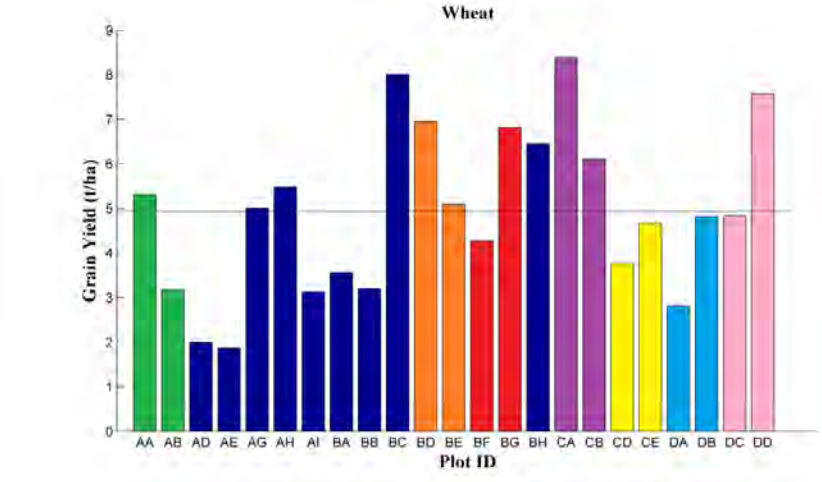


Figure 2.12- Grain yield measured for the 23 winter wheat plots collected on the 2011 field campaign. The black horizontal line indicates the mean yield for these plots.

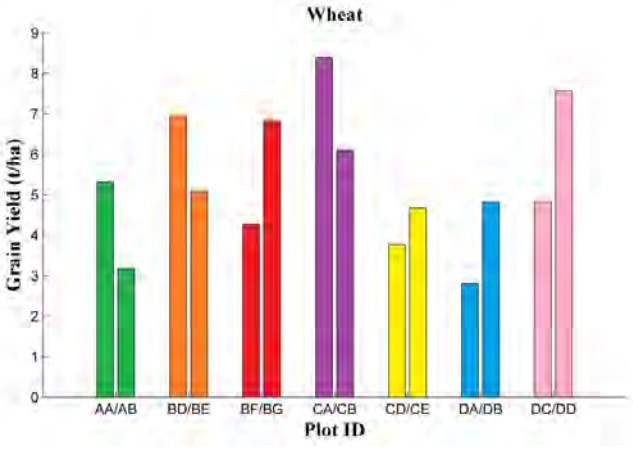


Figure 2.13 Zoom on the grain yield measured for the 7 paired ESUs located in fields characterized as spatially heterogeneous, and thus which were sampled twice.

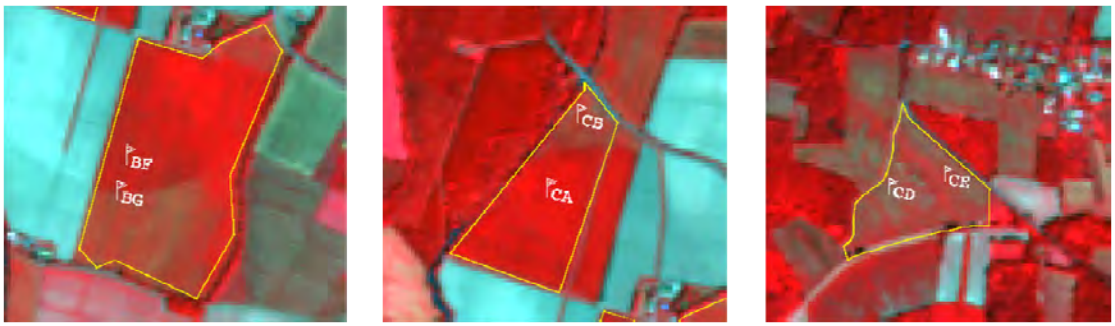


Figure 2.14- Example of the observed heterogeneity within-fields. The location of the ESU center measurements is indicated by the flags. A false-color composition of a SPOT5 image (10m resolution) taken on 21/05/2011 is displayed on the background. The yellow lines indicate the contours of each field.

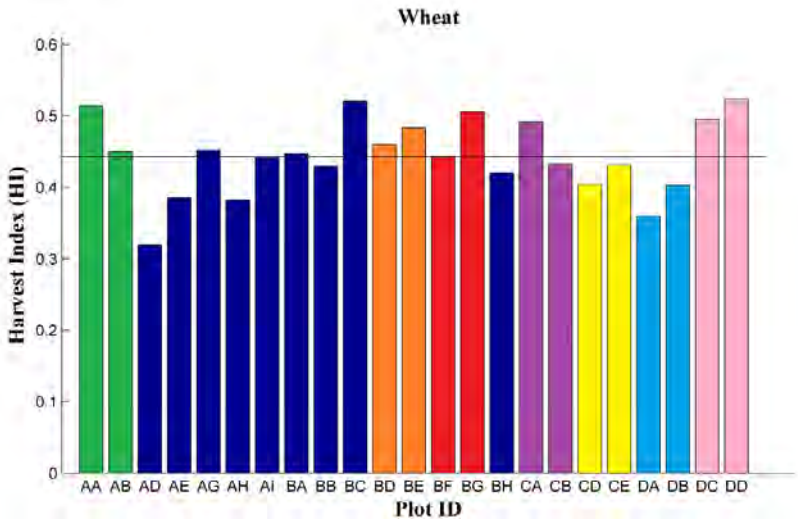


Figure 2.15- Harvest index calculated for the 23 winter wheat plots. The red horizontal line indicates the mean harvest index (HI=0.45±0.05).

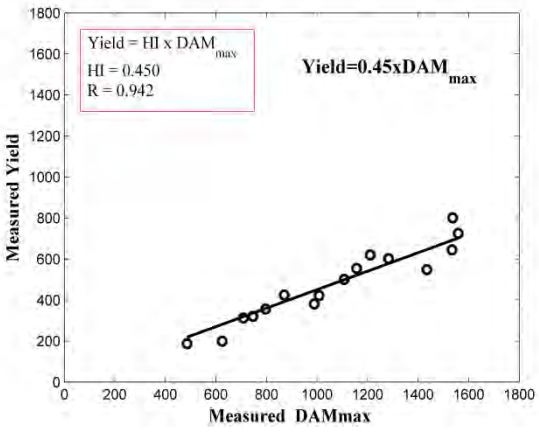


Figure 2.16- Relationship between measured maximum biomass (at the end of season) and observed yield. The slope of the linear regression gives the harvest index (HI), which was found to be HI=0.45, with a correlation of R=0.94.

2.4. Data from collaboration networks

2.4.1. Farmers' surveys

Another important *in-situ* dataset we have comes from the farmers' surveys. The Regional Space Observatory (OSR), recently created and managed by CESBIO, conducts these surveys. One of the OSR tasks is to collect and organize *in-situ* and remotely-sensed data in order to support pluridisciplinary research works. The interviews intend to obtain current information on crop fields surface, on farming practices (as crops grown, tillage practices, nutrient and pesticide application), on crop varieties, dates of sowing and harvest, irrigation practices, and on final yield at harvest. The OSR developed particularly good relations with two famers' networks: one in the zone around the experimental site of Lamasquère, including around 20 fields, and the other one next to Auradé site, where we cooperate with the Association des Agriculteurs d'Auradé since 1993, an organization composed by a group of 36 farmers.

The information collected through these surveys can be useful for a better understanding of the agricultural practices of our study area, as well for validating/comparing our models outputs or even for data input in the models. For this work, we are mainly interested in the final grain yield data.

2.4.2. Yield monitors

Yield monitors are a recent development in agricultural equipment that allows farmers to assess the effects of weather, soil properties, and management on grain yield. They are an essential tool for those who want to practice "precision agriculture", performing site-specific crop management. The accuracy of these devices depends on appropriate installation, calibration, and operation [http://www2.ca.uky.edu/agc/pubs/pa/pa1/pa1.htm].

The yield monitor aims at giving the farmer an accurate assessment of how yields vary within a field. The system (mounted on harvesting machines) is able to measure grain mass (wet and dry) or volume, grain moisture content (and temperature), area harvested, productivity (i.e. area harvested per hour), and eventually other parameters. The information can be displayed for instantaneous use. Therefore these devices can provide useful information for agricultural researches.

For our study, we obtained a data set of intra-field yield maps, measured with yield monitors installed on the combine harvesters. This data set was provided by Agri-Intranet [http://www.agri-intranet.com/]. However, they are located out of our Fomosat-2 window, but inside the SPOT window (Figure 2.17). We could only exploit the data acquired in 2011, since for this year we have enough SPOT images to cover

the winter wheat vegetative period. A total of seven wheat fields were analyzed and used for validation of the yield estimations provided by our models. Figure 2.18 (left) shows an example of the yield map produced by the yield monitor for one of the seven fields. For these called “raw” maps, we have a yield measurement every 1.5m or 2m. In order to have 10m resolution pixels (the same as SPOT5 images), we aggregated the yield maps by calculating the mean of all the measurements points within each 10m-side pixel. Figure 2.18 (right) illustrates the aggregated map of a selected field. The borders of the fields were excluded of the analysis in order to avoid the deviant points related to the maneuvering of the combine harvesters.

The great advantage of this set of yield *in-situ* data is to provide an information at intra-field scale, allowing to study the within field variability. Besides, these data allow validating the models outputs over a large range of yield values, wider than the mean yield information provided by the farmers’ surveys.

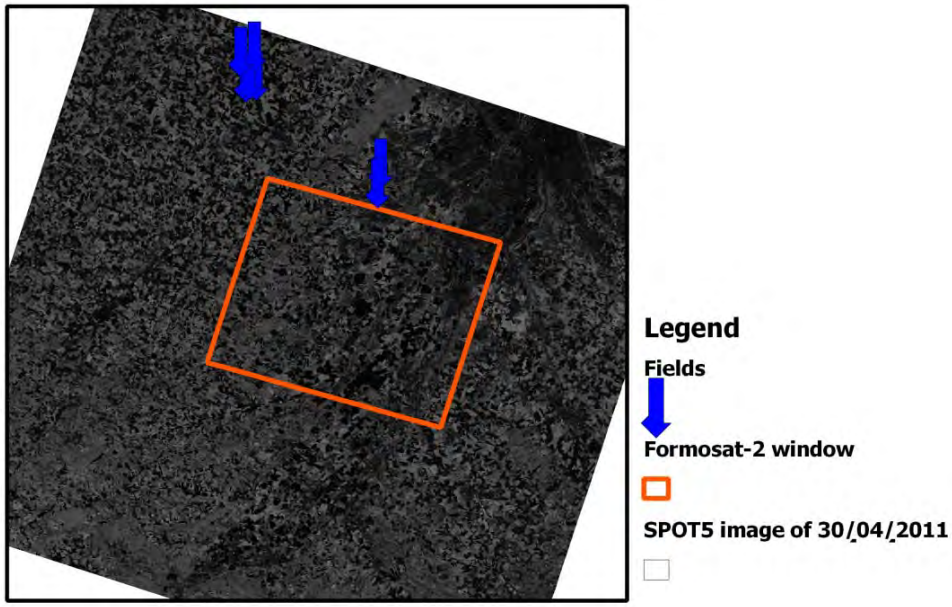


Figure 2.17- Location of the fields where yield maps are available. The Formosat-2 window (24km^2) is displayed in orange. The background image (60km^2) was acquired by the SPOT5 satellite on 30/04/2011.

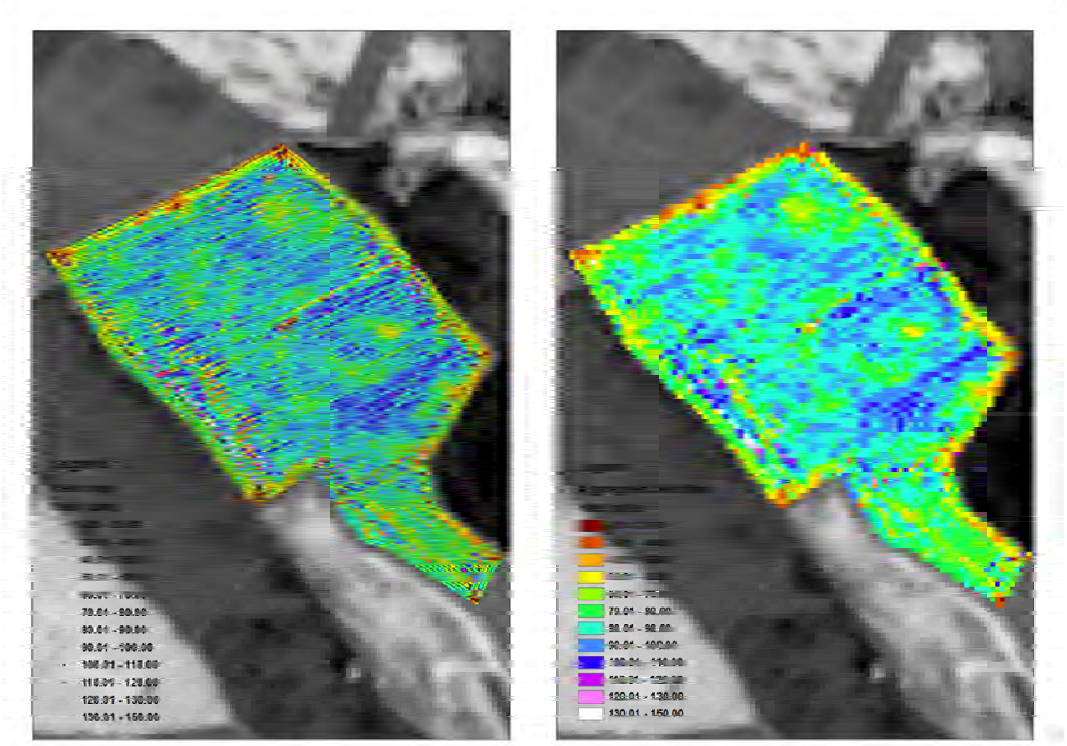


Figure 2.18- Example of yield map produced by the yield monitor (left). Measurement points are irregularly distributed and separated by approximately 1.5m or 2m. On the right, the yield values are averaged in order to have aggregated maps with 10m pixels resolution (equivalent to SPOT5 pixels resolution).

2.5. Regional data

2.5.1. Soil maps

The soil characteristic maps of our study area (Figure 2.19) were generated by Jean-François Dejoux (CESBIO), by the combination of different sources of soil features. Therefore these maps are the synthesis of different soil datasets, each one presenting specific spatial characteristics (resolution, location, accuracy, ...). The datasets sources and their main features are listed in Table 2.5. Through this work the soils were classified in five classes, representing the major soil types of our zone. These classes are: the alluvial soils of the Garonne and Touch rivers, the Garonne terrasses (boulbenes soils), the clayey limestone hillsides of Gascogne and of the Volvestre, the colluvial deposits and the non-calcareous molasses soils. The available data permitted establishing a soil clay percentage map and a soil depth map, as well. It should be noted that these maps demarcate the general soil characteristics and that, actually, a wider spatial variability is observed in the field.

These soil clay and depth maps will be needed in Chapter 5, as inputs of the SAFYE-CO_2 models for estimating crop water requirements. They were resampled in order to have spatial resolution of 8m, like the satellite data.

Table 2.5- Sources of different soil maps used for the compilation of a unique soil map of our study area [adapted from Claverie,(2012)].

Denomination	Source	Extent / Scale	Features
Wide morphopedological data set of the Midi-Pyrénées region	Chambre Régionale Agriculture Midi-Pyrénées (CRAMP) (1995)	Midi-Pyrénées region (1/500000)	Unique homogeneous data set, covering the whole study area.
NESTE data set of the small Pyrenees' and Saint-Lys terraces	CACG: soil recognition study n° 31 (1968) and Ignacio Soto thesis (2005)	200 000 ha in Haute Garonne department (1/500000)	More accurate data, covering the west part of the study area.
SAVE data set	CACG: soil recognition study n° 26 (1965)	114 Mha (1/500000)	Simplified map describing the extreme west of the zone of study.
Expertise	Brice Corrège (Chambre d'agriculture 31)	7 Mha (1/200000)	Expertise data that identify the superficial zones within the east part of the study area

2.5.2. Land use maps: the RPG

The cultivated plots were identified by means of the RPG (*Registre Parcellaire Graphique*). The RPG is a detailed land use database describing the cultivated fields (surface, location, crop species and management). It is available every year, since 2006, and the Agency for Services and Payment (ASP) is in charge of its distribution. The ASP is a French public institution responsible for the implementation of national and European public policies.

Every year the farmers have to assign the shape/limits of their fields and also to inform a series of characteristics of their exploitations, as: cultivated crops and associated surfaces, and presence or not of irrigation. The ASP commercializes these databases, keeping the identities of the farmers confidential.

The RPG elementary unit called *îlot* (in French) consists in an ensemble of adjacent fields cultivated by the same farmer. An *îlot* can enclose more than one crop type (if two or more fields with different cultivated crops are contiguous). They are called mixed *îlots*. Given that it is difficult to automatically distinguish and separate the crops inside a mixed *îlot* and the fact that mixed *îlots* correspond to a small percentage of the study area, we decided to work only with the 'pure *îlots'*, where a single crop type is cultivated. Figure 2.20 shows an extract of the RPG 2011, where the main crops are exhibited.

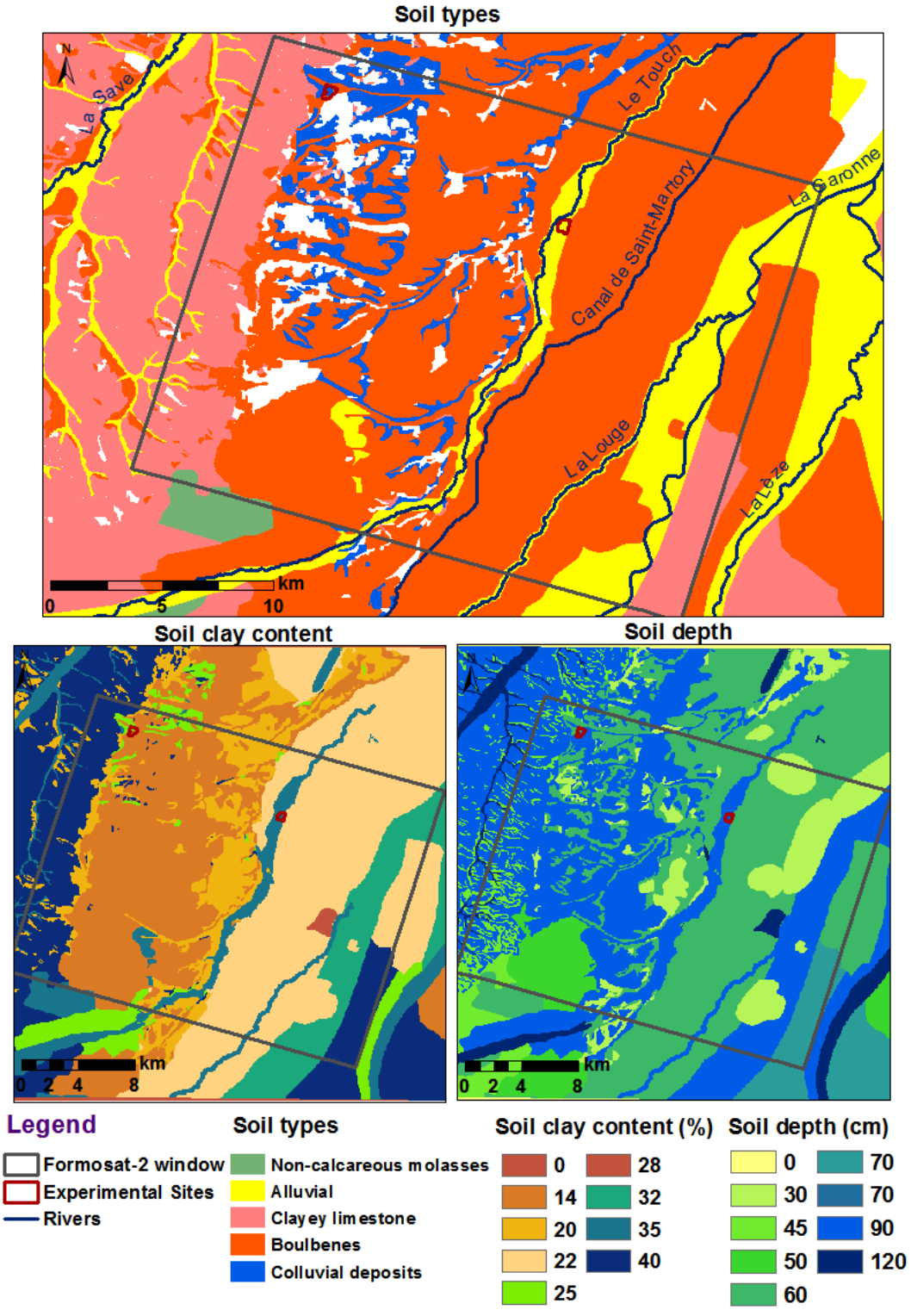


Figure 2.19- Soil maps. O the top the map illustrates the five main types of soil present in the study area. On the bottom: the soil clay content (left) and the soil depth (right) maps. The position of the experimental sites is indicated by the red contours.

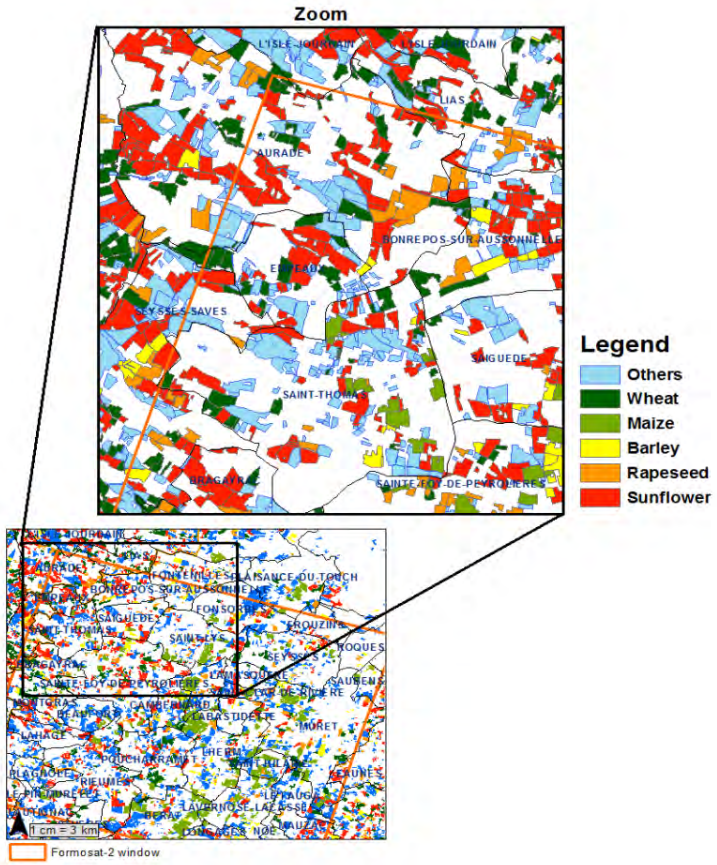


Figure 2.20- Extract of the RPG 2011 over our study area. The top window shows a zoom over few fields. The wheat, maize, barley, rapeseed and sunflower crops are discriminated; all the other crops types are classified as 'Others'. Only the 'pure îlots' are displayed.

2.5.3. SAFRAN Meteorological data

The SAFRAN meteorological data (Durand et al., 1993), produced by Météo-France, were used for this study. The SAFRAN data provide air temperature at 2 m above the ground (Ta), incoming global radiation (Rg), precipitation (P), air relative humidity at 2m (RH) and wind speed at 10m height (u_{10}) based on weather stations measurements and modeling. The data are available every 6 h over a 8 km spatial resolution grid (see Figure 2.1, the SAFRAN grid is designated by the black plus symbol). The climatic variables daily mean Ta, Rg and P (which are the variables required by our crop models) were calculated for each pixel over the study area using a bilinear interpolation.

2.5.4. Agreste Yield Statistics

In order to have yield data at a regional scale we used the Agreste agricultural statistics data presented as Annual Agricultural Statistics (SAA, Statistique Agricole Annuelle, www.agreste.agriculture.gouv.fr). The regional institutes of the Ministry of Agriculture, Food and Forestry are in charge of these statistics. They collect information concerning land use, cultivated surfaces and agricultural production (yield). Data are classed by crop type and are available for all the French departments.

For this thesis, we are interested in the yield statistics for winter wheat over the Haute-Garonne (mainly) and Gers departments, given that our Formosat-2 window covers them. Wheat data are separated in two groups: durum wheat (*Triticum durum*) and common wheat (*Triticum aestivum)*. As we don't differentiate varieties for our studies, we calculated the weighted average yield data, according to the surface area cultivated with each one of the varieties. We obtained thus a single wheat data set. Data from 2006 until 2011 were used. The mean-weigthed yield values for this period in Haute Garonne and Gers department are displayed in Figure 2.21. We observe that mean wheat yield can vary from 4.2 till 5.9 $t.ha^{-1}$ in the study area, depending on the year (and its climatic conditions) and on the department (soil and management characteristics).

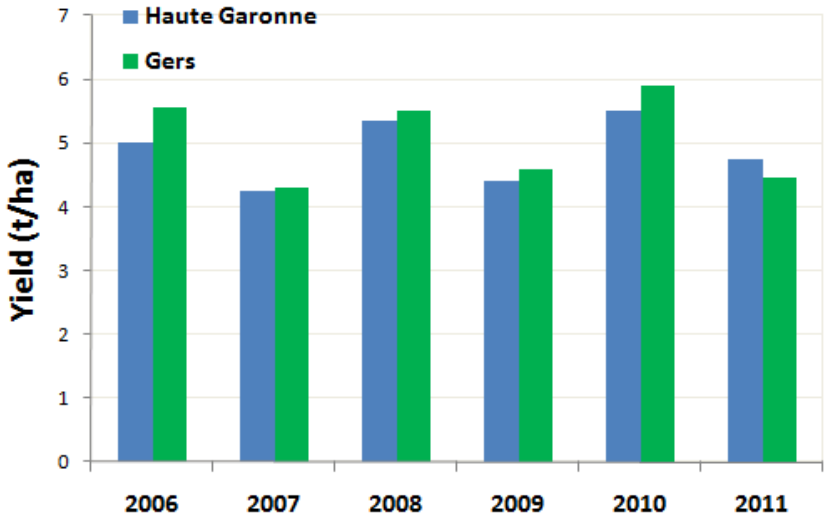


Figure 2.21- Yield statistics from the Agreste database for the Haute-Garonne and Gers departments of France, from 2006 to 2011.

3. Retrieving Green Area Index

This chapter presents the methods used for retrieving green area index (GAI) from satellite images. We start by a brief description on the subject. Next section gives a description of the two investigated approaches: a) an empirical approach for estimating GAI from NDVI images and ground data; and b) a physical approach, based on the inversion of a radiative transfer model using neural networks. Subsequently the validation of both methods by means of an *in-situ* data set over our study area is presented. The chapter is concluded with an inter-comparison of approaches and conclusions of the study.

3.1. Introduction

Recent advances in technology have opened the possibility to have Earth observation imagery combining both high spatial and temporal resolutions (Formosat-2, RapidEye, future missions Venμs and Sentinel-2). This kind of data may provide new perspectives for land surface monitoring, as biophysical information can be retrieved from it.

Remote sensing observations sample the radiation reflected or emitted by the surface, and thus do not provide directly the biophysical characteristics required by the crop models for describing some state variables of the surface (GAI, biomass…). An intermediate step is therefore necessary to transform the remote sensing measurements into estimates of the surface biophysical parameters (Baret and Buis, 2008).

The surface of green foliage is the main interface between atmosphere and vegetation, governing most of the exchanges of energy, water and carbon (Duveiller et al., 2011). The total green leaf surface of a canopy is often quantified by a dimensionless variable called Leaf Area Index (LAI). LAI is defined as one half the total developed area of green leaves per unit of ground horizontal surface area (Chen and Black, 1992). If the others canopy elements are considered, then the appropriate term is PAI (Plant Area Index), which is defined as the total developed area of all plant elements per unit horizontal ground area, *i.e*. including trunks, branches, stems and reproductive elements (Baret et al., 2010). Since the satellite observations allow to detect the total green contribution of vegetation, but not to distinct the different green plant elements (leaves, stems,…), we adopt the concept of Green Area Index (GAI), proposed by Baret et al., (2010). The GAI concept appears as an alternative to LAI and PAI. GAI considers only the photosynthetically (green) active plant area with no distinctions between leaves, stems and reproductive organs.

GAI is a difficult variable to measure, due to its spatial and temporal variability. Direct and indirect methods for ground-based GAI measurements do exist (see (Bréda, 2003; Demarez et al., 2008; Jonckheere et al., 2004)), but much expectation is given to GAI estimations from satellite remote sensing, due to the large spatial extent and dynamic coverage that can be provided by satellite imagery (Duveiller et al., 2011).

Many methods have been proposed to retrieve GAI from remote sensing observations. In literature we can find from simple empirical models with calibration over experimental data sets, up to more complex ones based on the use of radiative transfer models (Baret and Buis, 2008).

In this chapter we present two of these methods, an empirical and a complex one, their comparison against ground measurements and finally a comparative analysis inter-methods.

3.2. Methodology

3.2.1. Empirical approach

This approach was one of the first methods to link satellite reflectance information with canopy variables. The reflectance bands are generally combined into vegetation indices (*VI*), designed to minimize the influence of confounding factors such as soil reflectance and atmospheric effects (Baret and Guyot, 1991). Therefore the use of spectral vegetation indices intends to enhance the information contained in spectral reflectance data, by extracting the variability due to vegetation characteristics and to minimize soil, atmospheric, and sun-target-sensor geometry effects (Moulin, 1999; Viña et al., 2011).

The relationships between the vegetation indices and canopy variables are calibrated using *in-situ* observations (Baret and Buis, 2008). Nevertheless this approach is limited by the need of a large and representative amount of ground data for establishing a valid empirical relationship. The use of this approach is thus restricted by the difficulty in obtaining a training data base that represents the whole range of possible conditions encountered over the studied surfaces, such as combinations of geometrical configurations, type of vegetation, variability in vegetation development stages and several stress levels.

The NDVI (Normalized Difference Vegetation Index, Myneni et al., (1995); Rouse et al., (1974)) is an index often used by the scientific community to characterize vegetation (Asrar et al., 1984; Huete, 1988). The NDVI is defined as the ratio between the difference of the infrared and red reflectance bands and the sum of the same ones [Eq. (3.1)]. The use of the NDVI reduces the anisotropic effects from the surface because the directional signatures are similar in its wavebands (red and

infrared). For its ease of computation and recognized value in crop monitoring, we chose the NDVI in spite of other indices (as EVI, TSAVI, NDWI, and so on).

For this study, an exponential relationship was used to relate NDVI and GAI [Eq. (3.2)]. This formulation differs from the commonly used logarithmic equation proposed by Baret et al., (1989). Nonetheless, the chosen formulation fits correctly with the *in-situ* measurements of GAI, as it was also showed by Claverie (2012) for the same study area.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{3.1}$$

$$GAI = \alpha \times (e^{\beta \cdot NDVI} - e^{\beta \times NDVI_s}) \tag{3.2}$$

where $NDVI_s$ is the bare soil NDVI. It was fixed to 0.13.

The parameters α and β of equation (3.2) were calibrated using *in-situ* GAI estimated from the digital hemispherical photographs (with Can-EYE), by minimizing the Root Mean Square Error (RMSE) between estimated and *in-situ* GAI.

Given that data from two satellites (Formosat-2 and SPOT (2, 4 and 5)) were available and jointly used to parameterize the empirical law, the compatibility between these two data sets was firstly investigated. This pre-processing step aimed at having a complementary and homogeneous set of remote sensing data, which would allow appropriate calibration and validation of the empirical method.

3.2.1.1. Comparison of Formosat-2 and Spot data

The use of multi-sensor data is strongly recommended for having continuous time series of satellite images. However, there are several differences between the sensor systems in addition to their spectral bands responses differences. These additional differences are mainly due to differing orbital characteristics and spatial resolution.

As we are interested in satellite NDVI products for establishing the empirical relationship linking NDVI and GAI, the NDVI products of Formosat-2 and SPOT satellites were compared. For this purpose, the SPOT pixels (20m or 10m) were re-sampled to the F2 pixels size of 8m × 8m, using the nearest neighbor method. A field scale comparison was performed, by calculating the spatial average of NDVI of each crop field over the study area. The mean NDVI was calculated based on the field's contours determined by the RPG (*Registre Parcellaire Graphique* in French, which corresponds to land parcel information system).

Because of the difficulty of having SPOT and Fomosat-2 acquisitions exactly on the same day, images acquired within time intervals of +/- 3 days between the sensors

were accepted. Dates of acquisition are displayed in Table 3.1. View and solar angles of the paired images differed (listed on Table 3.1). Crop characteristics were assumed to be unchanged during the period between the paired acquisitions (maximum 3 days). No rain event was observed between the acquisitions.

The NDVI products were compared for the different crops encountered in our area of study. Only the dates within the period of vegetative development were used, from 2006 until 2010. Considering the Formosat-2 NDVI as the standard NDVI, a linear correction of the type $y = ax + b$ was applied to the SPOT NDVI to match the standard F2 (Steven et al., 2003). A linear relationship could be used for the intercalibration because both images datasets have been previously ortho-rectified and corrected for atmospheric effects. This intercalibration aims at obtaining a homogeneous NDVI dataset, regarding the future application in GAI estimates. The same could have be done to establish a SPOT standard product. This step allowed having an estimate of the correlation between the two sets of NDVI values. The possible influence of the different view angles on the residual errors between NDVI products was also evaluated (see section 3.3.2).

Table 3.1- Specifications of Spot and Formosat-2 paired dates of acquisitions: view and solar zenith (θ_v ,θ_s) and azimuth (ϕ_v , ϕ_s) angles

Date	Sensor	θ_v(°)	θ_s (°)	ϕ_v(°)	ϕ_s(°)
04/04/06	**SPOT**	-28	98	41	149
05/06/06	**F2**	22	349	27	133
23/06/06	**SPOT**	20	-75	21	152
23/06/06	**F2**	22	348	27	130
17/07/06	**SPOT**	-29	98	27	135
17/07/06	**F2**	22	348	28	131
23/11/06	**SPOT**	-1	101	64	168
23/11/06	**F2**	23	346	65	161
15/02/07	**SPOT**	-24	99	59	155
15/02/07	**F2**	26	226	60	153
06/07/07	**SPOT**	27	-72	22	150
07/07/07	**F2**	28	233	26	133
11/02/08	**SPOT**	21	-75	59	163
11/02/08	**F2**	34	245	61	155
20/06/08	**SPOT**	-16	103	25	134
19/06/08	**F2**	31	240	26	133
28/06/08	**SPOT**	-16	100	26	133
26/06/08	**F2**	31	240	26	132
19/03/09	**SPOT**	-22	99	48	150
17/03/09	**F2**	29	236	51	145
24/06/09	**SPOT**	26	-74	22	149
23/06/09	**F2**	29	236	29	124

3.2.2. Physical approach

Radiative transfer models (RTM) summarize our understanding of the physical laws governing the interactions between solar radiation and the canopy. They simulate the radiation field reflected or emitted by the surface for given observational configurations, once the vegetation and the background as well as possibly the atmosphere are specified (Baret and Buis, 2008). For one-dimensional RTM, the canopy is approximated to a turbid medium which is horizontally infinite and homogeneous but vertically finite and variable (Duveiller, 2011).

Retrieving canopy characteristics from remote sensing data demands an *inverse* of the causal relationship between canopy structure and reflectance that is usually addressed by radiative transfer models. The inversion process is inherently problematic since the number of unknowns is most often larger than the available independent remote sensing measurements. This undetermined nature compounded with the many uncertainties attached to the models and measurements causes the inversion problem to have many solutions that lead to the same results, making it an ill-posed problem (Combal et al., 2002; Duveiller, 2011).

Different techniques can be used to construct the inverse models. Artificial neural networks (NNT) are characterized by their efficient interpolation capacity over non-linear processes (Leshno et al., 1993). Several studies (listed by Baret and Buis, (2008)) demonstrated that this technique is suitable for generating and calibrating the inverse model, since it allows modeling complex non-linear and multi-variate functions.

The approach we present here, denominated physical approach, is based on the inversion of the radiative transfer model PROSAIL using neural networks (NNT). For this purpose, we used the *BV-NET* tool *(Biophysical variable neural network tool),* which was developed by the EMMAH team, from INRA Avignon. [www4.paca.inra.fr/emmah]. This tool is based upon the algorithm conceived by Baret et al., (2007) to derive the global LAI product developed within the CYCLOPES (Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites) project from SPOT/VEGETATION data. This algorithm allows building and calibrating an inverse model, by using a database simulated by the PROSAIL model (Baret et al., 1992b). The *BV-NET* tool is currently operational; still validation of the outputs is required. Since this tool was firstly developed for low/medium resolution applications, at present few studies have validated the GAI estimates obtained by this method using high spatial resolution images. We can quote: Bsaibes et al., (2009) for grasslands and crops, using Formosat-2 data; Verger et al., (2011) for crops using Chris/PROBA data; Duveiller et al., (2011) for wheat crop using SPOT data and Claverie, (2012) for summer crops using Formosat-2 data. For this study, we aim

at validating the BV-NET tool estimates for different crops (maize, sunflower, soybean and wheat) over our study area, using Formosat-2 and SPOT data. Some model parameters were modified to be compatible with the characteristics of our study area and remote sensing data.

This approach relies on three main steps (illustrated in Figure 3.1): a) the generation of a learning database, b) the training of the neural networks and finally c) the application of the created networks for estimating the requested biophysical variables (GAI in our case).

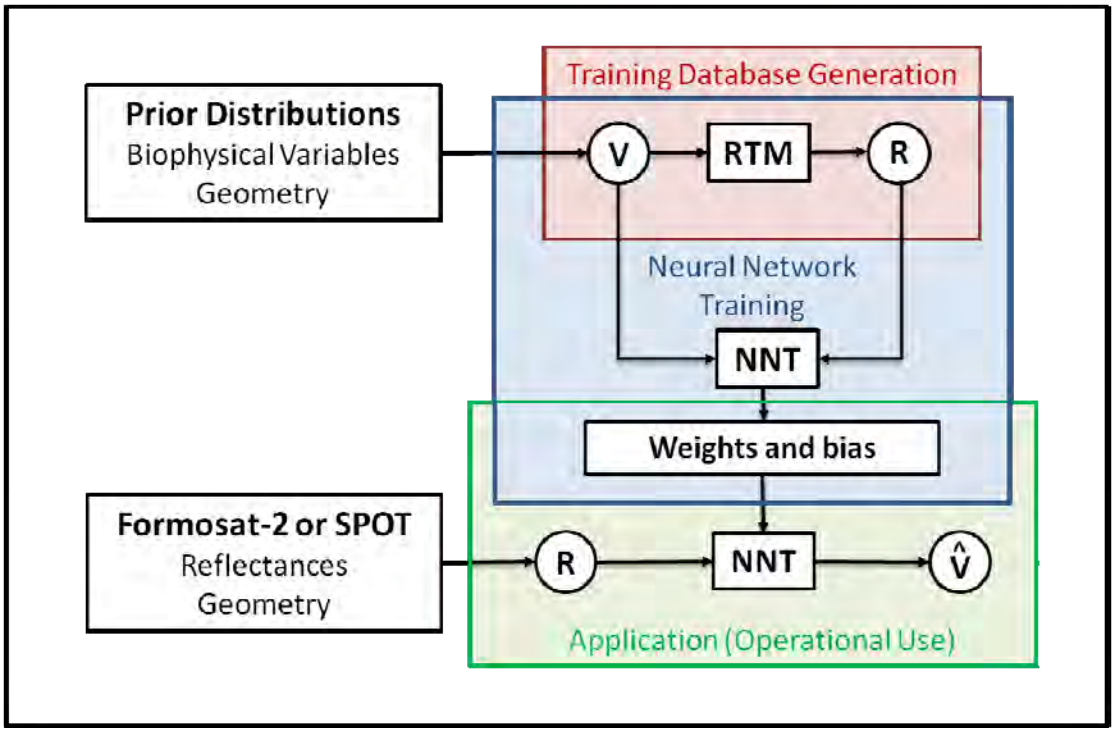


Figure 3.1- Diagram showing how the products $\widehat{V}$ (= estimated GAI, in our study) are generated operationally. NNT corresponds to the Artificial Neural Network characterized par its structure and its coefficients (synaptic weights and bias). V corresponds to the biophysical variable (=GAI) in the training data base and estimated by running the NNT over the simulated Formosat-2 or SPOT reflectance and geometry (extracted from Weiss and Baret (2010b)).

3.2.2.1. Generation of the training database

The training database consists on simulating a wide variety of canopy reflectances (which are the input variables) with the PROSAIL model, and linking then with the output variables (the biophysical variables: GAI, Fcover, Fapar). This work focuses on the GAI as the main output variable, but estimates of Fapar and Fcover were accomplished too.

3.2.2.1.1. The PROSAIL Radiative Transfer Model (Baret et al., 1992a)

PROSAIL is a combination of PROSPECT ((Jacquemoud et al., 2009)) and SAIL (Verhoef, 1984) models. The first one simulates leaf reflectances and transmittances, which are used as input variables for the SAIL model. PROSPECT requires input parameters describing leaf properties: the mesophyll structure parameter (*N*), chlorophyll content, both *a* and *b* (*Cab*), water content (*Cw*), brown pigment content (*Cbp*) and dry matter content (*Cdm*). The SAIL model provides directional reflectances, Fapar and Fcover variables. Input parameters describing structure of canopy and background soil reflectance are: green area index (*GAI*), average leaf angle (*ALA*), hotspot and brightness coefficient (*Bs*).

For all the cited input variables, we defined a variation interval. A full experimental plan was adopted to combine the input variables by splitting the range of variation of each variable into a small number of classes. The distribution of the input variables (minimum, maximum, mode, std and type) were based on the studies of Duveiller et al., (2011) and Claverie, (2012) and are presented in Table 3.2. Each variable distribution was divided into a number of classes, which depends on the importance of the variable on the radiative transfer processes. The product of all classes determines the total number of simulations. An amount of 204,800 cases were simulated.

Table 3.2- Input variables distribution of PROSAIL model, used for the learning database generation.

	Variable	Minimum	Maximum	Mode	Std	Class	Law
Canopy strucutre	GAI	0	8	2	2	8	Gauss
	ALA (°)	5	80	40	20	5	Gauss
	Hotspot	0.1	0.5	0.2	0.5	1	Gauss
Leaf optical properties	N	1.2	2.2	1.5	0.3	4	Gauss
	Cab	20	90	45	30	5	Gauss
	Cdm	0.003	0.011	0.005	0.005	4	Gauss
	Cw	0.6	0.85	0.75	0.08	4	Uniform
	Cbp	0	2	0	0.3	4	Gauss
Soil background property	Bs	0.16	1.3	0.586	0.14	4	Log-Normal

3.2.2.2. Neural networks training

The neural network is characterized by the type of neurons (the transfer function) and their architecture. In this work, a back-propagation network with two hidden layers is applied. The first layer has five neurons with sigmoid transfer function and the second layer has one neuron, with linear transfer function (Figure 3.2). The

combination of sigmoid and linear functions is capable of fitting any type of function (Weiss and Baret (2010a); Demuth and Beale, (1998)). The neural network input layer is composed of a) the angles characterizing the observation geometry (view and solar zenithal angles, $\cos(\theta_v)$ and $\cos(\theta_s)$, and the relative azimuth angle $\cos(\Delta\phi)$) and b) the top of canopy reflectances in the different wavebands of the satellite sensor. The Levenberg-Marquadt optimization algorithm is used in the training process to obtain the best agreement between the output simulated by the network and the correspondent value of the training database.

SPOT satellites have two kinds of sensor instruments on board (for example, HRG1 and HRG2 for SPOT5). The analysis of their spectral responses showed that differences between instruments on board of a same satellite are not significant (as can be seen in Figure 3.3, for the red and near-infrared bands). For the sake of simplicity and for computing-time gain, paired-sensors were considered as one (*e.g.* SPOT4HRV1 and SPOT4HRV2 considered as SPOT4). Thus, a different network is generated and trained for each one of the sensors, counting a total of four networks (one for Formosat-2, one for SPOT2, one for SPOT4 and one for SPOT5) for each of the output variables (GAI, Fcover and Fapar).

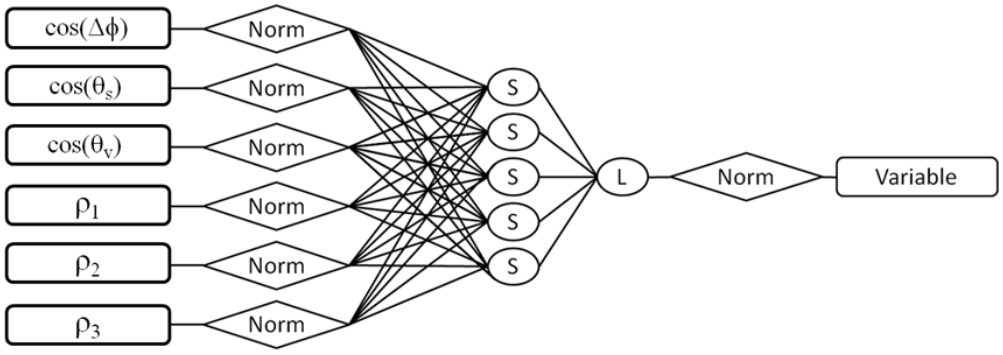


Figure 3.2- Neural network architecture developed for the estimation of the biophysical variables considered from the Formosat-2 or SPOT bands (ρ_n) and the three angles defining the geometry of observation. The network is composed of 2 hidden layers: 1 of 5 neurons S (Sigmoid) and 1 L (linear) output neuron. The inputs and outputs are normalized (adapted from Weiss and Baret (2010a)).

3.2.2.3. Applying the neural networks

After the training step, a relationship is established between inputs (reflectances and observation geometry) and outputs (biophysical variables) of the network. The neural networks are then applied to the Formosat-2 and SPOT images. This step can be defined as the operational use of the neural network. Output products are the

images of the estimated biophysical variables (GAI, Fapar and Fcover), having the same resolution as the input satellite images. For this thesis, we are interested in the GAI variable, given it will be further used for our crop modeling approaches.

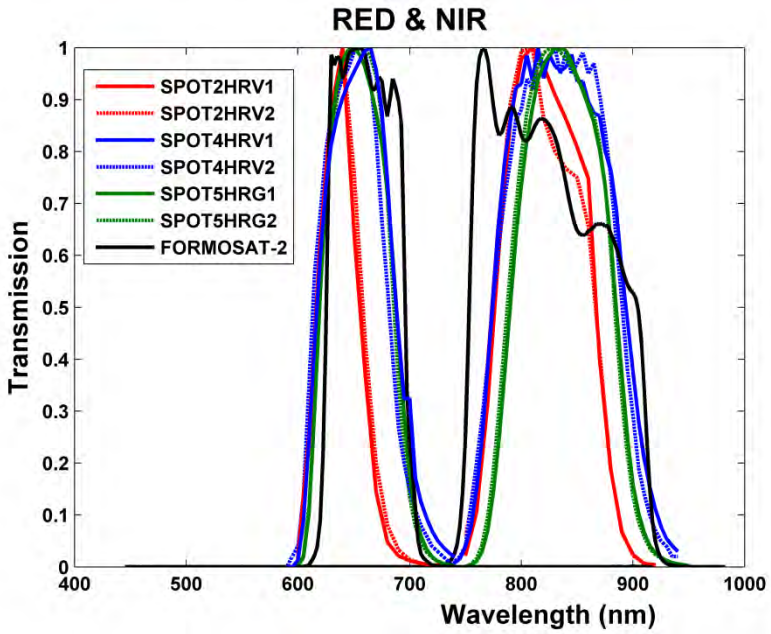


Figure 3.3- Spectral sensitivities for the red and infrared spectral bands for the instruments SPOT2HRV1, SPOT2HRV2, SPOT4HRV1, SPOT4HRV2, SPOT5HRG1, SPOT5HRG2 and Formosat-2.

3.2.3. Data pre-processing: interpolation in time

Despite the availability of an extensive temporal series of Formosat-2 and SPOT images, it is extremely difficult to have the *in-situ* measurements acquired concurrently with satellite observations. In order to validate the GAI estimations, in a point to point comparison, it was necessary to interpolate in time one of the time series. A temporal interpolation was then applied to the ground measurements, so they could match with the dates of remote sensing acquisitions. We applied a simple linear interpolation to the *in-situ* observations. Only the interpolated values for dates comprised within a window of 8 days (+/-4 days) around the *in-situ* measurement date were selected for our study.

3.3. Results: Validation and Inter-comparison

3.3.1. Evaluation of performances

In order to evaluate the performances of the studied methods, the deviation between estimated variables ($\hat{y}$) and ground measurements (y) are quantified by the statistical indicators listed in Table 3.3.

Table 3.3- Description of statistics indicators for evaluating the methods performances

Statistic	Formula
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$
Relative Root Mean Square Error	$RRMSE = 100 \times \frac{RMSE}{\tilde{y}}$
Bias	$B = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i - y_i$
Correlation coefficient	$R = \frac{\sum_{i=1}^{N}(y_i - \tilde{y}) \times (\hat{y}_i - \tilde{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \tilde{y})^2 \times (\hat{y}_i - \tilde{\hat{y}})^2}}$

N : number of observations ; $\tilde{y}$: mean value of ground measurements;
$\tilde{\hat{y}}$: mean value of estimated variables.

3.3.2. Comparison of Formosat-2 and Spot data

Figure 3.4 shows the scatter-plot of the Formosat-2 and SPOT NDVI products over 3957 crop fields (wheat, rapeseed, sunflower, maize and soybean), for 11 different dates. A consistent linear relationship is observed. Some points appear more dispersed, but as it can be seen on the related density plot in Figure 3.4, they represent a quite insignificant percentage of the dataset. The comparison of F2 and SPOT vegetation indices presented a strong correlation and low errors (R^2=0.98, RMSE = 0.034 and RRMSE = 6.76%). This is probably related to the accuracy of the images pre-processing (described in section 2.2.3). Besides, the results suggest there is no particular crop effect on the relationship. Slope *a* and interception *b* of the linear relationship found are 1.001 and -0.018, respectively. These coefficients can be applied to adjust the NDVI of SPOT to match the standard NDVI of the F2 satellite (or vice versa), aiming to obtain a relatively uniform NDVI dataset.

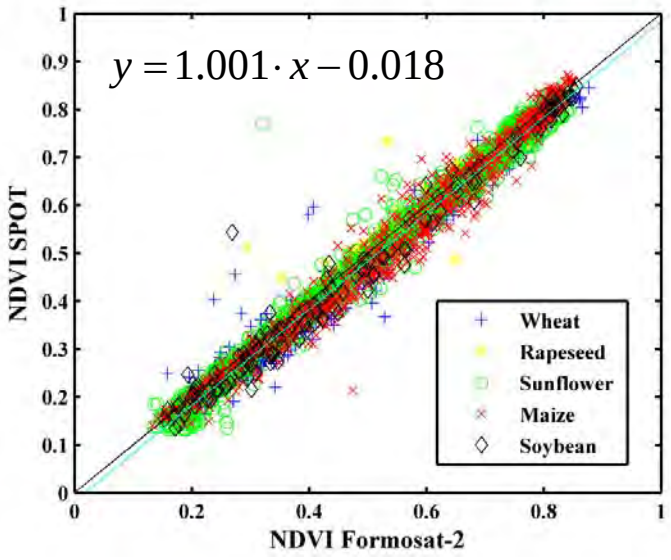


Figure 3.4- Comparison of F_2 and SPOT NDVI products over 11 different paired dates and 3957 plots cultivated with wheat, rapeseed, maize, sunflower and soybean crops

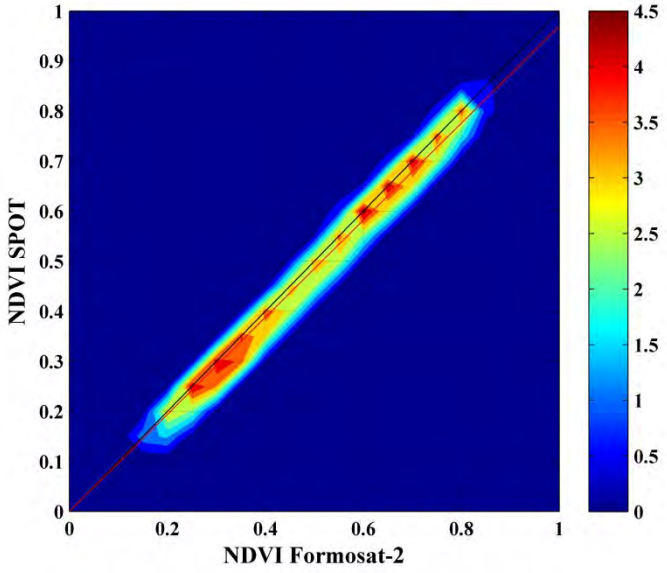


Figure 3.5- Density plot of NDVI Formosat-2 *vs* NDVI SPOT (same data as in Figure 3.4).

These results revealed that despite the differences in geometrical (solar and view angles) configuration of the acquisitions, NDVI from F2 and SPOT sensors are quite similar. Figure 3.6 shows the relationship between NDVI residual errors and view angles differences. No significant correlation (R^2=0.0136) was found, suggesting a poor effect of observation geometries on these NDVI products.

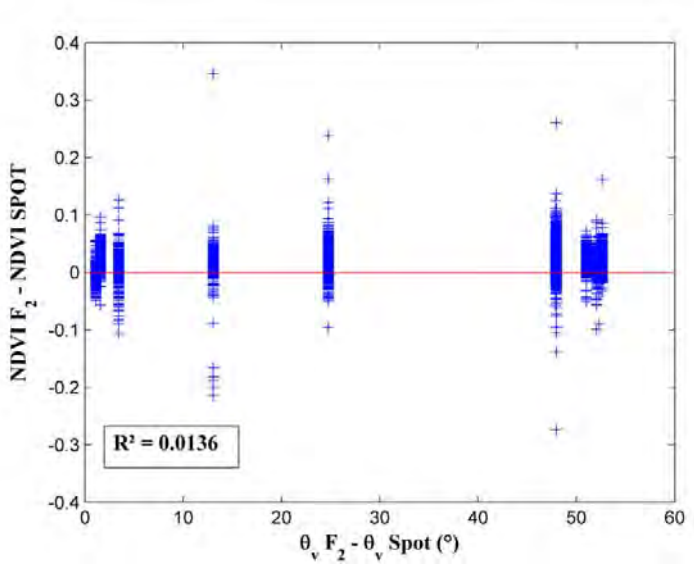


Figure 3.6- NDVI residues plotted against difference between Formosat-2 and SPOT view angles (θ_v).

3.3.3. Empirical method

The effective GAI data, obtained from the *in-situ* (indirect) measurements performed over several plots on the study area (detailed in Chapter 2), allowed establishing a NDVI-GAI exponential relationship for Formosat-2 and corrected SPOT NDVI (Figure 3.7). Only the remote sensing images acquired concurrently (± 4 days, *c.f.* interpolation method in section 3.2.3) to ground measurements (hemispherical photographs) were used for establishing the empirical relationship (dates different from those cited on section 3.2.1.1). The retrieved exponential law is written in equation (3.3).

A single exponential law was used for all crop types. This relationship covers a considerable range of NDVI values (between 0.26-0.9). A total of 89 points (pairs NDVI/effective GAI) were used to establish the relationship, all crop types included.

$$GAI = 0.16 \cdot e^{3.61 \cdot NDVI} - 0.26 \quad (3.3)$$

The empirical GAI was then calculated by applying the established equation to each NDVI point. Figure 3.8 shows the scatter-plot of the empirical estimated GAI and the effective GAI (GAI_{CanEye}), obtained from hemispherical photographs. A strong correlation is observed (R^2=0.87) and low errors (RMSE=0.39 and RRMSE=28.1%) as well. We observe a slightly stronger dispersion for the maize and sunflower crops and some bias for GAI values greater than 3, for soybean and maize. The dispersion observed for maize and sunflower, especially for the higher values of GAI, might be

due to a 'row effect', since the rows are quite pronounced when the vegetation is well developed (which is not the case for wheat or soybean). This induces more uncertainty over the GAI estimation using CAN-EYE because we deviate from the hypothesis of a turbid medium assumed by this approach.

In order to analyze the impact of the robustness of the NDVI-GAI law parameterization, we performed some tests. The data set was divided into a calibration group and a validation group. The number of observation points used for calibration varies between 6 and 44 points to establish the NDVI-GAI law. The rest of the points (89-n) was used for validation. A total of 10 000 combinations was tested for each dataset. We calculated then the relative root mean square error (RRMSE) between GAI_{CanEye} and GAI_{NDVI}, for the group of validation points. Afterward the mean and standard deviation (std) RRMSE over the 10 000 tests were also computed for each option. Figure 3.9 shows the mean (left), and standard deviation (right) of the RRMSE according to the number of observation points used for calibration. We observe the mean and std RRMSE reduce exponentially with the increase of amount of calibration points. Figure 3.10 illustrates the histogram of the RRMSE when 44 points (approximately 50% of available points) are used for calibration and the other 45 points used for validation. It is interesting to note that the mean RRMSE (29.24%) is quite close to the RRMSE obtained when using the all 89 points for calibration (28.1%), and also that the associated standard deviation (3.1%) is small. The analysis of these graphics suggests that 30 points (around 30% of the data set) would be sufficient for establishing a robust empirical relationship, with a mean and std RRMSE of 29.74% and 2.44%, respectively. As expected, this approach yields overall good results.

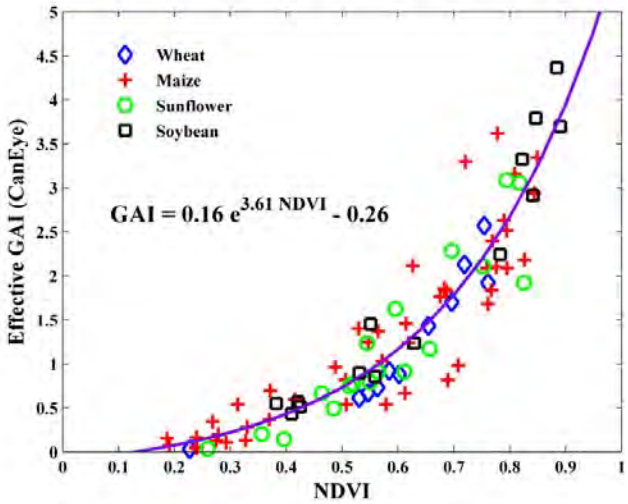


Figure 3.7- NDVI-GAI exponential relationship for SPOT and F2 data. The purple line represents the exponential law, characterized by the parameters α=0.16 and β=3.61. All crops (wheat, maize, sunflower and soybean) were used on the calibration process.

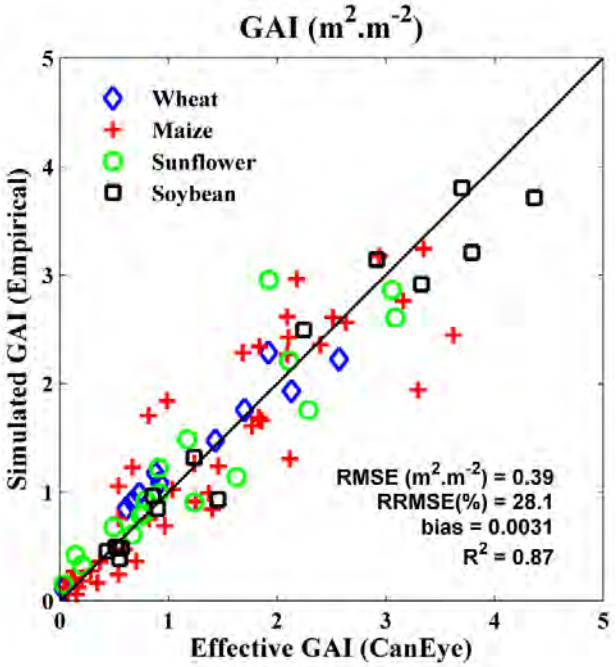


Figure 3.8- Validation of results obtained using the empirical relationship for wheat, maize, sunflower and soybean crops. Measured effective GAI plotted against GAI estimated trough the empirical approach. The performances (RMSE, RRMSE, bias and R^2) are indicated.

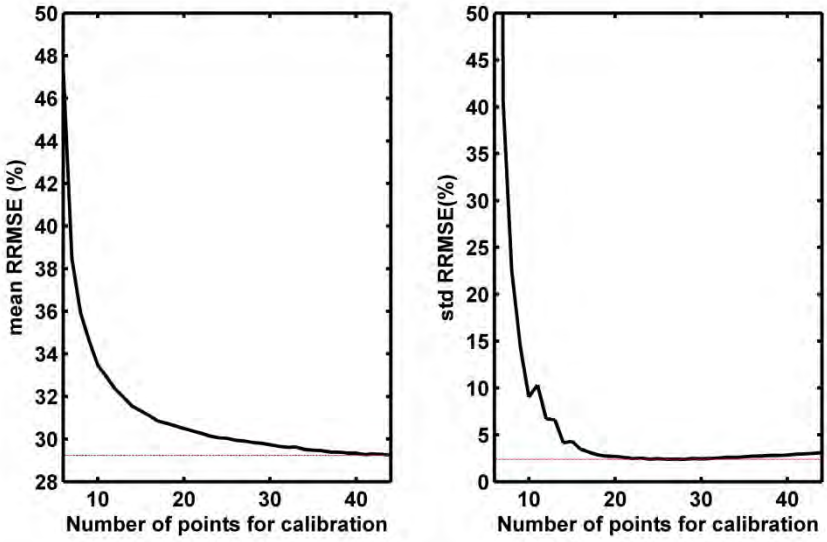


Figure 3.9- Analysis of number of points needed for calibrating the exponential relationship NDVI-GAI. The number of observation points used for calibration varies, going from 6 up to 44 points (abscissa). On the y-axis the mean (on left) and standard deviation (right) of the RRMSE are shown according to the number of observation points used for calibration. A total of 89 observation points (calibration + validation) was used.

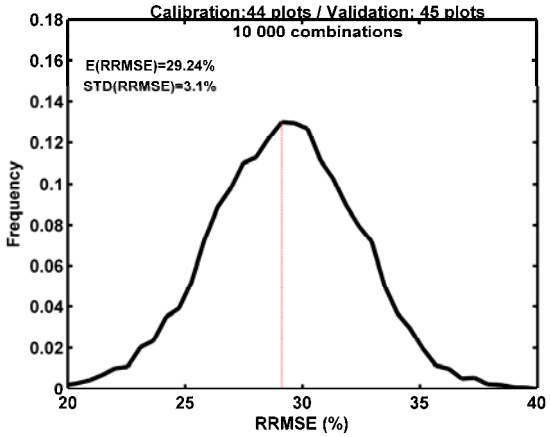


Figure 3.10- Histogram of the RRMSE obtained when 44 points are used for calibration and the other 45 points left are used for validation. RRMSE calculated for the validation datasets. A total of 10 000 combinations was tested (permuting the points used for calibration and validation).

3.3.4. Physical method

The GAI values estimated by the Neural Networks approach were compared with the *in-situ* measurements. Figure 3.11 shows the GAI retrieved by this approach against the effective GAI estimated by CAN-EYE. The performances in terms of correlation, bias, absolute and relative root mean square errors are indicated. We observe a strong correlation (R^2=0.86) and an absolute root mean square error (RMSE) of 0.51 $m^2.m^{-2}$ and relative RMSE of 31.2%. We observe that larger values of GAI (superior than 2 $m^2.m^{-2}$) are overestimated by this method, as indicates the bias of -0.15 $m^2.m^{-2}$. However, different trends for the investigated crops can be distinguished. Note that for wheat and maize (*blue* and *red symbols*) the simulations agree very well with observations. For sunflower and soybean (*green and black symbols)*, results suggest the simulations are biased, especially for greater values of GAI. The hemispherical photographs are supposed to be taken at different heights during all the season, in order to have the minimum distance of 1m between the top of the canopy and the camera (Demarez et al., 2008). However, when the vegetation is well developed the distance between the camera and the top of the canopy may be shorter (than in the beginning of the season). It probably generates a saturation effect, resulting in lower estimations of effective GAI. As a result, there would be an underestimation of the GAI estimated with CAN-EYE, and not an overestimation of NNT estimates.

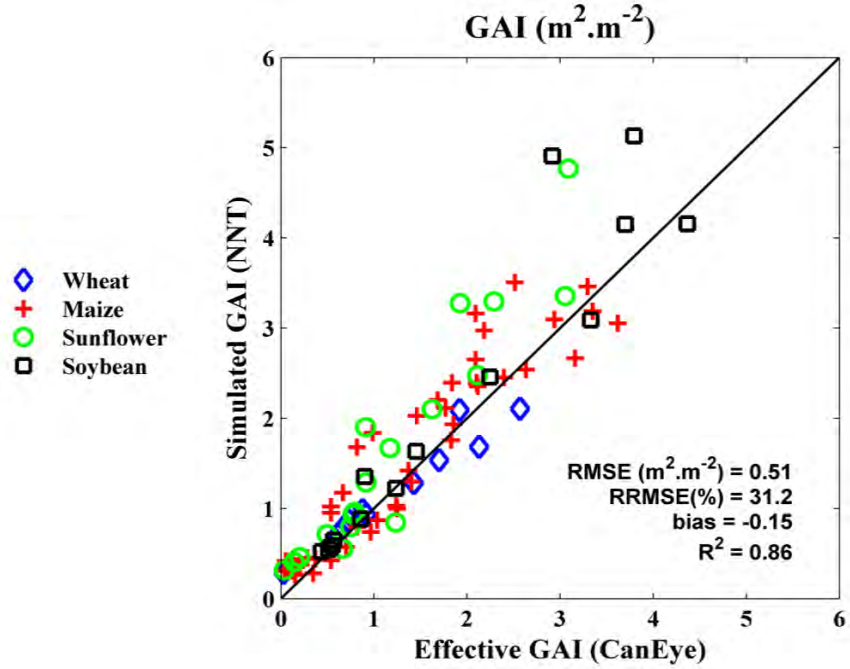


Figure 3.11- Validation of results obtained using the physically-based approach (NNT) for wheat, maize, sunflower and soybean crops. Measured effective GAI plotted against GAI estimated trough the physical approach. The performances (RMSE, RRMSE, bias and R^2) are indicated.

3.3.4.1. Influence of ALA parameter over GAI estimation

Based upon the previous results, we considered that it could be helpful to study the influence of the input parameter average leaf angle (ALA, Table 3.2) over the GAI estimation using BV-NET. Indeed, this could maybe be a cause of the likely GAI overestimation for sunflower and soybean crops by BV-NET.

Initially, we defined the same ALA configuration for all the crops (the same as showed earlier), so the ALA parameter distribution was set to the values showed in Table 3.2. Next we performed a study of sensibility for testing different configurations (narrower intervals) for sunflower and soybean crops. In order to establish optimal ranges of variation of the ALA parameter intervals, we considered, at first, the standard ALA configuration ("Config1", from Claverie (2012)) and then three other configurations ("Config2", "Config3", "Config4"), centered around low ALA values, given sunflower and soybean are rather planophile crops. The tested distributions are exhibited in Table 3.4 and the performances (RMSE, RRMSE, bias, R^2) for each configuration are presented in Table 3.5.

Table 3.4- List of the tested distributions of the ALA parameter as input for the BV-NET tool, for sunflower and soybean crops.

	Sunflower				Soybean			
ALA (°)	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**
Config1	5	80	40	20	5	80	40	20
Config2	10	50	30	20	10	60	30	15
Config3	25	35	30	5	10	20	15	5
Config4	15	25	20	5	15	25	20	5

Table 3.5- Performances of the different ALA distributions over GAI estimation with BV-NET tool. The statistics (RMSE, RRMSE. Bias, R^2) are calculated from the *in-situ* GAI measurements and the GAI simulated by the neural network method. Lowest errors are displayed in green and highest errors, in red.

	Sunflower				Soybean			
	RMSE	**RRMSE**	**Bias**	**R^2**	**RMSE**	**RRMSE**	**Bias**	**R^2**
Config1	**1.12**	**65.99**	0.57	0.88	**1.12**	**65.86**	0.41	0.96
Config2	0.93	55.20	0.40	0.82	1.13	66.39	0.37	0.96
Config3	0.98	57.96	0.42	0.82	**1.02**	**60.27**	0.23	0.95
Config4	**0.79**	**46.79**	0.27	0.83	1.08	63.54	0.29	0.96

By analyzing the obtained results, we observe that the change in the ALA parameters yields to better performances in GAI estimation with the BV-NET tool. For sunflower the "Config4" appeared to be the best one, resulting in a significant improvement in GAI estimation (from initial RRMSE 65% to 46%), especially for lower GAI values; we still note an overestimation for the greater GAI values. For soybean, even for the best configuration ("Config3"), we have only a slight improvement (5%), nevertheless the RMSE remains strong and overestimation of greater GAI values is observed as well. Results suggest the ALA parameter is probably not associated with the supposed overestimation of GAI by BV-NET. This question remains nevertheless open. Overestimation of BV-NET GAI or underestimation of our ground truth (GAI estimated with CAN-EYE)? More researches should be addressed on this subject; still the obtained results are satisfactory and appropriate for our purposes, since the use

of GAI maps for running the SAFY crop model (Chapter 4) mostly concerns winter wheat in this study.

The performances obtained with the physical approach for our study area (displayed in Figure 3.11) can be compared to other previous studies. Some authors employed the same approach, using the BV-NET tool, but with some differences in the model parameterization. Claverie (2012) found a lower RRMSE (21.15%), for summer crops using only Formosat-2 images, probably because Formosat-2 images have constant viewing angle (and SPOT do not). Bsaibes et al., (2009) obtained a RRMSE of 34.63% for wheat, meadow, maize and rice crops using Formosat-2 images as well. Duveiller et al., (2011) found a RRMSE around 25% for wheat applying a linearization operation coupled to the NNT tool and using SPOT images. Therefore we note that our results are in agreement with those obtained in similar studies.

3.3.4.2. Comparison with ground destructive measurements

The GAI simulated by BV-NET tool were also compared with ground destructive measurements (Figure 3.12). It is important to note that the points of this graphic are not the same than for the previously cases, since the destructive measurements were not necessarily performed on the same day as the hemispherical photographs (effective GAI). We observe that NNT retrievals are underestimated regarding the ground data. This was expected since remotely sensed products do not take into account the aggregation of the leaves (known as clumping effect), and since the ground destructive samplings consider all the green organs of collected plants.

We chose the effective GAI data for validating the BV-NET GAI estimates, because the remote sensed data are better correlated with it (then with the destructive GAI). Besides, both the SAIL radiative transfer model (part of the PROSAIL model, used for estimating GAI with neural networks) and the CAN-EYE algorithm (used for estimating the GAI effective from hemispherical photographs) are based upon the Poisson's law [http://www4.paca.inra.fr/can-eye].

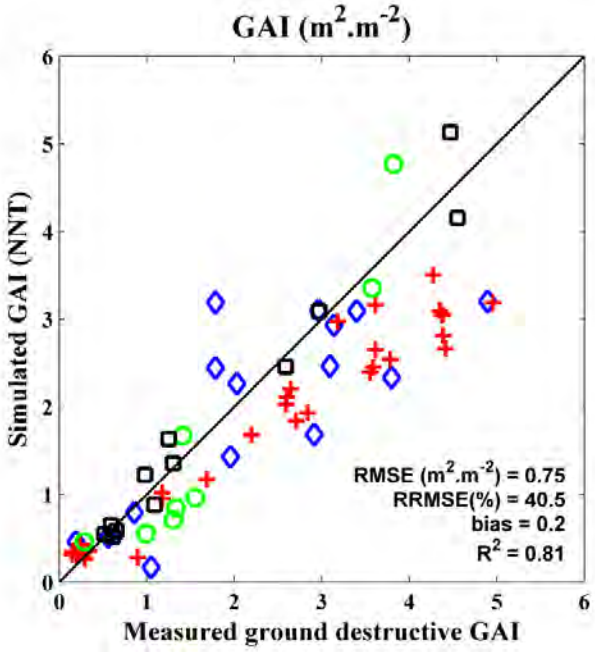


Figure 3.12- Comparison of GAI simulated through the NNT method against ground destructive GAI measurements.

3.3.5. Inter-comparison of both approaches

The empirical approach usually requires a group of samples for calibration and another one for validation, while the physical one needs only the validation sample. Given that we have a total of 89 available points, 30 points were assigned for calibrating the empirical law, while the left points (59) were used for validation of both methods. As described previously, the calibration and validation points were chosen randomly, and 10 000 combinations were tested. For each of these combinations, and each of the approaches, we calculated the RRMSE between estimated GAI and GAI *in-situ* data using the 59 validation points. It results in 10 000 RRMSE values for each method. The mean RRMSE and associated standard-deviation (std) were then calculated in order to compare the two investigated approaches. Figure 3.13 exhibits the histograms with the distributions of the 10 000 RRMSE values for the empirical (left) and physical (right) approaches. The mean RRMSE and std are displayed.

The performances indicate empirical and physical approaches yield to very similar results, when using 30 points for calibrating the empirical law, and the 59 left points for validating both approaches. We observe the NNT-based approach presents practically the same results as the empirical one ($RRMSE_{NNT}$ =29.04% and std_{NNT}=2.16% against $RRMSE_{emp=}$29.74% and std_{emp}=2.48%), which is quite satisfying. When all the observations are considered (the same 89 points are used for

calibration and validation), the empirical approach slightly overcomes the physical one (compare Figure 3.8 and Figure 3.11). Still, performances are quite similar and the physical method should be privileged since it does not require specific calibration.

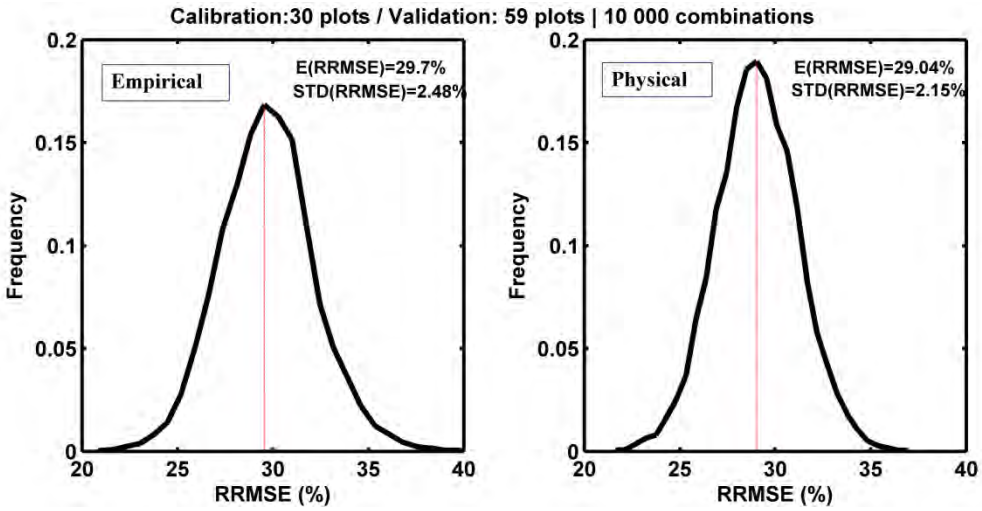


Figure 3.13- Histograms of the RRMSE obtained when 30 points are used for the calibration of the empirical relationship and the other 59 points are used for validation. A total of 10 000 combinations was tested (permuting the points used for calibration and validation). Results for empirical approach on the left, and for the physical approach on the right.

3.4. Conclusions

In this chapter, NDVI derived from F2 and Spot satellites were compared. The results revealed a strong correlation between them and low influence of observation geometries on NDVI products. Thus with appropriate atmospheric and geometrical corrections, linear intercalibration is valuable and vegetation indices issued from the two sensors could be combined, increasing the opportunity of having cloud-free acquisitions for continuous crop monitoring. This homogeneous NDVI data set was then used to establish the empirical relationship between NDVI and GAI.

When comparing the investigated methods for GAI retrieval, it is noticeable that the physical and empirical approaches yield to similar performances (R^2=0.86 for both and $RMSE_{emp}$= 0.41 vs $RMSE_{NNT}$= 0.40 $m^2.m^{-2}$, $RRMSE_{emp}$=29.7% vs $RRMSE_{NNT}$=29.04%). Nevertheless, the empirical approach requires a large amount of field data, demanding time-consuming measurements, which is a common constraint. Moreover, as the relationship NDVI-GAI is site-dependent, it would have to be calibrated when applied to landscapes that have different experimental conditions than for the one the law was established. The neural network method

leads to good performances as well, and presents the advantage of not being dependent of *in-situ* measurements or site-specific calibration.

The empirical relationship NDVI-GAI is known for saturating for strong values of NDVI, but as the empirical estimates were compared against *in-situ* GAI data estimated from hemispherical photographs using an exponential Poisson law (that also saturates), the saturation effect did not affected the performances. The physical approach, on the other hand, is less subjected to saturation issues. Still, it should be noted that physically-based GAI remain underestimated when compared to the destructive measurements, because of the clumping effect. Therefore, the physical method seems to be quite promising as it does not depend on any field measurements, and offers great perspectives to regional scale applications. This method will be then kept for the sequence of this thesis.

The availability of accurate, homogeneous and complete series of GAI estimates over the study area will allow the assimilation of these products into crop models, yielding better estimates of the crop–related variables: as biomass, yield and components of the carbon and water cycles.

4. Crop Models Description

This chapter is dedicated to the description of the crop models used and developed during the thesis for estimating the multiple variables that we are interested in. The first section will describe the original SAFY model (Section 4.1), which was applied for estimating biomass production and yield. On section 4.2, the SAFY-CO_2 model, which is an adaptation of the SAFY model, is described for estimating the components of the carbon budget, *i.e.* photosynthesis (GPP), NPP, yield, NEE and ecosystem respiration. On section 4.3 the SAFYE-CO_2 model is presented. SAFYE-CO_2 is a model resulting from the coupling of SAFY-CO_2 model with a water module, based on the FAO-56 method. It is able to simulate the same outcomes as the SAFY-CO_2 model (C fluxes) plus the components of the crop water cycle (like transpiration, evaporation and soil water content). Figure 4.1 presents a scheme illustrating the three investigated models and their respective outputs.

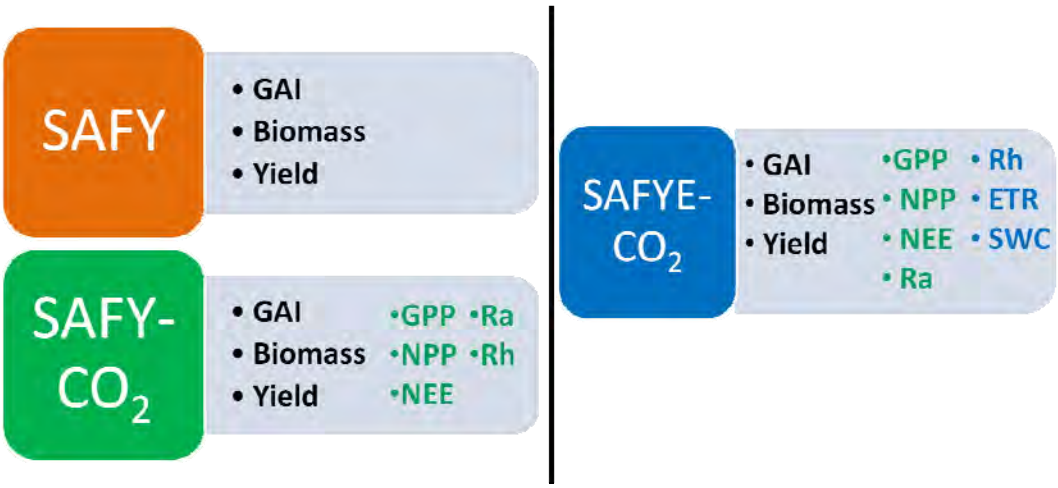


Figure 4.1- Scheme of the crop models presented in this chapter with their associated main estimated variables.

4.1. The SAFY model

4.1.1. Model description

The SAFY model (Simple Algorithm For Yield estimates, Duchemin et al., (2008)) is a daily time step crop model that simulates time series of GAI, dry aboveground biomass (DAM) and grain yield. The approach for biomass estimation is based on the Monteith's light-use efficiency theory (Monteith and Moss, 1977), coupled with a temperature stress function. Monteith's theory links the production of total dry aboveground biomass with the photosynthetically active portion of the solar radiation (PAR) absorbed by the plants. The SAFY model is driven by daily incoming global radiation (Rg) and cumulative daily mean air temperature (Ta). The biomass production starts on the plant emergence day (D_0) and ends on the day of complete leaf senescence.

The daily biomass production is calculated by multiplying the photosynthetically active radiation absorbed by the canopy (APAR) by the effective light-use efficiency (ELUE), and is constrained by the temperature-stress-function F_T [Eq. (4.1)]. The APAR [Eq. (4.2)] is calculated using the Rg, the climatic efficiency (ε_c) and the fraction of photosynthetically active portion of solar radiation absorbed by the plants (FAPAR). The FAPAR is calculated from the effective green area index (GAI) and light-extinction coefficient (k_{ext}) according to Beer's law [Eq. (4.3)]. The increase in GAI is controlled by a total biomass to leaf partitioning function. It is important to observe that in SAFY, the Monteith's LUE (light-use-efficiency) constant is replaced by an effective LUE, which corresponds to the plant's potential in converting the APAR into dry aerial biomass (DAM), regardless the agronomical and/or environmental limitations, such as water and nitrogen availability.

The biomass production can be affected by extreme (too high or too low) temperatures. This effect is taken into account by the temperature-stress-function F_T, which is a β-degree polynomial function varying according to the air mean daily temperature (Ta). The F_T function is derived from the STICS model (Brisson et al., 2003) and it depends upon three parameters: the optimal temperature for crop functioning (T_{opt}) and two extreme values (T_{min} and T_{max}) beyond which the plant growth stops [Eq. (4.4)].

$$\Delta DAM = APAR \times ELUE \times F_T(Ta) \tag{4.1}$$

$$APAR = Rg \times \varepsilon_c \times FAPAR \tag{4.2}$$

$$FAPAR = 1 - e^{-k_{ext} \times GAI} \tag{4.3}$$

$$\begin{aligned} F_T(Ta) &= 1 - \left[\frac{T_{opt} - T_a}{T_{opt} - T_{\min}} \right]^{\beta} \quad if\ T_{\min} < T_a < T_{opt} \\ F_T(Ta) &= 1 - \left[\frac{T_{opt} - T_a}{T_{opt} - T_{\max}} \right]^{\beta} \quad if\ T_{\max} > T_a > T_{opt} \\ F_T(Ta) &= 0 \quad if\ T_a < T_{\min} or\ T_a > T_{\max} \end{aligned} \tag{4.4}$$

The plant emergence day (D_0) depends on agricultural practices, date and depth of sowing, and also on the soil characteristics and climatic conditions. This parameter determines the beginning of the biomass production and thus the GAI development. We observe that there is a difference between the “agronomical emergence”, when the first aerial plant organs appear above the soil, and the “SAFY emergence”, which corresponds to the moment when green organs are perceived by the satellite.

During the growth stage, the aerial biomass production is distributed into leaf and non-leaf biomass, according to the partition function P_L. The partition-to-leaf function P_L [Eq. (4.6)] is an empirical function of the sum of daily mean air temperature (SMT), accumulated since plant emergence, with two parameters (P_{La} and P_{Lb}), adapted from Maas (1993). It varies between 0 and 1. As the P_{La} parameter is close to 0, P_L exponentially decreases with thermal time from a value close to 1 at plant emergence and to a value of 0 at the end of the leaf production phase. The daily leaf biomass production is then converted into leaf area [Eq. (4.5)], according to the value of the specific leaf area (SLA) parameter.

The senescence of leaves starts when the accumulated sum of temperature (SMT) reaches a given threshold (STT, Sum of temperature for senescence). The leaf senescence (ΔGAI^-, Eq. (4.8)) increases with thermal time and it is function of the rate of senescence (Rs) parameter. The GAI is then updated from the difference between ΔGAI^+ (growing) and the ΔGAI^- (senescence), plus GAI of the previous day [Eq. (4.9)]. The senescence phase ends when the GAI reaches a value lower than the initial one, indicating the vegetative period is totally ended.

$$\Delta GAI^+ = \Delta DAM \times P_L \times SLA \tag{4.5}$$

$$P_L = 1 - P_{La} \times e^{P_{Lb} \times SMT} \tag{4.6}$$

$$SMT = \sum_{D_0}^{j} (Ta - T_{\min})\, dj \tag{4.7}$$

$$\Delta GAI^{-} = GAI \times \frac{SMT - STT}{Rs} \tag{4.8}$$

$$GAI_{j} = GAI_{j-1} + \Delta GAI^{+} - \Delta GAI^{-} \tag{4.9}$$

The main variables and parameters, and associated notations and units, of the SAFY model are summarized in Table 4.1.
PS: The equations for estimating grain yield will be described in paragraph 4.1.5.

4.1.2. Spatial units

The SAFY model can be applied to different spatial units, from the pixel scale, passing by the field scale and even for a group of consecutive fields, cultivated with the same crop. In order to study the impact of intra-field variability over the SAFY biomass estimations, different approaches for running the model were tested. The comparison of *pixel-based* and *aggregative-based* approaches revealed that taking into account the intra-field spatial variability has no significant effect on the whole field biomass estimates. Thus, the *aggregative* approach might be favored due to its better time-consuming performance if we are interested in the mean field biomass estimates. For more details, see the paper in Appendix B. The study was performed only for the sunflower, characterized as a heterogeneous crop. Thus, we consider that the results are also valid for the winter wheat, which normally is more homogeneous than sunflower. Hereafter the chosen spatial unit is the field scale, which contours are provided by the RPG (Chapter 2). However for the validation/calibration of the model for evapotranspiration and CO_2 fluxes estimates considering a *pixel-based* approach may be more appropriate. It will be discussed in section 5.7.

4.1.3. Meteorological forcing variables

The analysis of SAFY equations shows that the model is driven by two meteorological variables: the daily incoming radiation (Rg) and the daily mean air temperature (Ta). In order to run the model, the daily values for these variables

come from either the towers installed on the experimental sites (Chapter 2.3.1.3) or from the SAFRAN dataset, provided by Météo-France (Chapter 2.5.3).

4.1.4. Model's parameterization

According to Maas (1992), parameterization is a modeling technique using an empirical function to approximate the response of a physical system. This technique reduces the complexity of models, simplifies the input requirement of models, and makes them easier to use for operational purposes.

The SAFY model was initially designed for studying irrigated winter wheat crop in the semi-arid Marrakech plain (Duchemin et al., 2008). The model was later adapted by Claverie et al., (2012) for the summer crops over a temperate climate area. Having the cited works as base, our objective is to parameterize the model for winter wheat over our study area (a temperate site). We aim at developing a robust and general calibration method that could be easily transposed to other regions.

The great advantage of the SAFY model is its low complexity level, which facilitates the optimization of unknown parameters using few observations (Duchemin et al., 2008). The parameters of the SAFY model (listed on Table 4.1) were divided into three main classes. The way those parameters are fixed or estimated is described in the following sections.

The first category includes the parameters that are fixed according to the values found in the literature. This is the case for the climatic efficiency (ε_C), the specific values of air temperature related to plant functioning (T_{min}, T_{max} and T_{opt}), the initial dry aboveground biomass value (DAM_0), the polynomial degree (β) of temperature-stress-function F_T, the sum of temperature for starting grains filling (STY), and the rate of grain filling (Py).

The second category includes the parameters estimated from *in-situ* measurements. These parameters depend on the genetic characteristics of the crop type. In this class, we have the light-interception coefficient (K_{ext}).

The third class includes the parameters that are calibrated by using the remotely sensed GAI time series.They allow fitting the SAFY simulated GAI courses to match with the neural networks GAI time series (from Chapter 3). They are: the specific leaf area (SLA), the two parameters of the partition-to-leave function (P_{La} and P_{Lb}), the two parameters of the senescence function (STT and Rs), the effective light-use efficiency (ELUE) and the plant emergence day (D_0). The calibration of these parameters is based on the minimization of the Root Mean Square Error (RMSE) between the remotely sensed GAI time series (derived from BV-NET) and those estimated by the SAFY model. A scheme of SAFY model process is shown in Figure 4.2. The optimization method is described in details on paragraph 4.1.4.3.

The following sections describe the approaches implemented for setting the parameters for each of the three categories.

Table 4.1- List of the main variables and parameters of the SAFY model: notation, units and values used for winter wheat crop.

	Description	Notation	Unit	Value	Method
Input Variables	Daily incoming global radiation	Rg	$MJ.m^{-2}.day^{-1}$		
	Daily mean air temperature	Ta	°C		
Parameters	Initial dry above-ground mass	DAM_0	$g.m^{-2}$	4.5	Literature
	Climatic efficiency	ε_c	-	0.48	Literature
	Light-interception coefficient	K_{ext}	-	0.76	*In-situ* measurem.
	Minimal temperature for growth	T_{min}	°C	0	Literature
	Optimal temperature for growth	T_{opt}	°C	20	Literature
	Maximal temperature for growth	T_{max}	°C	37	Literature
	Polynomial degree	β	-	2	Literature
	Rate of grain filling Sum of temperature for starting grain filling	Py STY	$°C^{-1}$ °C	0.0051 796	Literature
	Specific leaf area	SLA	$m^2.g^{-1}$	[0.0154-0.286]	Calibration
	Partition-to-leaf function: parameter a	P_{La}	-	[0.01-0.5]	Calibration
	Partition-to-leaf function: parameter b	P_{Lb}	-	[0.0001-0.02].	Calibration
	Sum of temperature for senescence	STT	°C	[200-2000]	Calibration
	Rate of senescence	Rs	$°C.day^{-1}$	$[10^3$-$2x10^4]$	Calibration
	Day of plant emergence	D_0	day		Calibration
	Effective light-use efficiency	ELUE	$g.MJ^{-1}$	[1.4-2.6]	Calibration
Output Variables	Green area index	GAI	$m^2.m^{-2}$		
	Dry above-ground biomass	DAM	$g.m^{-2}$		
	Grain Yield	GY	$q.ha^{-1}$		

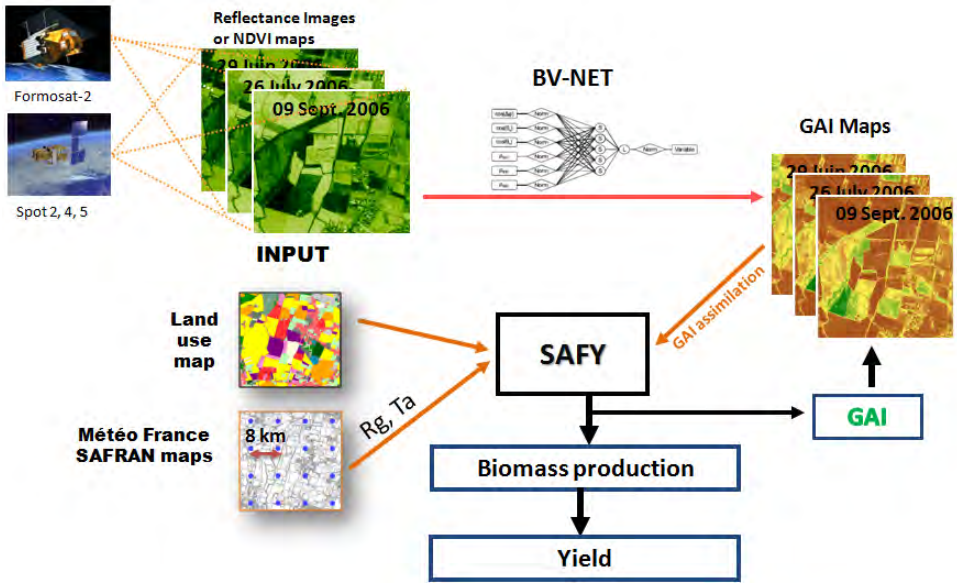


Figure 4.2- SAFY model diagram illustrating the main inputs of the model and the assimilation of series remotely sensed GAI maps for calibrating the model parameters for estimating biomass and yield products.

4.1.4.1. Parameters from the literature

The climatic efficiency (ε_C) corresponds to the photosynthetic radiation fraction of the global radiation. Many studies and our *in-situ* data (not shown) have shown that incoming PAR is roughly half of the incoming global radiation (Rg), independently of atmospheric conditions (Duchemin et al., 2008). Therefore, the climatic efficiency is practically constant in space and time, and its value is fixed to 0.48 (Varlet-Grancher et al., 1982).

The values of air temperature related to plant functioning (T_{min}, T_{max} and T_{opt}) were settled according to the standard parameters of the STICS model [*http://www.avignon.inra.fr/agrocilm_stics/*]. Wheat is generally considered to have an optimum temperature range of 17-23 °C over the entire growing season, with extreme temperatures of 0 °C and 37 °C beyond which growth stops (Porter and Gawith, 1999). Thus the minimal, optimal and maximal temperatures for wheat growth have been set up to 0, 20 and 37 °C, respectively.

The sum of temperature for starting grain filling (STY) was also determined according to standard parameters set for the STICS model. STY parameter is crop and variety-dependent. Nevertheless, this parameter was not available for the varieties cultivated on our experimental sites and for other plots in the study area this information (crop variety) is usually unknown.

We tested then setting different values for STY. From the STICS database for the common and durum wheat varieties, the STY can vary from 657 °C to 960 °C. First, we set STY to the minimum STY value of the database (657 °C); second to the maximum one (960 °C); and third, we set STY equal to the weighted average of the STY from common and durum varieties present in the database (STY=796 °C). The weights were settled according to the percentage of surface cultivated with each of the wheat types in the Haute-Garonne department (from the Agreste statistics, Chapter 2).

The polynomial degree (β) of temperature-stress-function F_T was set after Duchemin et al., (2008) and Claverie, (2012). The initial dry aboveground biomass value (DAM_0) was set to correspond to a GAI of approximately 0.1 $m^2.m^{-2}$, as showed Duchemin et al., (2008).

The rate of grain filling (Py) parameter determines the partition of the daily dry above-ground biomass converted into grains. A Py value equal to 0.0051 $°C^{-1}$ was fixed by Duchemin et al., (2008), using yield data from 9 wheat fields.

4.1.4.2. Parameters from *in-situ* measurements

The light-extinction coefficient (K_{ext}) was computed by inverting Beer's law [Eq. (4.3)] using the fraction of absorbed photosynthetically active radiation (FAPAR) and the effective green area index (GAIeff), both obtained from hemispherical photographs taken over wheat crops (at different growth stages) and processed using the CAN-EYE software. Figure 4.3 illustrates the exponential relationship between the two measured variables, FAPAR and the effective GAI. A value of K_{ext}=0.76 was found by minimizing the root mean square error between the established law and the measured data. This value is higher than the 0.5 used by Duchemin et al., (2008) for the wheat in semi-arid Morocco. Our value nevertheless is in the range 0.5-0.8, found by Bhattacharya et al., (2011) for the wheat in India, and also it is close to the value found by O'Connell et al., (2004) (K_{ext}=0.82 ±0.05) for the wheat in Australia.

4.1.4.3. Calibrated parameters

Calibration is considered as the procedure applied for identifying a single optimum parameter set resulting in a simulation that best reproduces observed variables. The calibration of models is a complex issue since the parameters are often inter-dependant (Duchemin et al., 2008). The SAFY parameters can compensate each other. The parameters compensation leads to different sets of parameter values that can yield similar simulations of a particular variable (GAI, for example), while other simulated variables present values completely distinct (*e.g.*

biomass). Thus it is crucial to identify the parameters having a biophysical interpretation, and to constrain them based on experimental results.

This category of parameters, including the ones to be calibrated, includes: the specific leaf area (SLA), partition-to-leave function parameters (P_{La} and P_{Lb}), the senescence parameters (STT and Rs), the effective light-use efficiency (ELUE) and the plant emergence day (D_0).

Different optimization algorithms can be applied for solving calibration problems. Hereafter we describe the optimization approach adopted for calibrating SAFY parameters. We also provide a succinct description of others well-known optimization methods that were tested for calibrating SAFY.

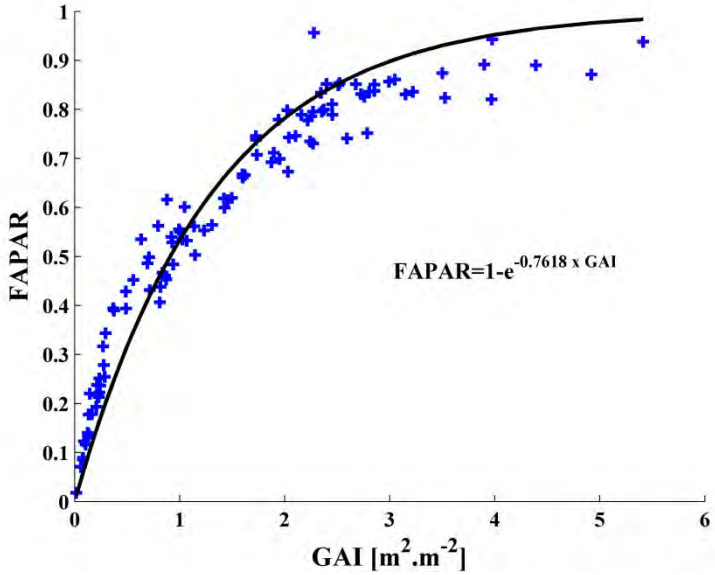


Figure 4.3- Relationship between FAPAR and GAI obtained from hemispherical photographs taken at different growth stages during wheat crops season and processed using the CAN-EYE software. The value of the k_{ext} coefficient is extracted from the exponential law established between the two variables.

4.1.4.3.1. Optimization methods

The SAFY traditional calibration method is based on the minimization of the Root Mean Square Error (RMSE) between the remotely sensed GAI time series and those estimated by the SAFY model. As it is shown in Figure 4.4, through the optimization process we try to find the best set of parameters that will generate a GAI profile that best matches with the observed GAI.

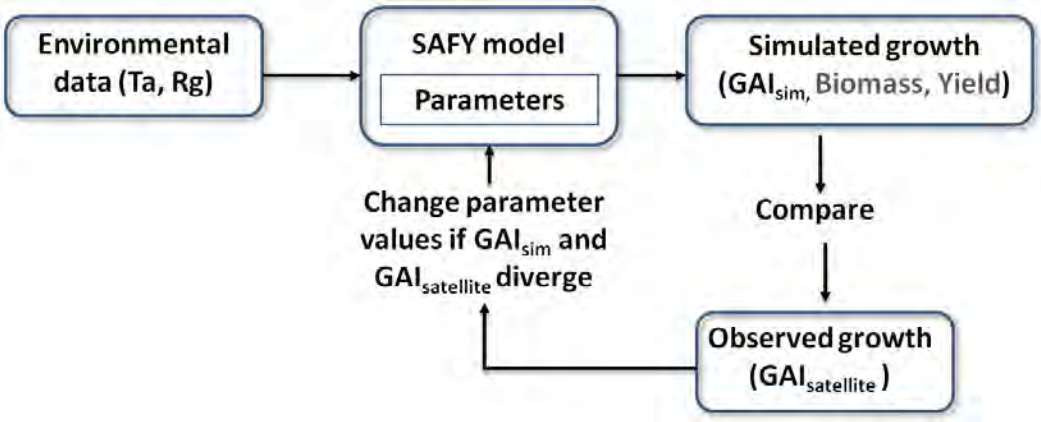


Figure 4.4- Diagram of the SAFY parameters calibration method. The simulated GAI profile is compared to the observed one in order to find the set of parameters that minimizes the divergence between them.

a) The modified simplex

A modified version of the Nelder-Mead simplex method (Lagarias et al., 1998) was initially chosen by Claverie (2012) for calibrating the model (parameters D_0, ELUE, P_{La}, $P_{Lb,}$, STT and Rs).

The simplex method is a very popular direct search method for multidimensional unconstrained minimization. The method attempts to minimize a scalar-valued nonlinear function of *n* real variables using only function values, without any derivative information.

A simplex is a geometric figure in *n* dimensions that is the convex hull of *n+1* vertices. The method generates a sequence of simplexes to approximate an optimal point of the function *f* (cost function). The initial working simplex *S* must not have its vertices in the same hyperplane. The method then performs a sequence of transformations of the working simplex *S*, aiming at decreasing the function *f* values at its vertices. At each step, the transformation is determined by computing one or more test points, together with their function values, and by comparison of these function values with those of the vertices. This process is terminated when the working simplex *S* becomes sufficiently small.

Our modified simplex approach allowed constraining the range of the parameters values to find solutions within realistic intervals. This adapted version runs the optimization process 30 times, from different starting points for each of the parameters, and then retains the set of parameters that gives the best solution (smaller RMSE). This technique intends to avoid finding a minimum local solution. For the sequence, this method is also referred as “fminsearch adapt”, since in Matlab, it is based on the *fminsearch* function, but with some adaptations we integrated.

b) Other optimization methods

We tested a few other optimization methods in order to compare the performances of our method (modified simplex) against other approaches and thus to confirm our choice for the best optimization method. A brief description of the tested methods is given below.

1) ***Fminsearch* (fmin):** It consists in the standard simplex, available in Matlab software. The intervals for parameters optimization are not limited. For low-dimensional problems, *fminsearch* is simple to use. For this method the user specifies a starting vector x_0 rather than a starting interval.

2) ***Dmin* (dmin):** This method is used on the soil moisture retrieval algorithm for the SMOS mission. It is based on the Levenberg-Marquardt (LM) method (Moré, 1978), which is a standard technique used to solve nonlinear least squares problems. The particularity of the *dmin* approach, compared to the LM, is to account for *a priori* knowledge on the parameters, considering the uncertainty associated to them. Therefore the cost function is calculated from the distance between retrieved and *a priori* parameters and the error between estimated and observed variable.

3) ***Non-Linear Approach (*nl*):*** This algorithm allows to solve nonlinear least-squares problems. It is based on the solver implemented on Matlab, the function *lsqnonlin* (more details in Coleman and Li,(1996)).

4) ***MultiStart (*ms*):*** MultiStart is a gradient-based solver characterized by its fast convergence to local optima for smooth problems. It is associated with the local solver *fmincon,* but offers other options as *lsqcurvefit*. MultiStart uses uniformly distributed start points within bounds, or user-supplied start points. Besides, it runs from all start points. MultiStart can run in parallel, distributing start points to multiple processors for local solutions (Ugray et al., 2007).

5) ***GlobalSearch* (gs):** GlobalSearch is a method similar to MultiStart. The main difference is that GlobalSearch uses a scatter-search mechanism for generating start points and then analyzes start points and rejects those that are unlikely to improve the best local minimum found so far. GlobalSearch is most focused on finding a global solution, and has an efficient local solver (*fmincon*) (Global Optimization Toolbox User's Guide R2013b).

6) ***Pattern Search (*ps*):*** This method provides convergence to local optimum, but slower than the gradient-based solvers. It is characterized by deterministic iterates. The start points are provided by the user.

Except for the *dmin*, *nl* and *fminsearch_adapt*, the cost function *f*, which is minimized through the optimization process, is defined as the quadratic difference between the simulated and observed GAI series, divided by the square of the standard deviation of the observations. For *dmin* and *nl*, the cost function is based on the error between estimated and observed GAI but also on distance between retrieved and *a priori* parameters. For *fminsearch_adapt*, the cost function is defined by the rmse between simulated and observed GAI profiles. For all the described methods only the GAI variable is used for calculating the cost function.

c) Comparison of optimization methods

The seven above-cited methods were compared through their performances obtained when optimizing the SAFY parameters over 22 wheat fields, for which biomass measurements were available. We used the relative root mean square error (rrmse) as performance indicator. The variables GAI and biomass estimated by SAFY (for each plot, for each optimization method) were compared against satellite and destructive data, respectively.

Figure 4.5 contains box-plot graphics, calculated using the results for all the 22 studied fields. The red lines indicate the median rrmse and the red plus, the outliers. The spacing between the different parts of the box helps to indicate the degree of dispersion in the results. The first column represents the results when the model is executed with the initial parameters values, with no optimization. Then the following columns show the results for the listed methods (1-7). Our primary method (modified simplex) is the so-called “fmin adapt”, and it is represented on the last column of the box-plots.

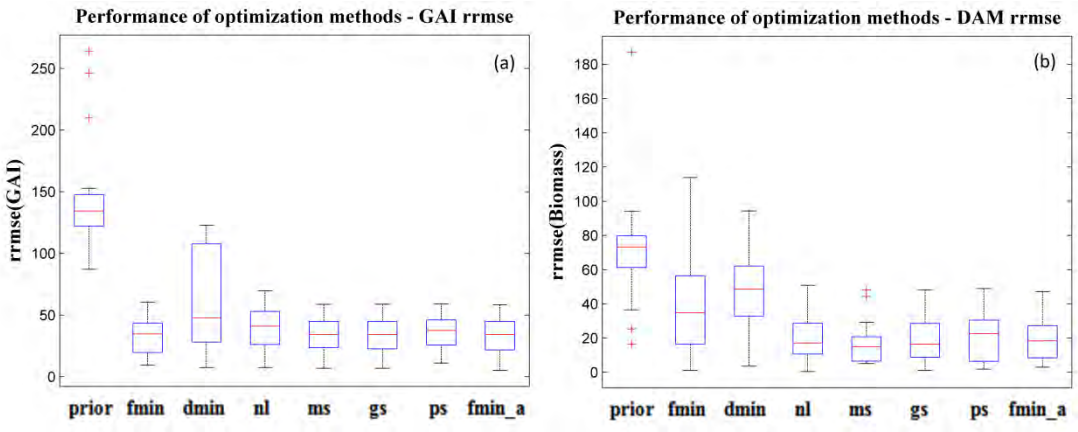


Figure 4.5- Comparison of performances of the seven tested optimization methods based on GAI estimates. (a) Box-plot of rrmse on GAI estimation, calculated using data from 22 wheat fields. (b) Box-plot of rrmse on biomass estimation, for the same fields.

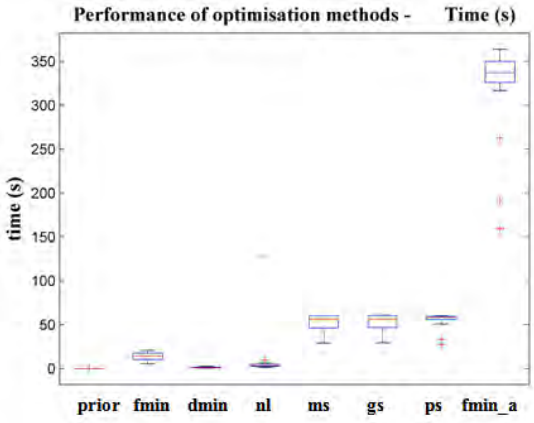


Figure 4.6- Comparison of the computation-time for the different optimization methods.

Those figures show that in terms of GAI the methods MultiStart (*ms*), GlobalSearch (*gs*) and modified simplex (*fmin_adapt*) have very similar performances, with a rrmse around 30% and approximately the same dispersion. We can also note that when no optimization is performed, the initial parameters yields in very poor results.

In terms of biomass (Figure 4.5(b)), the *MultiStart* method presents very good performances (rrmse around 15%), but with two outliers. The methods *fmin_adapt* and *GlobalSearch* are quite close, with similar dispersion and errors of 18% and 16%, respectively.

In general, the methods MultiStart (*ms*), GlobalSearch (*gs*) and modified simplex (*fmin_adapt*) seem to be the best adapted to our optimization problem. However the methods *nl* and *pt*, are not very far, providing good estimates as well. Finally, the *fmin* and *dmin* methods present the worse results. For the *fmin*, it is probably because the solver finds a local minimum, but not a global minimum (which is different for the *fmin_adapt*, that tests different starting points). The *dmin* method could probably yield better results if its coefficients were better adjusted. As this comparison of optimization methods was not supposed to be exhaustive, but only indicative, we let this possibility for improvement as a perspective.

Regarding the computation time (Figure 4.6) of each method to find the best solution, we observe a great disparity between them. The *fmin_adapt* is by far the most *time-consuming* approach. If better calibrated, the *dmin* may present great potential, being the faster one. It is followed by the *nl* method, very efficient as well. The methods *ms*, *gs* and *ps* display similar time performances. Since we have purely scientific objectives, the running-time is not a criterion of interest. However it can be useful for future operational use of the model.

Despite the *MultiStart* yields slightly better results; we decide to keep the *fmin_adapt* as our chosen method, for technical reasons. Given that we use Matlab

software for our simulations, we shall note that *fmin_adapt* employs functions present on the standard Optimization Toolbox, while the *MultiStart* approach requires another toolbox, called Global Optimization, which was not available during the thesis.

d) Influence of using biomass measurements for optimization

All the methods described so far considered only the GAI output variable for calculating the cost function, inner to each approach. As the biomass is one of the major outputs of SAFY and since in some cases biomass measurements are available, we thought about integrating biomass on the optimization cost functions, in addition to the GAI.

In Figure 4.7 and Figure 4.8, we can see the comparison between the seven optimization methods for the two cases: 1) only GAI on the cost function (white boxes); 2) GAI and biomass on the cost function (cyan boxes).

We conclude that the inclusion of biomass information on the function cost for optimizing SAFY parameters does not improve necessarily the results. For estimation of the GAI the results are slightly deteriorated (Figure 4.7). In terms of biomass, we observe performances can be improved, depending on the optimization method. For the modified simplex method (*fmin_a,* last column) a significant improvement on biomass estimation is observed, when using the available biomass measurements for the optimization.

As we aim at developing an approach depending exclusively on remotely sensed GAI data for future regional applications, we will not include the biomass in the cost-function calculation. However, for other studies it may be useful to integrate biomass measurements on the calibration of given parameters or still to integrate CO_2 and ETR measured fluxes for the SAFY-CO_2 and SAFYE-CO_2 models (that will be presented in sections 4.2 and 4.3).

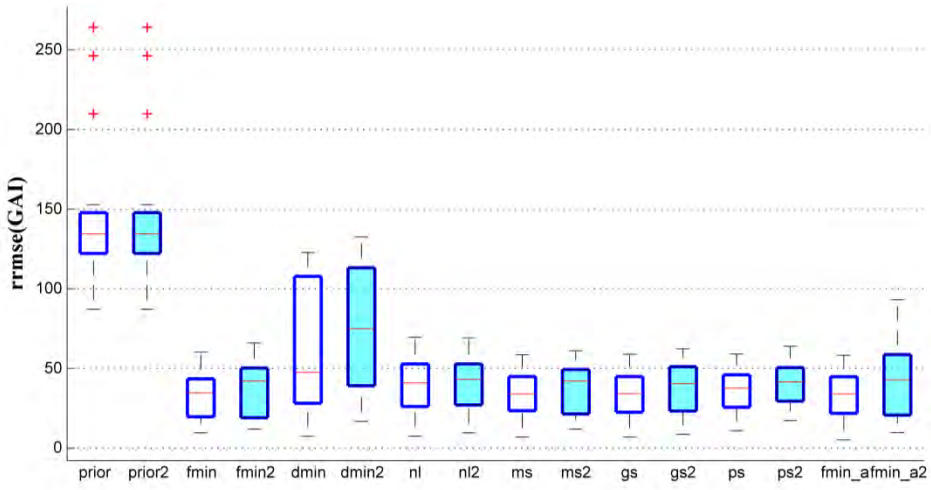


Figure 4.7- Performances of the optimization methods over GAI estimation when changing the cost function. White blocks represent the results obtained by applying a cost function depending only on the GAI; cyan blocks indicate performances obtained when GAI and biomass are account for the cost function assessment.

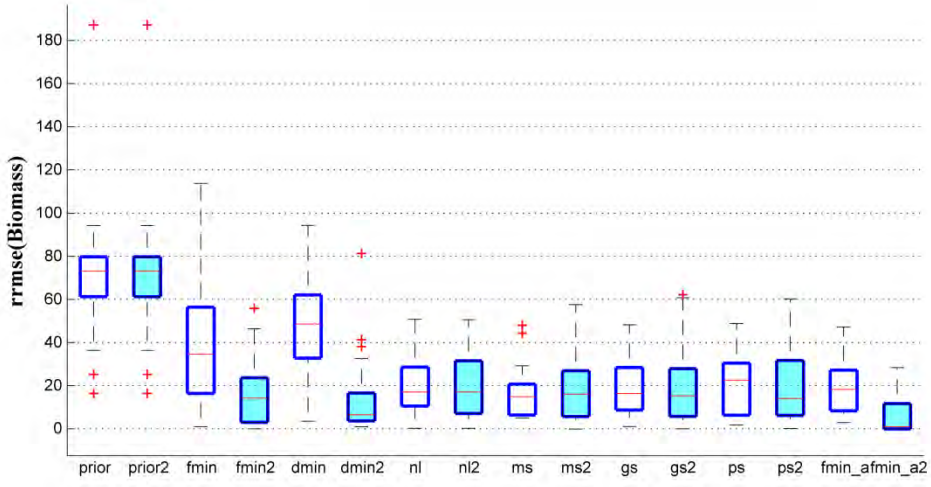


Figure 4.8- Performances of the optimization methods over biomass estimation when changing the cost function. White blocks: cost function depending only on the GAI; cyan blocks cost depending on GAI and biomass.

4.1.4.3.2. SAFY calibration method: further features

In this study, we applied a calibration methodology that differs from the one developed by Duchemin et al., (2008) and applied by Claverie (2012). They established a two-step calibration method. In Claverie (2012) the following parameters P_{La}, P_{Lb}, STT, Rs, D_0 and ELUE were retrieved using GAI time series derived from Formosat-2 images. In order to limit compensation during the optimization procedure, the parameters were classified into two groups: crop-

specific (first phase) and field-specific parameters (second phase). The four crop-specific parameters (P_{La}, P_{Lb}, STT, Rs), which constrain the shape of the GAI time course, were calibrated, on phase 1, separately for each crop type. The two field-specific parameters (D_0 and ELUE) were calibrated, on phase 2, for each simulation unit (see Claverie (2012) for more details). This approach was successfully applied to the summer crops.

Initially we tested the same method over the wheat experimental fields, in the same study area as Claverie's. The parameters values intervals were defined according to the above cited studies. This approach, however, did not yield to accurate results.

Afterward, we developed a new calibration approach, better adapted to the winter wheat crop over our study area. Its consists in calibrating all the cited parameters (P_{La}, P_{Lb}, STT, Rs, D_0 and ELUE) in one single step, but also in constraining the intervals of parameters values to narrower ranges. Besides, the SLA parameter, which was constant for the cited authors, will here be also calibrated with the other six parameters. Letting these seven parameters "free" to be calibrated, but finding solutions in narrower intervals, allowed the optimization process finding better results than with the two-step method.

The following section gives details on how the intervals for optimization of P_{La}, P_{Lb}, STT, Rs, D_0,ELUE and SLA parameters were set.

4.1.4.3.3. Defining parameters values intervals for calibration

Partition-to-leaf parameters (P_{La} and P_{Lb})

Contrary to Claverie (2012), we found that it was not possible to set the parameters P_{La} and P_{Lb} constant for all the wheat fields. Claverie ran the SAFY model over a number of fields available in the study area, letting these parameters free. In a second phase, for each crop type, P_{La} and P_{Lb} were set equal to the median of the histogram of P_{La} and P_{Lb} values found through the first phase (same methodology applied to senescence parameters Rs and STT). It means that these parameters were crop-specific (and not field-specific), being fixed to a same value according to the crop type. We tested this same approach, and the P_L (partition to leaf) function obtained from the median P_{La} and P_{Lb} from SAFY simulations over a number of wheat fields and different years can be seen in Figure 4.9 (represented by the orange line). In the same figure we plot the P_L function estimated from wheat *in-situ* measurements (for different years and sites as well), which is represented by the purple line. We observe that the P_L function from the median approach does not match with the function estimated from field measurements. As it is shown in Figure

4.10, this was not the case for Claverie (2012), for which both *in-situ* and simulated P_L functions corresponded well. This divergence may occur because we work with a winter crop instead of summer crops and also because more data were available and on 6 years with very contrasted climatic conditions.

Therefore, we could not use the same optimization as Claverie's. For this study, the parameters P_{La} and P_{Lb} are let free to be optimized for each specific field. The values for the optimization boundaries were determined according to our field *know-how* and to the studies performed by Duchemin et al., (2008). The chosen intervals for P_{La} and P_{Lb} optimization are: P_{La}:[0.01-0.5] and P_{Lb}:[0.0001-0.02].

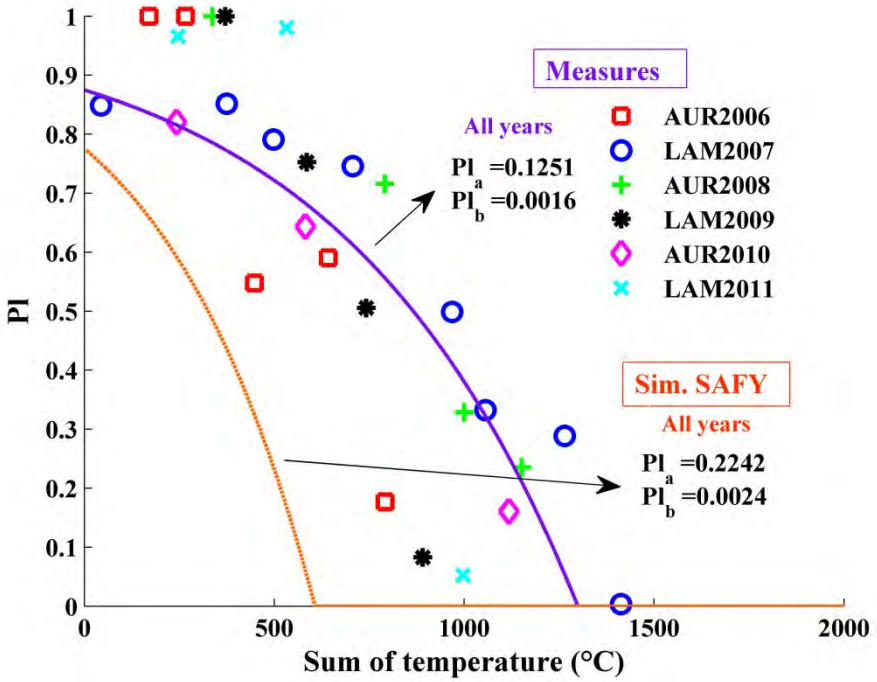


Figure 4.9- Comparison of partition-to-leaf function P_L estimated from *in-situ* measurements (represented by the purple line) and established from SAFY simulations (two-step calibration, represented by the orange line). The measured P_L is derived from destructive measurements performed over the Auradé and Lamasquère sites, for different years.

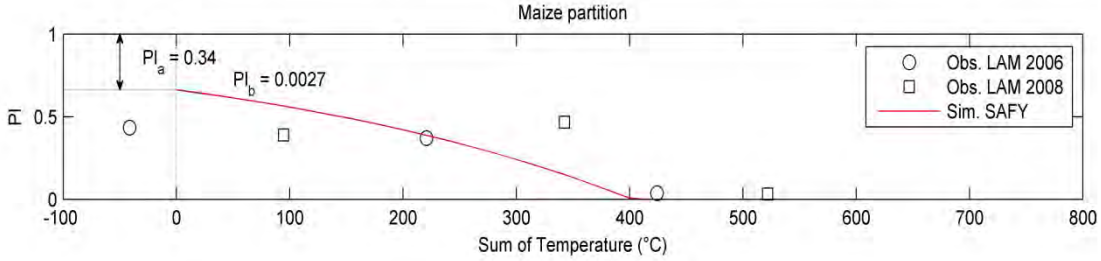


Figure 4.10- Extracted from Claverie (2012). Comparison of P_L function estimated from *in-situ* measurements and simulated by SAFY. Values established for maize, based on two years and one site measurements.

Senescence parameters (STT and Rs)

The sum of temperature for senescence (STT) and senescence rate (Rs) are the parameters the least sensitive on the equifinality issue. The limits were set according to Duchemin et al., (2008) and are: STT:[200 -2000] and Rs:[10^3-2x10^4].

Specific Leaf Area (SLA)

The SLA parameter was measured at our experimental sites from 2006 until 2011. SLA is defined as the regression coefficient between the leaf area (LA) and the leaf dry mass (LDM), obtained from destructive measurements. It's known that the SLA varies over the crop cycle. However for SAFY it remains constant. As in SAFY, the SLA parameter interferes only in the growing phase [Eq.(4.5)], we use sole destructive data collected during the growing stage to determine the SLA. Figure 4.11 illustrates the relationship between LA and LDM for the winter wheat. The slope value corresponds to the SLA, which is 0.0159 $m^2.g^{-1}$. This value is slightly lower than the values found in the literature (0.022, 0.024 or 0.025 $m^2.g^{-1}$ in Duchemin et al., (2008), Maas (1993), and Arora and Gajri, (1998), respectively). It is probably due to different genotypes and to environmental factors, such as air temperature (Hotsonyame and Hunt, 1998).

As we aim at establishing a general methodology, we choose to optimize the SLA (not initially fixed). Therefore SLA was initially set to 0.022 $m^2.g^{-1}$ as a starting value, since it is an intermediate value between values from the literature and from destructive measurements. We took 30% over and below this value to establish the limits for the optimization process. We obtain then SLA_{min} =0.0154 and SLA_{max}=0.286.

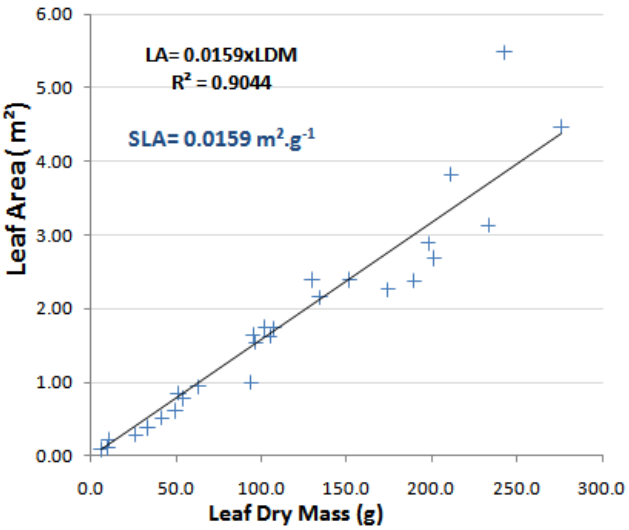


Figure 4.11- Relationship between leaf area (LA) and leaf dry mass (LDM) for the winter wheat, based on destructive measurements performed over the Auradé and the Lamasquère sites. The SLA value (0.0159) is obtained from the linear regression, and corresponds to the slope of the curve.

Day of emergence (D_0)

The day of plant emergence (D_0) strongly depends on agro-environmental conditions. D_0 occurs generally 1 to 3 weeks after sowing depending on the soil temperature and moisture. The values for the boundaries of this parameter will be determined according to the filter applied to the observed GAI profile (from satellite) for each of the studied fields.

Thus, before performing the calibration process, we filter the GAI profile in order to delimit the vegetative period. First we chose the temporal windows within begin and end of the growing season. For example, we know that for winter wheat, the growth period should start between September and January, and the end of the season should occur between May and August. Then, applying conditions related to the maximum and minimum GAI, we set the initial D_0 and day of end of the simulation, and also the limits for optimization of the D_0 parameter: D_{0min} and D_{0max} (Figure 4.12). The GAI values that were not included in the identified vegetative period were excluded.

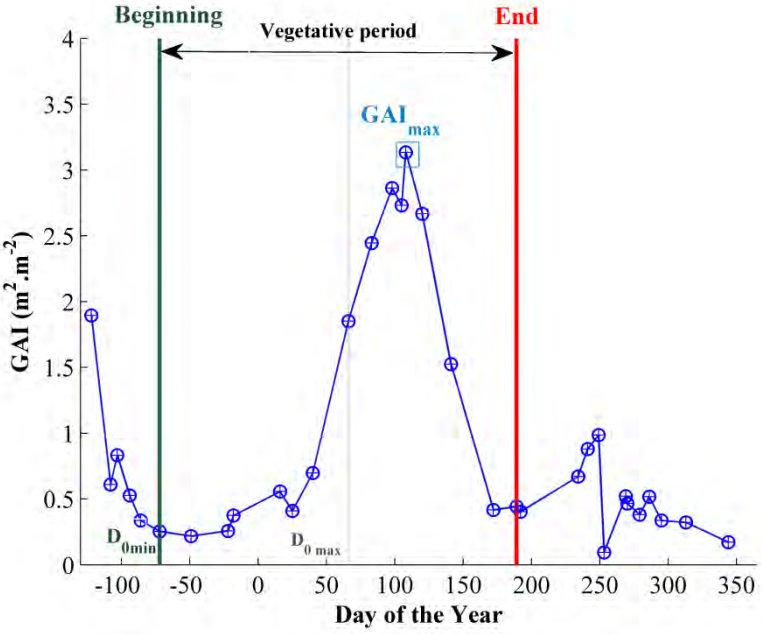


Figure 4.12- Example of the delimitation of the growing season of a wheat GAI profile. The circles indicate the dates for which satellite acquisitions were done. The maximum GAI is framed in light blue. The green vertical line indicates the earlier possible date for beginning crop simulation (correspondent to D_{0min}) and the gray dot-line indicates the later possible date (D_{0max}). The red line indicates the end of the growing season.

Effective light-use efficiency (ELUE)

In this study, the effective light-use efficiency (ELUE) parameter is supposed to account for all agro-environmental stresses, except for temperature. This parameter is expected to give in a simple manner a global level of all these agro-environmental stresses integrated together, which could be an indicator of the performance of agricultural practices, such as the irrigation and fertilization schedules. Preliminary simulations using the range of variation of ELUE between [0-10], as documented by Duchemin et al., (2008), yielded in overestimated biomass estimates. Further analysis showed that when choosing a narrower interval [1.4-2.8], better results were retrieved.

The SAFY parameters and their respective ranges of variation are summarized in Table 4.1.

4.1.5. Methodology for calculating yield

Estimating yield constitutes one of the main objectives of this thesis. Apart the fact that yield estimates are important for agronomical and economical issues at regional, national and global levels, yield estimates are also essential for calculating the carbon budget of the plots, since the carbon export term drives most of the carbon budget, followed by the net ecosystem exchange (NEE) (Ceschia et al., 2010). Our approach, as modelers, is based upon the crop growth SAFY model for estimating aboveground biomass production, from which yield will be estimated. Prior to the description of yield estimates with SAFY, we present a preliminary study that aims to establish a linear regression model between remotely-sensed NDVI indices and yield during the crop vegetative period.

4.1.5.1. Yield estimates from NDVI

Through this succinct study, we aim at establishing an empirical linear regression model linking yield and NDVI data for our three different sources of yield data: i) from the field campaign carried out in 2011, ii) from the farmers' surveys and, iii) from the yield monitors mounted on the combine harvesters.

As we have these different data sets of yield *in-situ* data, we chose working initially with only two of them: field campaign and data from yield monitors. Because of the incertitude related to the data acquired through the farmers' surveys, this data set is not considered for the moment.

The regression approaches for predicting yield must necessarily be calibrated using appropriate reference information. In most cases, agricultural statistics and, specifically, crop yield are used as reference information. This pre-requisite limits its applicability in many regions of the world (Rembold et al., 2013). Several studies have examined how the relationship between remotely sensed data and crop yields

varies as a function of time during the growing season (Mkhabela et al., 2011; Wall et al., 2008). Thus regression coefficients show strong temporal and spatial variations.

Various authors postulated that accumulated radiometric data are more closely related to crop production than instantaneous measurements (Becker-Reshef et al., 2010; Labus et al., 2002; Rojas, 2007), since it can capture the effect of adverse events which occur after flowering. Therefore several choices of instantaneous NDVI value or temporal *NDVI* integration are possible for establishing the crop production regression model. One of the simpler options is to use the maximum *NDVI* value of the season, especially if the objective is having an early prediction of production.

An alternative approach is to integrate the NDVI values around the maximum NDVI, or even to integrate the *NDVI* values over the total crop cycle. Yet, other authors found that reliable yield forecasts and predictions were determined using measurements at the end of the growing season (Jégo et al., 2012). For this work, we tested two different methods for linking NDVI information to yield data.

The first approach we adopted consisted in using a single NDVI date to establish a regression model with yield data [Eq.(4.10)]. For this purpose, we performed a regression for each date of the vegetative period (from sowing until harvest), in order to find the date for which the correspondent NDVI is best correlated to the yield. To determine this 'most correlated' NDVI day, we estimated the coefficients of correlation (R^2) for all the available plots from the emergence day until the end of the senescence. We then selected the NDVI date that provided the highest R^2 for estimating a linear regression based on the available points.

The second approach consisted in using NDVI integrated over different periods, mainly around critical growth stages (Belaqziz et al., 2013). In order to find the accumulated-NDVI that better correlates with the yield data, we tested thirteen ways of accumulating NDVI over 10 days at different stages. Initially, around the maximum seasonal NDVI, and then from March until end of May (which corresponds to active growing phase and just after flowering). Knowing that the flowering is a critical stage of the wheat development, and that the crop conditions around this period may summarize the history of crop growth and development (that will affect yield), we tried to identify when the flowering phase occurs (day of start of flowering). For this purpose, we ran the STICS model for wheat over one of our experimental sites, in 2011. We used then the 'flowering day' found by STICS to integrate NDVI values around this date. We also tested integrating NDVI over larger periods: from flowering stage until the end of the season and also over the whole vegetative period. A total of 15 different periods of NDVI accumulation were tested.

For each test, we established an empirical model based on the integrated NDVI and the yield data, having the form expressed in Eq.(4.11). Therefore, we

determined the period of accumulated-NDVI the 'most correlated' to the yield data, by evaluating the correlation coefficients.

Initially, we applied each of the described approaches over the two selected data sets separately ('field campaign' and data from 'yield monitors'), and then we merged these two data sets into one, for establishing a joint regression model based upon a larger data set. In order to check if the established relationships could be applied over other sites and years, we applied the joint 'best' models over a third yield data set: the farmers' surveys data set.

$$Yield = aNDVI_{best} + b \quad (4.10)$$

$$Yield = a\sum NDVI + b \quad (4.11)$$

4.1.5.2. SAFY yield estimation

In the initial version of the SAFY model, developed by Duchemin et al., (2008), the grain yield is calculated from a constant fraction (Py) of the daily increase of the total above-ground biomass (DAM) that is partitioned to grains [Eq. (4.12)]. The grain filling phase is bounded by the day when foliage production ends and the day when total senescence occurs.

$$\Delta Yield = \Delta DAM \times Py \quad (4.12)$$

However, Claverie (2012) used a simpler approach for estimating final yield, by multiplying the final total aboveground biomass by a harvest index (HI) [Eq. (4.13)]. This harvest index is crop-dependent and it is fixed according to field measurements of biomass (sites + ESUs) and yield (farmers' surveys).

For establishing the harvest index of winter wheat in our study area, we performed a traditional regression analysis, by plotting the destructive biomass against yield measurements, derived from our database. We obtained a harvest index equal to 0.45±0.05 (Chapter 2.3.2.1). This value is included in the range found by Unkovich et al., (2010) for wheat in Australia (HI=[0.08-0.56]). It is also in agreement with the values determined by Li et al., (2011) with HI between 0.39 and 0.44 for the wheat in China. In Southwest France, inside our study area, Debaeke et al., (1996) found HI between 0.19 and 0.439 for winter wheat cultivated on different soil types and under different water stress conditions.

$$Yield = DAM_{\max} \times HI \quad (4.13)$$

We also tested a different approach for calculating yield, by including in SAFY the equations used in the STICS model (Brisson et al., 2003). According to this approach, the dry biomass accumulated into the grains is calculated from a harvest index that evolves through the season (the IRCAB index, according to STICS notation). This index may increase in two different ways. First, according to a linear function of time [Eq. (4.14)], from the stage 'day of start of grain filling' (DGF) until the harvest. Second, according to a linear function of the crop development (in thermal time, Eq. (4.15)). From the DGF until harvest as well.

The fixed indexes VITIRCARB and VITICARBT, used to calculate the index IRCAB, were set according to the values found in the file *.plt* (contained in STICS, version 6.2). They correspond to the rate of increase of the harvest index. Their values are listed in Table 4.2.

The day of start of grain filling (DGF) was determined as the day when the sum of temperature (SMT) reaches a threshold STY, corresponding to sum of degree-days from emergence (D_0) for starting grain filling. It corresponds to the *stlevdrp* parameter in STICS.

The *stlevdrp* was available for six wheat crop varieties (Arminda, Talent, Thesee, Soissons, Promentin, and Sideral), including early and later types. The crop variety information is seldom available, thus the STY threshold is set as the weighted average value from the available varieties (Table 4.2). The daily grain yield increase is computed from the biomass production multiplied by the evolving harvest index IRCAB (Eq. (4.16)].

Table 4.2- Indexes for yield estimation accord to STICS model

Parameter	Value	Unit
VITIRCARB	0.01100	$g_{grain}.g_{plant}^{-1}.day^{-1}$
VITICARBT	0.00070	$g_{grain}.g_{plant}^{-1}$ °C-day^{-1}
STY	796	°C-day

$$IRCAB\ (i) = VITIRCARB\ \times (i - DGF\ ) \quad (4.14)$$

where $(i - DGF\ )$ represents the number of days spent from the day of start of grain filling (*DGF*) and the current day *i* and VITICARB is a constant (see Table 4.2)

$$IRCAB\ (i) = VITIRCARBT \times \sum_{DGF}^{i} f(Ta) \quad (4.15)$$

where $\sum_{DGF}^{i} f(Ta)$ represents the sum of degree-days from DGF until day *i* and VITICARBT is a constant (see Table 4.2)

$$\Delta Yield = IRCAB \times \Delta DAM \quad (4.16)$$

4.2. The SAFY-CO_2 model

4.2.1. Estimation of the CO_2 fluxes components

In order to simulate the components of the carbon budget, the SAFY-CO_2 model has been developed by modifying the SAFY crop model. We aim at establishing a robust and simple method, which requires few data, for estimating the gross primary production (GPP) and the net ecosystem exchange (NEE), as well as the autotrophic and heterotrophic respirations (Ra and Rh, respectively) of croplands. Figure 4.13 is a scheme representing the main terms of the NEE and of the carbon budget for our area of study. Next paragraphs describe the methods used for estimating these components.

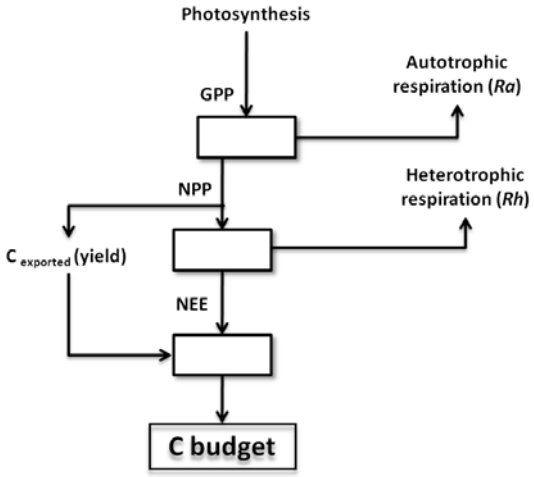


Figure 4.13- Diagram representing the main components of the CO_2 fluxes at the ecosystem/atmosphere interface and the main terms of the C budget in our study area.

4.2.1.1. GPP estimation

An accurate quantification of GPP in crops is essential for studies of carbon budgets at regional and global scales. The gross primary productivity (GPP) can be defined as the total amount of carbon dioxide that is fixed by the plants (Gitelson et al., 2012). One of the most widely applied concepts for modeling GPP is the light use efficiency approach of Monteith (1977).Thus in SAFY-CO_2 model the estimation of GPP is also based on the Monteith's theory, like the biomass estimation in the standard SAFY model (see 4.1.1), except that in this case the ratio GPP/Ra was supposed to be constant, which is often not true. The main difference is therefore that, instead of calculating the dry aboveground biomass (DAM), we calculate at first the GPP. The GPP is function of the incoming global radiation (Rg), the fraction of radiation absorbed by photosynthetically active elements of plants (fAPAR), the

climatic efficiency (ε_c), the efficiency of the absorbed radiation being converted to fix some CO_2 by plant photosynthesis (ELUE) and a temperature-stress function [Eq. (4.17)]. Note that in this case, the ELUE has a different meaning than for the original SAFY model.

$$GPP = Rg \times \varepsilon_c \times FAPAR \times ELUE \times F_T(Ta) \quad (4.17)$$

$$FAPAR = 1 - e^{-k_{ext} \times GAI} \quad (4.3)$$

To take into account the fraction of green tissues of the plant remaining during the senescent phase, we added a multiplicative factor ($sR_{10,}$ Beziat et al., 2009) to the GPP estimation [Eq. (4.18)]. The sR_{10} coefficient is set to 1 from the emergence day (D_0) until the end of the plateau, i.e., the day when the senescence begins. From this day until the end of the simulation, sR_{10} consists in the quotient between the GAI value of day *i* and value of maximum GAI multiplied by a corrective factor *Cs* [Eq. (4.19)]. Throughout our studies we realized that the coefficient sR_{10} was not sufficient to correct the effect of senescence over the simulated fluxes, so the *Cs* coefficient was included on the sR_{10}'s determination. It might be explained by the fact that at first the senescence phase acts on the lower parts of the plant (closer to the soil), and then on the higher canopy elements. It means that in the real (phenological) senescence may be more accentuated than what is detected by the satellite observations, requiring thus a corrective factor. Some empirical tests were performed in order to set an optimal value for *Cs*, establishing *Cs=1.2* for the sequence of our works.

$$GPP = Rg \times \varepsilon_c \times FAPAR \times F_T(Ta) \times fELUE \times sR10 \quad (4.18)$$

$$sR10 = \begin{cases} 1 & , \ from\ emergence\ until\ start\ of\ senescence \\ \dfrac{GAI}{GAI_{\max} \times Cs} & , \ from\ senescence\ until\ end \end{cases} \quad (4.19)$$

Effect of diffuse fraction over GPP estimation

The effects of the fraction of diffuse radiation over canopy photosynthesis are not usually considered by the crop models when estimating crop productivity. However, measurements, particularly from forest ecosystems, have shown that the efficiency of canopy gas exchange is very sensitive to the diffuse component of the incoming

solar radiation (Hollinger et al., 1998; Roderick et al., 2001). Besides, these effects have been highlighted over the experimental sites of Lamasquère and Auradé by Béziat et al., (2009). These remarks suggest that models which ignore the diffuse component of solar radiance may probably incorrectly simulate the changes in CO_2 uptake (Pury and Farquhar, 1997; Roderick et al., 2001). One way to observe the direct effect of the diffuse irradiance on canopy photosynthesis is to use the Monteith's light-use efficiency model.

It is known that the diffuse fraction of incoming solar radiation (*Rdf/Rg*) is strongly related to the fractional transmission of the solar irradiance through the atmosphere (*Rg/Ra*), where Ra= extraterrestrial radiation) (Bindi et al., 1992; Roderick, 1999; Spitters, 1986). As these two variables (*Rdf/Rg* and *Rg/Ra*) are negatively correlated, the diffuse fraction increases when the atmospheric transmission of solar radiation declines. Meanwhile, when R_{df} / R_g increases, there is a reduction in the proportion of shaded leaves within the canopy. Further, because the photosynthetic rate of leaves usually saturates at high incoming radiation, it follows that leaves in low irradiance will have a higher efficiency, and the reduction in the volume of shade within the canopy means that the canopy as a whole will also be more efficient in low irradiance (Roderick et al., 2001). It is then expected that efficiency should generally increase as *R*df/*R*g increases (and *Rg/Ra* decreases). As stated by Goudriaan, (1977), for a given value of solar radiation an increase in the diffuse component improves the distribution of light over the canopy and crop assimilation rate increases.

Given these remarks, and based on preliminary simulations, we developed an approach for correcting the effective-light use efficiency (*elue* parameter), based on the fraction of diffuse and global incoming radiation (*Rdf/Rg*).

First step consisted in calculating the ratio *Rdf/Rg*. Based on the literature, we chose to evaluate three models linking *Rdf/Rg* and *Rg/Ra*. The study was performed using daily values of the referred variables.

De Jong (1980) proposed a relationship [Eq. (4.21)] between the diffuse fraction of global radiation and the atmospheric transmission of solar radiation using data from a site in Netherlands. His model was retaken by Spitters et al., 1986, who has spread it within the scientific community. The second model is the one described by Collares-Pereira and Rabl, (1979). They established a model using multi-years data over five sites in the USA. The third method is more recent and was determined by Roderick (1999) using data measured in several sites in Australia and Antarctica. The Roderick's generic model was set to 43°N, latitude for our study area.

Figure 4.14 shows the relationship between diffuse fraction (*Rdf/Rg*) and atmospheric transmission (*Rg/Ra*) for the three cited models, using daily data from

our experimental sites. The extra-terrestrial radiation (*Ra*) was calculated according to Allen et al., 1998 (FAO, equation 21).

From 2005 to 2010 our experimental sites were not equipped for measuring the diffuse incoming global radiation. On the other hand, measurements of the total and diffuse incoming photosynthetic photon flux density (PPFD or PAR) were performed. It was thus necessary to convert the PPFD data (spectral response: 400nm-700nm) into global radiation (400-2700nm). It is known that photosynthetic activity is almost twice as great in the red part as in the blue part of the spectrum. More precisely, according to Spitters et al., (1986), the diffuse fraction in the PAR wavebands can be obtained from the diffuse fraction of global radiation, as shown in Eq. (4.20). By inverting this equation, from the available PAR_{df}/PAR_g data we obtained the ratio *Rdf/Rg* for our sites since 2005. The *in-situ Rg/Ra vs Rdf/Rg* data are shown in Figure 4.14Figure 4.15 (black dots).

$$PAR_{df}/PAR_g = \left[1 + 0.3 \times \left(1 - \left(R_{df}/R_g\right)^2\right)\right] \times \left(R_{df}/R_g\right) \tag{4.20}$$

From Figure 4.14 we note the observations from our sites (relating *Rdf/Rg* and *Rg/Ra*) are in agreement with the literature models. Still, for the interval $R_g/R_a < 0.25$ we observe the three models overestimate the *Rdf/Rg* values. The performances of the models against the in-situ data were evaluated in terms of correlation (r^2) and relative root mean square errors (rrmse). The results are listed in Table 4.3.

As the performances are satisfactory (and very close), we can affirm that the models are valid for our study area. The relationships are remarkably similar over different climates and latitude. For the rest of this work, we choose the DeJong's method, which is detailed in equation (4.21).

$$\begin{cases} R_{df}/R_g = 1 & for\ R_g/R_a < 0.07 \\ R_{df}/R_g = 1 - 2.3(R_g/R_a - 0.07)^2 & for\ 0.07 \le R_g/R_a < 0.35 \\ R_{df}/R_g = 1.33 - 1.46(R_g/R_a) & for\ 0.35 \le R_g/R_a < 0.75 \\ R_{df}/R_g = 0.23 & for\ 0.75 \le R_g/R_a \end{cases} \tag{4.21}$$

where R_{df}, R_g, R_a are diffuse, global and extra terrestrial global radiation, respectively.

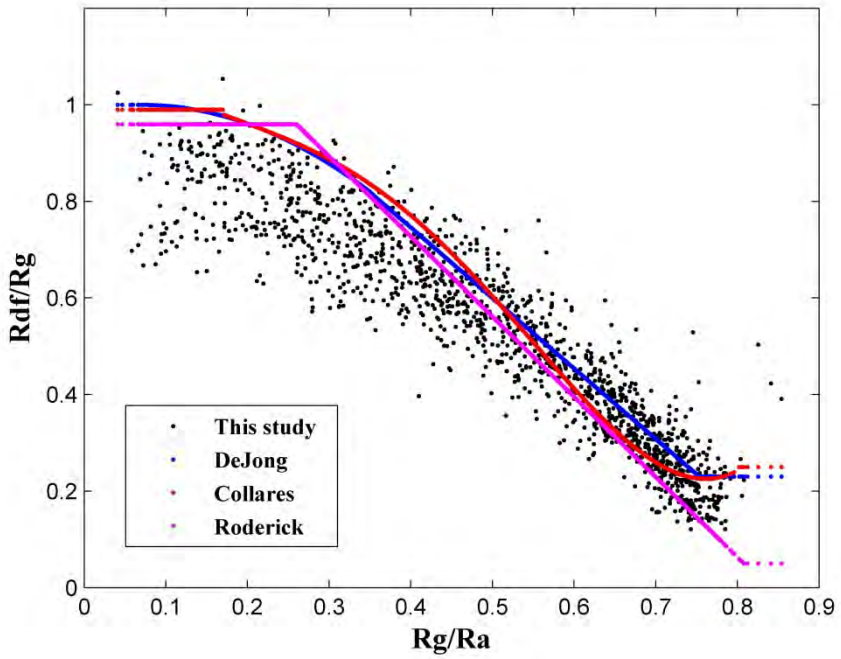


Figure 4.14- Comparison of daily models used to separate the diffuse component from the global radiation. In black, the data from our Auradé (years 2006, 2008, 2010) and Lam asquère (2007) sites. A total of N=1944 observations was available. In blue, the model from De Jong (1980), (described in Eq.(4.21)), established using data from a single location in the Netherlands. In red, the model based on the equations determined by Collares-Pereira and Rabl (1979), from multi-years data over five USA cities. In magenta, the generic model established by Roderick (1999) using data from several sites in Australia and Antarctica. The Roderick's model was set to 43°N, latitude for our study area.

Table 4.3- Performances of the comparison of the literature models laws against *in-situ* data collected in our study sites

Statistics	**Models**		
	DeJong	**Collares**	**Roderick**
rrmse (%)	19.76	20.68	20.87
r^2	0.89	0.89	0.89

Once the fraction of diffuse radiation is calculated, we aim at establishing a relation to link the diffuse fraction to the efficiency of GPP production. Field data from the radiation sensors mounted on our experimental sites allow establishing an exponential function between the ratio *Rdf/Rg* and the radiation use efficiency (RUE) (see Figure 4.15). The RUE was defined here as the ratio between GPP and Rg. The relationship between RUE and *Rdf/Rg*is defined as an exponential function, with parameters *a* and *b* [Eq. (4.22)].

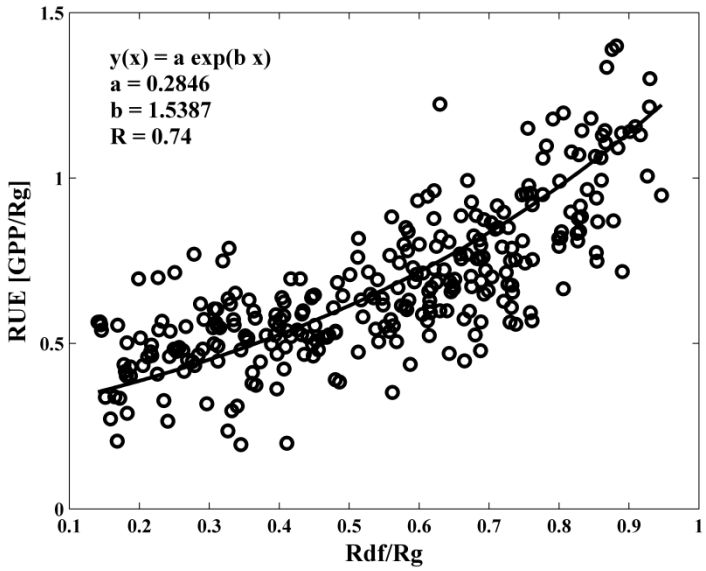


Figure 4.15- Relationship between the diffuse fraction of radiation (Rdf/Rg) and the radiation use efficiency (RUE). Data are derived from Auradé (2006, 2008 and 2010) and Lamasquère (2007) experimental sites (black circles). Only data acquired during well developed vegetation were used (GAI>1.5). The black line represents the fitted exponential law, characterized by the coefficients *a* and *b*. A correlation of R=0.74 was found between the two variables.

In SAFY-CO_2, and by comparison with SAFY, we replaced the ELUE by a function called *fELUE* [Eq. (4.23)] defined as the product between the exponential equation described previously and the effective-light use efficiency (ELUE). This term was thus incorporated on the estimation of the daily GPP estimates [Eq. (4.18)].

$$RUE = \left(0.28 \times e^{R_{df} / R_g \times 1.54}\right) \tag{4.22}$$

$$fELUE = \left(0.28 \times e^{R_{df} / R_g \times 1.54}\right) \times ELUE \tag{4.23}$$

4.2.1.2. NPP, Ra and DAM and Yield estimation

The Net Primary Productivity (NPP) is an important variable in studies of global carbon cycle. NPP refers to the net production of organic carbon by the plants (Ruimy et al., 1999) and corresponds to the carbon accumulation over a year basis, although the actual duration is shorter for many crops (emergence to harvest) (Choudhury, 2001). It is defined as the GPP minus the amount of carbon respired by plants themselves, which is the autotrophic respiration (R_a). The equation (4.24) shows how to calculate the daily net primary productivity.

For estimating autotrophic respiration, we chose an approach (McCree, 1974) that separates the R_a into two components: the maintenance respiration (R_m) and the growth respiration (R_g)[Eq. (4.25)]. Growth respiration is defined as the respiration that is required to produce new structures and maintenance respiration as the CO_2 released from the activities that maintain the existing biomass alive. According to Amthor (2000) the maintenance respiration depends on the total biomass cumulated at a given day and the growth respiration depends mainly on the product of the net photosynthesis of the previous day.

R_m can be calculated from the NPP and the maintenance coefficient m_R .The coefficient m_R corresponds to the fraction of maintenance respiration per NPP unit. m_R responds strongly to temperature (Amthor, 2000) and is estimated by a "Q_{10} type" equation [Eq. (4.26b)]. The Q_{10} coefficient is typically about 2. R_{10} stands for the reference respiration at 10°C. This parameter was calibrated by Béziat (2009) using the *in-situ* measurements over wheat, and it was set to 0.0025 gC released per gC of living biomass of the plant (for more details, see Béziat (2009) and Vries (1989)).

The growth respiration (R_g) was calculated according to the method used by Amthor (1989) and retaken by Choudhury (2000), as showed in (4.26c). The constant Y_G is the growth conversion efficiency, here fixed at 0.74, as the average of three values for wheat given in Amthor (1989). For estimating net carbon accumulation over the global land surface, Ruimy et al., (1996) had considered Y_G to be 0.72. According to Johnson and Thornley (1983), Y_G is typically about 0.75.

$$NPP = GPP - Ra \tag{4.24}$$

$$Ra = R_m + R_g \tag{4.25}$$

$$R_m = NPP \times m_R \times sR10 \tag{4.26a}$$

$$m_R = R_{10} \times Q_{10}^{\left(\frac{T_a - 10}{10}\right)} \tag{4.26a}$$

$$R_g = \left(1 - Y_G\right) \times \left(GPP - Rm\right) \tag{4.26a}$$

$$NPP = NPP_r + NPP_a \tag{4.27}$$

$$NPP_r = NPP \times RtS \tag{4.28}$$

$$RtS = fr = fr_{\infty} + (fr_0 - fr_{\infty})e^{-c\cdot\left(\frac{SMT_{D_0}^{D_S} - SMT_{D_0}}{SMT_{D_S} - SMT_{D_0}}\right)} \quad (4.29)$$

Where SMT: Sum of temperature; D_0: emergence date;
Ds: senescence start day.

$$DAM = \frac{NPP_a}{C_{veg}} \quad (4.30)$$

The total NPP can be defined as the sum of the aerial (NPP_a) and root (NPP_r) components [Eq. (4.27)]. It was necessary to estimate the aboveground component of the NPP, in order to calculate the DAM and also the yield (see below). Therefore, we use a root-to-shoot ratio (*RtS)* to quantify NPP_a and NPP_r. We chose a root-to-shoot ratio, according to the parameterization proposed by Baret et al., (1992b). Equation (4.29) shows the way the *RtS* is estimated. The root fraction (*fr*) is expressed in growth degree days (°Cd) since emergence; fr_0 is the extrapolated *fr* value at emergence; fr_∞ the asymptotic value of *fr* ; and *c* is the relative rate of decrease.

It is important to remark that the simple growth day scale was modified into a phenological time scale. Instead of a simple degree day unit, we use a relative growth degree day scale, that depends on the sum of temperature at emergence day (D_0) and at the day of start of senescence (D_S), as it is shown in (4.29). The related parameters were initially set according to Baret's results (fr_0*=0.6,* fr_∞*=0.1* and *c=1.5).*and then they were slightly modified based on Béziat' studies (unpublished), to better fit with our study sites. They were fixed to fr_0*=0.63,* fr_∞*=0.11* and *c=1.48.*

Finally the dry aboveground biomass (DAM) is estimated by dividing the aerial net primary productivity (NPP_a) per a coefficient that represents the carbon content of plants. This coefficient (C_{veg}) was set to 0.46 gC / g_{veg} , according to measurements performed over our experimental sites, for different crop types. Table 4.4 summarizes the new parameters of SAFY-CO_2 model, in addition to those described for the SAFY model.

The yield estimation is carried out using the same methods described for SAFY model (see section 4.1.5.2), i.e., yield depends on the biomass production and on a harvest index that can evolve (or not) through the season, according to equations (4.12), (4.13) or (4.16).

Table 4.4- List of the additional parameters of the SAFY-CO_2 model: notation, units and values fixed for winter wheat crop.

Description	Notation	Unit	Value	Method
Corrective factor over GPP during senescence	C_S	-	1.2	*In-situ* measurem.
Correction of the ELUE function (*fELUE*): parameter *a*	*a*	-	0.28	*In-situ* measurem
Correction of the ELUE function (*fELUE*): parameter *b*	*b*	-	1.54	*In-situ* measurem.
Maintenance respiration parameter: Q_{10}	Q_{10}	-	2	Literature
Maintenance respiration parameter: R_{10}	R_{10}	gC_{resp} / $gC_{tissus\ viv.}$	0.0025	Literature
Growth respiration conversion efficiency parameter	Y_G	-	0.74	Literature
Carbon content coefficient	C_{veg}	gC / g_{veg}	0.46	Literature
Root fraction parameter: fr_0	fr_0	°Cd	0.63	Literature
Root fraction parameter: fr_∞	fr_∞	°Cd	0.11	Literature
Root fraction parameter: *c*	*c*	-	1.48	Literature
Straw export coefficient	*sc*	-	0.29[LAM2007] 0.31[LAM2009]	*In-situ* measurem.

4.2.1.3. NEE and Rh estimation

The Net Ecosystem Exchange (NEE), also known as Net Ecosystem Production (NEP) when cumulated on an annual basis, can be calculated as the difference between the net primary production and the carbon losses by heterotrophic respiration (*Rh*) [Eq.(4.31)].

$$NEE = NPP - Rh \tag{4.31}$$

$$Rh = a \times e^{b \times T_S}$$
$$Rh_{ref} = a \quad and \quad Q_{10} = e^{b \times 10} \tag{4.32}$$

The heterotrophic respiration refers to the carbon lost by organisms in ecosystems other than the plants. It mostly represents the respiration of the microorganisms that live in the soil and decompose the organic matter that has reached the soil. For our study we calculate *Rh* through an empirical equation that was tested on our flux sites (provided by Delogu (2013)).

The *Rh* equation is an exponential type, dependent on the soil temperature. The parameters *a* and *b* were calibrated using the NEE data from the flux towers (Auradé and Lamasquère), from 2006 to 2010, over periods of bare soil. A total of 680 days of data were used for parameters calibration. The parameter *a* corresponds to the 'reference *Rh',* and its unit is gC m^{-2}.day^{-1}. The parameter *b* is linked with the Q_{10} constant [Eq.(4.32)]. Two methods were tested: one leaving both parameters *a* and *b* free, and the second one by setting Q_{10} to 2.3 (equivalent to *b*=0.0833). The last one (Q_{10} constant and Rh_{ref} free) resulted in more accurate results, so we kept it for all the further simulations. Table 4.5 shows the adjusted parameters for both cases.

The soil temperature (Ts) was calculated from the air temperature (Ta) multiplied by a corrective factor. This factor was found by the relationship between air and soil surface temperatures data measured by the meteorological towers in the experimental sites of Auradé and Lamasquère (Figure 4.16). We used data measured during the whole year (including vegetative and bare soil periods), from 2006 until 2010. Given the large amount of data used, the density plot (on the right of Figure 4.16) helps visualizing the relation between the two variables. Therefore, a factor of 1.07 was used for estimating Ts from Ta.

Table 4.5- Heterotrophic respiration (Rh) parameters (extracted from Delogu (2013)).

	Q_{10} and Rh_{ref} free	**Q_{10} constant and Rh_{ref} free**
Rh	0.68	0.34
Q_{10}	1.6	2.3
RMSE	0.29	0.25
R	0.67	0.73

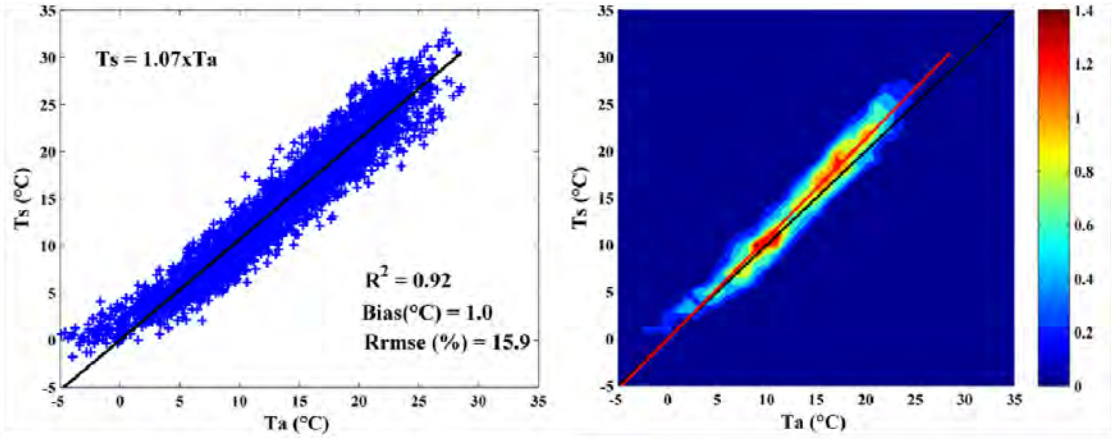


Figure 4.16- Relationship between air (Ta) and soil (Ts) temperatures. Data measured from 2006 to 2010, over the whole year, including vegetative and bare soil periods. On the right, the density plot of the left graphic.

4.2.2. Crop carbon budget

The net ecosystem production (NEP) is defined as the annual integration of the daily NEE values. As it was done by Béziat et al., (2009), the NEP was computed from the 1st October until the 30th September; because this period usually corresponds to an agricultural year, beginning after harvest of the summer crops and before the winter crops sowing.

In croplands, there might be a carbon input through organic fertilization (OF). For the Lamasquère site, the farmer provided us with the quantities he added of organic fertilizers during the wheat cultivation in 2007 and 2009. For the Auradé site, there is no organic fertilization, however we include a OF term corresponding to the carbon content of the sowed seeds (also considered at Lamasquère). This term was set to 6.25 gC m^{-2}.$year^{-1}$, corresponding to an average value for the study area.

At regional scale, it is not possible to obtain an estimate of the OF input from remote sensing products. The solution therefore remains in accessing regional/national inventories and database to have average estimates of the OF term. Over our area of study very few plots receive OF because most of the farms are cereal production farms.

The carbon output term (Cexp) corresponds to the amount of carbon of the plant biomass that is exported from the plot ecosystem at harvest. The Auradé site is a cereal production farm, where only the wheat grains are exported; thus, Cexp corresponds to the final grain yield. On the other hand, the Lamasquère site is a livestock breeding farm, where most of the aboveground biomass is exported, for cattle bedding and feeding. In this case, the Cexp term is computed from the grain yield plus the exported straws [Eq. (4.34)]. The SAFY-CO_2 model is able to simulate

the total aboveground biomass and the grain yield. It is therefore necessary to estimate the exported straw biomass from these two outputs. Since for the Lamasquère site we have all this information available (grain yield, total biomass and exported straw biomass), we determine a coefficient *sc* relating the exported amount of straw by the total straw biomass (defined as the difference between total biomass and grain yield, Eq.(4.33)). Besides, the *sc* coefficient could also be set using the data obtained through the 2011 field campaign; which can be useful for running the model at plots where this information in not available.

At larger scales, it remains a great challenge to differentiate the crop fields for which the straw is exported from those for which the straw is left on site. Studies were performed for mapping crop residue (Daughtry and Hunt Jr, 2008). Still it remains difficult to quantify the amount of residues on different fields at regional scale. As a consequence, the regional estimates of the Cexp term present high incertitude, which is then transferred on the calculation of the crop carbon budget. In future researches, we expect that the combination of multispectral (optical and radar) and hyperspectral data at high temporal and spatial resolutions will provide the means to overcome these limitations.

Finally, for the computation of the annual net ecosystem carbon budget (NECB), we sum the contribution of each of the described terms [Eq.(4.35)]. As in Hollinger et al., (2005) and Béziat et al., (2009), the micrometeorological convention is adopted, with NEP negative when the ecosystem is fixing carbon and positive when it is losing carbon. Cexp is considered an instant release of carbon to the atmosphere; thus it is positive. The OF is an input of carbon into the field, thus it is negative. The sign of the annual NECB indicates if the ecosystem is a carbon sink (NECB negative) or a carbon source (NECB positive).

$$sc = \frac{straw_{Exp}}{straw_{total}} = \frac{straw_{Exp}}{DAM_{\max} - Yield} \tag{4.33}$$

$$C\exp = Yield \quad \text{if only grain exported;}$$
$$C\exp = Yield + (DAM_{\max} - Yield) \times sc \quad \text{if grain + straw exported;} \tag{4.34}$$

$$NECB = NEP + OF + C\exp \tag{4.35}$$

4.3. The SAFYE-CO_2 model

This section describes the coupling of the SAFY-CO_2 model with a water balance module, based on the FAO-56 method. The resulting model is named SAFYE-CO_2, which is able of simulating the plant development and growth, the components of the CO_2 fluxes (as GPP, Reco, NEE…) but also the crop evapotranspiration and the soil water content. Initially the main principles and equations of the SAFYE-CO_2 model are presented. Next, the calibration and validation approaches using *in-situ* data are discussed.

4.3.1. The SAFYE-CO_2 description

The SAFYE-CO_2 model is essentially composed of two modules: the SAFY-CO_2 module (described in section 4.2) and the water budget module, which describes the water transfers in the soil-vegetation-atmosphere continuum. The 'plant' module is driven by the remotely-sensed GAI (as described for SAFY and SAFY-CO_2) and the water budget module is based on the SAFYE model (Duchemin et al., 2005), which is a modified version of the FAO-56 method. In comparison to the existing SAFYE model, the SAFYE-CO_2 model, developed through this thesis, is also capable of simulating the CO_2 fluxes, in addition to the standard outcomes: crop production, evapotranspiration and soil water content.

The plant module remains essentially the same as SAFY-CO_2, but here it also allows evaluating the crop water requirements; *i.e.*, the amount of water needed by the soil-vegetation system for evapotranspiration. The 'water module' allows an assessment of the water availability and it characterizes the crop water stress (Duchemin et al., 2005). In our modeling approach, the evapotranspiration is estimated base on the dual-crop coefficient (introduced in section 1.1.2), separating the evaporation and transpiration contributions. In the water module calculations, only the vertical flows are considered, neglecting the runoff horizontal flows. The soil is represented by two horizontal 'infinite' layers: the surface (top) and the deep layers. During the vegetative period, the soil deep layer is divided in two, for generating a root layer just between the superficial and the deep layers (as illustrated in Figure 4.17). Therefore, three soil layers are involved in soil water transfer mechanisms: (1) a superficial layer that works as an interface with the atmosphere, and supplies water to the deeper soil layers; (2) an intermediate layer that extends with the root zone; and (3) a deep layer that produces base flow.

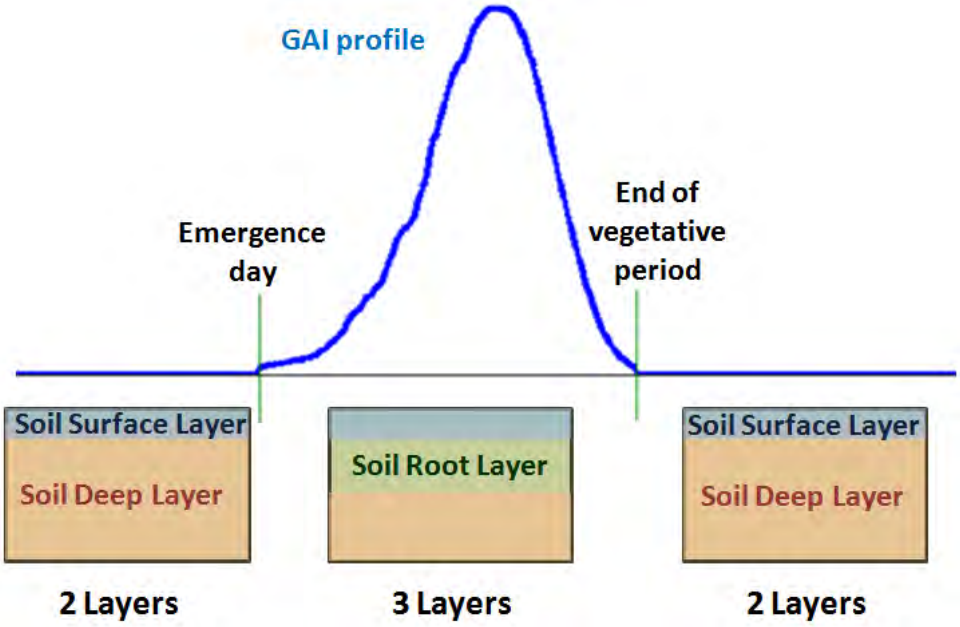


Figure 4.17- Illustration of the soil layers characterized in the SAFY-CO_2 model. The surface and deep layers are present during all the simulation phases. The soil root layer is only present through the vegetative period.

A diagrammatic representation of the model's processes is presented in Figure 4.18. In SAFYE-CO_2, the water balance is calculated at daily step, over a period of time chosen by the user. The water balance module is composed of three main steps: a) assessment of the evolution of the roots development and update of the soil water content and of the available water in the root and deep soils reservoirs; b) computation of the gravity fluxes between the surface, root and deep layers; c) computation of the diffusive fluxes. These processes are detailed in the sequence. Most of the components of the soil water budget are expressed here in terms of water depth (mm).

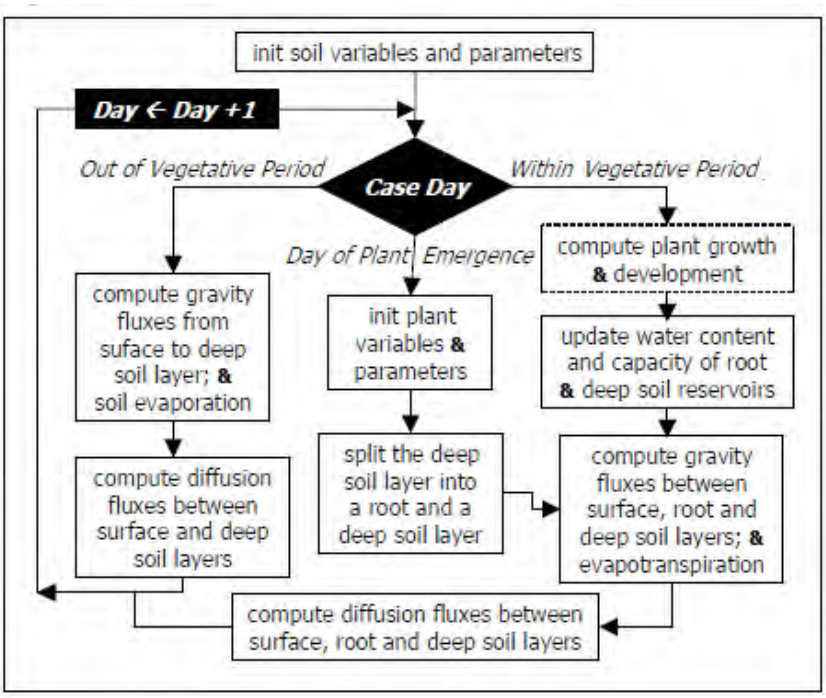


Figure 4.18- SAFYE-CO_2 processes flow chart [extracted from Duchemin et al., (2005)].

4.3.1.1. Description of the water module

The soil water availability refers to the capacity of a soil to retain some water available to plants and evaporation. The water capacities of the evaporative layer (TEW), of the root compartment (TAW) and of the deep compartment (TDW) are given by the equations (4.36), (4.37) and (4.38); besides, they depend upon the effective depth of the surface, root and deep soil layers (Z_e, Z_r and Z_d, respectively).

After heavy rainfall or irrigation, the soil will drain until field capacity is reached. Field capacity (θ_{fc}) is the amount of water that a well-drained soil should hold against gravitational forces, or the amount of water remaining when downward drainage has markedly decreased (Allen et al., 1998). In the absence of water supply, the water content in the root zone decreases as a result of water uptake by the crop. As water uptake progresses, the remaining water is held to the soil particles with greater force, lowering its potential energy and making it more difficult for the plant to extract it. The water uptake becomes zero when the wilting point is reached. Wilting point (θ_{wc}) is the water content at which plants will permanently wilt. As the water content above field capacity cannot be held against the forces of gravity and will drain and as the water content below wilting point cannot be extracted by plant roots, the total available water in each soil layer is the difference between the water content at field capacity and the wilting point (in $m^3.m^{-3}$), weighted by the effective layer depth (in mm).

The water currently contained in the three soil layer is represented by the variables named AEW, AAW, ADW, for the surface, root and deep soil layers, respectively.

The update of the current water content of to superficial layer (AEW, Eq.(4.39)) depends upon the daily effective precipitation and irrigation (PE), the daily evaporation (E) and transpiration (T.fts). The fraction of water that is taken from the superficial layer for transpiration (fts, Eq. (4.40)) refers to the ratio between the actual water content of the superficial layer (AEW) and the total actual water content of the two first reservoirs (superficial and root). The effective precipitation and irrigation (PE, Eq. (4.41)) are a fraction of the daily cumulated rain (P) and irrigation (I), which are function of the fraction of soil surface that is effectively exposed to evaporative energy (1-Fcover), and that is limited by the evaporative demand (represented by the reference evapotranspiration ET_0).

According to the dual-crop coefficient theory (Allen et al., 1998), the soil evaporation (E, Eq. (4.42)) is function of ET_0 and of a soil water evaporation coefficient (K_e). This coefficient is function of the surface that is not covered by the vegetation (1-Fcover) and of the superficial relative humidity (He). He [Eq. (4.52)] is defined as the ratio between the available evaporable water (AEW) and the total evaporable water (TEW). Besides, during dry extreme episodes, the evaporation is limited by a β function, which approximates a bilinear behavior, as established in Allen et al., (1998).

The transpiration (T, Eq. (4.43)) is determined by the product of the reference evapotranspiration ET_0 by a basal crop coefficient (Kcb). Moreover, the plant transpiration is reduced by the water stress coefficient (Ks), when stress conditions are met. The water stress function is described by Allen et al., (1998). Ks is a bilinear function, ranging between 0 (when the vegetation is completely stressed) and 1 (no water stress), and depending upon a critical relative humidity parameter (Hcrit) and the soil maximal relative humidity (He or Hr).

Through the vegetative period, the root layer is limited by the root effective depth (Z_r, Eq. (4.45), expressed in m). The Z_r evolves during the season as a function of the root growth speed (Vpr), which is modulated by the air temperature (Ta) and the water stress coefficient (Ks).

The update of the available water content of the root layer (AAW, Eq.(4.39) (4.46)) depends upon the fraction of the transpiration not extracted of the superficial layer (1-fts) and on the excess of water drained of the superficial zone (the depletion term DP_e). In this way, for each compartment, the total available water is split into the water currently contained in the soil (AEW, AAW, ADW) and the complementary empty space or depletion (DP_e, DP_r, DP_d, Eqs. (4.49),(4.50),(4.51)). When all compartments are full, the excess water flows out of

the system as deep drainage. The deep soil layer refers to a water storage reservoir. The deep available water (ADW, Eq. (4.47) or (4.48)) is thus connected to the root or the superficial layer, depending on the presence of vegetation or not. Finally, the deep water excess (DP_d) is definitely lost for the soil-plant system.

Water diffusion due to capillarity water movement, either upwards or downwards, is modeled between the superficial, root and deep compartments (or only superficial and deep out of vegetative periods) on the basis of their relative water content. It means the diffusive flux between adjacent layers is based on the moisture gradient between the two layers. The diffusive fluxes (ϕ_{xy}, Eqs. (4.53)(4.54)(4.55)) are calculated based on the method proposed by Devonec and Barros (2002),and used by Duchemin et al., (2005), for which the relative moisture gradient (θ_x-θ_y) is normalized by the soil field capacity (θ_{fc}).

$$TEW = \left(\theta_{fc} - \frac{\theta_{wp}}{2}\right) \cdot Ze.1000 \quad (4.36)$$

$$TAW = \left(\theta_{fc} - \theta_{wp}\right) \cdot Zr.1000 \quad (4.37)$$

$$TDW = \left(\theta_{fc} - \theta_{wp}\right) \cdot Zd.1000 \quad (4.38)$$

$$AEW\ (i) = AEW\ (i-1) + PE\ (i) - E\ (i) - T\ (i) \cdot fts \quad (4.39)$$

$$fts = \frac{AEW}{AEW + AAW} \quad (4.40)$$

$$PE = \max\left\{\left(P + I\right) - \min\left\{\left(P + I\right) \times \left(1 - Fcover\right), ET_0\right\}, 0\right\} \quad (4.41)$$

$$E = (1 - Fcover).ET_0.\underbrace{(1 - (1 - He)^{\beta})}_{Ke} \quad (4.42)$$

$$T = Kcb.Ks.ET_0 \quad (4.43)$$

$$Ks = \min\left\{1 - \frac{HRS - Hcrit}{-Hcrit}, 1\right\} \quad (4.44)$$

$$HRS = \max\{He, Hr\}$$

$$Z_r(j) = \max\left\{Z_r(j-1) + Ta(j) \times Ks(j) \times Vpr\ , Z_d - Z_e\right\} \quad (4.45)$$

$$AAW\ (i) = AAW\ (i-1) + DP_e - T(i) \times (1 - fts) \quad (4.46)$$

$$ADW\ (i) = ADW\ (i-1) + DP_r \quad (4.47)$$

$$ADW\ (i) = ADW\ (i-1) + DP_e \quad (4.48)$$

$$DP_e = \max\left\{TEW - AEW\ , 0\right\} \quad (4.49)$$

$$DP_r = \max\left\{TAW - AAW\ , 0\right\} \quad (4.50)$$

$$DP_d = \max\left\{TDW - ADW\ , 0\right\} \quad (4.51)$$

$$He = \frac{AEW}{TEW}; Hr = \frac{AAW}{TAW}; Hd = \frac{ADW}{TDW} \quad (4.52)$$

$$\Phi_{13} = K_{dif} \times \left(\frac{\theta_1 - \theta_3}{\theta_{fc}}\right)^{E_{dif}} \quad (4.53)$$

$$\Phi_{12} = K_{dif} \times \left(\frac{\theta_1 - \theta_2}{\theta_{fc}}\right)^{E_{dif}} \quad (4.54)$$

$$\Phi_{23} = K_{dif} \times \left(\frac{\theta_2 - \theta_3}{\theta_{fc}}\right)^{E_{dif}} \quad (4.55)$$

4.3.1.2. SAFYE-CO_2: Coupling SAFY-CO_2 with the water module

SAFYE-CO_2 is a model resulting from the coupling between the presented SAFY-CO_2 model (Section 4.2), representing the vegetation module, with the water balance module, presented in the previous paragraph. This coupling is achieved by means of three main variables (at daily step): the GAI, the water stress (Ks) and the superficial soil moisture (He).

The water stress Ks calculated by the water module is taken into account in the photosynthesis estimation [Eq. (4.56)], limiting the GPP when water stress conditions are met. The photosynthesis estimation could already be impacted by temperature stresses (by means of the *F(Ta)* function); but all the others stresses (water, nitrogen, …) were "included", in some way, in the light-use efficiency term (fELUE). The introduction of the water stress function allows better accounting of the effects of water stress episodes over the crop production; the Ks is therefore decoupled from the fELUE.

In addition, the GAI variable, simulated by the vegetation modules, is primarily used for the the estimation of the vegetation cover fraction (Fcover, Eq. (4.57)). Fcover is estimated according to the method presented by Welles and Norman, (1991), for which the Fcover is a function of the plant green area index (GAI), of the total plant area index (AI), and of two empirical coefficients(K_{cov} and E_{cov}). However, for the sake of simplicity, and given that the model is not capable of specifically simulating the total plant area, we consider the term GAI equals to AI, which makes the term $\left[GAI + AI/2 \times AI\right]$ negligible. The fcover term participates to the estimation of the effective intercepted rain and irrigation (PE) and to the evaporation estimates. The (vegetation module) GAI is also used for the crop transpiration estimates, by means of the crop basal coefficient (Kcb). Therefore, the Kcb [Eq. (4.58)] is calculated according to . It is a function of the GAI and of two other coefficients: the Kcb_{max} and the E_{trp}.

$$GPP = Rg \times \varepsilon_c \times FAPAR \times F_T(Ta) \times fELUE \times sR10 \times Ks \quad (4.56)$$

$$Fcover = K_{cov} \times (1 - e^{-E_{cov} \times GAI}) \times \left[\frac{GAI + AI}{2 \times AI}\right] \quad (4.57)$$

$$Kcb = Kcb_{max} \times (1 - e^{-E_{trp} \times GAI}) \quad (4.58)$$

The last difference in the SAFYE-CO_2 model relative to SAFY-CO_2 consists in the heterotrophic respiration (Rh) estimation. For SAFY-CO_2, Rh was calculated using an

empirical exponential equation depending exclusively on the soil temperature (Ts). With the inclusion of the water module, SAFYE-CO_2 is capable of simulating the evolution of the superficial soil humidity. Therefore, a new Rh function is implemented, depending on both soil temperature and soil moisture estimates [Eq. (4.59)]. This equation's formalism is based on that present in the CENTURY model (Parton et al., 1987), which was used (and calibrated) by Delogu (2013). The parameters *a* and *b* were calibrated using the NEE and SWC data from the flux towers (Auradé and Lamasquère), from 2006 to 2010, over periods of bare soil. The parameter *b* is linked with the Q_{10}, set to be equal to 2.3. The parameter *a* corresponds to the 'reference *Rh'.* A different value was set for each site. Table 4.5 presents the adjusted parameters (and associated statistical indicators R and RMSE) found for both sites. For the other fields within our study area extent, a mean value of Rh_{base}=0.49 is used, and Q_{10} remains fixed to 2.3.

$$Rh = a \times e^{b \times T_S} \times \frac{1}{1 + 30 \times e^{-8.5 \frac{\theta s - \theta wp}{\theta fc - \theta wp}}} \qquad (4.59)$$

$$\text{with} \quad \theta_S = He \ (\text{superficial soil moisture})$$

Table 4.6- Heterotrophic respiration (Rh) parameters (extracted from Delogu, 2013). The Rh_{base} and RMSE units are gC m^{-2}.day^{-1}.

	Auradé	**Lamasquère**
Rh **base**	0.45	0.52
Q_{10}	2.3	2.3
RMSE *(StD)*	0.64	0.20
R *(StD)*	0.68	0.52

4.3.2. Model's parameterization

The SAFYE-CO_2 *vegetation module* includes the same parameters as those of the SAFY-CO_2 model (Section 4.2). These parameters are essentially related to the plant phenologycal and physiological processes, and allow estimating the plant development, production and associated CO_2 fluxes for given climatic conditions. Therefore, the methodology for setting these parameters is maintained, as described for the SAFY-CO_2, excepting for one single parameter: the ELUE (part of the *fELUE* function, Eq. 4.35). Since for SAFYE-CO_2 the light-use efficiency ELUE does not comprise anymore the plant water stress, this parameter needs to be re-calibrated.

The SAFYE-CO_2 *water module* presents 13 new parameters, which are listed in Table 4.7. These parameters can be related to the soil processes or to the vegetation processes. Besides, they can be fixed in different ways: i) according to values found in the literature; ii) according to *in-situ* measurements/observation; iii) by means of a calibration process.

Table 4.7- List of the main parameters of the SAFYE-CO_2 water module: notation, units and values fixed for winter wheat crop.

	Description	Notation	Unit	Value	Method
Vegetation	Root growth rate	Vpr	m.°C	0.0072	Calibration
	Cover fraction coefficient	K_{cov}	-	0.91	*In-situ* measurem.
	Exponent of the cover fraction	E_{cov}	-	0.31	
	Maximal transpiration coefficient	Kcb_{max}	-	0.98	Calibration
	Exponent of the transpiration	E_{trp}	-	0.36	Calibration
	Transpiration reduction coefficient	*Hcrit*	-	0.55	Literature
Soil	Field capacity	θ_{fc}	$m^3.m^{-3}$	-	Soil maps +*in-situ* param.
	Wilting point	θ_{wp}	$m^3.m^{-3}$	-	
	Soil deep layer maximal depth	Z_d	m	[0.2-0.15]	Soil maps
	Soil surface layer depth	Z_e	m	0.05	Literature
	Evaporation reduction coefficient	β	-	0.48	Calibration
	Diffusive flux coefficient	K_{dif}	$mm.day^{-1}$	1.06	Calibration
	Diffusive flux exponent	E_{dif}	-	6.38	Calibration

Table 4.8- Soil hydraulic characteristics of the Auradé and Lamasquère sites.

Parameter	Auradé	Lamasquère
Field capacity ($\theta_{fc,}$ $m^3.m^{-3}$)	0.35	0.46
Wilting point ($\theta_{wp,}$ $m^3.m^{-3}$)	0.17	0.26
Soil deep layer maximal depth ($Z_{d,}$ m)	1.5	2.0

Parameters from the literature and from *in-situ* measurements

This category of parameters can be divided into the soil (θ_{fc}, θ_{wp}, Z_e, Z_d) and vegetation (K_{cov}, E_{cov} and *Hcrit*) related parameters.

The soil intrinsic properties are characterized by means of the following parameters : the soil field capacity (θ_{fc}) and wilting point (θ_{wp}), the soil depth (Z_d) and the soil superficial layer depth (Z_e).

The Z_e determines the water volume that can be assigned to evaporation. It was set to 5 cm, according to the value found in models of the literature, like ISBA (Noilhan and Planton, 1989) and ICARE (Gentine et al., 2007).

The three other parameters (θ_{fc}, θ_{wp}, Z_d) are field-specific, varying spatially and depending upon the soil type. For the experimental sites (Auradé and Lamasquère) these values were obtained from *in-situ* measurements and soil analysis. The values are listed in Table 4.8. The Lamasquère site is characterized by a clay and deep soil (around 2m), presenting a large capacity of water storage. The Auradé soil presents a high spatial variability and a lower average depth (between 0.6 and 1.5m). Besides, its texture properties (θ_{fc}, θ_{wp}) yield to an inferior capacity of total water content.

For the other fields within our study area, the given parameters are estimated from the regional soil maps (Section 2.5.1). The parameters θ_{fc} and θ_{wp} are extrapolated from the soil clay content maps, according to a relationship established over a number of field measurements next to the Lamasquère site and analyzed in laboratory. The obtained relationship relating soil humidity at field capacity or at wilting point and clay content is displayed in Figure 4.19.

The vegetation parameters related to the vegetation cover fraction estimation (fcover) are the K_{cov} and E_{cov} coefficients [Eq. (4.57)]. They were set by fitting a exponential law relating the GAI and Fcover, both obtained from hemispherical photographs taken over wheat crops (at developed growth stage, when GAI=AI) and processed using the CAN-EYE software. Figure 4.20 illustrates the exponential relationship between the two measured variables, GAI and Fcover. Values of K_{cov}=0.91 and E_{cov}=0.31 were found by minimizing the root mean square error between the established law and the measured data. These values are relatively close to those reported by Duchemin et al., (2006) for irrigated winter wheat in Morocco (K_{cov} =0.95 and E_{cov}=0.53).

Finally, the transpiration reduction coefficient (*Hcrit*) that represents the critical humidity for activating the plant water stress for transpiration is set constant for all the winter wheat fields. *Hcrit* is fixed according to the value reported by Allen et al., (1998), for which *Hcrit*=0.55.

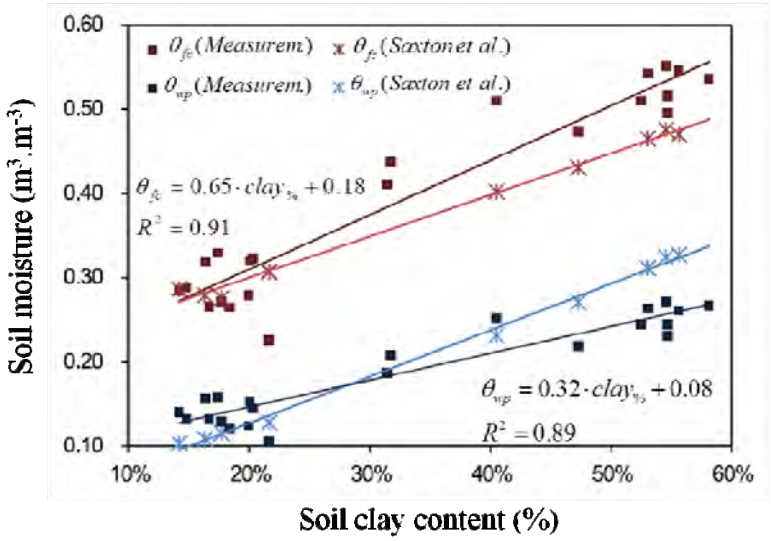


Figure 4.19- Relationship between soil clay content (%) and soil moisture at field capacity θ_{fc} and at wilting point θ_{wp} (both in $m^3.m^{-3}$). The square symbols represent the data derived from in-situ measurements at 4 different soil depths. In red the measurements relative to field capacity and in blue those relative to wilting point. For comparison, the cross symbols represent the critical soil moistures (θ_{fc} ,θ_{wp}) obtained when applying the relationship established by Saxton et al., (1986) [Image extracted from Claverie (2012)].

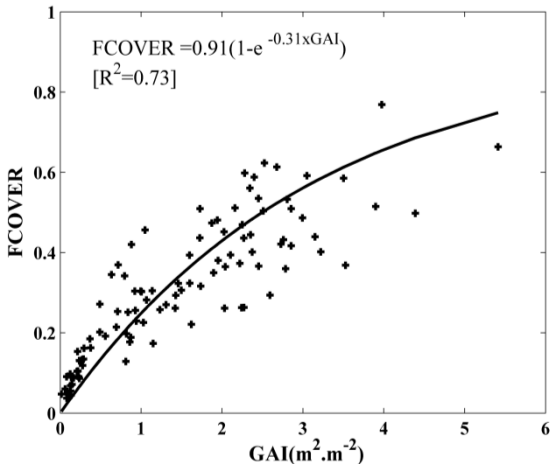


Figure 4.20- Relationship established between GAI and Fcover biophysical variables for winter wheat. Both variables were estimated from hemispherical photographs, processed with the CanEye software.

Calibrated Parameters

The SAFYE-CO_2 water module parameters that were fixed through a calibration approach are: the parameter related to the plant root growth rate (*Vpr)*, the parameters related to the transpiration estimates (Kcb_{max} and E_{trp}), the parameter

related to the soil evaporation (β) and finally the parameter concerning the diffusive fluxes (K_{dif} and E_{dif}). They were calibrated in the chronological order for which they are cited in the text.

The parameters calibration methodology is based upon evapotranspiration (ETR) and soil water content (SWC) data measured over the experimental sites. Five site-years of data are available: AUR2006, LAM2007, AUR2008, LAM2009 and AUR2010. The calibration approach consists in using the ETR and SWC data of two site-years for calibrating the six cited parameters (*Vpr*, Kcb_{max}, E_{trp}, β, K_{dif} and E_{dif}) and the data of the resting three site-years for validation. In order to choose the best pair of years for calibrating the parameters, all the combinations are tested (a total of 10 possibilities). The cost function used in the parameter's optimization is calculated from the sum of relative errors (RRMSE) between the model estimated and the measured ETR (or SWC) for the two selected years. The optimization process searches the set of parameters that minimizes the dual cost function.

The simplex method, with restricted parameters values boundaries, is chosen as optimization approach. The optimization of the parameters is performed through four sequential steps: 1) Vpr; 2) Kcb_{max} and E_{trp}; 3) β; 4) K_{dif} and E_{dif}. These six parameters are fixed to values found in the literature, as a start point value for the optimization.

The first step aims at calibrating the root growth rate parameter (*Vpr*). The initial value of *Vpr* is set as 0.0012 m.°C, according to the STICS model parameterization for the winter wheat crop. For setting the interval within the optimal *Vpr* is searched, we set a threshold of ±40% around the start value; *i.e* the optimal *Vpr* must be included in the range of [0.00072-.0017]. Through this step, the resting parameters remain constant.

The second step is dedicated to the optimization of the parameters related to transpiration: Kcb_{max} and $E_{trp.}$ (which define Kcb, Eq.(4.58)). The initial values for these two parameters were set according to *in-situ* measurements. The basal crop coefficient Kcb can be defined as the ratio of ETR to ET_0 when the soil surface layer is dry but when the average soil water content of the rootzone is adequate to sustain full plant transpiration. Figure 4.21 illustrates the relationship between Kcb and GAI, established from ETR fluxes measured by the flux towers and from destructive measurements of GAI performed at the Auradé and Lamasquère sites. The ETR data were filtered according to two criteria: i) no water supply (precipitation) was observed during a period of 4 days, in order to reduce the influence of soil evaporation; ii) no observed vegetation stress, based on a minimum threshold of soil root water content superior to 0.65 of the field capacity $\theta_{fc.}$ The GAI data were interpolated to fit the Kcb data. From this relationship, we obtain Kcb_{max} =1.07 and E_{trp}= 1.25. These values are set as starting points for the optimization process.

Besides, for Kcb_{max} we set a minimum–maximum interval of [0.6 -1.15] according to Allen et al., (1998) and to field expertise. The optimization interval for E_{trp} is empirically set to [0.2 - 1.72].

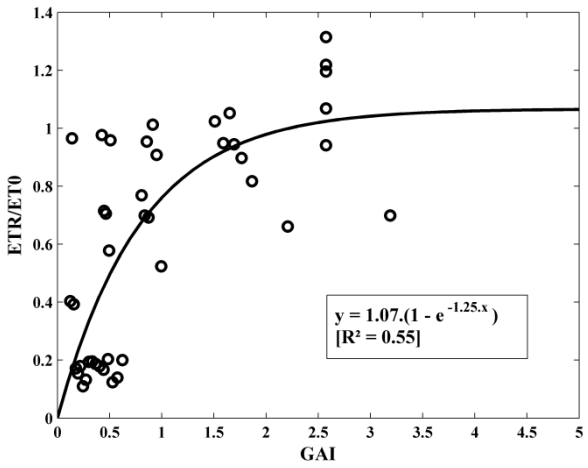


Figure 4.21- GAI-Kcb relationship. The crop coefficients Kcb are computed as the ratio of actual and reference evapotranspiration (ETR/ET_0, y-axis). The obtained coefficients of the exponential relationship, linking Kcb and GAI, are displayed.

The third step of the calibration procedure concerns the β parameter, related to the evaporation stress function estimate. Therefore, for this step the cost function is based on the integration of the errors (RRMSE) relative to ETR and also to superficial SWC between measurements and estimates. The initial value of β is set to an average value found by Claverie (2012) through a calibration procedure using data from Lamasquère (2006) and Auradé (2007). The interval for optimization was set to β within [0.3-1.5], according to empirical tests performed over the whole data set (from 2006 to 2010).

The fourth, and last, step of the calibration process aims at setting the parameter related to the diffusive fluxes estimation: K_{dif} and $E_{dif.}$ The calibration of these parameters is important for the soil water content estimations but also for the simulation of the evapotranspiration. These parameters can be very different, depending on the soil properties. Devonec and Barros (2002) have found optimal values of K_{dif} =0.27 and $E_{dif.}$=1; meanwhile Duchemin et al. (2005) found quite higher values: K_{dif}=175.15 and $E_{dif.}$=3.18. For our study area, Claverie (2012) has determined K_{dif}=3.5 and $E_{dif.}$=1. We use these last values as starting points for the optimization process. Given the large extension of the reported values for setting K_{dif} and $E_{dif,,}$ we decide to state an extremely large interval for this optimization of [0-500] for both parameters (to give a high degree of freedom and avoid negative solutions).

Therefore, this 4-steps optimization procedure was applied for each of the ten possible combinations of two years used for the parameters calibration. Each set of

obtained calibrated parameters (*Vpr*, Kcb_{max}, E_{trp}, β, K_{dif} and E_{dif}) was then applied to the three years validation data sets. The resulting evapotranspiration and superficial soil water content estimates for three validation-years were compared against the *in-situ* measurements. The error (RMSE) and the Nash-criteria were calculated for each case, and are displayed in Table 4.9. According to these results, which are differentiated by a color code: from the worst, in red, to the best results, in green, we observe that the pair of years 2007 (LAM) and 2008 (AUR) yields to good results in terms of both ETR and top-layer SWC. As a consequence, we choose these two site-years as the calibration sets and the other site-years (2006, 2009, 2010) constitute the validation data set.

The set of calibrated parameters are: *Vpr*=0.0072, Kcb_{max}=0.98, E_{trp}=0.36, β=0.48, K_{dif}=1.06 and E_{dif}=6.38. They were fixed to these values for all the investigated fields within our study area (including the experimental sites), regardless the year (from 2006 to 2011).

Table 4.9- Performances of the parameters optimization. A set of five site-years was available: AUR2006, LAM2007, AUR2008, LAM2009, AUR2010. The first column indicates the years used for calibrating the parameters (*Vpr*, Kcb_{max}, E_{trp}, β, K_{dif} and E_{dif}). The performances (RMSE and Nash-efficiency Eff) were calculated using the three remaining years for the evapotranspiration (ETR) and the superficial soil water content (SWC). For each column (2 to 5) the values are classed from the worst (in red) to the best (in green). The performances for the pair of years selected as the best one (Y2007_2008) are highlighted.

Calibration Years	RMSE ETR	Eff ETR	RMSE SWC	Eff SWC
Y2006_2007	0.64	0.34	0.04	0.54
Y2006_2008	0.55	0.61	0.05	0.10
Y2006_2009	0.54	0.61	0.06	0.14
Y2006_2010	0.50	0.67	0.07	-0.24
Y2007_2008	**0.48**	**0.66**	**0.05**	**0.33**
Y2007_2009	0.50	0.59	0.05	-0.20
Y2007_2010	0.51	0.62	0.05	0.09
Y2008_2009	0.51	0.66	0.06	-0.01
Y2008_2010	0.55	0.63	0.07	-0.38
Y2009_2010	0.56	0.58	0.07	-0.36

5. Crop Modeling Results & Discussions

This chapter contains the main results obtained during this thesis. Initially, we present the results of a preliminary study for estimating yield by means of an empirical linear regression model from NDVI data. Then, following the same order as Chapter 4, we start by presenting the SAFY model results; at first, GAI and biomass estimates, and then yield estimates by using different methods. Next, the performances of the SAFY-CO_2 model are evaluated, in terms of GAI, biomass and yield, but also in terms of the net CO_2 fluxes components (GPP, Reco, NEE) and net ecosystem carbon budgets. Finally, the results for the SAFYE-CO_2 model are presented, including the same outputs as the SAFY-CO_2 model, but also the variables related to the water balance: evapotranspiration (ETR) and soil water content (SWC). The scheme of the models and their outputs is reminded in Figure 5.1. The results of the different versions of the models (SAFY, SAFY-CO_2, SAFYE-CO_2) are compared in order to evaluate the possible improvements (or degradations) of their performances in terms of GAI, biomass and yield estimates. This inter-comparison is essential because the new versions, which include estimates of CO_2 and ETR fluxes, should not degrade the performances of the original SAFY model but should also allow the assessment of new variables of interest for studies linking agronomical and environmental issues.

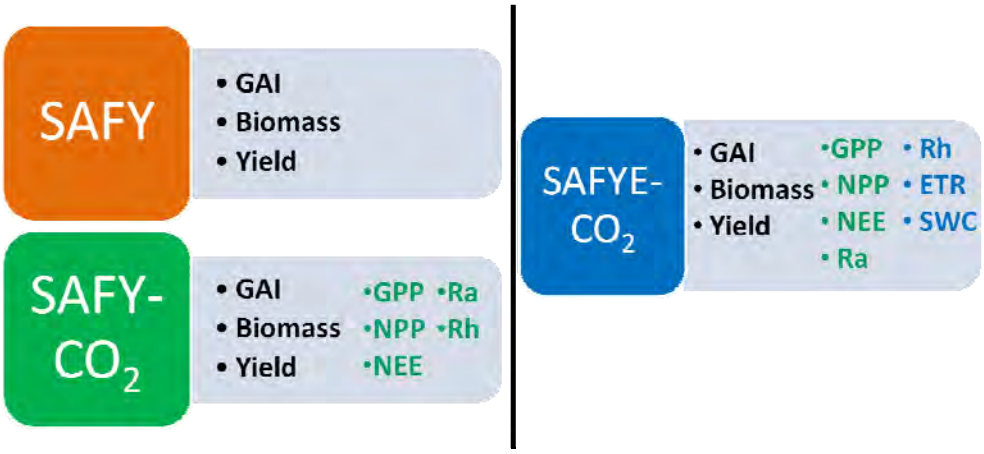


Figure 5.1- Scheme of the investigated crop models with their associated main estimated variables.

5.1. Empirical approach: Yield estimates from NDVI

In this section we present the results for establishing an empirical linear regression model linking yield and NDVI data, using data from the '2011 field campaign' (destructive measurements) and 'yield monitors' data sets (see Section 4.1.5.1). The established relationship is then tested over a third data set: farmers' surveys data set. For this study, the 'experimental sites' data set is not used, since it

includes only 6 points and since these data were provided by the farmers; thus the farmers' surveys data set can represent this category of yield data.

a) Single- date NDVI relationship

The first approach considered a single NDVI date to establish the linear relationship with yield data. For each one of the data sets, a regression was performed for each date of the vegetative period (from sowing until harvest). Figure 5.2 shows the Taylor diagrams illustrating the performance of all models (one for each date) for the 2011 campaign (left) and the yield monitors (right) data sets. Taylor diagrams provide a statistical tool for visualization of the correspondence between multiple model simulations and an observed variable in terms of their correlation, their root-mean-square difference and the ratio of their variances. Each model performance is visualized by a point on the polar plot. The radial distance from the origin is proportional to the standard deviation of the observations. The centered root-mean-square difference (SEE) between the test and the reference is proportional to their distance apart. The correlation between the two sets is given by the azimuthal position of the point (Taylor, 2001). In our diagrams, the plot colors represent the relative root mean square errors (rRMSE).

We note that the choice of a single data for establishing a model linking NDVI and yield has a great impact on yield estimations, given the large distribution of the plots on the Taylor diagrams (Figure 5.2). Besides, we observe the correlation coefficients (R) ranges from 0 up to 0.9 for the first data set, and from 0.4 up to 0.98 for the second one. The relative errors (rRMSE) vary between 0.08 and 0.3, according to the selected NDVI date.

We found that the best linear models for estimating yield are the ones using NDVI values of the May 31st for the '2011 field campaign' data set, and the of the 15th of June for the 'yield monitors' data set. The Figure 5.3 shows the relationships between observed yield and best NDVI date for both data sets. Even if the 'best dates' are different for the two yield datasets, we observe also good performances (R^2=0.95) for the 'yield monitors' data set when using the May 31st NDVI value for the regression model

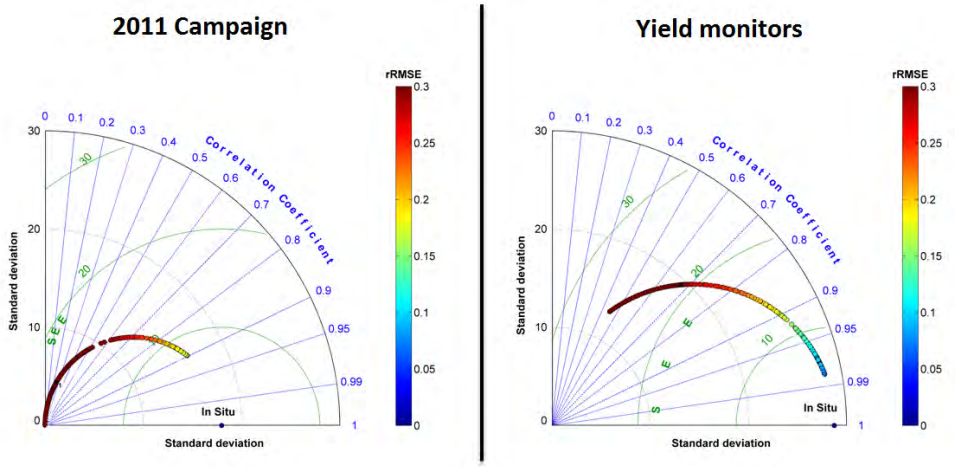


Figure 5.2- Taylor diagrams illustrating the performances of the regression models linking NDVI and yield. For each date of the vegetative cycle (thus each NDVI value) a model was established. On the left, statistics were calculated over 16 fields, part of the '2011 Campaign' data set. On the right, statistics were calculated over 7 fields of the 'yield monitors' data set.

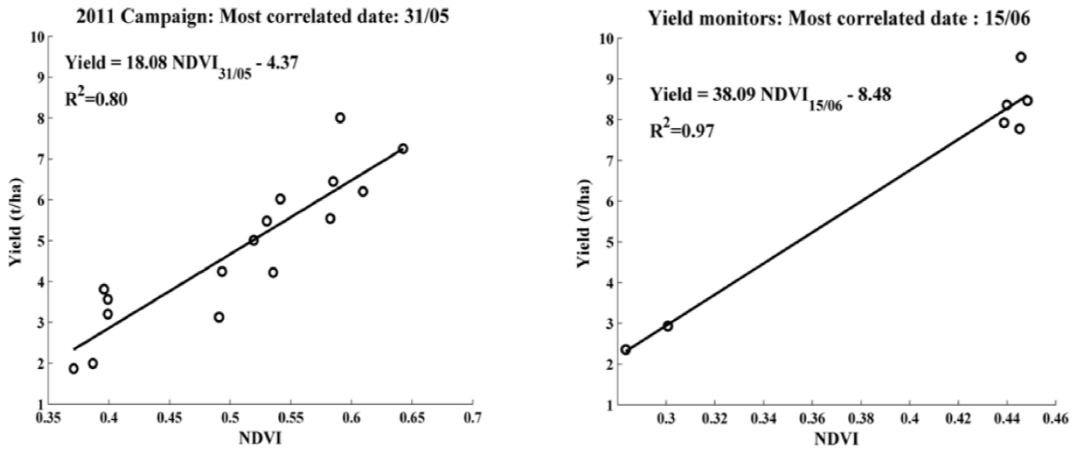


Figure 5.3- Relationship between observed yield and NDVI of the date the most correlated with the yield data. On the left, the results for the '2011 field campaign' data set, for which the best solution uses NDVI of May 31st. On the right, the results for the 'yield monitors' data set, for which the best solution uses NDVI of June 15th.

Next step consisted in merging the two data sets into one, so a wider data set, with an extensive dynamics, was available. Figure 5.4 shows the linear model, linking yield and NDVI, established for the combined data set using the May 31st NDVI values. The model's equation is displayed in Eq.(5.1). As it can be seen in the graphic, the joint model (black line) is similar to the models obtained separately for each dataset (displayed in red and blue), which indicates that the two datasets are in agreement , even if they come from different sources.

$$Yield = 19.82\ NDVI_{31/05} - 5.20 \qquad (5.1)$$

At last, we applied this joint model over a third yield data set, the farmers' surveys data set, in order to check if the established relationship could be applied over other fields and years. For this purpose, we used data from 2006, 2007, 2010 and 2011, the years presenting complete and gap-free satellite images series.

Figure 5.5 displays the yield estimated with the equation (5.1) using the 31st May NDVI. We observe there is a clear inter-annual trend, since the results are quite different depending on the year. An overall dispersion is remarked and an overestimation tendency as well. Nonetheless, a reasonable correlation (R around 0.63) is obtained independently of the year. The data collected in 2011 (yellow crosses) appear detached of those of other years. In addition, the yield estimates for 2011 present low errors, with RRMSE of 26.2%, while for the others years RRMSE are superior to 60%. This particular behavior may be due to the particular climatic conditions of 2011 (hot and dry spring), since the linear model (Eq. 5.1) was established using data collected exclusively in 2011 as well (from 2011 field campaign and yield monitors data sets).

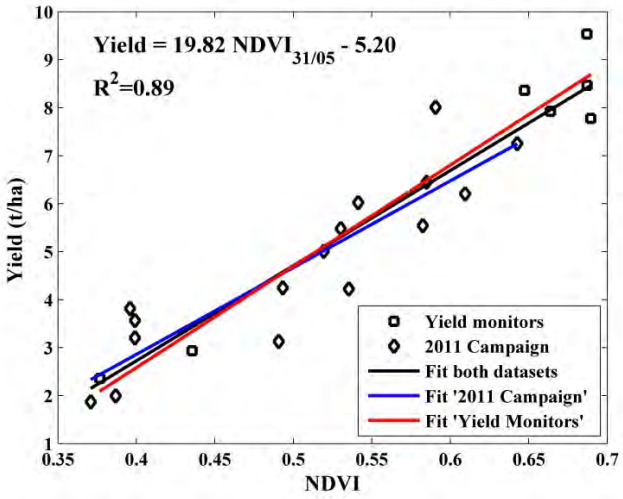


Figure 5.4- Relationship between observed grain yield (y-axis) and NDVI value of the May 31st. The '2011 field campaign' data are represented by the diamond symbol and data from 'yield monitors' data set by the square symbol. The linear regression model considering both data sets is indicated by the black line. The established relationship and correlation coefficient (R^2) are displayed on the left top corner. The blue and red lines represent the linear fit when considering only one of the data sets (the 2011 campaign, and yield monitors, respectively).

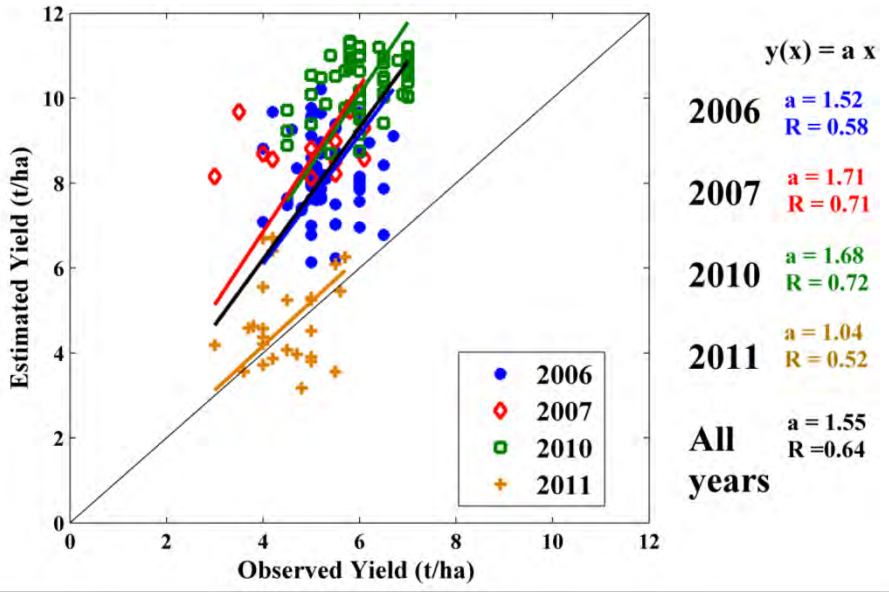


Figure 5.5- Comparison between observed (farmers surveys) and estimated yield, using a linear model linking yield and NDVI values. Data are represented by different colors and symbols according to the year, as shown in the legend. On the right the slopes of the estimated *vs* observed yield relationship for each year are displayed, as the correlation coefficients (R). The performances of the whole data set together (all years) are displayed in black.

Note that the range of yield data corresponding to the farmers' surveys data set is much smaller than with both other data sets. To further investigate this result, we firstly evaluated the distribution of the yield data gathered through farmers' surveys. The analysis of Figure 5.6 reveals that the median yield of the 2011 fields (4.35 $t.ha^{-1}$) is lower than those found for the others years (5.2, 5.25 and 6 $t.ha^{-1}$, in 2006, 2007 and 2010, respectively). In 2011, the reported yield values don't exceed 5.7 $t.ha^{-1}$, while for 2010 they often reach 7 $t.ha^{-1}$. This might be due to the exceptional dry character of the 2011's winter and spring, which probably affected the number and the filling of the grains. In addition, the analysis of the mean NDVI profiles (Figure 5.7), calculated over the available fields for each year, points out the inter-annual differences on the crop development. We observe a faster growth and decline of the vegetation in 2011 (sharp slope), while for the other years the growing phase is slower. Therefore the maximum development stage is shifted in time according to the year. It should also be noted that the mean NDVI values at May 31st are remarkably low for the 2011 crop season. As it can be seen in Figure 5.8, the median NDVI at May 31st is 0.67, 0.70, 0.80 and 0.49 for 2006, 2007, 2010 and 2011, respectively. This finding helps better understanding the overestimated yield results obtained for 2006, 2007 and 2010 (Figure 5.5) when applying the regression model that was calibrated with 2011 data uniquely. It means that the model is not suitable for predicting yield over climatically different years.

To corroborate the fact that the maximal seasonal NDVI did not appeared as a good solution for generating the linear relationship between yield and NDVI, we looked at the distribution of the maximum NDVI values of the fields included on the farmers' surveys data set (Figure 5.8). We note that no particular trend appears for 2011, having a median NDVImax similar to those of 2007 and 2010. Given the diverse yield distributions for those years, we conclude the maximum NDVI may not be a good predictor of yield. Since the grain filling usually begins after this stage (NDVImax), it seems important to consider later NDVI (after NDVImax) for predicting yield.

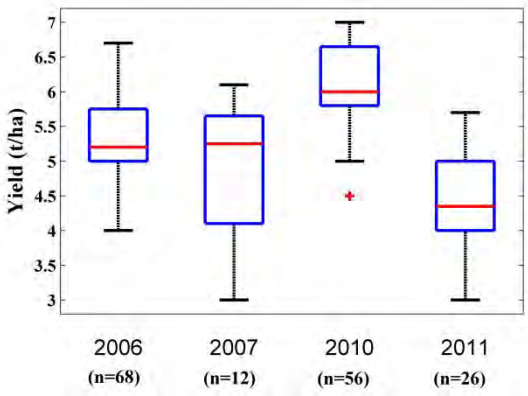


Figure 5.6- Box-plots illustrating the distribution of the yield data obtained by farmers' surveys in 2006, 2007, 2010 and 2011. The number of observations (n) used for generating each of these plots is displayed on the bottom between parenthesis.

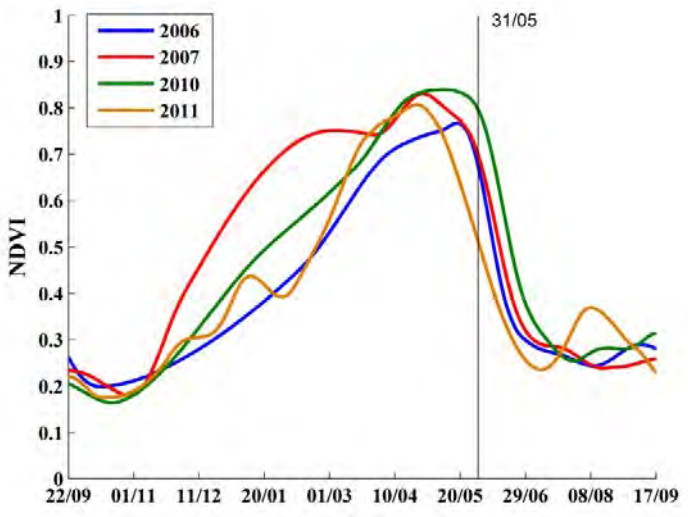


Figure 5.7- Mean NDVI profiles of the farmers' dataset fields of 2006 (blue), 2007 (red), 2010 (green) and 2011(yellow). The 31st of May is indicated by the vertical black line.

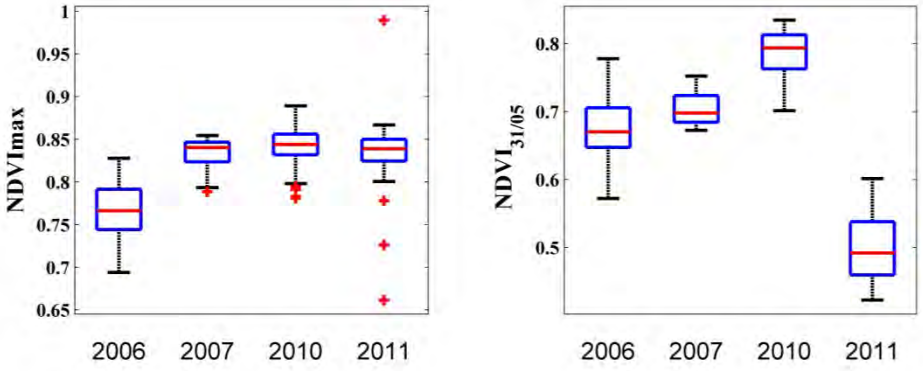


Figure 5.8- Box-plots illustrating the distribution of the values of the maximum NDVI (on the left); and NDVI values at the 31st of May (on the right), for the NDVI time series of the plots from 2006, 2007, 2010 and 2011 databases.

b) Integrated NDVI relationship

The second approach consisted in establishing an empirical linear regression model linking yield and NDVI accumulated through different periods. We tested integrating NDVI over 10 days at different stages (T1 to T13) and also over larger periods: as the whole vegetative period (T15) and from flowering stage until end of season (T14). A total of 15 tests were performed. Table 5.1 lists the periods of NDVI accumulation and obtained performances (R^2 and RRMSE) for both investigated data sets ('2011 field campaign' and 'yield monitors').

We observe that the methods related to the maximum NDVI value (T1, T2 and T3) yield to poor results for the '2011 campaign' data set. This indicates that the seasonal maximum value of NDVI is not the best yield predictor for our area of study. On the other hand, these methods present good performances over the 'yield monitors' data set (right column, Table 5.1). This may be partly because of the gap in the yield ground data dynamics. Since no data are available for the range of 3.0 to 7.0 $t.ha^{-1}$, it seems simpler establishing a regression model that encompasses all the observation points (Figure 5.9).

A closer look at the performances of the 'yield monitors' data set shows overall good results, with similar performances (R^2 around 0.9 and RRMSE around 11%), except for the early methods T4, T5 T6 and for the method T15, integrating the complete crop season. The method T14 presents the best performances (R^2= 0.95 and RRMSE=8.74%), which integrates NDVI values from the flowering day until the end of the season. Still, the method T11 has quite similar and good performances (R^2= 0.95 and RRMSE=8.59%), thus it can be considered as a 'best' method as well. The advantage of method T11 is to integrate NDVI over the last 10 days of May, while T14 integrates over a larger period.

Table 5.1- Performances of the different periods of accumulation of NDVI of winter wheat for establishing a linear relationship with observed yield production. The '2011 Field Campaign' and 'yield monitors' data sets are evaluated. Best solutions are colored in green.

		2011 Field Campaign		Yield monitors	
	Accumulation of NDVI	**R^2**	**RRMSE (%)**	**R^2**	**RRMSE (%)**
T1	$NDVI_{max}$ ±5 days	0.34	29.77	0.93	10.50
T2	$NDVI_{max}$ +10 days	0.35	29.64	0.94	9.68
T3	$NDVI_{max}$ - 10 days	0.40	28.39	0.91	11.97
T4	Middle 10-days of March	0.23	32.22	0.47	28.62
T5	Last 10-days of March	0.22	32.26	0.57	25.89
T6	First 10-days of April	0.23	32.18	0.79	18.14
T7	Middle 10-days of April	0.30	30.72	0.91	11.61
T8	Last 10-days of April	0.43	27.76	0.92	11.24
T9	First 10-days of May	0.65	21.79	0.91	11.51
T10	Middle 10-days of May	0.74	18.59	0.94	9.61
T11	Last 10-days of May	**0.79**	**16.89**	**0.95**	**8.74**
T12	Flowering day ±5 days	0.70	19.95	0.93	10.30
T13	Flowering day ±10 days	0.70	20.19	0.93	10.54
T14	Flowering day until end of season	0.72	19.33	**0.95**	**8.59**
T15	From emergence until end of season	0.43	27.72	0.79	17.89

For the '2011 campaign' data set, we note the late methods (T9 to T14), which integrate NDVI around mid-May, present the better performances. These results are consistent with the phenology calendar of the winter wheat of the study area, since the period from flowering until maturity is the most correlated with grain production. The accumulation of NDVI over the last 10 days of May (method T11) provided the best relationship with the yield measured data (R^2= 0.79 and RRMSE=16.89%).

Given the method T11 yields to good performances for both data sets, we combined them into one for establishing a single model. Figure 5.9 shows the relationship between yield and accumulated NDVI for the method T11 when using both data sets jointly. The established linear model is written in (5.2).

$$Yield = 1.82 \sum NDVI - 5.41 \quad (5.2)$$

In the end of May, we observe winter wheat has reached full vegetative development and yield formation has started. The status of the crop at this stage summarizes the crop growth and development conditions. From this stage, the grain filling stage lasts until maturity is achieved. According to the existing databases and to our field expertise, in the study area the wheat crops harvesting usually takes place from mid-late June to late July. It means that the established linear relationship would be able to predict winter wheat grain yield production between 15 days and 2 months before harvesting, depending on the climatic and management conditions of the studied fields.

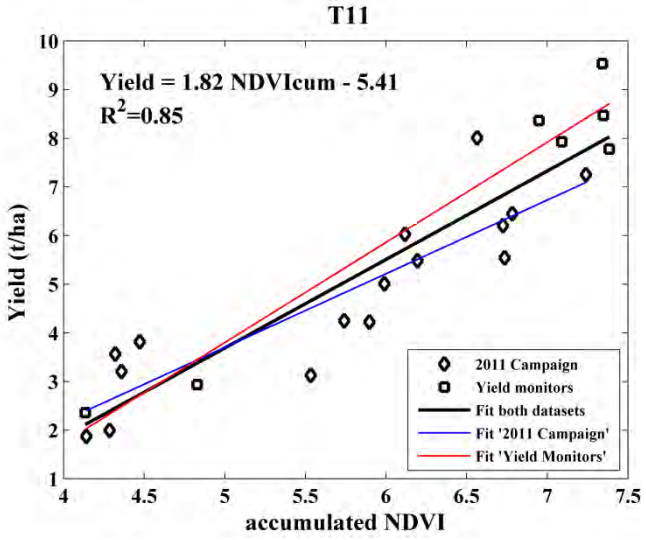


Figure 5.9- Relationship between observed grain yield (y-axis) and NDVI accumulated over the last 10 days of May (method T11). The '2011 field campaign' data are represented by the diamond symbol and data from 'yield monitors' data set by the square symbol. The linear regression model considering both data sets is indicated by the black line. The established relationship and correlation coefficient (R^2) are displayed on the left top corner. The blue and red lines represent the linear fit when considering only one of the datasets (the 2011 campaign, in blue, and yield monitors, in red).

As it was done for the first approach considering a single-date NDVI, here we also applied the joint model linking yield and accumulated NDVI (Eq. (5.2) over the farmers' surveys data set. The comparison between estimated and observed yields is displayed in Figure 5.10. We observe the results are quite similar to those obtained when using the previous model for a single-NDVI date (Eq. (5.1), Figure 5.5). The same inter-annual trends can be noted. In the same way, a reasonable correlation (R around 0.65) is obtained independently of the year. Besides, the 2011 yield estimates present low errors as well, in the range of 26%. In general, the NDVI-integrated model presents slightly better results than the single-date NDVI model. For example, when considering all years together, the integrated model presents

R=0.69 and the slope a=1.45, against R=0. 64 and a=1.55 for the single-date model. It suggests that better yield estimates are done when using NDVI integrated over the last 10 days of May, instead of using NDVI of May 31st. Still, these comments should be considered with caution.

Like the single-data model, the integrated model was calibrated using only data collected in 2011, which bring us to the same conclusion: the established relationship is not appropriate for being directly applied to other years.

From the results presented at this section, we may conclude that the empirical models linking yield to NDVI can provide accurate yield predictions. However the good performances will depend upon the yield data used for calibrating the model, and thus upon the period or date chosen for getting the NDVI values. In 2011, for our area of study, we observed that NDVI data of the period of late May, just after the flowering phase, were the most correlated to yield, and thus the most suitable for predicting yield production. However, since the wheat phenological cycle depends on the climatic characteristics, data from different years can be shifted in time, making it difficult to use a single model for different years. To overcome these limitations we could envisage establishing a relationship for each of the past years, and then creating a database with these models and the associated climatic characteristics of the year. For predicting yield for upcoming years, it would be necessary to select the relationship of the past year that has the climatic condition the closer to those of the on-going year. In any case, the SAFY crop model is expected to surmount these limitations. In the following section, we evaluate the performances of the SAFY model for GAI, biomass and yield estimation.

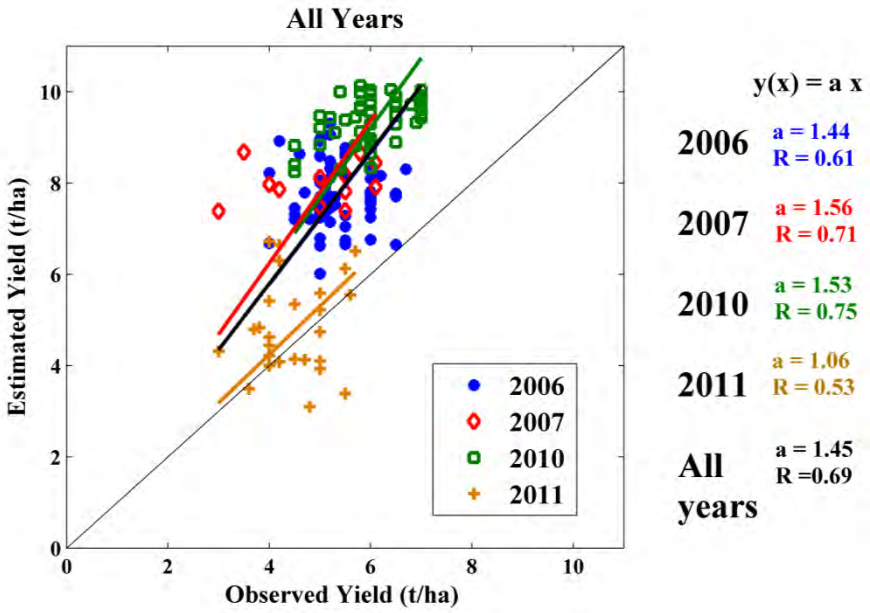


Figure 5.10- Comparison between observed (farmers' surveys) and estimated yield, using a linear model linking yield and accumulated NDVI (method T11). Data are represented by different colors and symbols according to the year, as expressed in the legend. On the right the slopes of the estimated *vs* observed yield relationship for each year are displayed, as the correlation coefficients (R). The performances of the whole data set together (all years) are displayed in black.

5.2. The SAFY model results and discussions

This section presents the main results of the SAFY model, which main outputs are the GAI, biomass production and yield. Initially the GAI and biomass estimates are validated for the experimental sites data set and for the 2011 intensive campaign plots. Next section presents the performances of different methods for estimating yield, for the different data sets. Finally the mean yield obtained for our study area is compared against regional statistics.

5.2.1. GAI and Biomass estimates: local scale

The GAI and biomass validation data set is divided in two groups. The first, called the 'experimental sites' group, contains biomass destructive measurements performed during the vegetative growth period (temporal dynamics) over the Auradé (AUR: 2006, 2008, 2010) and Lamasquère (LAM: 2007, 2009, 2011) sites. For this set of data, the yield data were obtained from farmers' inquiries. The second data set is the so-called the '2011 field campaign'. This set includes the final biomass and grain yield destructive measurements performed over 16 wheat fields in our

study area in 2011 (details in Chapter 2). Note that the biomass temporal dynamics are not available for this set, only the final biomass.

For both groups, the GAI profiles simulated by the SAFY model are compared with the remotely-sensed GAI dynamics (from Chapter 3), except for AUR2008 and LAM2009. For these two growing seasons, the satellite GAI series present significant gaps. Therefore the destructive GAI measurements for these sites were combined with the remotely-sensed GAI data, for building complete gap free GAI series.

The subsequent results were obtained with the optimization of the following parameters: the specific leaf area (SLA), the partition-to-leave function parameters (P_{La} and P_{Lb}), the senescence parameters (STT and Rs), the effective light-use efficiency (ELUE) and the plant emergence day (D_0). The list of these parameters values and intervals for optimization are displayed in Table 4.1. Figure 5.11 to Figure 5.13 present the GAI profiles simulated by SAFY (in blue) and the satellite GAI (black circles) on the left, and the SAFY simulated biomass (in blue) compared with destructive measurements, and associated standard deviations (in black) for the experimental sites on the right. Figure 5.14 and Figure 5.15 present the results for the '2011 field campaign' data set.

The general analysis of the GAI simulated profiles (for experimental sites + fields campaign) reveals that the SAFY model provides accurate estimates of the GAI temporal dynamics, for various climatic years, and for fields with different soil characteristics and under different management conditions. We observe that the growth (ascending slope) and the senescent (descending slope) phases are generally well represented. Nevertheless, in some cases the model may 'miss' the peak of maximum GAI (as it can be seen for AUR2006 and LAM 2009); still this divergence does not seem to significantly affect the GAI seasonal profile nor the biomass estimates.

In the SAFY model, the periods before the day of emergence and after the harvest day are characterized as bare soil periods, with a null GAI. However, the remotely-sensed GAI values range approximately from 0.1 to 0.3 $m^2.m^{-2}$ during these bare soil periods, being rarely equal to zero. Therefore for the optimization process and for statistics estimates only the data within the vegetative growth period is considered, so this systematic, but not constant, bias does not affect the simulations.

The SAFY model presents good performances in GAI estimation (including both data sets), with a mean square error (RMSE) of 0.17 $m^2.m^{-2}$ and a correlation coefficient of 0.98 (Figure 5.16).

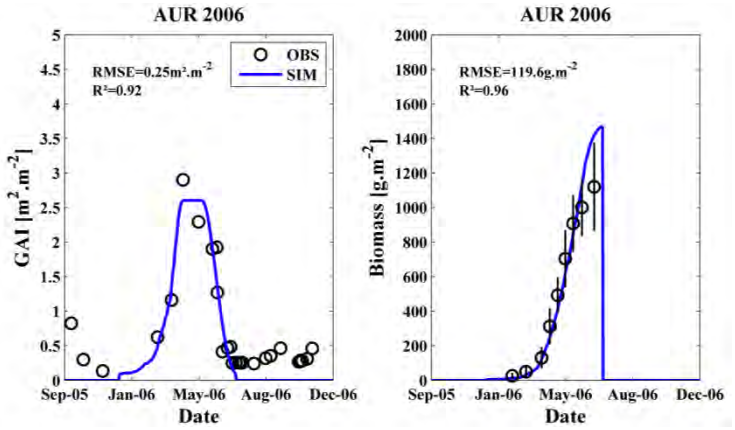


Figure 5.11- GAI and biomass simulated by the SAFY model (blue line) and observed (black circles) for the Auradé site, in 2006. The standard deviations of the biomass destructive measurements are indicated by the black vertical bars.

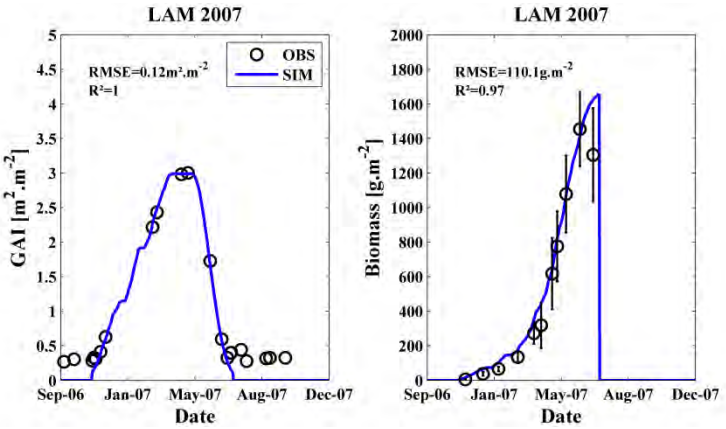


Figure 5.12-GAI and biomass simulated by the SAFY model (blue line) and observed (black circles) for the Lamasquère site, in 2007. The standard deviations of the biomass destructive measurements are indicated by the black vertical bars.

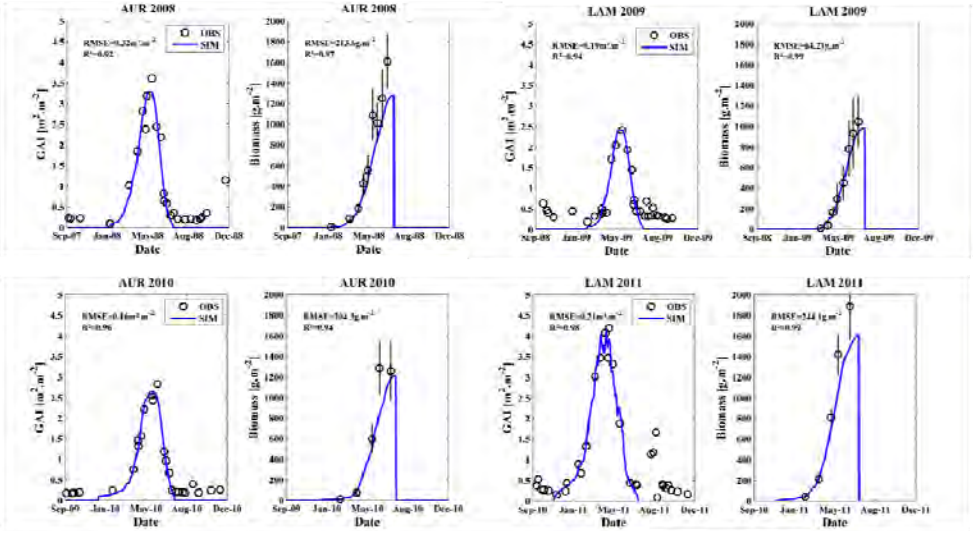


Figure 5.13- GAI and biomass simulated by the SAFY model (blue line) and observed (black circles) for the Auradé 2008 (top left), Lamasquère 2009 (top right), Auradé 2010 (bottom left) and Lamasquère 2011 (bottom right). The standard deviations of the biomass destructive measurements are indicated by the black vertical bars.

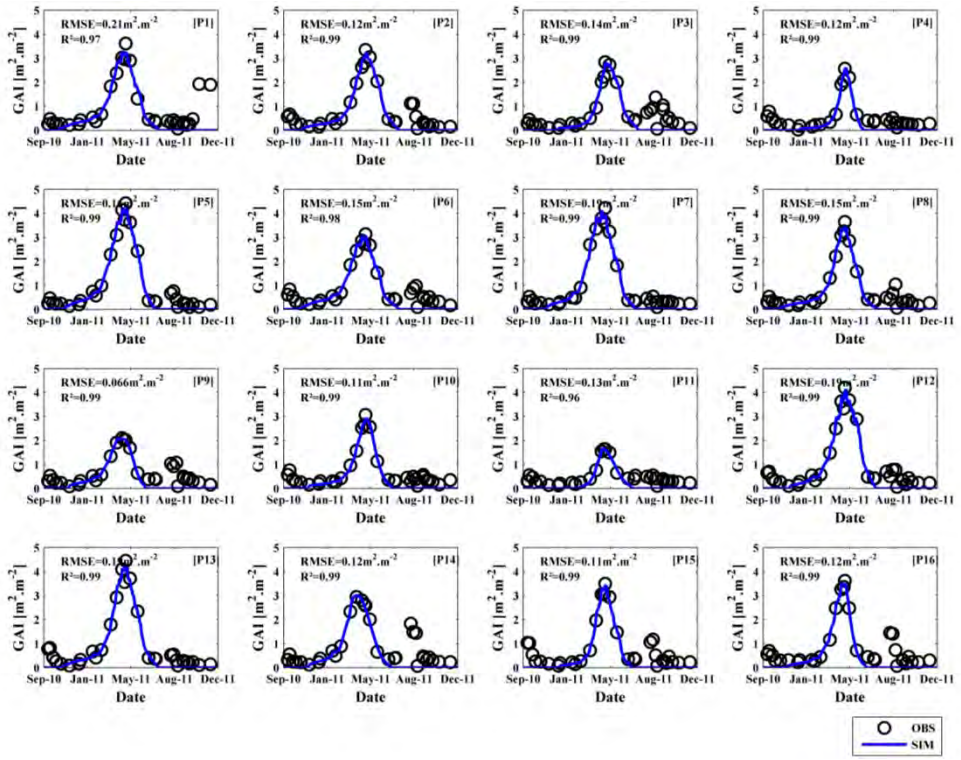


Figure 5.14- Simulated GAI profiles (blue line) and satellite observations (black circles) for the 16 fields of the '2011 campaign' data set.

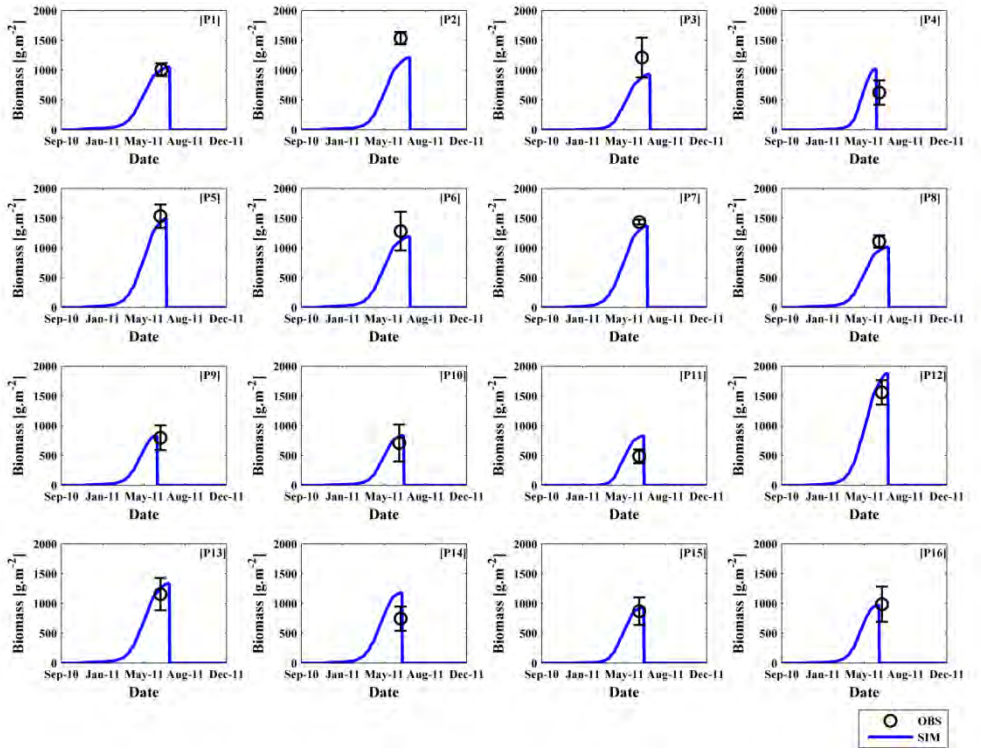


Figure 5.15- Simulated biomass profiles (blue line) and final biomass measurements and associated standard deviations (black circles and bars) for the 16 fields of the '2011 campaign' data set.

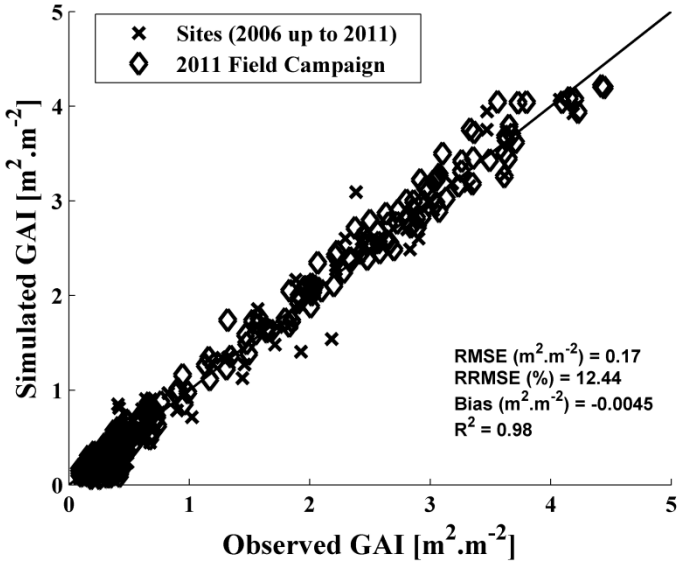


Figure 5.16- Comparison between the green area index simulated by the SAFY model and the satellite-derived GAI (observation) for the 'experimental sites' group (represented by the x symbol) and '2011 field campaign' set (represented by the diamond symbol). The line 1x1 is displayed. The performances (RMSE, RRMSE, Bias and R^2) are indicated on the lower right corner.

The analysis of the day of emergence (D_0) simulated for all available fields (from both data sets) indicates the median D_0 at the 12th November, the earlier D_0 at 16th October and the later D_0 around 13th December (Figure 5.19(a)). This range of dates corresponds to the typical sowing period in our study area, ranging from mid October until early/middle December (emergence occurs generally 1 to 3 weeks after sowing). However, we obtain three very late D_0s: the 10th January for AUR2008, the 04th February for LAM2009 and the 13th February for the plot [P11], part of the 2011 field campaign data set.

For the two experimental sites, this 'delay' might be explained by the GAI profiles used for the optimization, which combine satellite-derived and destructive GAI. Since the GAI destructive measurements started later in the season (January for AUR2008 and March for LAM2009), it is likely that some bias exist between them and the remotely-sensed GAI. Therefore the simulated D_0 for AUR2008 and LAM2009 depend mostly upon the destructive GAI measurements. The [P11] plot presents the later D_0 of its data set group. By analyzing the '2011 campaign' GAI profiles (Figure 5.14), we note that the plot [P11] has the lowest value of maximum GAI (around 2 $m^2.m^{-2}$), which might be due to a late sowing, that is reflected in the late simulated D_0.

It is important to keep in mind that there is a difference between the "agronomical emergence'' (first aerial plant organs appear above the soil) and the "SAFY emergence'' (green organs perceived by the satellite). Despite, the SAFY model was parameterized in order to have an approximated correspondence between the simulated emergence dates and the actual sowing/emergence days.

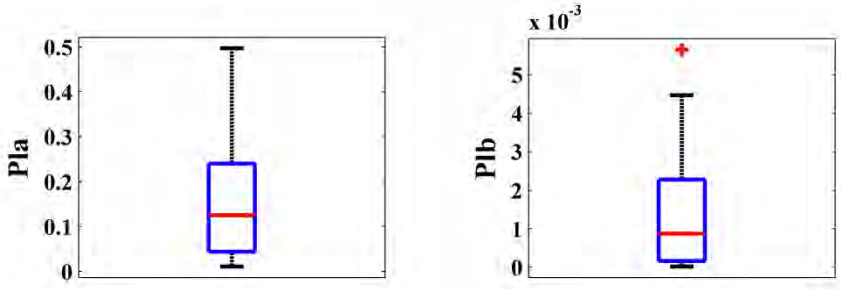


Figure 5.17- Box plots illustrating the distribution of the optimized parameters of the partition-to-leaf function (Pla and Plb) found for the experimental sites and 2011 field campaign data sets.

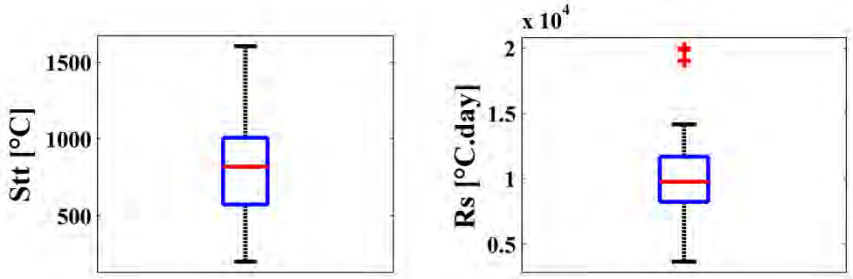


Figure 5.18- Box plots illustrating the distribution of the optimized senescence parameters (STT and Rs) found for the experimental sites and 2011 field campaign data sets.

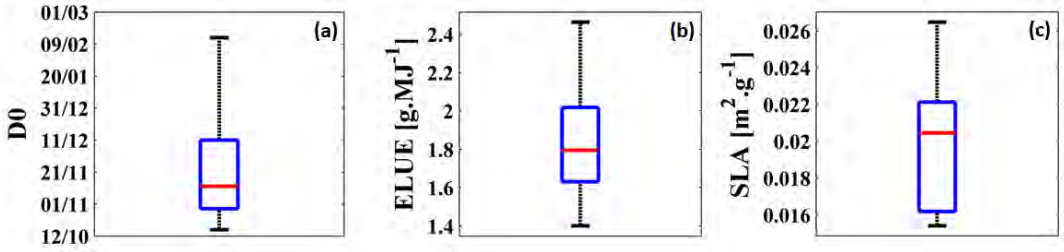


Figure 5.19- Box plots illustrating the distribution of the optimized parameters (D_0, ELUE and SLA) found for the experimental sites and 2011 field campaign data sets.

The box plots in Figure 5.17 represent the distribution of the optimized values for Pla and Plb parameters over 22 fields. The median values for Pla and Plb are 0.1248 and 0.00087, respectively. These values are close to those found from ground measurements (0.1251 and 0.0016), however the P_L function, if it had been established from the median Pla and Plb, would not be realistic, reaching the end of senescence too late in the season. Hence the lower value found for Plb would mean that in simulations it would take more time to the leaves to reach complete senescence. It confirms our choice of optimizing Pla and Plb for each plot, instead of setting them to constant (median) values. In the SAFY model, even if the P_L function parameters have a physical meaning, they remain empirical, and they do not condition other crop processes of the model.

The median temperature threshold for starting the senescence (STT) found over the 22 plots is 820°C, with a standard deviation of 334°C. The STT can be reached at different periods, depending on the climatic conditions of the year. The plot [P18] presents an extremely weak and unlikely STT of 201°C, still the simulated GAI profile and final biomass are correct. It is probably due to compensation with a strong senescence ratio (Rs) parameter (slower senescence).

We obtain a median specific leaf area (SLA, Figure 5.19(c).) of 0.020 $m^2.g^{-1}$, with standard deviation of 0.0034 $m^2.g^{-1}$. This value is slightly stronger than that derived

from measurements (SLA=0.159 $m^2.g^{-1}$). However, it remains close to values found in the literature (for example: 0.022 $m^2.g^{-1}$ in Duchemin et al., (2008)).

The distribution of the optimized light-use efficiency (ELUE) parameter is displayed in Figure 5.19(b). We observe a median ELUE of 1.81 gMJ^{-1}, with a standard deviation of 0.30 gMJ^{-1}. For the sake of simplicity, the ELUE is fixed to the value found in the optimization process for the whole crop season, despite in reality it is known to change through the plant development.

The temporal evolution of biomass production simulated by SAFY, for the experimental sites, are overall in good agreement with the destructive measurements (Figure 5.11, Figure 5.12, Figure 5.13). For AUR2006 (Figure 5.11) we observe that the simulated and observed biomass dynamics match well during the growing period, but during senescence the simulated biomass continues to increase, while the observations tend to stagnate. Still, when considering the standard deviation of the ground measurements, we note that the simulated final biomass is in the range of the observations. Note that the complete maturity stage (harvest) is reached a little later with the model. It is related to the simulated GAI (Figure 5.11, left), which presents a plateau (that is not present on the observations) that causes model to start senescence later.

For LAM2007 (Figure 5.12), we observe a decrease of the *in-situ* measured biomass values at the end of the season, which is not usual, since biomass tends to increase until the maturity phase. This phenomenon is due to the loss of leaves during the senescent phase. The fallen leaves are not taken into account when the biomass samples are collected, inducing lower biomass values. Still, this phenomenon cannot be reproduced by the model.

For AUR2008 and LAM2011 (Figure 5.13) the biomass production is slightly underestimated, even if the GAI profiles are well simulated. This underestimation can be seen in the scatter plot of the simulated against measured biomass for the experimental sites (Figure 5.20), and confirmed by the calculated bias (bias=-50.37 g/m^2) from 2006 to 2011. Moreover, the performances indicate a good correlation (R^2=0.92) and acceptable error (RRMSE=26.63%).

For the '2011 field campaign' data set, only the final biomass measurement is available, which stands for the biomass just before harvest. From Figure 5.15 we observe that the final biomass estimates are quite close to the ground measurements. For this data set, we found a relatively poor correlation between estimated and simulated biomass (R^2=0.58), but we obtained a RRMSE of 21.45% (lower than the 26.6% found for the experimental sites). The statistics indicate a slightly overestimation by the model, with a bias of 68.5 g/m^2 (against -50.27 for the other data set).

When considering both data sets at once (experimental sites + 2011 field campaign) we obtain a good correlation coefficient of 0.88 between estimations and observations, and a relative root mean square error of 25% (Figure 5.21). The under and overestimation tendencies of the experimental sites group and 2011 campaign group, respectively, seem to compensate each other, resulting in a low joint bias of -20.g/m².

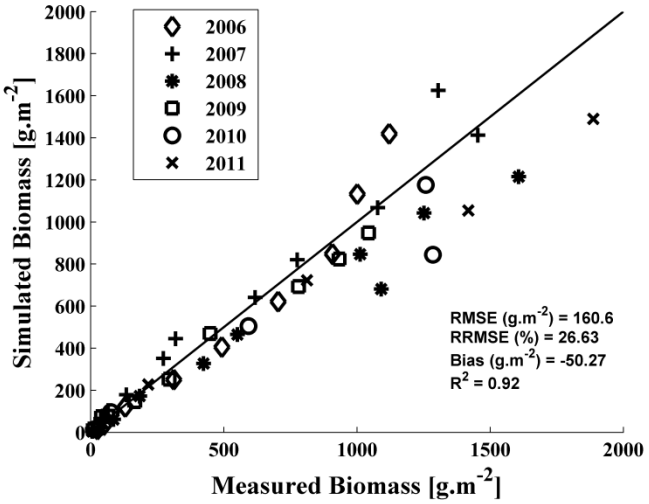


Figure 5.20- Comparison between biomass simulated by the SAFY model and ground measurements for the 'experimental sites' data set. The data for each site-year (AUR2006, LAM2007, AUR2008, LAM2009, AUR2010) are indicated by a different symbol, displayed in the legend.

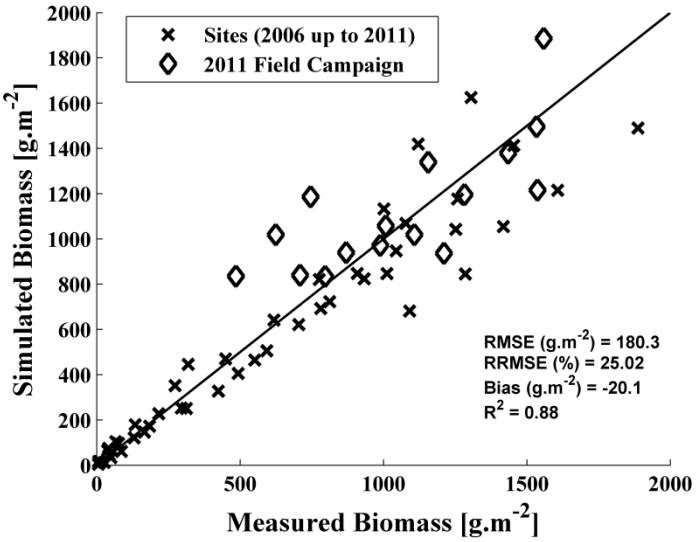


Figure 5.21- Comparison between biomass simulated by the SAFY model and ground measurements for both data sets: the 'experimental sites' group (represented by the x symbol) and '2011 field campaign' set (represented by the diamond symbol). The line 1x1 is displayed. The performances (RMSE, RRMSE, Bias and R^2) are indicated on the right corner.

5.2.2. Yield estimates: local scale

In this section we evaluate the yields estimates with SAFY by using four different approaches. As described in section 4.1.5.2, the first method, hereafter called 'method 1', estimates yields from a constant fraction of the daily biomass production (Eq. 4.12). The 'method 2' and 'method 3' are the ones based upon the STICS model equations, where yield is calculated from an evolving harvest index, depending on a sum of days (Eq. 4.14) or sum of temperature (Eq. 4.15) from the day of start of grain filling. Finally, the 'method 4' calculates yield from the product of the final biomass by a fixed harvest index (Eq. 4.13).

For the methods 2 and 3, the parameter STY (sum of temperature for starting the grain filling) has a key role on the yield estimates. Since the varieties are not known for the '2011 field campaign' data and since the 'experimental sites' varieties are not present on the STICS database, we tested setting different values for STY: to the minimum STY value of the database (657 °C); to the maximum one (960 °C); and to the weighted average STY (796 °C).

Figure 5.22 shows the comparison of measured against yield estimated by SAFY using the 'method 3` when setting STY to the three above cited values. Only the results for the '2011 campaign' data set are displayed. We observe that if the threshold for starting the grain filling is too low (STY=657), the yield production starts too early and thus the simulated yield is overestimated (blue circles). On the contrary, if STY is too high (STY =960), we note an underestimation of the estimated yield (red crosses) because the yield production began too late. For an intermediate value of STY (=796), we obtain better yield estimates, with a low bias and a relative root mean square error of 24.5%. Therefore, for the results presented in the sequence, the STY parameter was set to 796 °C.

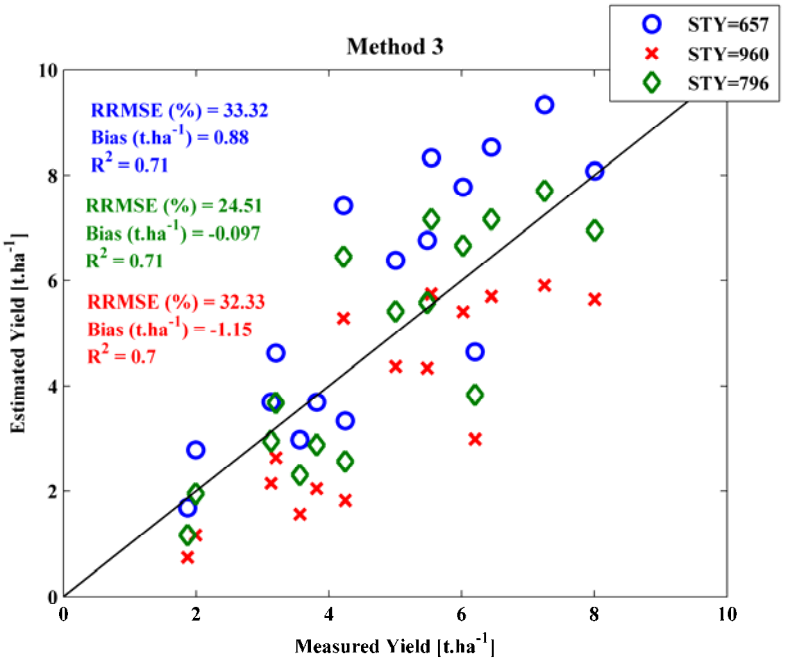


Figure 5.22- Analysis of the impact of the STY parameter over the SAFY yield estimates using the 'method 3' over the '2011 field campaign' data set. Estimated yield for three STY values are displayed: 1) STY=657°C, represented by the blue circles; 2) STY=960°C, represented by the red crosses; 3) STY=796°C, represented by the green diamonds. The performances are indicated on the left, with the same associated colors.

The comparison between the four methods for estimating yield with the SAFY model is displayed in Figure 5.23 and Figure 5.24, for the 'experimental sites' and the '2011 field campaign', respectively. As the yield data come from different sources for each data set (from destructive measurements and from surveys), we prefer not to merge them.

For the experimental sites (Figure 5.23), we observe that methods 1 and 4 have similar performances in terms of RRMSE (23.9% and 24.9%) and of correlation coefficient (R^2=0.31 and 0.39). These methods are very close, since they depend only upon a fixed harvest index (HI) and the produced biomass. The differences in the performances might be due to the different ways for setting HI. Even if the HI used in method 1 is derived from the literature, it yields to good estimates. The methods 2 and 3, on the contrary, present poor performances, with much dispersed points (almost null correlation) and high errors. This might happen because of the low number of available points (only 6), making it difficult to draw further conclusions, and also because of the uncertainty in the yield data. This information is gathered from the farmers, who are likely to average the yield of several fields, so we do not have the yield data at the actual field scale.

For the '2011 field campaign' data set we observe different behaviors (Figure 5.24). In this case, methods 1 and 3 present alike performances, with close relative errors (RRMSE=23.1% and 24.5%) and correlation coefficients (R^2=0.62 and 0.71). The comparison of methods 2 and 3, both having an evolving harvest index, indicates the method 3 yields to more accurate results. It means that considering a sum of temperature for the yield estimation (method 3) is better than considering a factor based on an amount of days (method 2). The method 4 presents an acceptable relative error of 28.5% but a poor correlation (R^2=0.44). As the harvest index used in this method (HI=0.45) was obtained from this same data set of ground measurements, better results were expected for method 4. These errors in yield estimation can be related to the slight overestimation of biomass estimates (Figure 5.21, diamond symbol), which is directly reflected on the yield estimates.

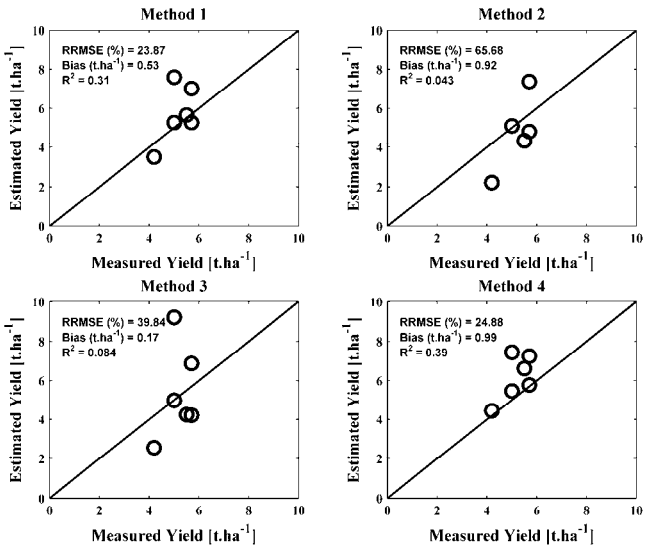


Figure 5.23- Comparison of the methods for estimating yield with the SAFY model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the experimental sites (AUR2006, AUR2008, AUR2010, LAM2007, LAM2009 and LAM2011).

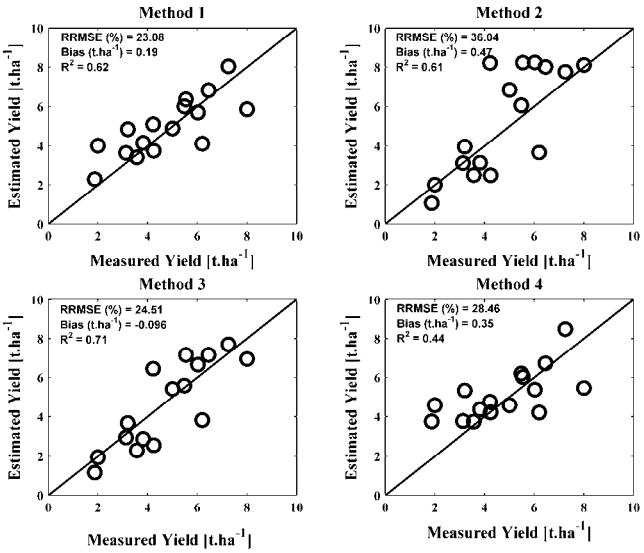


Figure 5.24- Comparison of the methods for estimating yield with the SAFY model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the 2011 campaign fields.

Next step is to evaluate the performances of SAFY model on yield estimation for two other data sets: yield from farmers' surveys (section 2.4.1) and from the yield monitors (section 2.4.2).

Figure 5.25 shows the performances of the four methods for estimating yield for the farmers' surveys data set. Here, the 'measured yield' corresponds to the yield informed by the farmers (thus equivalent approach as the experimental sites yield data). The scatter plots include data from 139 winter wheat fields, concerning 2006, 2007, 2010 and 2011. In general, we note that the observed yield covers a narrow interval, from 3.0 t.ha^{-1} to 7.0 t.ha^{-1}, with a mean yield of 5.4 t.ha^{-1}. Methods 1 and 4 provide good results in terms of relative error (RRMSE = 24.5 % and 22.4%), but very low correlation (R^2 around 0.1). Besides, we observe a slight trend of underestimation for method 1 and of overestimation for method 4; still these trends remain minor.

The methods 2 and 3 present poor performances, with elevated errors and almost null correlation. For these methods, the estimated yields have wider dynamics (with estimates that exceed 10 t.ha^{-1}), while the measured ones remain concentrated (between 3.0 and 7.0 t.ha^{-1}). One of the sources of the significant dispersion of the results may be the precision of the 'measured' yield data. As discussed earlier, the yield information collected from the farmers is probably averaged across the nearby wheat fields cultivated by the same farmer. For illustration, when a farmer finishes harvesting one of his fields, and starts harvesting a nearby field, if the container (with the harvested grains) is not full, he continues to fill it with the grains from the

second field. Thus the final grains weight (yield) is an average of the two fields. Consequently, if present, the inter-field heterogeneity will be reproduced by the model, but flattened by the farmers reported values (which could also explain the narrow range of informed yield values). Even if the scattering aspect is much more noticeable for methods 2 and 3, it can also be observed for methods 1 and 4.

Since we have the farmer identity associated to each yield value of the farmers' surveys data set, we tested averaging the model outputs for the fields that belonged to the same farmer and that presented the same measured yield value. As a result, we expect to minimize the scattering effect, especially for methods 2 and 3.

Figure 5.26 displays the obtained results. We observe the new averaged yield estimates are less dispersed than the previous ones (in Figure 5.25) and that the relative errors decreased a little for all four methods. This suggests our alternative averaging approach is pertinent. Still, it remains an approximate solution, since we do not know if the fields that we used to calculate the mean yield values correspond exactly to the fields that the farmers considered for assessing their yield estimates. For example, if a farmer owns three contiguous wheat fields, but only two of them are included in our database; the farmer's average yield is then calculated across three fields, and for us, across only two fields.

In addition, the analysis of the results per year indicates particular trends for 2007 (red diamond) and 2010 (green square). It can be mostly viewed for methods 2 and 3, and slightly for method 1. These methods strongly depend upon the climatic conditions (temperature), since the yield production starts when a cumulated temperature threshold is reached. Besides, methods 2 and 3 are even more affected by the climate because of their dynamic harvest indexes. The analysis of the mean daily temperature dynamics of the region (Figure 5.27(b)) revealed that, in fact, the years 2007 and 2010 had higher temperatures for the period of September (of year-1) to January. This could lead to early attainment of the temperature threshold for grain filling (STY) and thus stronger final yield estimates. Furthermore, jointly with the temperature conditions, the simulated day of emergence (D_0) may have an effect on the yield estimates, since the beginning of the sum of temperature depends upon D_0. The analysis of the box plots of the simulated D_0 for each year (Figure 5.27(a)) indicates that in 2007 the crops development started earlier than for the others years. Few field data are available for this year and the D_0's are within a narrow range. In 2010, we observe a high dispersion of D_0, which can also be seen on the yield estimates (methods 2 and 3, Figure 5.25).

According to these results, we can conclude that the method 4 is the better for estimating final yield with the SAFY model, in order to obtain estimates in agreement with yield data gathered from farmers' surveys. Besides, it should be noted that this kind of data set ought to be used with caution; and sometimes a pre-processing step

might be necessary (like the described averaging approach) for better interpretation of the results.

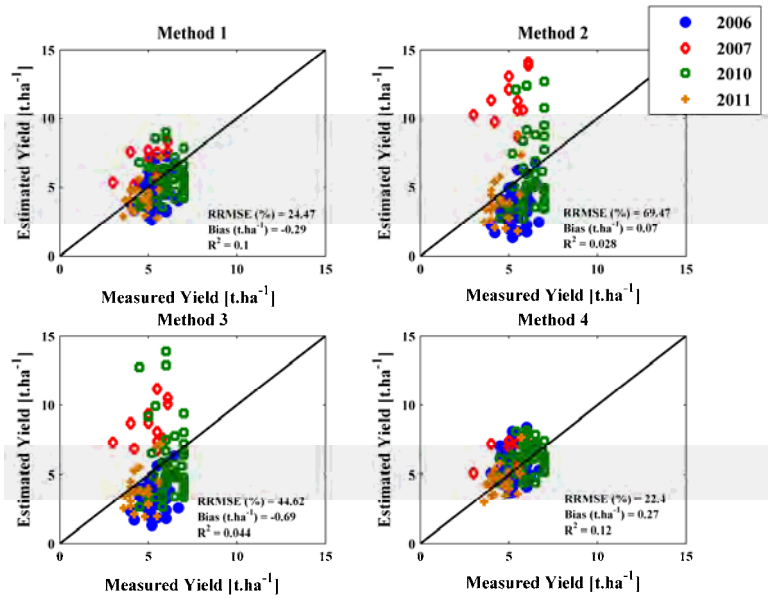


Figure 5.25- Comparison of the methods for estimating yield with the SAFY model with the famers' surveys yield dataset. For each year the results are displayed in a different color and symbol: 2006 (blue circle), 2007 (red diamond), 2010 (green square) and 2011 (yellow plus). The statistics (RRMSE, Bias and R^2) were calculated over the whole dataset.

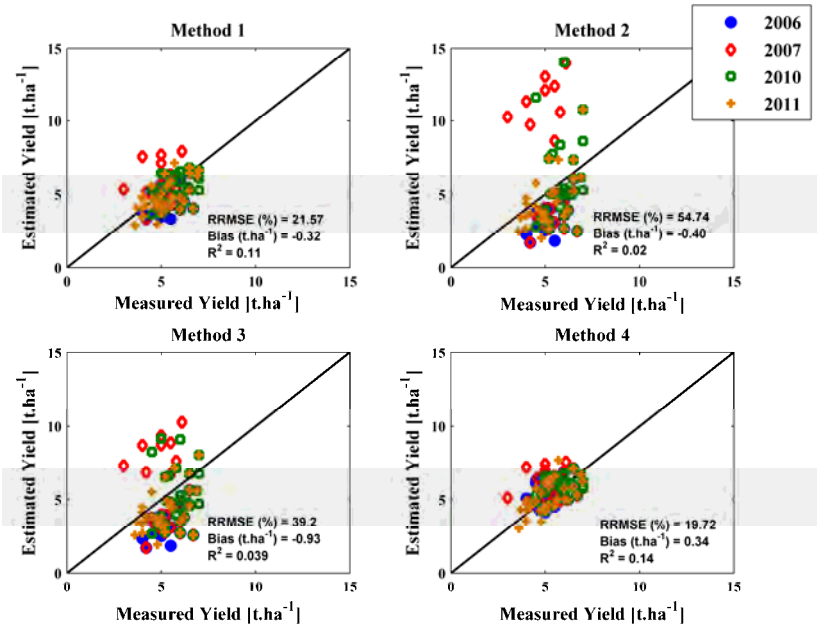


Figure 5.26- Comparison of the methods for estimating yield with the SAFY model with the famers' surveys yield dataset using the averaging-approach. Same description as Figure 5.25.

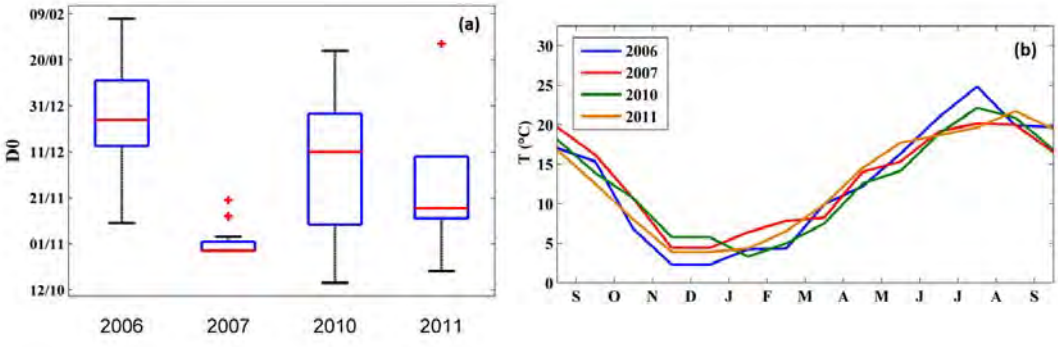


Figure 5.27- (a) Distribution of the simulated emergence dates (D_0) for the farmers surveys' dataset. (b) Dynamics of the mean daily temperature from September to September for the 2006 (blue), 2007 (red), 2010 (green) and 2011 (yellow) crop seasons.

The last yield results to be investigated are those concerning the data from the yield monitors mounted on the harvesting machines (section 2.4.2). Figure 5.28 shows the estimated against the measured yield for the four methods. We observe this data set encompasses a large range of yield values (from 2.3 $t.ha^{-1}$ to 9.5 $t.ha^{-1}$), which is well reproduced by the model, but systematically underestimated. Compared to the other datasets, we note that here we obtain extremely high correlation between estimated and measured data (R^2 from 0.76 to 0.93, depending on the method). Nevertheless a strong bias is observed, mostly for the high yield values. Even if we only have a small set of data (7 observation points), we are tempted to state that the significant bias can come from an omitted, or inadequate, calibration of the yield sensors. A way to correct this bias would be to consider the final yield values noted by the farmers (which are not available for these fields), in order to re-calibrate the yield measured by the yield monitors.

It is important to remind that this data set is constituted of yield maps where intra-field yield variability is accessible. For this study, we considered the mean yield (yield averaged across all pixels within each field), because we work at the field scale. Future works will explore the richness of this data set: for validating the model outputs at the pixel scale, for further understanding of the within-field heterogeneity, and so forth.

In this section, we investigated the performances of four methods for estimating yield with the SAFY model. For this, we considered yield data sets provided by different sources, each of them presenting specific characteristics. We can class them in three main yield data types (with the associated data sets): a) destructive measurements (2011 field campaign); b) yield reported by the farmers (experimental sites + farmers' surveys data sets); c) yield maps (yield monitors). According to the results that were presented, we cannot choose 'the best method' for estimating

yield with SAFY, since it depends upon the yield observation data category. Nonetheless, we observe that, in general, the method 1 (for which yield = fixed harvest index by daily biomass production) yields to good results, independently of the data's origin. The particularity of this method consists in the fact that the harvest index was fixed by Duchemin et al., (2008), and it works well for our study area. The method 4 (yield = fixed harvest index by the final biomass) presented overall good and rather stable performances, for all data sets. It revealed to be a consistent method, even if it takes into account only the final biomass production (and not the dynamics). Methods 1 and 4 can be considered as equivalents, except that their respective harvest indexes were set differently.

In all cases, the method 2 presented the poorest performances. The method 3, on the contrary, presented motivating results. This method provided good yield estimates for the '2011 field campaign' data set (RRMSE=24.5% and R^2=0.7) and estimations were well correlated with observations (R^2=0.93) for the 'yield monitors' data set. Given the obtained results, it seems worthy to further investigate the method 3. We propose testing this method over yield data sets like the '2011 field campaign', but for various climatic years and sites. Here we could only test it for few plots in 2011. Besides, there is room for improvement on the setting of the STY (temperature threshold for starting grain filling) and the VITICARBT (a fixed constant, part of evolving harvest index) parameters. Both parameters were set according to the STICS database, therefore their optimization remains plausible.

With this study, we realized the high values of yield data obtained from destructive measurements for validating the crop model outputs. Indeed, destructive measurements on contrasted years and plots (temporal and spatial variability) allow having a range of yield data suitable for validation of the model. The farmers' surveys data set presents too much uncertainty, because of the likely aggregation of each plot's yield information. The 'yield monitors' data set presents an interesting spatial variability for model validation at the pixel scale. This issue should be further investigated. Finally, we recommend strong effort for validating yield estimates in a range of conditions (spatial and temporal) with this kind of modeling approach since yield estimation remains a key component of the carbon budget modeling.

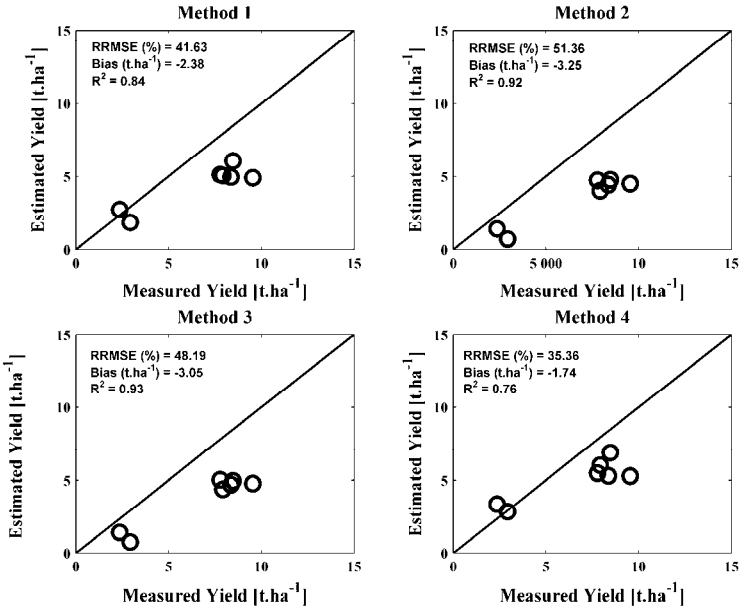


Figure 5.28- Comparison of the methods for estimating yield with the SAFY model for the yield monitors dataset. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis). Data collected through the 2011 crop season.

5.2.3. Yield estimates: regional scale

The evaluation of the performances of yield estimation at regional scale is not easy, since there are no available data that match exactly the study area. Nevertheless the Agreste yield statistics provide annual yield data at the departmental scale (Section 2.5.4). Our study area is defined by the Formosat-2 extent (of 24km x 24km), which covers approximately 10% of the Haute-Garonne department area. Therefore we cannot perform a validation or a direct comparison between the Agreste statistics and the yield estimates obtained for the wheat fields within the study area (that are identified by the land use maps RPG, 2.5.2). Still, we can evaluate inter-annual trends and try to draw general conclusions.

Figure 5.29 shows the comparison between the Agreste yield statistics and the yield estimated by the SAFY model using method 4 (yield =DAM_{max} x HI), for 2006, 2007, 2010 and 2011. We observe that the SAFY estimates have the same inter-annual dynamics as the Agreste statistics, following the increase/decrease yield trends according to the year. The yield simulated in 2007, however, presents the highest overestimation. That year biomass production and therefore grain production (because method 4 was used) probably started too early (see previous section) leading to highest yield overestimation. Besides, the two sets of data are well correlated (R^2=0.85). The model however seems to systematically overestimate

the yield production by approximately 12%. This might be due to an elevated value of the harvest index, which was set using destructive data collected just in 2011. Otherwise it could be due to biomass overestimation by the model, which is reflected into the yield estimation. Still, the overestimation remains not significant. Furthermore it should be noted that the two data sets don't have direct equivalence. At last, these results suggest that the fields within our study area are a representative sample of the Haute-Garonne department fields.

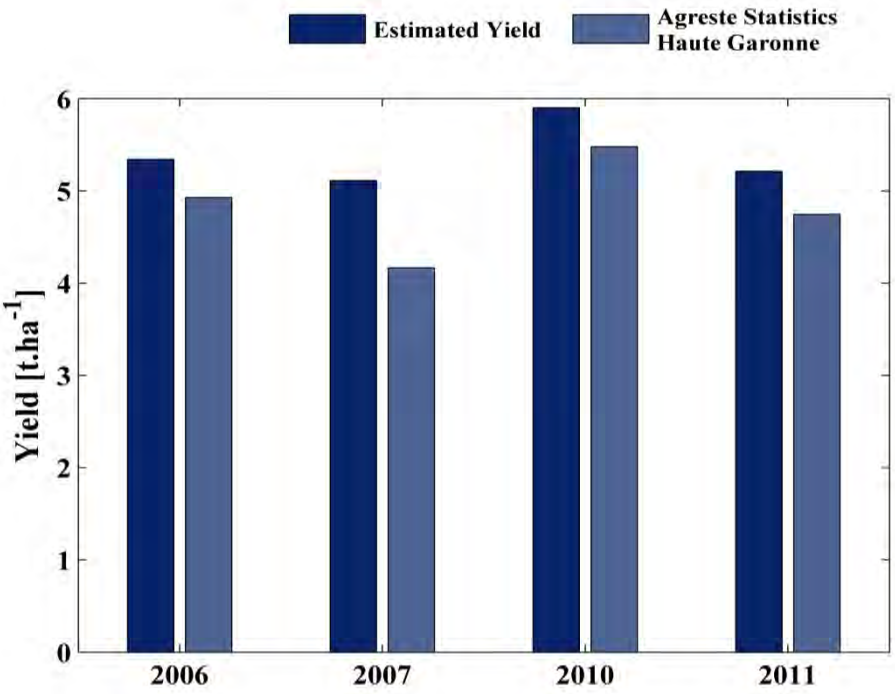


Figure 5.29- Graphic illustrating the mean grain yield estimated for the winter wheat fields included on the RPG (dark blue) with the SAFY model and the Agreste statistics for the Haute-Garonne department (light blue). Results are compared for 2006, 2007, 2010 and 2011.

5.3. The SAFY-CO_2 model results and discussions

In this section we present the results obtained for the SAFY-CO_2 model, which was described in Chapter 4.3. We remind that the main purposes of the SAFY-CO_2 model are to estimate seasonal dynamics of the components of the CO_2 fluxes: such as GPP, Ra, Rh, NEE. Besides, we aim at obtaining accurate estimates of crop growth (GAI) and production (biomass and yield), and then at evaluating and comparing these results with those obtained by the SAFY model. Finally, we aim at establishing and evaluating the annual net ecosystem carbon budgets for winter wheat crops.

The simulated fluxes are compared with the *in-situ* flux data, which were estimated from the measured turbulent fluxes at the Auradé and Lamasquère experimental sites. For the winter wheat crop, a total of five years-site data were available for this work: Auradé in 2006, 2008 and 2010; Lamasquère in 2007 and 2009. For the evaluation of production estimates, we make use of this same 'experimental sites' data set plus the one referred as the '2011 field campaign'.

The results regarding the components of the CO_2 fluxes (GPP, Reco, NEE) are presented at first and then those concerning the GAI, biomass and yield, in the same order as these terms are calculated by the SAFY-CO_2 model.

5.3.1. Estimation of the CO_2 fluxes components

5.3.1.1. GPP estimates

The dynamics of simulated gross primary production (GPP) were compared to the in-situ GPP data. Figure 5.31 to Figure 5.32, and Figure 5.35 to Figure 5.38 present the simulated and observed fluxes for the five investigated crop seasons. Table 5.2 summarizes the obtained performances, in terms of correlation (R^2) and root mean square error (RMSE).

In general, we observe that the estimated GPP dynamics are in agreement with the measured ones, presenting a correlation coefficient around 0.93 and a RMSE about 1.3 gC $m^{-2}d^{-1}$. The dynamics of GPP are very close to those of GAI. In addition, regardless the year, we note that the growing period estimates, approximately until early May, match well with the observations; however, a slight deviation is observed through the senescent phases at Auradé in 2006 and Lamasquère in 2007 and 2009.

In terms of absolute observed GPP values, the 2007 and 2010 crop seasonal dynamics present higher values than the others years. For these years, despite the good correlation between observations and estimates (R^2=0.94), the GPP estimates are underestimated by the model through the whole growth development period (Figure 5.32 and Figure 5.38, top). Higher measured GPP might be caused by the

higher precipitation during the vegetation season in 2010 and a very early development of the wheat in the 2006-2007 season, because of a warm winter leading to high GAI and biomass values during spring.

For LAM2007, we observe that the maximum GAI values both simulated by SAFY-CO_2 and observed (satellite-derived) are equal to 3 $m^2.m^{-2}$ (Figure 5.42); however, the destructive measurements performed over this site indicate a maximum GAI of 5.4 $m^2.m^{-2}$. Even if the satellite-derived GAI are known to be inferior to destructive measurements, this significant difference (higher than 40%) might be due to a possible saturation of the GAI retrieval method. Therefore the assimilation of the lower remotely-sensed GAI by SAFY-CO_2 may induce an underestimation of the GPP estimates.

Besides, in 2007 we observe two important peaks in the in-situ GPP dynamics that are not represented by the model simulations (within the black contour in Figure 5.33). A closer look at the net global radiation (Rg) values at the time of the GPP peaks indicates that high values of radiation are measured as well. Nevertheless, the analysis of the *fELUE* function, which corrects the diffuse radiation effects on GPP, shows that *fELUE* is negatively correlated to the GPP and Rg peaks (Figure 5.33). It means that the investigated periods are characterized by strong radiation and clear sky conditions. This suggests that the modeled *fELUE* function over-corrects the diffuse radiation effects, *i.e.* that the plant light-use efficiency is not enough important in some cases when clear sky conditions are met.

Still analyzing the LAM2007 results, we observe that the GPP estimates are equal to zero during some days around end of December and again by the end of January (periods enclosed in red in Figure 5.32). These periods correspond to days with very low mean daily temperatures (below 0°C), which are inferior to the minimal temperature for crop functioning (Tmin). Therefore the temperature stress function [$F_T(Ta)$] is equal to zero, and given that it has direct effect over the photosynthesis, the simulated GPP is zero as well. These effects, however, are not observed in the measured GPP dynamics. This might happen for two possible reasons: i) the wheat crop keeps a photosynthetic activity below 0°C, suggesting that the threshold Tmin should be better adjusted in SAFY-CO_2; ii) since the observed GPP are not directly measured, being estimated from a partitioning method of the NEE measurements into GPP and Reco, the in-situ GPP estimates might be compensated with the Reco estimates. It means that, in fact, the observed GPP values might be lower than exhibited in the graphics and thus that the GPP values provided by the SAFY-CO_2 for these periods might be correct.

In AUR2010 (Figure 5.38), the GPP estimates are overall underestimated. The analysis of the correspondent GAI dynamics shows a maximum GAI of about 3 $m^2.m^{-2}$, which can be considered as low GAI. Still for AUR2010, we observe that the model

is very sensitive, presenting strong variations in short periods of times. Figure 5.39 shows a zoom of the AUR2010 GPP daily dynamics, where some of these abrupt decrease in GPP can be observed. This closer look shows that for almost all the cases, the decrease in GPP is present on both observed and simulated estimates. An exception occurs at May 3rd, when the observation drops to 8.2 gC $m^{-2}d^{-1}$ while the model goes further reaching 1.65 gC $m^{-2}d^{-1}$ (encircled in red). We observe that in all cases (except the 3rd of May) these decreases correspond to net incoming radiation peaks, indicating that the model responds well to climatic events. For the May 3rd, however, the model strongly underestimates GPP. The analysis of the *fELUE* function shows that this day was characterized by a strong fraction of diffuse radiation (very cloudy day). Even if this function is supposed to take the diffuse effect into account, by increasing the plant light-use efficiency in cloudy days, the drop in global radiation was too strong and the *fELUE* function did not compensate it enough. Still, this behavior remains punctual.

For AUR2006, from the scatter plot (Figure 5.31, right) we observe an overall overestimation of the GPP estimates, but by evaluating the daily dynamics it can be noted that this overestimation mainly occurs during the senescent phase. This indicates that the growing phase is better represented by the model than the senescent one; nonetheless this can be related to climatic conditions. Indeed, a water stress is observed at the end of 2006 crop season (see section 5.4). Besides, we observe an increase in the observed GPP in mid-August until late September (enclosed in red in Figure 5.31), which is not reproduced by the model. This observed photosynthesis is due to a re-growth of the wheat crop and also to the presence of weed, which is illustrated in Figure 5.30. In the exhibited photographs (taken late September) we can see the significant presence of growing green vegetation around the flux measurement tower and also over the site. The sudden decrease in the GPP flux in early October is due to soil labor, which removed all the existing vegetation. An increase of the observed GPP is also observed at the beginning of the season, in early September (enclosed in red in Figure 5.31). It probably corresponds to re-growths, but from the previous summer crop, which were removed late September when the soil was ploughed.

In AUR2008 (Figure 5.35, top), we observe a good agreement between observed and simulated GPP dynamics. A slight underestimation can be noted in the beginning of the growing phase (ascending slope) and also in the senescent phase. During the senescence the model is less sensitive than the observations, since the descending slope is relatively constant, while certain variations are still present on the measured data.

The results for LAM2009 (Figure 5.36, top, Table 5.2) are, in general, similar to those obtained for the others years, thus correct. A GPP overestimation from

maximal development stage until mid-senescent phase can be observed, but no other particular anomaly is remarked in the model estimates. It is also possible to observe that in the end of the senescence phase (in June), the modeled GPP drops, until it reaches zero, corresponding to the crop harvest. On the other hand, the photosynthesis observations remain relatively active (around 3 gC $m^{-2}d^{-1}$) until September. According to field annotations, in this year they detected weed development through the senescence phase, which remained in the field even after harvesting (on 13th July).

For all the investigated years, we note that the model is not designed to simulate GPP after harvest. Thus GPP is zero from harvest until the end of the cropping year. Nevertheless, the observed GPP values often remain "active" and may increase again, ranging from 0.5 up to 3gC $m^{-2}d^{-1}$. This happens due to the presence of weed or to the re-growth of seeds that eventually fell on the soil during harvest. It can also be an artifact of the partitioning method or the influence of nearby summer crop plots over the fluxes measurements observations. Indeed, during those periods as the soil is bare (after harvest), the *footprint* (area that contributes to the flux measurements) increases and may encompass more the fluxes coming from adjacent plots. The re-growth phenomenon usually occurs in case of important precipitation events combined with appropriate temperatures. Some studies (Béziat et al., (2009); Ceschia et al., 2010) showed that these re-growths can have a significant impact on the ecosystem carbon assimilation (through photosynthesis), and thus on the ecosystem carbon budget calculations; this encourages future works to take into account the re-growths and weed events on the crop modeling approaches, and also to consider intercropping practices (that can have similar impact as re-growths). Furthermore, satellite missions like Sentinel-2 would provide the necessary satellite data series to detect and take into account these events by remote sensing products assimilation by the models.

Table 5.2- Performances of the carbon fluxes seasonal dynamics estimated by the SAFY-CO_2 model for winter wheat over the experimental sites, from 2006 to 2010.

	GPP		Reco		NEE	
	R^2	**RMSE***	**R^2**	**RMSE***	**R^2**	**RMSE***
AUR2006	0.91	1.42	0.77	0.80	0.85	1.21
LAM2007	0.94	1.4	0.80	1.23	0.87	1.06
AUR2008	0.94	1.26	0.74	0.89	0.89	1.05
LAM2009	0.93	1.13	0.71	0.90	0.79	1.19
AUR2010	0.94	1.27	0.82	0.85	0.88	1.28

* RMSE in gC $m^{-2}d^{-1}$.

Figure 5.30- Photographs of the Auradé site, taken the 29th September 2006, which indicate the presence of crop re-growths and weeds.

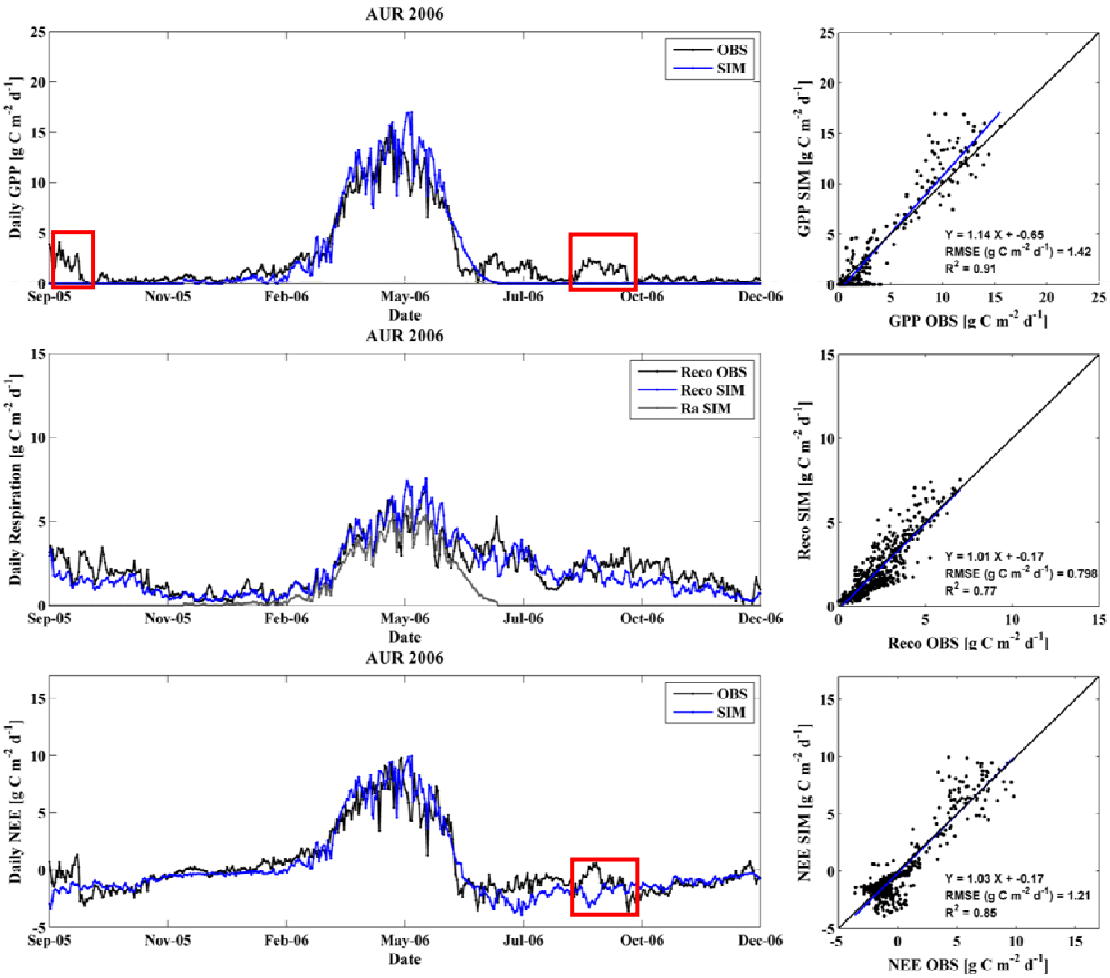


Figure 5.31- Evolution of the carbon fluxes for the Auradé site, 2005-2006 winter wheat crop season. Observed fluxes are represented in black and simulated fluxes in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.

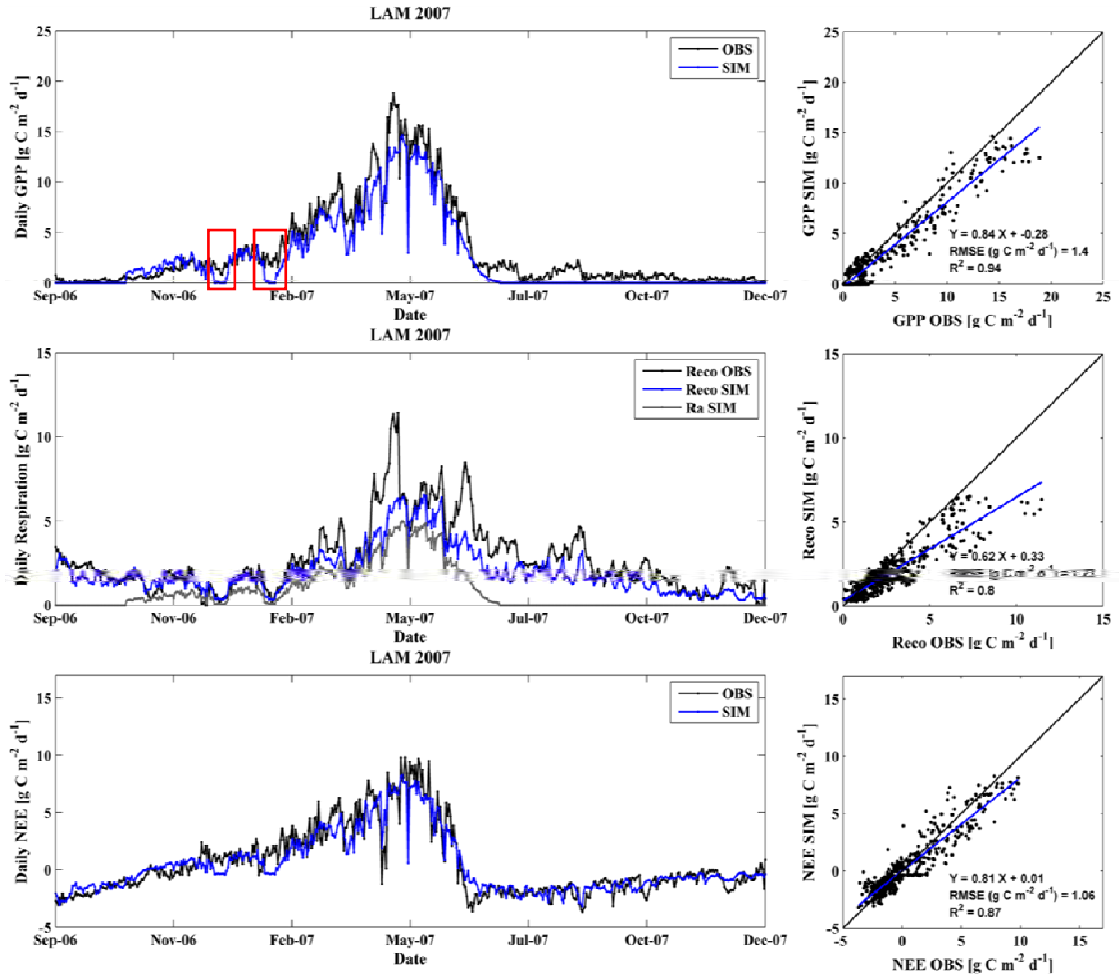


Figure 5.32- Evolution of the carbon fluxes for the Lamasquère site, 2006-2007 winter wheat crop season. Same legend as Figure 5.31.

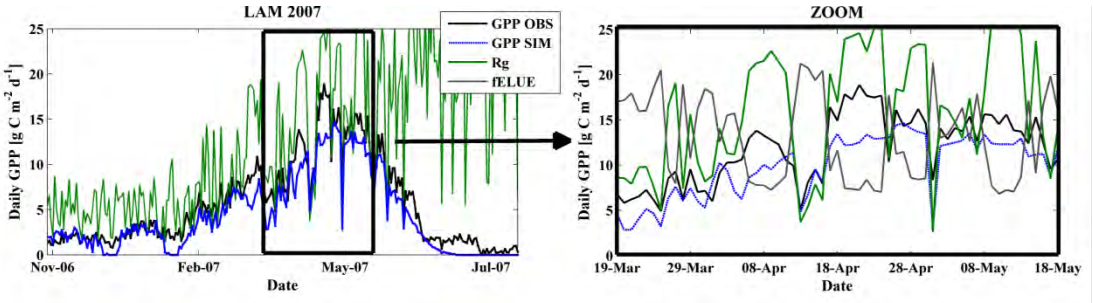


Figure 5.33- Zoom over the GPP dynamics for the Lamasquère site, 2006-2007 winter wheat season, with the global incoming radiation (Rg, in $W.m^{-2}$) and the *fELUE* function (x5, for visualization purposes).

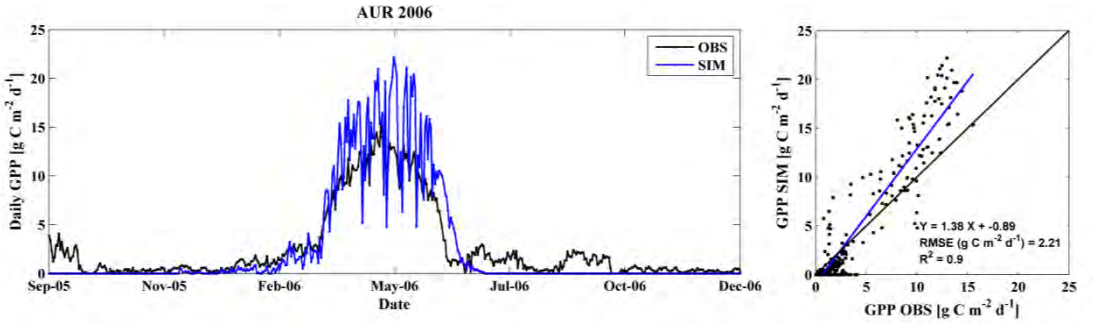


Figure 5.34- Example of the GPP dynamics simulated by SAFY-CO_2 when the effect of the diffuse global radiation fraction over GPP is not considered (no *fELUE* function). Results for the Auradé site, in 2006.

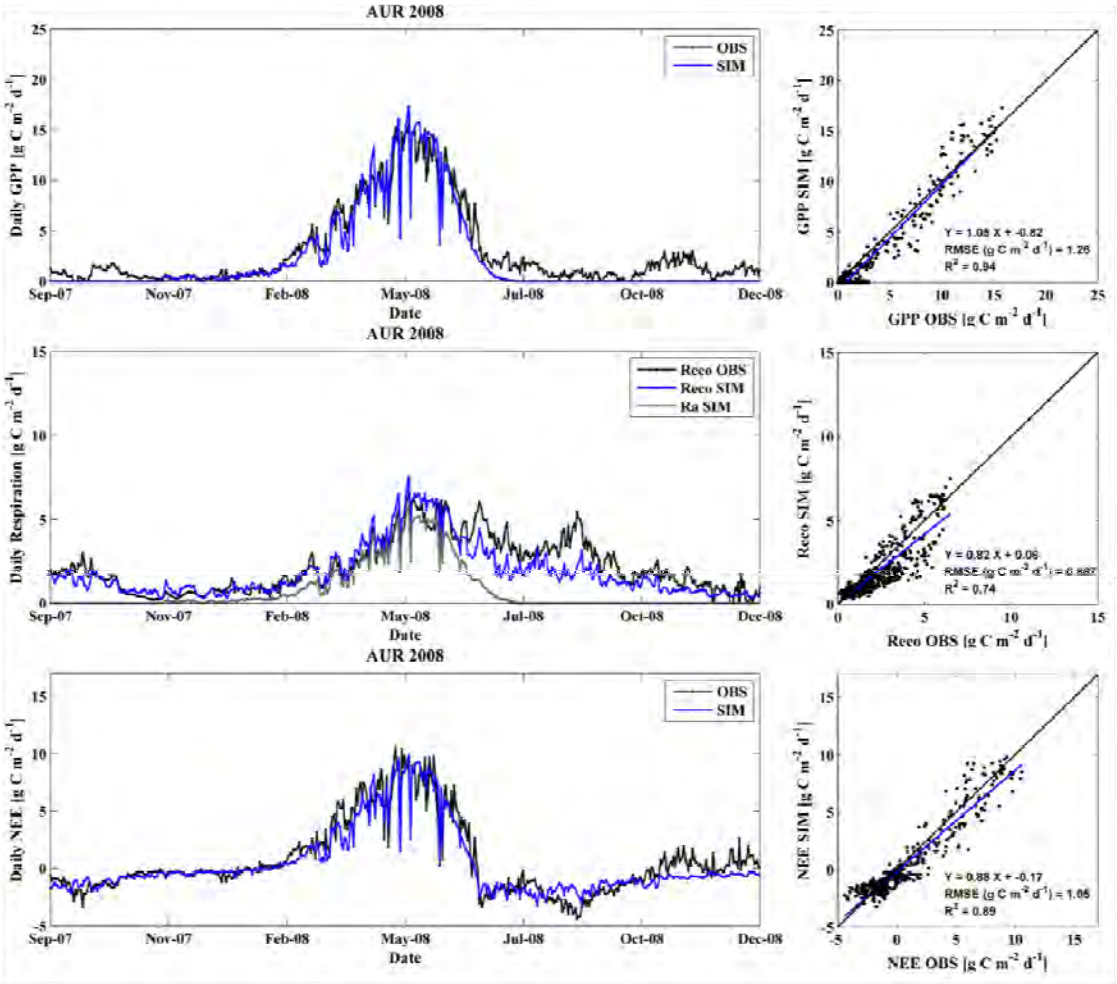


Figure 5.35- Evolution of the carbon fluxes for the Auradé site, 2007-2008 winter wheat crop season. Same legend as Figure 5.31.

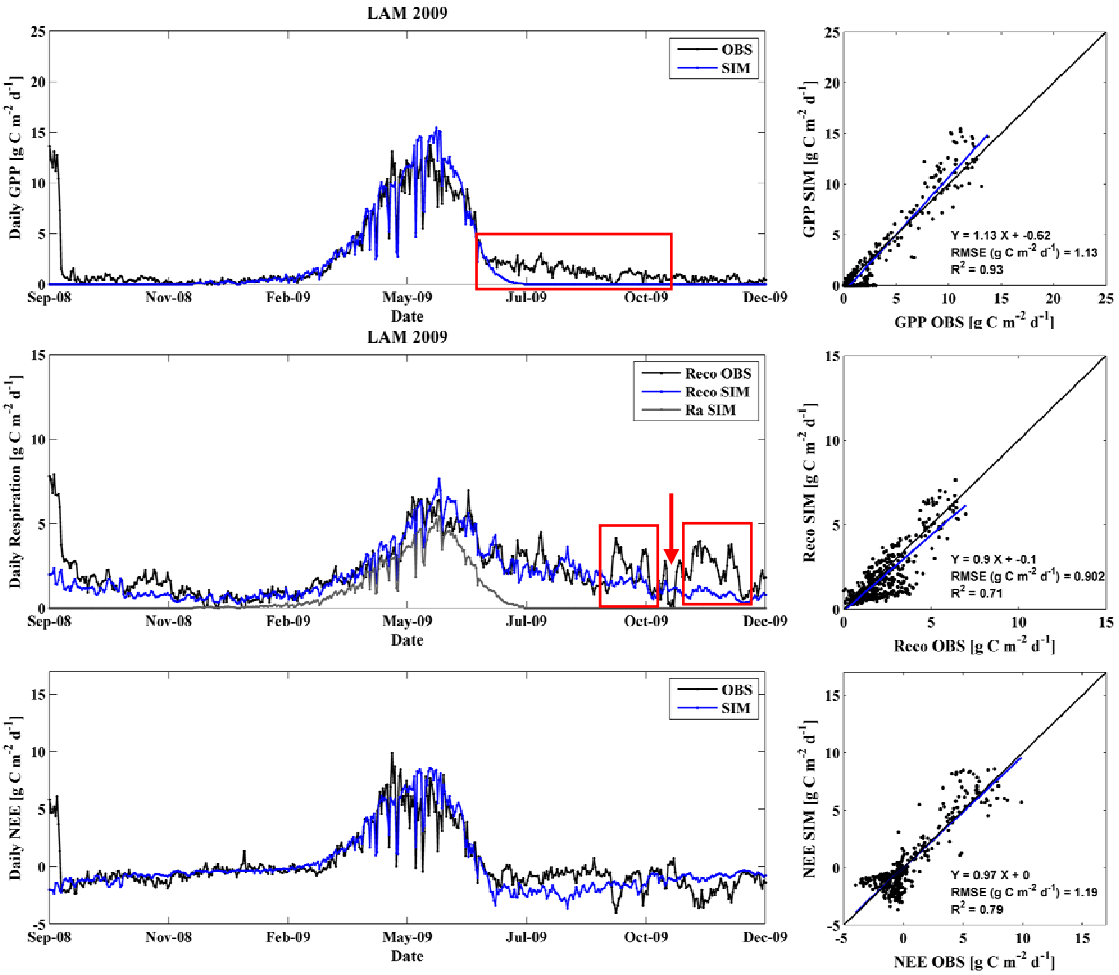


Figure 5.36- Evolution of the carbon fluxes for the Lamasquère site, 2008-2009 winter wheat crop season. Same legend as Figure 5.31.

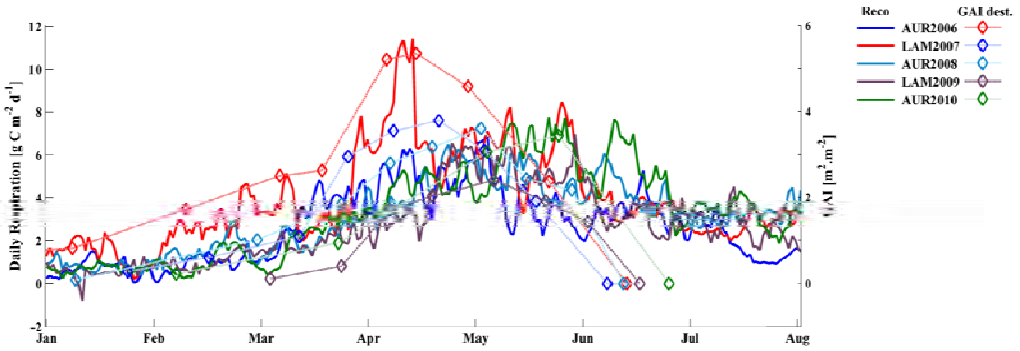


Figure 5.37- Comparison of the inter-annual dynamics of the observed ecosystem respiration (left axis) from 2006 to 2010 and the crop growth development, represented by destructive GAI measurements (right axis).

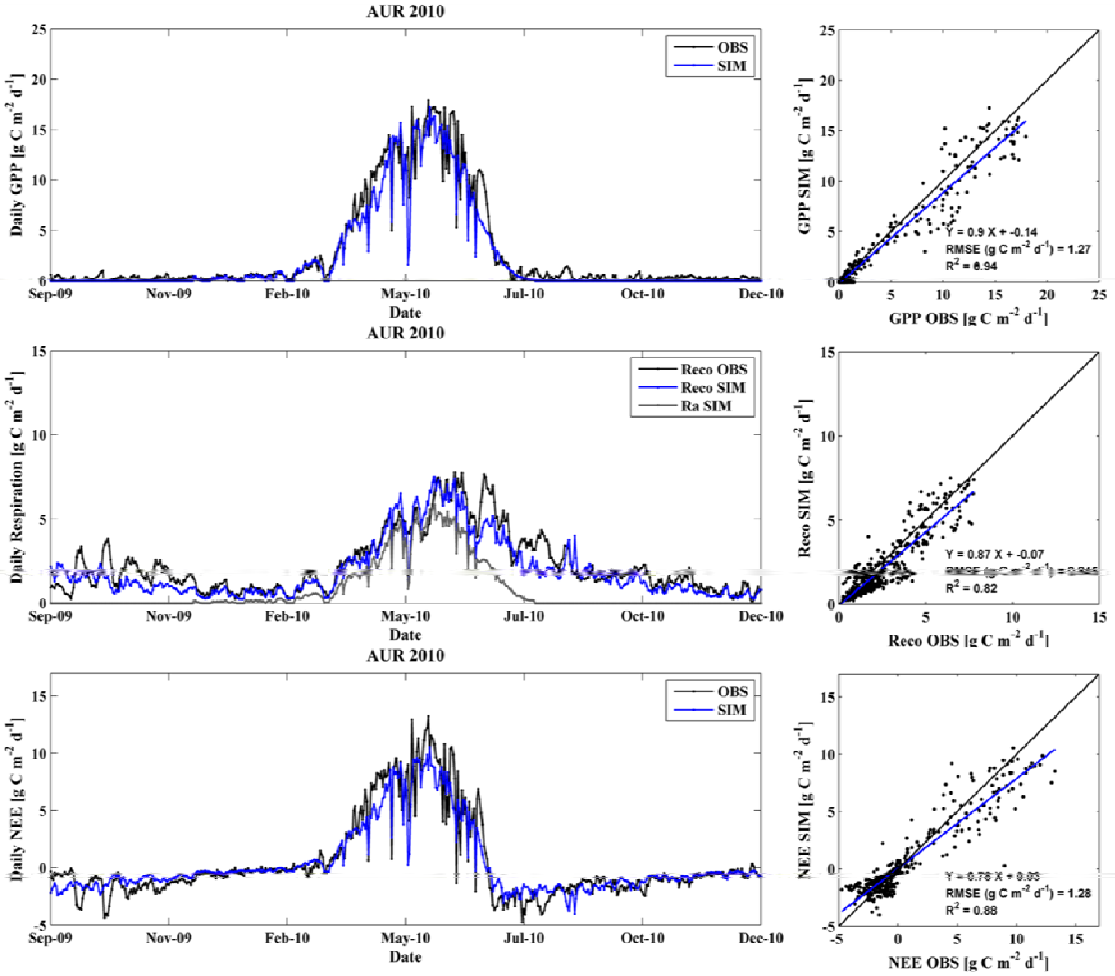


Figure 5.38- Evolution of the carbon fluxes for the Auradé site, 2009-2010 winter wheat crop season. Same legend as Figure 5.31.

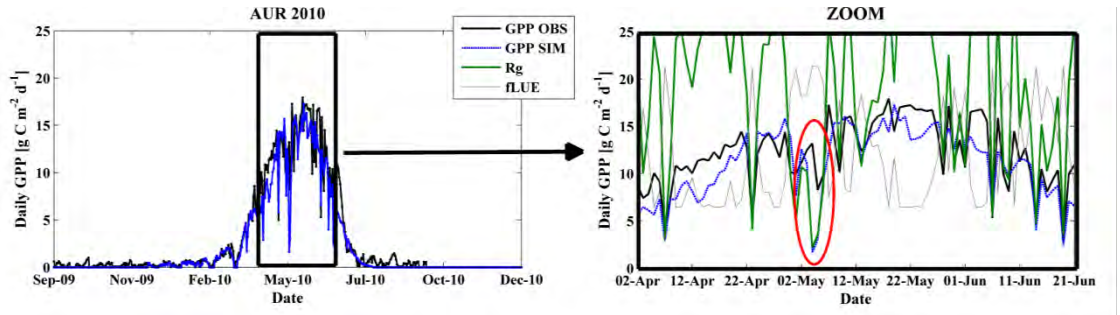


Figure 5.39-Zoom of the GPP dynamics for the Auradé site, 2009-2010 winter wheat season, with the global incoming radiation (Rg, in W.m^{-2}) and the *fELUE* function (x5, for visualization purposes).

5.3.1.1.1. Effect of diffuse fraction of global radiation over GPP estimation

In order to visualize the impact of the new included *fELUE* function over the GPP estimation with the SAFY-CO_2 model, we ran the model without the function, using only the regular effective light-use (ELUE) parameter.

Figure 5.34 shows the results obtained for AUR2006. We observe that the GPP estimates are noisier than before, varying more abruptly and thus presenting high variations in short periods of times. This effect is especially marked when the vegetation is well developed, *i.e*., when the GPP fluxes are stronger. By comparing the scatter plots of Figure 5.31 (with *fELUE*) and Figure 5.34 (no *fELUE*), and for other years (not shown), we conclude that the inclusion of a function that takes into account the effect of the fraction of diffuse radiation over canopy photosynthesis is valuable. By adapting the light-use efficiency of the canopy according to the irradiance conditions, we tend to have a better representation of the plant photosynthesis processes. Besides, it is essential to have an accurate estimation of the GPP fluxes, since they have an impact over the respiration (through biomass production) and thus the NEE fluxes estimates, and thus over the carbon budgets calculation.

Through this work we developed and tested our approach over winter wheat crops; next steps will consist in extending the method for other significant crops as maize and sunflower.

5.3.1.2. Ecosystem respiration estimates

Overall, we observe that the ecosystem respiration follows the same dynamics as GPP, but in an attenuated way; *i.e*., differences between bare soil and vegetative periods are smaller. Besides, simulated Reco matches usually well with observations (except in 2007), but in 2007, 2008 and 2010 after maximum growth development, the estimates seem to diverge from the observations. A strong Reco underestimation is thus observed in this period for those years. The obtained correlation coefficients are overall satisfactory: in average, R^2=0.78, with an exception in LAM2009, for which R^2=0.71 (Table 5.2). On the other hand, the obtained errors (RMSE) differ considerably depending on the year. For example, for AUR 2006 we obtain RMSE=0.80 gC $m^{-2}d^{-1}$, while for LAM2007, RMSE=1.2 gC $m^{-2}d^{-1}$.

In LAM2007 we observe a poor Reco estimation, significantly underestimated in comparison to the observations. This behaviour is aberrant if compared to those of other years. Figure 5.37 shows the *in-situ* observations of daily Reco dynamics for the 5 years-site of study. Together, we plot the dynamics of GAI, measured in the field. From this figure, we observe that the ecosystem respiration is stronger for LAM2007 (red line) than for the other years, especially in the beginning of the season until the crop maximum development. To investigate this effect, we evaluate

the correspondent GAI profiles. We note that the vegetation had a much faster development for LAM2007, reaching high GAI values earlier than usual. For example, in mid-March the LAM2007 site presented a GAI of 2.5 $m^2.m^{-2}$, while the others site-years had barely attained 1 $m^2.m^{-2}$ in the same period. The stronger Reco values in 2007 can then be explained by the fast (and strong) crop development, which can be justified by the extremely soft autumn and winter for this crop season. In the 2006-2007 season, warm temperatures were measured from the sowing (October/ November) until late March (see Figure 5.27(b)). This, jointly with relatively well distributed rain events, led to an earlier development of the wheat crop. Nevertheless, the model was not capable of reproducing this phenomenon correctly. As discussed for the GPP estimates, we remember it might be due also to low satellite-GAI values assimilation, inducing here lower autotrophic respiration (Ra) estimates. Furthermore, due to the described climatic condtions, there was a strong biomass production earlier than usual in the vegetative cycle, and consequently a high mortality rate and turnover of the plant organs after tillering, phenomenon that is not taken into account by the destructive measurements of (alive) biomass. Thus, the increase of the ecosystem respiration might be due to an early decomposition process of the senescent and death plant organs. This experience was observed for the Lonzee site at the same year (Aubinet et al., 2009). Besides, since Reco is the sum of plant respiration (Ra) plus the the heterotrophic respiration (Rh), the Reco underestimation may be partly because in SAFY-CO_2 model, Rh depends only on the soil temperature, and not on the soil humidity. Thus rainfall events that can contribute to an increase of soil micro-organisms activity are not taken into account. Nevertheless, in general, the Ra simulated dynamics seem correct (except for the two pics in 2007). According to the model simulations, the Ra represents the major part of the Reco during the vegetative period, which is in agreement with which was observed by Moureaux (2008) and Aubinet et al., (2009).

For AUR2006 (Figure 5.31), at the end of the growth periods (May to June) the simulated Reco is superior to the observations probably because of the slightly overestimated GPP. For AUR2008 (Figure 5.35), a very light overestimation can be observed during the vegetative period, but at the end of the season the simulated Reco is greatly inferior to the observations. For LAM2009 (Figure 5.36), the Reco estimates are overall in good agreement with the observations during the crop growth cycle, except in June, when Reco was underestimated. However, we observe that, from September, there is an increase of the observed Reco. It corresponds to manure application. After, in mid-October, a sudden Reco decrease can be noted, corresponding to soil labor, being followed by a Reco raise, caused by since the manure incorporation into the soil (which augments the soil respiration). For AUR2010 (Figure 5.38), the Reco provided by the model is in good agreement with

the observations, except from mid-September till November (Reco is underestimated) and for a Reco peak during senescence that is not reproduced by model. It is due to an increase of the Rh just after a rainy period.

In general, during the bare soil periods (or intercropping), two different behaviours can be observed. First, before the crop seeding: Reco values are relatively low and the model estimates agree well with the observations. During these periods, the Reco is equivalent to Rh (since there is no autotrophic respiration), showing that the model is, in general, capable of simulating correct estimates of Rh by means of a simple temperature-based exponential function. Second, for the period after harvesting: the Reco observations present some peaks, whereas the model esimates present little variations, due to temperature conditions. These peaks in observations can be explained by rainfall events, soil labour, residue and organic manure incorporation into the soil and re-growth (Béziat et al., 2009). All these factors are not (directly) taken into account by the SAFY-CO_2 modeling approach.

5.3.1.3. NEE estimates

In general, the daily dynamics of net ecosystem exchange (NEE) are well simulated by the SAFY-CO_2 model (see Figure 5.31,Figure 5.32,Figure 5.35,Figure 5.36,Figure 5.38, bottom). The average statistics indicators, calculated over the 5 site-years, indicate a correlation coefficient (R^2) of 0.84 and an error of 1.2 gC $m^{-2}d^{-1}$ (Table 5.2). The visual analysis of the simulated and observed NEE dynamics shows that they are coherent with the GPP dynamics; besides, some of the unusual behaviors noted for the GPP estimates can also be found in the NEE estimates. For example, the re-growth observed in the GPP profile in early September 2006 (Figure 5.31), can be clearly seen in the NEE profile. Since NEE represents the integration of ecosystem incoming CO_2 (via photosynthesis=GPP) and the leaving CO_2 (via respiration =Reco), some effects present on NEE measurements can be either compensated or accentuated on the GPP and/or Reco calculated during the partitioning process.

In 2006 (Figure 5.31, bottom) the NEE dynamic is overall good, but the model underestimates NEE (negative values) from July till October because the GPP of the re-growth is not accounted for by the model. In 2007 (Figure 5.32, bottom), the simulated NEE dynamic is in agreement with the measurements, even if the GPP and the Reco are underestimated. It happens because those two underestimations compensate each other. It can also be noted the small drops of NEE in winter because GPP=0 during the two cold events. In 2008 (Figure 5.35, bottom) good performances are obtained, except for a divergence in August, because Rh is underestimated by the model, and also after October, because the observed GPP is

not accounted for by the model. In 2009 (Figure 5.36, bottom), it can be noted that after the crop development period the quality of the simulated NEE decreases. From July to September, the NEE underestimation (NEE is too negative) is linked to the GPP estimates (equal to zero), while observations remain at low values. From October till early December, the NEE mismatch is due to the Rh underestimated by the model (manure application and incorporation into the soil). In 2010 (Figure 5.38, bottom), we observe that the NEE is underestimated, mostly during the crop growth periods, as it can be seen in the scatter plot (on the right) : the higher values of NEE (between 5 and 15 gC $m^{-2}day^{-1}$) are underestimated by the model (confirmed by the regression slope of 0.78). This tendency is essentially due to the GPP underestimation over the same period.

Figure 5.40 shows the cumulated NEE values measured by the flux towers and those estimated by the SAFY-CO_2.model. They are cumulated for the period of 1st September to 31 December of the following year. The analysis of the cumulated NEE allows the identification of phases of carbon storage or carbon release. Here we use the micrometeorological (or atmospheric) convention, with NEE negative when the ecosystem is fixing carbon (flux moving downward) and positive when it is losing carbon (flux moving upward). A negative slope on the cumulated NEE curve means that the ecosystem behaves as a carbon sink (GPP>Reco), and a positive slope means that the ecosystem behaves as a carbon source (GPP<Reco). The NEE dynamics reveal that, depending on the year, different crop developments can be noted. For AUR2006 we observe that the NEE cumulated value became negative in early April, whereas it happens a month earlier for LAM2007 (1st March), and a month later for LAM2009 and AUR2010 (around 21th April). As it was previously discussed for the GPP, in 2007 the wheat had an earlier development that year, which can also be clearly seen in the NEE cumulated values. This means that for different site-years the crop ecosystem started storing carbon at different times.

We note that the maximum absolute cumulated NEE values can be very different from one year to another; and also, the time in the season at which this maximum is reached. The 2007 and 2008 seasons are the ones that attained the higher carbon storage (around 526 gC m^{-2}), in early June. For LAM2007, it is explained by the fact that the crop growth has started earlier than usual, due to the favorable climatic conditions, and thus lasted longer. In AUR2008, the wheat developed well, reaching a maximum GAI around 4 $m^2.m^{-2}$. The LAM2009 season presents the lowest rate of carbon storage during the crop growth, achieving only 314 gC m^{-2} as maximum absolute NEE. Given that the emergence was low and delayed, because of the excess of water in the soil (high precipitation), the plants had a shorter period for assimilating carbon and the net assimilation was less intense due to lower plant density.

The final period of carbon release (positive slope on the cumulated NEE curve) is characterized by the change of sign in NEE, which becomes positive; it means that the Reco is superior to the GPP fluxes, and that the ecosystem changes from a sink to a carbon source. It corresponds to the senescence phase (decrease of green photosynthetic elements), passing to harvesting and then to the bare soil period.

The visual comparison between the cumulated NEE derived from the flux tower measurements and those provided by SAFY-CO_2 indicates that they match well, with a slight degradation in the end of the season. From these results we observe that the model is capable of reproducing the NEE evolution for different years and sites; it means that the model could simulate the NEE evolution for winter wheat crops under different management conditions, soil properties and for contrasted climatic years. A very good correlation coefficient (around 0.98) was obtained for all the investigated cases, and a good efficiency (Nash criteria) as well. However the associated relative errors, expressed in terms of RRMSE, vary from 23% (LAM2007) up to 67% (LAM2009). Since it is an integrative variable, small deviations through all the cycle may yield to important divergences at the end of the season. Also those good results are partly obtained because both GPP and Reco are underestimated at some stages for most years.

For AUR2006, we observe the model started the carbon sequestration a bit later than the observations and due to a slight overestimation of GPP in the final development period, the simulated cumulated NEE reaches the maximum absolute value at the same time as observations. Still, after mid-July (corresponding to harvest) the simulations and observations deviate because of the remaining photosynthesis activity in the observations due to the crop re-growth and weed presence (encircled in red in Figure 5.40). The same behavior can be observed for AUR2008 and LAM2009. For AUR2008, we observe that the estimates and observations match perfectly well until early March, and then they diverge slightly because the simulated carbon assimilation by the vegetation is a little underestimated. Besides, from early August the observed cumulated NEE changes its dynamics due to, first, an increase of the ecosystem respiration dynamics, probably due to soil labor and, second, due to the sowing of the following crop (rapeseed) that happened the 4th September (highlighted by a red circle). As it can be noted, these processes are not reproduced by the model.

For LAM2007 and AUR2010 we note a shift between model estimates and observations. For both years, the change of slope of the cumulated NEE happens approximately at the same time for estimates and observations; however, due to a small underestimation of the daily NEE estimates during the growing period, the model started the period of carbon release with a lower quantity of net carbon assimilated by the plot. Furthermore, for LAM2007 it is interesting to note that the

episode of simulated GPP equal to zero (due to extreme temperatures) can also be seen on the cumulated NEE profile (indicated by the red arrows in Figure 5.40).

Besides, since the heterotrophic respiration (Rh) drives the NEE estimates in periods of bare soil, the shift in the end of the season can be partly due to the modeling of the Rh in SAFY-CO_2. As it was observed in the daily dynamics of simulated respiration, a frequent mismatch with the observations was present at the end of the season (when Reco=Rh). Future works should improve the modeling of the heterotrophic respiration. As a start, an empirical function that depends not only upon soil temperature (as it is currently done) but also upon soil humidity could be used. The coupling of SAFY-CO_2 with a water budget module based on the FAO-56 would allow that by simulating new variables such as soil water content (and evapotranspiration), in addition to those already investigated here. Another potential option would be to couple SAFY-CO_2 with a more complex model that is able to simulate the evolution of the organic matter content in the soil (such as CENTURY). Finally, we should keep in mind that NEE is the sole variable that is directly measured by the flux towers instruments, being a 'true' observation; the GPP and Reco variables are estimated from the NEE measurements by means of a partitioning method (Béziat et al., 2009). Therefore, the comparison of the model estimates with these GPP and Reco observations should be carried out with caution, since they may contain artifacts due to the partitioning method (Reichstein et al., 2005).

5.3.1. Crop growth and production: GAI, DAM and Yield estimates

Figure 5.41 presents the GAI profiles simulated by the SAFY-CO_2 (in blue) model and the satellite GAI (black circles) on the left, and the simulated biomass (in blue) compared with destructive measurements, and associated standard deviations (in black) for the experimental sites, from 2006 to 2011. We observe that, in general, the temporal evolution of both GAI and biomass simulated by SAFY-CO_2 are in good agreement with the measurements. For AUR2008 and LAM2011 we note a marked underestimation of the final biomass; still the temporal dynamics of the estimates match with the observations. Since the biomass production depends upon the simulated photosynthesis (GPP), this underestimation is probably due to a GPP underestimation, mostly at the end of the development phase as it could be seen for 2008 in Figure 5.35. Unfortunately, as the 2011 flux data could not be used for analyzing the models outputs because of a gap of 2 months in the data in the beginning of the year, it is difficult to conclude on the causes of this biomass underestimation that year. In addition, the underestimated final biomass could also

be caused by an underestimation of the satellite-derived GAI in 2011: as it can be seen in Figure 5.41, for this year the destructive measurements of GAI are considerably stronger than those derived from satellite observations during the period of full crop development.

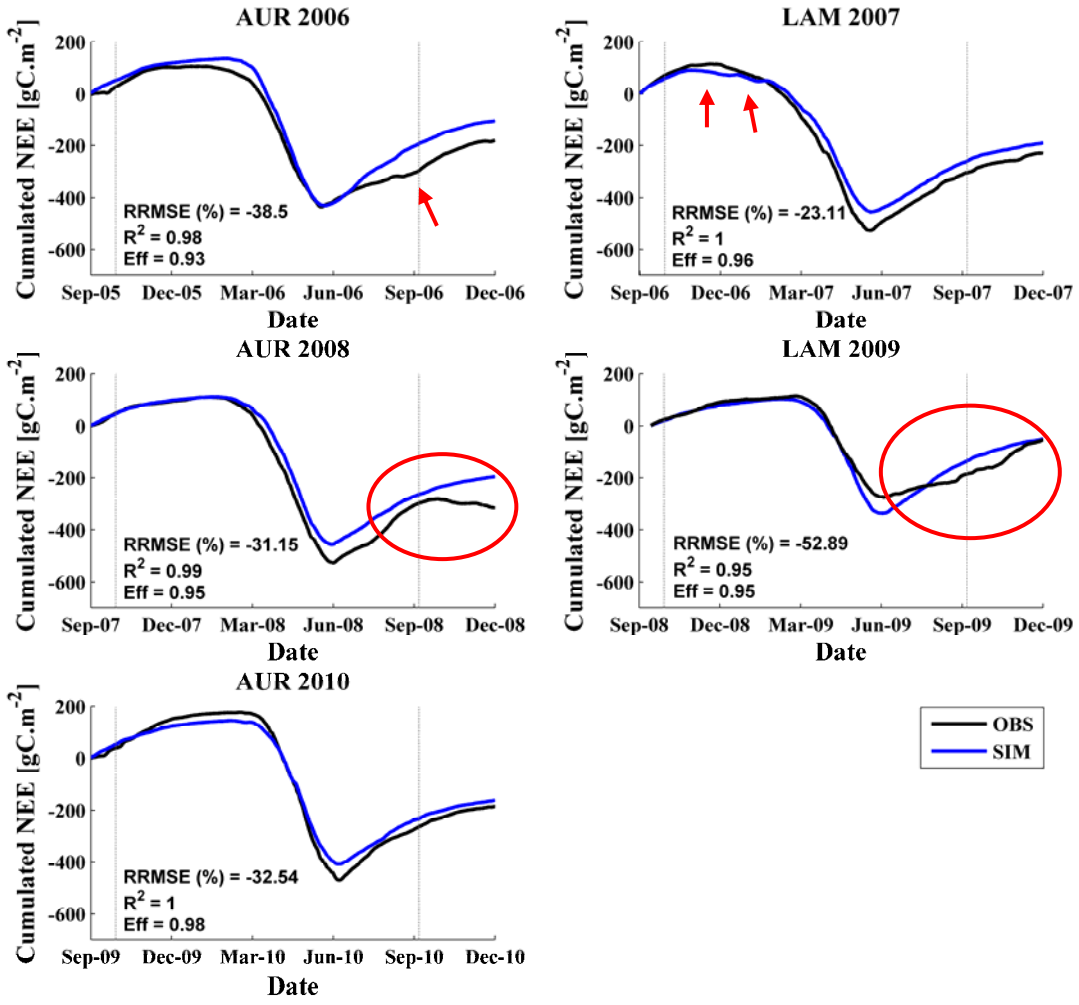


Figure 5.40- Cumulated values of NEE for the 5 site-years: AUR2006, LAM2007, AUR2008, LAM2009 and AUR2010. The observations are displayed in black and the model outputs are in blue. The performances, in terms of relative mean square error (RRMSE), linear coefficient of correlation (R^2) and Nash criterion (Eff) are shown in the left bottom corner. The grey vertical bars indicate the period when the NEE fluxes are integrated for estimating the net ecosystem production (NEP, used for calculating the net ecosystem carbon budgets). Particular events encircled or indicated by the red arrows are commented in the text.

Figure 5.42 and Figure 5.43 illustrate the scatter plots of observations against SAFY-CO_2 estimates of GAI and biomass for the 'experimental sites' and the '2011 field campaign' data sets together. The analysis of these figures reveals that the SAFY-CO_2 model provides accurate estimates of the GAI, with RRMSE of 11.5% and R^2=0.98, and of the biomass as well, with RRMSE of 23.7% and R^2=0.91. However, the high values of biomass are underestimated. Those values correspond mostly to the 2008 and 2011 data on the experimental sites.

The results of the yield estimates provided by the SAFY-CO_2 model (using the four investigated methods) over the '2011 field campaign' data set are displayed in Figure 5.44. The performances vary according to the chosen method. The methods 1 and 4 present nearly equivalent performances, with a relative error of about 23% and a coefficient correlation of 0.74. We remind that these methods depend upon a constant harvest index. On the other hand, the methods that use a dynamic harvest index for estimating yield (methods 2 and 3) present poorer performances, but correct as well. The methods 2 and 3 present here a stronger bias, but better relative errors; even though, the estimates and observations keep good correlation (around R^2=0.65). The inclusion of the method 3 in our modeling approach, by estimating yield from the biomass production and a harvest index that evolves with a thermal time function, is expected to improve the yield estimates. Despite this method provides the lowest RRMSE of the four methods (22.3%), it presents an important bias. This might be due to an inadequate threshold for starting grain filling (STY), here fixed to 796°C, which can induce, for example, an early beginning of grain production, and thus a stronger final grain yield. Besides, the parameter VITICARBT, which is also part of the yield function, was set equal to the value present in the STICS model, without further examination. In order to achieve better and more robust estimates of yield, taking into account the thermal evolution during the season, we recommend further investigations for establishing enhanced approaches for setting optimal STY and VITCARBT parameters.

The analysis of the yield estimates for the experimental sites, from 2006 to 2011 (Figure 5.45), shows that, in general, the same conclusions drawn earlier are valid for this data set as well. However, we note that for the experimental sites data set, an almost null correlation is found between estimates and observations, independently of the chosen method. This might be partly due to the few available points (only six) and because the range of observation is much smaller than that of the data set from the 2011 campaign, making it difficult to obtain a proper validation of the modeled yield. Other possible reasons were earlier discussed throughout the SAFY model results presentation. It is important to note that method 4 is the one that provides the best grain yield estimates for the experimental sites data set, with a RRMSE of

11.97%. Therefore we will use further the yield results from method 4 for the computation of the annual crop carbon budgets of the experimental sites.

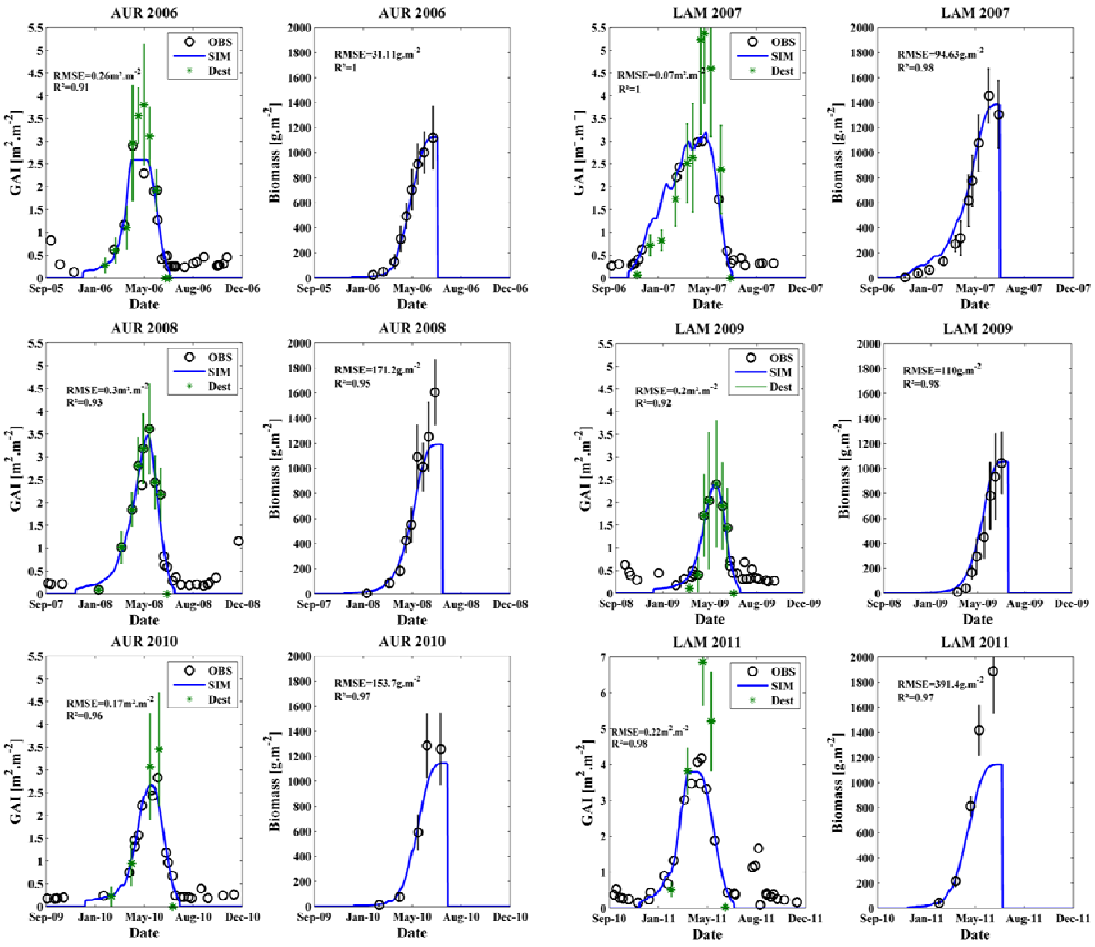


Figure 5.41- GAI and biomass simulated by the SAFY-CO_2 model (blue line) for the Auradé site (2006, 2008, 2010) on the left, and Lamasquère site (2007, 2009, 2011), on the right. The GAI observed by remote sensing is represented by black circles and GAI from destructive measurements by green stars. The destructive measurements of biomass are represented by black circles.

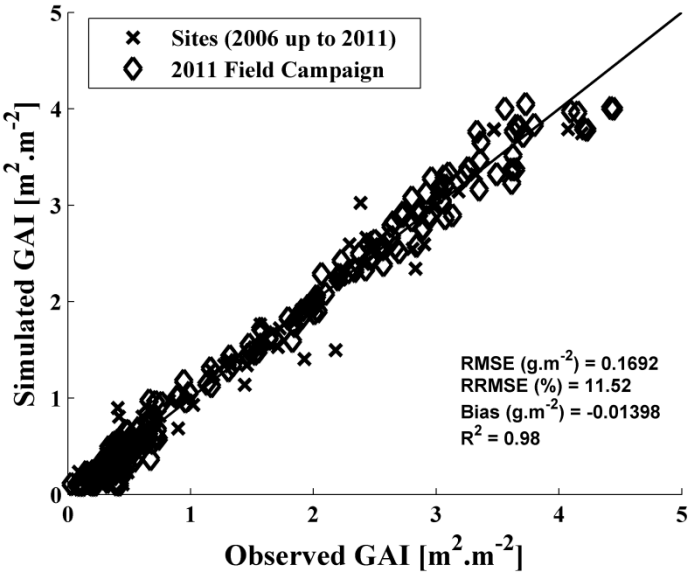


Figure 5.42- Comparison between the green area index simulated by the SAFY-CO_2 model and the satellite-derived GAI (observation) for the 'experimental sites' group (represented by the x symbol) and '2011 field campaign' set (represented by the diamond symbol). The line 1x1 is displayed. The performances (RMSE, RRMSE, Bias and R^2) are indicated on the right corner.

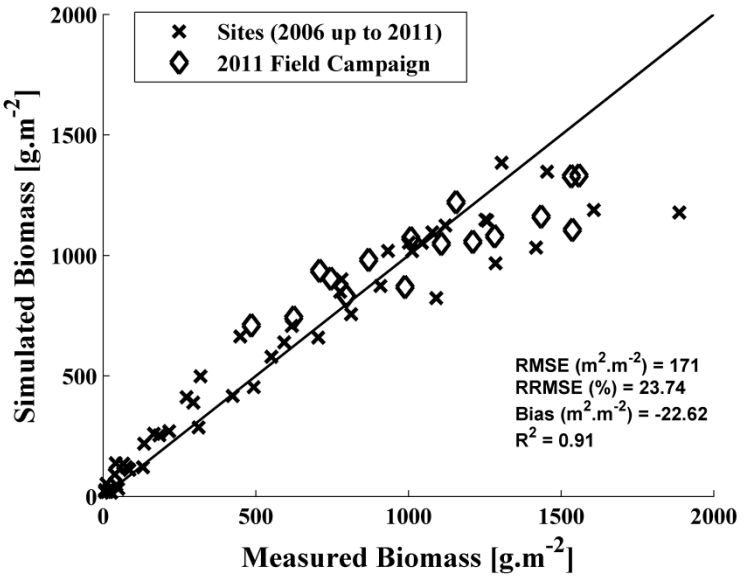


Figure 5.43- Comparison between biomass simulated by the SAFY-CO_2 model and ground measurements for both data sets: the 'experimental sites' group (represented by the x symbol) and '2011 field campaign' set (represented by the diamond symbol). The line 1x1 is displayed. The performances (RMSE, RRMSE, Bias and R^2) are indicated on the right corner.

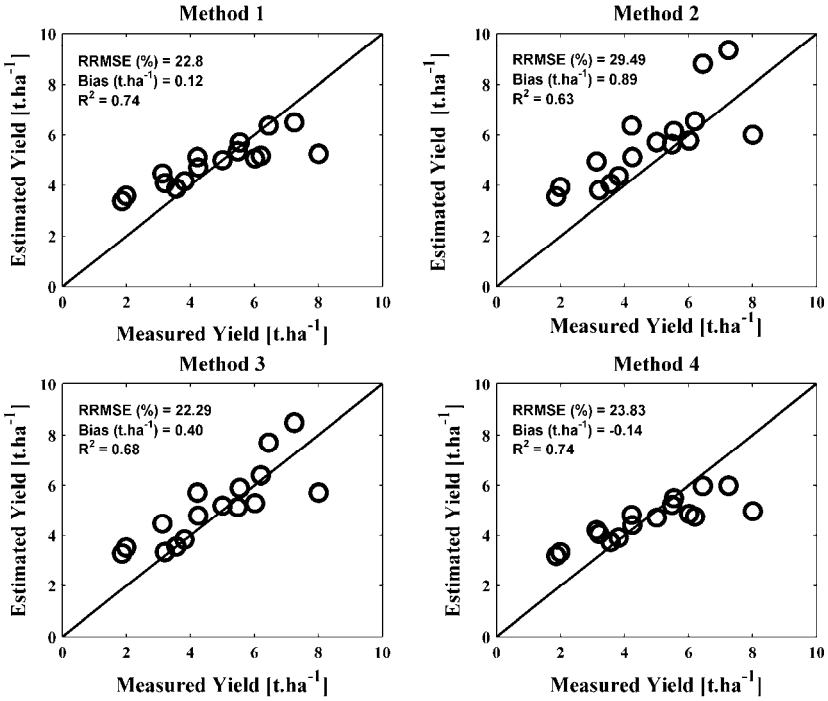


Figure 5.44- Comparison of the methods for estimating yield with the SAFY-CO_2 model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the 2011 campaign fields.

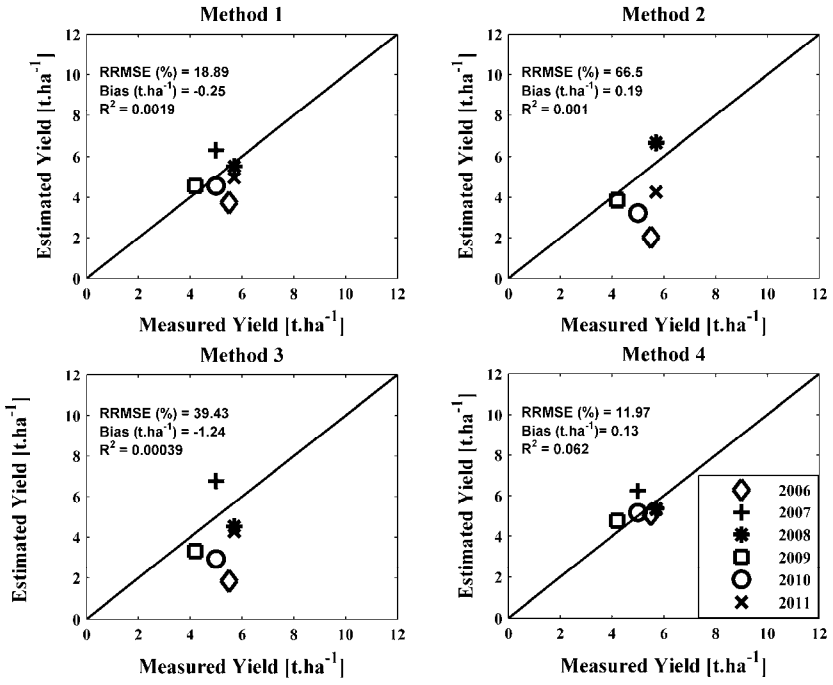


Figure 5.45- Comparison of the methods for estimating yield with the SAFY-CO_2 model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the experimental sites (AUR2006, AUR2008, AUR2010, LAM2007, LAM2009 and LAM2011).

5.3.2. Crop carbon budget

The annual crop carbon budget was calculated for the five studied sites-years. The net ecosystem carbon budget (NECB) estimates, and their components, were compared to those derived from the flux towers measurements and from farmers inventories (Figure 5.46).

We remind that the daily NEE simulated values were integrated from early October to late September, to establish the net ecosystem production (NEP). It is compared with the NEP derived from the flux tower measurements, integrated over the same period. The OF (organic fertilization) term is the same for the model and observations, since the model cannot simulate this term. These OF values were provided by the farmers. The observed Cexp term was also computed according to farmers' information.

The simulated NEP ranged from -153 gC $m^{-2}y^{-1}$ (LAM2009) to -318 gC $m^{-2}y^{-1}$ (LAM2007). This range is in partial agreement with that of the observations: from -208 gC $m^{-2}y^{-1}$ (LAM2009) to -374 gC $m^{-2}y^{-1}$ (LAM2007). In general, we observe that the NEP estimates provided by SAFY-CO_2 are close to the observations, despite a slight underestimation tendency (simulated NEP are less negative than the observations). The minimal divergence is observed for AUR2010, for which the simulated NEP is only 6% lower than the observation. For AUR2006, the model simulates a NEP of -244 gC $m^{-2}y^{-1}$ against a value of -323 gC $m^{-2}y^{-1}$ computed from field measurements (24% of difference). This difference is mainly due to the effect of the re-growth event over the NEE observations that are not taken into account by the model. The same phenomenon caused a 26% difference between observations and simulations for LAM2009. The LAM2007 results present the highest absolute NEP value and relatively low underestimation of 15%. As it was discussed earlier, despite the strong underestimation of the Reco by the model, and also a given underestimation of the GPP, at the end we obtain reasonable estimates of the NEE and then of the NEP. This double underestimation effect over NEP is also observed for AUR2010, where the estimated NEP is 6% lower than the observed NEP. For AUR2008, we note an underestimation of 10% of the simulated NEP compared to the observations (-310 against -346 gC $m^{-2}y^{-1}$). For this year, despite the good agreement between the NEE daily dynamics (Figure 5.40), a divergence appears in late September probably due to the emergence of the subsequent rapeseed crop, present in the measurements but not simulated by the model.

From those results, we infer that the NEP differences between the two studied sites over different years are mostly due to climatic variations (radiation, temperature and precipitation) and to management practices. These climatic conditions may affect the length of the cropping season, *i.e.* the length of the carbon

assimilation period that directly influences the NEP estimations, but it can also be influenced by a re-growth event or the development of weeds that are also driven by climatic factors. The effect of climatic conditions could be clearly seen for the Lamasquère site, which presented contrasted behaviors in 2007 and 2009. The crop growing season length was particularly long in 2007 because of the soft winter (inducing a very negative NEP value), while in 2009 the crop emergence occurred particularly late, thus the length of the cropping season was shorter (inducing low NEP value). In addition, in 2009, the emergence was poor, inducing low cover density, thus low GAI, thus low GPP that finally induces low NEP. Besides, this comparison demonstrates that long periods of bare soil can have an important effect on the annual carbon storage by the carbon release through heterotrophic respiration. The SAFY-CO_2 model has proven to be capable of reproducing most of these diverse behaviors.

The analysis of the annual amount of carbon exported by the ecosystem (Cexp) shows that it varies considerably from year to year. The simulated Cexp range from 232 gC $m^{-2}y^{-1}$ (AUR2006) to 388 gC $m^{-2}y^{-1}$ (LAM2007). The simulated dynamics are similar to those of observations, for which Cexp ranged from 205 gC $m^{-2}y^{-1}$ (AUR2010) to 336 gC $m^{-2}y^{-1}$ (LAM2007). The model tends to overestimate the exported amount of carbon. The differences between observations and model estimates of Cexp vary between 4.2% and 16%; these differences are inferior to those obtained for the NEP (between 6% and 26%). Still, the comparison of model Cexp estimates and observations should be performed with prudence. For both sites, the values of observed grain yield (component of Cexp) are provided by the farmers. The precision of this information is nonetheless questionable, since it might be relative to several fields cultivated by the same farmer, and not field-specific information. For the Lamasquère site, the uncertainty of the Cexp term is even more important since most part of the straw is exported.

In addition, we observe that the Lamasquère site presents stronger carbon export terms than Auradé does, in absolute values. The difference in Cexp between the two sites is linked to differences in straw management. As stated before, the Lamasquère is a livestock breeding farm exporting most of the aboveground biomass for cattle bedding and feeding, whereas Auradé exports only the grains, for being a cereal production farm.

From the NEP, OF and Cexp terms, the net ecosystem carbon budgets (NECB) were computed. The NECB estimated by SAFY-CO_2 characterize all the crop seasons as carbon sinks, except for LAM2009. The annual NECB estimates range from -70 gC $m^{-2}y^{-1}$ (AUR2008) to +70 gC $m^{-2}y^{-1}$ (LAM2009). Meanwhile, the NECB measurements range from -125 gC $m^{-2}y^{-1}$ (LAM2007) to 0 gC $m^{-2}y^{-1}$ (LAM2009). The Lamasquère site presents the worse estimated NECB: in 2007 a carbon sink of -16 gC $m^{-2}y^{-1}$ and in

2009 a carbon source of 70 gC $m^{-2}y^{-1}$. The carbon inputs (NEP+OF) of the Lamasquère site are stronger than those of Auradé, because of the important contribution of organic fertilization; however, given that most part of the biomass is exported (grains + straws), the final carbon budget may be close to neutrality (LAM2007) or represent a carbon source (LAM2009). For the Auradé site, the three wheat seasons (2006, 2008 and 2010) are carbon sinks, with estimated annual carbon storages of -18, -70 and -52 gC $m^{-2}y^{-1}$, respectively.

Finally, according to the model outputs, the NEP are overall underestimated and the Cexp are generally overestimated, resulting in cumulated errors of the NECB estimates. Since the under and over-estimations of both terms do not compensate each other, stronger divergences between NECB estimates and observations are found. As it can be seen in the Figure 5.46 the relative differences between modeled and in-situ NECB range from 40.3% (AUR2008) up to 1530% (LAM2009). In absolute terms, these differences range from 48 gC $m^{-2}y^{-1}$ (AUR2008) up to 108 gC $m^{-2}y^{-1}$ (LAM2007). These divergences highlight the importance of having accurate estimates of each of the terms that compose the NECB (NEP, OF and Cexp) for obtaining precise estimations of the crop carbon budgets.

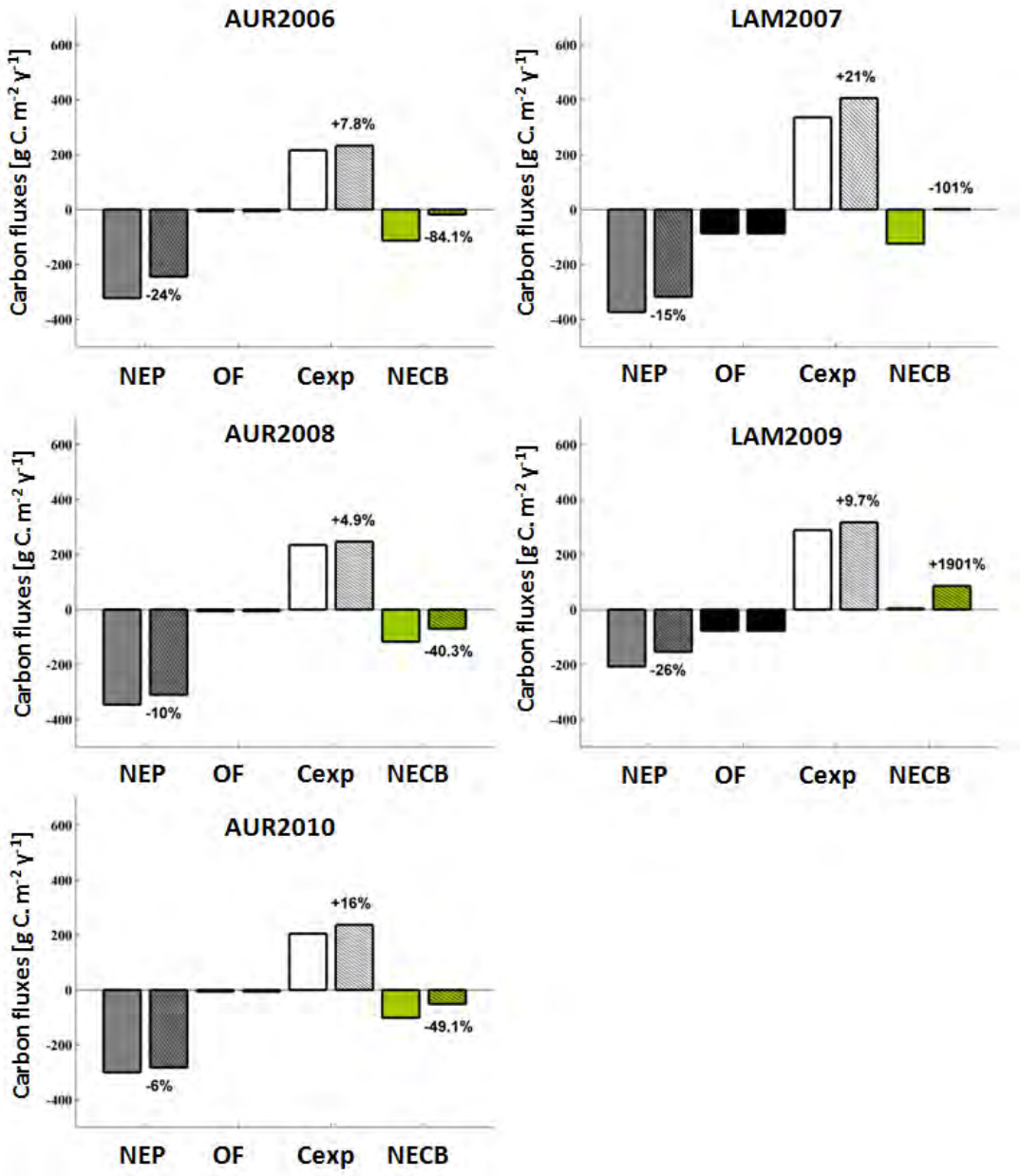


Figure 5.46- The annual carbon budget for the winter wheat crops, at Auradé (2006, 2008, 2010) and Lamasquère (2007, 2009) sites. The first bar of each pair represents the measurements and the second bar, hatched, represents the results provided by the SAFY-CO_2 model. The grey bars represent the net ecosystem production (NEP), the black bars represent the organic fertilization (OF), the white ones: the part of the crop biomass that is exported and the green bars represent the net ecosystem carbon budget (NECB).

5.4. The SAFYE-CO_2 model results and discussions

In this section we present the results obtained with the SAFYE-CO_2 model, which was described in Chapter 4.3. The SAFYE-CO_2 model provides the same outputs as SAFY-CO_2 (seasonal dynamics of the components of the carbon budgets: such as GPP, Ra, Rh, NEE, and of GAI, biomass and yield) plus the outputs related to the water balance: evapotranspiration (ETR) and soil water content (SWC).

The main differences between the two models are that, for SAFYE-CO_2: 1) the water stress is directly taken into account in the GPP estimates; 2) the heterotrophic respiration is a function of the superficial soil water content in addition to the standard soil temperature.

The simulated CO_2 and water fluxes are compared with the *in-situ* flux tower measurements, estimated using the Eddy-Covariance method at the Auradé and Lamasquère sites. The simulated soil water contents are compared with soil measurements performed on each site at depths of 0.05, 0.10, 0.30 and 1 m.

5.4.1. ETR and SWC estimates

The dynamics of simulated evapotranspiration (ETR) were compared to those measured at the Auradé and Lamasquère sites. As presented in section 4.3.1, the data of the site-years LAM2007 and AUR2008 were used for calibrating the parameters of the water balance module, and the other three site-years data sets (AUR2006, LAM2009, AUR2010) for validation (Figure 5.47). Table 5.3 summarizes the obtained performances, in terms of root mean square error (RMSE), correlation (R^2) and efficiency (Eff).

In general, we observe that the estimated ETR dynamics are in agreement with the measured ones, presenting a RMSE around 0.46 $mm.day^{-1}$, correlation coefficient (R^2) around 0.74 and efficiency (Eff) around 0.66. From the ETR dynamics, it is possible to identify the crop growth period, since they are similar to the GAI dynamics. We observe that in periods of bare soil: ETR is lower, since only the evaporation term contributes to the ETR estimation (no plant transpiration). Indeed, during the vegetation cycle, a gradual increase of the ETR (mainly because of plant transpiration) can be noted, which is directly related to the crop development (Béziat et al., 2013).

For LAM2007 (Figure 5.47, top), we observe that the evapotranspiration is well simulated, during periods of bare soil or low vegetation. As it was observed for the GPP dynamics, the ETR is underestimated by SAFYE-CO_2 over the crop growth maximal development phase. As previously discussed, it is due to the early development of the wheat this year because of the particular climatic conditions,

which led to very high true (destructive) GAI values at full crop development. This underestimation of higher values of ETR is clearly seen in the scatter plot (ETR_{SIM} *vs* $ETR_{OBS,}$ on the right), and is also confirmed by the regression slope of 0.74. In August, we can note a divergence between the model and the observations (enclosed in red), as the simulated ETR abruptly increase and the observations remain low. This increase in the simulated ETR corresponds to an increase in the reference evapotranspiration (ET_0), *i.e.* a strong increase of the climatic evaporative demand. It means that the model responded more intensively to the ET_0 variation than the measurements. During this period, the superficial SWC was overestimated by the model (Figure 5.50) which led to overestimated simulated evaporation.

For AUR2008 (Figure 5.47, 2nd), the simulated and observed ETR dynamics are quite well correlated (R^2=0.78) and present a low mean divergence (RMSE=0.42 $mm.day^{-1}$). A very slight overestimation can be noted around May (affecting the regression slope of 1.03), but it is not significant. LAM2007 and AUR2008 are the two data sets that were used for the calibration of the SAFYE-CO_2 water balance module parameters. As expected, the established set of optimized parameters yields to good results for both years.

The data sets of 2006, 2009 and 2010 constitute the 'validation' data set. For AUR2006 (Figure 5.47, 3rd), we observe an overall agreement between simulated and measured dynamics of ETR. However, as for LAM207, the ETR estimates are underestimated during the period of maximal development (from April until end of May). This is explained by low transpiration estimates at high GAI, which are related to the Kcb parameters. The results suggests that the value found for the Kcb_{max} parameter through the calibration process (Kcb_{max} =0.98) is a lit bit low for the AUR2006 site. Two other particular divergences can be noted (enclosed in red): one in November (just after sowing), that corresponds to a strong rain event; the second occurs in October (after harvest), when the observed ETR are stronger than the simulated by SAFYE-CO_2. This divergence is probably caused by the crop re-growth development this year. Still, a good correlation coefficient (R^2=0.75) and a low root mean square error (RMSE= 0.47 $mm.day^{-1}$) are found.

For LAM2009 (Figure 5.47, 4th), an overall good agreement is obtained between observations and model estimates. A slightly underestimation tendency is shown by the analysis of the scatter plot (on the right), however, contrarily to the LAM2007 and AUR2006 years, the underestimation is not particularly concentrated during the periods of high plant development, but scattered during the cycle. A short peak in the simulations was observed in August that year, but we could not find a reason for this divergence compared to the observations.

Finally, for AUR2010 (Figure 5.47, bottom), despite the overall good results (R^2=0.75 and RMSE= 0.50 $mm.day^{-1}$), an overestimation tendency by the model is

observed (regression slope =1.1), which was not observed for the others years. It can be noted that in June, three peaks of ETR are simulated by the model and are not present on the measurements. These peaks correspond to peaks in the reference evapotranspiration (ET_0). It means that the model was more sensitive to ET_0 extreme values than the measurements.

From these results, we can conclude that in terms of ETR dynamics the calibration obtained using ETR data from two different sites and two contrasted climatic years (LAM2007 and AUR2008) yields to rather accurate ETR estimates.

Table 5.3- Performances of the water balance components (ETR and SWC) dynamics estimated by the SAFYE-CO_2 model for winter wheat over the experimental sites, from 2006 to 2010. For each year, the statistical indicators are calculated for the data collected from September until December of the following year.

	ETR			SWC		
	RMSE*	**R^2**	**Eff**	**RMSE°**	**R^2**	**Eff**
LAM2007	0.52	0.73	0.73	0.08	0.30	-2
AUR2008	0.42	0.78	0.69	0.06	0.31	-1.7
AUR2006	0.47	0.75	0.67	0.05	0.43	-0.9
LAM2009	0.46	0.71	0.68	0.05	0.64	-0.9
AUR2010	0.50	0.75	0.55	0.04	0.67	0.47

* RMSE in $mm.day^{-1}$; ° RMSE in $m^3.m^{-3}$.

In order to calculate the cumulated evapotranspiration, we integrated the daily ETR over an agricultural year (from 1st October to 30th September). Figure 5.48 shows the comparison between the measurements and the SAFYE-CO_2 estimates. For the five investigated years, we note that the simulated cumulative sum of ETR matches well with the observations during the crop season, especially for LAM2007, LAM2009 and AUR2010. Besides, for LAM2009 and AUR2010 (and a bit for AUR2008) it can be noted that at the beginning of the senescence phase (approximately early June), the model estimates diverge from the observations. It is due to an overestimation of the ETR by the model at the end of the season. This suggests that the model might not be optimally parameterized for periods of high evaporative climatic demand (mainly in summer). Still, for all the years an overall good agreement between model and measurements is obtained at the end of the season (with relative differences ranging from 3.9% to 16%).

Finally, if we evaluate the cumulated ETR in terms of absolute values, we observe that 2007 was the climatic year that presented the higher water demand by the wheat crops (without considering the wheat varieties), because of its particularly

warm autumn and winter seasons that led to strong GAI values and long vegetative period. Besides, we observe that there is not a particular site-dependent under or over-estimation trend. For AUR2006 and LAM2007 the model overestimates the total cumulated ETR, while for the others years (AUR2008, LAM2009 and AUR2010) the model provides underestimated ETR. When evaluating the model in terms of cumulated ETR inter-annual dynamics (see the bar plot, Figure 5.48 bottom right), without considering the absolute values, but their increase/decrease trends, we observe that for Auradé, from 2006 to 2008, according to the measurements the ETR decreases significantly (from 497 to 374 mm.year^{-1}) while for the model, ETR remain practically constant for both years (≈420mm.year^{-1}). It is due to the ETR underestimation during crop development in 2006. Comparing AUR2008 to AUR2010, the simulated trend is in agreement with the observations. For Lamasquère, from 2007 to 2009, both model and observations show a decrease in the total cumulated ETR.

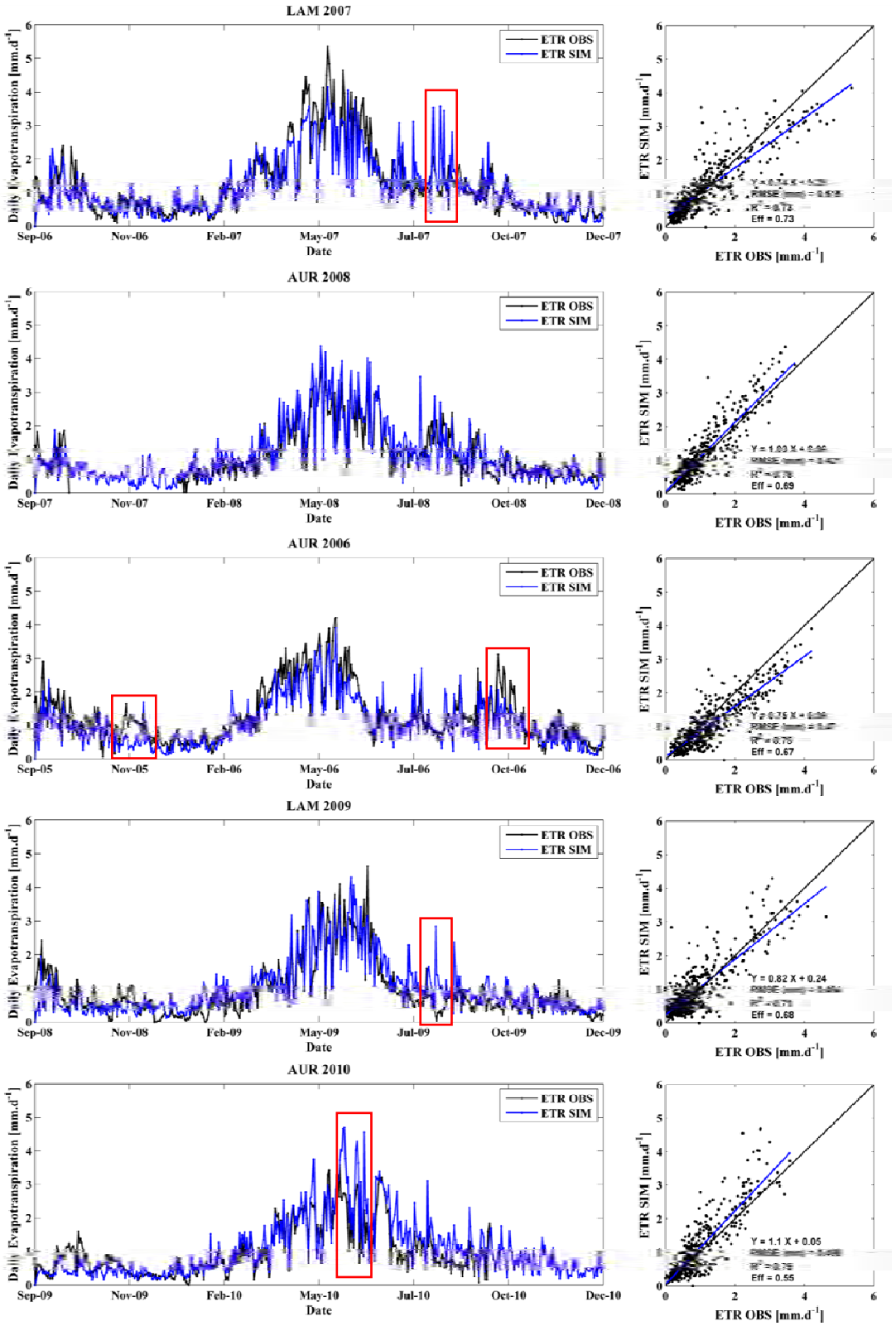


Figure 5.47- Evolution of the evapotranspiration (ETR) dynamics over winter wheat crop seasons for the calibration site-years (LAM2007 and AUR2008), followed by the validation site-years (AUR2006, LAM2009 and AUR2010). Observed ETR are represented in black and simulated ETR in blue. On the right, the scatter plots relating observed vs. simulated ETR and associated statistics (regression, RMSE, R^2 and efficiency Eff) are displayed.

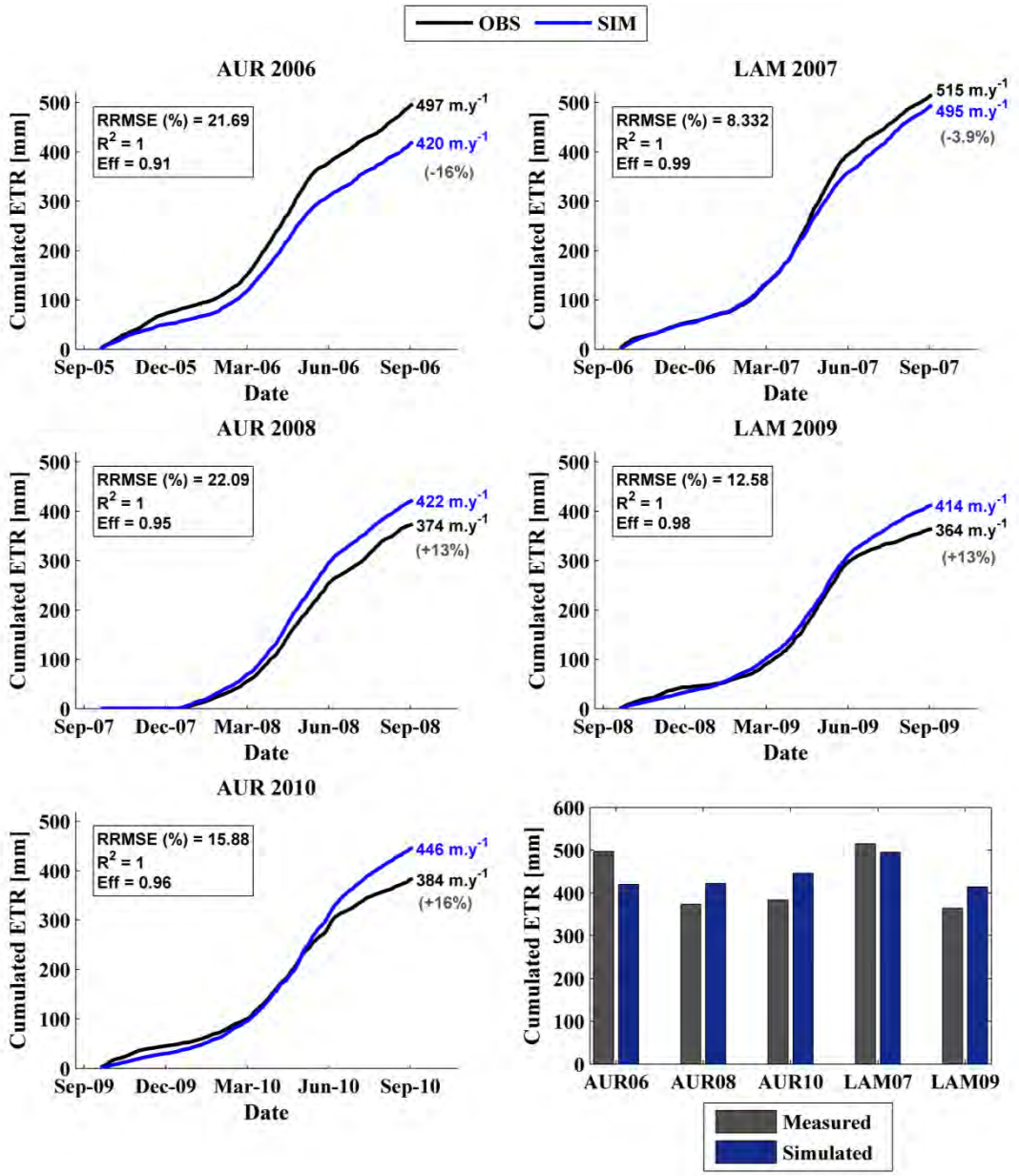


Figure 5.48- Cumulated evapotranspiration (ETR) during the agricultural year, for AUR2006, LAM2007, AUR2008, LAM2009 and AUR2010. Measured cumulated ETR are represented in black and simulated in blue. The cumulated ETR reached at the end of the season are indicated with the same representative colors and the difference between the measured and simulated values (in %) is displayed in grey. The calculated statistical indicators (RRMSE, R^2 and efficiency Eff) are displayed on the left top corner. On the right bottom corner the interannual dynamics of the total crop season cumulated ETR (observed and simulated) is shown for the Auradé (AUR2006, AUR2008, AUR2010) and Lamasquère (LAM2007, LAM2009) sites.

The dynamics of simulated soil water content of the superficial (H1), root (H2) and deep (H3) layers were compared with those measured at the Auradé and Lamasquère sites (Figure 5.49, Figure 5.50, Figure 5.51, Figure 5.52, Figure 5.53). The performances, in terms of root mean square error (RMSE), correlation (R^2) and efficiency (Eff) are summarized in Table 5.3. However, given the high spatial variability of the soil surfaces and relatively limited spatial representativeness of the *in-situ* measurements of soil humidity, the analysis here will be more qualitative than quantitative.

The soil superficial humidity (presented in the first line of the figures) is globally well reproduced by the model in terms of dynamics, however a bias on the absolute values is observed for all years. The models estimates are overestimated in comparison to the measurements, especially in 2007. The observed bias might be related to a poor initialization of the water content (for the three soil layers), conditioning the water balance results. We can also observe that both model and measurements respond instantly to precipitation events. Besides, we note that the observations have a “smoother” aspect than the model estimates. It might be explained by the fact that the half-hourly measurements are integrated over 24 hours to obtain daily information, to be comparable to the daily model outputs.

The root layer soil moisture dynamics are exhibited only during the crop growth, given that this layer is not simulated by the models during bare soil periods. Contrary to the top-layer, for the root layer the simulated moisture dynamics are smoother than the measurements. In general, as for the superficial layer, the SWC content is overestimated by the model. A possible reason for this fact is that transpiration does not take enough water from the root layer, inducing underestimated evapotranspiration for some years.

The simulated deep soil humidity remained practically constant over the whole year for LAM2007 and AUR2008. For AUR2006, a decrease in the deep soil moisture can be noted around May for both measurements and model estimates. It caused a water stress during this period, so the deep water reservoirs were required for maintaining the plant transpiration. A good agreement was observed between the model and the observations for LAM2009 and for AUR2010. Those years, the dynamic is well reproduced by the model, even if the SWC decrease between May and October is simulated with a lower intensity. In general, since this layer is driven exclusively by diffusive fluxes, it might be necessary to review the diffusion-related parameters, for better deep soil humidity estimates.

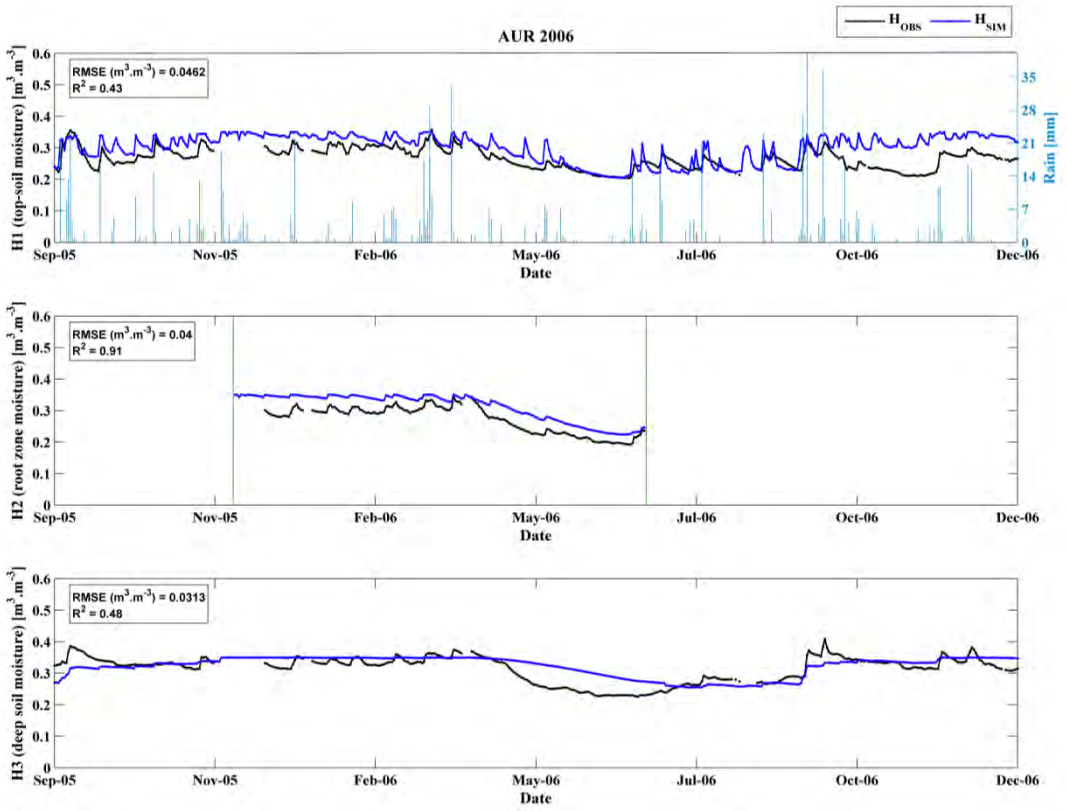


Figure 5.49- Comparison between the measured and simulated soil moistures for the Auradé site, 2006. The three soil layers are represented: H1, the superficial soil, H2, the root soil layer, and H3, the deep layer. The precipitation measurements are also indicated, and correspond to the right y-axis. The performances in terms of root mean square error (RMSE, $m^3.m^{-3}$) and correlation (R^2) are displayed in the top left corner.

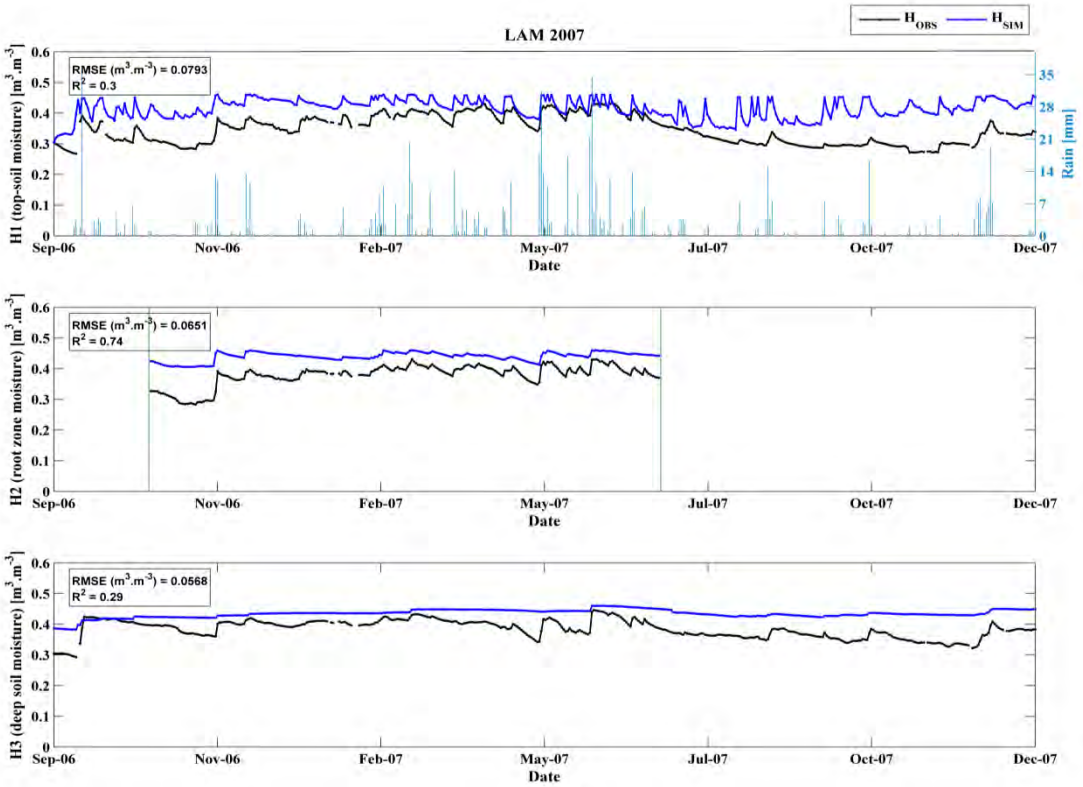


Figure 5.50- Comparison between the measured and simulated soil moistures for the Lamasquère site, 2007. Same legend as Figure 5.49.

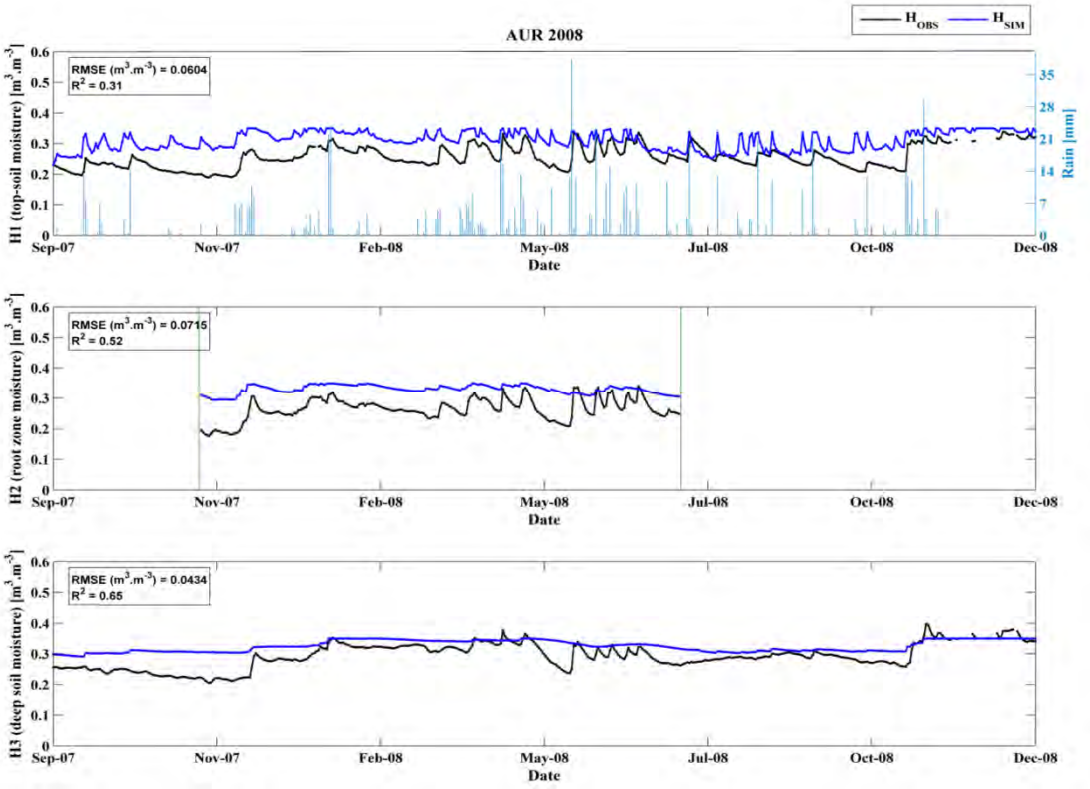


Figure 5.51- Comparison between the measured and simulated soil moistures for the Auradé site, 2008. Same legend as Figure 5.49.

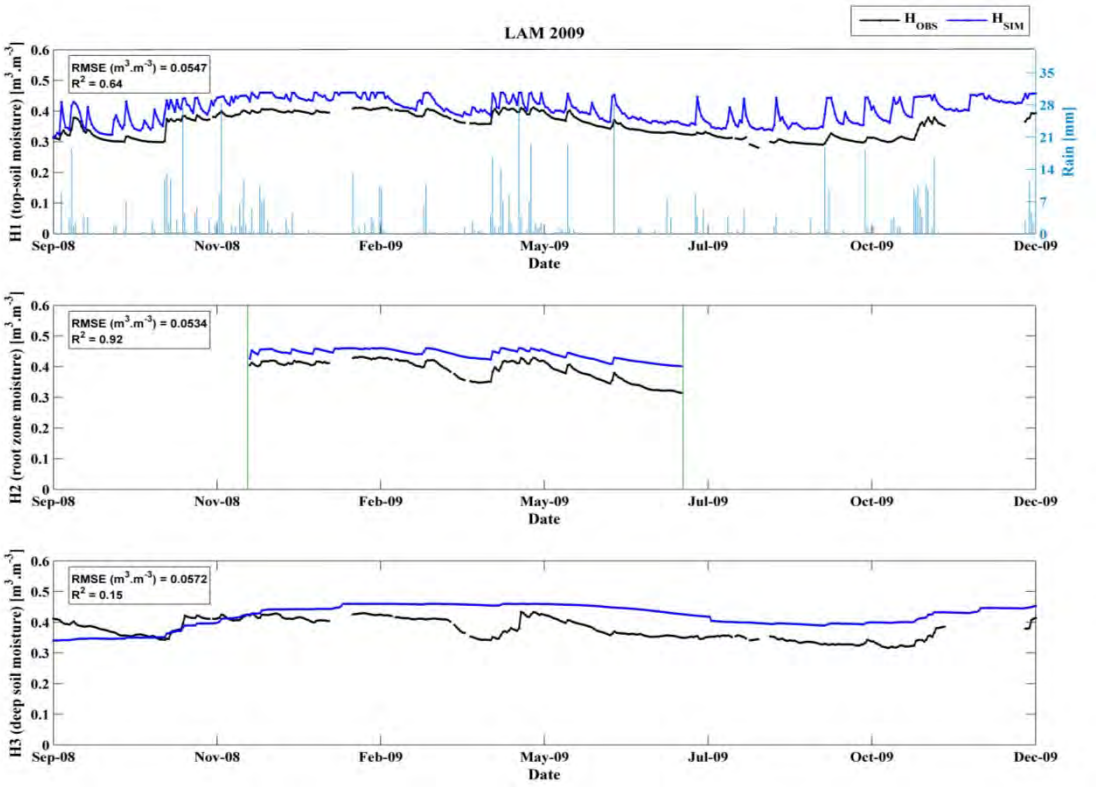


Figure 5.52- Comparison between the measured and simulated soil moistures for the Lamasquère site, 2009. Same legend as Figure 5.49.

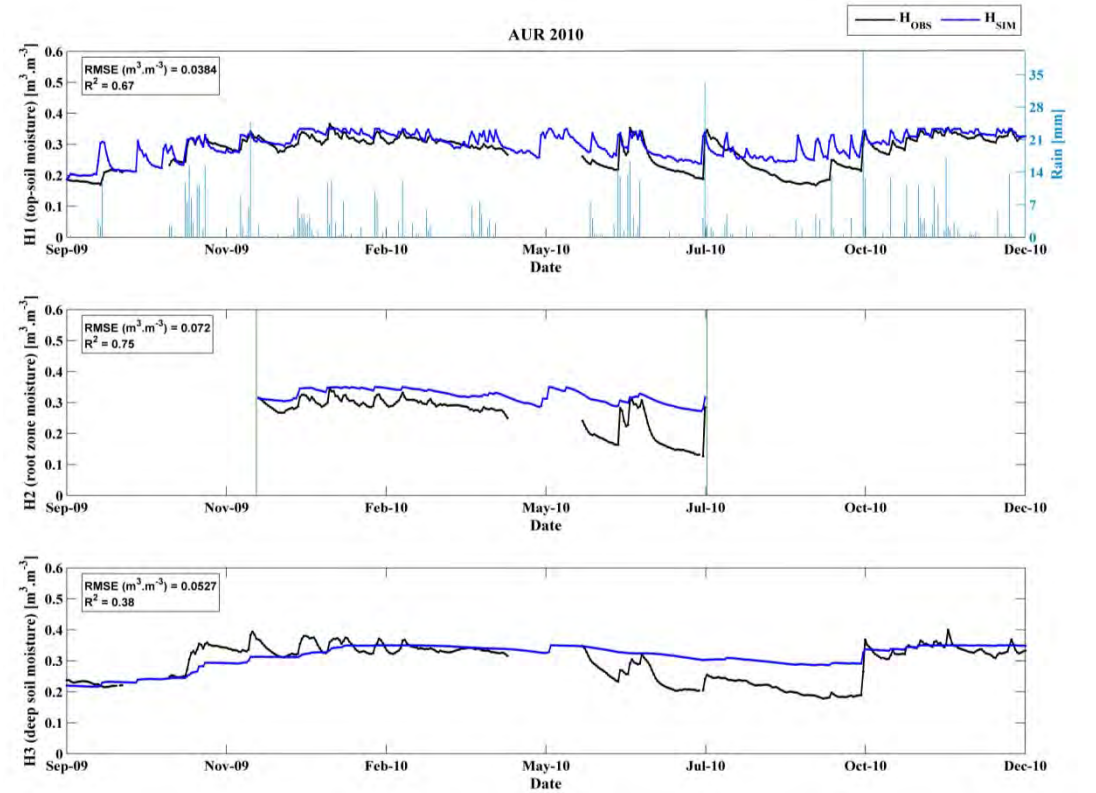


Figure 5.53- Comparison between the measured and simulated soil moistures for the Auradé site, 2010. Same legend as Figure 5.49.

5.4.2. GPP, Reco and NEE estimates

The dynamics of simulated gross primary production (GPP) obtained with the SAFYE-CO_2 model were the same as those of SAFY-CO_2, except for AUR2006 due to the impact of water stress during the crop development this year. Thus, in the absence of water stress the GPP dynamics estimated by SAFYE-CO_2 remain the same as for SAFY-CO_2. For this reason, only the results for AUR2006 will be presented here. The results for the other years are displayed in Appendix A.

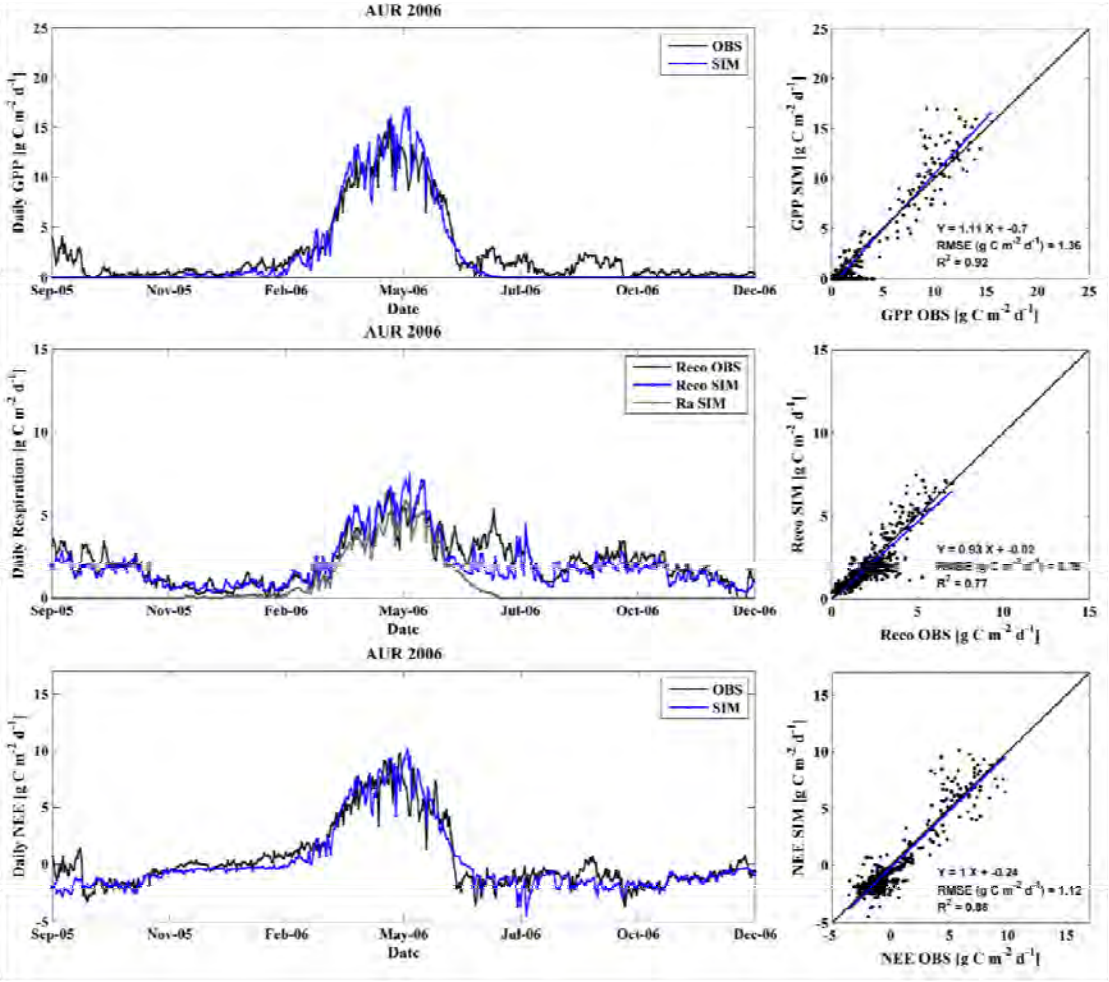


Figure 5.54- Evolution of the carbon fluxes for the Auradé site, 2005-2006 winter wheat crop season. Observed fluxes are represented in black and the fluxes simulated by the SAFYE-CO_2 model are represented in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.

For AUR2006, from the scatter plot (Figure 5.54, right) we observe an overall overestimation of the GPP estimates, especially for high values (GPP >10 gC $m^{-2}d^{-1}$). By evaluating the daily dynamics it can be noted that this overestimation mainly occurs from maximal development period until the mid-senescent phase. When compared with the SAFY-CO_2 GPP estimates (Figure 5.31), we observe that this overestimation was already noticeable, and it was even stronger. As it can be seen in Figure 5.55, illustrating the water stress function simulated by SAFYE-CO_2, for AUR2006 the water stress function started acting from May until the end of June. The water stress function is comprised between 0 and 1. The lower this function is, the more stressed is the vegetation. It means that for this period, the GPP estimates were constrained by the water stress function, getting the estimates closer to the *in-situ* measurements values. Even if for the other investigated years the crops were not affected by water stress, the results for Auradé in 2006 show the importance of considering possible water stress over the flux estimates. Besides, we note that the effect of crop re-growth and weed development increase GPP from mid-August until late September, but that phenomenon is not reproduced by the model (as for SAFY-CO_2).

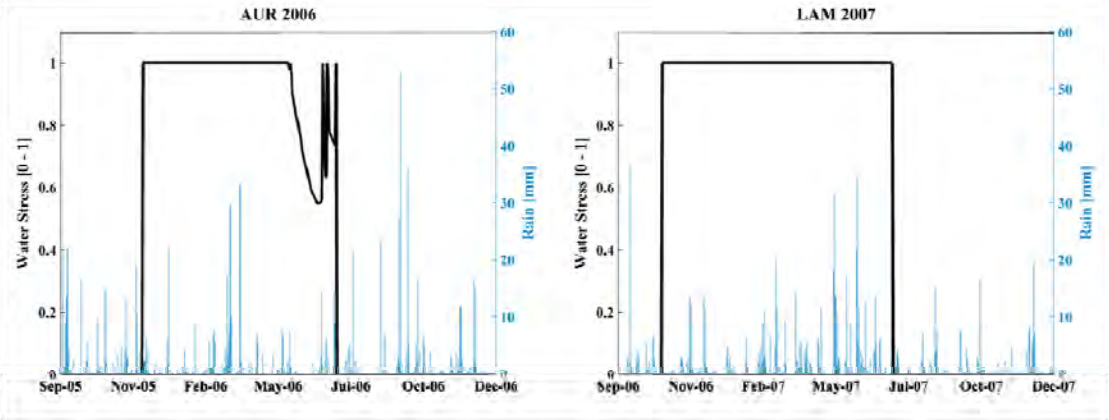


Figure 5.55- Water stress function simulated by the SAYE-CO_2 model for Auradé 2006 (on the left) and Lamasquère 2007 (on the right). The water stress is comprised between 0 (maximal stress) and 1 (no stress).The precipitations, measured by the meteorological instruments installed in both sites, are displayed in the right y-axis.

Concerning the ecosystem respiration estimates (see Appendix A), we observe that in general the simulated Reco matches well with the observations, still an overall underestimation tendency can be observed. We observe that for periods of low vegetation, in the beginning of the crop development, the Reco estimates are in very good agreement with the observations. Then, at maximal growth development, some divergences between model and observations may occur, as the peaks

observed for LAM2007 (see Appendix A for the GPP, Reco and NEE dynamics estimates of SAFYE-CO_2). For post-harvest periods, nevertheless, the performances can be degraded due to re-growths, weed development or manure application/soil labor, processes that may affect the ecosystem respiration and that are not considered by the model. The obtained correlation coefficients are overall satisfactory: in average, R^2=0.79 (Table 5.4). The obtained errors (RMSE) may vary considerably depending on the year. For instance, a RMSE=0.85 gC $m^{-2}d^{-1}$ for AUR2006 and 1.11 gC $m^{-2}d^{-1}$ for LAM2007 were found. In Table 5.4, the performances of the SAFYE-CO_2 model are compared with those of the SAFY-CO_2 model. We observe that Reco performances are thus improved by SAFYE-CO_2. For the non-stressed years (2007 to 2010) this improvement is entirely owed to the SAFYE-CO_2 heterotrophic respiration function, which depends upon soil temperature but also upon soil water content (simulated by the water module).

Table 5.4- Performances of the CO_2 fluxes estimated by the SAFYE-CO_2 model for winter wheat over the experimental sites, from 2006 to 2010. For comparison, the performances obtained by the SAFY-CO_2 model are indicated on the left side of each field, in grey color.

	GPP		Reco		NEE	
	R^2	RMSE*	R^2	RMSE*	R^2	RMSE*
AUR2006	0.91/0.92	1.42/1.36	0.77/0.77	0.80/0.75	0.85/0.86	1.21/1.12
LAM2007	0.94	1.4	0.80/0.81	1.23/0.97	0.87/0.88	1.06/1.16
AUR2008	0.94	1.26	0.74/0.75	0.89/0.83	0.89/0.90	1.05/1.12
LAM2009	0.93	1.13	0.71/0.75	0.90/0.84	0.79/0.81	1.19/1.14
AUR2010	0.94	1.27	0.82/0.85	0.85/0.79	0.88/0.89	1.28/1.27

* RMSE in gC $m^{-2}d^{-1}$.

Concerning the NEE estimates, In general, the daily dynamics of the net ecosystem exchange (NEE) are well simulated by the SAFYE-CO_2 model. The average statistics indicators, calculated over the 5 site-years, indicate a correlation coefficient (R^2) of 0.88 and an error about 1.17 gC $m^{-2}d^{-1}$ (Table 5.4). The performances of the SAFY-CO_2 model are also indicated. We observe that the performances are quite similar. Still, the SAFYE-CO_2 NEE estimates are slightly better correlated to in-situ measurements than those of SAFY-CO_2. However, for the error (RMSE) it depends upon the year. For AUR2006, thanks to the water stress function, the GPP estimates were improved. The Reco estimates were also improved due to the new Rh function of SAFYE-CO_2, resulting in better NEE estimates as well (RMSE =1.12 gC $m^{-2}d^{-1}$). On the other hand, for example, in LAM2007, the GPP estimates remained the same (no

water stress), but the Reco estimates were improved by SAFYE-CO_2 (RMSE =0.97 against 1.23 gC $m^{-2}d^{-1}$ for SAFY-CO_2). Since the GPP is underestimated by both models, the improved Reco resulted in a slight "degradation" of the NEE estimates by SAFYE-CO_2 (RMSSE=1.16 against 1.06 gC $m^{-2}d^{-1}$ for SAFY-CO_2). Given that the GPP and Reco terms compensate each other, higher (and better) Reco estimates highlighted the underestimation of GPP, which degraded the estimates of the NEE.

The analysis of the profiles of cumulated NEE (Figure 5.56) simulated by SAFYE-CO_2 shows a good correlation with the *in-situ* measurements, but the errors in terms of absolute values are quite important. In general, Reco was better simulated (and less underestimated) with the SAFYE-CO_2 version than with SAFY-CO_2. In LAM2007 and AUR2010 the apparent degradation in the modeled cumulated NEE are in reality caused by the GPP underestimation in both versions of models, but their effect is more visible on the NEE estimated by SAFYE-CO_2 (because Reco is higher). For LAM2009 the degradation in the NEE estimates comes from a slight overestimation of the modeled Reco with SAFYE-CO_2 in December/January that leads to an underestimated cumulated NEE value in June, when senescence occurs. Then the divergence with the measurements (caused by the re-growth event that is not accounted for by the models) is more visible with SAFYE-CO_2 than with SAFY-CO_2, because in June the SAFY-CO_2 cumulated NEE was lower than observations, as Reco was overall underestimated during this period. For AUR2008, the slight overestimation of Reco during the winter yielded to a small degradation of the cumulated NEE by SAFYE-CO_2 compared to SAFY-CO_2. Also, during the senescence phase, GPP was underestimated by both versions of the model, but Reco was less underestimated by SAFYE-CO_2. Therefore, in June (when senescence occurs), the cumulated NEE value for the SAFYE-CO_2 version was less negative, diverging more from the observations than SAFY-CO_2. This difference in the cumulated NEE in June between the two versions of the model leads to better estimates at the end of the cropping season with SAFY-CO_2. For AUR2010, SAFYE-CO_2 underestimates the Reco from September to December, even if the Reco dynamics are better represented compared to SAFY-CO_2. This underestimation leads to lower performances at the end of the season with SAFYE-CO_2. For AUR2006, the difference between the two versions of the model comes from that fact that none of them simulates the GPP event in September 2005 (re-growth event) but the SAFYE-CO_2 version simulates better (and higher) Reco during this period, causing higher divergences between observations and modeled values throughout the year. Also the water stress in June affects the GPP simulated by SAFYE-CO_2 while this effect is not accounted for with SAFY-CO_2.

In conclusion, the better estimation of Reco with SAFYE-CO_2 highlights the problems encountered in simulating GPP correctly (*e.g.* re-growth event in 2005, low

GPP at senescence in 2008, overall GPP underestimation in LAM2007) with both versions of the model. Therefore, some strategies will be implemented to overcome these issues (see discussion in section 5.7).

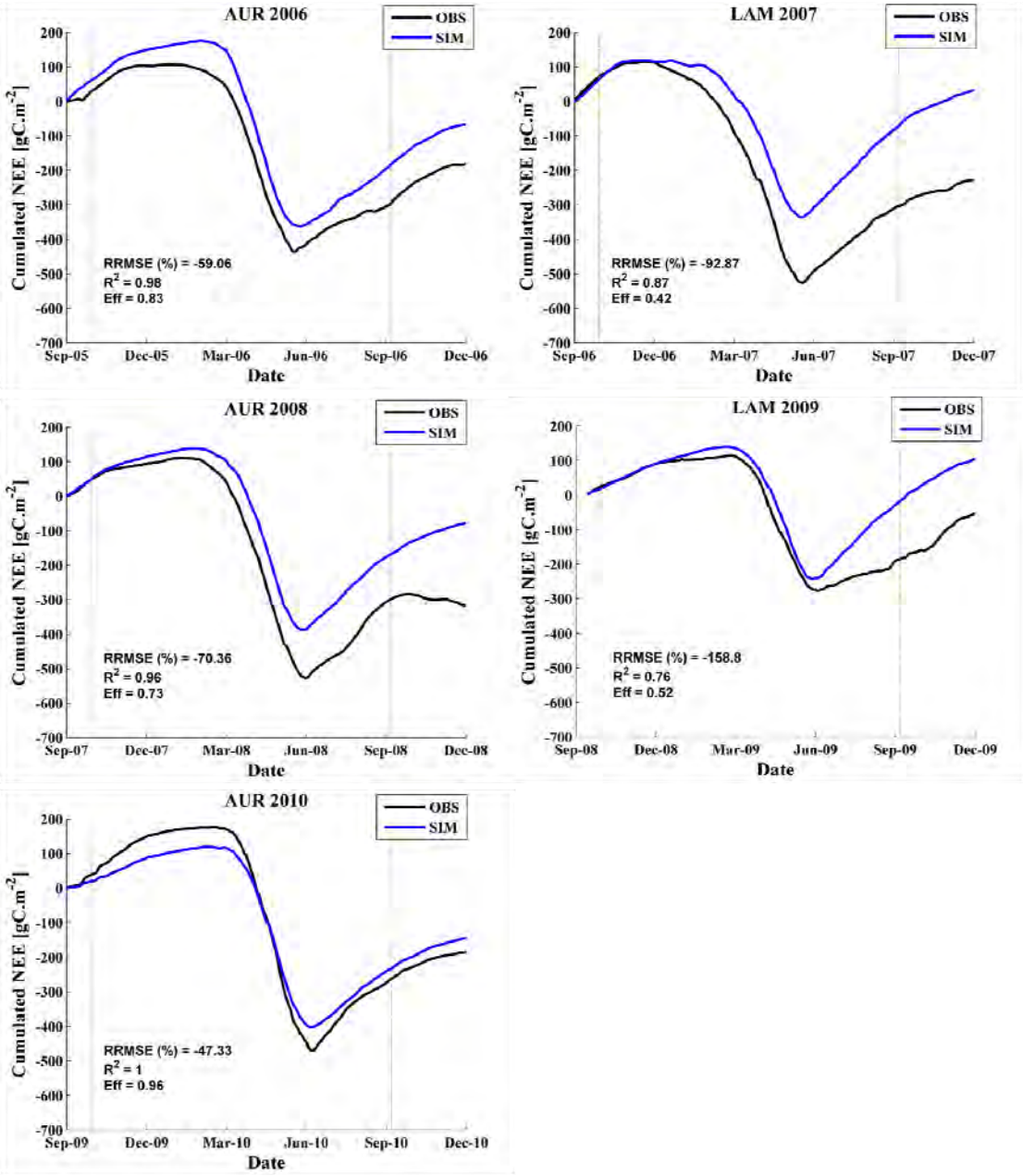


Figure 5.56- Cumulated values of NEE for the 5 site-years: AUR2006, LAM2007, AUR2008, LAM2009 and AUR2010. The observations are displayed in black and the SAFYE-CO_2 model outputs are in blue. The performances, in terms of relative mean square error (RRMSE), linear coefficient of correlation (R^2) and Nash criterion (Eff) are shown in the left bottom corner. The grey vertical bars indicate the period for which the NEE fluxes are integrated for estimating the net ecosystem production (NEP, used for calculating the net ecosystem carbon budgets).

5.4.3. Crop growth and production: GAI, DAM and Yield estimates

Figure 5.57 illustrates the scatter plots of SAFYE-CO_2 estimates against observations for GAI and biomass for the 'experimental sites' and the '2011 field campaign' data sets together. The analysis of these figures reveals that the SAFYE-CO_2 model provides accurate estimates of the GAI, with RRMSE of 11.8% and R^2=0.98, and of the biomass as well, with RRMSE of 25.9% and R^2=0.89.

It is important to note that the coupling of the water module with the SAFY-CO_2 model did not affected much the biomass and yield estimates. However, for LAM2011 we note a marked underestimation of the final biomass, that was already present for SAFY-CO_2, but accentuated by SAFYE-CO_2 due to the effect of the water stress. For the others years (2006-2010) the results are essentially the same (since no stress occurred, except in 2006).

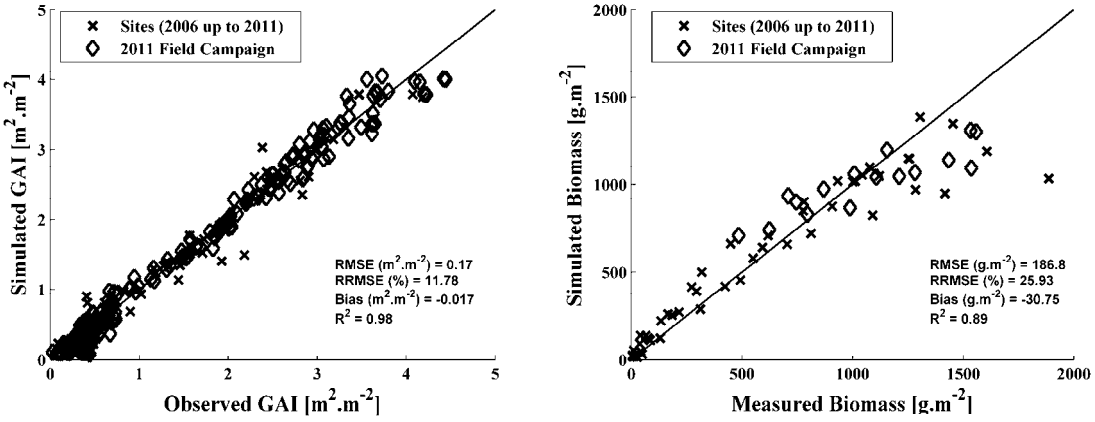


Figure 5.57- Comparison between the green area index (GAI, on the left) and the biomass (on the right) simulated by the SAFYE-CO_2 model and the satellite-derived GAI and ground measurements (observations), respectively. Two data sets are used: the 'experimental sites' data (represented by the x symbol) and the '2011 field campaign' data (represented by the diamond symbol). The line 1x1 is displayed. The performances (RMSE, RRMSE, Bias and R^2) are indicated on the right corner.

The results of the yield estimates provided by the SAFYE-CO_2 model (using the four investigated methods) over the '2011 field campaign' data set are displayed in Figure 5.58. The performances vary according to the chosen method. The methods 1 and 4 present nearly equivalent performances, with a relative error of about 24% and a coefficient correlation of 0.73. The methods 2 and 3 present poorer performances, but still acceptable. Note that for those two methods the points are in general closer to the 1:1 line. These results corroborate overall those found with SAFY-CO_2 and are important to validate the use of SAFYE-CO_2 as a main model. The

yield estimated by SAFYE-CO_2 for the experimental sites (2006-2011) are illustrated in Appendix A. As it was found for the SAFY and SAFY-CO_2 models, the method 4 is the one yielding to the best performances for this data set. Still, the correlation coefficients remain low for all the methods, partly due to the uncertainty associated to this data set.

The results of this section will be further discussed in section 5.5, which will present the comparison between the three investigated models (SAFY, SAFY-CO_2 and SAFYE-CO_2) in terms of GAI, biomass and yield estimates.

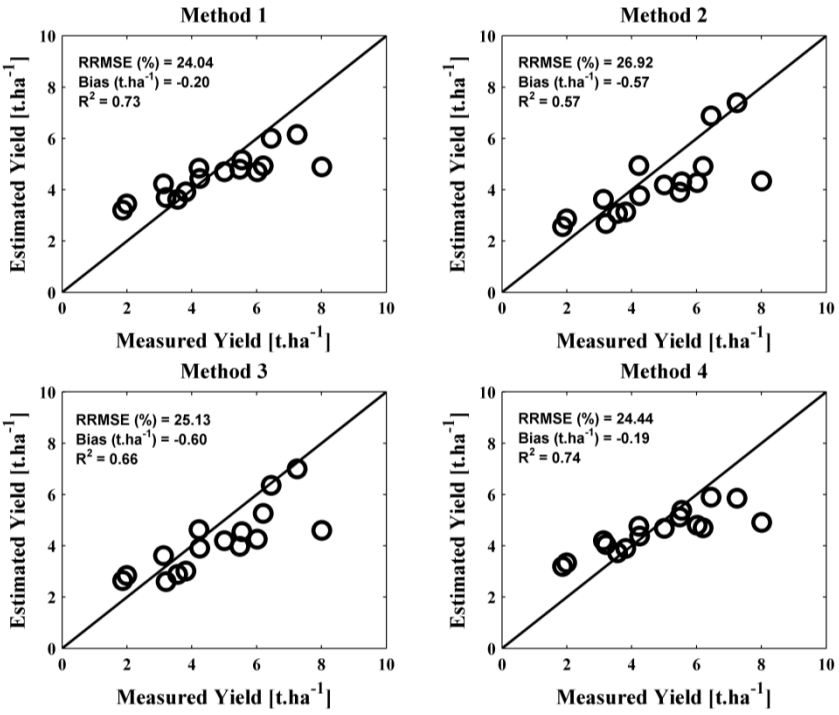


Figure 5.58- Comparison of the methods for estimating yield with the SAFYE-CO_2 model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the 2011 campaign fields.

5.4.4. Crop carbon budget

As for the SAFY-CO_2 model, the annual crop carbon budget was calculated for the five studied sites-years. The net ecosystem carbon budget (NECB) estimates, and their components, were compared to those derived from the flux towers measurements and from farmers inventories (see Figure 5.59).

The simulated NEP ranged from -33 gC $m^{-2}y^{-1}$ (LAM2009) to -252 gC $m^{-2}y^{-1}$ (AUR2010). This range is in poor agreement with the observations: from -208 gC $m^{-2}y^{-1}$ (LAM2009) to -374 gC $m^{-2}y^{-1}$ (LAM2007). In general, we observe a large

underestimation of the NEP estimates provided by SAFYE-CO_2 in comparison to the observations (simulated NEP are less negative than the observations). As discussed in the previous section, with its water module, the SAFYE-CO_2 improves the Reco estimates, but at the same time highlights the underestimation of the GPP term. Therefore, to improve the performances of the NEE estimates, it will be necessary to improve the GPP estimates. Even if these last present good agreement with observations, it still needs to be improved in order to obtain more accurate estimates of the NEE, and thus of the carbon budget. Still, for 2010 for example there is only 16% difference between the modeled NEE and the measurements.

In terms of the annual amount of carbon exported by the ecosystem (Cexp), SAFYE-CO_2 has performances quite similar to those of SAFY-CO_2. Indeed, a very good agreement is observed between the SAFYE-CO_2 estimates and the *in-situ* data. Since the biomass estimates were just slightly impacted by the water module (stress observed only for 2006), so were the yield estimates, leading to good estimates of the Cexp term.

Finally, according to the model outputs, the NEP are overall underestimated and the Cexp are generally overestimated, resulting in cumulated errors over the NECB estimates. As NEP and Cexp have opposite signs, the under and over-estimations of both terms do not compensate each other, which makes the errors in the NECB estimates even stronger. In terms of absolute differences between NECB estimates and observations, a minimum of 81 gC $m^{-2}y^{-1}$ was found for AUR2010, and a maximum of 292 gC $m^{-2}y^{-1}$ for LAM2007. These divergences are important and problematic, since they determine whether if the plot acts as a sink or a source of carbon. The solution for improving the NECB estimates should be mainly focused on the improvement of the NEP (and so on the GPP) term.

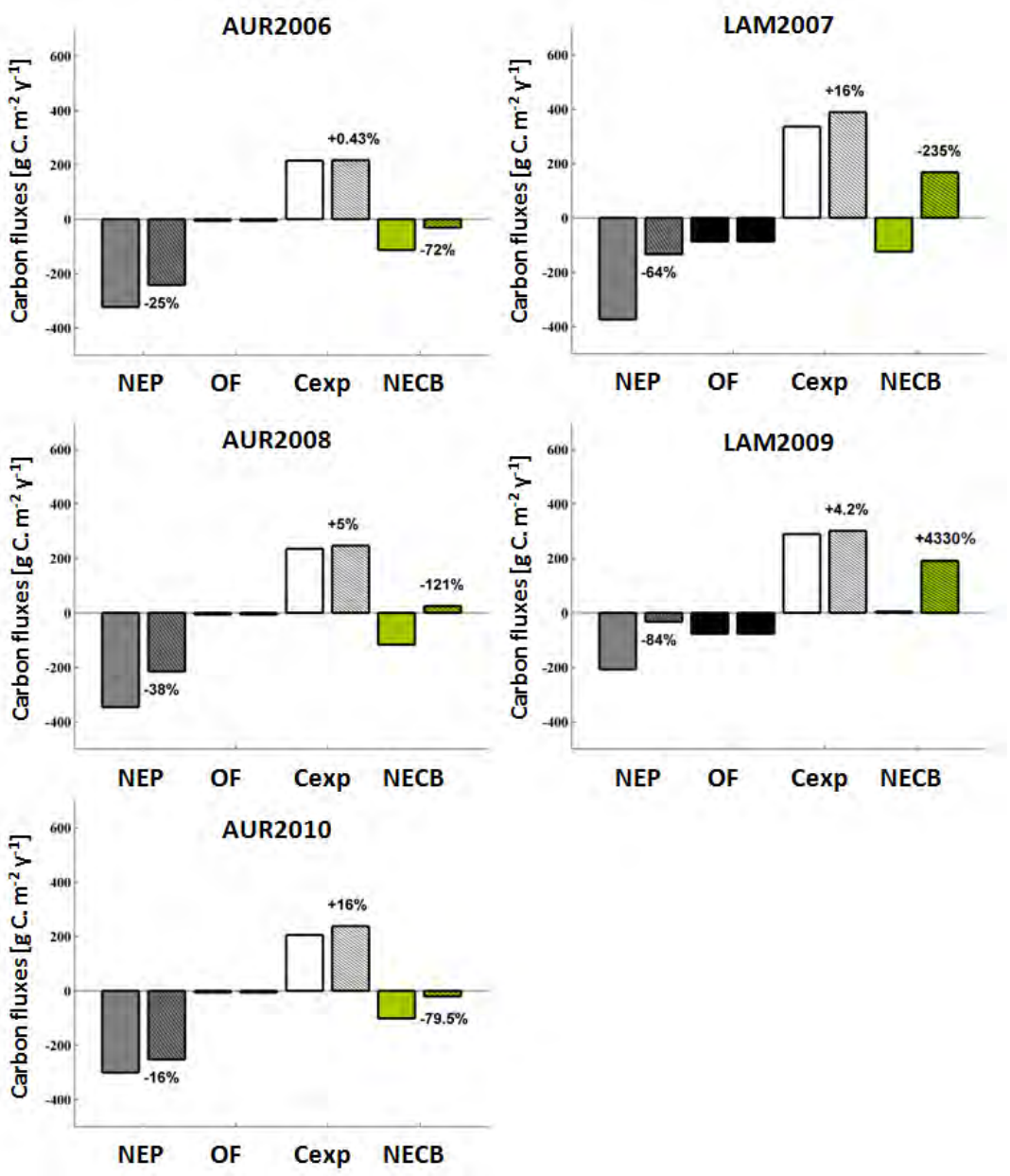


Figure 5.59- The annual carbon budget for the winter wheat crops, at Auradé (2006, 2008, 2010) and Lamasquère (2007, 2009) sites. The first bar of each pair represents the measurements and the second bar, hatched, represents the results provided by the SAFYE-CO_2 model. The grey bars represent the net ecosystem production (NEP), the black bars represent the organic fertilization (OF), the white ones represent the amount of carbon in the fraction of the crop biomass that is exported at harvest and the green bars represent the net ecosystem carbon budget (NECB).

5.5. Models inter-comparison: SAFY, SAFY-CO_2 and SAFYE-CO_2

This section presents the inter-comparison between the three investigated models in terms of: GAI, biomass and yield. It is an important step to evaluate the evolution of the model from its original version (SAFY) up to most complete version (SAFYE-CO_2).

The GAI, biomass and yield estimates provided by the developed SAFY-CO_2 and SAFYE-CO_2 models (Figure 5.42 to Figure 5.45 and Figure 5.57) can be compared to those provided by the standard SAFY model (Figure 5.16, Figure 5.21, Figure 5.23 and Figure 5.24). Table 5.5 summarizes the performances obtained by the models in terms of mean error (RMSE and RRMSE), bias and correlation (R^2). From these statistics, we observe that, in terms of GAI, the performances are basically the same for the three models. It means that the newly developed SAFY-CO_2 and SAFYE-CO_2 models do not degrade the GAI estimates compared to SAFY, which is crucial in the perspective of adopting one of the new models as a common tool. In terms of biomass, the SAFY-CO_2 model provides better estimates than SAFY (RRMSE =23.7% against 25% and R^2=0.91 against 0.88). Even if the improvement is slight, it shows that SAFY-CO_2 is good at estimating biomass production and therefore can replace the SAFY model with no performances deterioration. However, the SAFYE-CO_2 model presents slightly inferior performances (RRMSE =25.9% and R^2=0.89, see Table 5.5). Even if the degradations are not significant, they may point out that there are improvements to be performed, maybe in the coupling with the water module.

Table 5.5- Comparison of the performances of the SAFY, SAFY-CO_2 and SAFYE-CO_2 models on the GAI and biomass estimates.

	GAI			Biomass		
	SAFY	**SAFY-CO_2**	**SAFYE-CO_2**	**SAFY**	**SAFY-CO_2**	**SAFYE-CO_2**
RMSE*	0.17	0.17	0.17	180.3	171.0	186.8
RRMSE (%)	12.4	11.5	11.8	25.0	23.7	25.9
Bias*	-0.005	-0.013	-0.017	-20.1	-22.6	-30.7
R^2	0.98	0.98	0.98	0.88	0.91	0.89

* Units: GAI: $m^2.m^{-2}$; Biomass: $g.m^{-2}$.

In order to further investigate the possible causes of these differences, the biomass estimates for the three models were plotted against the *in-situ* measurements (Figure 5.60). From the analysis of this graphic, it can be promptly

noted a deviant point (observed DAM=1887g.m^{-2}) that is underestimated by the three models, but mostly by the SAFYE-CO_2 model. This point corresponds to a measurement performed at the Lamasquère site, at the end of the season (just before harvest, on 16/06/2011). This point is thus the main cause for the lower performances of the SAFYE-CO_2 model for estimating biomass. We can also observe another point that is more underestimated by SAFYE-CO_2 than by the other two models (indicate by the arrow), corresponding to a measurement around the flowering stage (12/05/2011). For the observations points for which the SAFY-CO_2 estimates (in red) cannot be seen, it means they correspond exactly to the same estimates as SAFYE-CO_2 (in blue), so both models outputs are superposed, and only the estimates that were plotted at last (SAFYE-CO_2) are visible.

To better understand these results, it is helpful to look at the GAI profiles observed and simulated for this site in 2011 (Figure 5.41). The simulated GAI are in good agreement with the satellite-derived GAI observations, which are used for calibrating the model, showing that the calibration method works. However, the comparison between these satellite-derived GAI against destructive measurements indicates that the ground measurements are largely superior to the satellite-derived GAI. At maximal development stage (GAI_{max}), the destructive measurements indicate a GAI of 6.86 $m^2.m^{-2}$, while the GAI obtained from satellite observation attains only 4.18 $m^2.m^{-2}$. Even if it is known that the satellite-derived GAI are inferior to destructive measurements (between 20-30%), here the discrepancy is unusual. The underestimated satellite GAI causes lower GAI estimates by the three models, which results in lower biomass estimates. This issue could be improved by the use of radar data (Radarsat-2, Sentinel-1,..) for estimating GAI in periods of strong vegetation development, since the radar signal is less sensitive to saturation (Revill et al., 2013).

Therefore, the underestimation of the satellite derived GAI explains why the three models underestimates the biomass for the two LAM2011 observations. Next step is to find an explanation for the differences observed between the three models. It can be noted that, for the point LAM 16/06/2011, SAFY-CO_2 and SAFYE-CO_2 have poorer biomass estimates than the SAFY model. This difference is probably due to the effect of the fraction of diffusive radiation over the photosynthesis estimates (*fELUE* function) by the two last models. This means that the *fELUE* function over-corrects the diffuse radiation effects when clear sky conditions are met, as it was previously discussed for SAFY-CO_2 (Figure 5.33). In addition to this aspect, it is important to note that the 2011 season was impacted by a severe spring drought in the southwest region. Since the SAFYE-CO_2 model has a “water module”, the soil dryness is accounted for in the biomass estimates by means of a water stress function. It explains why SAFYE-CO_2 provides lower biomass estimates than the SAFY-CO_2 model (for LAM2011 12/05 and 16/06). There are two other points (encircled in green,

Figure 5.60) for which the SAFY(E)-CO_2 models provide better estimates than SAFY. They correspond to observations for LAM2007 and AUR2008 and this improvement is due to the *fELUE* function that worked well for these cases.

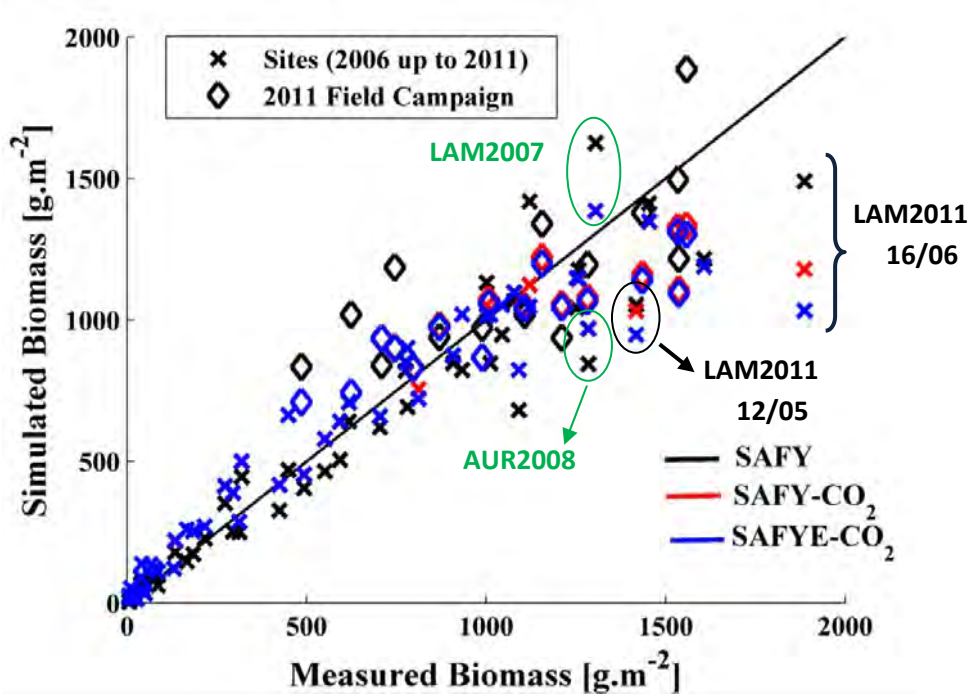


Figure 5.60- Comparison of biomass estimated by the three models: SAFY (in black), SAFY-CO_2 (in red) and SAFYE-CO_2 (in blue). Two data sets are represented: the experimental sites (cross symbol) and the 2011 field campaign (diamond symbol).

The yield performances are only compared between the different versions of the model for the '2011 field campaign' data set, as this data set is more significant and robust (Table 5.6). For the methods 1 and 4, the comparison of the results yields to similar conclusions than those found for the biomass estimates, since these methods depend upon the biomass production and a constant harvest index. Still, even if the SAFY-CO_2 provides the best performances, SAFYE-CO_2 presents good performances as well, better than those of SAFY model. Since here we only consider the '2011 field campaign' data set, the two atypical points of Lamasquère (cited above) are not accounted for. For the method 3, the yield estimates are a little worsened by the SAFY-CO_2 model, in terms of bias and R^2, and also by the SAFYE-CO_2 model, in terms of bias and R^2 and RRMSE. However, the differences between the performances are not important (see Table 5.6). For the method 2, the RRMSE is improved by SAFY-CO_2 and SAFYE-CO_2, but the bias is increased and the correlation coefficient slightly degraded. As it was discussed previously, the methods 2 and 3 depend upon a harvest index that evolves with time (method 2) or thermal time (method 3), which is expected to improve the estimation of the yield production. However, to achieve

this objective it would be necessary to enhance the calibration of the parameters associated with the yield function (the STY, sum of the temperature for starting grain filing, and the parameters VITICARB and VITCARBT, linked to the dynamic harvest index). Finally, for the moment, those results corroborate our choice of keeping the method 4 (or 1 as well) for yield estimation for upcoming applications and of encouraging further investigations of more complex (and complete) functions for simulating yield production.

Table 5.6- Comparison of the performances of the SAFY, SAFY-CO_2 and SAFYE-CO_2 models on the yield estimates using four different methods.

	Yield- Method 1			**Yield- Method 2**		
	SAFY	**SAFY-CO_2**	**SAFYE-CO_2**	**SAFY**	**SAFY-CO_2**	**SAFYE-CO_2**
RRMSE (%)	23.1	22.8	24.0	36.0	29.5	26.9
Bias [t.ha^{-1}]	0.19	0.12	-0.20	0.47	0.89	-0.57
R^2	0.62	0.74	0.73	0.61	0.63	0.57

	Yield- Method 3			**Yield- Method 4**		
	SAFY	**SAFY-CO_2**	**SAFYE-CO_2**	**SAFY**	**SAFY-CO_2**	**SAFYE-CO_2**
RRMSE (%)	24.5	22.3	25.1	28.5	23.8	24.4
Bias [t.ha^{-1}]	-0.097	0.40	-0.60	0.35	-0.14	-0.19
R^2	0.71	0.68	0.66	0.44	0.74	0.74

5.6. Regional estimates

In this section, we present the results of the SAFYE-CO_2 model for the winter wheat plots present in our study area (extent of the Formosat-2 images). The model was run for the years 2006, 2007, 2010 and 2011, since for the 2008 and 2009 years important gaps in the satellite observation series did not allow having a complete description of the crop development that is necessary for calibrating some of the model's parameters. Figure 5.61 presents a general view of the area of study, displaying the maximum GAI observed for the wheat fields of the investigated years all superposed, in order to have an idea of the broad scene. This may show that there is no clear regional pattern concerning maximum GAI distribution. Some of the output variables were selected in order to analyze the SAFYE-CO_2 performances at regional scale. They are: maximum seasonal GAI, final (maximum) biomass production, final grain yield, the cumulated net CO_2 fluxes (NEE) at the end of the agricultural year (*i.e.* NEP, from October to the end of September the following year), the net ecosystem carbon budget (NECB), the total ETR cumulated during the same period as NEE, and the two WUE indicators, the agronomical WUE_{agro} and environmental WUE_{NECB}. For each year, the amount of available plots over the studied area varies: 373 for 2006, 369 for 2007, 512 for 2010 and 342 for 2011. For these plots, the minimal (min), maximum (max), mean, and standard deviation (std) values were calculated for each output variable and for each of the investigated years. These values are displayed in Table 5.7.

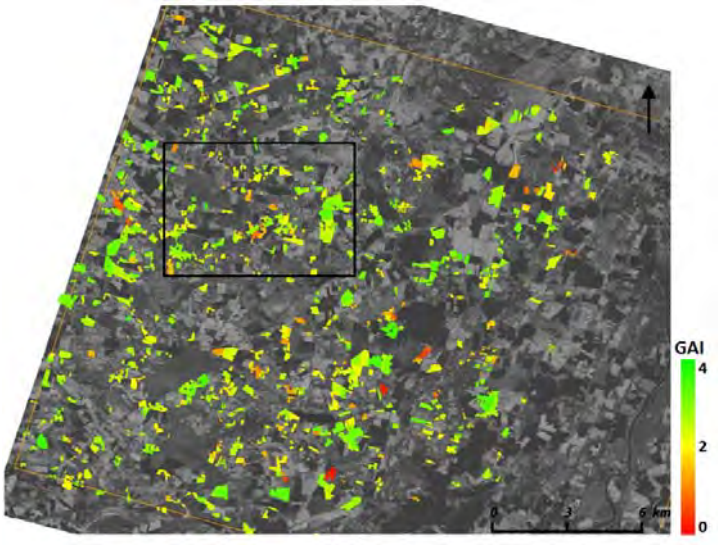


Figure 5.61- General view of the study area, showing the winter wheat fields cultivated in 2006, 2007, 2010 and 2011 (superposed). The colors are illustrative and represent the maximum GAI, ranging from 0 up to 4 $m^2.m^{-2}$. The black rectangle shows the region that was selected for a closer look at the SAFYE-CO_2 estimates.

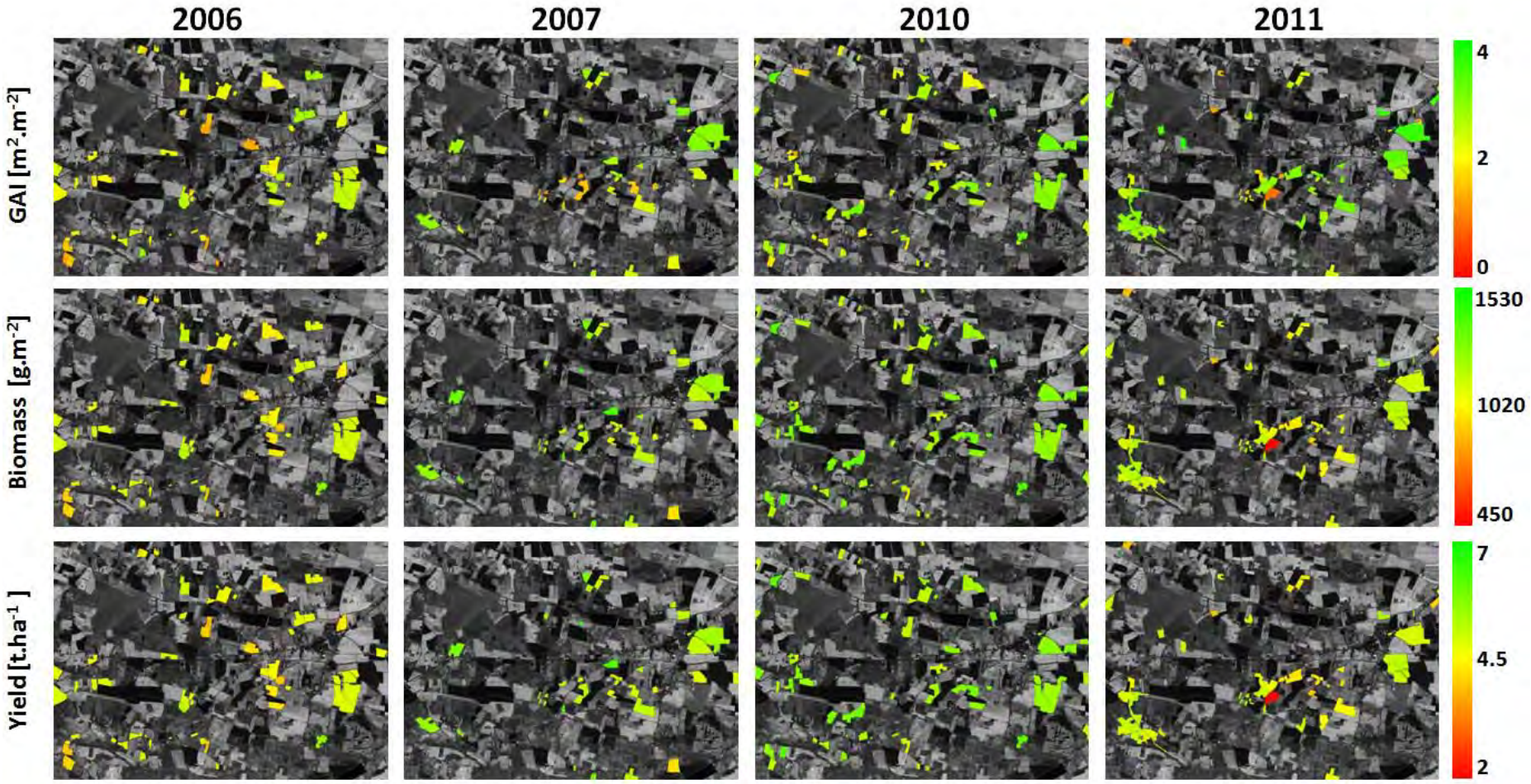


Figure 5.62- Maps of maximum seasonal GAI (top), final biomass (middle) and yield (bottom) estimated by the SAFYE-CO_2 model for a number of plots within the study area over the 2006, 2007, 2010 and 2011 years. The color scales are displayed on the right of each series.

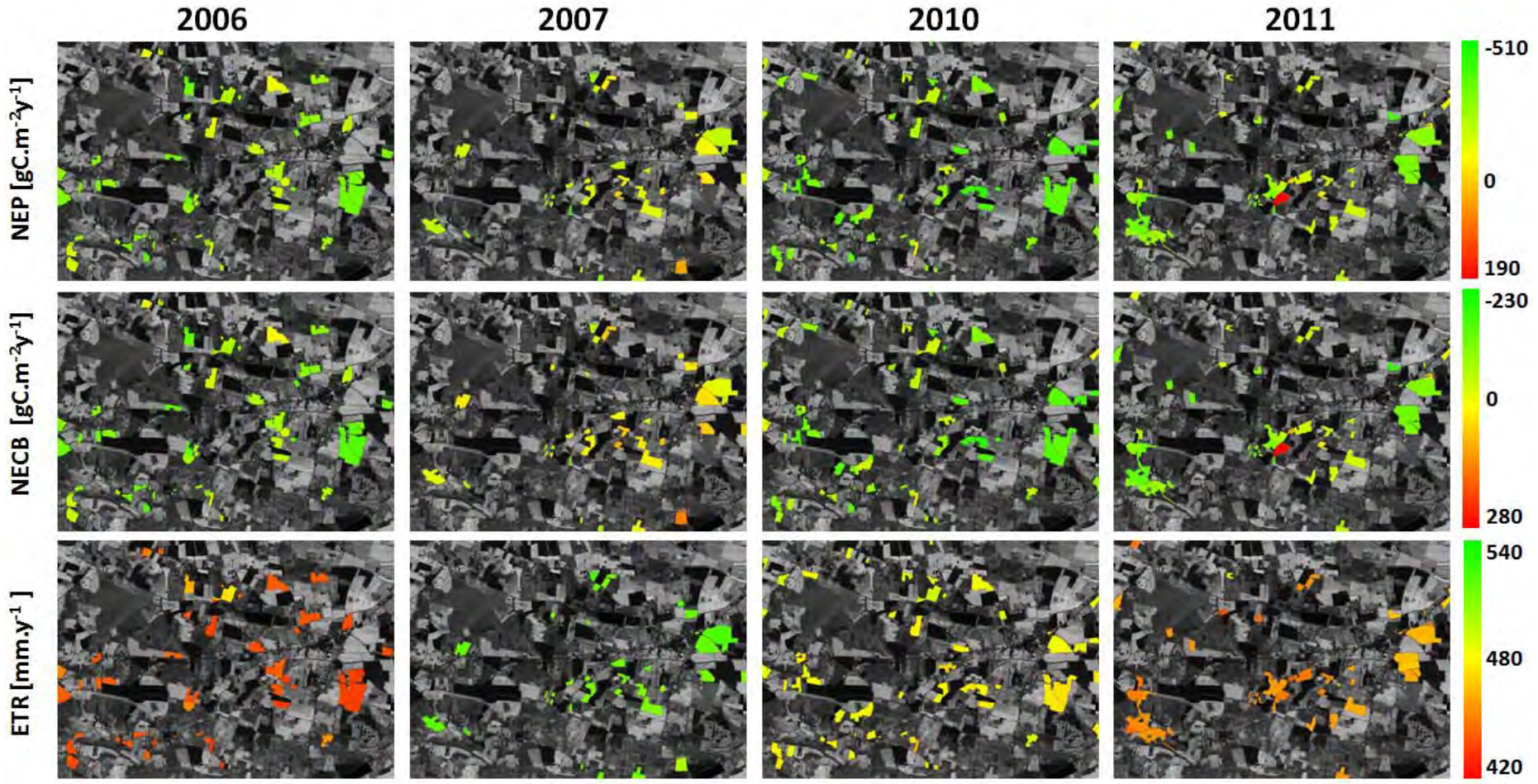


Figure 5.63- Maps of net ecosystem production at the end of the season (NEP, top), net ecosystem carbon budget (NECB, middle) and cumulated evapotranspiration (ETR, bottom) estimated by the SAFYE-CO_2 model for a number of plots within the study area over the 2006, 2007, 2010 and 2011 years. The color scales are displayed on the right of each series.

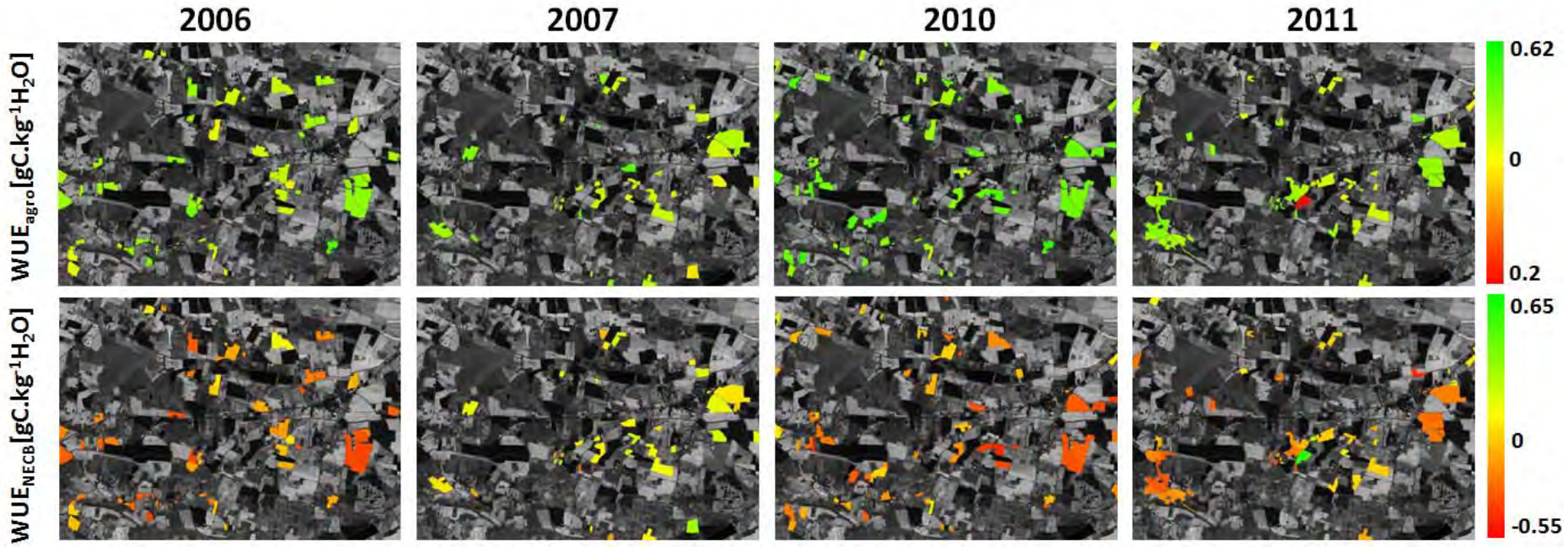


Figure 5.64- Maps of agronomical water use efficiency (WUE_{agro}, top) and environmental water (WUE_{NECB}, bottom) estimated by the SAFYE-CO_2 model for a number of plots within the study area over the 2006, 2007, 2010 and 2011 years. The color scales are displayed on the right of each series.

Table 5.7- Summary of statistical indicators: minimal (min), maximal (max), mean and standard-deviation (std) values calculated over a number (*n)* of plots for each of the investigated years (2006, 2007, 2010, 2011) for the following variables: maximal GAI, final biomass, grain yield, net ecosystem production (NEP), net ecosystem carbon budget (NECB), annual cumulated evapotranspiration during the crop season (ETR) and agronomical and environmental water use efficiencies (WUE_{agro} and WUE_{NECB}). The last two lines present the mean and std of each indicator considering the four years jointly.

	Max. GAI [$m^2.m^{-2}$]				Final Biomass [$g.m^{-2}$]				Yield [$t.ha^{-1}$]			
Year	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**
2006	0.73	3.76	2.44	0.20	577.8	1337.5	1049.4	47.7	2.60	6.02	4.72	0.21
2007	1.49	3.84	2.77	0.20	896.5	1518.6	1207.3	44.4	4.03	6.83	5.43	0.20
2010	1.04	3.74	2.80	0.19	896.5	1491.1	1230.3	45.9	3.27	6.71	5.54	0.21
2011	0.50	3.98	2.76	0.34	458.0	1295.2	996.6	62.1	2.06	5.83	4.48	0.28
Mean	*0.94*	*3.83*	*2.69*	*0.23*	*707.2*	*1410.6*	*1120.9*	*50.0*	*2.99*	*6.35*	*5.04*	*0.23*
STD	*0.43*	*0.11*	*0.17*	*0.07*	*224.0*	*110.7*	*115.5*	*8.1*	*0.85*	*0.50*	*0.52*	*0.04*

	NEP [gC $m^{-2}y^{-1}$]				NECB [gC $m^{-2}y^{-1}$]				ETR [$mm.y^{-1}$]			
Year	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**
2006	-467.5	45.3	-312.3	32.6	-231.49	158.67	-101.32	24.64	421.80	502.37	451.94	8.49
2007	-479.0	-33.6	-230.2	27.8	-184.64	148.11	13.48	20.56	499.00	536.55	518.06	3.17
2010	-502.3	4.3	-327.0	28.1	-223.60	148.66	-78.59	24.16	449.94	512.59	482.87	4.06
2011	-482.5	186.4	-218.6	50.7	-220.67	274.96	-18.59	38.24	439.71	498.88	468.97	5.02
Mean	*-482.8*	*50.6*	*-272.0*	*34.8*	*-215.10*	*182.60*	*-46.25*	*26.90*	*452.61*	*512.60*	*480.46*	*5.18*
STD	*14.5*	*96.1*	*55.5*	*10.8*	*20.82*	*61.76*	*52.95*	*7.78*	*33.04*	*17.00*	*28.08*	*2.33*

	WUE_{agro} [$gC.kg^{-1}H_2O$]				WUE_{NECB} [$gC.kg^{-1}H_2O$]			
Year	**Min**	**Max**	**Mean**	**Std**	**Min**	**Max**	**Mean**	**Std**
2006	0.27	0.58	0.48	0.02	-0.53	0.35	-0.22	0.06
2007	0.36	0.60	0.48	0.02	-0.35	0.30	0.03	0.04
2010	0.32	0.61	0.53	0.02	-0.47	0.32	-0.16	0.05
2011	0.21	0.55	0.44	0.03	-0.45	0.61	-0.04	0.08
Mean	*0.29*	*0.59*	*0.48*	*0.02*	*-0.45*	*0.39*	*-0.10*	*0.06*
STD	*0.07*	*0.03*	*0.04*	*0.00*	*0.07*	*0.15*	*0.11*	*0.02*

In order to have a closer look at the SAFYE-CO_2 estimates we selected a small area within the study area (indicated by the black rectangle in Figure 5.61) and then we produced a color map for each of the output variables. The color maps scales were fixed according to the minimum and maximal values found for each variable from the ensemble of the results (Table 5.7). The obtained maps are displayed in Figure 5.62, Figure 5.63 and Figure 5.64. It is important to keep in mind that the Table 5.7 values were calculated from the whole scene (and not from the selected region).

Concerning the maximum seasonal GAI (GAI_{max}) estimated by SAFYE-CO_2, we observe that the average GAI_{max} is similar for the four investigated years and ranges from 2.44 $m^2.m^{-2}$ (in 2006) up to 2.80 $m^2.m^{-2}$ (in 2010). The same is valid for the

maximal $GAI_{max,}$ no strong differences are observed between the years. However for the minimal estimated GAI_{max}, we observe that in 2007 it was equal to 1.49 $m^2.m^{-2}$, while for the other years lower values were found. It is probably due to the warm winter temperatures of the 2007 crop season, which induced an early and strong development of the vegetation (as discussed before for Lamasquère 2007). Still, in general, the inter-annual variability of GAI_{max} is not pronounced, even if those years had contrasted climatic conditions. For example, 2006 and 2011 were rather dry years for winter crops, and plants were subjected to water stress, while 2007 and 2010 were rather wet years. On the opposite, our destructive measurements showed high variability in GAI. This lack of inter-annual variability in the modeled maximum GAI may be caused by saturation effects in satellite-derived GAI estimates (as discussed in section 5.3.1), which does not allow the model to reproduce the real GAI temporal and spatial variability in our area of study.

However, it can be noted that in 2006 the GAI values are slightly inferior compared to the other years. Besides, we observe that the fields of the 2011 year seem to have more spatial variability than for the other years. This might be due to the exceptional dry character of the 2011 winter wheat season, so the soil types (and soil reservoirs) play a determinant role over the crop development, that may lead to higher inter-plots variability.

The analysis of the GAI plots within our selected region (Figure 5.62, first line) also shows that various of the plots cultivated with wheat in 2006, were also wheat plots in 2010; and a number of the 2007 wheat plots are present in 2011 as well. This illustrates a double-crop rotation, practice often adopted by the farmers in our study area.

For the final biomass production estimates, the values ranged from a minimum of 458 $g.m^{-2}$ (2011) to a maximum of 1518 $g.m^{-2}$ (2007). The mean biomass values varied from 996 $g.m^{-2}$ for 2011 to 1230 $g.m^{-2}$ for 2010. The standard deviation between biomass estimates from plots of the same year is not very strong (about 50 $g.m^{-2}$). However, from these min/max ranges the effect of the climatic conditions over the biomass estimates can be seen. According to the model, the lower biomass estimates were obtained for the dryer years (2006 and 2011), while for the wet years (2007 and 2010) the estimates are slightly superior. By the analysis of the maps, it can be noted that the strongest values of biomass estimates are related to those of GAI estimates; and the same for the low values: plots having low biomass, in general produce lower GAI. In 2011, a particular field captured out attention because of its low biomass estimate. This plot barely reaches a GAI_{max} of 1.5 $m^2.m^{-2}$ (which is low for winter wheat), inducing a final biomass estimate around 700 $g.m^{-2}$. Furthermore, since the yield here is estimated from the final biomass by a constant harvest index, the yield maps have basically the same characteristics as the biomass ones.

For the maximum GAI, biomass and yield outputs the model provides realistic estimates over the whole study area. For instance, concerning yield, regional statistics in Haute-Garonne give average values ranging from 4.2 to 5.5 $t.ha^{-1}$ between 2006 and 2010, which is in god agreement with the mean values provided by the model estimates.

Concerning the net CO_2 fluxes estimates, the cumulated NEE varies from -502 gC $m^{-2}y^{-1}$ (in 2010, strong net CO_2 assimilation) up to 186 gC $m^{-2}y^{-1}$ (2011, carbon release). According to the investigated year, the mean cumulated NEE may be quite different, *e.g.* NEP=-219 gC $m^{-2}y^{-1}$ for 2011 and about -312 gC $m^{-2}y^{-1}$ for 2006. Still, the mean NEP is negative for the four years, representing a sink for atmospheric CO_2 by the crops. Even so, some plots presented positive NEP, representing a CO_2 source, but these plots are few. These results are coherent with those observed by Béziat et al., (2009) for the crop rotations of the Auradé and Lamasquère sites. In this study, all NEPs were negative, corresponding to CO_2 sinks, which is also consistent with other studies on winter wheat and triticale (Ammann et al., 1996; Anthoni et al., 2004; Baldocchi and Collineau, 1994).

In Ceschia et al., (2010), on average, winter crops had rather similar NEPs, with -292 ± 170 gC $m^{-2}y^{-1}$ (for n=13 observations) and -358 gC $m^{-2}y^{-1}$ (n = 2), for common winter wheat and durum wheat, respectively, located over different sites in Europe. When considering all the wheat varieties together, the mean NEP and associated standard deviation was -326 ± 132 gC $m^{-2}y^{-1}$. Figure 5.65 shows the dynamics of cumulated NEE for the different sites investigated by Kutsch et al., (2010). High variability in cumulated NEE can be observed between sites and years. The range of values of the NEP simulated by SAFYE-CO_2 is included within the values reported by the Ceschia and Kutsch studies. When considering the estimates of the four years over our study area, we obtain an average NEP of -272 ± 55 gC $m^{-2}y^{-1}$. Even if this value is lower than that found by Ceschia et al., (2010) (-326 gC $m^{-2}y^{-1}$), it is included in the range of variation, due to the strong variability between plots (std=132 gC $m^{-2}y^{-1}$) in their study. This strong variability can be caused by contrasted climatic conditions and management regimes.

In Ceschia et al., (2010), the high NEP variability found for different sites cultivated with the same crop was largely caused by differences in latitude and climate variability influencing the length of the growing season and the amount of carbon assimilated. The effect of the length of growing season on NEP could be observed for the Lamasquère experimental site in 2007 compared to the other site-years. As it was discussed in previous sections, the 2007 crop season had a warm winter, inducing an early development of the vegetation. Since the growing season lasted longer, the period of carbon assimilation has lasted longer as well. At a regional scale, surprisingly, this behavior is not reproduced, since the mean NEP for

2007 is about -230 ± 28 gC $m^{-2}y^{-1}$, which is low if compared to other years (-327 ± 28 gC $m^{-2}y^{-1}$ in 2010 and -312 ± 33 gC $m^{-2}y^{-1}$ in 2006). On other plots, the warm winter may have led to higher Reco and thus lower NEP values. In addition to the climatic conditions, factors, such as management (*e.g.*, fertilization, amount of residues decomposing from the previous crop, crop varieties…) and soil properties, may also be involved in NEP variability and are discussed in Kutsch et al., (2010), Moors et al., (2010) and Eugster et al., (2010).

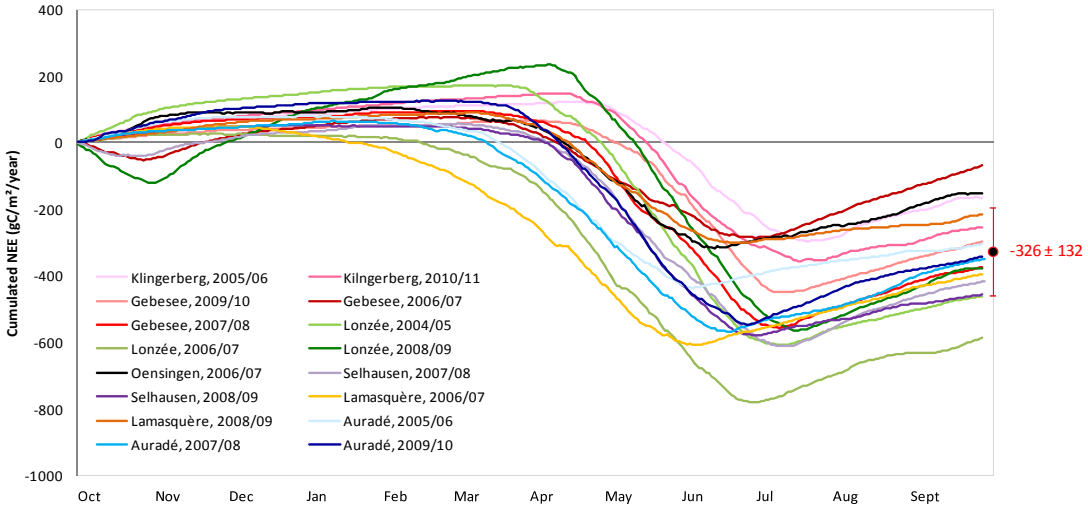


Figure 5.65- Profiles of cumulated NEE for winter wheat crops located in different sites in Europe. Comparison performed by Ceschia et al., (2010), in the scope of the CarboEurope-IP project. The mean cumulated NEE reached at the end of the season (NEP) and associated standard deviation are indicated.

In terms of carbon budgets, the average NECB calculated for each year for our study area indicates that, except for 2007, the wheat ecosystems acted as CO_2 sinks, presenting negative NECB. We obtained NECBs of -101 ± 25, -78 ± 24 and -19 ± 38 gC $m^{-2}y^{-1}$ for 2006, 2010 and 2011, respectively. For 2007, NECB was 13 ± 38 gC $m^{-2}y^{-1}$, being close to neutrality. Furthermore, when evaluating the minimal and maximal NECB values found for each year, we observe that some plots can have very different NECB (spatial variability) even if the climatic conditions on these plots were essentially the same. The largest interval of variation was found for 2011, for which plots NECB varied between -220.67 and 274.96 gC $m^{-2}y^{-1}$. For the SAFYE-CO_2 results presented in this section, we considered the hypothesis that only grain was exported and straw was left on site, what is true for the majority of fields within our zone of study. However, it is worth noting that in 2011 the straw was exported on many plots of our study area in order to compensate the lack of forage production at the

national level. Accounting for straw removal at some plots would have increased the spatial variability in NECB estimates in 2011.

To simulate the impact of this change in the straw management over the carbon budget, we recalculated the NECB for the 2011 plots by considering that the Cexp term was composed of grains plus a fraction of the straw (same method as described for the Lamasquère site in 2007 and 2009, see sections 4.2.2 and 5.3.2). For the first case, when only the grains are exported, the average NECB for 2011 was of -18.6 ± 38.2 gC $m^{-2}.y^{-1}$, representing a small CO_2 sink. For the second hypothesis (grain and straw exported), the mean NECB raise to 659.9 ± 50.7 gC $m^{-2}.y^{-1}$, representing a strong CO_2 source, showing the significant impact that this management practice may have on carbon budgets.

On average, the NECB for the different crop species examined by Ceschia et al., (2010) was 138 ± 239 gC m^{-2} y^{-1}, corresponding to a carbon source, but the variability in the observations (caused by differences in straw, organic fertilization management and climatic gradients) surrounding this estimate was larger than the source itself. Considering a mean soil organic C content of 5300 gC m^{-2} (53 t of organic C ha^{-1} to a depth of 30 cm; Smith et al.,(2000)) in European agricultural soils, the mean NECB would correspond to an annual loss of 2.6 ± 4.5% of the soil organic C content. Of course, this value should be considered with caution because the crop species, soil conditions and management practices in the Ceschia's study are probably not fully representative of all croplands found in Europe (see Osborne et al.,(2010)). Kutsch et al., (2010) found slightly lower but still positive values for the NECB (91 ± 203 gC m^{-2} y^{-1}) at eight sites with at least four years of continuous measurements. For our study, focused on the winter wheat crop, the average estimated NECB was -46 ± 53 gC $m^{-2}.y^{-1}$, corresponding to a carbon sink, but close to equilibrium and the variability of this term was higher than the estimate itself. In terms of soil organic C content, our results on winter wheat would correspond to an annual gain of 0.87 ± 1%. This variability was caused mainly by differences in climatic conditions, soil characteristics and some of the management practices (as dates of sowing and harvest). The management regimes concerning organic fertilization and straw removal were assumed to be similar for all the investigated plots, since this information is not available at the regional scale, and thus do not represent a source of variability between plots and years.

Even if our results contrast with those found by Ceschia et al. (2010) and Kutsch et al,. (2010), they agree with studies based on modeling and carbon inventories approaches that suggest that European cropland soils are close to equilibrium, being either small sources (Bondeau et al., 2007; Schulze et al., 2009; Smith et al., 2005) or small sinks (Gervois et al., 2008). Besides, studies from Ceschia et al., (2010) and Kutsch et al., (2010) included both winter and summer crops. The last ones tend to

act as carbon sources because of their short growing season. Here, the estimates were integrated over an agricultural year (from early October to late September). Longer integration periods, including crop rotations, would be necessary to assess the net biome production (NBP) and therefore to evaluate climatic variability effects on the NECB and NBP. However, because the detection of short-term changes in soil carbon stocks using conventional means (soil analysis) is problematic (Garten and Wullschleger, 1999), especially on large areas, and generally requires even longer integration periods to detect significant soil C changes (Smith, 2004), there is a real need for similar studies to evaluate the potential of croplands to store or release carbon under different soil conditions, crop species and management regimes.

The NEP is the result of two large and opposite terms, primarily photosynthesis (GPP) and ecosystem respiration (Reco). Therefore, a small increase in GPP or reduction in Reco would improve NEP, and consequently NECB, in a noticeable way. It was reported by Béziat et al. (2009) that on some occasions, re-growth events and weed development increased the number of days of active vegetation cover, increasing the net carbon assimilation. However, events are very dependent on climate and are usually interrupted by soil preparation prior to the sowing of the next crop. Béziat et al. (2009) estimated that re-growth events and weed growth caused a net fixation of approximately 50 gC m^{-2} after triticale at Lamasquère in 2005-2006. This re-growth occurred because the summer was relatively wet and because soil preparation occurred late in the season. Soil preparation, disking, stubble cultivation and use of herbicides may delay, prevent or interrupt voluntary re-growth and kill weeds. Therefore, postponing the operations or encouraging cover crops can improve the NEP term and thus improve carbon budget of agricultural ecosystems. The use of cover crops seems to be a powerful potential lever for improving agricultural plots carbon budget according to Ceschia et al., (2010).

Climate and management can cause large differences in yield, NEP and NECB among sites, even for the same or similar crops. In general, organic fertilization tended to improve the NECB of the present crop, but its effect on the subsequent crops is more difficult to assess. In most cases, the harvest index and the fate of the harvestable product drives the proportion of NPP that will be exported, thereby influencing the NECB. The C export term is usually the term having the biggest impact on the NECB, driving the plot to act as a source or a sink (Béziat et al., 2009; Ceschia et al., 2010; Kutsch et al., 2010). For farms specializing in cereal production, it is more likely that only the grains will be exported so that most of the biomass (approximately two thirds of the total biomass including roots) produced in the field could potentially remain there, with most of it being progressively decomposed and a small part of it increasing the soil carbon pool (see Osborne et al., 2010). However,

in a number of situations, baled straw may be removed for commercial and/or local reasons. Fields where the biomass is exported for silage or biomass energy will lose most of the carbon fixed by the plant during the growing season at harvest. Even if this loss of carbon may partly be compensated by animal manure application, it is more likely that the NECB will correspond to carbon losses from the soil (Ceschia et al., 2010). Thus, even a small increase in the exported carbon (Cexp) would substantially degrade the NECB. For this reason, the straw management (removal of not) play an important role over the carbon budget estimates.

Concerning the water consumption (cumulated ETR) by the winter wheat plots at regional scale, we observe low variability between the plots for a given year and a tendency of higher values of ETR_{cum} for wet year. For 2006 and 2011, characterized as dry years, the model yields to an average ETR_{cum} of 451.9 ± 8.5 and 469 ± 5 $mm.y^{-1}$ respectively. For 2010, a wet year, slightly higher values were estimated, with a mean ETR_{cum} of 482.9 ± 4.1 $mm.y^{-1}$. For 2007, the ETR is even more important: 518.1±3.2 $mm.year^{-1}$, probably due to stronger and earlier vegetation development this year and longer length of the vegetative cycle, which makes the transpiration term increase. Even if mean ETR_{cum} values for each year are not very different (maximal difference of 66 $mm.y^{-1}$), the inter-annual variability can be clearly seen for the selected area in Figure 5.63 (last line). On average, when considering all years together, a mean ETR_{cum} of 480.5 ± 28.1 $mm.y^{-1}$ is found. This value is within the interval observed for the ETR measured over the experimental sites from 2006 to 2010, which ranged from 364 to 515 $mm.y^{-1}$.

Once the yield, NECB and water consumption were computed, the water use efficiencies indicators could be calculated. For the agronomical WUE, the mean WUE_{agro} calculated for each year ranged from 0.44 ± 0.03 gC $kg^{-1}H_2O$ for 2011 to 0.53 ± 0.02 gC $kg^{-1}H_2O$ for 2010. The overall WUE_{agro} calculated from the four years was 0.48 ± 0.04 gC $kg^{-1}H_2O$, showing that similar performances are obtained independently of the year. Those stable WUE_{agro} values are within the range of what was found in Katerji et al., (2008) but lower than what was found experimentally at our sites (about 0.8 in 2006 and 1.0 in 2007) in Tallec et al., (2013). This difference might be partly explained by the yield underestimation in 2007 because of GAI saturation issues. As WUE_{agro} can be interpreted as the amount of produced biomass that is exported out of the plot (in terms of carbon) for a kg of water consumed, the more important this indicator is, the less water is necessary for producing a given amount of yield. Besides, we observe that for our small selected zone (Figure 5.64) no particular tendencies are observed between the different years.

For the environmental WUE_{NECB}, depending on the year the values ranged from a minimum of -0.53 gC $kg^{-1}H_2O$ in 2006 up to 0.61 gC $kg^{-1}H_2O$ in 2011. The mean

WUE_{NECB} calculated for all years is -0.10 ± 0.11 gC $kg^{-1}H_2O$. As the uncertainty indicates, a great variability was found between the investigated years. The best WUE_{NECB} was found for the 2006 year (-0.22 ± 0.06 gC $kg^{-1}H_2O$), meaning that, on average, each plot assimilated 0.22 gC per kg H_2O lost during the year. The worst year from an environmental perspective was 2007, for which each plot lost on average 0.03 gC per kg H_2O lost during the year. These trends can be visualized in the color maps displayed in Figure 5.64.

At Auradé in 2006 and Lamasquère in 2007, WUE_{NECB} values of -0.1 and 0.6 gC kg^{-1} H_2O were found experimentally. Even if the modeled values differ from the experimental data for those two site-years, mainly because of inaccuracy in NECB estimates, the range of modeled values in our area is realistic when compared with the literature (Tallec et al., 2013).

In Sinclair et al., (1984), several options were discussed to improve WUE_{agro}, one of the most realistic being to improve harvest index. However, improving the harvest index would probably decrease WUE_{NECB} because less biomass would remain on the field, leading to less carbon input into the soil (Tallec et al., 2013). Another way to improve both agronomical and environmental WUEs would be to reduce water loss through soil evaporation. Solutions have been proposed to reduce this term, including straw mulching, which can significantly reduce evaporation during the growing season (Li et al., (2008), Tallec et al., 2013). Varietal selection could also limit evaporation by selecting plants that rapidly cover the soil at the beginning of the growing season (Passioura, 2006) or by selecting plants that tolerate high sowing density (Ritchie and Basso, 2008). Tallec et al., (2013) showed that the annual WUE_{NECB} could be improved by limiting fallow periods and increasing CO_2 fixation. Indeed, the solution of an intermediate crop could limit the annual loss of E in ETR and increase gain of carbon by increasing the period of photosynthetic activity. Of course, some of these agricultural practice options should be carefully considered because they could increase TR and therefore limit the development of the next crop by mobilizing water resources and available nitrogen. Furthermore, these options should be analyzed from economic and societal perspectives and on the basis of technical feasibility.

5.7. Discussion concerning the potential improvements of carbon and water fluxes modeling with SAFYE-CO_2

Many measurements are necessary to assess the full carbon and water budgets of croplands, given that they are very diverse in terms of crop species grown, rotation, management, soil types, and climatic conditions (Smith et al., 2010). Through the work presented in this thesis, we aimed at developing a robust remotely-sensed based modeling approach that requires few input data and that is capable of correctly simulating the crop development and production, but also of assessing accurate crop carbon and water budgets estimates at regional level. Since it is difficult to directly assess results at large scales, we used the valuable dataset available over our study area to test this approach. It allowed us validating the models and better understanding the processes at plot scale; for then up-scaling it to larger croplands surfaces. Still this approach suffers from two main methodological constraints.

Firstly, if the carbon input by organic fertilization is essential for assessing the carbon budget, it remains an extremely complex term to estimate at regional scales, since there are no effective methodologies allowing to estimate it at the plot scale (for instance by remote sensing). Consequently, the open access to regional/ national databases, inventories and field expertise concerning organic fertilization supplies seem essential for better estimating budget at larger scales. Secondly, we have found that there is still a lot of uncertainty around the carbon export estimates. Even if the model is able to correctly estimate the grain yield contribution, it remains a challenge to first, identify the harvest practices, by detecting if only seeds or if seeds and straws are exported from the plot; and second: to quantify the amount of exported straws, when it is the case. Still some improvements are possible, for example: a) better validation of the model by the comparison of the yield estimates with the yield maps (from the combine harvester), at intra-field scale; and b) estimation of the straw removal or not by the combined use of optical and radar observations (need to be investigated).

Another key issue concerning the accuracy of the carbon budget estimates at regional scale concerns the ability of the model to correctly simulate the NEP term. In order to compare some of our results with other studies, we select the interesting study carried out by Wattenbach et al., (2010). In this paper, the authors present a multi-site model comparison for four cropland ecosystem models: DNDC (Li et al., 1992; Li et al., 2005), ORCHIDEE-STICS (de Noblet-Ducoudré et al., 2004; Gervois et al., 2008), CERES-EGC (Gabrielle et al., 2006; Lehuger et al., 2009) and SPA models

(Sus et al., 2010; Williams et al., 1996). Amongst others, they compare the accuracy of the models in estimating NEE, GPP, Reco, and also ETR for winter wheat and maize derived from eddy covariance measurements on five European sites. One of the investigated sites is the Auradé site, for the winter wheat 2006 season, that was also studied through this thesis.

Figure 5.66 presents the cumulative NEE fluxes measured by EC and those simulated by the four above cited models (extracted from Wattenbach et al., (2010)), plus those simulated by the SAFY-CO_2 and SAFYE-CO_2 models, for the Auradé site in 2006. The qualitative analysis of these results shows that the models developed through the thesis provide good NEE estimates relative to the other models. We note a mismatch for all models in the early stage of the growing season, when low fluxes predominate (Wattenbach et al., 2010). Then, during the mid-growing period they have an overall good NEE representation, except for SPA. From approximately DOY 140 the models trajectories tend to diverge. Besides, a common pattern for the six models is the difficulty to reproduce the fluxes during senescence and post-harvest periods.

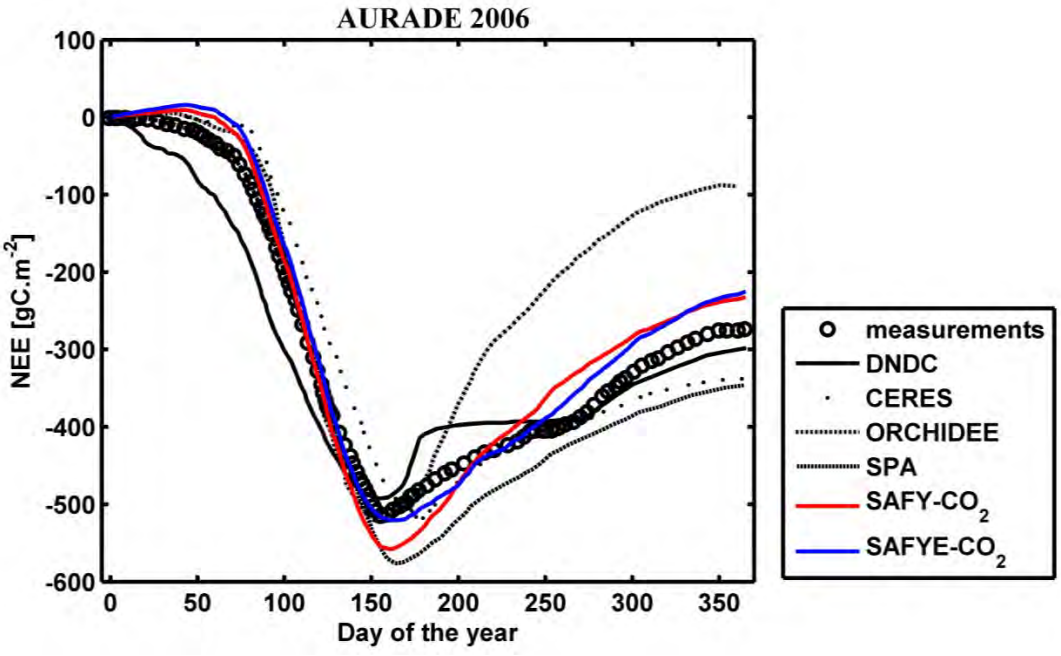


Figure 5.66- Comparison of the cumulative NEE for the winter wheat at the Auradé site year 2006 between in-situ flux measurements and different models: DNDC model (grey solid line), ORCHIDEE-STICS (dark grey dash dot line), SPA (black dash line) and measurements indicated by open circles [extracted from Wattenbach et al., (2010)], SAFY-CO_2 (red line) and SAFYE-CO_2 model (blue line) [found through this thesis].

For other components of the cropland carbon and water cycles, here in particular the Reco, GPP and ETR, different statistical indicators were used by both studies for evaluating the models estimates. Still, we can take a 'consultative' look at the correlation coefficients obtained by both studies (Kendall coefficient for Wattenbach and Pearson coefficient this study, see Table 5.8). We note that the results obtained using SAFY-CO_2 and SAFYE-CO_2 for NEE, GPP, Reco and ETR are rather good when compared to the other models.

Table 5.8- Correlation coefficients of the investigated models.

	NEE	GPP	Reco	ETR
DNDC*	0.68	0.59	0.38	0.43
ORCHIDEE-STICS*	0.57	0.68	0.35	0.59
CERES*	0.63	0.64	0.62	0.68
SPA*	0.63			
SAFY-CO_2°	0.92	0.95	0.88	
SAFYE-CO_2°	0.93	0.96	0.88	0.85

* r: Kendall correlation coefficient; ° R: Pearson correlation coefficient.

As stated by Wattenbach et al., (2010), a number of studies have evaluated SVAT models for simulating carbon fluxes from croplands (Adiku et al., 2006; Huang et al., 2009; Wang et al., 2007). However, the combined evaluation of water and carbon fluxes remains relatively rare in the literature (Adiku et al., 2006), especially for cropland ecosystems.

We could see that models like SPA and ORCHIDEE-STICS are highly accurate in simulating net carbon fluxes (NEE) and water fluxes. However, they are only able to simulate the seasonal carbon balance of a limited number of crops with a limited consideration of management (Wattenbach et al., 2010). On the contrary, models like CERES-EGC and DNDC are less accurate in simulating NEE and especially poor in reproducing latent heat fluxes. They are however capable of simulating full crop rotations and of considering the associated management effects.

Another interesting study using the SPA model (with modifications) was performed by Revill et al., (2013). In this paper the authors aimed at evaluating a framework for the assimilation of leaf area index (LAI) observations, retrieved from both optical and radar (SAR) sensors, in the SPAc model for simulating crop development and C fluxes for European landscapes. Amongst the various investigated sites, part of the CarboEurope-IP project, they also used the flux data from the Auradé (2006 crop season) and Lamasquère (2007) sites. The great innovation of this study was to combine data derived from high resolution optical and SAR sensors, thus increasing the number of available observations for improving the model performance. It also showed that LAI derived from SAR data are less

sensitive to saturation effects than the ones derived from optical sensors. Therefore, it confirms our future objective of adding radar data to our modeling approach for better estimation of GAI temporal series.

Furthermore, we can compare the performances of their approach against ours in terms of NEE estimates since we have two common data sets: Auradé 2006 and Lamasquère 2007. Table 5.9 presents the performances in terms of correlation coefficient (R^2), regression slope and root mean square error (RMSE) for the daily NEE estimates. We observe that the performances of both approaches are quite close. For AUR2006, the SAFYE-CO_2 model presented the best performances in terms of slope and RMSE, and SPAc in terms of correlation. For LAM2007, SAFY and SAFYE-CO_2 present better RMSE and correlation with measurements, however the SPAc has a regression slope closer to 1. These results show that two models from different natures (SPAc: an ecosystem model adapted for crop-specific parameterization and SAFY(E)-CO_2: a semi-empirical crop model), when driven by remote sensing products, can yield to good estimates of the components of the carbon budget (NEE, yield (not shown)).

Table 5.9- A comparison between the SPAc model performances (Revill et al., 2013) and those of SAFY-CO_2 and SAFYE-CO_2 model for the NEE daily and cumulated estimates for the 2006 crop season at the Auradé site and the 2007 season at the Lamasquère site. Differences in measured cumulated NEE between the two studies are caused by differences in periods of integration.

	R^2			**Slope**			**RMSE [gC m^{-2}.day^{-1}]**		
	SPAc	***SAFY-CO_2***	***SAFYE-CO_2***	***SPAc***	***SAFY-CO_2***	***SAFYE-CO_2***	***SPAc***	***SAFY-CO_2***	***SAFYE-CO_2***
AUR2006	0.91	0.85	0.86	0.87	1.03	1	1.56	1.21	1.12
LAM2007	0.82	0.87	0.88	0.92	0.81	0.81	1.98	1.06	1.16

Cumulative NEE

	Measured SPAc study	***SPAc***	***Measured this study***	***SAFY-CO_2***	***SAFYE-CO_2***
AUR2006	-476	-377	-323	-244.5	-242.4
LAM2007	-549	-602	-374	-317.9	-134

Most of those models however require a large number of input parameters (*e.g.* around 200 for STICS) that are generally very difficult to obtain over large areas and at a fine resolution compatible with the analysis at the plot scale. For instance, most of those models require information concerning management like day of sowing, dates and dose of fertilization, etc... Additionally, none of these models (including SAFY-CO_2 and SAFYE-CO_2) were able to reproduce the development of weeds o re-

growths events, like in summer 2006 at Auradé (they were even more pronounced at Lamasquère in 2005 and 2011, not shown here) and their impact on NEE and the final carbon budget. In 2005 at Lamasquère, the re-growth event that followed the triticale crop caused a 50gC net fixation (Béziat et al., 2009), which is not negligible. Spontaneous re-growth, development of weeds or the addition of an intermediate crop in the crop rotation are seldom or never accounted for in the modeling of CO_2 fluxes, even if Ceschia et al., (2010) showed that they significantly improve the carbon budget of crops by compensating part of the heterotrophic respiration during their development phase or by even allowing a net CO_2 fixation.

A modeling approach based on remote sensing could allow detecting and then accounting for such events in the modeling process. Therefore, the modeling approach proposed here with the SAFYE-CO_2 model consists in a first step for filling in the gap that exists in obtaining an accurate representation of the main components of the cropland carbon and water budgets over regional scales, by adopting a simplified but robust modeling methodology that requires few input parameters.

Still, from the modeling results we could identify some issues that could be improved in the proposed approach. Therefore, next steps would be:

a) To better adjust the *fELUE* function (that accounts for the effect of the fraction of diffuse radiation over photosynthesis), for avoiding the over-correction of the diffuse radiation effects for clear sky conditions (cf. section 5.3.1.1). This adjustment would have an effect over the biomass/yield estimates and also over the CO_2 flux estimates;

b) To include the effects of post-harvest events (re-growths, weeds, inter-crops). For this, it will be necessary to carry out a sensitivity analysis concerning the detection of these vegetation types from remote sensing observations;

c) To include radar data assimilation for better estimating GAI for strong vegetation development (when GAI from optical remote sensing data saturates) and also for filling eventual gaps in optical GAI series (as the gaps observed in 2008 and 2009 series in our data set, which avoided us of having regional analysis for these years), as it was shown by Revill et al., (2013).

d) To exploit the potential offered by the synergy of high spatial and temporal resolutions optical images (SENTINEL-2, Venus, Landsat, SPOT, …) with radar data (SENTINEL-1, RADARSAT, ALOS) for identifying the straw management (if removed of left on site) and for a more precise estimation of the date of crop emergence, which is a critical parameter for the SAFY model and the following versions presented in this thesis (SAFY-CO_2, SAFYE-CO_2). Indeed we could see in sections 5.3.1 and 5.4.2 that this parameter is essential for estimating

proper CO_2 assimilation during the first phases of the winter wheat development (see Figure 5.31, Figure 5.35, Figure 5.36).

e) To improve the modeling of the heterotrophic respiration (Rh) by the coupling of the SAFYE-CO_2 model with a more complex model like CENTURY (Parton et al., 1987) that is able to simulate the evolution of the organic carbon content in the soil. This topic would be related to the previous one (d), concerning the straw management.

f) To estimate crop carbon and water budgets over complete crop rotations. In this thesis we presented the work done for the winter wheat. Subsequently, we plan to apply the developed approach over other important crop types as maize, sunflower and soybean.

g) To use the full potential of high resolution data. For example, of the yield using the yield maps (from combine harvesters) for better model validation for different crops at pixel scale. Another interesting application would be to achieve a more precise validation of the flux estimates by weighting the model outputs according to the flux measurements footprint. A preliminary study has been performed on this topic for the Auradé site (summarized on the section 2.3.1.4).

Conclusions & Perspectives

The overall objective of this thesis was to establish and evaluate an approach for estimating the production and the main components of the carbon and water biogeochemical cycles of crop ecosystems over contrasted climatic years, from the local up to the regional scales, focusing specifically on the winter wheat crop. For this, we proposed a regional modeling approach combining high spatial and temporal resolutions remote sensing data with a simple crop growth model and a water balance module, and we used an extensive set of *in-situ* measurements for the calibration and validation of the approach.

The first part of this thesis was dedicated to the retrieval of the green area index (GAI) from temporal series of Formosat-2 and SPOT satellites images. For this, two methods were tested. The so-called empirical method consisted in establishing an exponential relationship linking the normalized difference vegetation index (NDVI) and *in-situ* measurements of GAI obtained from processing of hemispherical photographs (GAI_{eff}) using the CanEye software. The second approach was a physical one, based on the inversion of a radiative transfer model (PROSAIL, Baret et al., (1992a)) using neural networks. In this case, *in-situ* data are not necessary for calibrating the model. For the validation of this method, we also used *in-situ* GAI_{eff} data for the main crops grown in southwest France: wheat, maize, sunflower, and soybean. The investigated methods yielded to similar performances, with a RRMSE about 29% and correlation coefficient (R^2) of 0.86. The empirical relationship NDVI-GAI_{eff} is known for saturating for strong values of NDVI, but as the GAI empirical estimates were compared against *in-situ* GAI data estimated from hemispherical photographs that includes an exponential Poisson law (that also saturates), the saturation effect did not affected the performances of the empirical method. The physical approach, on the other hand, is supposed to be less sensitive to saturation issues, so its estimates were more related to the true values of GAI (destructive measurements) than those of the empirical method. Still, it should be noted that physically-based GAI remain underestimated when compared to the destructive measurements, because of the clumping effect. As the physical method does not depend upon any field measurements and since it is less sensitive to saturation effects, it offers great perspective for regional scale studies. Therefore, this method was chosen for establishing the GAI maps that were next integrated into the crop modeling approaches.

The crop modeling studies started with the original SAFY model (Duchemin et al., 2008), which is able to simulate the crop GAI, biomass and yield variables. The phenological parameters (day of emergence, start of senescence, partition-to-leaf,…) of SAFY were calibrated by using the temporal series of remotely-sensed GAI. Concerning the yield estimates, four different functions were tested (two based on harvest index and two based on the STICS model biomass allocation module), in order to identify the one providing the most accurate yield estimates. It is important to emphasize that en extensive field campaign was performed in the scope of this thesis in 2011. This *in-situ* campaign consisted in measuring winter crops biomass at maturity stage (just before harvest) and final grain yield as well over heterogeneous plots. This data set was essential for validating the yield estimated by the crop model since the only ground data existing so far were those provided by the farmers (through farmers' surveys). Indeed, the farmers data set present a high level of uncertainty and cover a low range of yield values (because of aggregation issue), making it of little use for validation of yield estimates. Later, the SAFY model could also be tested against a third data set corresponding to the yield maps provided by the yield monitors mounted on combine harvesters. These various available data sets showed us that yield ground data used for validating the crop model outputs must cover a large range of yield observations. The 'yield monitors' data set presents therefore an interesting spatial variability that could be used for model validation at the pixel scale. This issue should be further investigated in future works. In addition, we recommend strong effort for collecting yield data in a range of conditions (spatial and temporal) for the validation of this kind of modeling approach, especially, as yield estimates are a key component of the carbon budget.

Finally, our results show that the SAFY model is able of correctly reproducing the dynamics of GAI (RRMSE=12%), of biomass (RRMSE about 25%) and the yield (RRMSE=24% for the best method based on harvest index) for fields with contrasted climatic years (from 2006 to 2011) and under different management conditions. Still, the yield estimated by using the STICS module could be improved by better adjusting the yield-related parameters (temperature threshold for starting grain filling and evolving harvest index). This will require further works.

The SAFY-CO_2 model was developed for estimating seasonal dynamics of the components of the carbon budget: photosynthesis (GPP), autotrophic respiration (Ra), heterotrophic respiration (Rh), ecosystem respiration (Reco), net primary production (NPP) and net ecosystem exchange (NEE), in addition to the GAI, biomass and yield estimates already estimated with the SAFY version. In the original SAFY model, the Monteith's light-use efficiency theory is used for converting the photosynthetically active radiation absorbed by the plants directly into dry matter. For SAFY-CO_2 this function was modified, making the photosynthesis (GPP) the initial

process to be estimated. The remarkable advance in this methodology was the inclusion of a function that allows taking into account the effects of the fraction of diffuse radiation over canopy photosynthesis, by adapting the light-use efficiency of the canopy according to the irradiance conditions. The carbon budget components (GPP, Ra, Rh, …) were then computed. Firstly, Ra was calculated by estimating maintenance and growth respirations using standard equations from the literature, then Ra was subtracted from the GPP in order to calculate NPP. A simple Q_{10} function parameterized from the flux data of our experimental sites was used to calculate Rh and this term was added to the NPP, for calculating the NEE term.

Finally, the crop net ecosystem carbon budgets (NECB) were calculated and compared with the results from our experimental sites (Auradé and Lamasquère), by considering the simulated NEE, the carbon imported as organic fertilization and seeds (OF, data obtained from the farmers) and the carbon exported at harvest (Cexp) simulated by the model. However, some hypotheses had to be dressed for the other plots of our study area, since some of the required information for calculating NECB was not available. The first hypothesis concerns the straw management. At the moment this information cannot be derived from satellite observations and therefore it is not possible to differentiate the crop fields for which the straw is exported from those for which the straw is left on site. As a consequence, the estimates of the exported carbon term present high incertitude (reproduced on the carbon budgets). To overcome this limitation, future researches should exploit the combination of multispectral (optical and radar) and hyperspectral remote sensing data for providing the straw management information. Fortunately, for most of the plots in our area, straws are left on site (except in 2011), since most farms are cereal production farms. The second hypothesis that was made for the calculation of the carbon budget calculations concerns the organic fertilization input. Since it is not possible to obtain an estimate of this variable from remote sensing products and as few farms in our area are animal production farms we considered this term negligible in our regional approach and the OF term only corresponded to the amount of carbon in the seeds. The solution for areas where animal production is significant remains in accessing regional/national inventories and database to have average estimates of the OF term.

The net CO_2 fluxes components estimated with the SAFY-CO_2 were overall in agreement with the measurements, presenting very good correlations (R^2 about 0.9 for GPP, 0.75 for Reco and 0.84 for NEE). However, the analysis indicated that the performances of the model could be improved by considering weeds development and re-growths events after harvest or by estimating more accurately the remotely-sensed GAI for years with strong vegetation development, by using for instance radar data. Besides, an enhanced estimation of the date of emergence could yield to

better CO_2 fluxes estimates in periods of very low vegetation. The net ecosystem carbon budgets were then calculated for different climatic, environmental and management conditions. They allowed identifying the main factors by which they are influenced and the potential levers for improving carbon storage in the crop soils and thus reducing the impacts of crop production on climate change. Stronger divergences between NECB estimates and observations were found, ranging from 48 gC $m^{-2}y^{-1}$ up to 108 gC $m^{-2}y^{-1}$. This finding highlights the importance of having accurate estimates of each of the terms that compose the NECB especially the two main terms (NEE and Cexp) that largely compensate each other (because they are of opposite sign).

The last step of the modeling approach was to couple the SAFY-CO_2 model with a water balance module, based on the FAO-56 method. The resulting model was named SAFYE-CO_2. With the water module, the evapotranspiration (ETR) and the soil water content (SWC) estimates were added to the model outputs. This coupling allowed us to consider possible water stress effect on photosynthesis, to provide an estimate of the water consumption of the croplands (in form of ETR) and also to provide a more accurate representation of the soil heterotrophic respiration, by using the SWC for its computation. Besides, with the improved estimation of these terms we build the path for combined carbon and water cycles estimates, but also for agronomical and environmental indicators of the water use efficiency.

The SAFYE-CO_2 presented overall good results in terms of evapotranspiration estimates (RMSE about 0.46 mm.day^{-1} and R^2 around 0.74). In terms of annual crop water consumption, the model estimates and measurements presented absolute differences between 20 mm.$year^{-1}$ and 77 mm.$year^{-1}$, which can be considered as satisfactory. In terms of biomass, yield and CO_2 fluxes components estimates, the performances of SAFYE-CO_2 were slightly lowered compared to those of SAFY and SAFY-CO_2. This apparent “degradation” however mainly highlights the underestimation of GPP caused by the saturation of remote-sensing derived GAI for years with strong vegetation development. A possible solution to solve this issue would be to use GAI derived from radar data, since the radar signal is less sensitive to saturation at high values. Also annual GPP values were underestimated because re-growth events were not accounted for by the model.

Finally, from the computed yield, NECB and annual evapotranspiration, the agronomical and environmental WUE indicators could be calculated, allowing possible sustainable strategies to be considered for preserving the environmental resources and maintaining crop production.

From the work accomplished thorough this thesis, we could verify the potential of a simplified modeling approach, combining remote sensing observations with a semi-empirical crop model and a standard water balance module, in providing good

estimates of the main components of the carbon and water cycles. This work was possible thanks to the availability of an extensive set of remote sensing data and of ground data (*in-situ* measurements and farmers inventories databases) over the study area for consolidating and validating our modeling approaches.

Now that the approach is well established, its main advantages and drawbacks could be identified and analyzed. The next step would be to apply it to other crop species for our area of study and to improve it by considering the effect of intermediate crops, re-growths events and weed development on the CO_2 and water fluxes estimates, since those events prove, for instance, to have a significant effect on the annual NECB of the plots. The use of radar data for better estimation of GAI during strong cover development and in case of gaps in optical observations (cloudy periods) would surely improve the estimates of the components of the carbon and water budgets.

In the future, this approach could be extended to a global scale thanks to the upcoming satellite missions (Sentinel-1 and 2, …) and it could be validated in wider range of crops and management regimes by using data from international flux networks (ICOS and FLUXNET).

Conclusions & Perspectives

L'objectif général de cette thèse est d'établir une approche pour estimer la production et les principales composantes des cycles biogéochimiques du carbone et de l'eau des cultures pour des années climatiques contrastées, de l'échelle locales à régionales, avec un focus particulier sur la culture de blé. Pour ce faire, nous proposons une approche de modélisation combinant données à hautes résolutions spatiales et temporelles couplées à un modèle agronomique simple et une module de bilan hydrique, et nous nous sommes appuyés sur un jeu de données in-situ extensif pour la calibration et la validation de cette approche.

La première partie de cette thèse est dédiée à l'estimation de l'indice vert de la végétation (GAI) à partir de séries temporelles d'images satellitales provenant de Formosat-2 et SPOT. Pour cela deux méthodes ont été testées. La première méthode, empirique, consistait à établir une relation exponentielle entre l'indice normalisé de la végétation (NDVI) et des mesures in-situ de GAI (GAI_{eff}) obtenues à partir du traitement de photographies hémisphériques de la végétation analysées avec le logiciel CanEye. La seconde méthode, physique, était basée sur l'inversion du modèle de transfert radiatif PROSAIL (Baret et al., 1992a) en utilisant des réseaux de neurones. Dans cette approche, les données in-situ ne sont pas nécessaires pour calibrer le modèle. Pour valider cette méthode, nous avons aussi utilisé des données in-situ de GAI_{eff} obtenues pour les principales cultures du sud ouest de la France : blé, maïs, tournesol et soja. Les deux méthodes employées avaient des performances similaires pour l'estimation du GAI, avec des RRMSE d'environ 29% et des coefficients de corrélation (R^2) de 0.86. La relation empirique NDVI-GAI_{eff} est connue pour saturer à de fortes valeurs de NDVI, mais puisque les estimations de GAI avec cette méthode étaient comparées avec des données in-situ de GAI calculées à partir de photographies hémisphériques utilisant une loi de Poisson (qui sature aussi), le phénomène de saturation n'impactait pas les performances de la méthode empirique. Pour l'approche physiques, qui est sensée être moins sensible à ces effets de saturation, les estimations de GAI étaient mieux corrélées aux vraies valeurs de GAI obtenues à partir de mesures destructives que celles obtenues avec la méthode empirique. Cependant, les estimations de GAI obtenues avec la méthode physique sont globalement plus faibles que les valeurs de GAI obtenues avec les méthodes destructives à cause du phénomène d'agrégation foliaire.

Puisque la méthode physique ne nécessite pas de mesures terrain pour l'inversion du GAI et qu'elle est moins sensible aux phénomènes de saturation, elle offre de meilleures perspectives pour des études aux échelles régionales et globales. Nous avons donc choisi cette approche pour produire des cartes dynamiques de GAI qui ont ensuite été utilisées pour l'approche de modélisation spatialisée.

Les travaux de modélisation ont débutés avec l'utilisation du modèle original SAFY (Duchemin et al., 2008) qui permet de simuler l'évolution du GAI, de la biomasse et le rendement pour les cultures. Les paramètres phénologiques (jour de levée, début de sénescence, partition foliaire,…) de SAFY ont été calibrés à partir de séries temporelles de GAI obtenues à partir des données de télédétection. Concernant l'estimation des rendements, quatre fonctions ont été évaluées (deux basées sur des indices de récolte et deux issues de modules de partition provenant du modèle de culture STICS) afin de choisir la plus performante pour l'estimation des rendements. Il est important de préciser qu'une large campagne expérimentale avait été menée en 2011 sur le blé dans le cadre de cette thèse, et qu'elle a permis d'obtenir des jeux de données de validation pour la biomasse et les rendements sur un ensemble de parcelles hétérogènes juste avant que la récolte ait lieu. Ce jeu de données a été essentiel pour la validation des rendements estimés à partir du modèle puisque les seules données disponible alors étaient des rendements moyens obtenus à partir d'enquêtes auprès des agriculteurs qui présentent un fort degré d'incertitude et couvrent une faible gamme de rendement (pour des raisons d'agrégation) ce qui les rend difficilement utilisables pour la validation du modèle. Par la suite, le modèle SAFY a pu être testé à l'aide d'un troisième jeu de données correspondant à des cartes de rendement obtenues à partir de moissonneuses batteuses équipées de capteurs de rendement. Ces différents jeux de données de validation nous ont permis de réaliser que les données de rendement terrain utilisées pour la validation du modèle doivent couvrir une large gamme de valeurs. Les cartes de rendements à haute résolution issues des moissonneuses batteuses offrent donc d'intéressantes perspectives pour la validation du modèle, notamment à l'échelle pixellaire, car elles permettent de caractériser les fortes hétérogénéités intra-parcellaires. De plus, il nous apparait essentiel à l'avenir de collecter des données de rendement pour une large gamme de conditions spatiales et environnementales pour la validation de ce type d'approche de modélisation, en particulier pour estimer des bilans de carbone à l'échelle parcellaire puisque la composante rendement est un des principaux termes qui entrent en compte dans son calcul.

Finalement, nos résultats montrent que le modèle SAFY est capable de reproduire correctement les dynamiques de GAI (RRMSE=12%), de biomasse (RRMSE d'environ 25%) et le rendement (RRMSE=24% pour la meilleure méthode d'estimation basée sur des indices de récolte) pour un ensemble de parcelles soumises à des gestions et des années climatiques contrastées. Les estimations de rendements obtenues avec les modules d'allocation issus de STICS pourraient cependant être améliorées en ajustant les paramètres de seuil de température pour le démarrage de la phase de remplissage du grain et l'évolution de l'indice de rendement. Ce travail devra être poursuivi ultérieurement.

Le modèle SAFY-CO_2 a ensuite été développé afin de pouvoir estimer les composantes du flux net de carbone : la photosynthèse (GPP), les respirations autotrophe (Ra), hétérotrophe (Rh), de l'écosystème (Reco), la productivité nette de l'écosystème (NPP) et le flux net de l'écosystème (NEE), en plus du GAI, de la biomasse et du rendement déjà simulés par le modèle original SAFY. Cependant, dans SAFY, le paramètre d'efficience d'utilisation de la lumière basé sur la théorie de Monteith était utilisé pour convertir le rayonnement photosynthétiquement actif absorbé par la plante directement en biomasse. Dans SAFY-CO_2 cette fonction a été modifiée pour estimer tout d'abord la photosynthèse et une des avancées notable dans cette approche a été l'ajout d'une fonction qui prend en compte l'effet de la fraction diffuse du rayonnement sur la photosynthèse à l'échelle du couvert qui permet d'ajuster l'efficience d'utilisation de la lumière en fonction des conditions d'éclairement. Les autres composantes du flux net de carbone (GPP, Ra, Rh…) étaient ensuite modélisées. Tout d'abord, Ra était calculé en estimant les respirations d'entretien et de croissance en utilisant des équations standard de la littérature, puis Ra était soustrait à la GPP de manière à calculer la NPP. Ensuite, une simple équation de type Q_{10} paramétrée à partir des données expérimentales de nos sites d'études a été utilisée pour calculer Rh et ce terme était ajouté à la NPP pour calculer la NEE. Finalement, le bilan carbone des parcelles cultivées (NECB) a été calculé et comparé aux données pour nos parcelles expérimentales d'Auradé et de Lamasquère, en utilisant la NEE simulée, la quantité de carbone apportée sur la parcelle sous forme d'engrais organique et de graine (OF, données fournies par les agriculteurs) et la quantité de carbone exporté au moment de la récolte (Cexp) simulée par le modèle.

Certaines hypothèses on cependant du être faites pour les autres parcelles de notre zone d'étude, puisque certaines des informations requises pour le calcul du bilan de carbone n'étaient pas disponibles. La première concerne la gestion des pailles. A l'heure actuelle, cette information ne peut pas être obtenue à partir des images de télédétection et il est donc impossible de différencier les parcelles cultivées sur lesquelles la paille est exportée de celles sur lesquelles la paille est

enfouie. En conséquence, les estimations d'export de carbone présentent une large incertitude (qui se répercute sur les bilans de C). Des travaux s'appuyant sur une utilisation combinée de données multispectrales (optique, données hyperspectrales et radar) devront donc être menés à l'avenir pour tenter de fournir une information à la parcelle concernant la gestion des pailles. Heureusement, sur notre zone d'étude, les pailles sont enfouies sur la plupart des parcelles (excepté en 2011) puisque la plupart des fermes ont une vocation de production céréalière. La seconde hypothèse qui a été faite concerne les apports d'engrais organique. Puisque il n'est pas possible d'obtenir une estimation de cette variable à partir de données de télédétection et comme très peu de fermes dans notre zone d'étude sont dédiées à de la production animale, nous avons considéré que les apports d'engrais organiques étaient négligeables et le terme OF correspond uniquement à l'apport de carbone sous forme de graines. Pour des zones géographiques où la production animale est significative, la seule solution pour estimer les apports de carbone organique à la parcelle réside dans l'utilisation de données d'inventaires régionaux/nationaux pour estimer des apports moyens d'engrais organiques.

Les composantes du flux net de CO_2 estimées avec SAFY-CO_2 étaient globalement en accord avec les mesures, présentant de très bonnes corrélations (R^2 d'environ 0.9 pour la GPP, 0.75 pour Reco et 0.84 pour la NEE). Cependant, nos analyses ont indiquées que les performances du modèle pourraient être améliorées en considérant les phénomènes de repousses et de développement d'adventices après la récolte ou en estimant plus précisément le GAI par exemple à partir des données de télédétection radar pour les périodes de forts développement de la végétation. De plus, une meilleure estimation de la date de levée permettrait d'obtenir de meilleures estimations des flux de CO_2 durant les périodes de faible développement de la végétation. Les bilans de C qui ont été calculés pour différentes conditions environnementales et climatiques nous ont permis d'identifier les principaux facteurs qui les influencent et d'identifier de possible leviers d'action pour améliorer le stockage de carbone dans les sols agricoles et donc réduire les impacts de la production agricole sur le changement climatique. Toutefois, d'assez forts écarts entre les NECB simulés et observés ont été trouvés, allant de 48 gC $m^{-2}an^{-1}$ jusqu'à 108 gC $m^{-2}an^{-1}$. Ces résultats mettent en évidence l'importance d'estimer précisément chacun des termes qui composent le bilan de carbone, en particulier les deux principaux termes (NEE et Cexp) qui se compensent largement (car ils sont de signes opposés).

La dernière étape de modélisation consistait à coupler SAFY-CO_2 avec un module de bilan hydrique basé sur la méthode FAO-56. Le modèle qui en résulte, appelé SAFYE-CO_2 permettait de simuler en plus des variables citées précédemment

l'évapotranspiration (ETR) et le contenu en eau du sol (SWC). Ce couplage nous a permis de prendre en compte l'effet d'éventuels stress hydriques sur la photosynthèse, d'estimer la consommation en eau des cultures et aussi d'estimer plus précisément la respiration hétérotrophe en prenant en compte l'évolution du contenu en eau du sol. Ce couplage nous a par ailleurs permis d'avoir une approche couplée des cycles de l'eau et du carbone et d'estimer des indicateurs agronomiques et environnementaux de l'efficience de l'utilisation de l'eau (WUE).

Le modèle SAFYE-CO_2 nous a permis d'obtenir de bons résultats en terme d'évapotranspiration simulée (RMSE d'environ 0.46 $mm.j^{-1}$ et R^2 d'environ 0.74). D'un point de vue consommation annuelle en eau des cultures de blé, le modèle et les observations présentaient des écarts absolus compris entre 20 $mm.an^{-1}$ et 77 $mm.an^{-1}$, ce qui est satisfaisant. Les performances de la version SAFYE-CO_2 étaient un peu dégradées concernant les simulations de biomasse et composantes des flux nets de CO_2 par rapport aux modèles SAFY et SAFY-CO_2. Cependant, cette apparente « dégradation » met surtout en évidence les sous-estimations de GPP causées par la sous-estimation des GAI obtenues par télédétection par rapport aux données destructives pour les années à fort développement de la végétation. Une solution envisageable pour résoudre ce problème pourrait être d'assimiler des estimations de GAI obtenues à partir de données radar qui sont moins sensibles aux phénomènes de saturation du signal à fort GAI.

Finalement, à partir des rendements, bilans carbone et évapotranspirations annuelles calculées, des efficiences d'utilisation de l'eau agronomiques et environnementales ont été calculées. Elles pourront permettre l'indentification de stratégies de développement durable permettant de limiter les impacts de la culture de blé sur l'environnement tout en maintenant de bonnes capacités de production.

A partir du travail qui a été réalisé dans le cadre de cette thèse, nous avons pu vérifier le potentiel d'une approche simplifiée de modélisation agronomique combinée à des données de télédétection et un module de bilan hydrique pour l'estimation des principales composantes du cycle de l'eau et du carbone des cultures. Ce travail a été rendu possible grâce à la disponibilité d'un large jeu de données de télédétection et de données terrain (mesures in-situ et enquêtes auprès des agriculteurs) sur notre zone d'étude pour consolider et valider notre approche de modélisation. Maintenant que la démarche est bien établie, ses principaux avantages et limitations ont pu être identifiés et analysés. L'étape suivante consistera donc à l'appliquer à d'autres cultures sur notre zone d'étude et à l'améliorer en prenant notamment en compte l'effet des repousses spontanées, développement d'adventices et la mise en place de couverts intermédiaires sur les flux d'eau et de CO_2, puisque ces évènements peuvent avoir de forts impacts, notamment sur les bilans annuels de carbone des parcelles. L'utilisation de données

radar pourrait aussi permettre une meilleure estimation des GAI durant les phases de fort développement du couvert ou en cas de longue périodes de couverture nuageuses qui empêchent l'exploitation des données optiques comme en 2008. Cela améliorerait très certainement nos estimations des composantes des bilans d'eau et de carbone pour les cultures.

A l'avenir, cette approche pourra être étendue à une échelle globale grâce au lancement des satellites des prochaines missions spatiales (Sentinel-1 et 2,…) et elle pourrait être validée pour une plus grande gamme de cultures et de modalités de gestion en utilisant les données des réseaux internationaux de mesure de flux (ICOS, Fluxnet…).

References

Adiku, S. et al., 2006. PIXGRO: A model for simulating the ecosystem CO_2 exchange and growth of spring barley. Ecological Modelling, 190(3): 260-276.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome, 300: 6541.

Ammann, C., Meixner, F., Busch, J. and Lösch, R., 1996. CO_2 and H_2O gas exchange of a triticale field: II. Micrometeorological flux studies and comparison with upscaling from porometry. Physics and Chemistry of the Earth, 21(3): 151-155.

Amthor, J.S., 1989. Respiration and crop productivity. Springer-Verlag.

Amthor, J.S., 2000. The McCree–de Wit–Penning de Vries–Thornley respiration paradigms: 30 years later. Annals of Botany, 86(1): 1-20.

Anthoni, P.M., Freibauer, A., Kolle, O. and Schulze, E.-D., 2004. Winter wheat carbon exchange in Thuringia, Germany. Agricultural and Forest Meteorology, 121(1): 55-67.

Arora, V. and Gajri, P., 1998. Evaluation of a crop growth-water balance model for analysing wheat responses to climate-and water-limited environments. Field crops research, 59(3): 213-224.

Asrar, G., Fuchs, M., Kanemasu, E. and Hatfield, J., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. Agronomy journal, 76(2): 300-306.

Aubinet, M. et al., 1999. Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. Advances in ecological research, 30: 113-175.

Aubinet, M. et al., 2009. Carbon sequestration by a crop over a 4-year sugar beet/winter wheat/seed potato/winter wheat rotation cycle. Agricultural and Forest Meteorology, 149(3): 407-418.

Baldocchi, D., 2008. TURNER REVIEW No. 15.'Breathing'of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Australian Journal of Botany, 56(1): 1-26.

Baldocchi, D. and Collineau, S., 1994. The physical nature of solar radiation in heterogeneous canopies: spatial and temporal attributes. Exploitation of environmental heterogeneity by plants: 21-71.

Baldocchi, D. et al., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society, 82(11): 2415-2434.

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biology, 9(4): 479-492.

Baret, F. and Buis, S., 2008. Estimating canopy characteristics from remote sensing observations: review of methods and associated problems, Advances in land remote Sensing. Springer, pp. 173-201.

Baret, F., De Solan, B., Lopez-Lozano, R., Ma, K. and Weiss, M., 2010. GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5 zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. Agricultural and Forest Meteorology, 150(11): 1393-1401.

Baret, F. and Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote sensing of environment, 35(2): 161-173.

Baret, F., Guyot, G. and Major, D., 1989. Crop biomass evaluation using radiometric measurements. Photogrammetria, 43(5): 241-256.

Baret, F. et al., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. Remote sensing of environment, 110(3): 275-286.

Baret, F., Jacquemoud, S., Guyot, G. and Leprieur, C., 1992a. Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. Remote Sensing of Environment, 41(2): 133-142.

Baret, F., Olioso, A. and Luciani, J., 1992b. Root biomass fraction as a function of growth degree days in wheat. Plant and soil, 140(1): 137-144.

Bausch, W.C. and Neale, C.M.U., 1987. Crop coefficients derived from reflected canopy radiation: a concept. Transactions of the ASAE-American Society of Agricultural Engineers (USA).

Becker-Reshef, I., Vermote, E., Lindeman, M. and Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sensing of Environment, 114(6): 1312-1323.

Belaqziz, S. et al., 2013. A new irrigation priority index based on remote sensing data for assessing the networks irrigation scheduling. Agricultural Water Management, 119: 1-9.

Bernacchi, C.J., Hollinger, S.E. and Meyers, T., 2005. The conversion of the corn/soybean ecosystem to noâ€till agriculture may result in a carbon sink. Global Change Biology, 11(11): 1867-1872.

Bethenod, O., Katerji, N., Goujet, R., Bertolini, J.M. and Rana, G., 2000. Determination and validation of corn crop transpirationby sap flow measurement under field conditions. Theoretical and Applied Climatology, 67(3-4): 153-160.

Béziat, P., 2009. Effet des conditions environnementales et des pratiques culturales sur les flux de carbone et d'eau dans les agrosystèmes, Université Paul Sabatier, Toulouse III.

Béziat, P., Ceschia, E. and Dedieu, G., 2009. Carbon balance of a three crop succession over two cropland sites in South West France. Agricultural and Forest Meteorology, 149(10): 1628-1645.

Béziat, P. et al., 2013. Evaluation of a simple approach for crop evapotranspiration partitioning and analysis of the water budget distribution for several crop species. Agricultural and Forest Meteorology, 177: 46-56.

Bhattacharya, B.K., Mallick, K., Nigam, R., Dakore, K. and Shekh, A., 2011. Efficiency based wheat yield prediction in a semi-arid climate using surface energy budgeting with satellite observations. Agricultural and Forest Meteorology, 151(10): 1394-1408.

Bindi, M., Miglietta, F. and Zipoli, G., 1992. Different methods for separating diffuse and direct components of solar radiation and their application in crop growth models. Climate Research, 2: 47-54.

Bondeau, A. et al., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Global Change Biology, 13(3): 679-706.

Boulet, G., Chehbouni, A., Braud, I. and Vauclin, M., 1999. Mosaic versus dual source approaches for modelling the surface energy balance of a semi-arid land. Hydrology and Earth System Sciences Discussions, 3(2): 247-258.

Bréda, N.J., 2003. Ground‐based measurements of leaf area index: a review of methods, instruments and current controversies. Journal of experimental botany, 54(392): 2403-2417.

Brisson, N. et al., 2003. An overview of the crop model STICS. European Journal of agronomy, 18(3): 309-332.

Brisson, N. and Levrault, F., 2010. Changement climatique, agriculture et forêt en France: simulations d'impacts sur les principales espèces. Le Livre vert du projet Climator (2007-2010). op. cit.

Brisson, N. et al., 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(5-6): 311-346.

Bsaibes, A. et al., 2009. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. Remote sensing of environment, 113(4): 716-729.

Calvet, J.-C. et al., 1998. An interactive vegetation SVAT model tested against data from six contrasting sites. Agricultural and Forest Meteorology, 92(2): 73-95.

Campbell, J.B., 2002. Introduction to remote sensing. CRC Press.

CCRS, 2014. Fundamentals of Remote Sensing.

Ceschia, E., Damesin, C., Lebaube, S.p., Pontailler, J.-Y. and Dufrêne, E., 2002. Spatial and seasonal variations in stem respiration of beech trees (Fagus sylvatica). Annals of Forest Science, 59(8): 801-812.

Ceschia, E. et al., 2010. Management effects on net ecosystem carbon and GHG budgets at European crop sites. Agriculture, ecosystems & environment, 139(3): 363-383.

Chen, B. et al., 2012. Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis. Remote Sensing of Environment, 124: 742-755.

Chen, J. and Black, T., 1991. Measuring leaf area index of plant canopies with branch architecture. Agricultural and Forest Meteorology, 57(1): 1-12.

Chen, J.M., 1996. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. Agricultural and Forest Meteorology, 80(2): 135-163.

Chen, J.M. and Black, T., 1992. Defining leaf area index for non‐flat leaves. Plant, Cell & Environment, 15(4): 421-429.
Chen, J.M. and Cihlar, J., 1995. Plant canopy gap-size analysis theory for improving optical measurements of leaf-area index. Applied Optics, 34(27): 6211-6222.
Choudhury, B.J., 2000. A sensitivity analysis of the radiation use efficiency for gross photosynthesis and net carbon accumulation by wheat. Agricultural and Forest Meteorology, 101(2): 217-234.
Choudhury, B.J., 2001. Modeling radiation-and carbon-use efficiencies of maize, sorghum, and rice. Agricultural and Forest Meteorology, 106(4): 317-330.
Choudhury, B.J., Ahmed, N.U., Idso, S.B., Reginato, R.J. and Daughtry, C.S.T., 1994. Relations between evaporation coefficients and vegetation indices studied by model simulations. Remote sensing of environment, 50(1): 1-17.
Ciais, P., C. et al., 2013. Carbon and Other Biogeochemical Cycles. In: IPCC (Editor), Climate Change 2013: The Physical Science Basis. Cambridge University Press.
Claverie, M., 2012. Estimation spatialisée de la biomasse et des besoins en eau des cultures à l'aide de données satellitales à haute résolution spatiale et temporelle : application aux agrosystèmes du Sud-Ouest de la France.
Claverie, M. et al., 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.
Coleman, T.F. and Li, Y., 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on optimization, 6(2): 418-445.
Collares-Pereira, M. and Rabl, A., 1979. The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values. Solar Energy, 22(2): 155-164.
Combal, B., Baret, F. and Weiss, M., 2002. Improving canopy variables estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies. Agronomie-Sciences des Productions Vegetales et de l'Environnement, 22(2): 205-216.
Combal, B. et al., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. Remote sensing of environment, 84(1): 1-15.
Courault, D., Seguin, B. and Olioso, A., 2005. Review on estimation of evapotranspiration from remote sensing data: From empirical to numerical modeling approaches. Irrigation and Drainage systems, 19(3-4): 223-249.
Damesin, C., Ceschia, E., Le Goff, N., Ottorini, J.M. and DufrÃªne, E., 2002. Stem and branch respiration of beech: from tree measurements to estimations at the stand level. New Phytologist, 153(1): 159-172.
Daughtry, C.S.T. and Hunt Jr, E.R., 2008. Mitigating the effects of soil and residue water contents on remotely sensed estimates of crop residue cover. Remote sensing of environment, 112(4): 1647-1657.
De Jong, J., 1980. Een karakterisering van de zonnestraling in Nederland. Doctoraalverslag Vakgroep Fysische Aspecten van de Gebouwde Omgeving afd. Bouwkunde en Vakgroep Warmte-en Stromingstechnieken afd. Werktuigbouwkunde, Technische Hogeschool (Techn. Univ.), Eindhoven, Netherlands, 97: 67.

de Noblet-Ducoudré, N. et al., 2004. Coupling the soil-vegetation-atmosphere-transfer scheme ORCHIDEE to the agronomy model STICS to study the influence of croplands on the European carbon and water budgets. Agronomie, 24(6-7): 397-407.

De Wit, C.T. and Goudriaan, J., 1978. Simulation of ecological processes. Centre for Agricultural Publishing and Documentation.

Debaeke, P., Puech, J., Casals, M. and Petibon, P., 1996. Élaboration du rendement du blé d'hiver en conditions de déficit hydrique. I. Étude en lysimètres. Agronomie, 16(1): 3-23.

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

Delogu, E., 2013. Modélisation des flux de carbone dans les agro-systèmes, Université Paul Sabatier, Toulouse III.

Demarez, V., Duthoit, S., Baret, F., Weiss, M. and Dedieu, G., 2008. Estimation of leaf area and clumping indexes of crops with hemispherical photographs. Agricultural and Forest Meteorology, 148(4): 644-655.

Demuth, H., Beale, M. and Hagan, M., 1998. Neural Network Toolbox User's Guide, The MathWorks. Inc., Natick, MA.

Devonec, E. and Barros, A.P., 2002. Exploring the transferability of a land-surface hydrology model. Journal of Hydrology, 265(1): 258-282.

Droogers, P. and Bastiaanssen, W., 2002. Irrigation performance using hydrological and remote sensing modeling. Journal of Irrigation and Drainage Engineering, 128(1): 11-18.

Duchemin, B., Berthelot, B., Dedieu, G., Leroy, M. and Maisongrande, P., 2002. Normalisation of directional effects in 10-day global syntheses derived from VEGETATION/SPOT:: II. Validation of an operational method on actual data sets. Remote Sensing of Environment, 81(1): 101-113.

Duchemin, B. et al., 2005. Un modèle simplifié pour l'estimation du bilan hydrique et du rendement de cultures céréalières en milieu semi-aride. Deuxième Congrès Méditerranéen "Ressources en Eau dans le Bassin Méditérranéen": 14-17.

Duchemin, B., Maisongrande, P., Boulet, G. and Benhadj, I., 2008. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index. Environmental Modelling & Software, 23(7): 876-892.

Duchemin, B.t. et al., 2006. Monitoring wheat phenology and irrigation in Central Morocco: On the use of relationships between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices. Agricultural Water Management, 79(1): 1-27.

Dufrêne, E., Pontailler, J.-Y. and Saugier, B., 1993. A branch bag technique for simultaneous CO2 enrichment and assimilation measurements on beech (Fagus sylvatica L.). Plant, Cell & Environment, 16(9): 1131-1138.

Durand, Y. et al., 1993. A meteorological estimation of relevant parameters for snow models. Annals of Glaciology, 18: 65-71.

Duthoit, S., 2006. Prise en compte de l'agrégation des cultures dans la simulation du transfert radiatif : importance pour l'estimation de l'indice foliaire (LAI), de la parcelle au paysage, Université Paul Sabatier, Toulouse III.

Duveiller, G., 2011. Crop specific green area index retrieval from multi-scale remote sensing for agricultural monitoring, PhD thesis: Université catholique de Louvain (Belgique).

Duveiller, G., Weiss, M., Baret, F. and Defourny, P., 2011. Retrieving wheat Green Area Index during the growing season from optical time series measurements based on neural network radiative transfer inversion. Remote Sensing of Environment, 115(3): 887-896.

Epron, D., Le Dantec, V., Dufrene, E. and Granier, A., 2001. Seasonal dynamics of soil carbon dioxide efflux and simulated rhizosphere respiration in a beech forest. Tree Physiology, 21(2-3): 145-152.

Eugster, W. et al., 2010. Management effects on European cropland respiration. Agriculture, ecosystems & environment, 139(3): 346-362.

Fischer, W., Badgeley, P., Orr, D., Zissis, G. and al., e., 1975. History of remote sensing. Manual of Remote Sensing, I. American Society of Photogrammetry, 27-50 pp.

Frazer, G.W., Canham, C. and Lertzman, K., 1999. Gap Light Analyzer (GLA), Version 2.0: Imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, users manual and program documentation. Simon Fraser University, Burnaby, British Columbia, and the Institute of Ecosystem Studies, Millbrook, New York, 36.

Gabrielle, B., Denoroy, P., Gosse, G., Justes, E. and Andersen, M.N., 1998. Development and evaluation of a CERES-type model for winter oilseed rape. Field Crops Research, 57(1): 95-111.

Gabrielle, B. et al., 2006. Process-based modeling of nitrous oxide emissions from wheatâ€cropped soils at the subregional scale. Global biogeochemical cycles, 20(4).

Garten, C.T. and Wullschleger, S.D., 1999. Soil carbon inventories under a bioenergy crop (switchgrass): Measurement limitations. Journal of Environmental Quality, 28(4): 1359-1365.

Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G. and Duchemin, B., 2007. Analysis of evaporative fraction diurnal behaviour. Agricultural and Forest Meteorology, 143(1): 13-29.

Gervois, S.b. et al., 2008. Carbon and water balance of European croplands throughout the 20th century. Global Biogeochemical Cycles, 22(2).

Gitelson, A.A. et al., 2012. Remote estimation of crop gross primary production with Landsat data. Remote Sensing of Environment, 121: 404-414.

Goudriaan, J., 1977. Crop micrometeorology: a simulation study. Pudoc, Center for Agricultural Publishing and Documentation.

Granier, A., Biron, P., Bréda, N., Pontailler, J.-Y. and Saugier, B., 1996. Transpiration of trees and forest stands: short and long-term monitoring using sapflow methods. Global Change Biology, 2(3): 265-274.

Grant, R.F. et al., 2007. Net biome productivity of irrigated and rainfed maize-soybean rotations: modeling vs. measurements. Agronomy journal, 99(6): 1404-1423.

Hagolle, O. et al., 2008. Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images. Remote Sensing of Environment, 112(4): 1689-1701.

Hoffer, R.M., 1978. Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data. Remote sensing: The quantitative approach, 5.

Hollinger, D. et al., 1998. Forest–atmosphere carbon dioxide exchange in eastern Siberia. Agricultural and Forest Meteorology, 90(4): 291-306.

Hollinger, S.E., Bernacchi, C.J. and Meyers, T.P., 2005. Carbon budget of mature no-till ecosystem in North Central Region of the United States. Agricultural and Forest Meteorology, 130(1): 59-69.

Horst, T.W., 1999. The footprint for estimation of atmosphere-surface exchange fluxes by profile techniques. Boundary-Layer Meteorology, 90(2): 171-188.

Horst, T.W. and Weil, J.C., 1994. How far is far enough?: The fetch requirements for micrometeorological measurement of surface fluxes. Journal of Atmospheric and Oceanic Technology, 11(4): 1018-1025.

Hotsonyame, G. and Hunt, L., 1998. Seeding date, photoperiod and nitrogen effects on specific leaf area of field-grown wheat. Canadian journal of plant science, 78(1): 51-61.

Hoyaux, J., Moureaux, C., Tourneur, D., Bodson, B. and Aubinet, M., 2008. Extrapolating gross primary productivity from leaf to canopy scale in a winter wheat crop. Agricultural and forest meteorology, 148(4): 668-679.

Hu, Z. et al., 2009. Partitioning of evapotranspiration and its controls in four grassland ecosystems: Application of a two-source model. Agricultural and Forest Meteorology, 149(9): 1410-1420.

Huang, Y. et al., 2009. Agro-C: A biogeophysical model for simulating the carbon budget of agroecosystems. Agricultural and Forest Meteorology, 149(1): 106-129.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote sensing of environment, 25(3): 295-309.

IPCC, 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. (Working Group II Contribution to the IPCC 5th Assessment Report).

Iqbal, M.M. and Arif, M., 2010. Climate-change aspersions on food security of Pakistan. A Journal of Science for Development: 15.

Jacquemoud, S. and Baret, F., 1993. Estimating vegetation biophysical parameters by inversion of a reflectance model on high spectral resolution data.

Jacquemoud, S. et al., 2009. PROSPECT + SAIL models: A review of use for vegetation characterization. Remote Sensing of Environment, 113, Supplement 1(0): S56-S66.

Jarosz, N. et al., 2008. Carbon dioxide and energy flux partitioning between the understorey and the overstorey of a maritime pine forest during a year with reduced soil water availability. Agricultural and forest meteorology, 148(10): 1508-1523.

Jégo, G., Pattey, E. and Liu, J., 2012. Using Leaf Area Index, retrieved from optical imagery, in the STICS crop model for predicting yield and biomass of field crops. Field Crops Research, 131: 63-74.

Johnson, I. and Thornley, J., 1983. Vegetative crop growth model incorporating leaf area expansion and senescence, and applied to grass. Plant, Cell & Environment, 6(9): 721-729.

Jonckheere, I. et al., 2004. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. Agricultural and forest meteorology, 121(1): 19-35.

Katerji, N., Mastrorilli, M. and Rana, G., 2008. Water use efficiency of crops cultivated in the Mediterranean region: review and analysis. European journal of agronomy, 28(4): 493-507.

Kite, G.W. and Droogers, P., 2000. Comparing evapotranspiration estimates from satellites, hydrological models and field data. Journal of Hydrology, 229(1): 3-18.

Kleemann, S. and Gill, G., 2008. Variation in the response of canola cultivars to changes in row spacing, Proceeding of the 15th Australian Society of Agronomy Conference. Australian Society of Agronomy, Lincoln, New Zealand.

Krinner, G. et al., 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Global Biogeochemical Cycles, 19(1).

Kutsch, W.L. et al., 2010. The net biome production of full crop rotations in Europe. Agriculture, ecosystems & environment, 139(3): 336-345.

Kuusk, A., 1991. The inversion of the Nilson-Kuusk canopy reflectance model, a test case, Geoscience and Remote Sensing Symposium, 1991. IGARSS'91. Remote Sensing: Global Monitoring for Earth Management., International. IEEE, pp. 1547-1550.

Labus, M., Nielsen, G., Lawrence, R., Engel, R. and Long, D., 2002. Wheat yield estimates using multi-temporal NDVI satellite imagery. International Journal of Remote Sensing, 23(20): 4169-4180.

LaFrance, B., Lenot, X., Ruffel, C., Cao, P. and Rabaute, T., 2012. OUTILS DE PRETRAITEMENTS DES IMAGES OPTIQUES KALIDEOS. Revue française de photogrammétrie et de télédétection(197): 10-16.

Lagarias, J.C., Reeds, J.A., Wright, M.H. and Wright, P.E., 1998. Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. SIAM Journal on Optimization, 9(1): 112-147.

Lamaud, E., Brunet, Y. and Berbigier, P., 1996. Radiation and water use efficiencies of two coniferous forest canopies. Physics and Chemistry of the Earth, 21(5): 361-365.

Le Quéré, C. et al., 2013. The global carbon budget 1959-2011. Earth System Science Data, 5(1): 165-185.

Lehuger, S. et al., 2010. Predicting the net carbon exchanges of crop rotations in Europe with an agro-ecosystem model. Agriculture, ecosystems & environment, 139(3): 384-395.

Lehuger, S. et al., 2009. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agriculture, ecosystems & environment, 133(3): 208-222.

Leshno, M., Lin, V.Y., Pinkus, A. and Schocken, S., 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural networks, 6(6): 861-867.

Li, C., Frolking, S. and Frolking, T.A., 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. Journal of Geophysical Research: Atmospheres (1984â€“2012), 97(D9): 9759-9776.

Li, C. et al., 2005. Modeling impacts of farming management alternatives on CO2, CH4, and N2O emissions: A case study for water management of rice agriculture of China. Global Biogeochemical Cycles, 19(3).

Li, H. et al., 2011. A comparison of harvest index estimation methods of winter wheat based on field measurements of biophysical and spectral data. Biosystems Engineering, 109(4): 396-403.

Li, Q. et al., 2008. Effects of irrigation and straw mulching on microclimate characteristics and water use efficiency of winter wheat in North China. PLANT PRODUCTION SCIENCE-TOKYO-, 11(2): 161.

Liang, S., 2005. Quantitative remote sensing of land surfaces, 30. John Wiley & Sons.

Lobell, D.B. et al., 2013. The critical role of extreme heat for maize production in the United States. Nature Climate Change.

Maas, S.J., 1988a. Use of remotely-sensed information in agricultural crop growth models. Ecological modelling, 41(3): 247-268.

Maas, S.J., 1988b. Using satellite data to improve model estimates of crop yield. Agronomy Journal, 80(4): 655-662.

Maas, S.J., 1991. Use of remotely sensed information in plant growth simulation models. Adv. Agron, 1: 17-26.

Maas, S.J., 1992. GRAMI: A crop growth model that can use remotely sensed information. ARS-US Department of Agriculture, Agricultural Research Service (USA).

Maas, S.J., 1993. Parameterized model of gramineous crop growth: I. Leaf area and dry mass simulation. Agronomy Journal, 85(2): 348-353.

Maier, C.A., Zarnoch, S.J. and Dougherty, P.M., 1998. Effects of temperature and tissue nitrogen on dormant season stem and branch maintenance respiration in a young loblolly pine (Pinus taeda) plantation. Tree Physiology, 18(1): 11-20.

McCree, K., 1974. Equations for the rate of dark respiration of white clover and grain sorghum, as functions of dry weight, photosynthetic rate, and temperature. Crop science, 14(4): 509-514.

Medrano, H., Flexas, J. and Galmés, J., 2009. Variability in water use efficiency at the leaf level among Mediterranean plants with different growth forms. Plant and Soil, 317(1-2): 17-29.

Mkhabela, M., Bullock, P., Raj, S., Wang, S. and Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology, 151(3): 385-393.

Moncrieff, J.B. et al., 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. Journal of Hydrology, 188: 589-611.

Monteith, J., 1977. Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 281(980): 277-294.

Monteith, J.L., 1985. Evaporation from land surfaces: progress in analysis and prediction since 1948, National conference on advances in evapotranspiration, Hyatt Regency Chicago, Ill.(USA), 16-17 Dec 1985. American Society of Agricultural Engineers.

Monteith, J.L. and Moss, C.J., 1977. Climate and the efficiency of crop production in Britain [and discussion]. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 281(980): 277-294.

Moors, E.J. et al., 2010. Variability in carbon exchange of European croplands. Agriculture, ecosystems & environment, 139(3): 325-335.

Moré, J.J., 1978. The Levenberg-Marquardt algorithm: implementation and theory, Numerical analysis. Springer, pp. 105-116.

Moriondo, M., Giannakopoulos, C. and Bindi, M., 2011. Climate change impact assessment: the role of climate extremes in crop yield simulation. Climatic change, 104(3-4): 679-701.

Moulin, S., 1999. Impacts of model parameter uncertainties on crop reflectance estimates: a regional case study on wheat. International Journal of Remote Sensing, 20(1): 213-218.

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

Moureaux, C., 2008. Mesure des flux de CO2 et bilan carboné d'une rotation de quatre cultures.

Moureaux, C., Debacq, A., Bodson, B., Heinesch, B. and Aubinet, M., 2006. Annual net ecosystem carbon exchange by a sugar beet crop. Agricultural and Forest Meteorology, 139(1): 25-39.

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

Noilhan, J. and Mahfouf, J.F., 1996. The ISBA land surface parameterisation scheme. Global and Planetary Change, 13(1): 145-159.

Noilhan, J. and Planton, S., 1989. A simple parameterization of land surface processes for meteorological models. Monthly Weather Review, 117(3): 536-549.

O'Connell, M., O'Leary, G., Whitfield, D. and Connor, D., 2004. Interception of photosynthetically active radiation and radiation-use efficiency of wheat, field pea and mustard in a semi-arid environment. Field Crops Research, 85(2): 111-124.

Ogée, J. et al., 2003. Partitioning net ecosystem carbon exchange into net assimilation and respiration using $13CO_2$ measurements: A cost-effective sampling strategy. Global Biogeochemical Cycles, 17(2).

Olioso, A., Chauki, H., Courault, D. and Wigneron, J.-P., 1999. Estimation of evapotranspiration and photosynthesis by assimilation of remote sensing data into SVAT models. Remote Sensing of Environment, 68(3): 341-356.

Olioso, A. et al., 2005. Future directions for advanced evapotranspiration modeling: Assimilation of remote sensing data into crop simulation models and SVAT models. Irrigation and Drainage Systems, 19(3-4): 377-412.

Osborne, B., Saunders, M., Walmsley, D., Jones, M. and Smith, P., 2010. Key questions and uncertainties associated with the assessment of the cropland greenhouse gas balance. Agriculture, ecosystems & environment, 139(3): 293-301.

Parton, W.J., Schimel, D.S., Cole, C.V. and Ojima, D.S., 1987. Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Science Society of America Journal, 51(5): 1173-1179.

Passioura, J., 2006. Increasing crop productivity when water is scarce- from breeding to field management. Agricultural water management, 80(1): 176-196.

Pattey, E., Strachan, I.B., Desjardins, R.L. and Massheder, J., 2002. Measuring nighttime CO_2 flux over terrestrial ecosystems using eddy covariance and nocturnal boundary layer methods. Agricultural and Forest Meteorology, 113(1): 145-158.

Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193(1032): 120-145.

Porter, J.R. and Gawith, M., 1999. Temperatures and the growth and development of wheat: a review. European Journal of Agronomy, 10(1): 23-36.

Pury, D.d. and Farquhar, G., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big‐leaf models. Plant, Cell & Environment, 20(5): 537-557.

Qiu, G.Y. et al., 2008. Water use efficiency and evapotranspiration of winter wheat and its response to irrigation regime in the north China plain. Agricultural and forest meteorology, 148(11): 1848-1859.

Rayment, M.B. and Jarvis, P.G., 2000. Temporal and spatial variation of soil CO_2 efflux in a Canadian boreal forest. Soil Biology and Biochemistry, 32(1): 35-45.

Reichstein, M. et al., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biology, 11(9): 1424-1439.

Rembold, F., Atzberger, C., Savin, I. and Rojas, O., 2013. Using low resolution satellite imagery for yield prediction and yield anomaly detection. Remote Sensing, 5(4): 1704-1733.

Revill, A., Sus, O., Barrett, B. and Williams, M., 2013. Carbon cycling of European croplands: A framework for the assimilation of optical and microwave Earth observation data. Remote Sensing of Environment, 137: 84-93.

Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. Water resources research, 8(5): 1204-1213.

Ritchie, J.T. and Basso, B., 2008. Water use efficiency is not constant when crop water supply is adequate or fixed: The role of agronomic management. European journal of agronomy, 28(3): 273-281.

Roderick, M.L., 1999. Estimating the diffuse component from daily and monthly measurements of global radiation. Agricultural and Forest Meteorology, 95(3): 169-185.

Roderick, M.L., Farquhar, G.D., Berry, S.L. and Noble, I.R., 2001. On the direct effect of clouds and atmospheric particles on the productivity and structure of vegetation. Oecologia, 129(1): 21-30.

Rojas, O., 2007. Operational maize yield model development and validation based on remote sensing and agro‐meteorological data in Kenya. International Journal of Remote Sensing, 28(17): 3775-3793.

Roupsard, O. et al., 2006. Partitioning energy and evapo-transpiration above and below a tropical palm canopy. Agricultural and Forest Meteorology, 139(3): 252-268.

Rouse , J.J., Haas, R., Schell, J. and Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, 351: 309.

Rouse Jr, J., Haas, R., Schell, J. and Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, 351: 309.

Ruimy, A., Dedieu, G. and Saugier, B., 1996. TURC: A diagnostic model of continental gross primary productivity and net primary productivity. Global Biogeochemical Cycles, 10(2): 269-285.

Ruimy, A., Kergoat, L. and Bondeau, A., 1999. Comparing global models of terrestrial net primary productivity (NPP): Analysis of differences in light absorption and light‐use efficiency. Global Change Biology, 5(S1): 56-64.

Saito, M., Miyata, A., Nagai, H. and Yamada, T., 2005. Seasonal variation of carbon dioxide exchange in rice paddy field in Japan. Agricultural and Forest Meteorology, 135(1): 93-109.

Saxton, K., Rawls, W.J., Romberger, J. and Papendick, R., 1986. Estimating generalized soil-water characteristics from texture. Soil Science Society of America Journal, 50(4): 1031-1036.

Schulze, E.D. et al., 2009. Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance. Nature Geoscience, 2(12): 842-850.

Sellers, P.J. et al., 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science, 275(5299): 502-509.

Shuttleworth, W.J. and Wallace, J.S., 1985. Evaporation from sparse crop - an energy combination theory. Quarterly Journal of the Royal Meteorological Society, 111(469): 839-855.

Sinclair, T.R., Tanner, C.B. and Bennett, J.M., 1984. Water-use efficiency in crop production. BioScience, 34(1): 36-40.

Smith, P., 2004. How long before a change in soil organic carbon can be detected? Global Change Biology, 10(11): 1878-1883.

Smith, P. et al., 2005. Carbon sequestration potential in European croplands has been overestimated. Global Change Biology, 11(12): 2153-2163.

Smith, P. et al., 2010. Measurements necessary for assessing the net ecosystem carbon budget of croplands. Agriculture, ecosystems & environment, 139(3): 302-315.

Smith, P., Powlson, D.S., Smith, J.U., Falloon, P. and Coleman, K., 2000. Meeting Europe's climate change commitments: quantitative estimates of the potential for carbon mitigation by agriculture. Global Change Biology, 6(5): 525-539.

Soltner, D., 2006. Les bases de la production végétale : Tome 2 : Le climat : météorologie, conservation des sols, bioclimatologie, agronomie du carbone.

Soto, I.T., 2005. Mise en relation de la cartographie du rendement avec la distribution spatiale de l'état de surface du sol observée par télédétection, INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE.

Spitters, C., 1986. Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis Part II. Calculation of canopy photosynthesis. Agricultural and Forest meteorology, 38(1): 231-242.

Steduto, P. and Albrizio, R., 2005. Resource use efficiency of field-grown sunflower, sorghum, wheat and chickpea: II. Water use efficiency and comparison with radiation use efficiency. Agricultural and forest meteorology, 130(3): 269-281.

Steduto, P., Katerji, N., Puertos-Molina, H., Mastrorilli, M. and Rana, G., 1997. Water-use efficiency of sweet sorghum under water stress conditions Gas-exchange investigations at leaf and canopy scales. Field crops research, 54(2): 221-234.

Steiner, J.L. and Hatfield, J.L., 2008. Winds of change: A century of agroclimate research. Agronomy journal, 100(Supplement_3): S-132-S-152.

Steven, M.D., Malthus, T.J., Baret, F., Xu, H. and Chopping, M.J., 2003. Intercalibration of vegetation indices from different sensor systems. Remote Sensing of Environment, 88(4): 412-422.

Stockfors, J. and Linder, S., 1998. Effect of nitrogen on the seasonal course of growth and maintenance respiration in stems of Norway spruce trees. Tree Physiology, 18(3): 155-166.

Supit, I., Hooijer, A.A. and Van Diepen, C.A., 1994. System description of the Wofost 6.0 crop simulation model implemented in CGMS. Joint research centre; European commission.

Sus, O. et al., 2010. A linked carbon cycle and crop developmental model: Description and evaluation against measurements of carbon fluxes and carbon stocks at several European agricultural sites. Agriculture, ecosystems & environment, 139(3): 402-418.

Suyker, A.E. and Verma, S.B., 2009. Evapotranspiration of irrigated and rainfed maize-soybean cropping systems. Agricultural and forest meteorology, 149(3): 443-452.

Tallec, T., Béziat, P., Jarosz, N., Rivalland, V. and Ceschia, E., 2013. Crops' water use efficiencies in temperate climate: Comparison of stand, ecosystem and agronomical approaches. Agricultural and Forest Meteorology, 168: 69-81.

Tambussi, E.A., Bort, J. and Araus, J.L., 2007. Water use efficiency in C3 cereals under Mediterranean conditions: a review of physiological aspects. Annals of Applied Biology, 150(3): 307-321.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres (1984â€“2012), 106(D7): 7183-7192.

Tilman, D., Balzer, C., Hill, J. and Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences, 108(50): 20260-20264.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment, 8(2): 127-150.

Tucker, C.J. and Sellers, P.J., 1986. Satellite remote sensing of primary production. International journal of remote sensing, 7(11): 1395-1416.

Ugray, Z. et al., 2007. Scatter search and local NLP solvers: A multistart framework for global optimization. INFORMS Journal on Computing, 19(3): 328-340.

Unkovich, M., Baldock, J. and Forbes, M., 2010. Variability in harvest index of grain crops and potential significance for carbon accounting: examples from Australian agriculture. Advances in Agronomy, 105: 173-219.

Van Niel, T.G. and McVicar, T.R., 2004. Current and potential uses of optical remote sensing in rice-based irrigation systems: a review. Crop and Pasture Science, 55(2): 155-185.

Varlet-Grancher, C., Bonhomme, R., Chartier, M. and Artis, P., 1982. Efficience de la conversion de l'énergie solaire par un couvert végétal. Acta Oecologica. Oecologia plantarum.

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

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote sensing of environment, 16(2): 125-141.

Verhoef, W. and Bach, H., 2003. Remote sensing data assimilation using coupled radiative transfer models. Physics and Chemistry of the Earth, Parts A/B/C, 28(1): 3-13.

Vermote, E.F., Tanré, D., Deuze, J.L., Herman, M. and Morcette, J.-J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. Geoscience and Remote Sensing, IEEE Transactions on, 35(3): 675-686.

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

Vries, F.P.d., 1989. Simulation of ecophysiological processes of growth in several annual crops, 29. IRRI (free PDF download).

Wall, L., Larocque, D. and Léger, P.M., 2008. The early explanatory power of NDVI in crop yield modelling. International Journal of Remote Sensing, 29(8): 2211-2225.

Wallin, G.r. et al., 2001. Carbon dioxide exchange in Norway spruce at the shoot, tree and ecosystem scale. Tree Physiology, 21(12-13): 969-976.

Walter, J., 1994. CIMES. A package of programs for determining canopy geometry and transmission of PAR by hemispherical photographs. Université Louis Pasteur, Strasbourg.

Wang, J., Yu, Q. and Lee, X., 2007. Simulation of crop growth and energy and carbon dioxide fluxes at different time steps from hourly to daily. Hydrological processes, 21(18): 2474-2492.

Wattenbach, M. et al., 2010. The carbon balance of European croplands: a cross-site comparison of simulation models. Agriculture, ecosystems & environment, 139(3): 419-453.

Weir, A.H., Bragg, P.L., Porter, J.R. and Rayner, J.H., 1984. A winter wheat crop simulation model without water or nutrient limitations. The Journal of Agricultural Science, 102(02): 371-382.

Weiss, M. and Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. Remote sensing of environment, 70(3): 293-306.

Weiss, M. and Baret, F., 2010a. Venμs biophysical variable products algorithm theoretical basis document, EMMAH, INRA.

Weiss, M. and Baret, F., 2010b. Venμs Biophysical Variable Products Algorithm Theoretical Basis Document (ATBD), INRA-Emmah.

Welles, J.M. and Norman, J.M., 1991. Instrument for indirect measurement of canopy architecture. Agronomy journal, 83(5): 818-825.

West, T.O. et al., 2010. Cropland carbon fluxes in the United States: increasing geospatial resolution of inventory-based carbon accounting. Ecological Applications, 20(4): 1074-1086.

Williams, D.G. et al., 2004. Evapotranspiration components determined by stable isotope, sap flow and eddy covariance techniques. Agricultural and Forest Meteorology, 125(3): 241-258.

Williams, M. et al., 1996. Modelling the soil-plant-atmosphere continuum in a Quercus-Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties. Plant, Cell & Environment, 19(8): 911-927.

Xiao, Z., Wang, J. and Tong, T., 2008. Retrievals of canopy biophysical variables using multi-temporal remote sensing data.

Zeleke, K.T. and Wade, L.J., 2012. Evapotranspiration Estimation Using Soil Water Balance, Weather and Crop Data. Evapotranspiration–Remote sensing and Modeling. Ed. InTech: 41-57.

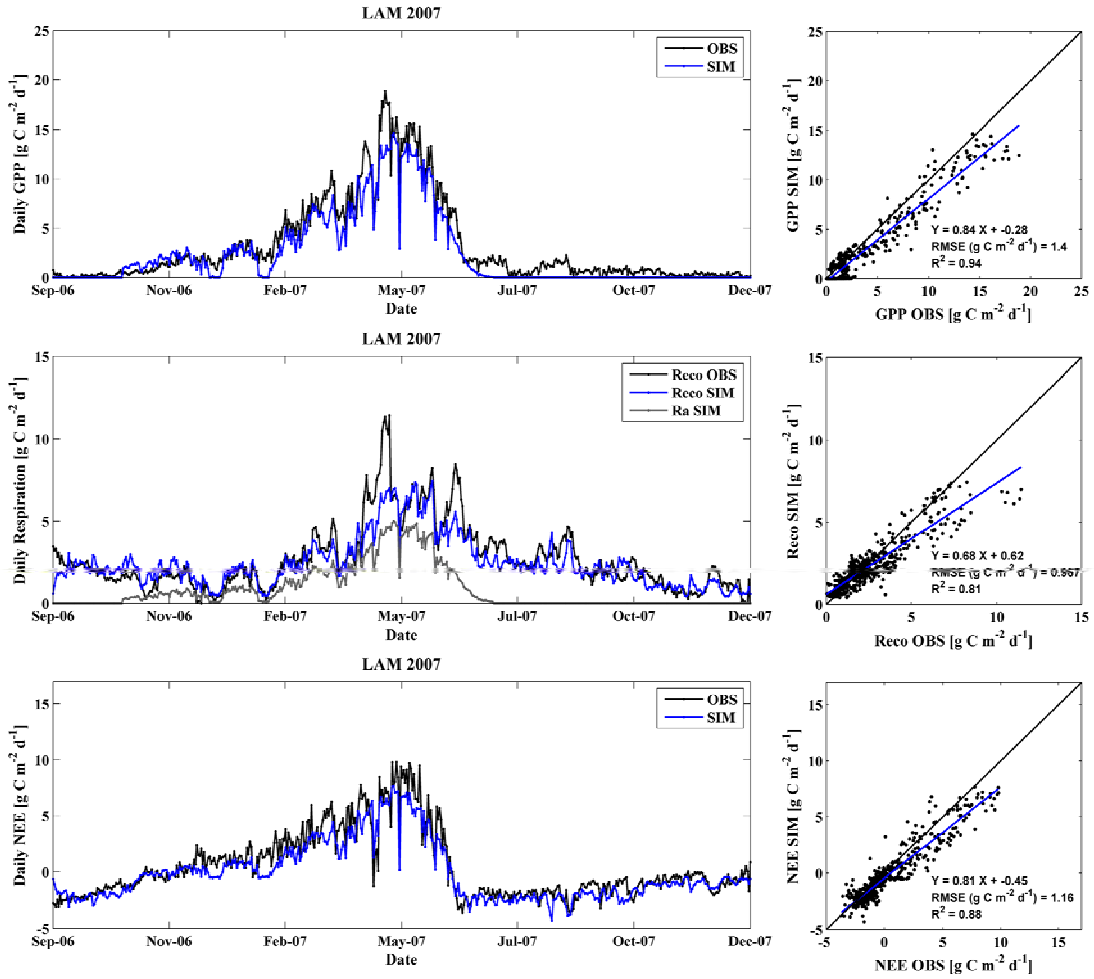
Zhang, Y., Li, C., Zhou, X. and Moore Iii, B., 2002. A simulation model linking crop growth and soil biogeochemistry for sustainable agriculture. Ecological Modelling, 151(1): 75-108.

List of Figures

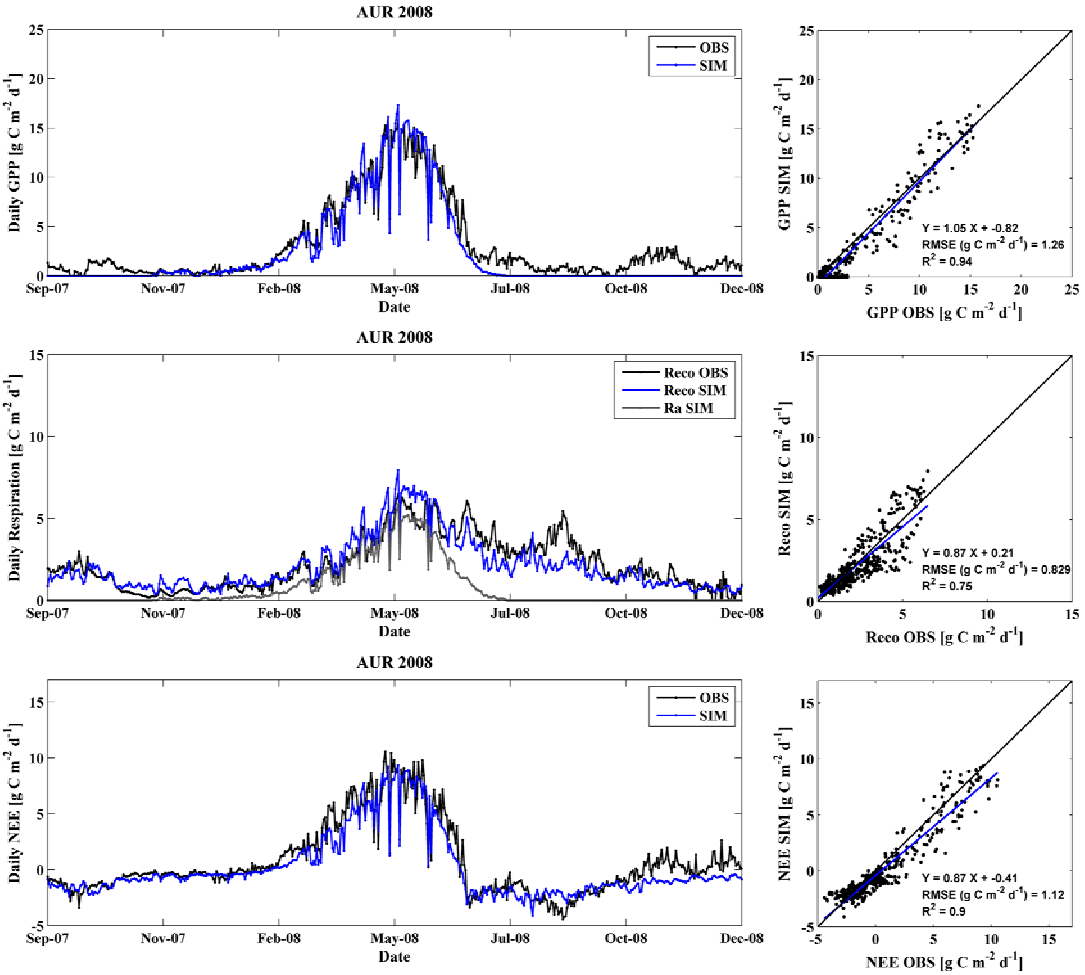
APPENDIX

APPENDIX A

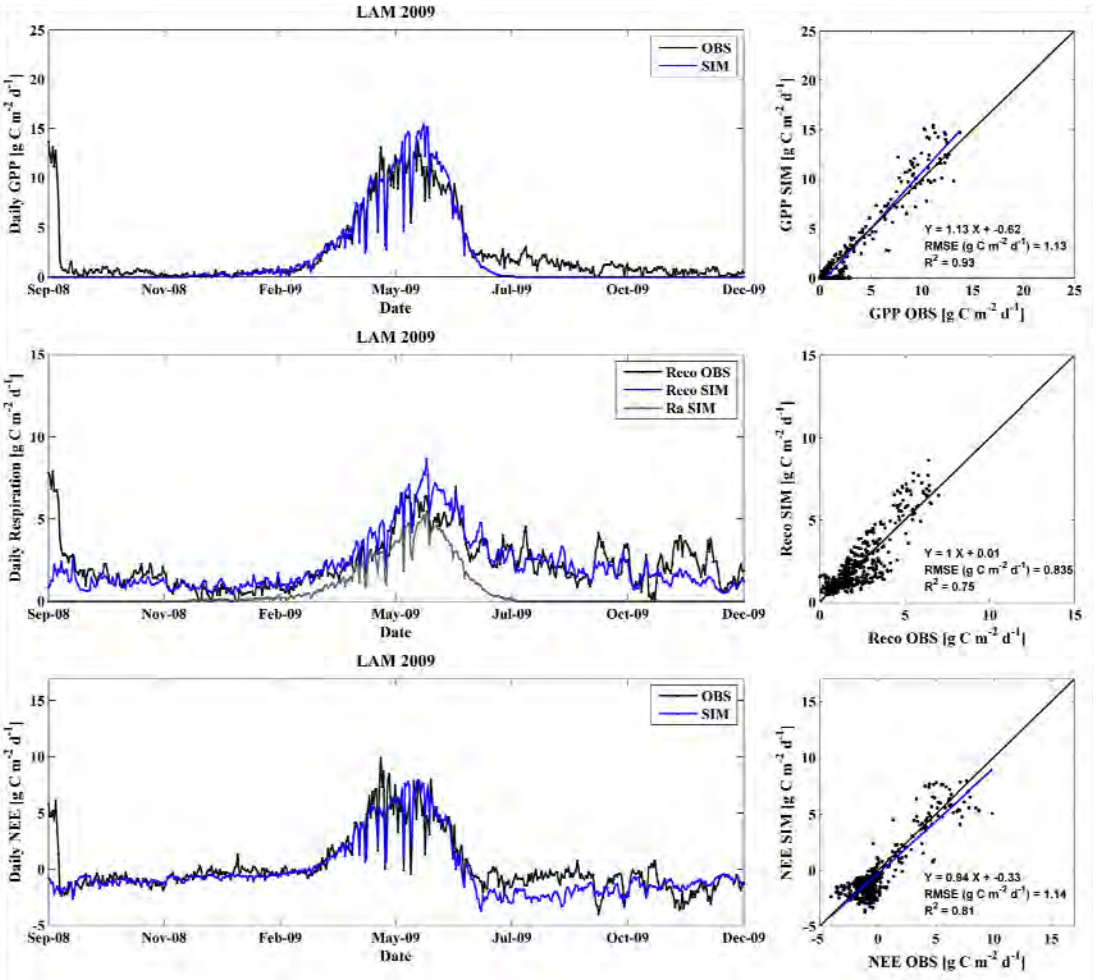
CO_2 flux estimates by the SAFYE-CO_2 model for the experimental sites: GPP, Reco and NEE dynamics



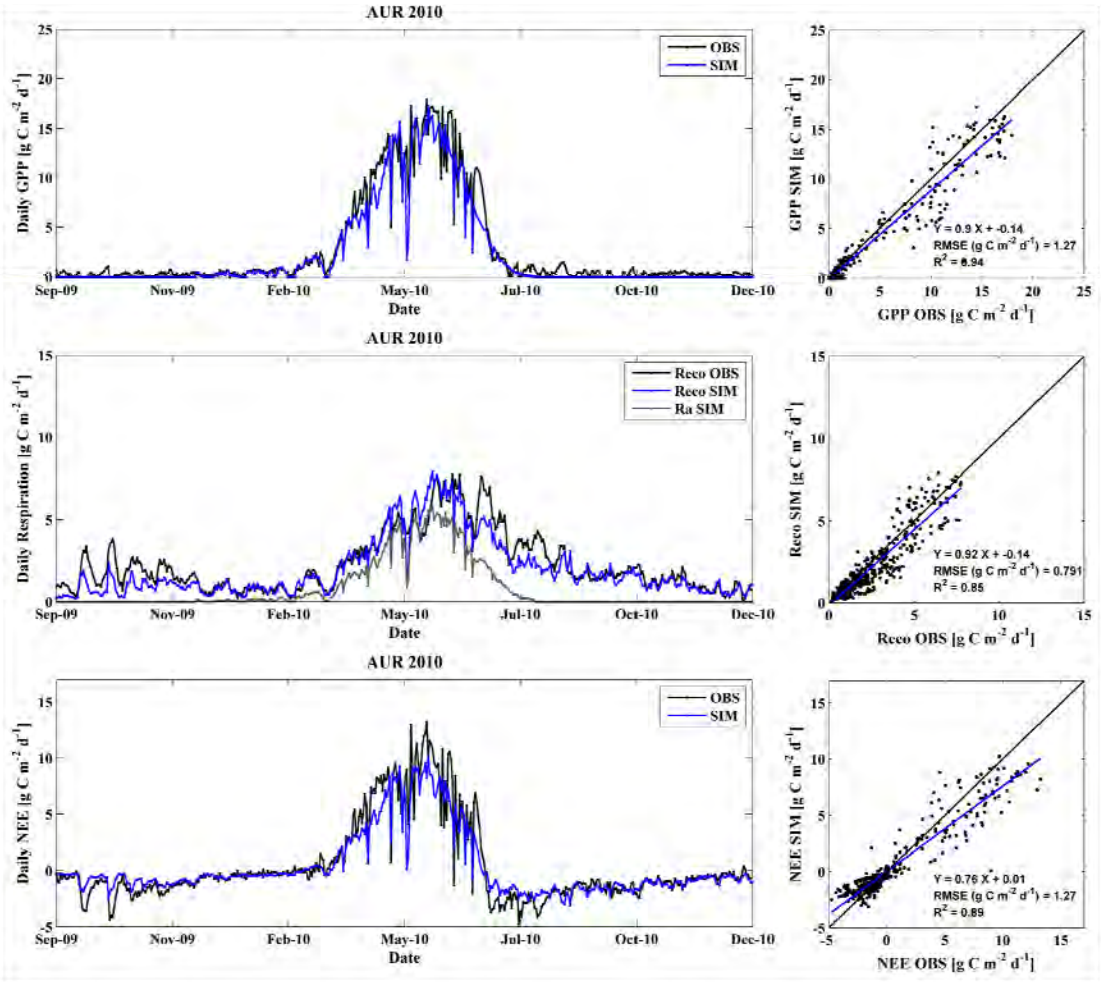
Evolution of the carbon fluxes simulated by SAFYE-CO_2 for the Lamasquère site, 2006-2007 winter wheat crop season. Observed fluxes are represented in black and simulated fluxes in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.



Evolution of the carbon fluxes simulated by SAFYE-CO_2 for the Auradé site, 2007-2008 winter wheat crop season. Observed fluxes are represented in black and simulated fluxes in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.

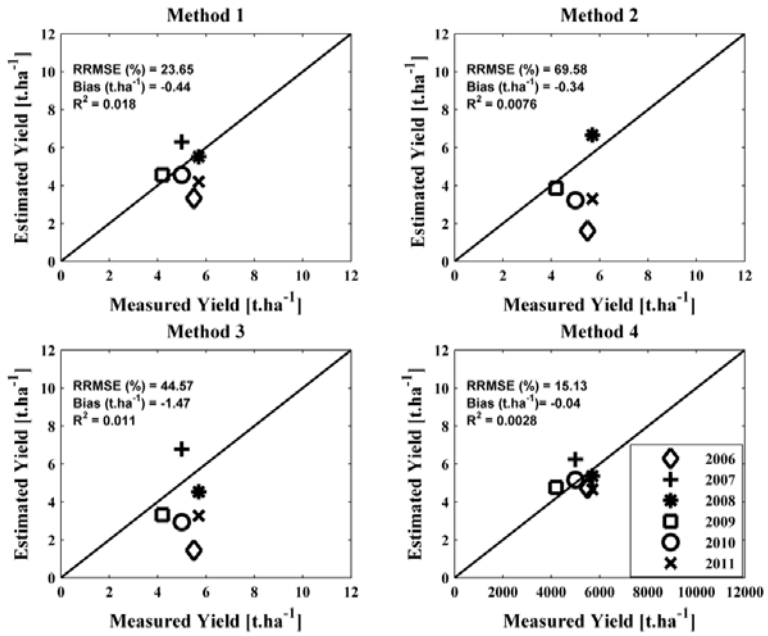


Evolution of the carbon fluxes simulated by SAFYE-CO_2 for the Lamasquère site, 2008-2009 winter wheat crop season. Observed fluxes are represented in black and simulated fluxes in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.



Evolution of the carbon fluxes simulated by SAFYE-CO_2 for the Auradé site, 2009-2010 winter wheat crop season. Observed fluxes are represented in black and simulated fluxes in blue. On the right, the scatter plots relating observed vs. simulated variables and associated statistics (regression, RMSE and R^2) are displayed. In the top: the GPP dynamics; in the middle: the ecosystem (Reco) and autotrophic (Ra) respirations; in the bottom: the NEE dynamics.

Yield estimates provided by SAFYE-CO_2 for the experimental sites.



Comparison of the methods for estimating yield with the SAFYE-CO_2 model. Scatter plots of the estimated yield (y-axis) against the measured yield (x-axis) for the experimental sites (AUR2006, AUR2008, AUR2010, LAM2007, LAM2009 and LAM2011).

APPENDIX B

Letter submitted to IEEE Geoscience and Remote Sensing Letters

Analysis of the Influence of GAI Intra-Field Spatial Variability on Sunflower Biomass Estimates Using SPOT and Formosat-2 Data

Amanda Veloso, Eric Ceschia, Marjorie Battude, Jean-François Dejoux, Martin Claverie, Valérie Demarez

***Abstract*— In this study, we used a Simple Algorithm For Yield estimate (SAFY) coupled with remotely sensed time series of Green Area Index (GAI) to obtain Dry Aboveground bioMass (DAM) estimations. The GAI time series were retrieved using the BV-NET tool and were validated against ground measurements acquired over sunflower fields. The GAI time series were used to calibrate the SAFY parameters. Regarding the biomass estimates, the model was run for five heterogeneous fields using a pixel-based and an aggregative-based approaches. The simulated biomasses were validated with ground data collected by means of two different sampling strategies: Elementary Sampling Unit (ESU) protocol and Transect protocol. The results allowed evaluating the effect of intra-field variability on biomass estimations. Similar DAM estimates were obtained when the model was run for pixel-based and aggregative-based approaches.**

***Index Terms*— Biomass, crop model, green area index, heterogeneity, remote sensing, sunflower**

I. Introduction

CROP modeling is a valuable tool for agricultural research and management. Numerous studies ([1],[2],[3]) revealed that the use of remote sensing data is mandatory for crop monitoring over large areas. In southwest France, spatial variability is observed between crops (mosaic of different crops) but also within crops (intra-field heterogeneities). As noted by [4], a surface property is characterized as heterogeneous if its measurements vary in space within the observed scene. The characterization of the spatial heterogeneity is therefore strongly dependent on the spatial resolution of the data. In France, croplands cover often small areas and show frequent temporal changes, demanding high spatial and temporal resolution data for an accurate crop monitoring. Nevertheless, in an era that emphasizes global scaled research, spatial data aggregation is widely used in environmental applications, for scaling-up from local to regional or global scales [5]. Thus, when aggregated data are used as input to crop models, the output estimations may be affected, due to the resolution differences.

This work aims at analyzing the potential of high temporal and spatial resolution satellite images to characterize the local (intra-field) spatial variability of the Green Area Index (GAI) variable and its effects on dry aboveground biomass (DAM) estimates. For this purpose, GAI times series were firstly retrieved from satellite images using the BV-NET tool [6]. The derived GAI products were then combined with an agrometereological crop model, SAFY (Simple Algorithm For Yield estimate) [7], for simulating GAI and DAM time series for different spatial extents (Elementary Sampling Unit (ESU), transect, or field). A *pixel-based* and an *aggregative-based* approaches were used to run SAFY simulations.

Manuscript received December 14, 2012.

A.Veloso, E. Ceschia, M. Battude, J.F. Dejoux and V. Demarez are with the CESBIO, 31401 Toulouse, France (e-mail: velosoa@cesbio.cnes.fr; ceschiae@cesbio.cnes.fr; battudem@cesbio.cnes.fr; dejouxjf@cesbio.cnes.fr; demarezv@cesbio.cnes.fr).

M. Claverie is with the Department of Geographical Sciences, University of Maryland, College Park, MD 20742 USA (e-mail: mcl@umd.edu).

II. Data and Methods

A. Study Site

The study was carried out in southwest France, within a 24 km x 24 km area (1°10'E, 43°27'N), from 2006 to 2010. The study area is mainly covered by arable lands (around 60%). Major cultivated crops are wheat, barley, rapeseed, maize, soybean and sunflower. The latter, a non-irrigated crop, exhibits higher intra-field spatial variability due to spatial variation of soil properties and thus water availability. Moreover, sunflower plant density and green-up rate are extremely variable because of seeding damages. As they present the highest spatial variability, this work focuses on sunflower crops.

B. Field Data

Two sampling methods were used to collect the *in-situ* data:

a) Transect Sampling protocol

The GAI and biomass measurements were performed along two transects crossing the field [Fig.1(b)]. The transects were centered in the zone where surface fluxes measurements were performed [8]. This protocol was applied to one field, hereafter named as the Auradé site.

Effective GAI values were estimated from Digital Hemispherical Photographs (DHP) taken from above the canopy. The DHPs were processed using the imaging software

CAN-EYE V.5.1 [http://www4.paca.inra.fr/can-eye], providing values of effective GAI [9]. An amount of 15 photographs (spaced of 20 meters) were taken along each transect.

For biomass sampling, one plant was manually harvested every 20 meters (total of 30 plants). The sampling was dried in an oven at 65°C during 48 hours and the dry aboveground biomass was measured.

b) Elementary Sampling Unit (ESU) protocol

Thirteen DHPs were taken over an Elementary Sampling Unit (ESU) of 20m × 20m area [Fig.1(a)], according to the VALERI protocol [http://w3.avignon.inra.fr/valeri]. This protocol was applied to 3 fields (thus, 3 ESUs), 9 times per season.

Biomass was estimated from 10 plants randomly collected near each ESU at different dates (between 4 and 6) during the vegetative period.

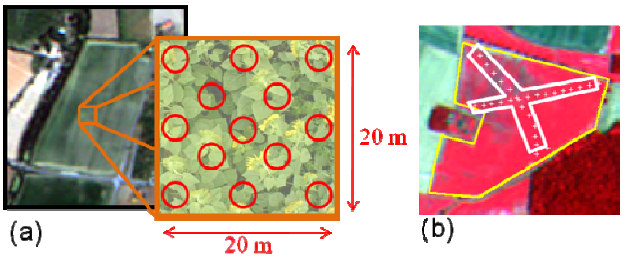


Fig 1 (a) ESU sampling protocol: spatial distribution of the 13 DHPs. (b) Transect sampling protocol (DHP and DAM) for the Auradé site.

C. Remote Sensing Data

1) Formosat-2

Formosat-2 (F2) satellite provides images with spatial resolution of 8m in four reflective bands centered at 488, 555, 650 and 830 nm. The sensor has a footprint of 24 km × 24 km and an orbital cycle of one day. Images are taken at near constant viewing angles (around 29°). A set of 112 images is available from 2006 to 2010 over the study area.

2) SPOT

The SPOT images were taken by the SPOT 2, 4 and 5 satellites. Data have spatial resolution of 20m (SPOT 2 and 4) or 10m (SPOT 5) in the green, red and near infrared spectral bands for SPOT 2 plus short-wave infrared band for SPOT 4 and 5. The field of view is 60 km. The viewing angles varied between +/- 27°. A total of 52 SPOT images were acquired over the 5 years period.

3) Images pre-processing

Both remote sensing datasets were processed with the KALIDEOS processing chain [http://kalideos.cnes.fr]. It provides valuable atmospheric, radiometric and geometric corrections [10], required for accurate time series studies (detailed information on [11]). The geometric correction accuracy obtained is on the order of 0.2 pixels.

D. Retrieving GAI from remote sensing data

The method used here is based on the inversion of the radiative transfer model PROSAIL [12] using Artificial Neural Networks (ANN). This approach relies on three main steps: a) creation of a learning database, b) training of the neural networks and finally c) application of the created networks for estimating the requested biophysical variables [13]. This approach is also denominated as BV-NET tool *(Biophysical variable neural network*). The ANN approach requires reflectance images as inputs (for all available spectral bands), plus the corresponding acquisition geometry (sun and view zenith angles, and sun-view relative azimuth angle). Moreover, it needs *a priori* distributions of input variables, which describe leaves and soil optical properties, and canopy geometry.

E. Meteorological data

The SAFRAN meteorological data [14], produced by Météo-France, were used for this study. SAFRAN provides air temperature at 2 m above the ground (Ta) and incoming global radiation (Rg) based on weather stations measurements and modeling. The data are available every 6 h over an 8 km spatial resolution grid. The daily mean Ta and Rg (which are the climatic variables required by SAFY) are calculated for each pixel over the study area using a bilinear interpolation.

F. Biomass estimation

1) The crop growth model

SAFY [7] is a daily time step crop model that simulates time series of GAI, dry aboveground biomass (DAM) and grain yield.

The approach for biomass estimation is based on the Monteith's light-use efficiency theory [15], coupled with a temperature stress function (not described here).

The model is driven by daily incoming global radiation (Rg) and cumulative daily mean air temperature (Ta). The biomass production starts on the plant emergence day (D_0). The daily biomass production is calculated by multiplying the photosynthetically active radiation absorbed by the canopy (APAR) by the effective light-use efficiency (ELUE), and is constrained by the temperature-stress-function F_T. The APAR is calculated using the Rg, the climatic efficiency (ε_c) and the fraction of photosynthetically active portion of solar radiation absorbed by the plants (FAPAR). The FAPAR is calculated from the effective green area index (GAI) and light-extinction coefficient (k_{ext}) according to Beer's law. The GAI growth is controlled by a leaf partition function. The GAI senescence starts at a certain level of cumulative Ta (parameter Stt) and is controlled by Rs (rate of senescence) parameter. A full description of SAFY is available in [7].

2) Calibration of the model

The values of the SAFY parameters applied to sunflower are given in [16]. In this study, four parameters are calibrated using the remotely sensed GAI time series data for each simulation unit (pixel, transect or field, Section II.G). These parameters are: the day of plant emergence (D_0), the effective light-use efficiency (ELUE), the rate of senescence (Rs) and the temperature threshold for starting senescence (S_{TT}). They allow fitting the simulated GAI courses to match with the BV-NET GAI time series.

The calibration method is based on the minimization of the Root Mean Square Error (RMSE) between the remotely sensed GAI time series (derived from BV-NET) and those

estimated by the SAFY model. An adapted version of the simplex method [17] was chosen.

The calibrated values of D_0, ELUE, Rs and S_{TT} parameters were used for running SAFY biomass simulations.

G. Methodology for SAFY simulations

In order to study the impact of the intra-field variability over the biomass estimations, the SAFY model was applied to different spatial units, according to different approaches, as detailed below:

a) Pixel-based approach

The model runs and is calibrated for each pixel of the chosen spatial extent, providing a biomass value for each pixel. This approach is applied for all the pixels encompassed in the whole field or the transect extent. A mean biomass and associated standard deviation could thus be computed for each spatial unit. For the ESU extent, if a SPOT image (20 m resolution) is used, the model runs for the pixel corresponding to the exact location of the ESU. For F2 images (8 m resolution), SAFY runs for a 3 x 3 pixels window, centered on the ESU, to match with the ESU *in-situ* sampled area.

b) Aggregative-based approach

An average GAI profile is computed from all the pixels of the field or the transect. SAFY runs once using the average GAI profile for the calibration, providing a biomass estimate for the whole field or transect.

III. RESULTS AND DISCUSSION

A. Evaluation of BV-NET GAI estimations

Fig. 2 shows the GAI retrieved by Neural Networks approach compared with the *in-situ* effective GAI estimated from hemispherical photographs. The performances in terms of correlation (R^2=0.89), bias (b=0.26 $m^2.m^{-2}$), absolute and relative root mean square errors (RMSE=0.62 $m^2.m^{-2}$, RRMSE=41.1%) indicate that estimations are satisfactory, considering no *a priori* field information was required by this method. We observe that sunflower GAI values estimated with BV-NET (GAI_{ANN}) are superior to GAI estimated from hemispherical photographs ($GAI_{CAN-EYE}$). The photographs are taken at the same height (using a pole) during all the season. Consequently, when the vegetation is well-developed the distance between the camera and the top of the canopy became shorter (than in the beginning of the season). It probably generates a saturation effect, resulting in lower estimations of effective GAI using DHPs [9]. This tendency is confirmed by the comparison between effective GAI values and GAI destructive measurements (provided in Fig 3). As expected, the latter were superior than the effective ones, since the CAN-EYE software tends to underestimate for clumped canopies [9]. The GAI_{ANN} seems to present a better correlation with the GAI derived from destructive measurements than with the $GAI_{CAN-EYE}$. However, further conclusions cannot be drawn, since concurrent acquisition of satellite image (obligatory for the ANN) and destructive measurements were available for only 4 points (indicated by the plus in Fig 3).

The analysis of the two highest GAI_{ANN} values indicates they were obtained from different sensors and within an interval of 2 days ($GAI_{ANN\text{-}SPOT}$= 4.4, in 28/06/2008 and $GAI_{ANN\text{-}Formosat\text{-}2}$= 3.3, in 26/06/2008). This unexpected difference might be due to different observation geometries. Furthermore, as the related NDVI values are quite similar ($NDVI_{SPOT}$= 0.79; $NDVI_{Formosat\text{-}2}$= 0.81), a potential improvement of BV-NET estimations could be achieved by adding the NDVI products as inputs, in addition to the reflectance images.

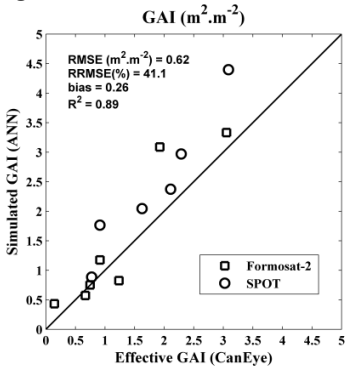


Fig. 2. Validation of sunflower GAI estimated by the Artificial Neural Networks approach.

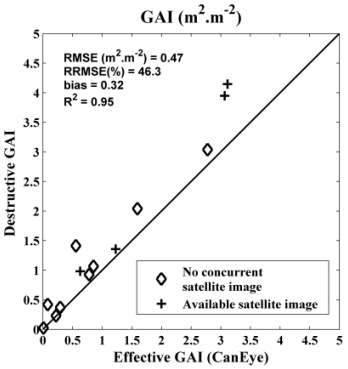


Fig 3. Comparison of ground destructive GAI and effective GAI estimated from digital hemispherical photographs using CAN-EYE.

B. Evaluation of GAI time series simulated by SAFY

SAFY model was run either for each pixel or for the averaged spatial extents (field and transect). Fig. 4 shows the GAI profile of the Auradé site simulated by SAFY (*lines*), and the corresponding *BV-NET* data used for calibration (*circles*). The standard deviations (std) associated to the *pixel-based* approach estimations (*black solid line*) illustrate the high spatial heterogeneity of this field. This characteristic can be visualized in the series of images presenting the temporal evolution of GAI [Fig. 5] during the vegetative season of 2007.

The analysis of the GAI profile of the "ESU2" on 2010 [Fig. 6] indicates that GAI values retrieved for the ESU extent (*green* line) are higher than the average GAI over the entire field (*black* and *blue* lines). The divergence between the GAI of the ESU and of the field extents confirms the intra-field variability. As it can be observed in Fig. 7, the zone where the ESU is located (*green square*) presents higher GAI values than other zones of the field (encircled in *red*).

The results for the 2 remaining ESU fields were analogous to those of Auradé site and are not presented here.

C. *Evaluation of the aboveground biomass estimations*

1) Biomass validation for transect and ESU sampling strategies

The SAFY biomass estimates can be validated by the comparison with the *in-situ* data collected using either the transect or the ESU protocol. Fig. 8 presents the resultant scatter plot from global comparison of simulated and measured DAM. The related performances in terms of correlation (R^2=0.91), bias (b=-34.59 $g.m^{-2}$) and absolute root mean square error (RMSE=85.97 $g.m^{-2}$) are satisfactory and show a good agreement between biomass estimates and the ground data.

Fig. 9-12 demonstrate that the dynamics of *in-situ* biomass data (*green circles*), collected using the transect or the ESU sampling protocols, are well reproduced by the SAFY estimations (*red* and *green* lines). Considering the significant standard deviations of the ground data, we notice that estimations are inside the ground range values. Additionally, it is interesting to notice that the standard deviations associated to the *in-situ* biomass increase during the growing season, reaching the highest values when the biomass production is maximum and when local heterogeneity tends to be more pronounced as well.

A decrease of the *in-situ* measured biomass values was observed at the end of the season [Fig. 9, 10, 12]. It is due to the loss of leaves during the senescent phase [16]. The fallen leaves are not taken into account when the biomass samples are collected, inducing lower biomass values. However, this phenomenon is not significant.

2) Impact of intra-field variability on biomass estimates

For the 4 investigated sunflower fields [Fig. 9-12], the biomass estimates calculated using the *aggregative-based* approach (*blue dashed* lines) fit with the biomass values estimated by the *pixel-based* approach (*black solid* lines). These results suggest that the intra-field variability has a low impact over the field total biomass estimates. We observe that the *aggregative* method can maintain the main characteristics of the original data, even if spatial details are lost during aggregation.

For the Auradé site, the heterogeneity observed on the GAI time series [Fig. 4] is reproduced on the biomass estimations as well, as seen in Fig. 9 (*black solid line* and related std). Even though Auradé is a very heterogeneous field, the average GAI profile used for the optimization process yields good biomass estimates. The comparison of the transect and field results suggests that the transect is fairly representative of the field, sampling correctly its spatial variability. However, the comparison should be carried out carefully, given that the standard deviations associated to ground biomass measurements are considerably high.

For the "ESU2" and "ESU3", on 2010, the differences of biomass estimates between ESU unit (*green line*) and field unit (*black and blue lines*), shown in Fig. 11 and 12, are expected since the ESUs are located in a zone with high GAI values [Fig. 7, for "ESU2"]. As demonstrate by [7], higher values of GAI yield in higher biomass estimates.

Globally, the achieved results suggest the high spatial resolution (8, 10 or 20m) seems to be less crucial than expected when performing biomass estimations. Given that similar estimates are obtained for *pixel-based* and *aggregative-based* approaches, the choice of running the model for aggregated units would allow gaining on time-consuming performance. Still, the high spatial resolution remains essential for detecting field boundaries.

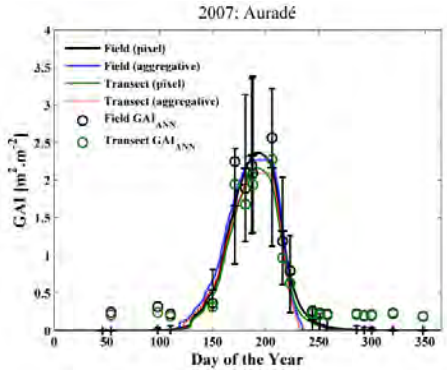


Fig. 4. Auradé site GAI profiles simulated by SAFY for the field (in *blue* and *black*) and transect (in *green* and *red*) extents by means of the pixel (*solid lines*) and aggregative (*dashed lines*) approaches. GAI derived from ANN method and used for the model calibration are represented by the circles.

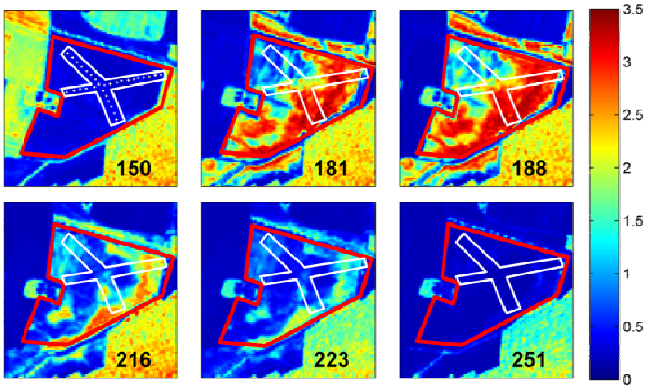


Fig. 5. Temporal evolution of the GAI over the Auradé experimental site (delimited in red), for six dates (DOY) in 2007.The transect shape is indicated in white.

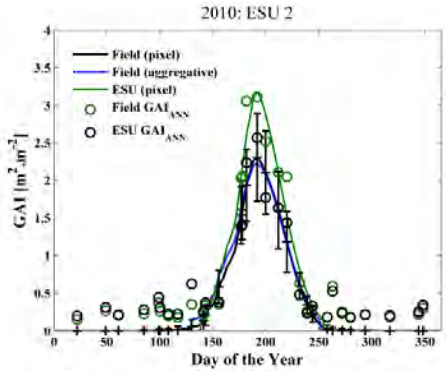


Fig. 6. GAI profiles simulated by SAFY for different spatial units (lines) and derived from satellite images (circles) over the field of ESU2, on 2010.

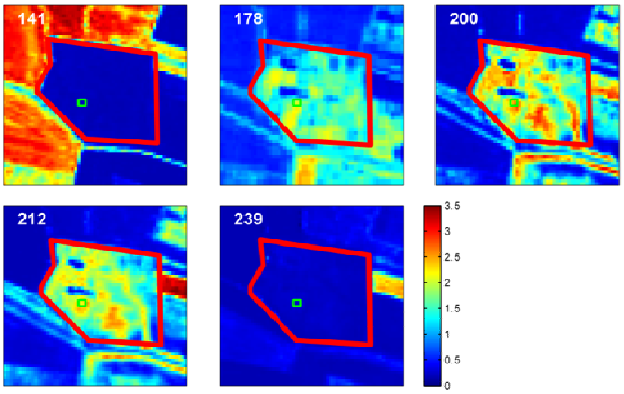


Fig. 7. Temporal evolution (DOY in white) of the GAI over the field (delimited in red) of ESU2, on 2010. The green square indicates the ESU specific location.

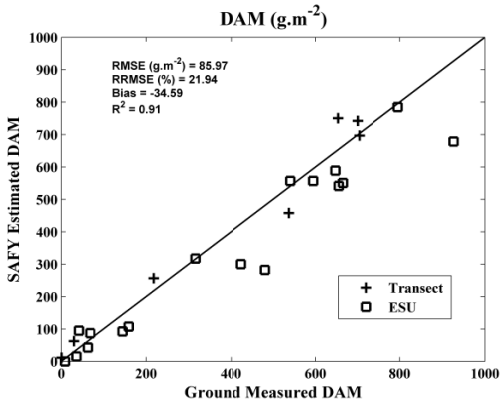


Fig. 8. Comparison between DAM simulated by SAFY and ground measurements (performed either using the transect or the ESU protocol).

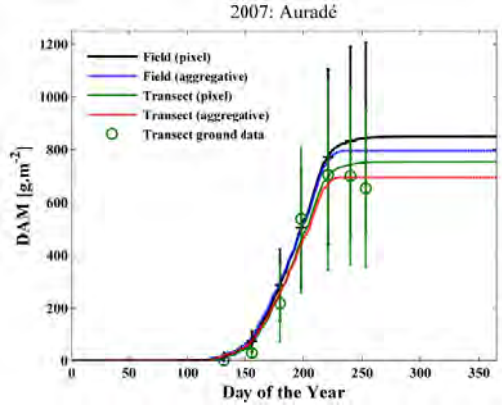


Fig. 9. Dry aboveground biomass (DAM) simulated by SAFY for field (in *blue* and *black*) and transect (in *green* and *red*) extents by means of the pixel-based (*solid lines*) and aggregative-based (*dashed lines*) approaches and *in-situ* measurements (circles) over the Auradé site, in 2007.

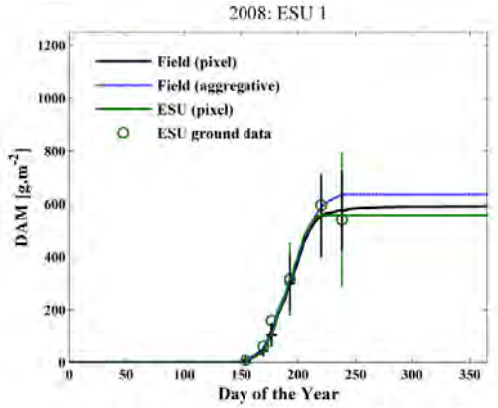


Fig. 10. Biomass estimates (lines) and ground measurements (circles) for ESU1, on 2008.

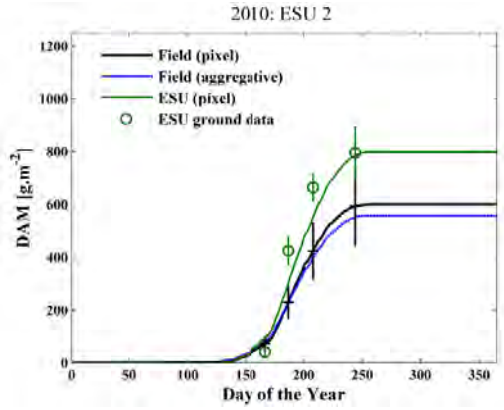


Fig. 11. Biomass estimates (lines) and ground measurements (circles) for ESU2, on 2010.

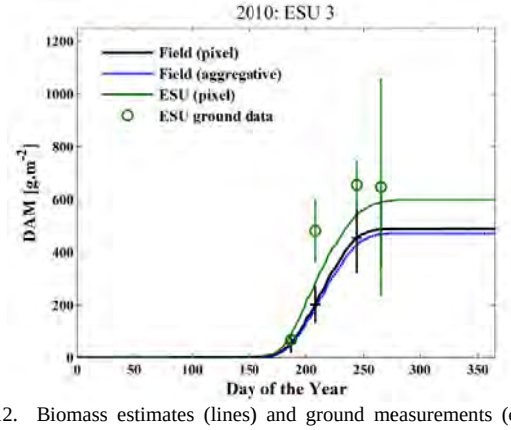


Fig. 12. Biomass estimates (lines) and ground measurements (circles) for ESU3, on 2010.

IV. Conclusion

This work has demonstrated that the BV-NET tool, based on the inversion of PROSAIL model using artificial neural networks, yields accurate GAI estimates for the sunflower crop. A major advantage of this method is that it is site-independent, can be easily applied to different sensors and requires no *a priori* calibration. Still, improvements over this approach might be accomplished by including NDVI information as input variable on the neural networks training step.

This study also revealed a saturation effect when estimating

GAI from digital hemispherical photographs. This effect is due to the short distance (< 1m) between the canopy and the camera. This inconvenient was partially avoided by putting the camera on the top of a 2.5m height pole. Nonetheless, the problem persists when canopy height is superior to 1.5m.

The dry aboveground biomass estimates provided by the SAFY model were validated against ground measurements. The results confirm the model is able to simulate the temporal and spatial dynamics of biomass production. Yet, the availability of high temporal resolution GAI dataset remains indispensable.

The comparison of the *pixel-based* and the *aggregative-based* approaches revealed that taking into account the intra-field spatial variability has no significant effect on the whole field biomass estimates. Thus, the *aggregative* approach might be favored due to its better time-consuming performance if mean field biomass estimates are targeted.

ACKNOWLEDGMENT

The Formosat-2 images used in this paper are © NSPO (2006) and distributed by Spot Image S.A. all rights reserved. The authors would like to acknowledge Olivier Hagolle, Mireille Huc and the KALIDEOS program for the satellite images processing, and the CNRM-GAME (Météo France) for providing the SAFRAN meteorological data. We also thank Marie Weiss and Frédéric Baret for the support on BV-NET tool and CAN-EYE software.

REFERENCES

[1] T. O. West, C. C. Brandt, L. M. Baskaran, C. M. Hellwinckel, R. Mueller, C. J. Bernacchi, V. Bandaru, B. Yang, B. S. Wilson, and G. Marland, "Cropland carbon fluxes in the United States: increasing geospatial resolution of inventory-based carbon accounting," *Ecological Applications*, vol. 20, pp. 1074-1086, 2010.

[2] S. Moulin, A. Bondeau, and R. Delecolle, "Combining agricultural crop models and satellite observations: from field to regional scales," *International Journal of Remote Sensing*, vol. 19, pp. 1021-1036, 1998.

[3] A. de Wit, G. Duveiller, and P. Defourny, "Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations," *Agricultural and Forest Meteorology*, vol. 164, pp. 39-52, Oct.

[4] J. Kolasa and C. D. Rollo, "The heterogeneity of heterogeneity," *Ecological Heterogeneity. Springer-Verlag, New York*, 1991.

[5] L. Bian and R. Butler, "Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data," *Photogrammetric Engineering and Remote Sensing*, vol. 65, pp. 73-84, 1999.

[6] M. Weiss and F. Baret, "Venμs biophysical variable products algorithm theoretical basis document," EMMAH, INRA 2010.

[7] B. Duchemin, P. Maisongrande, G. Boulet, and I. Benhadj, "A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index," *Environmental Modelling & Software*, vol. 23, pp. 876-892, 2008.

[8] P. Béziat, E. Ceschia, and G. Dedieu, "Carbon balance of a three crop succession over two cropland sites in South West France," *Agricultural and Forest Meteorology*, vol. 149, pp. 1628-1645, 2009.

[9] V. Demarez, S. Duthoit, F. Baret, M. Weiss, and G. Dedieu, "Estimation of leaf area and clumping indexes of crops with hemispherical photographs," *Agricultural and forest meteorology*, vol. 148, pp. 644-655, 2008.

[10] O. Hagolle, G. Dedieu, B. Mougenot, V. Debaecker, B. Duchemin, and A. Meygret, "Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: application to Formosat-2 images," *Remote Sensing of Environment*, vol. 112, pp. 1689-1701, 2008.

[11] B. Lafrance, X. Lenot, C. Ruffel, P. Cao, and T. Rabaute, "Outils de pretraitements des images optiques KALIDEOS," *Revue française de photogrammétrie et de télédétection*, pp. 10-16, 2012.

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

[13] F. Baret, O. Hagolle, B. Geiger, P. Bicheron, B. Miras, M. Huc, B. Berthelot, F. Niño, M. Weiss, and O. Samain, "LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm," *Remote Sensing of Environment*, vol. 110, pp. 275-286, 2007.

[14] Y. Durand, E. Brun, L. Mérindol, G. Guyomarc'h, B. Lesaffre, and E. Martin, "A meteorological estimation of relevant parameters for snow models," *Annals of Glaciology*, vol. 18, pp. 65-71, 1993.

[15] J. L. Monteith and C. J. Moss, "Climate and the efficiency of crop production in Britain [and discussion]," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 281, pp. 277-294, 1977.

[16] M. Claverie, V. Demarez, B. Duchemin, O. Hagolle, D. Ducrot, C. Marais-Sicre, J. F. Dejoux, M. Huc, P. Keravec, and P. Béziat, "Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data," *Remote Sensing of Environment*, 2012.

[17] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization*, vol. 9, pp. 112-147, 1998.