Early estimated wheat yield with a simple crop growth model through the assimilation of Landsat images.

Deborah Gaso [1]
Andres G. Berger[1]

[1] Instituto Nacional de Investigación Agropecuaria – INIA-La Estanzuela
CP 70006 - Colonia - Uruguay
dgaso@inia.org.uy
aberger@inia.org.uy

Abstract. On most areas where agriculture is developed in Uruguay, soil properties are highly variable in space and as a result the conditions for crop growth are also highly variable. Farmers have interest in knowing final yield pre-harvest as well as the existing spacial variability to decide management practices during the growing season. The objective of this study was to evaluate a method to optimize a simple crop model (pySAFY) to predict leaf area index through the assimilation of time series of Landsat images and later use the calibrated parameters (at the pixel level) to estimate pre-harvest yield and crop growth. The model used is based on the SAFY model (Simple Algorithm for Yield estimate), which was developed by Duchemin (2008), however major changes were made in the model to represent nitrogen accumulation and remobilization in the above ground biomass, which is restricting yield more than other factors represented in the original SAFY (water and heat stress). Both Landsat 7and 8 images free or nearly free cloud acquired during the growing season were used. The study sites were located at 33°34'44S latitude and 58°10'06W longitude in the south west of Uruguay and were planted with wheat. The model allowed to predicted average yield and spacial variability with acceptable accuracy. Yield was overestimated in areas of the field where Landsat 7 images have SLC gaps or underestimated in areas identified as low yield potential areas. Spacial (30x30m) and temporal resolution of Landsat was not enough to attain high accuracy in predicting micro variability. The method proved useful to use images readily available and allowed to predict pre-harvest yield with enough precision for most practical purposes.

Key words: vegetation index, Landsat images, model calibration, leaf area index.

1. Introduction.

Timely assessments of crop growth conditions and early yield estimations are demanded by crop insurance companies, and are used to organize logistics of harvest and support in-season decision making. During the last decade pre-harvest yield estimation has been a problem tackled from several approaches that in general used remote sensing, crop modeling or a combination of both. The application of crop models on large areas has been hampered by lack of sufficient and accurate information about inputs. The main problem to achieve rational accuracy with crop models is the difficulty to know input parameters, initial conditions and manager practices. Since the beginning the development of crop growth models was to simulate agricultural field where soil, climate and agricultural practices were well know spatially homogeneous (Mass, 1988; Guérif and Duke, 2000; Hatfield et al., 2008).

On most areas where agriculture is developed in Uruguay, soil properties are highly variable in space and as a result the conditions for crop growth are also highly variable. When using fully developed crop models to simulate growth and yield over large areas it is necessary to know the input and parameters for each homogeneous zone or grid cell. Therefore the application of crop models to predict growth and yield at field scale is still a challenge. Remote sensing reflectance of the crop over the growing season has been used to improve model capability to predict yield over large areas. While the data of remote sensing provides a quantification of the actual state of the plant community attributes (leaf area index, accumulated aboveground biomass, nitrogen status) in discrete time, the growth models supply a continuous description of plant growth. In this manner the growth models and remote sensing data are complementary (Baret & Guyot, 1991; Duchemin et al. 2008a; Moulin et al., 1998).

The objective of this study was to evaluate a method to predict early yield based on the assimilation of time series of Landsat images into a simple crop model (pySAFY).

2. Materials and Methods.

The study site was located at 33°34’44S latitude and 58°10’06W longitude in the south west of Uruguay. The site has been managed with continuous no-till agriculture with standard management practices, and a crop rotation that typically includes a winter cereal (wheat or barley) and a summer crop (soybean or maize). The sowing date ranged from mid-May to mid-June in all study sites during growing season 2012 and 2013.

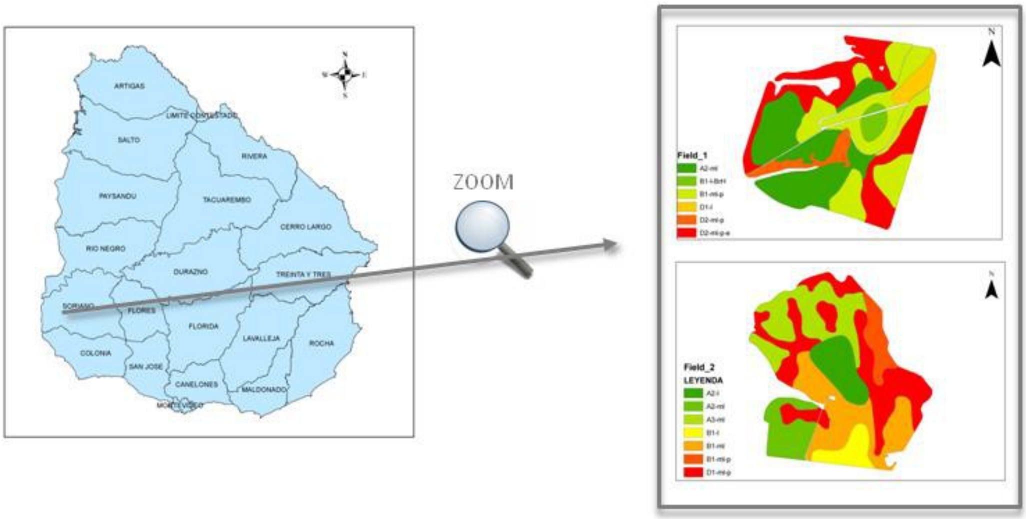

Figure 1. Localization of experimental sites.

2.1. Methodology.

Multi-temporal remote sensing data were acquired by the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 8 Operational Land Imager (OLI) sensors to monitor the growth conditions of wheat during winter growing season 2013. All images free or nearly free of clouds covering the experimental sites were identified and downloaded from the USGS-EROS archive. All images were atmospheric corrected following the methodology implemented in ENVI 5.1 and the mean surface reflectance values of red and NIR bands were extracted to calculate vegetation indices.

Within the site 13 sampling points were selected, based on prior knowledge of the location of low and high yielding areas, in an attempt to have representative samples of the different regions that could possibly be observed within that field. Each sampling point was visited at approximately the same date of image acquisition. At each sampling point total above ground biomass samples were collected in three locations within a 10m distance from the center of the sampling point. Also in the same locations, leaf area index and fractional ground cover was measured with a ceptometer (AccuPAR, Decagon Devices Inc.). The field data were used to estimate the relationship between LAI and vegetation indices.

Three vegetation indices were calculated to predict leaf area index: NDVI, WDRVI and CI_green. The evolution of these indices between emergence to flowering were obtained for each image according to equations 1, 2 and 3:

$$NDVI = {}_{\varrho}NIR - {}_{\varrho}RED \,/\, {}_{\varrho}NIR + {}_{\varrho}RED \quad (1)$$
$$WDRVI = [(\alpha+1)^{*} NDVI + (\alpha-1)]/[(\alpha-1)^{*}NDVI + (\alpha+1)] \quad (2)$$
$$CI\_green = {}_{\varrho}NIR/ {}_{\varrho}Green \quad (3)$$

where ${}_{\varrho}NIR$ is the reflectance of near infrared, ${}_{\varrho}RED$ is the reflectance of red, ${}_{\varrho}Green$ is reflectance of green, and the value of alpha was fixed at 0.1 based on Gitelson (2004),.

Time series course of LAI for each pixel was predicted using the relationship found between each index and measurement data at the sampling point. We selected the index what best predicted LAI, and this index was implemented along with the index-LAI function into the model. The model was inverted against the image time series, calibrating 2 parameters (nitrogen in leaf and initial above ground biomass).

2.2. ***PySAFY model description.***

In this work we used a semi-empirical model based on SAFY model (Simple Algorithm for Yield estimate), which was developed by Duchemin (2008). The model used a daily time step with meteorological data (maximum air temperature, minimum air temperature and solar radiation) obtained from a nearby meteorological station.

The model simulated the time series of dry aerial mass using Monteith's (1977) theory, where the total dry biomass is the product of integrating incoming solar radiation and three efficiency factors: i) photosynthetic active proportion (PAR) of total solar radiation (εc), which is the ratio between total radiation and photosynthetic active radiation; ii) the light interception efficiency (εi), which is the portion of photosynthetic active radiation that is absorbed; and effective light-use efficiency (ELUE), which is the ratio between produced dry biomass and absorbed photosynthetic active radiation. Daily total above-ground dry matter accumulation is calculated though the following equation:

$$DAM = Rg * \varepsilon c * \varepsilon i * ELUE \quad (2)$$

The light use efficiency is tightly associated with specific nitrogen content (SPLN) of the leaf, which is allowed to change over time according to the nitrogen balance of the leaves. To calculate daily ELUE we used the relationship proposed by Sinclair and Amir 1992 through the following equation:

$$ELUE = 1.5 * (1 - \exp(-1.7 * SPLN - 3)) \quad (3)$$

During vegetative growth total nitrogen incorporated into the plant was allocated between leaves and stems assuming allometric growth, therefore implying constant rates of proportional growth for each plant fraction. In the model green leaf area index is calculated directly as a function of total above-ground dry matter with the following function:

$$LAI = ((DAM(i) / DAM(i-1))^{lallomb}) * LAI(i-1) \qquad (4)$$

where DAM (i) and DAM (i-1) is dry aboveground matter of day i and i-1 respectively, LAI (i-1) is leaf area index of day i-1 and lallomb is the allometric growth proportionality parameter associated with partitioning between leaves and stems. Parameter lallomb was adjusted with field data from crops growing in the same local conditions (lallomb=0.8747).

Leaf senescence after flowering was the result of the nitrogen balance of the crop and the remobilization of nitrogen from leaf and steams to the growing grains. The amount of nitrogen available for translocation to seeds was calculated as total nitrogen accumulated in the leaves and stems on the end of vegetative growing discounting for the nitrogen which was structural and not available for translocation. The length of the grain filling period was governed by the pool of nitrogen in the plant and the rate of nitrogen translocation. The model considered to be the end of the grain filling period when LAI reached a value lower than 1.

The destination of daily production of dry matter (DAM) depends on phenological phases. For this reason to run the model it is necessary to know as an input parameter the beginning and end of the vegetative growing phase (emergence to flowering). The phenological phases were not simulated by pySAFY, these data should be incorporated by the user as inputs. Final yield would be the result of integrated total dry matter production throughout the growing season multiplied by a harvest index factor that was increasing linearly over the grain filling period as proposed by Sinclair and Amir, 1991. The equation to calculate grain dry matter is therefore:

$$GDM = DAM * HI \qquad (5)$$

where

HI = 0.011 * days_after_anthesis

2.3. *Assessment of model performance.*

Estimated yield was compared with data of yield monitor at the test sites. Data collected with the monitor were interpolated using ordinary kriging. The model accuracy for calibration and validation was tested using RMSE (root mean square error) and Bias statistics:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Di)^2}{N}}$$

$$Bias = \sum Di / N$$

where Di is the difference between Y – Yi, Yi is measurement yield for simulation scenario i, Y is corresponding value estimated for the model, and N is the total number of simulation scenarios.

3. Results and discussion.

3.1. Green LAI estimation.

In previous works crop biophysical descriptors, such as LAI have been estimated from vegetation indices to estimate the linking functions through empirical relationships. Equations describing these relationships have different mathematical forms (linear, exponential, power, inverse of exponential, etc). The empirical coefficient depends on field conditions, the indices used and the crop type (Gitelson, 2004; Haboudane et al., 2004; Matsushita and Tamura, 2002; Qi et al., 2000).

Using data from sampling points on multiple fields during year 2013, NDVI, WDVRI and CI_green vegetation indices were estimated to predict LAI from spectral data of Landsat images. Orthogonal regression was fitted to WDRVI and CI_green, while a logarithmic function was fitted to NDVI using JMP software (Figure 2).

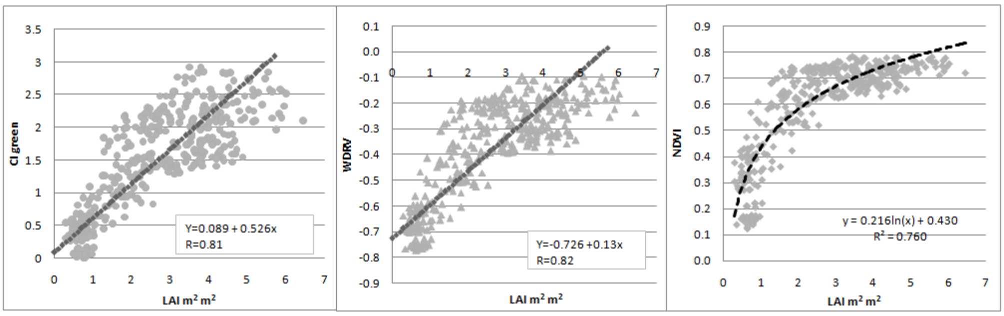

Figure 2. Relationship between vegetation indices and LAI during 2013 season in one field nearby of study area.

WDRVI and CI_green did not show saturation for values of LAI above 3 in contrast to NDVI, but deviations from linearity and fit error was greater in WDRVI than CI_green. Hence we select CI_green to predict LAI in model runs.

Times series of LAI were calculated using the above obtained empirical relationships between CI and LAI. These maps of LAI were assimilated into the pySAFY model. In the assimilation process two parameters were optimized using Levenberg-Marquardt optimization: initial above-ground biomass and specific leaf nitrogen content of the plant.

3.2. *Yield estimation.*

Running the model in forward mode over Landsat images with the corresponding estimated two parameter for each pixel, allowed the prediction of grain yield independently at the pixel level. The spatial yield variability was correctly predicted with the model, moreover average yield of all fields and average yield of different management zones (zones of different potential into the field) was well predicted. In areas where Landsat 7 images have SLC gaps (times series of LAI to predict parameters had fewer points), pySAFY model overestimated yield. Otherwise estimation yield in different areas of the field was acceptable.

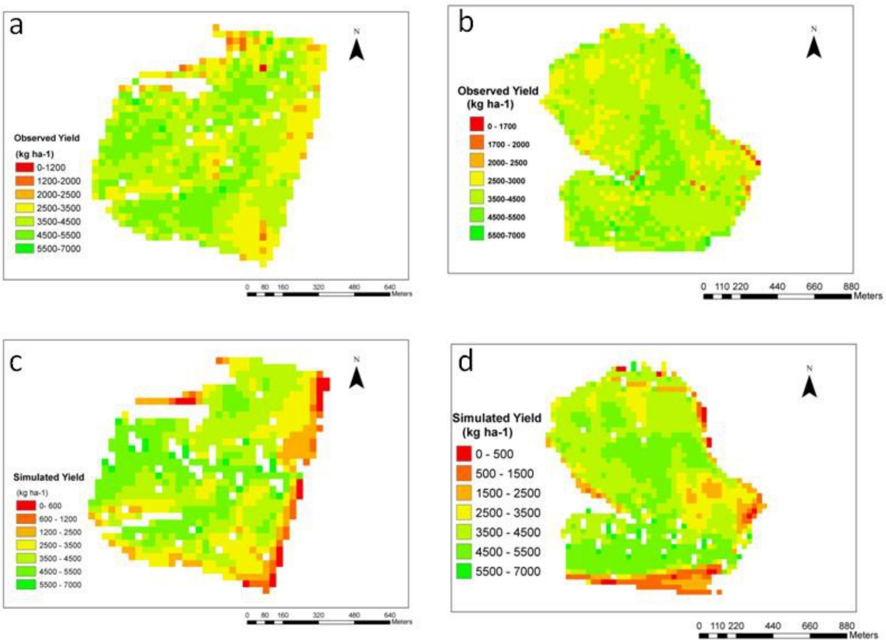

Figure 3. Estimated yield with pySAFY and observed yield (kg ha-1) for Field 1 (a and c) and for Field 2 (b and d).

The average Bias was low at each field (Bias = 75 and -159 in field 1 and 2 respectively).This was associated to areas where the model predicted yield accurately, and areas where the estimation showed large bias which corresponded with Landsat 7 SLC gaps or edges of the field.

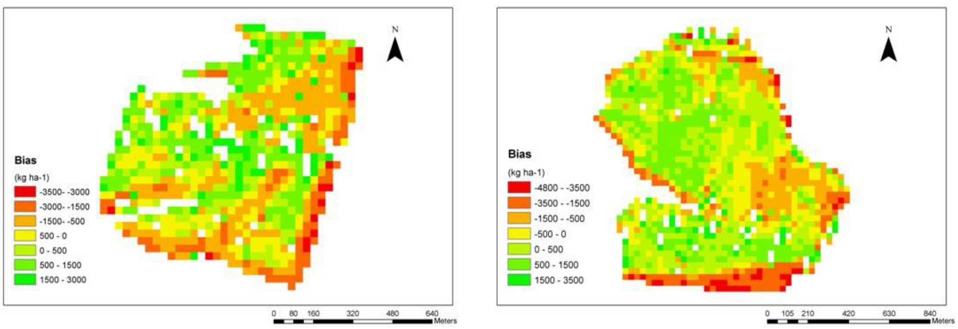

Figure 4. Difference between yields estimated and observed (kg ha-1), a corresponding with Field 1 and b Field 2.

Even though the model predicted with enough accuracy average yield and spacial variability, micro variability was not estimated correctly. The statistical RMSE was high (1517 and 1206 in field 1 and 2 respectively). The model was more sensitive, and tended to estimate lower yields in areas of poor crop development, where peak LAI over the growing season was small. This derived in a tendency of the model to underestimate yield in areas that were classified as lower potential areas.

4. Conclusions

The method of assimilating remote sensing data of low resolution into the pySAFY model allowed achieving enough accuracy to predict average yield and spacial variability for most practical purposes.

The model allowed to detect and differentiate the general trend of yield within a field, and to detect areas of high vs. low yield potential. However, the micro variability (pixel-to-pixel variability) was not well estimated.

The model was sensitive to the increase or decrease in the number of images in the time series with decreased error in the estimates when the time series were complete. Better accuracy would be achieved with higher spatial and temporal resolution of images.

5. References.

Bater, F. and Guyot, G. Potential and limits of vegetation indices for LAI and APAR assessment. Remote Sensing of Environment, v 35, p. 161-173, 1991.

Duchemin, B.; Maisongrande, P.; Boulet, G.; Benhadj, I. A simple algorithm for yield estimates: Evaluation for semi-arid irrigated winter wheat monitored with green leaf area index.Environmental Modelling Software, v 23, p. 876-892, 2008.

Gitelson, A. Wide Dynamic Range Vegetation Index for Remote Quantification of biophysical characteristics of vegetation. Journal of Plant physiology, v 161, p. 165-173, 2004.

Guérif, M. and Duke, C.L. Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. Agriculture, Ecosystems and Environment, v 81, p 57-69, 2000.

Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agricultura. Remote Sensing of Environment, v 90, p. 337-352, 2004.

Mass, S.J. Use of remotely – sensed information of agricultural crop growth model. Ecological Modelling, v 41, p. 247-268, 1988.

Matsushita, B. and Tamura, M. Integrating remotely sensed data with an ecosystem model to estimate net primary productivity in East Asia. Remote Sensing of Environment, v 81, p. 58–66, 2002.

Moulin, S.; Bondeau, A.; Delecolle, R. Combining agricultural crop models and satellite observations: From field to regional scales. International Journal of Remote Sensing, v 19, n 6, p. 1021-1036, 1998.

Robertson, A.L.N.; Peng, Y.; Gitelson, A.; Arkebauer, T.J.; Pimsteinc, A.; Herrmannc, I.; Karnielic, A.; Rundquist, D. C.; Bonfil, D.J. Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. Agricultural and Forest Meteorology, v 192–193, p. 140–148, 2014.

Qi, J.; Kerr, Y. H.; Moran, M. S.; Weltz, M.; Huete, A. R.; Sorooshian, S.; Bryant, R. Leaf Area Index Estimates Using Remotely Sensed Data and BRDF Models in a Semiarid Region. Remote Sensing of Environment, v 73, p. 18–30, 2000.