Estimation of rice production at regional scale with a Light Use Efficiency model and MODIS time series

Mirco Boschetti¹, Daniela Stroppiana¹, Roberto Confalonieri², Pietro Alessandro Brivio¹, Alberto Crema¹ and Stefano Bocchi²

¹IREA-CNR, Institute for Electromagnetic Sensing of the Environment, Via Bassini 15, 20133 Milano, Italy
²Department of Crop Science, Section of Agronomy, University of Milano, Via Celoria 2, 20133 Milano, Italy
E-mail: boschetti.m@irea.cnr.it

Abstract
This study presents the application of a Light Use Efficiency model that exploits MODIS satellite data to assess rice yield in Italy. Field experimental data acquired in 2003 and 2004 were used to define model’s parameters (Radiation Use Efficiency and Harvest Index) and to calibrate the relation between MODIS-EVI and the fraction of Absorbed Photosynthetically Active Radiation (fAPAR). Satellite images were used to spatially estimate fAPAR and key phenological rice stages that control biomass accumulation. The maps of rice yield estimates for the year 2002, 2003 and 2004 were compared to official statistics and showed a good agreement with an inter-annual relative RMSE ranging from 15 to 17%.

Keywords: Paddy rice, yield estimation, MODIS, LUE model.

Introduction
The assessment of Above Ground Biomass (AGB) is fundamental to support sustainable development, to monitor the response of vegetated ecosystems to climate change, and to provide crop yield forecast for food security policies. Moreover, in developed countries, such as the members of the European Union, timely information on crop productivity is necessary for the implementation of the Common Agricultural Policy. Crop growth models are useful for estimating growth, development and yield but they need high-quality spatially distributed input data to derive reliable estimates. Indeed, the use of remotely sensed data in crop models can improve the accuracy of yield forecast over large areas [Moulin et al., 1998; Doraiswamy et al., 2004]. This idea is strengthened by international initiatives focusing on crop monitoring, such as the Global Rice Science Partnership (GRiSP) program (http://grisp.irri.org/product-line-5-2, last access September 2011), which underlines the strategic role of remote sensing in providing spatially and temporally distributed data for modeling rice growth.

Light Use Efficiency (LUE) models [Kumar and Monteith, 1981] are often used for providing maps of biomass and crop yield since they require a limited set of input data; moreover, they are suitable for the integration with Earth Observation (EO) data. LUE models were successfully applied in different contexts and they proved to be particularly suited for the assessment of Net Primary Production (NPP) over large and remote areas where other sources of data are lacking [Bartholomè, 1990; Prince, 1991; Potter, 1993;...
Veroustraete et al., 1994; Prince and Goward, 1995].


Reeves et al. [2005] used satellite data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the NASA-Terra platform and meteorological inputs in a plant growth model [Running et al., 2000] to produce daily estimates of wheat Gross Primary Productivity (GPP). GPP estimates were then successfully used to compute crop biomass accumulation and to assess wheat yield for two states of the United States. Recently, Maselli et al. [2011] presented a method based on the use of MODIS NDVI data for estimating the cropped area and the production of winter wheat in Tuscany (Italy). Finally, it is worth mentioning two operational EO products, provided at global scale at 1 km resolution: the VGT - Dry Matter Productivity (DMP) and the MOD17 - Net Photosynthesis and Primary Productivity. The VGT-DMP is produced from SPOT-VEGETATION data (http://web.vgt.vito.be/documents/BioPar/g2-BP-RP-BP053-ProductUserManual-DMPV0-I1.00.pdf, last access September 2011) and it is currently used by the Food Security Action of the European Commission Joint Research Centre (http://mars.jrc.ec.europa.eu/mars/About-us/FOODSEC, last access September 2011) for systematic monitoring of crop and pasture production in developing countries. The MOD17 (http://modis.gsfc.nasa.gov/data/dataprod/pdf/MOD_17.pdf, last access September 2011) provides a measure of terrestrial vegetation growth and production activity; compared to the VGT-DMP, the MOD17 NPP is computed as the difference between GPP, estimated from the Montheit approach, and respiration costs, which are modelled separately.

**Objectives of the work**

The main objective of the present research was to evaluate the performance of a LUE model coupled with time series of MODIS data at 250 meter spatial resolution for the estimation of rice yield at regional scale over northern Italy. In order to achieve this goal, we:

1) evaluated the performance of MODIS vegetation indices (VIs) for the estimation of rice fAPAR by comparing them with in-situ measurements;

2) used maps of the dates of rice emergency and maturity derived from MODIS VI time series for biomass accumulation in the LUE model;
estimates the Radiation Use Efficiency (RUE or $\varepsilon$) parameter and the Harvest Index (HI) from field experimental data. The estimates of crop biomass production and yield provided by the LUE model were compared to experimental field data acquired in the period 2003-2004 and information provided by farmers. A quantitative accuracy assessment of regional yield estimates was also conducted by comparing model results with official statistics from the Ente Nazionale Risi (ENR) (http://www.enterisi.it/ser_statistiche.jsp, last access September 2011).

**Materials**

**Study area**

Figure 1 shows the study area within the main Italian Rice District [Bocchi et al., 2003], located in Northern Italy and delimited by the Provinces of Milan, Pavia, Vercelli and Novara. Rice represents the major crop in this area, with 90% of the total Italian rice production, in turns representing 60% of the total European production. The Eastern part of the Italian Rice District, which includes rural areas of the Milan and Pavia provinces and belongs to the Ticino River Regional Park and to the South Milan Agricultural Park, was monitored via direct field measurements. In two intensively rice cultivated areas, Besate (45° 19’N; 8° 57’E) and Opera (45° 23’N, 9° 11’E), experimental field campaigns were carried out in 2003 and 2004, respectively. Figure 1 presents the location of the farms where secondary data\(^1\) were collected via farmer’s interviewing and the validation sites monitored in 2005 in Opera, Carpiano (45° 34’ N, 9° 27’ E) and San Giuliano (45° 24’ N, 9° 17’ E) municipalities.

---

\(^1\)Bastiansen and Ali [2003] classified the available data as primary when they arrived from field experiment and secondary when they comes out from interviews. This second category is more representative of real domain situations but it is difficult to assess their quality because they are not retrieved by controlled situation such as those produced by self-measurements in field.
Rice production Estimation with MODIS

Rice information and data
Table 1 summarizes the dataset used for this study which comprises:
1) Field experiments:
   − detailed agronomic field measurements to define model’s crop parameters [Boschetti et al., 2006];
   − field spectral data collected to study VIs fAPAR relation for rice [Stroppiana et al., 2009].
2) Farm information:
   − field measurements of rice LAI at farm level for calibrating and validating the satellite-based fAPAR estimations;
   − available geo-coded information on cultivated varieties, fields extent, sowing dates, management and final yields obtained by interviewing nine rice farmers.
3) Regional information:
   − official statistics provided by ENR over the entire Italian Rice Area available as aggregated at administrative level.

All data were assigned, according to the crop characteristics and the length of the growing season (sowing to maturity), to three main classes of rice variety: Indica (I), Japonica Early (JE) and Japonica Medium-late (JM) [Boschetti et al., 2009]. This classification, based on the morpho-physiological characteristics of rice, was proposed by Confalonieri and Bocchi [2005] to group the different varieties for crop model calibration and application. This simplified classification is useful for large scale simulations of rice growth because it allows to account for morpho-physiologic heterogeneity by using only three sets of parameters.

Field campaigns
In 2003 and 2004 experimental plots were set up for acquiring field spectral data, LAI measurements and agronomic data. The data collected allowed the estimation of rice crop parameters such as RUE and HI and to study the relation between vegetation indices and LAI/ fAPAR. A detailed description of the 2003-2004 field campaigns can be found in Boschetti et al. [2006] and Stroppiana et al. [2009]. Specifically for studying the MODIS-VI fAPAR relation with real satellite data, we measured LAI in rice fields greater than 2 ha which were contiguous to the experimental plots and covering a total area of about 27 ha in 2003 and 14 ha in 2004. These measurements allowed the extraction of representative mean LAI values to be compared with MODIS vegetation indices at 250 m pixel resolution. Mean values were calculated in order to take into account the natural spatial variability of paddy rice and the possible error in geolocating the single LAI measurements on MODIS pixels.

The 2005 field campaign aimed to collect data for validating the VI-fAPAR relationship. Bi-weekly measurements from tillering to flowering (n = 7) of LAI were collected with the LAI2000 instrument on five plots (plot size between 3 ha and 6 ha) within three rice farms characterized by large paddy rice areas easily identified on the MODIS imagery (Fig. 1); plot measurements were aggregated for each farm summing up a total of 21 samples (7 measurements x 3 farms) available to validate the results. We acquired LAI measurements along 20-meter transects. For each transect, four measurements were collected in diffuse radiation conditions (at sunset or on overcast days) with the LAI2000 instrument using single-sensor mode with the following sequence: one above the canopy at the beginning of the transect, four measurements regularly distributed (4 in 20 m) under-canopy and again one measurement above at the end of the transect.
Table 1 - Available data for the model calibration and validation.

<table>
<thead>
<tr>
<th>Measured variable</th>
<th>Yield</th>
<th>LAI</th>
<th>Biomass</th>
<th>RUE</th>
<th>HI</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Aim

<table>
<thead>
<tr>
<th>Data Set</th>
<th># data</th>
<th>Period</th>
<th>Location</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Field experiment

The acquisition protocol and the post-processing of the LAI2000 measurements are described in detail by Stroppiana et al. [2006]; in particular, we decided to discard the 4th ring measure to produce more accurate LAI estimates. For each homogeneous area of 2 × 2 MODIS pixels, a mean LAI value was obtained by averaging all the transect measurements acquired in the different fields.

The CORINE Land Cover Map
A detailed digital land cover map represents a fundamental data to analyse coarse/moderate remotely sensed images, thus allowing to localize the different cultivated surfaces. Several
publications indicate that the resolution and accuracy of crop maps is a strong limitation for the application of crop growth models [Turner et al., 2002; Bastiaanseen and Ali, 2003; Lobell et al., 2003; Doraiswamy et al., 2005]. However, the extent of rice fields in the Italian Rice District, where rice is often the solely crop and it can cover more than 90% of the agricultural land, is such that homogenous fields can be identified also at the MODIS spatial resolution. Moreover, in those highly specialised cropping systems, fields often do not present any crop rotation. As a consequence, rice fields remain the same for several years and a recent land cover map can be considered representative of the current situation. The CORINE Land Cover (CLC) digital map has a spatial resolution of 100 m that is well suited to analyse MODIS 250 m images; in particular, we used the map produced in 2000 at the scale of 1:100,000 (http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-raster-1, last access September 2011) to identify the distribution of rice fields within the study area.

**Meteorological data**

Meteorological data were derived from the meteorological stations available in the study area. These stations include the national meteorological network of the Ufficio Centrale di Ecologia Agraria (UCEA; http://old.politicheagricole.it/ucea/forniture/index3.htm) and the regional services of the Agenzia Regionale Protezione Ambiente Lombardia (ARPA) and of the Rete AgroMeteorologica del Piemonte (RAM). Other stations were located in experimental sites for the year 2003-2005 (Agriculture Faculty, State University of Milan and Institute for Environment and Sustainability, Joint Research Centre). For the entire study area, daily mean air temperature and global solar radiation were estimated by spatial interpolation of the available meteorological recordings. The interpolation procedure used was a weighted mean of the values collected by the available surrounding stations, with the weight inversely proportional to the square of the distance. Since the topography of the study area does not present significant altitude differences, altitude and aspect were not taken into account in the interpolation procedure. It is important to remind that the solar radiation measurements can be affected by uncertainty due to, for example, sensor degradation; however, the weighted spatial interpolation adopted could compensate for measurement inaccuracy.

**Remotely sensed data**

Multi-temporal analysis has been based on the 250-m MODIS Vegetation Indices coded as TERRA product MOD13Q1 validated version collection V004 (http://edcimswww.cr.usgs.gov/pub/imswelcome/). Four complete years from 2002 to 2005 of MODIS data were downloaded from the Land Processes Distributed Active Archive Center (LP DAAC) of the U.S. Geological Service (USGS). More details on the characteristics of the MODIS sensor can be found in Tucker and Yager [2011] in this special issue. The MOD13Q1 (http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf, last assess September 2011) product contains 16-day temporal composites of the surface reflectance of MODIS bands 1-3 and 7 (blue, red, NIR and shortwave infrared spectral regions), NDVI and Enhanced Vegetation Index (EVI) as well as VIs quality information. The EVI was specifically developed for the MODIS sensor [Huete et al., 2002] and, compared to NDVI, offers a better sensitivity in high biomass regions, it de-couples the signals from the canopy and the background and reduces the atmospheric influence on the signal. In order to overlay farms field boundaries
and GPS positions of ground measurements, the MODIS dataset was re-projected to UTM 32 WGS84 coordinate system using the MODIS Re-projection Tool (https://lpdaac.usgs.gov/tools/modis_reprojection_tool, last access September 2011).

**Methodology**

**LUE model**

When water and nutrient availability are not limiting factors for plant growth, Above Ground Biomass (AGB), or NPP, can be estimated as a function of the absorbed Photosynthetically Active Radiation (PAR in the wavelength range 400 to 700 nm), Radiation Use Efficiency (RUE) and a factor accounting for thermal limitations to photosynthesis [1]. A LUE model is often also called Radiation Use Efficiency model however, in this paper we use the term LUE to identify the model and RUE to refer to the parameter ε* of Monteith equation.

\[
AGB = \sum_{d=\text{SOS}}^{\text{Maturity}} \varepsilon^* \times 0.45 \times \text{Rad}_d \times f\text{APAR}_d \times T_{\text{lim}d} \quad [1]
\]

where:
- AGB (g m\(^{-2}\)) is the total dry matter production computed from daily AGB cumulated between start of season (SOS) and maturity;
- ε* is the potential RUE (g MJ\(^{-1}\)) not affected by limiting factors [Prince, 1991];
- Rad\(_d\) (MJ m\(^{-2}\) day\(^{-1}\)) is the daily global solar radiation that, multiplied by 0.45, can be assumed as a proxy for the incident PAR;
- fAPAR\(_d\) (-) is the daily fraction of PAR absorbed by vegetation that varies daily as a function of canopy characteristics and plant development;
- T\(_{\text{lim}d}\) (-) is the daily air temperature-dependent limiting factor;
- d is the day.

The dates of start of season (SOS) and maturity were derived by analyzing MODIS NDVI time series as described in Boschetti et al. [2009]. The T\(_{\text{lim}}\) function was determined by using a linear function and setting the proper cardinal temperatures [2].

\[
\begin{align*}
T_{\text{lim}} &= 0 \quad \text{when } T_a \leq T_b \\
T_{\text{lim}} &= 0.0588 T_a - 0.6471 \quad \text{when } T_b < T_a \leq T_{\text{opt}} \\
T_{\text{lim}} &= 1 \quad \text{when } T_{\text{opt}} < T_a \leq T_c \\
T_{\text{lim}} &= 0 \quad \text{when } T_a > T_c
\end{align*}
\]

[2]

where:
- T\(_a\) is the daily average air temperature;
- T\(_b\), T\(_{\text{opt}}\) and T\(_c\) are the base, optimum and cutoff temperatures for growth.

In this study, T\(_b\), T\(_{\text{opt}}\) and T\(_c\) were set to 11°C, 42°C and 28°C [Confalonieri and Bocchi, 2005]. Crop yield can be estimated from the total AGB by applying the specific dry Harvest Index (HI) and by considering the relative grain humidity of commercial rice [3].
\[
\text{Yield} = \frac{\text{AGB} \times \text{HI}}{1 - m_0} \quad [3]
\]

where:
- \( \text{Yield} \) (g m\(^{-2}\)) is the grain mass commercially produced;
- \( \text{HI} \) is the ratio of dry grain mass to aboveground biomass specific for each group of rice cultivars;
- \( m_0 \) is the water content present in commercial grains, here considered equal to 0.12 (12\%, ENR communication).

**RUE and HI calibration**
The maximum RUE (\( \varepsilon^* \)), not limited by stressors, was calculated using field data from seven agronomic experiments carried out on different rice cultivars and by regression between periodic AGB samplings and cumulated APAR [Boschetti et al., 2006]. Statistical analysis showed that, when temperature limitation (\( T_{\text{lim}} \)) is taken into account, there is no statistical difference between the different rice cultivars thus, \( \varepsilon^* = 2.9 \text{ g MJ}^{-1} \text{ m}^{-2} \) was used for all the three groups of (I, JM and JE) [Boschetti et al., 2006].

HI values estimated in controlled field experiments are generally higher than those measured on commercial farms. Therefore, specific HI values were calibrated for the three rice functional groups in order to be more representative of real farming condition. Bastiaanseen and Ali [2003] proposed to use data derived from farmers’ interviews to define specific HI values that best fit the model yield predictions. According to this approach, three HI were calibrated by minimization of the Root Mean Squared Error (RMSE) between the yield data provided by the farmers and the AGB estimates provided by the LUE model. The tolerance range for the HI adjustments was set to the min-max values derived by Italian field experiment [Boschetti et al., 2006] and provided by experts personal communication.

**GIS analysis**
The estimates provided by the model for the 2002-2004 period were aggregated at the level of administrative units in order to be comparable to the ENR statistics. Following the method proposed by Reeves et al. [2005], for each administrative unit a weighted mean yield was calculated where the weight for each group is computed based on the fraction of area occupied within the unit.

The final rice biomass map was derived by taking into account only the MODIS pixels assigned to the rice agricultural class by the CLC map. Since this map does not provide specific spatial information on different cultivated varieties and recent varieties (e.g., Indica, Japonica early) have higher HI compared to the traditional (low productive Japonica medium late genotypes), we estimated spatially distributed HI values though GIS analysis. A weighted HI (\( \text{HI}_w \)) for each administrative unit, derived from the intersection of the ENR information and rice class of CLC map, was calculated on the basis of the percentage of surface occupied by the different rice group. This value was multiplied by the cumulated AGB estimated by the LUE model to derive mean yield per administrative unit.

**fAPAR model**
We defined a semi-empirical model [Turner, 2002; Kiniry et al., 2004] to estimate fAPAR
from MODIS data. Although the traditional approach exploits NDVI [Asrar et al., 1984; Myneni and Williams, 1994], we preferred to use MODIS-EVI, less affected by saturation effects compared to NDVI in high biomass ecosystems and by the influence of the background [Baret et al., 1989; Myneni et al., 1995; Gao et al., 2000]. However, the analysis of MODIS NDVI was also conducted as a comparative exercise to analyse the different VIs performance on real MODIS satellite data.

The MODIS VIs vs fAPAR relations were also compared with those derived from field spectral data. In order to simulate the response of the MODIS sensor, the reflectance spectra collected on the ground with the Field Spec FR at 1 nm spectral resolution were resampled to the blue, red and NIR band equivalent reflectance of the MODIS sensor. Field daily fAPAR measurements simultaneous to the MODIS EVI composite images were extracted from the dataset of field data. Daily LAI time series were derived by interpolating field data using a sigmoid function in the form of $\text{LAI} = a / \{1 + \exp[-(\text{DOY} - x_0)/b]\}$, where $a$, $x_0$ and $b$ are the equation parameters and DOY refers to the Day Of the Year. Coefficient of interpolated regression functions for green LAI were always significant ($p < 0.001$) and the relationship presented always $R^2_{adj}$ greater than 0.94 and a Standard Error of Estimate lower than 0.32. The corresponding fAPAR series were estimated using the Lambert Beer’s formula [Monsi and Saeki, 1953] [4].

$$f\text{APAR} = 1 - e^{-k\text{LAI}} [4]$$

A variety-specific value of the coefficient of light extinction $k$ [Boschetti et al., 2006] was derived as a function of development stage.

An example of field LAI values, interpolated LAI time series and fAPAR for rice fields of Besate 2003 and Opera 2004 are presented in Figure 2.

Figure 2 - Field mean LAI values (black spots), the interpolated LAI profile and the corresponding fAPAR profile for 2003 (Besate rice fields) (a) and 2004 (Opera rice fields) (b).

Following the approach proposed by Turner et al. [2002], the relation between MODIS EVI and fAPAR was obtained by extracting from the 16-day composite images the temporal VI signal of rice. The corresponding fAPAR were derived using the function previously described by selecting the values at the end of the 16-day composite window as suggest by Turner et al. [2002] and Kiniry et al. [2004]. A temporal smoothing procedure based on Savitzky–Golay polynomial filter function [Chen et al., 2004] was applied to the original
16-day composite VI data (EVI and NDVI) in order to eliminate spurious data which affect the time series and to produce an interpolated daily VI temporal profile. The 2002, 2003 and 2004 daily fAPAR profiles for the entire study area were then created by applying the predictive relation VI vs fAPAR on the daily smoothed VI series.

Results and discussion

fAPAR time series

The linear regression between MODIS-EVI and field fAPAR resulted statistically significant (n = 17 for the two years of data pooled together; *** $p<0.001$) with a high coefficient of determination ($R^2 = 0.82$). This relationship (Fig. 3a) is in agreement with that obtained from field spectroradiometric data by Stroppiana et al. [2009]. The EVI-fAPAR relationship derived by satellite data falls within the 99% confidence limit of the one derived from field observations. These results confirm the general validity of the satellite-derived relationship. The same analysis was carried out also for NDVI and the relationship resulted significant (n = 17 for the two years of data pooled together; *** $p<0.001$) with a very good coefficient of determination ($R^2 = 0.94$) but it is significantly different from the one developed from field data (Fig. 3b). The NDVI-fAPAR model derived is similar to others previously published however the relation coefficients, for both field and satellite data, are different from the theoretical one provided by Myneni and Williams [1994] (slope = 1.1638 and intercept = -0.1426) and the ones used by Eerens et al. [2000] to derive fAPAR from SPOT-VGT data (slope = 1.54 intercept = -0.247) and from NOAA-AVHRR (slope = 1.68 and intercept = -0.27). On the contrary, it is interest to notice the similarity of the coefficient derived by field measurements by Fensholt et al. [2004] (slope = 1.51 and intercept = -0.40) with the ones derived in field by our experiment (slope = 1.41 and intercept = -0.49).

The difference observed in the VI-fAPAR relations shown in Figure 3 confirms that changes in NDVI are likely due to factors other than crop conditions (i.e., soil or background conditions) thus reducing its reliability for monitoring activities. Validation was performed using independent field fAPAR measurements collected in 2005 and MODIS VI data (n = 21 from different locations).

![Figure 3 - Comparison between fAPAR-EVI (a) and fAPAR-NDVI (b) relations derived from field data and satellite observations.](image-url)
Results showed that fAPAR estimated from the EVI model has a good agreement with field data although the satellite-based model seems to slightly underestimate the biophysical parameter. The NDVI model, as expected, presents a problem for low fAPAR values due to the influence of background in periods preceding the close canopy stage, but it does not underestimate high fAPAR values due to saturation. The indices of agreement [Martorana and Bellocchi, 1999] calculated during the validation of the two VI-based models identify EVI-fAPAR as the best model: RMSE = 0.13 (0.17 for NDVI), rRMSE = 20% (24% for NDVI) and modelling efficiency (EF) 0.78 (0.65 for NDVI) close to optimum values. The derived equation based on EVI was used to spatialize fAPAR over the entire study area for the growing seasons of 2002, 2003 and 2004.

**Above Ground Biomass estimation**

Biomass production was estimated by applying [1] for each pixel on a daily basis and by cumulating daily AGB values between the start of season (SOS) and physiological maturity which were available spatially distributed from the map produced by Boschetti et al. [2009]. These phenological dates estimates resulted in agreement with those simulated by a traditional approach based on meteorological measurements, resulting in an average error of only one week [Boschetti et al., 2009]. The use of these data allowed us to dynamically set the start and the end of the period of biomass accumulation rather than relying on fixed dates. Indeed, in the study area agricultural practices can vary from one farm to another thus leading to the choice of different rice varieties and therefore different crop calendars. In our previous study we observed a difference of more than 20 days between the sowing date of the traditional and new varieties. Figure 4 shows daily AGB estimated by the LUE model (black continuous line) for the growing seasons of 2003 and 2004 compared to biomass measurements from field campaigns (grey circle markers).

![Figure 4](image-url)

\textbf{Figure 4} – The profile of estimated AGB based on the LUE model (continuous line) compared to field biomass samplings acquired in Besate 2003 (a) and Opera 2004 (b) (grey point); error bars represent standard deviation of field measurements.

The model, thanks to the daily estimation of the biophysical variable (fAPAR), is able to accurately reproduce rice growth, especially during the stem elongation phase, from DOY 180 (200) to DOY 200 (230) in 2003 (2004). At the first stages of the crop cycle, the LUE model overestimates biomass production, especially in the 2003 dataset; it is likely that, as
the signal from the canopy increase with the growth of the plants, the fAPAR estimation is more reliable. Field biomass measurements presented in Figure 4 were collected in a single experimental field covering only a portion of the MODIS 2 by 2 pixel window used for deriving VI values which could be instead the average response of several fields. Therefore, these results do not represent a real validation exercise but rather a qualitative evaluation of the model temporal behavior.

**HI calibration and evaluation**
Table 2 compares HI derived from field experiment for the I and JM varieties [Boschetti et al., 2006], values calibrated for each group and HI values proposed in the literature.

Table 2 - Parameters used for model application compared to international literature values. HI values were obtained by a calibration procedure.

<table>
<thead>
<tr>
<th>Rice group</th>
<th>RUE\textsubscript{max} (g MJ\textsuperscript{-1})*</th>
<th>HI\textsubscript{dry} (Kg/Kg)**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Field exp.</td>
<td>Mean</td>
</tr>
<tr>
<td>I</td>
<td>3.30</td>
<td>0.53</td>
</tr>
<tr>
<td>JE</td>
<td>2.80</td>
<td>2.2-5.0</td>
</tr>
<tr>
<td>JM</td>
<td>2.60</td>
<td>0.39</td>
</tr>
</tbody>
</table>

*RUE\textsubscript{max} describes the potential maximum rue non affected by stressors

** HI\textsubscript{dry} corresponds to the ratio between oven dry biomass and grain

*** information derived from other field experiment or provided by expert

The calibrated HI ranges from 0.38 to 0.50, values which are in agreement with those proposed in the international literature and with those usually suggested for Italian rice. However, calibrated HI is lower than estimates from the 2004 experimental dataset, when rice grew in optimal condition (production level 1; Penning de Vries et al. [1989]). However, lower values should better depict ordinary, hence not always optimal, agronomic conditions.

Figure 5 shows the comparison between yields declared by farmers and estimates obtained from the MODIS-LUE model. The model resulted efficient (EF = 0.41) and provided accurate estimates (rRMSE = 9.7% and RMSE = 0.58 t/ha).

**Regional maps of biomass and yield**
Figure 6a shows pixel-based AGB estimates for 2004 over the study area at the full spatial resolution of the MODIS data (250 m); estimates range from 10 t/ha (blue colours) to a maximum of 18 t/ha (red colours). The more productive areas (South-East and North West) correspond to the zones where are cultivated the traditional varieties (i.e. Arborio, Carnaroli, Volano, Vialone nano) whose morphological characteristics (plants are taller than the new varieties) determine a greater amount of total biomass.

However, we observed that the greatest AGB estimates are mainly distributed along the boundary of the study area as provided by the Corine Land cover map. It is likely that in these area some boundary effects, such as contamination from non-agricultural lands, produce less reliable AGB figures. Hence, part of the uncertainty in the satellite-based
estimates might be driven by the coarse spatial resolution of the MODIS data and by the thematic accuracy of the CLC map which should be comparable with the heterogeneity and fragmentation of the observed surfaces. However, none of these problems was observed in homogeneous rice area such as the central part of the Vercelli province. The selection of pure MODIS pixel (i.e. > 80% of rice) or the use of un-mixing techniques to extract uncontaminated rice signal could be applied to reduce the influence of different cover types on the temporal profiles of the VI.

Figure 5 - Model estimates vs yield values provided by the farmers for three studied years 2002, 2003 and 2004. The values are grouped referring to the farm administrative locations - Basile (BAS), Carpiano (CAR), San giuliano Milanese (SAN) e Besate (BES) – and cultivated variety - Indica (I), Japonica Early (JE) and Japonica Medium-late (JM).

Figure 6b presents an example map of the rice yield estimates for 2004 aggregated per administrative unit. Rice grain yield estimates range from 4.0 to 7.3 t/ha, which are reasonable figures for the rice production in Italy. The most productive regions at the center of the study area (green patches) belong to the areas cultivated with the new varieties (I or JE). It is interesting to observe that these areas correspond to relatively low AGB (Fig. 6a) which does not necessarily implies low yields if the HI is high such as in the case of these new varieties. In fact, lower yields are spatially relegated to the south eastern part of the district, where traditional varieties are predominant.

Validation

Compared to traditional average statistics, thematic maps derived from the integration of satellite data within an appropriate model supply information about the spatial variability of the analysed system. The relevance of such information is even more evident considering that it is accompanied by the uncertainty analysis. It is widely acknowledged that any thematic map is worthless without proper accuracy assessment [Foody, 2002].
Figure 6 – The map of estimated AGB at the MODIS 250 m pixel resolution for 2004 (a). Rice yield at the level of administrative units (b) and map of percentage errors of yield estimates (c).
For this reason, a validation exercise was carried out for the three year period 2002-2004 at the administrative level: yield estimates were compared to those provided by the ENR official statistics. For each year, indices of overall accuracy/error, MAE (Mean Average Error), RMSE and rRMSE were calculated. Moreover, the Percentage Error (PE) expressed as [(Estimated –Observed)/Observed]*100 was calculated for each spatial unit and represented as a map. Results show a satisfactory agreement between MODIS yield estimates and ENR official statistics. RMSE (MAE) for 2002, 2003 and 2004 data are respectively of 0.86 (0.69), 0.96 (0.74) and 0.88 (0.68) t/ha, corresponding to a rRMSE of 15.3, 16.4 and 15.0%. The analysis of the PE maps (Fig. 6c shows the 2004 map) highlights similar patterns for the three analysed years: an overestimation in the South-East areas (Milano and Pavia Provinces, see Fig. 1) and in the extremely North-West part of the study area (Vercelli Province), and a slight underestimation in the central part (Novara Province) of the main Italian Rice District. A possible reason for the overestimation along the borders of the study area could be due to the above discussed issues of the mixture of the coarse MODIS pixels. PE histograms (Fig. 7) show a normal distribution centred on zero with more than 25% of the data in the error range 0-10 %; another 50% falls in the range 10-20%. The histograms are right skewed indicating a general overestimation of the model. This situation is particularly evident for 2003 data: this year was the hottest of the century (i.e., high solar radiation and air temperature) and for the simplified assumption of the LUE model those conditions correspond to high biomass production but not necessary correspond to high yields. This observation underlines that despite the satisfactory estimates, further improvements of the model are required to take into account other stressors such as saturation of the enzymatic chains under high radiative conditions and RUE decrease due to the reduction of stomata conductivity when high vapour pressure deficit (VPD) values occur.

Moreover, there is scientific evidence [Hasegawa and Horie, 1996; Confalonieri et al., 2007] that RUE decreases from its maximum value after flowering and, during the ripening
phase, until maturity while in the present research a constant RUE value was used for
the entire growing season. A RUE correction factor as a function of the development
stage could represent a further improvement of the model proposed. Finally, as quoted
by Kiniry et al. [2001], crop modelling needs a more realistic description of HI due to
varietals differences and environmental effect on actual rice production, factors not taken
into account in this work. A description of HI based on the simulation of the processes
that determine the number of panicles, seeds per panicle, weight per seed could provide
a better yield estimation.

Conclusions
The objective of the present study was to evaluate the performance of a LUE model based
on MODIS satellite data for the estimation of the production of paddy rice in Northern Italy.
The Italian rice district is an ideal case study because over an almost contiguous area of
about 200,000 ha more than 30 rice varieties are cultivated and the production covers 90%
of the total Italian yield. The main results achieved with this study are:

− A predictive relationship between EVI and fAPAR which proved to be more reliable than
the one based on NDVI. The equation applied pixel by pixel allowed the estimation of
the spatial and temporal distribution of the biophysical parameter fAPAR that governs
the plant capacity of absorbing radiation.
− The implementation of spatially distributed crop phenological information retrieved
from the MODIS time series that allowed to dynamically set the start and end of the
period of biomass accumulation.
− The assignment of an Harvest Index specific for each of the three morpho-physiological
groups.
− The integration of the above mentioned results in a LUE model for rice yield estimation
which provided reliable biomass estimates in terms of both temporal trend and absolute
values for the three year period 2002-2004. Rice yield was estimated with rRMSE
below 20%.

Finally, the analysis of the spatial distribution of the percentage error suggested that in
some areas AGB overestimation could be due to inaccuracy of the thematic Land Cover
map and its spatial heterogeneity with respect to MODIS spatial resolution of 250 m.
A deeper analysis should be carried out to assess the influence of spatial heterogeneity
on the estimates provided by the model. The use of detailed agriculture census data at
cadastral level, currently available and yearly updated, could be a solution to update
thematic agricultural information. Despite our satisfactory results, in the future we
would like to improve the model by reducing errors due to the adopted simplifications
which do not take into account other potential biotic and abiotic stressors acting on
plant productivity.

Acknowledgements
Authors wants to thank all the colleagues that provide a precious helps in conducting field
campaign. In particular we want to acknowledge Mr. Mauro Musanti and Dr. Giacomo
Parma for the costly LAI2000 measurements. A particular thanks to Mr. Fedeli that guests
our experiments in his farms and to Bruno Marabelli (ENR) for providing the official
statistics of Italian Rice production.
References


4257(94)90016-7.


**Received 27/06/2011, accepted 18/07/2011**