



HAL
open science

Outils, références et méthodes pour la construction d'un simulateur pour la prévision du rendement et de la qualité du tournesol à l'échelle territoriale mobilisant la télédétection satellitaire

Luc Champolivier, Philippe Debaeke, Jean-François Dejoux, Caroline Dizien, Arnaud Micheneau, Céline Colombet, Hervé Gibrin, Célia Pontet, Al Bitar Ahmad, Ronan Trépos, et al.

► To cite this version:

Luc Champolivier, Philippe Debaeke, Jean-François Dejoux, Caroline Dizien, Arnaud Micheneau, et al.. Outils, références et méthodes pour la construction d'un simulateur pour la prévision du rendement et de la qualité du tournesol à l'échelle territoriale mobilisant la télédétection satellitaire. *Innovations Agronomiques*, 2019, 71, pp.15-34. 10.15454/qmd4zh . hal-02105013

HAL Id: hal-02105013

<https://hal.inrae.fr/hal-02105013>

Submitted on 19 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Outils, références et méthodes pour la construction d'un simulateur pour la prévision du rendement et de la qualité du tournesol à l'échelle territoriale mobilisant la télédétection satellitaire

Champolivier L.¹, Debaeke P.², Dejoux J.F.³, Dizien C.⁴, Micheneau A.¹, Colombet C.², Gibrin H.³, Pontet C.¹, Al Bitar A.³, Trépos R.⁵, Ansart A.³, Marais Sicre C.³, Garric B.¹, Mestries E.¹, Casadebaig P.², Fernandez-Diclo Y.²

¹ Terres Inovia, Centre de recherche INRA de Toulouse, Bât. AGIR, 24 Chemin de Borde Rouge, Auzeville, CS52627, F-31326 Castanet Tolosan Cedex

² INRA - UMR AGIR, 24 Chemin de Borde Rouge, Auzeville, CS52627, F-31326 Castanet Tolosan Cedex

³ CESBIO, 18 av. Edouard Belin, bpi 2801, F-31401 Toulouse cedex 9

⁴ Agrosolutions - 83 avenue de la Grande Armée, F-75016 Paris

⁵ INRA - UR MIAT, 24 Chemin de Borde Rouge, Auzeville, CS52627, F-31326 Castanet Tolosan Cedex

Correspondance: l.champolivier@terresinovia.fr

Résumé

L'objectif du projet est de proposer des méthodes permettant à terme de construire un outil de prévision du rendement et de la qualité (teneur en huile et teneur en acide oléique) du tournesol quelques semaines avant sa récolte, à l'échelle d'un bassin de collecte, à l'aide de modèles plus ou moins complexes associés à des observations d'états de culture (GAI : green area index) acquises par voie satellitaire et supposées améliorer la qualité de prédiction des modèles. Il s'agit donc d'identifier les types de modèles les mieux adaptés à la prévision à l'échelle territoriale parmi une gamme de modèles allant des modèles corrélatifs les plus simples (à une ou deux variables explicatives) aux modèles de cultures dynamiques plus ou moins spécifiques (SUNFLO, SAFY). L'intérêt des techniques d'assimilation de données de GAI acquises par télédétection dans les modèles dynamiques de culture pour en améliorer la qualité de prédiction est également évalué.

Des sujets périphériques à ces problématiques, mais incontournables dans l'objectif d'élaborer les modèles et de construire un outil de prévision opérationnel, sont étudiés : (i) détermination de méthodes simples et rapides d'estimation du GAI au sol afin de faciliter la constitution de réseaux de dispositifs de « vérité terrain » nécessaires à la calibration des modèles biophysiques, (ii) constitution d'une base de données d'itinéraires techniques, de contextes de culture (données d'entrée des modèles de culture), de production (rendement, teneurs en huile et en acide oléique : données indispensables pour construire et évaluer les modèles de prédiction) et d'images satellitaires sur plusieurs centaines de parcelles agricoles, (iii) estimation de la date de levée de la culture (rarement disponible, mais indispensable pour certains modèles de culture) par voie satellitaire et (iv) localisation de toutes les parcelles de tournesol d'un territoire en cours de campagne, à l'aide de la télédétection.

Le principal résultat est que les modèles corrélatifs simples à une ou deux variables explicatives (GAI maximal au cours du cycle et persistance du GAI pendant la phase reproductive du cycle) paraissent les plus pertinents pour prédire le rendement à l'échelle de parcelles au sein d'un territoire.

Mots-clés : prévision, rendement, télédétection satellitaire, territoire, tournesol

Abstract: Tools, references and methods for the construction of a simulator for forecasting grain yield and quality of sunflower crop at the territorial level using satellite remote sensing

The aim of the project was to elaborate methods making it possible to build a tool for predicting grain yield and quality (oil and oleic acid concentrations) of sunflower a few weeks before crop harvest, at the scale of a production area, using a range of models associated with observations of green area index (GAI). GAI was estimated by remote sensing that is supposed to improve the quality of prediction models. It was therefore necessary to identify the types of models that were best suited to territory-scale forecasting among a range of models from the simplest correlative ones (with one or two explanatory variables) to dynamic crop models (SUNFLO, SAFY). The interest of using assimilation techniques of GAI data into dynamic crop models to improve the accuracy of prediction was also evaluated.

With the perspective of building an operational forecast tool, other related issues were studied: (i) simple and rapid methods of GAI estimation at field level to facilitate the set up of networks of "ground truth" devices for the calibration of biophysical models, (ii) creation of a database of cultural practices, initial crop conditions (input data of crop models), production (yield, oil and oleic acid contents: key data for building and evaluating prediction models) and satellite images on several hundreds of agricultural plots, (iii) method for estimating the date of crop emergence (seldom available, but essential for some crop models) by satellite remote sensing and (iv) mapping with remote sensing all the sunflower fields of a territory during the growing season.

The main result was that simple correlative models with one or two explanatory variables (maximum GAI at anthesis and duration of the GAI after flowering) seem to be the most relevant ones for predicting sunflower performance at the scale of a territory.

Keywords: prediction, yield, satellite remote sensing, territory, sunflower

Introduction

La prédiction du rendement et de la qualité des graines (huile, protéines) à l'échelle territoriale, quelques semaines avant la récolte, présente un intérêt stratégique pour les coopératives qui collectent, stockent et commercialisent le tournesol. Ainsi, la prévision du rendement doit permettre de gérer au mieux leur logistique et réduire ainsi les coûts financiers et environnementaux du stockage. La prévision de la teneur en huile revêt quant à elle encore plus d'importance. En effet, dans la majorité des cas, les coopératives ne déterminent pas cette teneur à la réception des graines livrées par les agriculteurs car il n'existe pas encore de méthodes de mesure rapides et peu coûteuses au silo. Pour des raisons de gestion de lots en prévision de leur commercialisation, elles ont néanmoins besoin d'avoir une estimation prévisionnelle de la teneur en huile de leurs différents lots issus des différentes situations de leurs bassins de production. La précision requise sur la valeur absolue n'est pas très forte. Pouvoir classer les situations serait déjà très utile. Enfin, la teneur en acide oléique conditionne l'accès à des marchés à plus forte valeur ajoutée avec des effets de seuils.

La prévision du rendement et de la qualité de la récolte des cultures quelques semaines à quelques mois avant leur récolte a fait l'objet de nombreux travaux s'appuyant sur des méthodes dont la nature varie en fonction de l'échelle de prévision (tant spatiale que temporelle).

La première série de méthodes s'appuie sur l'observation d'un réseau de parcelles de référence. Dans ce cas, un grand nombre de parcelles (plusieurs centaines) sont enquêtées ou suivies à l'échelle d'un territoire, après échantillonnage selon la distribution des sols et des pratiques. Des diagnostics agronomiques peuvent être conduits pour déterminer les facteurs limitants du rendement et de la qualité pour situer et comprendre l'écart au rendement accessible (Doré et al., 2008 ; Champolivier et al., 2011 ; Hall et al., 2013). Ces données sont utilisées pour établir des relations entre le rendement, ses composantes et les principaux facteurs limitants de la production (Supit et Wagner, 1999 ; Mercu

et al., 2001). Ces enquêtes agronomiques ne permettent pas en général de prévoir le rendement ou la qualité une année donnée, mais elles permettent de mieux comprendre les causes d'écart au potentiel dans un territoire donné.

En cours de culture, des indicateurs précoces peuvent cependant être utilisés pour prévoir la qualité de la récolte (par ex teneur en protéines du grain) ou même le rendement à partir de mesures sur la plante, par exemple l'indice de nutrition azotée ou la teneur en chlorophylle des feuilles (Le Bail et al., 2005 ; Barbottin et al., 2008). Ces méthodes s'appuient sur un réseau de parcelles de référence dont la représentativité doit être bien établie au préalable en lien par exemple avec une typologie des situations de production.

Une deuxième approche consiste à utiliser des modèles de prévision du rendement et de la qualité à l'échelle de la parcelle puis de procéder à la régionalisation des prévisions par des techniques dites de « changement d'échelle ». Les modèles appliqués à l'unité de base (parcelle ou situation homogène) peuvent être des modèles de simulation dynamique ou des modèles statistiques. Cette régionalisation devient plus complexe dès lors que l'on ne considère pas seulement le climat et les sols mais aussi les pratiques culturales dont l'accès en temps réel est difficile. La littérature est assez riche en cas d'étude où des modèles développés à la parcelle sont utilisés à une échelle régionale, qu'il s'agisse de modèles agroclimatiques simples (Potgieter et al., 2005) ou de modèles à base de processus écophysologiques (Jagtap et Jones, 2002 ; Launay et Guérif, 2003 ; Champolivier et al., 2012). Utilisés seuls, ces modèles sont plus souvent utilisés pour estimer les rendements accessibles et la variabilité spatio-temporelle sur de longues séries climatiques que pour prédire le rendement de l'année. Les modèles qui simulent l'élaboration de la teneur en huile et en acide oléique à cette échelle font largement défaut (Pereyra-Irujo et al., 2009).

Une troisième approche couple observations pendant la campagne (par imagerie satellitaire) et modèle de culture : les données observées (souvent l'indice foliaire mais aussi l'humidité du sol) sont assimilées dans le modèle dynamique pour corriger les incertitudes liées au modèle ou aux données d'entrée et améliorent ainsi les prévisions (Launay et Guérif, 2005 ; Dorigo et al., 2007 ; de Wit et van Diepen, 2007). Cette approche combinant observations et modèles, sur un territoire important, permet d'envisager des applications de prédiction du rendement et de la qualité en temps réel. Le modèle SAFY (Duchemin *et al.*, 2008) par exemple ne considère que le GAI satellitaire (dédit du NDVI) et le climat journalier pour prédire le rendement des cultures.

Les approches couplant modèles de culture et données de télédétection sont souvent restreintes au rendement (Belaqziz *et al.*, 2013 ; Mkhabela *et al.*, 2011 ; Ren *et al.*, 2008). Nous n'avons pas identifié de travaux portant sur la prévision de la teneur en huile et en acide oléique à une échelle régionale. Par ailleurs, le tournesol est peu souvent retenu comme espèce d'application pour la télédétection (Pena-Baragan et al., 2010 ; Claverie et al., 2012) alors qu'il y aurait un intérêt évident à mesurer le GAI pour cette espèce.

Le projet a reposé sur un partenariat entre Terres Inovia, l'INRA, le CESBIO et InVivo Agrosolutions. Il a bénéficié de la dynamique partenariale mise en place depuis plusieurs années dans le cadre de l'UMT Tournesol.

L'objectif du projet est de proposer des méthodes permettant, à terme, de construire un outil de prévision de collecte (quantité et qualité) du tournesol quelques semaines avant sa récolte, à l'échelle d'un bassin de collecte et de ses différentes situations de culture. Nous faisons l'hypothèse que l'utilisation combinée de différents types de modèles plus ou moins complexes (modèles corrélatifs simples, modèles de culture spécifique ou non) et d'observations issues de la télédétection satellitaire (indices de végétation et indice foliaire) permettront de réaliser de bonnes prédictions. Nous avons choisi de réaliser les simulations à l'échelle de la parcelle agricole et de passer à l'échelle du bassin de collecte en agrégeant les résultats parcellaires.

Ce projet s'est déroulé sur les campagnes 2014 à 2016 et se situe dans un contexte de préparation à l'arrivée des satellites Copernicus européens Sentinel-1 (radar) et surtout Sentinel-2 (optique). Les 2 satellites Sentinel-2 ont été lancés respectivement en juin 2015 et en mars 2017. A eux deux, ils permettent une prise de vue systématique de toute la France tous les 3 jours à 5 jours. Les images sont gratuites et libre d'accès, en temps réel.

Les travaux ont été conduits dans la région toulousaine avec l'appui des coopératives Arterris et Terres de Gascogne, mais l'outil a vocation à être utilisé partout.

1. Méthodes d'estimation de l'indice foliaire par voie satellitaire

En amont des aspects de modélisation de la production de la culture en quantité et qualité, l'opérationnalité future des outils qui découleront de ce projet est conditionnée par notre capacité à répondre à différentes questions d'ordre méthodologique. Ainsi, nous prévoyons de mobiliser des données de GAI (green area index = indice de surface verte = surface verte déployée des feuilles, tiges et capitules / surface occupée par les plantes au sol) potentiellement accessibles par télédétection satellitaire. Cela nous a amené à travailler sur deux sujets strictement liés à la télédétection :

- Deux types de modèles et méthodes de calcul du GAI à partir des signaux de réflectance issus des capteurs satellitaires : modèles corrélatifs basés sur des indices de végétation et méthode d'inversion de modèle de transfert radiatif permettant d'accéder directement à des variables biophysiques comme le GAI.
- Ces modèles nécessitent d'être évalués voire paramétrés à l'aide de résultats de mesures réalisées au sol qui sont jusqu'à maintenant lourdes et donc peu compatibles avec un déploiement à l'échelle d'un réseau de « vérité terrain » sur un grand territoire. Nous avons donc évalué différentes méthodes de mesures indirectes et rapides (par capteurs) du GAI au sol.

1.1 Méthodes indirectes de mesure du GAI au sol

Plusieurs méthodes indirectes de mesure du GAI au sol (photos hémisphériques au-dessus de la végétation, LAI-2000 de chez Li-Cor, appareil Pieton V1 issu des travaux de l'UMT Capte) ont été comparées à la méthode de référence historiquement employée (méthode allométrique basée sur la mesure de la longueur et de la largeur des feuilles vertes) dans des dispositifs expérimentaux en micro-parcelles mis en place à l'INRA chaque année du projet. Toutes ces méthodes sont non destructives et permettent, par exemple, un suivi dans le temps des mêmes plantes et mêmes zones. Les modalités expérimentales ont été définies de façon à créer une diversité de structures du couvert végétal (variétés à ports différents, niveaux d'alimentation hydrique et azotée, écartements entre rangs, densités de peuplement, stades de mesure, hétérogénéité du peuplement variés).

Un tableau de comparaison des méthodes selon différents critères a été produit (Tableau 1). Les résultats montrent que plusieurs d'entre-elles sont utilisables en pré-floraison et que leurs performances sont assez comparables. Cela suppose néanmoins un respect des conditions d'éclairement lors de la prise de mesures. À partir de la floraison, il est difficile d'éviter le biais causé par la présence des capitules. Une correction standard peut cependant être apportée. En période de sénescence, la méthode des photographies hémisphériques semble la mieux adaptée car elle se concentre sur la strate supérieure. En revanche, cette méthode nécessite un post-traitement et est sujette aux interprétations de l'opérateur, ce qui peut affecter les résultats. La méthode Pieton est prometteuse et pourrait constituer un bon compromis entre faisabilité, fiabilité et précision sur l'ensemble du cycle cultural. Elle offre la possibilité de faire des mesures dynamiques en continu sur des surfaces de quelques dizaines de mètres carrés, ce qui permet d'améliorer l'échantillonnage de la zone utilisée pour

la calibration des données satellitaires (les autres méthodes ne permettent de réaliser que des mesures ponctuelles). Elle semble moins sensible aux conditions environnementales et à la présence du capitule.

Tableau 1 : Comparaison des différentes méthodes de mesure du GAI à la lumière des résultats du projet

Méthode	Allométrie	LAI 2000 (Li-Cor)	PIETON (UMT CAPTE)	Photos hémisphériques (Obj. Fish Eye)	Photos drone (capteur Airphen)
Protocole de mesure (1 parcelle)	3 x 4 plantes	1 mesure au-dessus + 4 en-dessous du couvert	Mesure en dynamique dans l'interrang	8 photos	Prise de photos aériennes
	1 feuille / 2				
	Densité parcellaire Coeff.de forme à déterminer (par défaut 0.7)				
Points forts	Mesure de référence	Rapidité	Rapidité	Prise de vue par-dessus ou par-dessous	Vue globale (parcelle, essai)
	Suivi plante	Accès direct au GAI	Accès direct au GAI		
Points faibles	Long et fastidieux : 15 – 45 mn selon stades	Sensible aux conditions d'éclairement	Sensible aux conditions d'éclairement	Sensible aux conditions d'éclairement	Sensible au vent
	Traitement post-mesure (optimisé)	Valeur approchée du GAI	Prototype en cours de développement	Traitement post-mesure (logiciel CAN-EYE etc...)	Traitement post mesure

1.2 Méthode d'estimation de l'indice foliaire par voie satellitaire

La démarche consiste à paramétrer les modèles permettant d'estimer le GAI à partir des images satellitaires grâce aux observations réalisées au sol par méthode indirecte (photos hémisphériques). Deux types de modèles ont été testés : corrélation entre un indice de végétation (NDVI : normalized difference vegetation index) et le GAI observé, et inversion de modèle de transfert radiatif.

Le dispositif expérimental était constitué de 14 (2014), 13 (2015) et 14 (2016) placettes d'environ 400 m² (l'équivalent d'un pixel d'image satellitaire) dites « ESU » (elementary sampling units) (Baret *et al.*, 2014) localisées dans des zones homogènes de parcelles agricoles de la Haute-Garonne et du Gers. Chaque année, les ESU ont été choisies de façon à représenter une large gamme de conditions de culture (type de sol, variété, date de semis, densité de peuplement, potentiel de croissance...). En plus de ces mesures, des observations ont été réalisées de façon à identifier d'éventuels facteurs limitants (densité de peuplement, maladies, enherbement...).

Sur toutes ces ESU, des mesures de GAI au sol ont été réalisées à l'aide de plusieurs méthodes :

- Photographies hémisphériques prises à 1.20 m par-dessus le sommet de la végétation (traitées avec le logiciel CAN-EYE développé par l'INRA d'Avignon) et mesures avec le Pieton, chaque année (seulement en 2016 pour Pieton) sur toutes les ESU et toutes les deux semaines de mi-juin à début août (soit de l'ordre de 5 dates de mesure).
- Méthode allométrique aux mêmes dates que les photographies hémisphériques en 2014 et 2016 et seulement à deux dates en 2015 (avant et après floraison).

Parallèlement, des images satellitaires ont été acquises sur ces parcelles à 7 dates en moyenne en 2014 (satellite FormoSat), à 13 dates en moyenne sur l'ensemble du cycle en 2015 (de 8 à 22 selon les parcelles ; issues de l'expérience Spot 5-Take 5), à 8 dates sur l'ensemble du cycle en 2016 (satellite Sentinel-2A) sur le dispositif 'Nailloux' (deux fois moins sur le dispositif 'Montoussé'). Quelques données Landsat 8 ont été acquises en conditions non nuageuses sur chacune des 3 campagnes. C'est donc ce

capteur satellitaire qui permet une comparaison entre les 3 années. Ce jeu de données est complété par une campagne de mesures réalisées en 2013, donc hors projet, par le CESBIO.

La méthode consistant à corrélérer le GAI au NDVI n'a pas donné satisfaction (valeurs de GAI trop faibles). Le modèle d'inversion BVNET (Biophysical Variable Neural Network : outil développé par l'INRA d'Avignon, UMR EMMAH, qui se base sur un calcul par inversion d'un modèle de transfert radiatif et résolution numérique grâce à des réseaux de neurones), recodé sous Orfeo Toolbox par le CESBIO afin de permettre une utilisation plus opérationnelle, présente l'avantage théorique de permettre de calculer le GAI pour différents capteurs satellitaires sans nécessité de données de calibration de terrain. Son utilisation a donné satisfaction. La corrélation avec les données terrain 2014 et 2015 a été meilleure, en particulier en 2015, avec Spot 5-Take 5 et pour les valeurs élevées de GAI (de 2 à 4), qu'avec le modèle corrélatif disponible.

2. Méthodes d'acquisition d'informations contextuelles par imagerie satellitaire

2.1 Méthode d'identification et de géolocalisation des parcelles de tournesol

L'identification et la géolocalisation de toutes les parcelles de tournesol sont des préalables à la prévision de la production à l'échelle d'un bassin de collecte. À l'avenir, ces informations devraient pouvoir être fournies par les coopératives. C'est encore rarement le cas à l'heure actuelle. Par ailleurs, la question se poserait également pour des études à une échelle administrative (canton, département...). Dans le contexte du projet, le besoin est de disposer de cartes d'occupation du sol de la campagne culturale en cours 4 à 6 semaines avant la récolte. Il s'agit d'une estimation en temps réel. Des travaux sont en cours au CESBIO sur ce sujet. La possibilité de produire des estimations en temps réel et avec un haut niveau de précision nécessite de combiner en particulier des informations historiques (cartes d'occupation du sol des 3 à 6 années passées obtenues par la PAC Graphique ou/et les images satellitaires des années considérées) et quelques images satellitaires acquises sur la campagne en cours.

Des méthodes de classification combinant l'utilisation de séries temporelles d'images satellitaires exploitées en NDVI, des données d'apprentissage acquises sur le terrain et un masque issu du registre parcellaire graphique (RPG) permettant de distinguer les zones portant des grandes cultures ont été testées.

Les performances de ces méthodes sont très bonnes sur de petites zones (500 à 1000 km²), quelle que soit l'année climatique comme l'illustre la Figure 1. Les performances de la classification augmentent au fil de la saison. Si on utilise les images jusqu'à fin mai, la précision pour le tournesol est de 62% uniquement puisque à ce niveau de précocité les parcelles de tournesol se trouvent injustement classées dans d'autres classes de cultures d'été (maïs irrigué, sorgho ou soja). L'ajout de nouvelles images satellites, permet d'augmenter la précision pour le tournesol en diminuant les mélanges avec les autres cultures. Les meilleurs résultats sont obtenus fin août. À cette date la classification permet d'identifier 84% des tournesols. Ces tests ont été réalisés sur plusieurs années, ce qui permet de dire que la tendance visible sur la Figure 1 est représentative, pour la zone étudiée et qu'elle n'est pas très impactée ni par l'année climatique, ni par la date des images utilisées (Marais Sicre *et al.*, 2016).

Travailler sur de grandes zones de plus de 10 000 km² dégrade les résultats. L'hypothèse est que sur une grande zone, la très grande diversité des cultures, des conditions météorologiques, des types de sols et des pratiques (travail du sol, période de semis...) entraîne une grande hétérogénéité des stades au sein de chaque culture et donc de nombreuses confusions possibles.

On peut penser que ces méthodes sont assez robustes sous réserve d'obtenir au moins 3 ou 4 images satellitaires bien réparties entre février et fin juillet, ce qui est réaliste avec les satellites aujourd'hui

disponibles. Des travaux récents du CESBIO confortent cette hypothèse de robustesse par l'ajout d'images radar comme celles fournies par Sentinel-1. Dans l'attente de nouveaux résultats, il est prudent d'utiliser ces méthodes de classification dans des bassins de production agricole restreints (moins de 2 000 km²) et assez homogènes en systèmes de production. La distinction du tournesol nécessite la collecte d'un échantillon de données de terrain sur l'ensemble des cultures (nature et géolocalisation).

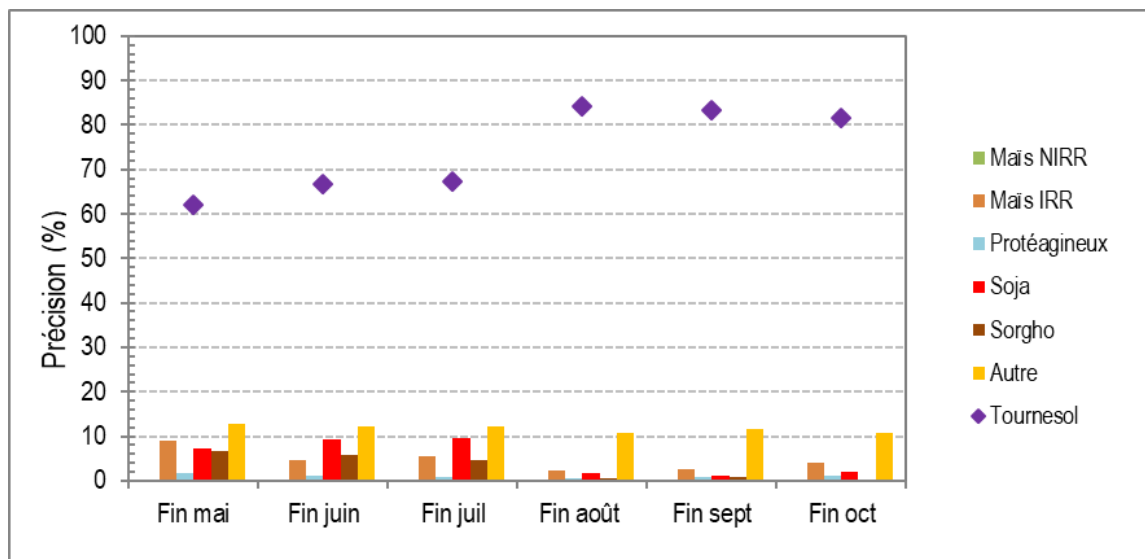


Figure 1 : Taux de parcelles de tournesol bien classées (losange violet). Taux de parcelles de tournesol classées à tort dans d'autres classes de culture (histogrammes). Année 2015, images Landsat 8, « vérité terrain » du CESBIO. Les classifications sont réalisées pour chaque mois avec l'ajout des nouvelles images de ce mois. Ces résultats ont été obtenus sans utilisation d'informations des années précédentes. NIRR : non irrigué. IRR : irrigué.

2.2 Méthode d'estimation de la date de levée par voie satellitaire

Les modèles de culture nécessitent la plupart du temps de disposer de la date de levée comme variable d'entrée. À ce jour, celle-ci n'est jamais disponible pour l'ensemble des parcelles d'un bassin. Comme pour la localisation des parcelles, on peut imaginer qu'à terme, elle le sera dans les bases de données de certaines coopératives. Nous avons néanmoins choisi de travailler sur une solution qui permettrait de calculer la date de levée de chaque parcelle à partir d'images satellitaires.

Plusieurs jeux de données ont été obtenus en 2014 et 2015 : sur plusieurs centaines de parcelles agricoles géoréférencées, les dates de semis déclarées par les agriculteurs ont été collectées et des notations de stades précoces (levée et 2 à 8 feuilles) réalisées.

Cette base de données d'observation « terrain » n'a pas encore été valorisée en lien avec les données satellitaires. Une limite forte à sa valorisation est le faible nombre d'images satellitaires disponibles entre fin avril et fin juin en 2014. Ce problème est fréquent dans le sud-ouest où la nébulosité est maximale pendant ces 2 mois. Le jeu de données 2015 est plus exploitable grâce à l'expérience Spot 5-Take 5, même s'il n'y a aucune image entre le 10 avril et le 3 juin 2015.

3. Constitution d'une base de données pour le paramétrage et l'évaluation des modèles de prévision du rendement

Le constat a été fait avec les deux coopératives partenaires du projet (Arterris, Terres de Gascogne) qu'il n'était pas possible d'obtenir des valeurs de collecte (en quantité et qualité) sur une aire

correspondant à un ou plusieurs silos. Le lien entre le lieu de livraison des graines par les agriculteurs et les parcelles récoltées n'est pas suffisamment étroit et la teneur en huile n'est jamais estimée à cette échelle par les coopératives. Pour pallier cette difficulté, nous avons choisi de travailler à l'échelle d'un réseau d'exploitations agricoles, en assimilant un groupe de quelques exploitations à un bassin de collecte sur lequel il était possible de mesurer la collecte de façon relativement fiable et exhaustive (sur toutes les parcelles de tournesol des exploitations) et de récupérer les informations permettant de faire tourner les modèles (dates de semis, variétés...).

Un réseau de parcelles agricoles a ainsi été mis en place chacune des trois années du projet dans la région toulousaine (en tout sur trois ans, 312 parcelles de tournesol, représentant 1760 hectares, ont été suivies ; Tableau 2). Chaque parcelle a fait l'objet d'une collecte d'informations auprès des agriculteurs et de leurs techniciens afin d'obtenir les données d'entrée nécessaires à la mise en œuvre des modèles : réserve utile en eau du sol, date de semis, variété, dates et doses d'apports d'azote, dates et doses d'irrigation. Les parcelles ont également fait l'objet de 2 ou 3 visites au cours desquelles ont été observés les stades phénologiques, les structures de peuplement, les maladies, l'enherbement, les carences en éléments minéraux et tout autre facteur non pris en compte par les modèles de culture. Des fonctions de nuisibilité de ces facteurs limitants vis-à-vis du rendement et de la teneur en huile ont été établies sur la base des connaissances antérieures de Terres Inovia et de l'INRA afin de corriger les résultats des simulations par les modèles de culture. Les rendements ont été collectés auprès des agriculteurs, si possible à l'échelle de la parcelle, sinon à l'échelle de l'ilot cultural ou de l'exploitation selon la disponibilité. Lors de la récolte, un échantillon de graines a été prélevé sur la quasi-totalité des parcelles pour détermination de la teneur en huile, et les échantillons issus des parcelles ensemencées en tournesol oléique ont fait l'objet d'une mesure de teneur en acide oléique.

Tableau 2 : Caractéristiques des bassins d'étude pour chacune des trois années.

Année	Coopérative	Nb exploitations	Nb parcelles	Surface cumulée (ha)	Surface moyenne parcellaire (ha)	Nb parcelles moyen par exploitation
2014	Arterris	8	58	230	3,94	7
	Val de Gascogne	9	64	335	5,28	7
	Total	17	122	565	4,64	7
2015	Arterris	13	70	375	5,35	7
	Val de Gascogne	10	49	414	8,45	5
	Total	23	119	789	6,63	6
2016	Arterris	5	26	156	5,99	5
	Val de Gascogne	6	45	250	5,57	8
	Total	11	71	406	5,72	6

Le rendement moyen annuel aux normes le plus élevé est obtenu en 2016 (26.1 q/ha). Il n'est que de 21.9 q/ha en 2015 et 25.0 q/ha en 2014. La variabilité intra-annuelle du rendement est élevée : son amplitude (rendement maximal moins rendement minimal) va de 26 (2016) à 30 q/ha (2014) (Figure 2). Les rendements sont parfois très faibles (inférieurs à 10 q/ha). Cela correspond à des situations avec des dégâts de grêle ou avec une très mauvaise levée. Les rendements les plus élevés sont atteints en 2016 (jusqu'à 42.9 q/ha).

La teneur en huile aux normes moyenne (pondérée par les surfaces parcellaires) varie de 44.2 % en 2016 à 46.2 % en 2014. La variabilité intra-annuelle de ces teneurs en huile couvre une amplitude d'environ 10 points, quelle que soit l'année. Les teneurs en huile sont parfois très faibles, puisque 11 % des parcelles présentent une teneur en huile inférieure à 42%. Sur les 3 campagnes, 70 % des parcelles présentent une teneur en huile supérieure à 44 % (norme commerciale). La teneur moyenne en acide oléique (pondérée par les surfaces) est très stable d'une année à l'autre (88.3 % en 2015 et 2016 et 88.8 % en 2014). L'amplitude de la variabilité annuelle couvre 10 points en 2014 et 2015 et seulement 7 points en 2016.

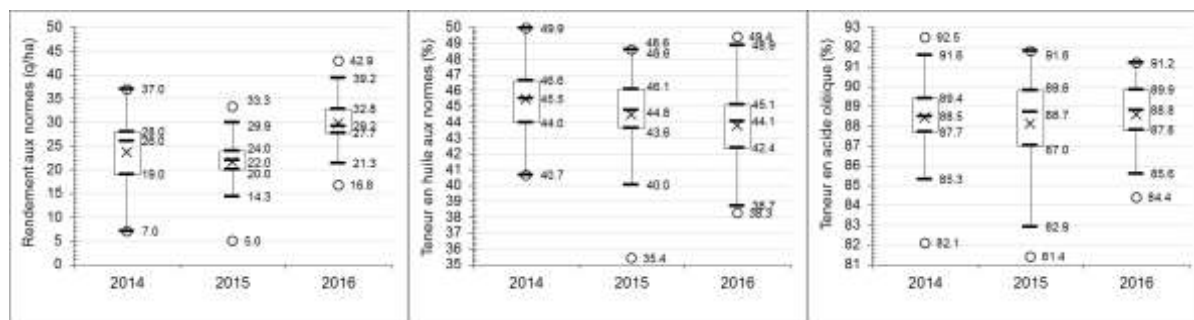


Figure 2 : Distribution des rendements, de la teneur en huile et de la teneur en acide oléique dans les réseaux de parcelles de chacune des 3 années du projet.

Chaque année, des images satellitaires ont été collectées pendant la période de culture du tournesol : en 2014, 6 images exploitables en moyenne pendant la période du cycle (principalement Landsat 8 et quelques images Deimos) ; en 2015, 13 images exploitables en moyenne (Spot 5-Take 5) ; en 2016, 5 images exploitables en moyenne (Sentinel-2A).

Pour chaque image satisfaisante, l'indice de végétation NDVI a été calculé sur chaque parcelle agricole. Nous disposons ainsi d'un profil de valeurs de NDVI pour 102 parcelles en 2014, 119 parcelles en 2015 et 71 parcelles en 2016 (Figure 3). La valeur de NDVI retenue par parcelle correspond à la moyenne des valeurs de NDVI de l'ensemble des pixels couvrant cette parcelle, exceptés ceux des bords de parcelles potentiellement partiellement en dehors de la parcelle. On ne prend donc pas en compte l'hétérogénéité intra-parcellaire. Ce choix est justifié par le fait que les valeurs de rendement, de teneur en huile et en acide oléique dont nous disposons ont été acquises à l'échelle de la parcelle entière et non pas du pixel.

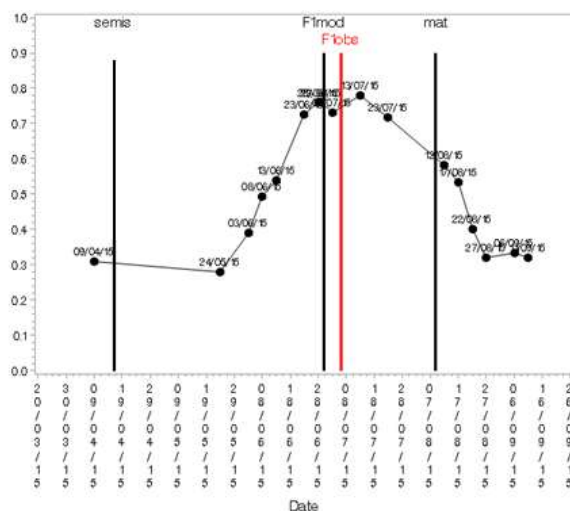


Figure 3 : Exemple de profil de NDVI sur une parcelle suivie dans le cadre du projet en 2015 (F1 mod et F1 mat : dates de début floraison et de maturité physiologique calculée à l'aide d'un modèle de somme de températures issu de SUNFLO ; F1obs : date de début floraison estimée à l'aide des observations au sol).

En 2014 et 2015, les GAI ont été obtenus à l'aide de BVNET recodé sous Orfeo Toolbox. Cet outil n'est pas encore paramétré pour le capteur du satellite Sentinel-2A et ne peut donc pas être utilisé pour les images 2016. Les valeurs de GAI satellitaire de 2016 ont donc été obtenues par la méthode indirecte suivante : (i) calcul de l'équation de la régression linéaire entre les valeurs de NDVI issues de Landsat 8 et de Sentinel-2A sur les parcelles du bassin pour lesquelles les deux types d'images étaient disponibles à des dates voisines en 2016, (ii) conversion des valeurs de NDVI « Sentinel-2A » en équivalent NDVI « Landsat 8 » grâce à cette équation pour toutes les parcelles de l'étude et toutes les images, (iii) calcul de l'équation de la régression exponentielle entre le NDVI « Landsat 8 » et le GAI

BVNET « Landsat 8 » sur les données 2015 du projet et enfin, (iv) estimation d'un équivalent GAI BVNET « Sentinel-2A » à partir du NDVI « Sentinel-2A » sur les données 2016 grâce à la régression exponentielle précédente.

4. Modèles de culture et adaptations pour le projet

Dans le cadre du projet, il était prévu d'utiliser deux modèles de cultures : i) SAFY (CESBIO) et SUNFLO (INRA – Terres Inovia). SAFY est un modèle agro-météorologique générique simple basé sur l'assimilation des GAI obtenus par télédétection satellitaire (Duchemin *et al.*, 2008). Des paramètres d'efficacité de conversion du rayonnement intercepté sont calibrés et certains facteurs limitants sont intégrés au modèle. SUNFLO est un modèle de culture spécifique au tournesol issu de la collaboration entre l'INRA (UMR AGIR) et Terres Inovia (Casadebaig *et al.*, 2011). Sa principale spécificité par rapport à des modèles plus génériques est qu'une attention particulière a été apportée à la simulation de l'indice foliaire, facteur explicatif majeur de la production de la culture du tournesol. Une autre spécificité de SUNFLO est qu'il a été construit avec le souci de ne mobiliser que des données d'entrée relativement faciles d'accès. Enfin, il tient compte de spécificités variétales mesurables.

Avec une initialisation unique à l'implantation de la culture, la précision des modèles de culture comme SUNFLO se situe en général aux alentours de 15 à 20%. Nous faisons l'hypothèse que l'utilisation des techniques d'assimilation de données observées dans les modèles au cours du cycle de croissance de la culture pourrait permettre de franchir cette limite. Un module d'assimilation de données de GAI satellitaire dans SUNFLO a donc été construit et l'amélioration qu'il apporte sur la prédiction du rendement évaluée. Par ailleurs, l'intérêt des coopératives pour la prévision de la teneur en acide oléique nécessite que le modèle SUNFLO soit complété par un module qui permette de simuler cette variable.

Pour des raisons de charge en travail, il n'a pas été possible de travailler avec SAFY, ainsi que sur l'assimilation de données de GAI dans SUNFLO pour la prédiction de la teneur en huile. Néanmoins pour ce second point, les avancées réalisées sur le modèle d'assimilation de données pour le rendement pourront être remobilisées pour la teneur en huile.

4.1 Module d'assimilation de données de GAI issues de la télédétection satellitaire dans SUNFLO

Un inventaire des méthodes d'assimilation de données observées existantes a d'abord été réalisé à l'aide d'une étude bibliographique. Ces méthodes ont ensuite été évaluées en fonction des objectifs du projet afin d'être en mesure de sélectionner la ou les plus adaptées. C'est la méthode du « filtre de Kalman d'ensemble » qui a été retenue (Kalman, 1960).

Une étude de sensibilité de deux variables de sortie du modèle (indice foliaire et rendement) à l'incertitude sur les valeurs de 37 paramètres et variables d'entrée du modèle a été réalisée. Deux méthodes d'exploitation des données et d'analyse des résultats ont été utilisées successivement : la méthode « Morris » pour une première sélection grossière des paramètres les plus influents, puis la méthode « Extended Fast », plus poussée, sur les paramètres identifiés avec la méthode « Morris », pour une meilleure évaluation des indices de sensibilité.

La méthode d'assimilation retenue a été implémentée sur la plateforme de modélisation de l'INRA (RECORD). Dans cette approche, il s'agit, lors de chaque « arrivée » d'une donnée observée, de modifier les valeurs d'indice foliaire simulées par le modèle à l'aide des valeurs observées afin de réduire l'écart (observé – simulé) et de répercuter ces modifications sur une autre variable clé : la biomasse produite, très explicative de la production. On considère une incertitude sur les variables d'entrée les plus influentes (identifiées lors de l'analyse de sensibilité) dans le processus d'assimilation

de données par filtre de Kalman d'ensemble. Les algorithmes d'assimilation implémentés sous Record sont opérationnels. Ces différentes méthodes ont permis d'améliorer nettement la prédiction du rendement par rapport à la version native du modèle SUNFLO (Figure 4).

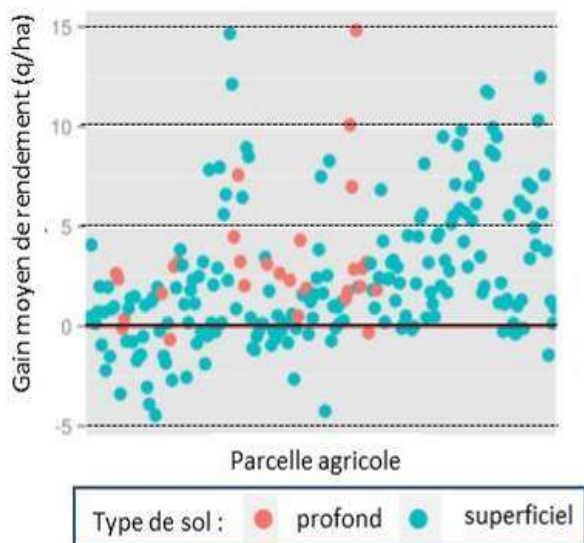


Figure 4 : Exemple d'amélioration de la prédiction du rendement grâce à une des méthodes d'assimilation testées, par rapport à la forme native de SUNFLO (mean improvement : amélioration moyenne par situation simulée en q/ha ; un point = 1 parcelle agricole ; données 2015 et 2016 du proiet).

4.2 Module de simulation de la teneur en acide oléique dans SUNFLO

La démarche de modélisation s'est appuyée sur une analyse bibliographique qui a mis en évidence que la température minimale ou nocturne moyenne pendant une période d'une vingtaine de jours pendant la phase de remplissage des graines pouvait permettre de prédire la teneur en acide oléique (Merrien et al., 2005, Izquierdo et al., 2008). Les différentes hypothèses issues de ces publications ont été testées sur une base de données issue d'essais d'évaluation variétale de plein champ conduits par Terres Inovia de 2005 à 2015. Vingt-sept variétés présentes dans au moins quinze essais chacune ont été retenues. Pour chaque essai, les variables disponibles sont le nom de la commune d'implantation, la date de semis, le rendement, la teneur en huile et la teneur en acide oléique. Les données météorologiques utilisées sont issues de la station météo de référence la plus proche de la commune d'implantation car il n'y avait pas de station sur les essais. La distance de la station météo à l'essai peut aller de 3 à 30 km environ. Les dates de début de floraison (F1) ont été calculées sur les essais pour lesquels elles n'avaient pas été observées, à l'aide d'un modèle « somme de températures » inspiré de SUNFLO. Nous avons ensuite calculé les variables explicatives mentionnées dans les deux publications mentionnées précédemment.

Il n'est pas apparu de relation suffisamment étroite entre la teneur en acide oléique et les variables explicatives citées dans la littérature. De nombreuses autres variables explicatives (quantitatives et qualitatives) ont été testées une par une ou en combinaison, mais cela n'a pas apporté d'amélioration.

Nous ne sommes donc pas en mesure de proposer un formalisme et un paramétrage robuste pour un module d'estimation de la teneur en acide oléique. Ce travail a permis de mettre en évidence que les relations issues de la littérature, et établies dans des conditions optimales de contrôle des facteurs du milieu, ne sont pas toujours immédiatement utilisables dans des modèles destinés à fonctionner avec les données produites dans des situations moins contrôlées et non instrumentées.

5. Modèles corrélatifs simples

Dans un contexte de changement d'échelle et pour l'application à la prévision de la collecte, nous faisons l'hypothèse que d'autres types de modèles (de type corrélatif) pourraient être plus adaptés que

les modèles de culture dynamiques. Il s'agit de modèles à une ou deux variables explicatives issues de la télédétection : indice de végétation (NDVI) ou variables agronomiques. Pour des raisons de charge en travail, seule la prévision du rendement a été abordée dans le cadre du projet.

Compte tenu des connaissances antérieures de Terres Inovia, deux variables agronomiques potentiellement explicatives du rendement ont été sélectionnées : le GAI maximal (GAI-max), normalement atteint au moment du début de la floraison, et la durée de GAI du début de la floraison à la date de réalisation de l'estimation du rendement (le 10 août dans cette étude : GAD-10/08). Ces variables ont été utilisées seules (modèles à une seule variable explicative : soit GAI-max, soit GAD-10/08) et combinées (modèle à deux variables explicatives : GAI-max et GAD-10/08). La durée de GAI du début de la floraison à la maturité physiologique (GAD-mat), qui est l'indicateur utilisé dans les démarches de diagnostic *a posteriori*, a également été utilisée. Cette variable ne présente pas d'intérêt dans une démarche de prévision, mais permet de chiffrer l'effet de la perte d'information liée à l'évolution du GAI du 10 août à maturité sur la prévision du rendement.

Des modèles corrélatifs simples utilisant directement comme variable d'entrée un indice de végétation (NDVI) ont aussi été étudiés. Cette voie présente l'avantage de ne pas nécessiter de disposer de modèles d'estimation du GAI à partir d'images satellitaires qui ne sont pas toujours très bien calés. À l'image des choix réalisés sur le GAI, nous avons utilisé le NDVI maximal (NDVI-max), la durée de NDVI du début de la floraison au 10 août (DNDVI-10/08) et de début de la floraison à maturité (DNDVI-mat).

Les variables explicatives ont été calculées pour chaque parcelle des bassins d'études pour laquelle nous disposons d'un profil de NDVI et de GAI. Dans un premier lieu, un calcul journalier du GAI (resp. NDVI) a été réalisé de la date de semis à la date de maturité en réalisant une interpolation linéaire entre les valeurs obtenues à chaque date de disponibilité des images satellitaires (Figure 5). Les variables explicatives ont été calculées de la façon suivante :

- Le GAI-max (resp. NDVI-max) correspond à la valeur maximale du GAI (resp. NDVI) pendant la période de 20 jours centrée sur la date de début floraison. La plupart du temps, cette valeur correspond à une date de disponibilité d'image (Figure 5A). Lorsque ce n'est pas le cas (le plus souvent lorsque la valeur maximale observée est plus tardive que la période retenue), la valeur interpolée la plus élevée dans la période de 20 jours est retenue (Figure 5B). Ceci permet d'écartier des valeurs de GAI-max qui pourraient être trop tardives par rapport au cycle du tournesol (défauts sur les images, enherbement tardif...).
- Le GAD-mat (resp. DNDVI-mat) correspond à la somme des GAI (resp. NDVI) journaliers de la date de début de la floraison à la date de maturité physiologique (Figure 5D).
- Le GAD-10/08 (resp. DNDVI-10/08) correspond à la somme des GAI (resp. NDVI) journaliers de la date de début de la floraison à la date de maturité physiologique calculée avec les données disponibles le 10 août. Nous plaçant en condition de prévision, nous faisons comme si nous ne disposions pas des images acquises après le 10 août. De plus, il y a rarement une image disponible précisément le 10 août. Pour obtenir une valeur le 10 août par interpolation linéaire, nous avons fait l'hypothèse que le GAI à maturité physiologique était de 0.7 (resp. 0.46 pour le NDVI). Ces valeurs correspondent aux moyennes des valeurs de NDVI et GAI obtenues à maturité physiologique par interpolation linéaire avec les profils de GAI et NDVI complets (non tronqués au 10 août). La valeur de GAI (resp. NDVI) au 10 août est donc obtenue par interpolation linéaire entre la valeur calculée à partir de la dernière image avant le 10 août et la valeur standard retenue à maturité physiologique (Figure 5C). Dans ce cas, la date de maturité physiologique est calculée sur la base de somme de températures depuis le semis, en utilisant les données météo de l'année, jusqu'au 10 août, et des données moyennes au-delà.

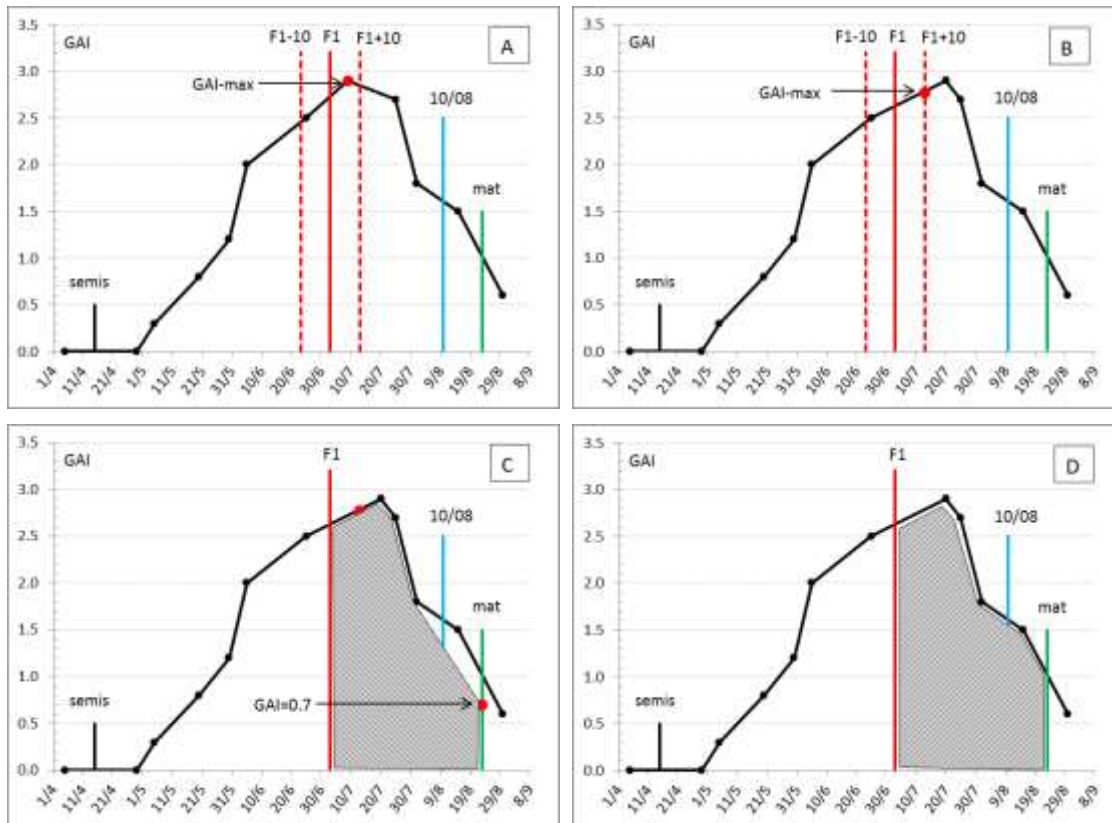


Figure 5 : Illustration des méthodes de calcul du GAI-max (A : cas où le GAI maximal observé est dans la période F1-10 j à F1+10j ; B : cas où le GAI maximal observé n'est pas dans cette période), du GAD-10/08 (C : GAD = aire grisée) et du GAD-mat (D : aire grisée). F1 : date de début floraison ; mat : date de maturité physiologique. Données non réelles.

Toutes les variables précédentes ont été déterminées sur la base des valeurs de GAI brutes, non ajustées selon un modèle mathématique. Or, il arrive couramment que les profils bruts aient une allure plus ou moins irrégulière. Un « lissage » a donc été réalisé par ajustement des points à une fonction double sigmoïde (ajustement réalisé à l'aide d'un module du modèle SAFY) pour voir si cela permettait d'améliorer la prédiction (comme illustré sur la Figure 6). Cette approche n'a été conduite que sur les parcelles de 2014 et 2015. Les variables explicatives suivantes ont été calculées : GAI maximal selon la fonction d'ajustement (GAI-max_aj), GAD pendant la période allant du début de la floraison à la maturité physiologique (GAD-mat-aj) et GAD selon la fonction d'ajustement pendant la période allant du début de la floraison au 10 août (GAD-10/08-aj).

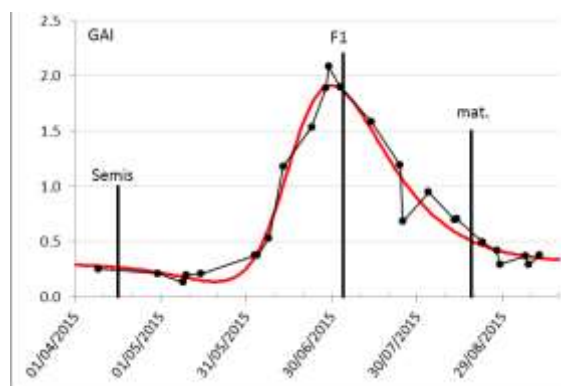


Figure 6 : Profil d'évolution du GAI pour la parcelle 15-50-01-00 du projet (en rouge : ajustement selon une fonction double sigmoïde).

Quatre types de formalismes ont été étudiés pour mettre en relation ces variables et le rendement : modèles linéaire, linéaire + plateau, quadratique et quadratique + plateau.

Chaque type de modèle a été paramétré pour chacune des trois années, pour chaque couple d'années (2014 + 2015, 2014 + 2016 et 2015 + 2016) et pour les trois années ensemble (Figure 7), afin de pouvoir ensuite appliquer chaque jeu de paramètres aux données non utilisées pour les calculer afin de réaliser une évaluation indépendante de la qualité prédictive des modèles (par exemple : application des paramètres calculés sur les données de 2014 et 2015 aux données de 2016).

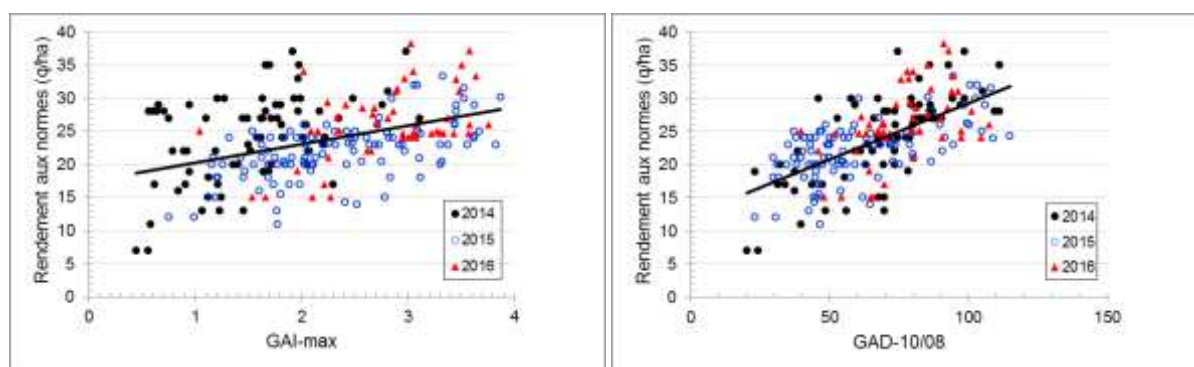


Figure 7 : Exemples de relations entre le rendement aux normes et les variables explicatives de la famille du GAI (GAI-max à gauche et GAD-10/08 à droite) pour les 3 années ensemble.

Les résultats des simulations indiquent que les modèles qui présentent les meilleures qualités de prédiction sont la plupart du temps les modèles ayant comme variable explicative, le GAD-10/08 (Tableaux 3 et 4).

Tableau 3 : Meilleurs modèles corrélatifs obtenus (R^2 : coefficient de détermination ; RMSE : root mean square error ; RRMSE : relative root mean square error).

Meilleurs modèles	R^2	RMSE (q/ha)	RRMSE (%)
Rendement = 0.144 GAD _{10/08} + 13.92	0.368	4.00	17.1
Rendement = 2.046 GAI _{max} + 0.089 GAD _{10/08} + 12.44	0.402	3.90	16.7

Tableau 4 : Résultats de l'évaluation de deux types de modèles (linéaire et quadratique) obtenus avec différentes variables explicatives (RMSEP : root mean square error of prediction ; paramétrage avec 2 années et évaluation sur la troisième année).

Années paramétrage	Année évaluation	Modèle	rdt moyen observé	Variables explicatives														
				GAI-max			GAD-mat			GAD-10/08			GAI-max et GAD-mat			GAI-max et GAD-10/08		
				rdt moyen simulé	Biais	RMS EP	rdt moyen simulé	Biais	RMS EP	rdt moyen simulé	Biais	RMS EP	rdt moyen simulé	Biais	RMS EP	rdt moyen simulé	Biais	RMS EP
2014 + 2015	2016	Linéaire	26.1	24.6	-1.5	5.07	27.3	1.1	5.23	24.9	-1.2	4.80	27.8	1.6	5.17	25.0	-1.2	4.75
		Quadratique		24.6	-1.5	5.13	27.1	1.0	5.11	25.0	-1.2	4.77	27.6	1.5	5.25	24.8	-1.3	5.23
2014 + 2016	2015	Linéaire	22.1	25.9	3.8	5.18	20.9	-1.2	3.58	22.4	0.4	4.05	21.9	-0.2	3.33	23.1	1.1	4.30
		Quadratique		26.0	3.9	5.26	19.9	-2.2	4.55	22.3	0.2	4.09	20.5	-1.5	4.43	22.9	0.8	4.41
2015 + 2016	2014	Linéaire	24.2	19.5	-4.7	7.52	24.4	0.2	5.16	24.0	-0.2	4.87	21.9	-2.3	5.66	21.8	-2.4	5.53
		Quadratique		19.3	-4.9	7.62	24.5	0.3	5.12	24.0	-0.2	4.85	22.1	-2.1	5.59	21.3	-2.9	5.81

De plus, avec ces modèles, le sens de variation des rendements moyens annuels observés est très bien conservé au niveau des rendements moyens simulés lorsque deux années sont utilisées pour le

paramétrage. Dans ces conditions, le biais est toujours inférieur ou égal à 1.2 q/ha. Le GAD-mat ne donne pas de meilleurs résultats que le GAD-18/08. Cela signifie qu'il n'y a pas de perte d'information consécutive à la non prise en compte des GAI postérieurs au 10 août. Les modèles avec GAI-max seul ou GAI-max combiné à GAD-10/08 ou GAI-mat sont moins performants. Il en est de même lorsque l'on utilise les variables explicatives de la famille du NDVI. Il n'apparaît pas d'amélioration de la qualité de prédiction lorsqu'on utilise les GAI ajustés selon la fonction double sigmoïde.

6. Comparaison des résultats des simulations réalisées à l'échelle du bassin d'étude avec les différents types de modèles

Le modèle corrélatif le plus performant (modèle linéaire ne mobilisant que le GAD-10/08 et paramétré sur 2 années) a été comparé au modèle SUNFLO avec assimilation de données observées de GAI satellitaire. Ces dernières sont les données brutes (non ajustées à une fonction double sigmoïde). Les résultats des simulations avec SAFY n'ont pas pu être mis à disposition pendant le temps du projet.

La base de données a été réduite aux parcelles permettant d'utiliser les deux types de modèles (disponibilité d'images satellitaires en nombre suffisant, de la date de semis, de la conduite de fertilisation azotée et hydrique). Cela réduit la base principale à 65 parcelles en 2014, 98 parcelles en 2015 et 50 parcelles en 2016.

La qualité de prédiction du modèle SUNFLO est très inférieure à celle du modèle corrélatif. Le biais va de -3.8 à -10 q/ha selon les années (contre -0.4 à 1.2 pour le modèle corrélatif) et la RMSEP est comprise entre 7.6 et 11.5 q/ha (contre 4.3 à 4.9 pour le modèle corrélatif).

Si la simulation des rendements moyens annuels à l'aide du modèle corrélatif est satisfaisante, la distribution des valeurs simulées peut parfois être très différente de celle des valeurs observées (Figure 8). C'est en particulier le cas en 2014, année pour laquelle les valeurs simulées sont nettement moins variables que les valeurs observées. C'est également le cas en 2016, mais dans une moindre mesure.

Les relations entre les rendements observés et les rendements estimés avec le modèle SUNFLO montrent que la forte surestimation moyenne constatée avec ce modèle résulte d'une surestimation sur une très forte proportion des parcelles en 2014 et 2015 (Figure 8). En 2016, le pourcentage de parcelles avec surestimation du rendement est moins élevé mais toujours nettement supérieur à 50%. Le modèle corrélatif a quant à lui tendance à surestimer les rendements inférieurs à 25 q/ha et à sous-estimer les autres, sauf en 2015.

Plusieurs raisons potentielles peuvent être avancées pour expliquer la supériorité des modèles corrélatifs :

- Ces modèles ont l'avantage d'être construits sur la base des rendements atteints dans les milieux étudiés pendant deux années. SUNFLO n'utilise pas cette connaissance. Il doit se « contenter » d'informations sur le sol, l'itinéraire technique, la météo et la cinétique d'évolution du GAI.
- SUNFLO est un modèle qui, à l'origine, ne prend pas en compte les facteurs limitants autres que climatiques et nutritionnels (eau et azote). L'hypothèse que nous avons formulée est que les autres facteurs limitants (essentiellement biotiques) provoquent souvent une réduction du GAI qui pourrait être prise en compte par SUNFLO par assimilation de données de GAI. Il est probable que ces facteurs limitants ne se traduisent pas tous par une réduction du GAI ou pas uniquement par une réduction du GAI (effet sur d'autres paramètres du modèle).
- SUNFLO a été paramétré avec des indices foliaires acquis par voie allométrique. Nous avons réalisé l'assimilation de valeurs de GAI issues de la télédétection satellitaire. Ces deux variables ne sont pas équivalentes. L'une des tâches de ce projet a pour objectif de paramétrer une relation entre ces deux grandeurs. Nous n'avons pas pu aboutir dans le temps du projet. Nous

disposons toutefois des données pour le faire. Une suite de ce projet sera de calculer des valeurs d'indice foliaire allométrique équivalentes au GAI satellitaire et d'assimiler ces nouvelles valeurs dans SUNFLO.

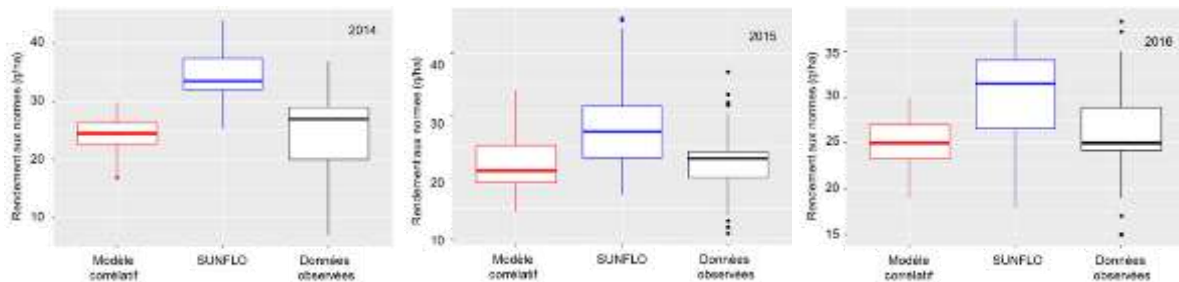


Figure 8 : Distribution des rendements estimés avec le meilleur modèle corrélatif et SUNFLO avec assimilation de GAI satellitaire et des rendements observés sur le terrain pour chacune des trois années de l'étude.

7. Construction d'un simulateur

Il n'a pas été possible d'aller jusqu'à la construction d'un simulateur opérationnel. Néanmoins deux de ses constituants ont été produits : une chaîne de traitement automatique des images satellitaires et un démonstrateur.

7.1 Chaîne de traitement automatique des images satellitaires

En 2015, les partenaires de trois projets (projet CASDAR n°1311, projet FUI Maiséo, projet ANR Rue des Sols) ont pris la décision de mutualiser leurs forces pour construire une plateforme informatique et des algorithmes qui permettent automatiquement de traiter les données d'images satellitaires. Cette plateforme est interfacée avec le portail Theia (www.theia-land.fr/) dont la vocation est de mettre « à disposition de la communauté scientifique et des politiques publiques une vaste panoplie d'images à différentes échelles, des méthodes et des services ». Un cahier des charges a tout d'abord été élaboré, puis la plateforme a été construite.

La plateforme informatique mise en place par le CESBIO réalise les opérations suivantes : (i) téléchargement des images optiques satellitaires (images de réflectance de surface) à partir du portail Theia, (ii) génération des indices de végétation (par exemple NDVI) et des paramètres biophysiques (par exemple GAI) en les harmonisant entre capteurs, (iii) extraction de ces paramètres biophysiques par pixel ou par entité géographique (parcelles) et (iv) constitution de séries temporelles continues par interpolation entre 2 images (fonction double logistique).

A terme, cette plateforme sera en mesure de réaliser les tâches complémentaires suivantes : détection des cultures d'été, génération des cartes d'occupation du sol (quelle culture dans telle parcelle avec la probabilité associée) et estimation de la date de levée des cultures.

7.2 Démonstrateur

Ce démonstrateur a pour objectif d'accélérer le potentiel de transfert vers les partenaires opérationnels futurs utilisateurs de l'outil.

Le démonstrateur est basé sur deux blocs. Le premier est un environnement « arrière-plan » ou communément appelé « backend ». Il est constitué d'un ensemble d'outils de traitement des données d'images satellitaires et de modèles statistiques (ajustement des GAI bruts à une fonction double sigmoïde...) et physiques (SAFY).

Le deuxième bloc est le « frontend » constitué d'une application web qui permet un accès fluide et interactif aux données. Le portail web montre la localisation géographique des parcelles (Figure 9). Lorsque l'utilisateur choisit une parcelle en cliquant sur la carte, la série temporelle de GAI satellitaire (sans interpolation), le résultat de l'interpolation et l'estimation par le modèle SAFY de cette même variable s'affichent. Une moyenne sur l'ensemble des parcelles du bassin est aussi fournie. L'interface permet également à l'utilisateur de procéder à une inspection plus précise des dates clés calculées à partir des données satellitaires interpolées (date de début de croissance végétative ...) et d'une estimation du rendement. L'utilisateur peut aussi télécharger les données de la parcelle choisie sous forme d'un fichier texte (.csv) directement à partir de l'application.

Ce démonstrateur n'est pas mis en ligne actuellement. L'ensemble des scripts sont disponibles sous forme de livrables.



Figure 9 : Capture d'écran du portail web du démonstrateur.

Conclusion

L'approche de prévision de collecte en quantité et en qualité à l'échelle d'un territoire mobilisant les modèles et la télédétection satellitaire est originale. Rares sont les travaux réalisés ou en cours dans ce domaine. Une collaboration pluri-disciplinaire a pu être mise en place associant des compétences en mathématiques appliquées, télédétection et agronomie et couplant modélisation et expérimentation. Une base de données pluriannuelle originale a été constituée à l'échelle de bassins de collecte d'étude. L'association de coopératives (Arterris et Terres de Gascogne) a permis de confirmer leur forte attente vis-à-vis de la prévision de collecte que ce soit pour la logistique de collecte ou pour la mise en marché.

Le projet a abordé des problématiques en amont de la prévision de collecte, mais qui sont indispensables pour que la variable GAI puisse être estimée par télédétection satellitaire sur les parcelles de tournesol d'un bassin : méthode d'estimation rapide du GAI au sol afin de pouvoir réduire le coût des réseaux « vérité terrain », mise en place de réseau « vérité terrain » pour pouvoir évaluer les méthodes d'estimation du GAI à partir des capteurs satellitaires, cartes d'occupation des sols, prévision de la date de levée...

La comparaison des modèles à l'échelle du bassin de collecte a montré que les modèles corrélatifs à une ou deux variables biophysiques ou agronomiques pouvaient présenter de meilleures performances qu'un modèle de culture de type SUNFLO, même lorsqu'il lui est adjoint un module d'assimilation de données satellitaires. Ces modèles corrélatifs requièrent toutefois de disposer de nombreuses données pour leur paramétrage. Il reste à préciser si ce paramétrage est très dépendant des conditions de

culture (région, systèmes de culture). Ces modèles doivent aussi faire la preuve qu'ils restent meilleurs lorsque l'on souhaite descendre au niveau de la prédiction de rendement infrarégional, par exemple par catégorie de sol, par type de conduite de culture...

L'expertise et les connaissances acquises au cours du projet ont permis de repenser le schéma d'organisation d'un futur outil de prévision de collecte.

Jugeant ce projet stratégique, les partenaires ont décidé de poursuivre le travail au-delà du projet CASDAR. Ils ont travaillé sur fonds propres en 2017 et ont répondu avec succès à l'appel à projets « ressourcement scientifique » 2017 de l'Institut Carnot « Plant2Pro ». Ce projet a commencé en mars 2018, il a pour objectif d'approfondir le travail prévu dans le cadre du présent projet CASDAR :

- Estimation du GAI à partir des données Sentinel-2A à l'aide de BVNET,
- Construction d'un modèle de conversion du GAI satellitaire en indice foliaire allométrique compatible avec SUNFLO,
- Amélioration de la capacité de SAFY à prédire le rendement du tournesol à partir de la biomasse par la prise en compte du stress hydrique,
- Développement et test de modèles statistiques de prévision de la qualité,
- Assimilation des informations satellitaires dans SUNFLO pour la prédiction de la teneur en huile,
- Etude de l'effet sur la qualité des prédictions du nombre d'images disponibles, de la date à laquelle la prédiction est faite, du choix du climat post-prédiction,
- Mise en place d'une chaîne de simulation complète à des fins de démonstration de faisabilité.

Ce projet a reçu l'agrément du pôle de compétitivité Agri sud-ouest Innovation.

Références bibliographiques

Barbottin A., Makowski D., Le Bail M., Jeuffroy M.-H., Bouchard C., Barrier C., 2008. Comparison of models and indicators for categorizing soft wheat fields according to their grain protein contents. *Europ. J. Agronomy* 29, 175–183.

Baret F., Camacho F., Fang H., Garrigues S., Gobron N., Lang M., Lacaze R., LeBlanc S., Meroni M., Martinez B., Nilson T., Pinty B., Pisek J., Sonnentag O., Verger A., Welles J., Weiss M., Widlowski J.L., Schaepman-Strub G., Roman M., Nickeson J., 2014. Global Leaf Area Index Product Validation Good Practices. Version 2.0; Fernandes R., Plummer S., Nightingale J., Eds.; *Satellite-Derived Land Product Validation, Land Product Validation Subgroup (WGCV/CEOS)*: Zurich, Switzerland, 2014; p. 76.

Belaqziz S., Khabba S., Er-Raki S., Jarlan L., Le Page M., Kharrou M.H., Adnani M.E., Chehbouni A., 2013. A new irrigation priority index based on remote sensing data for assessing the networks irrigation scheduling. *Agricultural Water Management*, 119, 1-9.

Casadebaig P., Guilioni L., Lecoeur J., Christophe A., Champolivier L., Debaeke P., 2011. SUNFLO, a model to simulate genotype-specific performance of the sunflower crop in contrasting environments. *Agricultural and forest meteorology* 151, 2, 63-178.

Champolivier L., Debaeke P., Thibierge J., Dejoux J.-F., Berger F., Casadebaig P., Ledoux S., Ludot M., Jouffret P., Lecomte L., Merrien A., Mestries E., Thiard J., Vogrincic C., Noël M., Caumes E., Edeline T., Provot M., 2011. Construire des stratégies de production adaptées aux débouchés à l'échelle du bassin de collecte. Actes du colloque CIAG « Tournesol et Agriculture Durable », Toulouse, 9 Juin 2011, *Innovations Agronomiques* 14, pp 39-57.

Champolivier L., Debaeke P., Thiard J., Thibierge J., 2012. An evaluation of sunflower production strategies in a supplying area of an agricultural cooperative using the simulator COLLECTO. *Proc. 18th Int. Sunflower Conf.*, Mar del Plata (Argentina), 26 Feb- 01 Mar 2012, pp. 742-747.

- Claverie M., Demarez V., Duchemin B., Hagolle O., Ducrot D., Marais-Sicre C., Dejoux J.F., Huc M., Keravec P., Béziat P., Fieuzal R., Ceschia E., Dedieu G., 2012. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment* 124, 844–857.
- Doré T., Clermont-Dauphin C., Crozat Y., David C., Jeuffroy M.-H., Loyce C., Makowski D., Malezieux E., Meynard J.-M., Valantin-Morison M., 2008. Methodological progress in on-farm regional agronomic diagnosis. A review. *Agronomy for Sustainable Development* 28, 151-161.
- Dorigo W.A., Zurita-Milla R., de Wit A.J.W., Brazile J., Singh R., Schaepman M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modelling. *International Journal of Applied Earth Observation and Geoinformation* 9, 165–193.
- Duchemin B., Maisongrande P., Boulet G., 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, 876–892.
- Hall A. J., Feoli C., Ingaramo J., Balzarini M., 2013. Gaps between farmer and attainable yields across rainfed sunflower growing regions of Argentina. *Field Crops Research*, 143, 119-129.
- Izquierdo N.G., Aguirrezabal L.A.N., 2008. Genetic variability in the response of fatty acid composition to minimum night temperature during grain filling in sunflower. *Field Crops research*, 106, 116-125.
- Jagtap S.S., Jones J.W., 2002. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystems and Environment* 93, 73–85.
- Kalman R.E., 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME*, 82 D.
- Launay M., Guérif M., 2003. Ability for a model to predict crop production variability at the regional scale: an evaluation for sugar beet. *Agronomie* 23, 135–146.
- Launay M., Guérif M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, Ecosystems and Environment* 111, 321–339.
- Le Bail M., Jeuffroy M.-H., Bouchard C., Barbottin A., 2005. Is it possible to forecast the grain quality and yield of different varieties of winter wheat from Minolta SPAD meter measurements? *Europ. J. Agronomy* 23, 379–391.
- Marais Sicre C., Inglada J., Fieuzal R., Baup F., Valero S., Cros J., Huc M., Demarez V., 2016. Early detection of summer crops using high spatial resolution optical image time series. *Remote Sensing*, 8(7), 591.
- Mercau J.L., Sadras V.O., Satorre E.H., Messina C., Balbi C., Uribelarrea M., Hall A. J., 2001. On-farm assessment of regional and seasonal variation in sunflower yield in Argentina. *Agricultural Systems* 67, 83-103.
- Merrien A., Pouzet A., Krouti M., Dechambre J., Garnon V., 2005. Contribution à l'étude de l'effet des températures basses sur la composition en acide gras de l'huile des akènes de tournesol (oléique et classique). *OCL*, 12 (5-6), 455-458
- Mkhabela M.S., Bullock P., Raj S., Wang S., Yang, Y., 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agricultural and Forest Meteorology*, 151, 385-393.
- Pena-Barragan J.M., Lopez-Granados F., Jurado-Exposito M., Garcia-Torres L., 2010. Sunflower yield related to multi-temporal aerial photography, land elevation and weed infestation. *Precision Agric* 11, 568–585.
- Pereyra-Irujo G.A., Izquierdo N.G., Covi M., Nolasco S.M., Quiroz F., Aguirrezabal L.A.N., 2009. Variability in sunflower oil quality for biodiesel production: a simulation study. *Biomass & Bioenergy* 33, 459-468.
- Potgieter A.B., Hammer G.L., Doherty A., de Voil P., 2005. A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia. *Agricultural and Forest Meteorology* 132, 143–153.
- Ren J., Chen Z., Zhou Q., Tang H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation*, 10, 403-413.

Supit I., Wagner W., 1999. Analysis of yield, sowing and flowering dates of barley of field survey results in Spain. *Agricultural Systems* 59, 107-122.

de Wit A.J.W., van Diepen C.A., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology* 146, 38–56.

Cet article est publié sous la licence Creative Commons (CC BY-NC-ND 3.0)



<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « *Innovations Agronomiques* », la date de sa publication, et son URL)