<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Spatialized Application of Remotely Sensed Data Assimilation Methods for Farmland Drought Monitoring Using Two Different Crop Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>Silvestro, Paolo Cosmo; Casa, Raffaele; Pignatti, Stefano; et al.</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2016-07-08</td>
</tr>
<tr>
<td><strong>Publication information</strong></td>
<td>Ouwehand, L. Proceedings of Dragon 3 Final Results and Dragon 4 Kick-Off</td>
</tr>
<tr>
<td><strong>Conference details</strong></td>
<td>Dragon 3 Final Results and Dragon 4 Kick-Off, Wuhan, China, 4-8 July 2016</td>
</tr>
<tr>
<td><strong>Link to online version</strong></td>
<td><a href="http://earth.esa.int/dragon-2016/default_1622.html">http://earth.esa.int/dragon-2016/default_1622.html</a></td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/10200">http://hdl.handle.net/10197/10200</a></td>
</tr>
</tbody>
</table>
SPATIALIZED APPLICATION OF REMOTELY SENSED DATA ASSIMILATION METHODS FOR FARMLAND DROUGHT MONITORING USING TWO DIFFERENT CROP MODELS

Silvestro Paolo Cosmo (1), Casa Raffaele (1), Pignatti Stefano (2), Castaldi Fabio (1), Yang Hao (3), Yang Guijun (3)

(1) Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy), Email:rcosa@unitus.it; p.c.silvestro@unitus.it
(2) Consiglio Nazionale delle Ricerche – Institute of Methodologies for Environmental Analysis (C.N.R. – IMAA), Via del Fosso del Cavaliere, 100, 00133 Roma, (Italy) Email: stefano.pignatti@cnr.it
(3) National Engineering Research Center for Information Technology in Agriculture (NERCITA), 2449-26, Beijing 100097, P.R. China Email: yanggj@nercita.org.cn; yangh@nercita.org.cn

ABSTRACT

The aim of this work was to develop a tool to evaluate the effect of water stress on yield losses at the farmland and regional scale, by assimilating remotely sensed biophysical variables into crop growth models. Biophysical variables were derived from HJ1A, HJ1B and Landsat 8 images, using an algorithm based on the training of artificial neural networks on PROSAIL. For the assimilation, two crop models of differing degree of complexity were used: Aquacrop and SAFY. For Aquacrop, an optimization procedure to reduce the difference between the remotely sensed and simulated CC was developed. For the modified version of SAFY, the assimilation procedure was based on the Ensemble Kalman Filter. These procedures were tested in a spatialized application, by using data collected in the rural area of Yangling (Shaanxi Province) between 2013 and 2015. Results were validated by utilizing yield data both from ground measurements and statistical survey.

1. INTRODUCTION

Assimilation of biophysical variables retrieved from remote sensing into dynamic process-based crop water response models is a promising methodology for the assessment of drought impact on crops, suitable for high spatial and temporal resolution applications [1]. In particular such methodology allows to estimate yield losses resulting from the occurrence of drought, provided that the crop models used do take into account the relevant crop physiological processes occurring during water stress. The models suitable to study the impact of drought on crops should simulate crop growth taking into account factors such as the availability of soil water and its impact on crop evapotranspiration. For operational assimilation schemes, a balance must be found between the accurate description of processes leading to very complex crop models with hundreds of parameters and input variables and simpler models, with a more coarse or empirical description of the relevant processes [2]. The former are usually more difficult to manage during assimilation, the latter could be more suitable operationally, with less uncertainty introduced during parameterization, but could be less accurate.

In the context of Sino-European Dragon-3 program funded by ESA and the Ministry of Science and Technology (MOST) of the P.R. China, the “Farmland Drought” project dealt with the development of crop variables assimilation schemes and their validation at the farmland scale for a test site in China. The work developed in this context focused on the application of a methodology of assimilation of biophysical variables into two crop models of different complexity and the comparison of validation results for a rural area in the Province of Shaanxi.

2. STUDY AREA AND MATERIALS

2.1 In situ data

The study site used in this work is located in the rural area of Yangling, in the Province of Shaanxi, Central China. The experimental site is characterized by a semi-arid climate with a long term average rainfall of 573 mm per year, a minimum temperature of -4°C in winter and a maximum temperature of 35°C. The climatic data sets included three growth seasons, from October 2012 to June 2014 and consisted of rainfall, temperatures (mean, maxima and minima), relative humidity and wind speed. The crop considered in this study is winter wheat, the main crop of the study area. Field measurements were carried out from March to June for each crop season. The details of field measurements are shown in Tab. 1.

Table 1. Field measurements of LAI (L), biomass (b) and yield (y) conducted in Yangling from 2013 to 2015. n. pts are the number of sample fields.

<table>
<thead>
<tr>
<th>Year</th>
<th>n. pts</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>49</td>
<td>30 Mar</td>
<td>27 Apr</td>
<td></td>
<td>1 Jun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L, b</td>
<td>L, b</td>
<td></td>
<td>b,y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 Mar,</td>
<td>22 Apr</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>2014</td>
<td>35</td>
<td>28 Mar</td>
<td>27 Apr</td>
<td>May</td>
<td>9 Jun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L, b</td>
<td>L, b</td>
<td>L, b</td>
<td>b,y</td>
</tr>
<tr>
<td>2015</td>
<td>28</td>
<td>27 Mar</td>
<td>25 Apr</td>
<td></td>
<td>5 Jun</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L, b</td>
<td>L, b</td>
<td></td>
<td>b,y</td>
</tr>
</tbody>
</table>

They consisted of observations in 28 to 49 points...
Images acquired by Landsat 8 were provided by the United States Geological Survey with atmospheric and geocorrection, as reflectance values. Images acquired by HJ were corrected geographically by a series of ground control points on the map and warping the images. The HJ data were converted from Digital Number (DN) to radiance using the coefficients provided in the metadata. Once the image data were converted into radiance, atmospheric correction was performed using the FLAASH atmospheric correction module in ENVI (Harris, FL, USA). A winter middle latitude atmospheric model was used for images acquired between January and March, while a summer middle latitude model was used for images acquired between April and June. The selected aerosol model was rural with no aerosol retrieval and an initial visibility between 20 km and 40 km, depending on the image.

3. METHODS
In order to obtain an estimation of yield and to evaluate the effect of drought, two crop growth models were chosen among those able to correctly represent crop response to water stress and suitable for regional scale applications. We compared two models with respect to three different criteria: complexity, plasticity and accuracy. Subsequently the assimilation of biophysical variables retrieved from remote sensing was carried out, validating the simulated yield with point field measurements. Yield values were also compared with those published by the National Bureau of Statistics of the People's Republic of China.

3.1 Crop Models
The two models considered are Aquacrop [3] and SAFY, the Simple Algorithm For Yield [4]. Aquacrop is a "water driven" productivity model that simulates Canopy Cover (CC), biomass and yield of a crop mainly as a function of the water productivity, i.e. the biomass produced per unit of water transpired by the vegetation [3]. A more detailed explanation about how the Aquacrop model works is presented by Raes et al. (2009) [5]. SAFY is based on the Monteith concept [6], which links the production of total dry phytomass to the photosynthetically active portion of solar radiation (PAR) absorbed by the crop. A detailed description of SAFY was presented by [4]. In this work a modified version of SAFY was used, introducing a dependence of biomass yield on crop water stress, as described by the FAO irrigation and drainage paper n. 56 [7].

3.2 Comparison Criteria: Complexity, Plasticity and Accuracy
Complexity is the parsimony of the model in representing the biophysical system, it gives information about amount and relevance of model parameters [8]. In this study it is evaluated using as sample index the proportion of relevant parameters (S) over all total number of model parameters (T):

$$R_p = \frac{S}{T}$$  

(1)

Plasticity is the tendency of the model to change its behaviour when employed in different conditions[9]. Changes in model behaviour can be quantified using an indicator of model plasticity (L):

$$L = TDCC \cdot e^{\sigma_{SAM}}$$  

(2)

where TDCC is the top-down concordance coefficient and $\sigma_{SAM}$ is the standard deviation of the normalized agrometeorological indicator.

Accuracy is the ability of the model to fit reference measured data [10], in this case the yield, using root mean square error (RMSE) as a sample index. The accuracy of the model is evaluated by measuring the RMSE between yield measured in the field and that estimated by the models. It has been reported ([11], [4]) that the accuracy of Aquacrop is higher than SAFY, but additional considerations apply for an application at a regional scale.

3.3 Global Sensitivity Analysis
Most assimilation methods update the parameters on the basis of observations of the state variables. For computational requirements the number of parameters should be limited. Sensitivity analysis can help to identify the parameters that most affect the output of the model. We used a combination of two different Global Sensitivity Analysis (GSA) methods: Morris and EFAST ([9]; [12]). The first is a screening method and it is used to identify negligible parameters, taking
advantage of its high computational speed. In this way, negligible parameters were excluded from the SA performed with the second method, in order to reduce its running time. EFAST is a variance based method, its accuracy is higher than Morris and using a reduced set of parameters the computational time is reduced, balancing accuracy and running speed. The SA is also useful to quantify the complexity and the plasticity of the models

3.4 Estimation of LAI and Canopy Cover

LAI and Canopy Cover were calculated using the algorithm developed by Baret et al. (2006) [13] which is based on the training of artificial neural networks (ANN) over a data base simulated using radiative transfer models. A coupling of PROSPECT and SAIL models [14] was adopted to simulate the reflectance spectra of vegetation, in this case winter wheat.

A wide range of possible canopy scenarios are described by the combination of 14 parameters which relate to canopy, leaves, soil and observation geometry. Selecting and sampling the range for each parameters, the model PROSAIL generates a sufficient number of simulation (in this study 41472) to describe all expected scenarios.

The set of simulated reflectance is resampled according to the characteristics of the used sensor, then it is divided in two parts, one of them with the purpose of training ANNs and the other to evaluate the theoretical performance.

To make the simulations more realistic and reduce the ANN training hyper-specialization, a Gaussian noise error is added on the basis of an additive and multiplicative model. After the training of the neural networks, the coefficients and bias are stored to be used later in the estimation of variables LAI, FAPAR and FCOVER from reflectance observed values.

3.5 Ensemble Kalman Filter Assimilation on SAFY

In order to improve the accuracy of SAFY for simulating Yield, an EnKF based assimilation method was adopted [1]. The first step of this method is the individuation of main parameters, as previously done by means of the SA. The ranking of the most important parameters (and relative range of variation) is shown in Tab. 3.

Table 3. List of the most influential parameters of the SAFY model, as detected by the SA, allowed to vary during the assimilation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min</th>
<th>Max</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMT_sen</td>
<td>2070</td>
<td>2530</td>
<td>GDD to complete senescence</td>
</tr>
<tr>
<td>Da</td>
<td>1.35</td>
<td>1.55</td>
<td>Soil bulk density (g cm(^{-3}))</td>
</tr>
<tr>
<td>Ptfn_Topt</td>
<td>20</td>
<td>22</td>
<td>Optimum temperature for plant development (°C)</td>
</tr>
<tr>
<td>Pfen_SenA</td>
<td>1260</td>
<td>1540</td>
<td>Temperature threshold to start senescence (°C)</td>
</tr>
<tr>
<td>Pfen_PrtA</td>
<td>0.16</td>
<td>0.2</td>
<td>Make vary the origin and slope of the partitioning function</td>
</tr>
<tr>
<td>Pfen_MrgD</td>
<td>288</td>
<td>300</td>
<td>Day of the year of emergence</td>
</tr>
<tr>
<td>Pgro_P2G</td>
<td>0.0098</td>
<td>0.012</td>
<td>Partition coefficient to grain</td>
</tr>
<tr>
<td>Pgro_Lue</td>
<td>1.75</td>
<td>1.95</td>
<td>Effective light-use efficiency: ratio of energy produced as DAM from APAR (g MJ(^{-1}))</td>
</tr>
<tr>
<td>Pgro_R2P</td>
<td>0.7</td>
<td>0.9</td>
<td>Climatic efficiency: ratio of incoming photosynthetically active radiation (PAR) to global radiation</td>
</tr>
</tbody>
</table>

All the model parameters were calibrated using, as a reference, 5 points of the measurement field campaign of Yangling 2013. Even the parameters that vary in the assimilation were initially calibrated in this way. The calibrated values are used as mean values of the range in which the parameters vary, the minimum and maximum of the ranges are chosen arbitrarily in a surrounding of the mean value. It is possible to divide the assimilation algorithm into the following steps:

1) The algorithm generates an ensemble of N (in this case N=100) values of parameters \( \mathbf{P}_p \)

\[
P^j_k = P_k + \delta^j_k \quad \text{where:}
\]

\[j = [1, \ldots, N] \quad N=\text{number of ensemble elements (100 in this case)} \]

\[k=\text{number of parameters} \quad \delta = \text{error drawn from N(0,} \sigma) \]

2) Running SAFY in order to obtain N values of LAI for the first observation date with the addition of an error \( \varepsilon \)

3) The algorithm generates an ensemble of N observations from RS data:

\[
M^j_i = M^0_i + \tau^j_i \quad \text{(4)}
\]

where \( M \) is the set of state variables measurements, \( i \) is the index in the temporal series, \( \tau \) is the measurements error drawn from N(0,\( \sigma)\)

4) Calculating the Kalman gain (using the covariance matrix)

\[
K_i = \frac{\sum_{j} R^j_l}{R^{\sum \sum} R^j_l + \text{var}(\tau^j_i)}
\]
5) Application of the Kalman filter to obtain a corrected LAI and parameters values estimated
6) Replacement of LAI simulated with the LAI calculated at the previous step, and further running of the model.
7) At each observation date repeat from step 3). When the last observation has been assimilated SAFY runs to the end and outputs the yield.

3.6 Optimization method on Aquacrop
Aquacrop is a water-driven crop growth model, for this reason it is linked with water balance and the yield simulated by this model is strongly influenced by water stress. These characteristics make it suitable to describe the links between drought and yield. Another advantage of Aquacrop is the use of CC as a main state variable which turns out to be estimated more accurately than LAI from remote sensing. The problems related to Aquacrop are mainly due to its complexity and the unavailability of the code to the scientific community. For these two reasons it was not possible to perform an updating assimilation of the state variable (as done for SAFY). The assimilation method employed was based on the recalibration, through a numerical optimization technique (simplex method), of a limited set of parameters chosen according to the sensitivity analysis study. The first 4 steps of this algorithm are the same of the previous used with SAFY, the difference is at the fifth step, the simplex optimization method is applied. This method finds the best parameters which minimize the following function:

\[
MSE = \sum_{i=1}^{n} \frac{(CC_i(t) - CC_m(i))^2}{n}
\]

where:
MSE = Mean Square Error
n = numbers of parameters
CC_i = simulated Canopy Cover
CC_m = measured Canopy Cover

3.7 Regional Scale Application
The application at the regional scale was done for the winter wheat growth cycle of Yangling of 2013 and 2014. We chose these two years because it was possible to validate the results with official yield data made available by the Shaanxi Provincial Bureau of Statistics and the National Bureau of Statistics of the People's Republic of China.

For each year the biophysical variables were assimilated using the values retrieved from 3 images acquired between March and May. For each image, only the wheat fields included in the rural area of Yangling were considered. To identify the pixels in which wheat is present, 5 images acquired between February and June (one for each month) have been used. The images were converted into LAI maps through the PROSAIL-ANN algorithm. The pixels of each image have been considered wheat if the following conditions were true:
- LAI of the 1st date smaller than LAI of 2nd date and this latter smaller than LAI of 3rd date
- LAI of 3rd date higher than 4
- LAI of 4th date higher than 3.5
- LAI of 5th date smaller than 0.5 (after harvesting)

After the application of a mask to limit the assimilation only to wheat fields, each image map contained approximately 600000 pixels, which made a pixel-by-pixel assimilation not feasible. To reduce the computational cost of the algorithm, a classification of the pixels based on the LAI of the images that were used in the assimilation was employed. The images values of LAI were rounded at the first decimal place, if the number in that place was odd, it was rounded at the nearest lower even number. In this way the range between the minimum and maximum value of LAI was divided into steps of 0.2. The total number of classes is the product of the number of steps between the minimum feature value of LAI for each image and the maximum. Each class is defined by the combination of 3 numbers, 0.2 multiples, belonging to the range min LAI-max LAI of each image. In this way, after running the algorithm for all the combination of LAI, it is possible to assign a value of yield at each class. Classifying the points of map at each pixel the corresponding yield value was assigned. In this way it was possible to obtain yield maps of the region.

4. RESULTS AND DISCUSSION

4.1 Complexity and Plasticity results
The complexity of the two crop models tested was quantified using the parameter ratio (R_p). The parameters were considered relevant if the sum of EFAST Sensitivity indices was greater than an arbitrary threshold value fixed at 0.15. This value was set exclusively to quantify and compare the complexity of Aquacrop and SAFY. It resulted that the R_p for Aquacrop was 0.11, indicating that the relevant parameters were on average 11% of the total, while for SAFY a R_p value of 0.08 was obtained. The complexity of Aquacrop resulted to be higher than SAFY, i.e. the former uses a higher number of equations and input parameters than the latter, making the calibration more complex and the running time longer.

The plasticity of Aquacrop and SAFY was assessed by calculating the L index. It was found that SAFY has a plasticity index (L) of 0.169 while the L index for Aquacrop is 0.299. This means that SAFY has a higher plasticity than Aquacrop (L ranges from 0 to about 1.51, with the highest plasticity at 0). The plasticity quantifies how much the climatic conditions change the influence of specific groups of parameters on the output. A high plasticity would limit the generalization
capability of a crop model, strengthening the need for site-specific calibrations and sensitivity analyses.

4.2 LAI and Canopy Cover validation

The ANN algorithm which use for training the simulation done by PROSAIL model [12] was applied on several images acquired by HJ1A, HJ1B and Landsat 8 (Tab. 2) in 3 years (2013-2015) from March to May of each years. This procedure allowed to obtain the LAI and Canopy Cover maps for the area of interest, which were compared to field measurements for validation.

In Fig. 1 the comparison between LAI retrieved from the satellite images, by the PROSAIL-ANN algorithm, and the LAI measured in field campaigns is shown. The results indicate that the algorithm tends to underestimate the value of LAI compared to the field measurements, especially for the 2013 images. This depends on the quality of the images. The region of interest is strongly characterized by the presence of clouds and fog, so the probability for acquiring clear image is very low. The atmospheric correction for these images is fundamental and it strongly influences the results.

The estimation error is greater for higher values of LAI. In fact for LAI between 0 and 2 the difference between estimates and measures is low (Fig.1). Between 3 to 4 there is the highest error, for LAI values higher than 4 the error between simulations and measurements is lower again. The total relative root mean squared error (RRMSE) was of 0.29 %.

Fig. 2 shows the results obtained for Canopy Cover. The field measured LAI values were converted into Canopy Cover using the relationship between CC and LAI for wheat proposed by by Nielsen et al. (2011) [15]:

\[ CC = 94 \cdot (1 - \exp(-0.43 \cdot LAI))^{0.52} \] (7)

In this case the RRMSE was lower than for LAI, i.e. 10% against 29% for LAI. Saturation issues affects less the estimation of CC. So the use of CC would be preferable as a biophysical variable to be used in the assimilation.

These validation results provided information on the estimation error of the observed biophysical variables, which was considered during the assimilation.

4.3 Point application of EnKF Assimilation: comparison of two different models

As described in the paragraphs 3.5 and 3.6, the inaccessibility of the Aquacrop code did not permit the application of an updating assimilation method, as for SAFY, but it obliged to use an optimization method. The assimilation has been applied on the points where the field measurements were carried out, in this way it was possible to test the two different methods, comparing their accuracy after the assimilation and their performances.

In Fig. 3 the accuracy of the yield estimates obtained after the assimilation are shown for SAFY and Aquacrop. There was a considerable difference in the computational time. In fact, using a PC with an Intel Core i7-4770 CPU at 3.4 GHz, for the optimization method with Aquacrop it was of around 40 minutes for each simulated data point, against the 30 seconds of the EnKF with SAFY.

For these reasons, whilst the EnKF-SAFY method was applied at every field measurements points (for 3 different years), the Aquacrop optimization algorithm was applied only to a limited sample of points (Fig. 3b). Considering only the points analyzed with the Aquacrop optimization method, the RRMSE for the EnKF-SAFY method was lower to 15.3%. Therefore the accuracy of EnKF-SAFY method for a limited sample of points is higher, albeit only slightly, than that of Aquacrop optimization method.

Figure 2. Validation results for the retrieval of canopy cover (CC) from HJ1A, HJ1B and Landsat 8 images, using field measurements of 3 years in Yangling rural area. The field measured LAI was converted into CC using the equation proposed by Nielsen et al. (2011)
4.4 Regional scale application results
The spatialised application of the assimilation was carried out only for SAFY, because of the above mentioned computational constraints with Aquacrop. Fig. 4 shows the Yangling winter wheat yield maps estimated for 2013 and 2014. The mean yield calculated for 2013 was 5.395 t ha\(^{-1}\), whereas the value provided by the Shaanxi Provincial Bureau of Statistics was 5.495 t ha\(^{-1}\). For 2014, the mean yield estimated was 6.2244 t ha\(^{-1}\), while the value published by National Bureau of Statistic of the People's Republic of China was 6.13 t ha\(^{-1}\). The estimates obtained with SAFY highlight lower yields in 2013 as compared to 2014, in line with official statistics. The yield estimation error of the EnKF with SAFY for the punctual application was around 17% and it is confirmed by the regional application, where for the 2013 it was about 18% and for 2014 about 16%.

The 2013 growth season was characterized by low rainfall between October (the month of sowing) and May (senescence period) and by low temperature in winter, particularly between December and January, when temperatures were always below zero. The yield in 2013 was around the 13% lower than 2014. This yield loss could be thus explained by low winter temperatures and water stress.

The winter wheat crop was exposed to thermal stress only between December and January, while temperatures were higher than the seasonal average for the remaining part of 2013. The period of low temperatures, even below zero, concerned a rather short period, therefore this does not appear to be the main cause of loss of yield. The Yangling rural area is classified as semi-arid, rainfall occur in the summer, while winters are almost totally lacking rain. Winter wheat crops have to be irrigated, but irrigation is not always sufficient to compensate for the lack of rain, especially in the fields far from rivers and lakes in the Yangling area. These observations allow to assume that the decrease in yield between 2013 and 2014 is mainly due to drought in the winter of 2013 which affected crop stand establishment.

5. CONCLUSIONS

Results obtained with the assimilation of CC into AquaCrop, have shown the adequacy of this method to provide yield estimation as affected by water availability, even though some drawbacks appear for the spatialized applications. Its highest complexity makes the calibration more difficult, the choice of values to assign to the parameters which remained fixed (not re-calibrated) was based on the literature and on a preliminary calibration, but for regional scale applications this could be a strong limitation. Its lower plasticity means that this model is less affected by the
boundary conditions, such as weather conditions or scenario characteristics, which makes it less suitable to establish a relationship between drought and yield. Its high accuracy is its strong point, but its high computational cost makes it less suitable for a spatialized application. Results obtained from the assimilation of LAI into the SAFY model seem to indicate that the EnKF assimilation procedure performed better. Given the lower number of parameters of the model, it proved to be quite robust despite the lower complexity of SAFY. The accuracy of SAFY is lower than that of Aquacrop, but using the EnKF assimilation method it is possible to increase it, reaching acceptable accuracy, albeit slightly less than that of Aquacrop.

The application on regional scale of EnKF with the SAFY model proved to be a good method, both for the accuracy (confirmed by official statistics) and for computational requirements. To further validate this approach it would be appropriate to repeat this study on different regions, with different climatic characteristics from each other, for which it is possible to have accurate statistical information about the yield.

6. REFERENCES


