Interreg
Sudoe
EUROPEAN UNION
climalert
European Regional Development Fund

CHAPTER 3.2: Crop biomass and yield estimation based on satellite images.

INDEX

1. Introduction.

Estimation of agricultural production (crop biomass and yield) is mostly based on the use of crop modelling techniques. Since the late 1960's, when crop modelling and agro ecosystem modelling was initiated, many agro ecosystem models have been developed at different scales, for a great variety of crop and plants and for many applications. Nowadays crop simulation models not only incorporate biological and physiological knowledge of plants, but also model the interactions between plants and their environment

There are different types of crop models which are characterized by their level of complexity and their ability to exploit different levels of information at different spatial scales. Remote sensing is used as a tool to feed these models and compensate for the lack of local information and simplifications of these models (Baghdadi and Zibri, 2016).

The use of crop simulation models has increased significantly in recent years, as it allows the determination of crop production potentials in a given agroclimatic area. Information is needed on the factors that define and limit crop growth and development, and which are often integrated into models as independent variables. The type of simulation model to be used depends on the objective to be achieved, the level of complexity of the process under study, and the nature and availability of data required as model inputs. The information and data requirements associated with crop simulation models in the field of crop biomass and yield estimation depend on the degree of detail to be achieved. Among the components that make up the crop simulation models are those related to the crop itself, climate, soil, management, and socio-economic factors.

The development that computational methods have experienced in recent years has provided a major impetus in the generalization of the use of models that integrate the effect of agricultural inputs on crop yields. In this way a large number of computer applications have emerged that integrate simulation models in the field of characterization of crop development. Among the models developed leveraging the power of computing techniques, a distinction can be drawn among them depending on their degree of complexity.

Among the "complex" models it is possible to find the so-called "ecophysiological" models that simulate the major plant processes in a mechanistic way (SUCROS (Penning de Vries and Van Laar, 1982), WOFOST (Van Diepen et al., 1989), etc.)

In this category of "complex" models it is also possible to find the "agronomical" models that describe the effect of agricultural practices on a large number of crops (CERES-maize (Jones and Kiniry, 1986), Crop-Syst (Stockle et al., 1994), EPIC (Williams et al., 1984), GOSSYM (Reddy et al., 1997), STICS (Brisson et al, 2003)). In these models some physiological processes have been simplified. They mostly use different versions of the Monteith equation that simulates the daily increase in biomass from global incident radiation and from three efficiency factors (climatic, light absorption and radiation use) (Monteith, 1972). These versions of Monteith equation are integrated with a set of a large number of other equations.

Apart from these "complex" models there are simple ones that just calculate the biomass from Monteith equation. The third category of crop models (semi-empirical approaches) combines the Monteith equation with some major processes (plant development, water dynamics in the soil, etc.). Nevertheless they have a restricted number of formalism and parameters compared with the "complex" crop models (AquaCrop (Steduto et al., 2009), SAFY (Duchemin et al., 2008).

In crop simulation models, vegetation state variables such as developmental phase, organ dry mass, and leaf area index (LAI) are linked to driving variables like weather conditions, nutrient availability and management variables. The output of the models usually is the final yield or accumulated biomass (Delécolle et al., 1992). Crop models use computational iterations that represent the time step of the model. At each iteration, vegetation state variables are updated based on the input driving variables and the values of the state variables at the previous time step (Delécolle et al., 1992).

The parallel development of agro ecosystem models and remote sensing techniques led to an early fusion of these fields and to the development of synergic applications.

The first civil satellite for earth observation, LANDSAT-1, was launched in 1972 and showed that RS (Remote Sensing) is an excellent tool to monitor the bio-geophysical processes that take place on our planet from global to regional scales (Goward and Williams, 1997). Only a few years later the North-American Large Area Crop Inventory Experiment (LACIE) and AgRISTARS programs proved that RS data could successfully assist in crop identification, estimation of some important crop canopy properties, and even help to forecast crop production (Moran et al., 1997). The launch of the Radarsat satellite in 1995 has provided opportunities for crop monitoring with radar images. The use of radar is particularly attractive because of its all weather capability and the sensitivity of microwaves to canopy structure and moistures (McNairn et al., 2002). More recently, new satellite missions emerged in the microwave domain (e.g., TerraSAR-X, Radarsat-2, Sentinel-1, or Alos-2).

Since these early days many scientists have retrieved canopy state variables over large areas using available sensors (see Table 2 for some examples). Leaf area index (LAI), fractional cover (fCOVER), the fraction of photosynthetically active radiation absorbed by the canopy (fPAR), and plant chlorophyll concentration are among the most important canopy state variables and therefore frequently assimilated in agro ecosystem models (Dorigo et al., 2007).

Canopy state and driving variables retrieved from remote sensing data and used in agroecosystem modeling studies

Biophysical parameter	Main indicator	Application	State (S) or driving (D)
Fraction of absorbed photosynthetically absorbed radiation (fAPAR)	Photosynthesis	Clevers (1997); Gobron et al. (2000)	S
Leaf Area Index (LAI)	Plant functioning	Bouman (1995); Doraiswamy et al. (2004); Mo et al. (2005); Moulin et al. (2003)	S
Fractional cover (fCOVER)	Plant development	Bouman (1995)	S
Chlorophyll and other pigments	Nitrogen stress/photosynthesis	Haboudane et al. (2002); Zhao et al. (2004)	S
Mineral content (K, P, Ca, Mg)	Crop quality	Mutanga et al. (2004)	S
Plant water content	Drought stress	Moran et al. (1994)	S
Above ground biomass/net primary production	Carbon storage; crop yield	Tucker et al. (1983)	S/D
Evapotranspiration	Drought stress	Bastiaanssen and Ali (2003); Hurtado et al. (1994)	D
Vegetation height	Plant development	Richardson et al. (1982)	S

Figure 1. Canopy state and driving variables usually retrieved from remote sensing data (Dorigo et al., 2007).

2. Crop state variables retrieval from RS.

A model is needed in order to estimate surface biogeophysical variables based on optical Remote Sensing (RS). This model should enable the interpretation of spectral observations and their translation into a surface bio-geophysical variable. From a statistical point of view this boilsdowns to a regression problem (Fernandes and Leblanc, 2005). Bio-geophysical variable retrievals, as described in terrestrial RS literature, are typically grouped in two categories: (1) the statistical category; and (2) the physical category (Baret and Buis, 2008). Over the last decade, however, both methodological categories expanded into subcategories. Exemplary is the increasing number of elements of both categories which have been integrated into hybrid approaches (Verrelst et al., 2015).

2.1. Statistical approaches.

Statistical approaches intent to reach a relationship between the spectral signature of an object, in general the leaf or canopy reflectance, and the biophysical or biochemical variable of interest (a state variable of a crop model for example). To establish such relationships, spectral, biophysical, and biochemical measurements have to be taken under varying field or laboratory conditions and for different plant species and phenological development stages. Depending on the accuracy of the measurements and the range of conditions considered, different degrees of validity and portability of the relationships will be reach (Sims and Gamon, 2002).

The retrieval of biophysical variables from RS data, requires a variety of data manipulations to enhance subtle spectral features and to reduce undesired effects caused by variations in soil reflectance, sun and view geometry, atmospheric composition, and other leaf or canopy properties. The most widespread method used to reduce background effects and enhance spectral features is to express spectral reflectance in a combination of a limited number of (transformed) spectral bands, to create what is known as a vegetation index (VI). Most VIs are focused on the red-edge region, which is the region between 680 and 800 nm. This spectral region is characterized by a sharp decrease of chlorophyll absorption from maximum absorption around 680 nm to almost zero absorption at 800 nm. This makes this wavelength range very well suited to study vegetation characteristics

(Baret et al., 1992). VIs can be subdivided into two main categories: (i) VIs designed for broadband multispectral sensors, and (ii) hyperspectral VIs based on discrete narrow bands.

The group of classic broadband VIs can be subdivided into ratios, and orthogonal indices (Broge and Mortensen, 2002). The ratios are calculated without considering the soil reflectance properties (eg. NDVI), while the orthogonal indices take into account the soil reflectance (they are referred to a specific baseline for the local soil background (eg. SAVI)). Finally, hybrid indices can be considered as a combination of ratios and orthogonal indices.

Regarding hyperspectral VIs based on discrete narrow bands, during the last decade new indices have been explored using the information contained in narrow absorption features. In this way it is possible to improve estimations of leaf constituents like chlorophyll and water (Haboudane et al., 2004) or even to explore biochemicals with more subtle spectral absorption features such as protein, lignin and phosphorus (Mutanga et al., 2004).

Furthermore, in recent times, novel approaches based on spectral shape and depth of spectral absorption characteristics have been developed. Although most of these new techniques were originally developed to identify leaf components, many of them have been successfully applied in estimating other biophysical variables such as LAI (Haboudane et al., 2004).

Examples of using statistical–empirical approaches for the estimation of canopy state variables

Vegetation index	Parameter	Vegetation type	Regional/global	Source
CRDR (Mutanga et al., 2004)	Nitrogen	Range land	Regional	Mutanga et al. (2004)
EVI (Huete et al., 1994)	LAI	Various	Global	Huete et al. (2002)
MGVI (Gobron et al., 2000)	fPAR	Green vegetation	Global	Gobron et al. (2000)
Multiple linear regression green, red, NIR + NDVI	LAI, shoot and leaf dry and fresh weight, plant height	Rice	Regional	Yang et al. (2004)
NDVI (Rouse et al., 1973)	Yield	Soybean	Regional	Liu and Kogan (2002)
PVI (Richardson and Wiegand, 1977)	fAPAR	Corn	Regional	Wiegand et al. (1991)
RVI (Pearson and Miller, 1972)	LAI	Maize	Regional	Gardner and Blad (1986)
MTCI (Dash and Curran, 2004)	Chlorophyll	Various	Global	Dash and Curran (2004)
TSAVI (Baret et al., 1989a,b)	Green Crop Area Index	Wheat	Regional	Broge and Mortensen (2002)
WDVI (Clevers, 1989)	fAPAR, LAI, Biomass	Rice	Regional	Casanova et al. (1998)

Figure 2. Examples of using statistical-empirical approaches for the estimation of canopy state variables (Dorigo et al., 2007).

Depending on the statistical method adopted for linking spectral information to the measured biophysical and biochemical variables, a distinction can be defined between two groups of method (Verrelst et al., 2015):

(1) Parametric regression methods: Parametric methods assume an explicit relationship between spectral observations and a specific bio-geophysical variable. Thus, explicit parameterized expressions are built, typically by relying on statistical or physical knowledge of the variable and the spectral response. Typically a band arithmetic formulation is defined (e.g., a vegetation index) and then linked to the variable of interest based on a fitting function.

(2) Non-parametric regression methods: Non-parametric methods directly define regression functions according to information from RS data. Hence, in contrast to parametric regression methods, a non-explicit choice is to be made on spectral band relationships, transformation(s) or fitting functions. These last ones can further be split into linear or non-linear regression methods.

2.2. Physical approach.

This approach involves the use of physical laws based on cause-effect relationships. The variables of the model are inferred based on specific knowledge, typically obtained with radiative transfer functions. These physics-based methods can also be combined with non-parametric statistical approaches like hybrid approaches. Hybrid models make use of the generic properties of physics-based methods combined with the flexibility and computational efficiency of nonparametric nonlinear regression methods (Dorigo el al., 2007).

The physical approach consists of inverting a radiation transfer model for the estimation of leaf and canopy properties. The most established way of modelling reflectance in canopies is to combine a leaf optical model with a canopy reflectance and a soil reflectance model and calculate the top-of-canopy reflectance.

Soil reflectance is an important element in modelling canopy reflectance, being the lower boundary condition and having its own spectral properties (e.g. absorption features and directional reflectance properties). Knowing soil reflectance is fundamental if sparse or low vegetated canopies are to be simulated. This model input is typically measured in the field, taken from the image itself, or can be simulated using soil reflectance models (Jacquemoud et al., 1992).

Regarding leaf optical models, the understanding of leaf microstructures and the distribution of biochemical components in leaves is still very limited. The same is true for the anisotropic scattering of leaves (Jacquemoud and Ustin, 2001). Nevertheless, various approaches have been proposed, successfully describing leaf scattering and absorption in a more or less simplified way.

Canopy reflectance models simulate the interactions between solar radiation and the elements constituting the canopy using physical laws. For applications in remote sensing this calculated reflectance should be in agreement with measured reflectance data corrected for atmospheric influences (Dorigo et al., 2007).

The traditional canopy reflectance models, which are based on the radiative transfer approach, assume that the canopy is a turbid medium where the canopy elements (leaves) are treated as small, randomly distributed absorbing and scattering elements with no physical size. A one-dimensional approximation (Verhoef, 1984) assumes the canopy to be horizontally homogeneous and infinite but vertically variable and finite. These assumptions, together with the fact that leaf area is explicitly taken into account, make this type of model well suited for describing radiance propagation in denser canopies where the single vegetation elements are smaller than the canopy height, which is the case for most agricultural crops (Dorigo et al., 2007).

Inverting a canopy reflectance model consists in finding the set of input parameters that leads to the best match between the bi-directional reflectance factor (BRF) simulated with a canopy reflectance model and the reflectance measured by the sensor (Combal et al., 2002). Different methods have been developed to solve this problem.

Examples of various approaches used to model leaf and canopy reflectance

Medium	Type	Leaf model	Canopy model
Homogeneous	1D radiative transfer	(Fukshansky et al., 1991)	SAIL (Verhoef, 1984), KUUSK (Kuusk, 1995a)
	Plate model	PROSPECT (Jacquemoud and Baret, 1990)	–
Heterogeneous	3D radiative transfer	–	DISORD (Myneni et al., 1992)
	Geometric	–	Chen and Leblanc (1997)
	Hybrid	–	DART (Gastellu-Etchegorry et al., 1996), GeoSAIL (Huemmrich, 2001), TRIM (Goel and Grier, 1988) INFORM (Schlerf and Atzberger, 2006)
	Ray tracing	RAYTRAN (Govaerts et al., 1996)	RAYTRAN (Govaerts and Verstraete, 1998), SPRINT (Goel and Thompson, 2000)
	Radiosity	ABM (Baranoski and Rokne, 1997)	PARCINOPY (Chelle and Andrieu, 1998)
	Stochastic	SLOP (Maier et al., 1999)	SMRT (Shabanov et al., 2000)

Figure 3. Examples of various approaches to model leaf and canopy reflectance (Dorigo et al., 2007).

Comparison of physical and statistical approaches

Statistical	Physical
Many field or laboratory measurements required for establishment of statistical relationship	Field or laboratory measurements only used for validation
Spectral data usually transformed	Original spectra used for inversion
Function usually based on a limited number of spectral bands	Inversion usually based on complete spectral information
Statistical function accounts for one variable at the time	Various parameters estimated at the same time
Not possible to incorporate information of other variables	Possibility to incorporate prior information on distribution of different variables
Computationally not very demanding	Computationally very intensive
Atmosphere, view, and sun geometry are not directly accounted for	Influences of atmosphere, view and sun geometry are directly incorporated
Statistical approaches normally based on nadir measurements	Possibility to use multiangular information
Little knowledge of user required	Knowledge of user required for the choice of appropriate canopy reflectance model, inversion technique, and distribution of variables

Figure 4. Comparison of physical and statistical approaches for crop state variables retrieval from RS (Dorigo et al., 2007).

3. Coupling crop models and crop state variable retrieved from RS data.

Crop models are widely used to describe the impact of climatic conditions and management strategies at field scale, and can be applied in a distributed mode at regional scale. The major problems with some crop and ecophysiological models could be related to a certain oversimplified description of the natural system, inaccurate parameterization and uncertainty, and hence a low prediction performance. The problems are particularly evident at regional scales where model input parameters have to be gathered from scattered point locations such as weather stations (de Wit et al., 2005). Boundary conditions (soil, management) are often poorly known and model parameters have to be estimated from limited experimental data. Remote Sensing (RS) offers the spatial observation of biophysical/biochemical variables. Therefore the combined use of RS derived biophysical/biochemical state variables and crop models is able to improve their predictive performance, especially at regional scale (Launay and Guérif, 2005).

Various methods have been developed to integrate remotely sensed observations in crop models. In general, three different strategies can be applied which have been described in various papers (Bach and Mauser, 2003; Delècolle et al., 1992; Houser et al., 1998; Makowski et al., 2003; Moulin et al., 1998).

<u>Calibration:</u> With the 'calibration' method, an optimal agreement between the simulated and the observed state variables of the crop model are otained by adjusting model parameters or initial conditions. The sensitive and uncertain model parameters are calibrated either manually or automatically by running the model with various combinations of parameter values within realistic ranges. Examples of this type are given by Maas (1988) and Bouman (1995).

The direct use of radiometric information to re-parameterise and/or re-initialize a crop model is another version of this calibration approach (instead of using the crop model state variable value). This version of calibration is based on considering that the temporal behaviour of canopy surface reflectance, as they can be observed from satellite, can be reproduced by coupling a radiative transfer model to the crop production model (Bouman 1992, Major et al. 1992, Fischer et al., 1997). Analytic reflectance models accounts for view and solar geometries, crop structure and crop and soil optical properties. The minimization of differences between the simulated and observed reflectance (not crop model state variables) is carried out by adjusting initial conditions or model parameters. The pertinent parameters are those which strongly constrain the behaviour of both satellite signals and biological variables of interest, i.e., the parameters which drive the canopy development.

<u>Forcing:</u> The 'forcing' strategy consists of the substitution of at least one state variable in the model using remote sensing data. The direct use of observed data to prescribe a state variable requires the availability of observations at each model time step, which is daily or weekly in case of most of the agroecosystem models. However remote sensing estimations are available only at acquisition dates, generally less frequent than the model step. To derive the state variable at model time step different interpolation techniques, such as linear interpolation, fast Fourier transformations (Roerink et al., 2000) and wavelet approaches, are used to fill the gaps between two observations.

Updating: The 'updating' method consists of the continuously updating of model state variables, whenever an observation is available. In this case there is not necessary a direct use of observed data and an interpolation process to derive the state variable at each model time step. The state variable is updated based on observations (according to different approaches) and the crop model is applied until the next update of the state variable. This strategy is more generally referred to as 'sequential data assimilation' and several algorithms have been developed for assimilating observations into models (McLaughlin, 2002). A better simulated state variable at one day will also improve the accuracy of the simulated state variable at succeeding days. This is the assumption of this method.

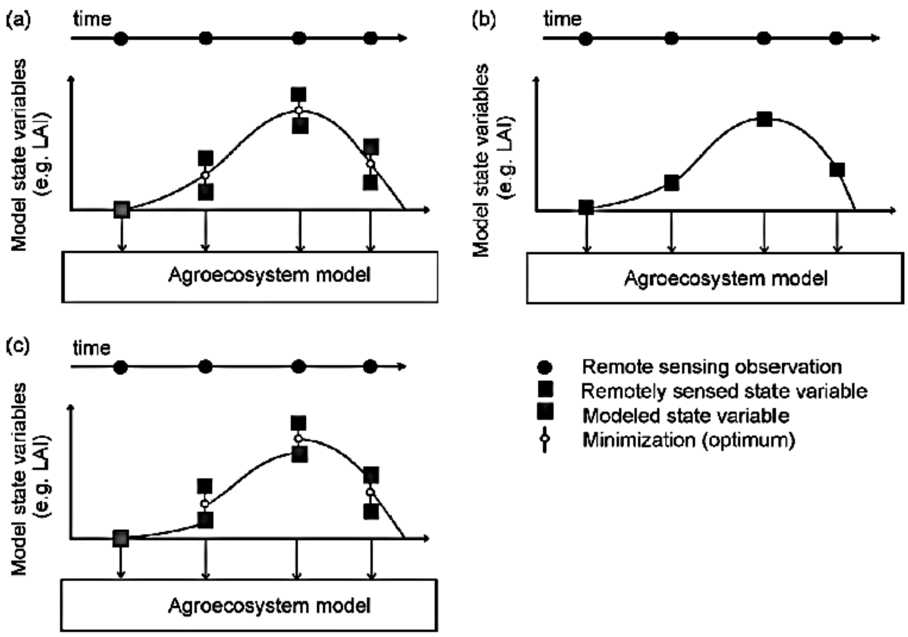

Figure 5. Schematic representation of different methods to integrate remotely sensed observations in crop models (Dorigo et al., 2007).

4. Examples of coupling.

4.1. Empirical approaches.

Among the different empirical approaches that convert "directly" data provided by the satellites into useful information (such as crop yield), machine learning plays a fundamental role. It allows to model complex patterns that cannot be discerned with a simple technique (linear models). For the development of these techniques, the increase of computational power and the lowering of computer equipment has been a key point. Below, some examples of the use of empirical approaches (machine learning techniques in this case) for modelling crop yields and other crop parameters based on remote sensing data.

The most succesful empirical techniques to achieve precise yield forecast are Artificial Neural Network (ANN), Random Forests (RF), Support Vector Regression (SVR) and k-nearest neighbour. For example, Pantazi et al. 2016 have used satellite information and soil information to estimate yield in a 22 ha wheat field. The satellite images were obtained on May 2 and June 3 from the UK-DMC-2 of the Disaster Monitoring Constellation for International Imaging (DMCii) with a resolution of 22 m. A modified version of ANN was used as algorithm to estimate the yield. NDVI was calculated from the satellite bands. This methodology allowed improving wheat yield forecasts.

Panda et al. (2010) implemented Back-propagation Neural Network (BPNN) to test the efficiency of the following four spectral vegetation indices: NDVI, green vegetation index (GVI), soil adjusted vegetation index (SAVI) and perpendicular vegetation index (PVI) in corn crop yield prediction. The results showed that the corn yield was best predicted using BPNN models that used the means and standard deviations of PVI grid images.

Han et al. (2020) proposed a multi-source data machine learning approach to improve wheat yield forecast in China. Kamir et al. (2020) propose a similar model to improve yield estimation from Australia.

Hunt et al. (2019) provide a novel demonstration of the use of freely available Sentinel-2 data to estimate within-field wheat yield variability in a single year. In this study, RS information was combined with environmental data to improve model prediction capacity. Unlike other authors, Hunt et al. (2019) used Random Forest (RF) algorithm to generate the model. RF results were compared with simple regression. 34 wheat plot were analysed in this study and the mean RMSE was 0.61T/ha. Applying this method they obtained a range of crop yield across the landscape of 4.09 to 12.22 t/ha, with a total crop production of approx. 289,000 T.

One of the latest published works has been done by Kayad et al. (2019). A 22-ha cornfield was monitored during 2016 to 2018. The yield was measured using grain yield monitor, mounted on the harvester machine. Vegetation indices obtained from 34 satellite images were used in this study. Each image was compared through correlation with yield image. Multiple regression and two different machine learning approaches were also tested to model corn grain yield. Green Normalized Difference Vegetation Index (GNDVI) provided the highest R^2. The most accurate yield prediction was obtained by Random Forest technique.

Some satellites provide Synthetic Aperture Radar (SAR) images, nevertheless, only few studies have addressed the possibilities of combining radar data and crop. Thanks to the Sentinel-1 mission (ESA), images are now provided routinely and freely all over the world which allow to develop operational services. This satellite constellation has a revisit period of 6 days and a spatial resolution of 20×22 m (10×10 m of pixel spacing). One of the greatest advantages of radar images over optical images is that cloudiness does not affect them. Therefore, they are useful every day of the year, making it easier to obtain a better time series. Simple empirical approaches have been validated on corn to retrieve basic biophysical variables such as dry biomass and leaf area index (Baup et., al 2019). Ndikumana et al. (2018) proposed a methodology to estimate rice height and dry biomass based on SAR retrievals. To do this, 3 machine learning techniques were compared: Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Random Forest

(RF). The study was done using multi-temporal dataset (may-17 to september-17) of Camargue region. Model validation was done with data acquired in 11 rice plots. Non-parametric methods (SVR and RF) had a better performance over the parametric MLR method for rice biophysical parameter retrievals. Dry biomass and rice high are strongly correlated ($R^2>0.9$) to dual-polarization signal of Sentinel-1 images.

4.2. Agronomical models.

As it was mentioned in the introduction agronomical, crop or ecophysiological models simulate major plant processes in a mechanistic way with different levels of detail. In the case of agronomical or crop models the description the effect of agricultural practices on a large number of crops is included. Remote sensing is a valuable tool to feed these models and compensate the lack of local information and the simplification usually introduced in these models due to their complexity.

Jin el al. (2018) in their review of data assimilation of remote sensing and crop models, presented a detailed compilation of studies in which crop models have been coupled with remote sensing data. Examples of these studies are presented in figures 6, 7 and 8, which are adapted from Jin el al. (2018). In these the studies crop model-RS data couples are divided into three groups depending on the coupling approach: calibration, forcing, update.

The progressive research of calibration method.

Crops	Crop models	State variable	Aims	References
Sugar beet	SUCROS	LAI	Yield	Clevers et al. (1994)
Sugar beet	SUCROS-SAIL	LAI	Yield	Guerif and Duke (2000)
Wheat	DSSAT	LAI	Yield	Dente et al. (2008)
Maize	DSSAT	LAI	Yield	Fang et al. (2008)
Maize, Wheat	EPIC	LAI	Yield	Ren et al. (2009) and Ren et al. (2010)
Maize	SAFY	LAI	Yield	Claverie et al. (2009)
Maize	STICS	LAI	Yiled	Jégo et al. (2012)
Wheat	WOFOST	LAI	LAI	Zhao et al. (2013)
Wheat	WOFOST	LAI	Yield	Ma et al. (2013)
Wheat	DSSAT	LAI	Yield	Jiang et al. (2014)
Rice	RiceGrow	LAI,LNA, LAI + LNA	Yield	Wang et al. (2014)
Wheat	VIP	FAPAR	Vcmax	Hu et al. (2014)
Sugarcane	MOSICAS	FAPAR	Yield	Morel et al. (2014a)
Wheat	SWAP	LAI, ET	Yield	Huang et al. (2015)
Rice	WOFOST	LAI	WRT	Jin et al. (2015a)
Rice	WOFOST	LAI	WRT	Liu et al. (2015)

Figure 6. Examples of coupling agronomical models with RS data based on calibration method (Jin et al., 2018).

The progressive research of forcing method.

Crops	Crop model	State variables	Aims	References
Maize	Maize model	LAI	AGB, LAI	Maas (1988)
Wheat, Sugar beet	SUCROS	LAI	Yield	Bouman (1995)
Rice	ORYZA1	LAI	Yield	Huang et al. (2001)
Wheat	ROTASK 1.5	LAI	YIeld	Clevers et al. (2002)
Plant	PROMET-V	LAI	AGB, LAI	Schneider (2003)
Wheat	STICS	LAI	ET	Duchemin et al. (2003)
Rice	ORYZA1	LAI	Yield	Abou-Ismail (2004)
Wheat	ROTASK	Flowering date	Yield	Jongschaap and Schouten (2005)
Potato	ROTASK 1.0	LAI, AGN	LAI, AGN	Jongschaap (2006)
Wheat	STICS	LAI	LAI	Hadria et al. (2006)
Wheat	DSSAT	LAI	AGB, ET	Thorp et al. (2010)
Sugar beet	MOSICAS	ε	Yield	Morel et al. (2012)
Wheat	WOFOST	LAI	Yield	Tripathy et al. (2013)
Sugarcane	MOSICAS	FAPAR	Yield	Morel et al. (2014b)
Maize	RS-P-YEC	LAI	Yield	Yao et al. (2015)

Note: The AGB, AGN, and ET represent aboveground biomass, aboveground nitrogen accumulation, and crop transpiration, respectively. The ξ and FAPAR represent interception efficiency index, and the fraction of absorbed photosynthetically active radiation, respectively.

Figure 7. Examples of coupling agronomical models with RS data based on forcing method (Jin et al., 2018).

Summary of updating method in data assimilation of crop model and remote s

Crops	Crop models	state variable	Aims	References
Maize	Maize model	LAI	AGB, LAI	Maas (1988)
Sugar beet	SUCROS	LAI	Yield	Clevers et al. (1994)
Maize	ROMET-V + GeoSAIL	SM	SM, Yield	Bach and Mauser (2003)
Wheat	DSSAT	LAI	Yield	Dente et al. (2008)
Wheat	STICS	LAI	LAI	Hadria et al. (2006)
Wheat, Maize	WOFOST	SM	Yield	De Wit and Van Diepen (2007)
	Palmer	SM	SM	Bolten et al. (2010)
Wheat	WOFOST	LAI	LAI	Curnel et al. (2011)
Wheat	WOFOST	LAI	Yield	Wu et al. (2011)
Wheat	DSSAT	NDVI	LAI	Li et al. (2011)
Wheat	DSSAT	LAI	Yield	Nearing et al. (2012)
Maize	WOFOST	LAI	Yield	Wang et al. (2013)
Maize	WOFOST	LAI	Yield	Zhao et al. (2013)
Wheat	WheatGrow	LAI, LNA	Yield	Huang et al. (2013)
Maize	DSSAT	LAI, SM	Yield	Ines et al. (2013)
Wheat	DSSAT + PROSAIL	LAI	LAI	Dong et al. (2013)
	DSSAT	SM	Yileld	Chakrabarti et al. (2014)
Wheat	DSSAT	LAI	Yield	Jiang et al. (2014)
Maize	WOFOST-HYDRUS-1D	LAI	Yield	Li et al. (2014)
Wheat	WOFOST	LAI	Yield	Liu et al. (2014)

Figure 8. Examples of coupling agronomical models with RS data based on updating (Jin et al., 2018).

4.3. SAFY-WB.

The agronomical models mentioned above can be considered as complex models given the number of input parameters and simulated processes. For example, the STICS model employs more than 227 parameters in order to simulate a wheat crop (129 parameters related with plant processes, 23 parameters for agricultural practices, and 75 parameters related with soil properties and dynamics). This information may available at a local scale (field scale) but it is difficult to get over large areas. As it was mentioned in the introduction there is a category of crop models, the semi-empirical approaches, that combines the Monteith equation with some major processes (plant development, water availability in the soil) but with a restricted number of formalisms and parameters compared with the "complex" crop models.

Monteith model (Monteith, 1977) provides the simulation of dry biomass production and the interception of light by vegetation. For the modelling of growth and crop yields, it can be advantageous to introduce the leaf area index (LAI), which is involved in the production of vegetative biomass and is a key variable in the functioning of crops. In this context, the Simple Algorithm for Yield Estimate (SAFY), a model developed by Duchemin et al. (2008) takes into account the main processes of cereal development and growth at the plot scale. This model is based on the light-use efficiency theory of Monteith model. It provides a simulation of the increase in dry above-ground phytomass. Also, it takes into account the influence of temperature and the dynamics of green leaves.

Recently this model (SAFY) has been combined with a simple water balance model (SAFY-WB) (Duchemin et al., 2015). This agro-meteorological model, named SAFY-WB (simple algorithm for yield estimates coupled with a water balance model) detailed in the literature (Duchemin et al., 2008, 2015), requires the following meteorological input variables: global solar radiation, air temperature, precipitation, and potential evapotranspiration. In the context of the present project the coupling of this model with RS data seems an interesting choice in order to estimate crop biomass and yield. Next there is a brief description of the model as it is described in the recent study of Baup et al., (2019).

The phytomass increases during the period of photosynthetic activity, from an initial value (TDM_0) at the day of plant emergence (D_0) to a final value when leaf senescence ends. The *TDM* (total dry biomass) and *GAI* (green leaf area index) are proportional to the absorbed photosynthetically active radiation, according to the effective light use efficiency, and a stress coefficient related to the meteorological conditions (relationship adapted from the literature (Monteith et al., 1977)).

$$TDM_{doy} = TDM_{doy-1} + \Delta TDM_{doy} \quad (1)$$

$$\Delta TDM_{doy} = ELUE \times fPAR_{doy} \times Sc \quad (2)$$

$$GAI_{doy} = GAI_{doy-1} + \Delta GAI_{doy} \quad (3)$$

$$\Delta GAI_{doy} = \Delta TDM_{doy} \times SLA_{doy} \times PLI_{doy} \quad (4)$$

Where *ELUE* represents the effective light use efficiency, *SLA* the specific leaf area, *PLI* the partitioning to leaf index, *Sc* the stress coefficient of water and temperature, and *doy* the day of the year.

The photosynthetically active fraction of solar radiation absorbed by the plants (*fPAR*) is proportional to *GAI* and *Rg* (global solar radiation).

$$fPAR_{doy} = 0.47 \times Rg \times \left(1 - e^{-k \times GAI_{doy}}\right) \quad 5$$

The leaf production and leaf senescence are controlled by a growing degree-day function. During the leaf growing period, a fraction of the daily *DM* (dry biomass) production is allocated to the leaf production

$$PRT_{doy} = max\left(0, Pl_A \times e^{\left(Pl_B \times \sum Temp_{doy}\right)}\right) \quad 6$$

$$LP_{doy} = TDM_{doy} \times PRT_{doy} \times SLA_{doy} \quad 7$$

where, Pl_A and Pl_B (empirical parameters) drive the partitioning to the leaves. The senescence occurred at a prescribed rate, when the air temperature accumulated from plant emergence ($\sum Temp_{doy}$) reaches a crop-specific threshold (*Stt*). A temperature-stress function affected the dry biomass production, considering a 2-degree polynomial specified by an optimal value (for example 30 ºC in the case of corn) and two extreme values (6 and 42 ºC for corn), beyond which the crop growth stopped.

The limited number of parameters of this model facilitates combining it with remotely sensed data. Six main parameters have been defined as target parameters through the sensitivity study described by Duchemin et al. (2008). Four of them describe the development stages of the crop as follows: partitioning to the leaf parameters (Pl_A and Pl_B) is effective during the growth phase (Equations (6) and (7)), while the cumulative temperature, which induces senescence and the rate of senescence (*Stt* and *Rs*), is used to describe the last phenological stages. Four phenological stages are simulated by the model, as follows: four to five leaves, flowering, ripening, and harvest. They are derived from D_0, from the day on which the *EDM* (ear dry biomass) starts growing, when *GAI* starts decreasing, and when *GAI* reaches 0, respectively.

Recent studies have shown that this model provides reliable predictions of biomass and correct estimations of crop yield in different water regimes (rainfed/irrigation) and for different crops: wheat, corn, and sunflower, using only optical images (Claverie et al., 2012; Duchemin et al., 2015; Battude et al, 2016), and for soybean, sunflower, and grain corn using both optical and SAR (Synthetic Aperture Radar images) [Betbeder et al., 2016; Baup et al., 2016; Fieuzal et al., 2017).

5. References

Bach, H., Mauser, W., 2003. Methods and examples for remote sensing data assimilation in land surface process modeling. IEEE Trans. Geosci. Remote Sens. 41 (7), 1629–1637.

Baghdadi, N., Zribi, M., 2016. Land Surface Remote Sensing in Continental Hydrology; ISTE Press: London, UK; Elsevier: Oxford, UK.

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

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

Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Sicre, C.M., Le Dantec, V., Demarez, V. 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. Remote Sens. Environ. 184, 668–681.

Baup, F., Ameline, M., Fieuzal, R., Frappart, F., Corgne, S., Berthoumieu, J.-F., 2019. Temporal Evolution of Corn Mass Production Based on Agro-Meteorological Modelling Controlled by Satellite Optical and SAR Images. Remote Sens., 11, 1978.

Baup, F., Villa, L., Fieuzal, R., Ameline, M. 2016. Sensitivity of X-Band (δ_0, γ) and Optical (NDVI) Satellite Data to Corn Biophysical Parameters. Adv. Remote Sens. 5, 103–117.

Betbeder, J., Fieuzal, R., Baup, F. 2016. Assimilation of LAI and Dry Biomass Data from Optical and SAR Images into an Agro-Meteorological Model to Estimate Soybean Yield. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 2540–2553.

Bouman, B. A. M., 1992, Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. Int. J. Remote Sens, 13, 2565-2581.

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

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère,J.P., Hénault,C.,Maraux, F., Seguin,B., Sinoquet, H., 2003. Anoverview of the crop model STICS. Eur. J. Agron. 18, 309–332.

Broge, N.H., Mortensen, J.V., 2002. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. Remote Sens. Environ. 81 (1), 45–57.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.-F., Huc, M., Keravec, P., Béziat, P., et al. 2012. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. Remote Sens. Environ. 124, 844 857.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnére, A., Myneni, R., Knyazikhin, Y., Wang, L., 2002. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. 84, 1–15.

de Wit, A.J.W., Boogaard, H.L., van Diepen, C.A., 2005. Spatial resolution of precipitation and radiation: the effect on regional crop yield forecasts. Agricult. Forest Meteorol. 135 (1–4), 156–168.

Delécolle, R., Maas, S. J., Gue'rif, M., and Baret, F., 1992, Remote sensing and cro production models: present trends. ISPRS Journal of Photogrammetry and Remote Sensing, 47, 145±161.

Dorigo, W., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhance agroecosystem modeling. Int. J. Appl. Earth Obs. 9, 165–193.

Duchemin, B., Fieuzal, R., Rivera, M.A., Ezzahar, J., Jarlan, L., Rodriguez, J.C., Hagolle, O., Watts, C. 2015. Impact of Sowing Date on Yield and Water Use Efficiency of Wheat Analyzed through Spatial Modeling and FORMOSAT-2 Images. Remote Sens. 7, 5951–5979.

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. Environ. Model. Softw. 23, 876–892.

Fernandes, R., Leblanc, S., 2005. Parametric (modified least squares) and nonparametric(theil-sen) linear regressions for predicting biophysical parametersin the presence of measurement errors. Remote Sens. Environ. 95 (3), 303–316.

Fieuzal, R., Sicre, C.M., Baup, F. 2017. Estimation of Sunflower Yield Using a Simplified Agrometeorological Model Controlled by Optical and SAR Satellite Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 5412–5422.

Fischer, A., Kergoat, L., and Dedieu, G., 1997, Coupling satellite data with vegetation functional models: review of different approaches and perspectives suggested by the assimilation strategy. Remote Sens Review, 15, 283-303.

Goward, S.N., Williams, D.L., 1997. Landsat and Earth system science: development of terrestrial monitoring. Photogrametric Eng. Remote Sens. 63, 887–900.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90 (3), 337–352.

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J., 2020. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. Remote Sens. 12, 236.

Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H., Goodrich, D.C., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour. Res. 34 (12), 3405–3420.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yieldmapping using Sentinel-2. Remote Sens. Environ., 233, 111410.

Jacquemoud, S., Baret, F., Hanocq, J.F., 1992. Modeling spectral and bidirectional soil reflectance. Remote Sens. Environ. 41 (2–3), 123–132.

Jacquemoud, S., Ustin, S.L., 2001. Leaf optical properties: a state of the art. In: Proceedings of the Eighth International Symposium Physical Measurements & Signatures in Remote Sensing, CNES, Aussois, France.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. Euro. J. Agron. 92 (2018) 141–152.

Jones, C.A. and Kiniry, J. R. (eds.). 1986 .CERES-Maize: A Simulation Model of Maize Growth and Development. Texas A. and M. University Press, College Station, TX. 194 pp.

Kamir, E., Waldner, F., Hochman Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J. Photo. Remote Sens. 160, 124-135.

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., Pirotti, F., 2019. Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. Remote Sens. 11, 2873.

Launay, M., Guérif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agricult. Ecosyst. Environ. 111 (1–4), 321–339.

Maas, S.J., 1988. Using satellite data to improve model estimates of crop yield. Agron. J. 80 (4), 655–662.

Major, D. J., Schaalje, G. B., Wiegand, C., and Blad, B. L., 1992, Accuracy and sensitivity analyses of SAIL model-predicted reflectance of maize. Remote Sens Review, 41, 61-70.

McLaughlin, D., 2002. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. Adv. Water Resour. 25 (8–12), 1275–1286.

McNairn, H., Ellis, J., Van Der Sanden, J.J., Hirose, T., Borwn, R.J., 2002. Providing crop information using RADARSAT-1 and satellite optical imagery. International Journal of Remote Sensing, 23:5, 851-870, DOI: 10.1080/01431160110070753

Monteith, J. L., 1972. Solar radiation and productivity in tropical ecosystems. Journal of Applied Ecology, 9, 744–766.

Moran, M.S., Clarke, T.R., Inoue, Y., Vidal, A., 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. Remote Sens. Environ. 49 (3), 246–263.

Moulin, S., Bondeau, A., Delécolle, R., 1998. Combining agricultural crop models and satellite observations: from field to regional scales. Int. J. Remote Sens. 19 (6), 1021–1036.

Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. Remote Sens. Environ. 89 (3), 393–408.

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., Hossard, L., 2018. Deep Recurrent Neural Networkfor Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. Remote Sens. 10, 1217.

Panda, S. S., D. P. Ames and S. Panigrahi. 2010. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sens. 2 (3): 673–696

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheatyield prediction using machine learning and advanced sensing techniques. Comput.Electron. Agric. 121, 57–65.

Penning de Vries, F.W.T., Van Laar, H.H. (editors), 1982. Simulation of plant growt and crop production. Simulation Monographs, Pudoc, Wageningen, The Netherlands, p. 3 308.

Reddy, V. R., Baker, D. N., and Jenkins, J. N., 1985. Validation of GOSSYM: Part II – Mississip Conditions. Agric. Sys. 17, 133–154.

Roerink, G.J., Menenti, M., Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. Int. J. Remote Sens. 21 (9), 1911–1917.

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81 (2–3), 337–354.

Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water. I. Concepts. Agron. J. 101, 426–437

Stöckle, C.O., Martin, S.A., Campbell, G.S., 1994. CropSyst, a cropping systems simulation model; water/nitrogen budget and crop yield. Agric. Syst. 46, 335–359.
Van Diepen C.A., Wolf J., Van Keulen H., Rappoldt C., 1989. WOFOST: a simulation model of crop production, Soil Use Manage. 5: 16-24.
Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote Sens. Environ. 16 (2), 125–141.
Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysica properties – A review. ISPRS J. Photogramm. Remote Sens. 108, 273–290.
Williams, J.R., Jones, C.A., Kiniry, J.R., Spanel, D.A., 1989. The EPIC crop growth model. Trans. ASAE 32, 497-511.