RESEARCH ARTICLE

Coupling remote sensing with a water balance model for soybean yield predictions over large areas

Daniela F. Silva Fuzzo[1] · Toby N. Carlson[2] · Nektarios N. Kourgialas[3] · George P. Petropoulos[4,5]

Abstract

In this study a new method for predicting soybean yield over large spatial scales, overcoming the difficulties of scalability, is proposed. The method is based on the so-called "simplified triangle" remote sensing technique which is coupled with a crop prediction model of Doorenbos and Kassam 1979 (DK) and the climatological water balance model of Thornthwaite and Mather 1955 (ThM). In the method, surface soil water content (Mo), evapotranspiration (ET), and evaporative fraction (EF) are derived from satellite-derived surface radiant temperature (Ts) and normalized difference vegetation index (NDVI). Use of the proposed method is demonstrated in Brazil's Paraná state for crop years 2002–03 to 2011–12. The soybean crop yield model of DK is evaluated using remotely estimated EF values obtained by a simplified triangle. Predicted crop yield by the satellite measurements and from archived Tropical Rainfall Measuring Mission data (TRMM) and European Centre for Medium-Range Weather Forecasts (ECMWF) data were in good agreement with the measured crop yield. A "d_2" index (modified Willmott) between 0.8 and 0.98 and RMSE between 30.8 (kg/ha) to 57.2 (kg/ha) was reported. Crop yield predicted using EF from the triangle were statistically better than the DK and ThM using values of the equivalent of EF obtained from archived surface data when compared with the measured soybean crop data. The proposed method requires no ancillary meteorological or surface data apart from the two satellite images. This makes the technique easy to apply allowing providing spatiotemporal estimates of crop yield in large areas and at different spatial scales requiring little or no surface data.

Keywords Soybean yield modeling · Satellite measurements · Remote sensing · Evapotranspiration · Crop yield in large areas · Triangle method · Geospatial data analysis techniques

Introduction

Soybean crop is one of the most economically significant crops worldwide (Mammadov et al. 2018). As such, establishing an easy to apply and time-effective modeling approach for predicting soybean yield is essential for stable markets at global scale (Pagano and Miransari 2016). In Brazil in particular, soybean crop has been the key crop in the increase of the area and grain

✉ Daniela F. Silva Fuzzo
silva.danielaf@gmail.com

Toby N. Carlson
tnc@psu.edu

Nektarios N. Kourgialas
kourgialas@nagref-cha.gr

George P. Petropoulos
petropoulos.george@gmail.com

[1] Paulista State University "Julio de Mesquita Filho", Department of Geography, Renato da Costa Lima, 451 - Ville de France, 19903-302 Ourinhos, SP, Brazil

[2] PennState University, 604 Walker Building, University Park, PA 16802, USA

[3] Institute for Olive Tree Subtropical Crops and Viticulture, Water Recourses-Irrigation & Env. Geoinformatics Lab., NAGREF - Hellenic Agricultural Organization (H.A.O.-DEMETER), Chania, Greece

[4] Department of Soil Water Resources, Institute of Industrial & Forage Crops, Hellenic Agricultural Organization (HAO) "Demeter" (former NAGREF), Directorate General of Agricultural Research, 1 Theofrastou St., 41335 Larisa, Greece

[5] School of Mineral Resources Engineering, Technical University of Crete, 73100 Chania, Crete, Greece

production, totaling a production of 111.5 million tons for the 2018 harvest year alone. In order to determine agricultural production, crop yield models are often used. Those models usually consider meteorological variables, agricultural practices, and soil parameters as the main physically-driven conditions for the agricultural cycle modelling (Melo et al. 2008; Ferreira and Rao 2011; Gusso et al. 2017). Specifically, for soybean multiple linear regression models are often utilized to explain yield variations (Tao et al. 2008; Mercante et al. 2010; Dalposso et al. 2016).

The most common approach for crop yield estimation is based on the crop water production functions (CWPF). CWPF relate crop yield and water in terms of evapotranspiration and applied water. Crop yield models can also incorporate Geographical Information Systems (GIS) and geospatial data analysis techniques to analyze agrometeorological data (Berka and Rudorff 2003). In addition, artificial neural networks (ANNs) are widely applied for estimating crop yield. This is due to their ability to model highly nonlinear systems in which the relationships between variables are unknown or very complex (Goyal 2013; Safa et al. 2015; Kourgialas et al. 2017). Recently, studies have evaluated the use of ANNs as an estimating tool of the productivity of soybean cultivars based on different agricultural practices (e.g. Alves et al. 2018). Some of the key challenges that may prevent scalability of such modeling approaches include the difficulty of establishing an extensive monitoring network on large spatial scale and long computational times of models. Moreover, meteorological data are often not available at the same time and spatial scale, while the aggregation of agrometeorological parameters at large spatial scale is leading to high uncertainty in modeling simulation (Sims et al. 2008).

To overcome these challenges, satellite remote sensing datasets could be used in synergy with ground-based observations (Petropoulos and McCalmont 2017; Bao et al. 2018). Specifically, the use of simplified spectral models based on the crop cycles has shown a promising potential in crop harvest prediction (Mercante et al. 2010; Gusso et al. 2013). Remote sensing can be used to obtain key parameters in large areas that can be included in crop productivity models (Huang et al. 2002). Multi-temporal satellite data can provide information on long-term changes in soil moisture or soil nutrients, surface heat flux, and evapotranspiration (ET) (Lei and Yang 2010; Evett et al. 2012; Liou and Kar 2014; Pandey et al. 2015; Petropoulos et al. 2016; Petropoulos et al. 2018a, b). Such information can be used in many fields such as climatology, meteorology, agriculture, hydrology, and geography. Estimates and map-based land surface parameters (such as available surface moisture and evapotranspiration in large areas) are necessary for efficient agricultural management and for predicting crop drought risk in large agricultural areas. Such decision making systems are extremely effective and low cost (Feng et al. 2014; Yang et al. 2015).

Several studies provide crop yields from conventional weather station data. Most of these studies are based on timely climatic data, usually derived from small experimental plots (Prasad et al. 2006; Er-Raki et al. 2007). Other researches seek to complement the mathematical models using remote sensing data and surface parameters (e.g. Meyer 1990; Pablos et al. 2017). Yet, it is sought to adapt the models using only obtained variables in orbital images, facilitating the coverage of large agricultural areas in a short period of time and at low cost. Recently, the reliability of the physiological meaning of the enhanced vegetation index (EVI) for the development of a remote sensing-based procedure to estimate soybean production prior to crop harvest was evaluated (Gusso et al. 2017). The key idea behind all these techniques is that surface radiant temperature (Ts) - and by association the surface turbulent energy fluxes - are sensitively dependent on the surface soil water content (Mo). A variety of approaches have been developed for deriving spatiotemporal estimates of Mo and the surface turbulent energy fluxes from satellite-derived Ts (see e.g. review by Petropoulos et al. 2009).

A widely used group of remote sensing based approaches uses the triangle configuration of pixels in Ts and fractional vegetation cover (Fr) space. This method is referred to as "the triangle model" (Gillies and Carlson 1995; Gillies et al. 1997; Carlson 2007; Minacapilli et al. 2016; Zhu et al. 2017). The triangular method allows expressing changes in land use, such as trajectories within a triangular domain with all surface moisture content, Fr and Ts. Lambin and Ehrlich (1996) were between the first to introduce the use of trajectories within the triangular domain to quantify the temporal movement of pixels associated with changes in soil use and cover. A series of studies have documented the triangular or trapezoidal shape related to the data of images obtained by remote sensing (Bastiaassen et al. 2005; Stisen et al. 2007; Petropoulos et al. 2009). Because of the triangle configuration of pixels in Ts/Fr space (Huang et al. 2002), the boundaries of the triangle, constrain the solution for soil moisture and evapotranspiration (ET). The basic idea of this method is that surface spatial heterogeneity soil moisture is implied by variations in Ts (Carlson 2007). Recently, Carlson (2013) and Carlson & Petropoulos (2019) proposed a new and simple version of this triangle method to obtain Mo and ET.

This study proposes a new method for predicting soybean crop yields, which utilises the triangle remote sensing method and the crop prediction model of (Doorenbos and Kassam 1979), (DK). DK requires knowledge of evapotranspiration fraction (EF), defined as the ET divided by the net radiation (Rn). Widely, this ratio is obtained from climatological water balance models (CWBs) (Thornthwaite and Mather 1995), which require a series of archived surface data. The methodology proposed here presents an advance over existing crop yield estimation methods, since the EF is derived by simple remote sensing techniques. Specifically, for assessing soybean crop yields, a simplified triangle technique is incorporated, overcoming the difficulties of scalability, giving cost-

effective predictions at large spatial scale. Use of this new method to predict soybean crop yield is demonstrated herein over a large area of Brazil.

Materials and methods

Study area

The study area encompasses the state of Paraná, located in the southern region of Brazil between latitudes 22°29'S and 26°43'S and meridians 48°2'W and 54°38'W. The area is characterized by temperate climate with well-distributed rainfall and hot summers. In winter, the average temperature is below 16 °C, while in the summer the maximum exceeds 30 °C. Paraná is located in a region of climatic transition from a mild, northern subtropical climate to a southern temperate climate with severe winters and a short growing season. The state of Paraná presents the second largest production of soybean crop in Brazil (18,307 tons). For this reason, the five municipalities that showed high levels of average soybean yield (kg ha^{-1} within 10 years), based on data from the Secretary of Agriculture of State Paraná, were selected to be studied. In the state of Paraná the regions that were considered are Cascavel, Toledo, Campo Mourão, Apucarana, and Jaguariaíva (Fig. 1).

Datasets Acquisition & pre-Processing

The study period covers the agricultural years 2002–2003 to 2011–2012, a period of 10 years for soybeans. Key parameters from satellite imagery were obtained, namely soil surface moisture (Mo) and EF. Those parameters were acquired from the MODIS MOD13A2 and MOD11A2 products (specifically for tile h13v11). These products are 16-day images of the vegetation index NDVI and 8 daily Ts, at 1 km spatial resolution. All data were obtained at no cost from NASA (https://wist.echo.nasa.gov/api) in a sinusoidal projection and Hierarchical Data Format (HDF). The acquired images were the first re-projected to WGS-84 projection (EPSG 4326) processed using the MODIS Re-projection Tool (MRT), (available from https://lpdaac.usgs.gov/lpdaac/tools/modis_reprojection_tool). Then all images were exported in Geo TIFF format.

In total, 14 NDVI and Ts images were composed in 16- and 8-day periods, respectively (Table 1). The images selected of each composition comprised the agricultural period of soybean growth (September to April) totaling 28 images for each year.

The MOD11A2 product pixel values were then converted to Kelvin temperature, and subsequently to degrees Celsius (°C) as follows:

$$T\ (K) = MOD11A2*0.002 \tag{1}$$

$$T\ (°C) = TK-273.1 \tag{2}$$

The MO13A2 products were also converted to a scale of −1 to 1 by dividing the image numerical values by 10,000, using the equation below:

$$NDVI = MOD13A2/10,000 \tag{3}$$

The process of mapping the compositions of the soybean crops was made in RGB multi-temporal for the NDVI images, in order to represent only pixels with the soybean crop. For this process, the images that represented the beginning of the harvest (sprout phase) were first determined. The images that represented the total development of the plant, from the image to the beginning of the vegetative cycle were allocated in the channel Green (G) and the image with the highest value in the vegetative cycle in the Red (R) channel, while the image in channel Blue (B) corresponded to the second lower vegetative value.

Among the various simulations of cuts of gray levels, for channels of RGB color composition, the best properties were obtained by determining the RGB composition range between the channels with the values of R-180 and G and B -110 for all crop years. This involved the separation of pixels with gray levels >180 in R and gray levels <110 in GB channels.

The climatological water balance model of Thornthwaite and Mather (1995) was parameterised using as inputs parameters that were obtained from meteorological data (rainfall and average temperature) for the five studied regions. These data were obtained from three sources: (a) the Technological Institute SIMEPAR for 2002–03 to 2010, using conventional meteorological stations for each of the five studied region, (b) rainfall from the TRMM (Tropical Rainfall Measuring Mission) satellite, and, (c) average temperature from the Global Atmospheric Model ECMWF (European Centre for Medium-Range Weather Forecasts) which covers the entire county. So, based on the TRMM and ECMWF the corresponding rainfall and temperature data were estimated for the five studied regions. The TRMM data are available at no cost (http://trmm.gsfc.nasa.gov/data_dir/data.html). The temperature data obtained in raster format from JRC are also free of charge (TIF) (https://ec.europa.eu/jrc/en/mars).

Modeling system development

The proposed decision making system incorporates the following components: (1) The soybean yield prediction model of Doorenbos and and Kassam (1979) (DK); (2) The
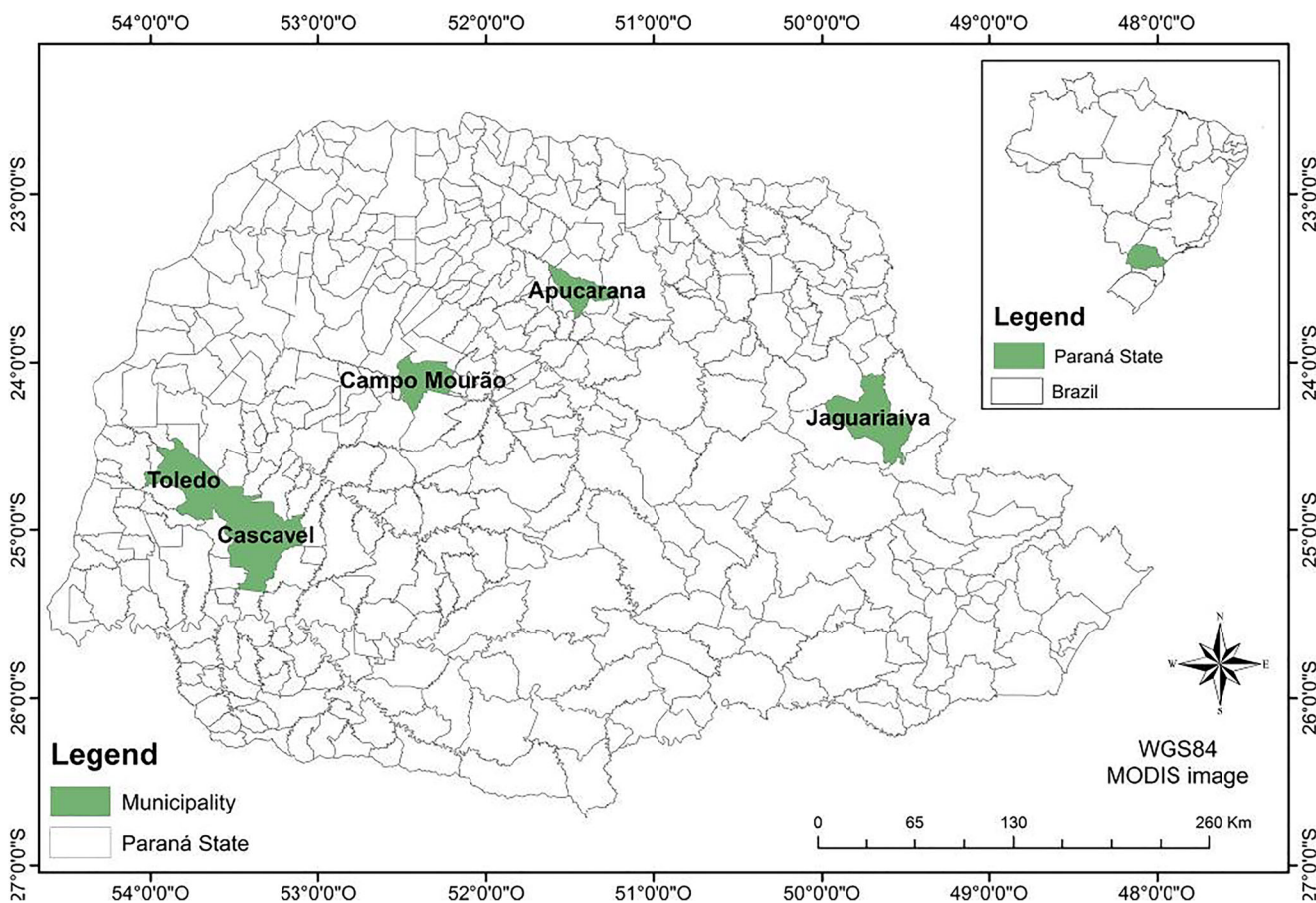
Fig. 1 Location of study area where are also shown the studied regions/municipalities

climatological water balance model of Thornthwaite and Mather 1995 (ThM), and, (3) the simplified triangle remote sensing method. All the data used in the decision making system underwent quality control to ensure data integrity, correctness, and completeness and to identify and address errors and omissions in the datasets.

The soybean yield prediction model of Doorenbos and Kassam 1979 (DK)

The multiplicative crop yield model used in this study is based on DK, as proposed by Rao et al. (1988). This model (described by Eq. 4 below) estimates the production based on the evapotranspiration ratio, ET/ETp, where ET is the actual evapotranspiration and ETp is the potential evapotranspiration. This ratio defines the production according to the water requirement of soybeans, whereby the water deficit, limits the

crop yield during some stages of the plant growth. This is expressed as follows:

$$\frac{Ya}{Yp} = \prod_{i=1}^{4} \left[1 - k_{yi} \left(1 - \frac{ET}{ETp} \right)_i \right] \tag{4}$$

- $Y\alpha$ is the estimated yield (kg/ha);
- Yp, the potential yield (kg/ha);
- ET and ETp (expressed in mm/day) are defined above; and
- k_y is the coefficient of penalization yield due to water deficiency for each physiological growth stage of the crop

The soybean crop phenological cycle consists of four different physiological growth stages: emergency/formation cotyledon, flowering, pod formation/grain filling, and full

Table 1 NDVI images and radiometric surface temperature (Ts) images selected on the same dates of satellite overpass

Images	Ts and NDVI
Scenes	257, 273, 289, 305, 321, 337,353, 001, 017033049065081 and 097
Dates	09/14 09/30 10/16 11/01 11/17 12/03 12/19 01/01 01/17 02/02 02/18 03/06 03/22

maturation. In this case each stage presents its water and thermal need for the best development of the plant that are, the yield coefficients (K_Y) that present the values of water predicted for each vegetative stage: vegetative development 0.2; flowering 0.4; grain filling 0.8 and maturation 0.2.

Potential yield (Yp) depends on the technological level applied to agriculture. A maximum value is established for the growing conditions provided there is no restriction. The value of the highest yield obtained was increased by 10% to eliminate any environmental effects that might interfere with the yield potential, according to Moraes et al. 1998 and Martins and Ortolani 2006 (Table 2). Potential yield is understood as the highest expected yield for a particular crop in the region in the condition of commercial cultivation, provided that no climatic restriction occurs. It depends on the region, the cultivar, the planting season and the technological level used. If these variables are fixed, so that always is considered the same technological level (fertilization, phytosanitary control), even cultivar, planting season and region, it is possible to experimentally estimate the potential yield from a series of crops, due to climate only.

The description of input typically, in agrometeorological modeling, the ET and ETp values were obtained from climatic water balance (CWB) models. Mathematically, ET/ETp (being a relative evapotranspiration) is very similar to EF (Nehal et al. 2017). Thus, replacing ET/ETp by the EF, the modified agrometeorological crop prediction model DK (Eq. 5) is expressed as:

$$\frac{Ya}{Yp} = \prod_{i=1}^{4} \left[1 - k_{yi} (1-EF)_i \right] \qquad (5)$$

The climatological water balance model

In this study, for computing the climatic water balance the Thornthwaite and Mather 1955 (ThM) model was used. This is one of the most widely used soil water balance estimation method for estimating the actual evapotranspiration, soil water deficit and excess. Based on this method, the actual ET and potential evapotranspiration (ETp) contributing as part of the DK model can be estimated from archived data. Detailed method description can be found in Thornthwaite and

Table 2 Values of predicted yield (Yp) plus 10% for the main soy-producing regions/municipalities of Paraná

Regions	Yp (kg/ha)
Apucarana	3.630
Campo Mourão	3.630
Cascavel	3.630
Toledo	3.955
Jaguariaíva	3.762

Mather (1995) and other works published (Pereira et al. 2002; Silva et al. 2006; Bruno et al. 2007; Sparovek et al. 2007; Dourado-Neto et al. 2010). Generally, to calculate ET/ETp obtained from climatological water balance models (CWBs) requires archived surface data. To overcome this, remote sensing triangle method was used, covered next.

The simplified triangle method

The solution for EF and Mo, obtained from the triangle approach was initially proposed by Carlson (e.g. Gillies et al. 1997), a model that does not require any external information or knowledge of land surface models. The triangular shape of the pixels indirectly constrains the solution. A triangular shape of the pixel envelope appears when the radiometric surface temperature (Ts) is plotted versus the fractional vegetation cover (Fr), which has been obtained from the normalized difference vegetation index (NDVI). Details of the triangle method implementation procedure can be found in Carlson (2013) and Carlson and Petropoulos (2019).

The simplified triangle method for estimating land surface moisture and energy fluxes uses the relationship between scaled Ts and Fr derived from remotely sensed data (Price 1990; Gillies and Carlson 1995; Carlson and Petropoulos 2019; Carlson et al. 1995). To construct the triangle (Fig. 2), cloud and standing water is required to be removed. Once this is done, several derivative parameters are determined from the pixel values of Ts and the NDVI. These are:

- the bare soil and dense vegetation values of NDVI (respectively, NDVIo and NDVIs);
- the Ts for dry/bare soil, which is representative of the highest values of Ts for pixels found over dry/bare soil (Ts [max]) and the value of the minimum Ts representative of cool, wet pixels (Ts[min]) such as found over dense vegetation.

Fr is derived from NDVI and the values of NDVIo and NDVIs (Carlson 2007). Because soil moisture has a large spatial and temporal heterogeneity, it is a difficult surface parameter to measure directly on a routine basis over large areas and longer temporal scales. Many techniques and instruments have been developed to measure soil moisture indirectly, and these continue to evolve and improve (Barrett and Petropoulos 2013; Zhang and Zhou 2016; Petropoulos et al. 2018a, b; Srivastava et al., 2019; Bao et al. 2018).

NDVIo, NDVIs, Ts[max] and Ts[min] are used to define the vertices of the triangle. NDVIs and Tmin, represent dense vegetation, define the lower left (cold) vertex and the so-called 'cold edge' of the triangle shown in Fig. 2. The cold edge represents the limit of soil wetness and corresponds to the values of Mo and EF equal to 1.0. Similarly, NDVIo and

Fig. 2 Simple geometry of the triangle. NDVI varies between its minimum and maximum values, respectively NDVIo and NDVIs, where NDVI is here scaled in Fr
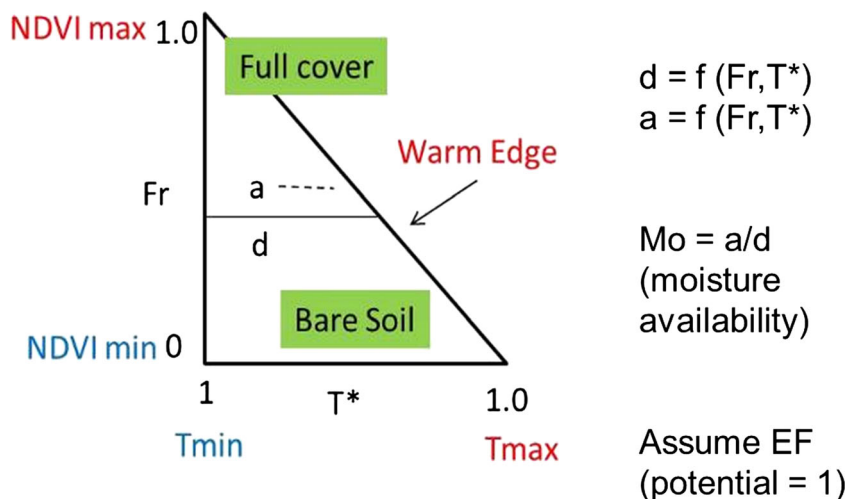
Ts[max] define the lower right vertex of the triangle. Another highly important feature, the 'warm edge', represents the limit of soil dryness where Mo = 0 and extends from Ts [max] and NDVIo to NDVIs, which, for a triangle with a well-defined upper vertex, occurs at Ts[min]. Note that while Mo equals zero along the warm edge EF itself is non-zero along the warm edge except at the lower right vertex.

Two important assumptions are made here. The first, which is also made in almost all Ts/VI methodologies (e.g., Jiang and Islam 2001), is that transpiration (evaporation from the leaves) that always equals potential, at least when the vegetation is not at the wilting point. The second assumption is that the relation between EF and Mo varies linearly across the triangle domain.

That bare soil fraction (equal to Mo), that Mo is the availability of moisture on the surface, is the ratio between the lengths of a / d, both of these lengths being functions of the scaled radiometric surface temperature (T*) and Fr. Ts is scaled to a variable T* where Ts varies between its limits of Ts[min] and Ts[max]. The variable T* is scaled between 0 and 1 as defined below.

As stated above, both Mo and EF vary linearly within the triangle between 0 and 1.0, such as (for Mo) between the cold and warm edges of the triangle. For each value of Fr and EF from the combined vegetation and bare soil, the canopy EF is assumed to be the weighted value of EF for the vegetation fraction of the pixel (EFveg = 1.0, by definition) Thus, the mathematical framework is expressed by the following simple definitions:

$$T^* = \{Ts–Ts\,(min))/(Ts\,(max)–Ts\,(min)\} \tag{6}$$

$$Fr = \left\{ \frac{NDVI–NDVI_0}{NDVIs–NDVI_0} \right\}^2 \tag{7}$$

$$Mo = 1–T^*(pixel)/T^*(warm\ edge) \tag{8}$$

$$EF = EF_{soil}(1–Fr) + Fr\ EF\ veg = Mo(1–Fr) + Fr \tag{9}$$

Where EFsoil refers to the ratio of soil evaporation to net radiation.

An overview of the proposed methodology implementation for predicting soybean yield is presented in Fig. 3 below.

Statistical analysis

Statistical analysis was conducted to verify the closeness of fit between measured crop yields and those derived from DK using three different sets of estimates of ET/ETp (EF): (1) from the triangle model, (2) for CWB by SIMEPAR (data surface), and (3) for CWB by TRMM and ECMWF (meteorological satellite). Model performance was evaluated by means of several statistical measures. In particular, the following statistical parameters were computed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), index of agreement (d_1) and d_2, Camargo and Sentelhas coefficient (c), determination coefficient (R^2), and Ea and Es errors. The average random error (non-systematic) (Ea) results from unpredictable factors and fluctuations, which cause approximately half of the measurements taken to deviate further, and the other half to deviate less, affecting the accuracy of the estimate. Systematic error (ES) is caused by identifiable sources and can in principle be eliminated or compensated. These errors cause the measurements made to be consistently above or below the actual value. MAE may be most appropriate for checking the correctness or accuracy of estimated data in relation to the measured data. To verify the final quality of the estimator model, the index of agreement, "d_1" (Willmott et al. 1985) was also used:

$$d_1 = 1 - \left[\frac{\sum(e_i–o_i)2}{\sum(|e_i–o| + |o_i–o|^2)} \right] \tag{12}$$

where e_i is the error, o_i the observation and o is the mean of observations.

Although the statistic d_1, as described by Willmott (1981), is widely used, Willmott et al. (1985) and Legates and Mccabe (1999), reported that using the quadratic function in Eq. 12
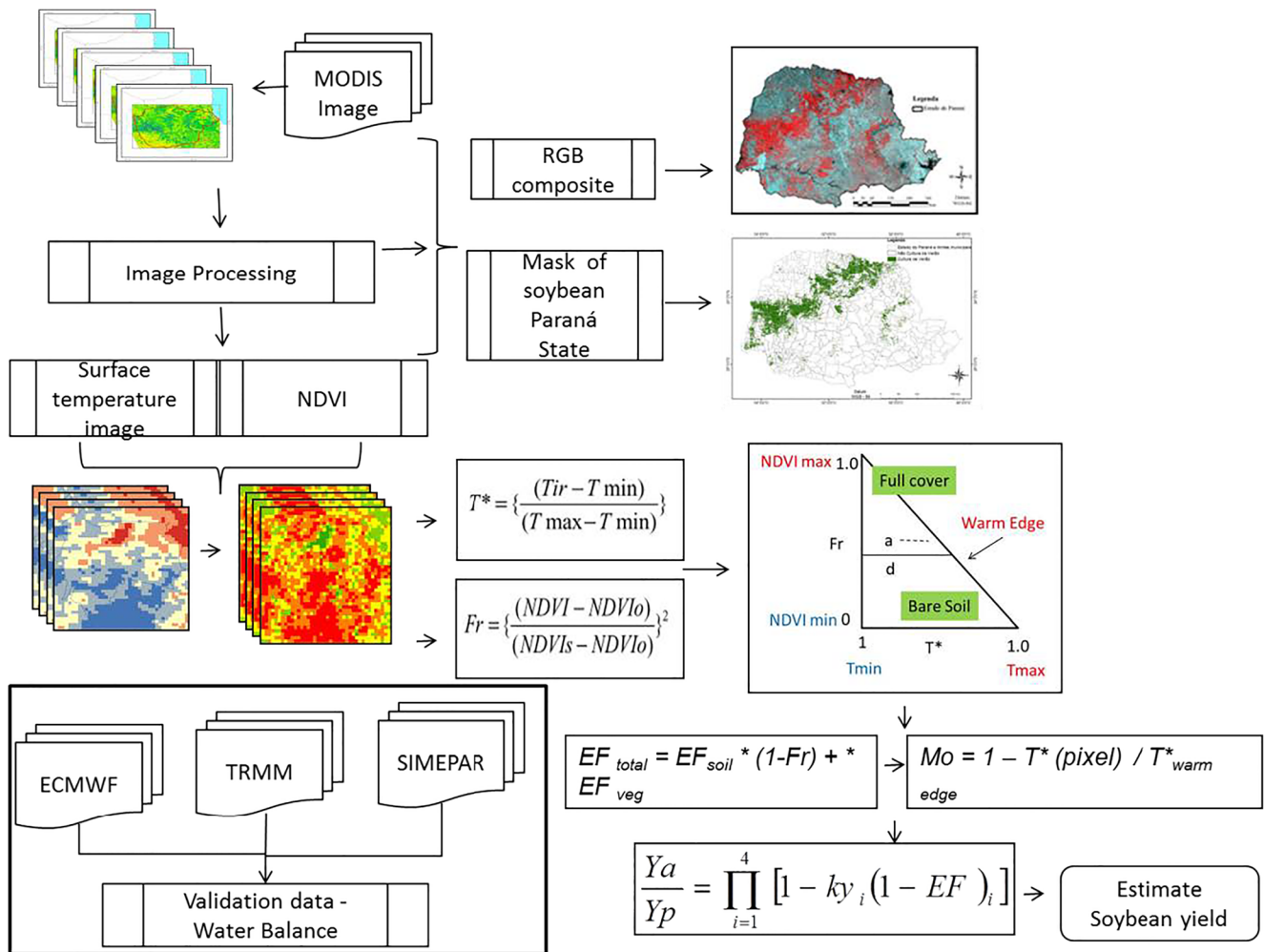
Fig. 3 Main steps comprising our methodology

may result in less reliable values of this ratio, even when there is a good performance of the estimator model. Thus, Willmott et al. (1985) propose an adaptation called index Willmott modified, expressed as follows:

$$d_2 = 1 - \left[\frac{\sum(e_i - o_i)}{\sum(|e_i - o| + |o_i - o|)} \right] \quad (13)$$

where d_1 and d_2 are dimensionless. The range of d_1 is bounded by 0 and 1, with values close to 1 indicating a near perfect fit. The values of d_2 are within −1 and 1.

In addition, the Camargo and Sentelhas coefficient (c) was computed to indicate the performance of the proposed method (Camargo and Sentelhas 1997), computed as shown in Eqs. 13–16 below:

$$R = \frac{\sum_{i=1}^{n}[(x_i - x)(y_i - y)]}{\sqrt{\sum_{i-1}^{n}(x_i - x)^2}\sqrt{\sum_{i-1}^{n}(y_i - y)^2}} \quad (14)$$

$$I_d = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(|y_i - x| + |x_i - x|)^2} \quad (15)$$

$$c = RI_d \quad (16)$$

where: x = measured values; y = estimated values; R = Pearson's correlation coefficient; Id = Willmott's index of agreement; c = Camargo and Sentelhas coefficient.

The subjective criteria for model assessing based on Camargo and Sentelhas coefficient are the following: values >0.85 indicate great performance, values between 0.85 to 0.76 suggest a very good performance, values between 0.75 to 0.66 a satisfactory performance, values between 0.65 to 0.61 median performance, values between 0.60 to 0.51 unsatisfactory performance, values between 0.50 to 0.41 poor performance, and values <0.40 very poor performance (Camargo et al. 2014).

A criterion of 5% was used to ascertain whether there was any significant difference between the measured crop yield data and those derived from the simplified triangle method. For the 5% criterion, it was determined that a total of 400 samples would be required (i.e., 400 points). Accordingly, 200 samples were randomly selected as points on the target of interest (mask of soybean crops) and 200 samples on other
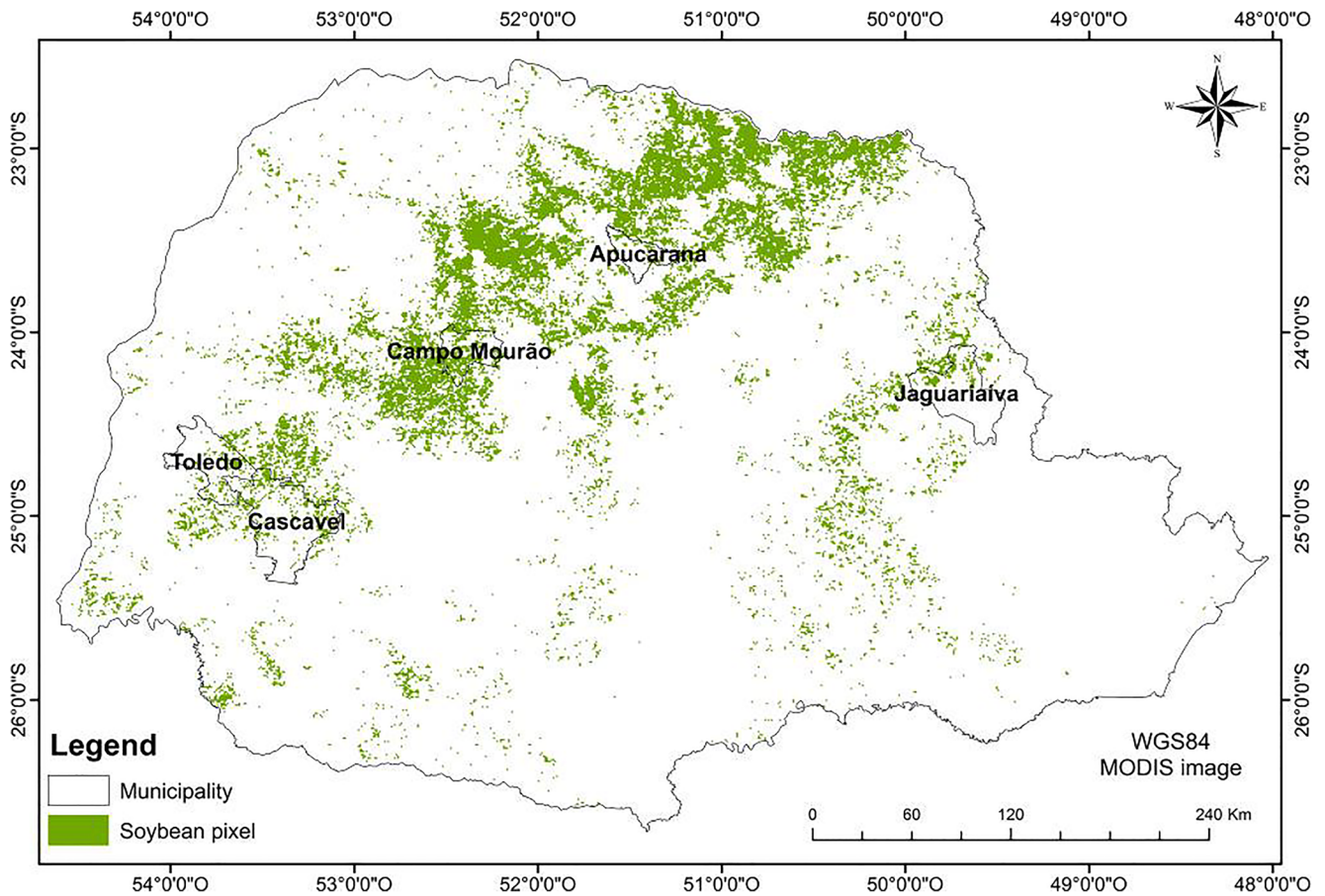
Fig. 4 Map with the soybean vegetation during summer for Paraná, crop year 2010–11 (Images NDVI- MODIS- 1 km)

targets; i.e., a stratified random sampling. Thus, the EG and IK statistical parameters computed using the equations shown below were determined for soybean production from the masked images in the analysis:

$$EG = \frac{A}{n} *100 \qquad (17)$$

$$IK = \frac{\left[n * \sum_{i=1}^{r} x_{ij} - \sum_{i=1}^{r} \left(x_{i.} * x_{.j} \right) \right]}{\left[n^2 - \sum_{i=1}^{r} \left(x_{i.} * x_{.j} \right) \right]} \qquad (18)$$

where: EG is the Global Accuracy, A is the general strike (sample point with correct answers), IK is the Kappa

coefficient of agreement, n is the number of observations, and r is the number of rows of the error matrix; x_j - Note that the subscripts refer to Row i and Column j, where $x_{i.}$ is the total marginal Row I, and x_j is the total marginal Column j.

An EG value (%) close to 100 indicates that the classification is significantly better than random. The Kappa Coefficient can range from −1 to 1. A value of 0 indicated that the classification is no better than a random classification. A negative number indicates the classification is significantly worse than random. A value close to 1 indicates that the classification is significantly better than random (Stein et al. 1998).

Table 3 Error matrix for two classifications: vegetation and rest-of-year vegetation. The columns are reference data, and the rows are predictions from the mask for the 2010–11 crop year

Classification	Vegetation	rest-of-year vegetation	Grand total	Error of Inclusion (%)
Vegetation	155	45	200	22.5%
Rest-of-year vegetation	30	170	200	15.0%
Grand total	185	215	400	EG = 81.25%
Error of Omission (%)	16.2%	20.9%	–	IK = 0.62

EG = Global Accuracy; IK = Kappa Index
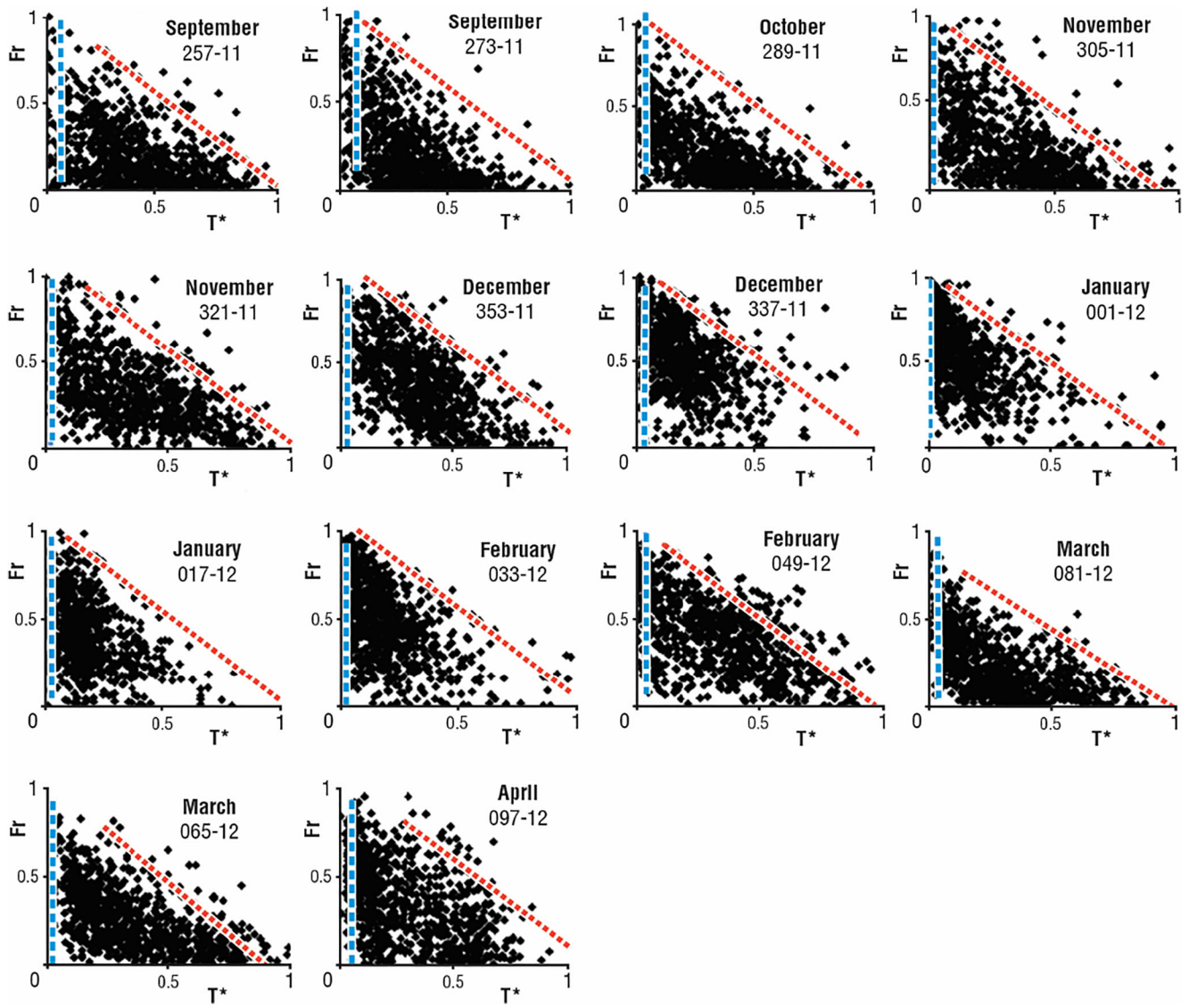
Fig. 5 Scatterplot examples obtained in our study from the implementation of the simplified triangle method for the region of Toledo, for different months of Crop Year 2011–12. The dotted red lines represent the warm (dry) edge which intersects the T* axis (the horizontal axis) at Fr = 0. The vertical axis (the cold edge) is marked by the dotted blue line and corresponds to the value of Mo = 1 and T* = 0

Fig. 6 Schematic representation of triangle method and phenological cycle of the soybean crop for region of Toledo

Results & discussion

One of the key study goals was to separate the soybean crop pixels from other crops, according to the planting time of the analyzed region. As such, a procedure was developed to implement this task in IDL (Interactive Data Language) programming environment. RGB color rendering has been transf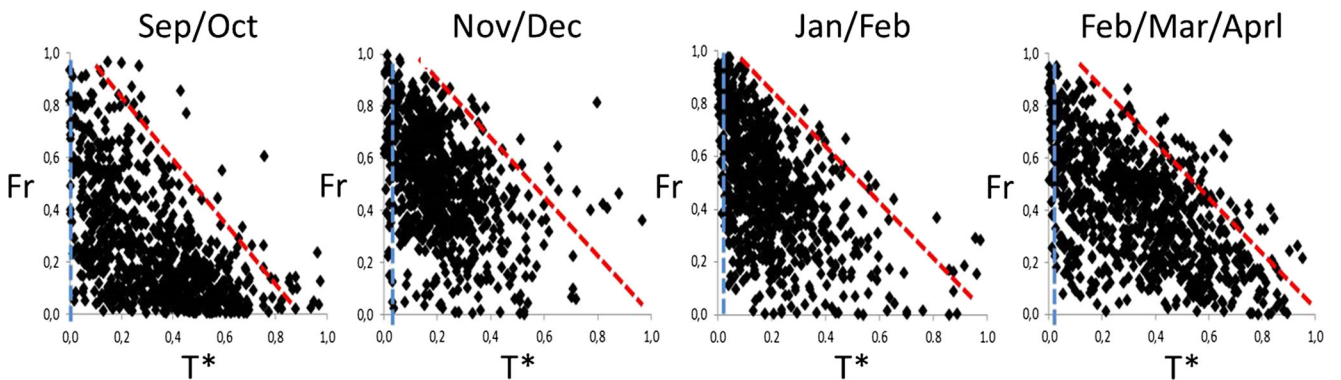ormed to gray levels (GL) ranging from 0 to 255. The objective of this transformation was to normalize the values according to the plus or minus variation, depending on the agrometeorological conditions in each crop year. Subsequently, the cutoff values were defined. This procedure was performed for each crop year to obtain the soybean crop mapping. It was found that the one that presented the best cutting result for mask generation was obtained with R-160_GB-150 for the crop years, that is, the pixels with GL > 160 on channel R and GL <150 on GB channels.

Pixels were extracted at lower gray values, creating the soybean crop map (Fig. 4). These were taken as the basis for comparison with LANDSAT 5/ TM image mosaics as ground reference.

Studying the behavior of different targets is relevant to remote sensing because this allows one to construct masks of these targets based on NDVI (Johann et al. 2012). The definition this mask, that define pixel of soybean were compared with high spatial resolution (mosaic images Landsat TM) used for ground reference. The result of this procedure was to identify a so-called "soy belt," that extends from the western to the northern region of the state with a concentration in the Mid-East region of the state.

The Kappa index (IK), here determined to be 0.62 (Table 3), corresponds in assessing agreement between measured surface class types in Fig. 4. This result was a match of 70% of the samples in the area of interest. So, it can be said that the

classification in the study area had a good representation of the reliability of the field measurements. Table 2 presents matrix errors of 400 samples used in the analysis for the crop year 2010/11. Overall accuracy (EG) is 81.25%, suggesting a good reliability of the created mask in identifying the soybean crops.

Ide and Baptista (2018) evaluated the applicability of time series of the enhanced vegetation index (EVI), from MODIS, in the mapping of irrigated areas in the Northeastern region of Brazil. The MODIS images were classified with the iterative self-organizing data analysis technique (Isodate) algorithm, generating a binary map of irrigated and non-irrigated areas for each year, presenting average Kappa coefficients of 0.26 and 0.00, respectively.

Figure 5, presents an example of scatter plots obtained for the simplified triangle method for different months for the area around the region of Toledo (crop year 2011–12). By constraining the pixels within the triangle (as shown in Fig. 5), interior values of Mo and EF all fall within the limits of 0 and 1.0. The warm side of the triangle, the sloping right-hand side, tends to be marked by a rather sharp border; this feature is referred to as the "warm edge." The warm edge, constituting the warmest pixels for each value of Fr, denotes a dry soil limit, likely corresponding to some minimum soil surface wetness, Mo = 0.

Table 4 Statistical analysis of index "d_1" and "d_2" Willmott, coefficient of determination Rsquared, index "c", RMSE, MAE, Ea, Es errors, and lower and upper limit with 95% confidence, the performance measured crop yields from 2002/03 to 2011/12 in kg/ha^{-1} versus those predicted and obtained from DK using estimated EF values obtained from the triangle method. The regions are listed as Apucarana (Apu), Campo Maurao (CM), Jaquariaiva (Jaq), Toledo (Tol), Cascavel (Cas). Values that were superior to those in Tables 5 and 6 are highlighted only for the statistical parameters d_1 and c

	Apu	CM	Jag	Tol	Cas
d_1	0.87	**0.98**	**0.92**	**0.96**	**0.92**
d_2	0.98	1.00	0.99	1.00	0.99
R^2	0.60	0.95	0.82	0.88	0.80
C	0.53	**0.93**	**0.75**	**0.85**	**0.73**
RMSE (Kg /ha^{-1})	109	78.8	145	149	156
MAE(Kg/ha^{-1})	47.3	30.8	48.5	51.8	57.2
Ea (Kg/ha^{-1})	45.4	26.2	43.5	45.5	54.9
Es (Kg/ha^{-1})	13.2	16.2	21.4	24.9	16.3

Table 5 Same as Table 3 but for results measured and obtained from DK using SIMEPAR data as input to the CWB

	Apu	CM	Jag	Tol	Cas
d_1	0.79	0.73	0.84	0.89	0.78
d_2	0.96	0.92	0.97	0.99	0.95
R^2	0.59	0.50	0.71	0.77	0.63
C	0.49	0.36	0.60	0.60	0.49
RMSE (Kg/ha^{-1})	149	209	261	245	246
MAE (Kg/ha^{-1})	56.4	72.9	80.5	102.9	76.6
Ea (Kg/ha^{-1})	46.4	60.8	70.9	90.3	65.9
Es (Kg/ha^{-1})	32.1	40.2	38.2	49.4	39.0

Table 6 Same as Table 3 but for results predicted and obtained from DK using TRMM and ECMWF data as input to the CWB

	Apu	CM	Jag	Tol	Cas
d_1	0.83	0.91	0.82	0.81	0.73
d_2	0.97	0.99	0.97	0.96	0.93
R^2	0.63	0.70	0.76	0.57	0.63
C	0.53	0.63	0.63	0.46	0.46
RMSE (Kg/ha^{-1})	133	137	260	300	209
MAE (Kg/ha^{-1})	47.5	74.5	76.6	92.6	65.6
Ea (Kg/ha^{-1})	31.5	57.3	60.1	68.6	46.7
Es (Kg/ha^{-1})	35.6	47.6	47.4	62.2	46.1

Table 7 predicted crop yield by the Triangle Method, 2002/03 to 2011/12 in kg ha^{-1}, and data measured by SEAB

Date (year/month)	Apucarana			Campo Mourão			Jaguariaiva			Toledo			Cascavel		
	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %
2002/03	2700	2793	−3.3	3000	3060	−2.0	3000	2965	1.2	3470	3569	−2.8	3300	3156	4.6
2003/04	3000	2985	0.5	3160	2975	6.2	3140	3211	−2.2	2400	2569	−6.6	2750	2658	3.5
2004/05	3300	3256	1.4	2750	2658	3.5	2000	2249	−11.,1	2650	2454	8.0	2355	2159	9.1
2005/06	2660	2980	−10.7	2280	2242	1.7	2800	2900	−3.4	2230	2349	−5.1	2700	2659	1.5
2006/07	2700	2600	3.8	2727	2689	1.4	3250	3150	3.2	3100	3136	−1.1	2851	2658	7.3
2007/08	3200	3156	1.4	3180	3000	6.0	3100	2875	7.8	3479	3258	6.8	2988	2956	1.1
2008/09	3060	3032	0.9	3000	2986	0.5	3000	3072	−2.3	2300	2632	−12.6	2580	2798	−7.8
2009/10	2800	2685	4.3	2500	2457	1.8	3200	3049	5.0	3500	3347	4.6	3322	3256	2.0
2010/11	2900	2859	1.4	3300	3256	1.4	3420	3008	13.7	3200	3185	0.5	3200	2986	7.2
2011/12	2800	2512	11.5	3173	3080	3.0	2700	2689	0.4	3124	2968	5.3	2726	2365	15.3

For a correct triangle with a well-defined vertex, the warm edge can be drawn as a straight line from the lower right-hand corner Ts[max] to the upper point where Fr = 1.0 at the upper vertex, as in the examples shown in Fig. 5. It can also be determined as the best fit through the warmest pixels. For a such triangle, the model assumes the value of T* along the warm edge is therefore equal to 1- Fr. More generally, in the case of a truncated vertex or where the warm edge line does not line up well with the edge of the pixel envelope, the warm edge line can be fit more snugly with the edge of the pixel envelope by adjusting the value of Ts[max].

The bare soil edge (Fr = 0), referred to as the soil line (Price 1990), is also well delineated by inspection. Similarly, the left-hand (cold) side of the pixel distribution often tends to have a fairly sharp edge (the cold edge) and is thought to correspond to a maximum ET, the potential ET (ETp) for each value of Fr (Mo = 1.0; EF =1.0). Here it is drawn as a vertical line from the upper vertex of the triangle to the base.

Estimates of EF and Mo, using the simplified triangle model were obtained from the triangular scatterplots for Ts* versus Fr, such as can be seen in Fig. 5 for the five regions of the state of Paraná, for 10 years of crop data (2002–03 to 2011–12). As a further example of the triangle shape, the region of Toledo and the agricultural year 2011–12, show a scattering of pixels for each satellite image, representing the entire phenological cycle of soybean (Fig. 6). This figure shows the typical triangular dispersion of pixels in T*/Fr space, where the features described above are the warm, dry edge (the dashed red line), the wet, cold edge (the dashed blue line, essentially the vertical axis of the triangle) and the soil line. Warm and cold edges, respectively, correspond to the driest and wettest pixels for a given value of Fr (Jiang and Islam 2001; Petropoulos et al. 2009; Garcia et al. 2014). Silva-Fuzzo and Rocha 2016 obtained good results with this method when compared to using the climatological water balance, which provided the basis for this methodology. Kasim (2015) indicated with some success that the triangle method can estimate Mo at the superficial layer of the soil.

Table 8 Measured crop yield with SIMEPAR, 2002/03 to 2011/12 in kg/ha^{-1}, and data measured by SEAB

Date (year/month)	Apucarana			Campo Mourão			Jaguariaiva			Toledo			Cascavel		
	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meass	Est	Diff %
2002/03	2700	2900	−6.9	3000	3200	−6.3	3000	3539	−15.2	3470	3500	−0.9	3300	3421	−3.5
2003/04	3000	3386	−11.4	3160	3547	−10.9	3140	3280	−4.3	2400	2800	−14.3	2750	2965	−7.3
2004/05	3300	3018	9.3	2750	2751	0.0	2000	2400	−16.7	2650	2100	26.2	2355	2411	−2.3
2005/06	2660	2689	−1.1	2280	2890	−21.1	2800	2900	−3.4	2230	2400	−7.1	2700	2963	−8.9
2006/07	2700	2800	−3.6	2727	2896	−5.8	3250	3575	−9.1	3100	2800	10.7	2851	3057	−6.7
2007/08	3200	3350	−4.5	3180	3048	4.3	3100	3210	−3.4	3480	3563	−2.3	2988	3356	−11.0
2008/09	3060	3100	−1.3	3000	3023	−0.8	3000	3258	−7.9	2300	2800	−17.9	2580	3210	−19.6
2009/10	2800	2900	−3.4	2500	2900	−13.8	3200	3300	−3.0	3500	3500	0.0	3322	3211	3.5
2010/11	2900	3000	−3.3	3300	3400	−2.9	3420	3235	5.7	3200	3300	−3.0	3200	3358	−4.7
2011/12	2800	2900	−3.4	3173	3106	2.2	2700	3153	−14.4	3124	3080	1.4	2726	3058	−10.9

Table 9   estimated crop yield with TRMM/ECMWF, 2002/03 to 2011/12 in kg/ha^{-1}, and data measured by SEAB

Date (year/month)	Apucarana			Campo Mourão			Jaguariaíva			Toledo			Cascavel		
	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %	Meas	Est	Diff %
2002/03	2700	2985	−9.5	3000	3104	−3.4	3000	3245	−7.6	3470	3614	−4.0	3300	3458	−4.6
2003/04	3000	3156	−4.9	3160	2985	5.9	3140	3281	−4.3	2400	3000	−20.0	2750	2863	−3.9
2004/05	3300	3256	1.4	2750	2689	2.3	2000	2569	−22.1	2650	2896	−8.5	2355	2896	−18.7
2005/06	2660	2630	1.1	2280	2189	4.2	2800	3152	−11.2	2230	2856	−21.9	2700	2896	−6.8
2006/07	2700	2956	−8.7	2727	3100	−12.0	3250	3426	−5.1	3100	2820	9.9	2851	3268	−12.8
2007/08	3200	3058	4.6	3180	3256	−2.3	3100	3180	−2.5	3479	3350	3.9	2988	3102	−3.7
2008/09	3060	3121	−2.0	3000	2985	0.5	3000	3156	−4.9	2300	2600	−11.5	2580	2890	−10.7
2009/10	2800	2900	−3.4	2500	2863	−12.7	3200	3575	−10.5	3500	3289	6.4	3322	3212	3.4
2010/11	2900	3059	−5.2	3300	3257	1.3	3420	3200	6.9	3200	3400	−5.9	3200	3156	1.4
2011/12	2800	2900	−3.4	3173	3106	2.2	2700	2984	−9.5	3124	2860	9.2	2726	2812	−3.1

Recently, Tian et al. (2013) implemented the triangle using MODIS images for the Heibe River basin, located in the arid region of northeastern China during the growing season of 2009. Their results showed that the pixel envelopes formed from Ts and Fr produced an estimate of the ET in which different domain sizes (land uses) had little effect on the spatial pattern of ET. In their study, the Pearson correlation coefficient (R) ranged from 0.94 to 1.0 between measured and predicted ET data in different domain scales.

Measured crop yields were compared to estimated yields derived from the simplified triangle model using the evapotranspiration fraction (EF) (in place of ET/ETp) in the agrometeorological crop model DK. This comparison was carried out for the five regions as shown in Fig. 1. Table 4 shows the validation results for the five regions (Apucarana, Campo Mourão, Jaguariaiva, Toledo, and Cascavel) between measured soybean yields and those calculated from remote estimates of the evapotranspiration fraction (EF), derived from the triangle method. In addition, estimated soybean yield by the ThM and DK modeling approaches were compared to measured soybean yields (Tables 5 and 6). In order the satellite image data to be compatible with the crop measurements, satellite estimates of EF for individual pixels were averaged over an area around each region.

To make these figures easier to evaluate, only the statistical parameters d_1 and c are highlighted in Tables 4-6. Highlighted numbers in Table 4 indicate that their values are higher (better) than their counterparts in the other two tables. Except for the case of "c" for Apucarana, which showed no difference in this parameter, the soybean yield estimation model presented generally better results when used with the data obtained by the simplified triangle method for the other parameters in the Table.

The results shown in Table 4 indicate that the values are higher (better) than their counterparts in the other two tables, that is, the soybean yield values estimated using data from the triangle method showed higher values than the other models, of Tables 5 and 6.

One reason for the slightly lower agreement between field measurements and those derived from TRMM or SIMEPAR data is that TRMM data pertains to a larger surface resolution of approximately 25 km, whereas SIMEPAR data are field-specific and therefore not as well-suited for a large areas.

In overall, Table 4 shows close agreement between measured and predicted crop yields, such as the index of agreement d_1 which values were reported between 0.8 and 0.95 in most cases. Similar trends were also observed for the coefficient c (values close to 1), except for Apucarana which showed lower values, indicating poorer performance. MAE, RMSE, Ea, Es, depict high variability from one region to another, while the values of these parameters generally express differences amounting to a very small fraction of the yield itself according to Tables 7, 8 and 9.

Conclusions

In this study a new method was proposed for predicting soybean yield over large spatial scale. This method is based on the simplified triangle remote sensing technique, coupled with the crop prediction model of Doorenbos and Kassam 1979 (DK) and the climatological water balance model of Thornthwaite and Mather 1995 (ThM). Use of the method was demonstrated for a large agricultural region in Brazil. Crop yield predictions by the technique proposed here were compared versus field measurements of soybean crop yield which formed our reference dataset. In addition, results were compared with the crop prediction model of Tao et al. (2008) and the climatological water balance model of Thornthwaite and Mather (1995). Soybean yields for five regions in Brazil, when compared to those estimated from DK using values of EF, obtained from the simplified triangle method, showed satisfactory agreement. Moreover, for the studied five regions, the soybean yields estimated by the triangle were slightly better than the

corresponding crop yields obtained by DK and ThM modeling approaches using surface data.

The challenge of acquiring a vast amount of surface data required for climatological water balance models implementation at large spatial scales were overcame by this new method, that is able to provide cost-effective predictions of soybean yield. The simplified triangle method and its derivative EF are potentially very useful and appropriate for use over large regions where there is lack of abundant surface observations. The technique is simple to apply, requires no ancillary surface or atmospheric information, and uses only remotely sensed images of NDVI and surface radiometric temperature (obtained, for example, from MODIS sensor). Yet, it is possible to use only data obtained by images from remote sensors, to adapt the models using only obtained variables in orbital images, facilitating the coverage of large agricultural areas in a short period of time and at low cost. This work has the potential to be downscaled to local scale or transferred to other countries with different production systems such as in the Mediterranean basin. So, using smaller scalability, for instance small properties, in which an extensive monitoring network could easier be established, the simplified triangle method can be further tested/evaluated based on different crop assessment models. Especially for tree crops, the implementation of such remote sensing techniques in combination with emerging wireless sensor network technologies remains a challenge for future work. The wider applicability and evaluation of the proposed method accuracy in other regions of the world, representative of a variety of climatological and environmental conditions, remains to be seen.

References

Alves GR, Teixeira IR, Melo FR, Souza RTG, Silva AG (2018) Estimating soybean yields with artificial neural networks. Acta Sci, Agron 40:35–50. https://doi.org/10.4025/actasciagron.v40i1.35250

Bao Y, Lin L, Wu S, Deng KAK, Petropoulos GP (2018) Surface Soil Moisture Retrievals Over Partially Vegetated Areas From the Synergy of Sentinel-1 & Landsat 8 Data Using a Modified Water-Cloud Model. Int J Appl Earth Obs Geoinf 72:76–85. https://doi.org/10.1016/j.jag.2018.05.026

Barrett BW, Petropoulos GP (2013) Remote sensing of energy fluxes and soil moisture content. CRC Press, Boca Raton, pp 85–120 ISBN 9781466505780

Bastiaassen WGM, Noordman EJM, Pelgrum H, Davids G, Allen RG (2005) SEBAL for spatially distributed ET under actual management and growing conditions. ASCE Journal of Irrigation and Drainage Engineering 131:85–93

Berka LMS, Rudorff BFT (2003) Shimabukuro, Y.E. soybean yield estimation by an agrometeorological model in a GIS. Sci Agric 60(3):433–440

Bruno IP, Silva AL, Reichardt K, Dourado-Neto D, Bacchi OOS, Volpe CA (2007) Comparison between climatological and field water balances for a coffee crop. Sci Agric 64:215–220

Camargo AP, Molle B, Tomas S, Frizzone JA (2014) Assessment of clogging effects on lateral hydraulics: proposing a monitoring and detection protocol. Irrig Sci 32(3):181–191

Camargo AP, Sentelhas PC (1997) Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. RevistaBrasileira de Agrometeorologia Santa Maria 5(1):89–97

Carlson T (2007) An overview of the "triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. Sensors 7:1612–1629

Carlson TN, Gillies RR, Schmugge TJ (1995) An interpretation of methodologies for indirect measurement of soil water content. Agric For Meteorol 77:191–205

Carlson TN (2013) Triangle models and misconceptions. International Journal of Remote Sensing Applications 3(3):155–158

Carlson TN, Petropoulos GP (2019) A new method for estimating of evapotranspiration and surface soil moisture from optical and thermal infrared measurements: the simplified triangle. Int J Remote Sens [in press]

Dalposso GH, Uribe-Opazo MA, Johann JA (2016) Soybean yield modeling using bootstrap methods for small samples. Span J Agric Res 14(3):e0207. https://doi.org/10.5424/sjar/2016143-8635

Doorenbos J, Kassam AH (1979) Yield response to water. Rome, FAO, 197p. (irrigation and drainage paper, 33)

Dourado-Neto D, de Jong van Lier Q, Metselaar K, Reichardt K, Nielsen DR (2010) General procedure to initialize the cyclic soil water balance by the Thornthwaite and Mather method. Sci Agric (Piracicaba, Braz.) 67(1):87–95

Er-Raki S, Chehbouni A, Guemouria N, Duchemin B, Ezzahar J, Hadria R (2007) Combining FAO-56 model and ground-based remote sensing to estimate water conumptions of wheat crops in a semi-arid region. Agric Water Manag 87(1):41–54

Evett SR, Kustas WP, Gowda PH, Anderson MC, Prueger JH, Howell TA (2012) Overview of the bushland evapotranspiration and agricultural remote sensing experiment 2008 (BEAREX08): a field experiment evaluating methods for quantifying ET at multiple scales. Adv Water Resour 50:4–19

Feng Z, Li-Wen Z, Jing-Jing S, Jing Feng H (2014) Soil moisture monitoring based on land surface temperature vegetation index space derived from MODIS data. Pedosphere 24(4):450–460

Ferreira DB, Rao VB (2011) Recent climate variability and its impacts on soybean yields in southern Brazil. Theor Appl Climatol 105:83–97. https://doi.org/10.1007/s00704-010-0358-8

Garcia M, Fernández N, Villagarcía L, Domingo F, Puigdefábregas J, Sandholt I (2014) Accuracy of the temperature–vegetation dryness index using MODIS under water-limited vs. energy-limited evapotranspiration conditions. Remote Sens Environ 149:100–117

Gillies RR, Carlson TN (1995) Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. J Appl Meteorol 34:745–756

Gillies RR, Carlson TN, Cui J, Kustas WP, Humes KS (1997) Verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index NDVI and surface radiant temperature. Int J Remote Sens 18:3145–3166

Goyal S (2013) Artificial neural networks in vegetables: a comprehensive review. Scientific Journal of Crop Science 2(7):75–94. https://doi.org/10.14196/sjcs.v2i7.928

Gusso A, Arvor D, Ricardo Ducati L (2017) Model for soybean production forecast based on prevailing physical conditions. Pesq Agrop

Brasileira 52(2):95–103. https://doi.org/10.1590/S0100-204X2017000200003

Gusso A, Ducati JR, Veronez MR, Arvor D, Jr S, Da LG (2013) Spectral model for soybean yield estimate using Modis/EVI data. Int J Geosci 4:1233–1241. https://doi.org/10.4236/ijg.2013.49117

Huang C, Wylie B, Yang L, Homer C, Zylstra G (2002) Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. Int J Remote Sens 23(8):1741–1748

Ide AK, Baptista GM (2018) Modis time series for irrigated-area mapping in hydrographic basins of the Brazilian northeastern region. Pesq. Agropec. Bras., Brasília 53(1):80–89

Jiang L, Islam S (2001) Estimation of surface evaporation map over southern Great Plains using remote sensing data. Water Resour Res 37(2):329–340

Johann JA, Rocha JV, Garbellini D, Lamparelli RA (2012) Estimating areas with summer crops in Paraná, Brazil using multitemporal EVI / Modis images. Pesq agropec bras 47(9):1295–1306

Kasim AA (2015) Derivation of surface soil water content using a simplified geometric method in Allahabad District, Uttar Pradesh India. Int J Sci Eng Res 6(10)

Kourgialas NN, Karatzas GP, Koubouris GC (2017) GIS policy approach for assessing the effect of fertilizers on the quality of drinking and irrigation water and wellhead protection zones (Crete, Greece). J Environ Manag 189:150–159

Lambin E, Ehrlich D (1996) The surface temperature-vegetation index space for land cover and land cover-change analysis. Int J Remote Sens 17:463–487

Legates DR, Mccabe GJ (1999) Evaluating the use of 'goodness- of-fit' measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241

Lei H, Yang D (2010) Interannual and seasonal variability in evapotranspiration and energy partitioning over an irrigated cropland in the North China plain. Agric For Meteorol 150:581–589

Liou Y-A, Kar SK (2014) Evapotranspiration estimation with remote sensing and various surface energy balance algorithms—a review. Energies 7:2821–2849

Mammadov J et al (2018) Wild relatives of maize, Rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. Front Plant Sci 9:886

Martins AN, Ortolani AA (2006) Estimativa de produção de laranja valência pela adaptação de um modelo agrometeorológico. Bragantia 65(2):355–361

De Melo RW, Fontana DC, Berlato MA, Ducati JR (2008) An agrometeorological-spectral model to estimate soybean yield, applied to southern Brazil. Int J Remote Sens 29:4013–4028. https://doi.org/10.1080/01431160701881905

Mercante E, Lamparelli RAC, Uribe-Opazo MA, Rocha JV (2010) Linear regression models to soybean yield estimate in the west region of the state of Paraná, Brazil, using spectral data. Eng Agríc 30(3):504–517

Meyer S J (1990) The development of a crop specific drought index for corn. Lincoln/USA. 165p. Doctoral dissertation - University of Nebraska

Minacapilli M, Consoli S, Vanella D, Ciraolo G, Motisi A (2016) A time domain triangle method approach to estimate actual evapotranspiration: application in a Mediterranean region using MODIS and MSG-SEVIRI products. Remote Sens Environ 174:10–23

Moraes AVC, Camargo MBP, Mascarenhas HAA, Miranda MAC, Pereira JCVNA (1998) Teste e análise de modelos agrometeorológicos de estimativa de produtividade para a cultura da soja na região de Ribeirão Preto. Bragantia, Campinas 57(2): 393–406

Nehal L, Abderrahmane H, Abdelkader K, Zahira S, Mansour Z (2017) Evapotranspiration and surface energy fluxes estimation using the Landsat-7 enhanced thematic mapper plus image over a semiarid Agrosystem in the North-west of Algeria. Revista Brasileira de Meteorologia 32(4):691–702. https://doi.org/10.1590/0102-7786324016

Pablos M, Martínez-Fernández J, Sánchez N, González-Zamora Á (2017) Temporal and spatial comparison of agricultural drought indices from moderate resolution satellite soil moisture data over Northwest Spain. Remote Sens 9:11. https://doi.org/10.3390/rs9111168 (1168)

Pagano MC, Miransari M (2016) The importance of soybean production worldwide. In book: Abiotic and Biotic Stresses in Soybean Production 1:1–26. https://doi.org/10.1016/B978-0-12-801536-0.00001-3

Pandey PC, Mandal V, Katiyar S, Kumar P, Tomar V, Patairiya S, Ravisankar N, Gangwar B (2015) Geospatial Approach to Assess the Impact of Nutrients on Rice Equivalent Yield using MODIS Sensors' Based MOD13Q1-NDVI Data. Sensors Journal, IEEE 15(11):6108–6115

Pereira AR, Angelocci LR, Sentelhas PC (2002) Agrometeorologia: fundamentos e aplicações práticas. Guaíba: Agropecuária 478

Petropoulos G, Ireland G, Lamine S, Griffiths HM, Ghilain N, Anagnostopoulos V, North MR, Srivastava PK, Georgopoulou H (2016) Operational evapotranspiration estimates from SEVIRI in support of sustainable water management. Int J Appl Earth Obs Geoinf 49:175–187

Petropoulos G, Carlson TN, Wooster MJ, Islam S (2009) A review osTs/VI remote sensing based methods for the retrieval of land surfasse energy fluxes and soil surface moisture. Prog Phys Geogr 33:224–250

Petropoulos GP, Srivastava PK, Piles M, Pearson S (2018a) EO-based operational estimation of soil moisture and evapotranspiration for agricultural crops in support of sustainable water management. Sustainability MDPI 10:181-1–181-18120. https://doi.org/10.3390/su10010181

Petropoulos GP, McCalmont JP (2017) An operational in-situ soil moisture and soil temperature monitoring network for West Wales, UK: the WSMN network. Sensors 17:1481–1491. https://doi.org/10.3390/s17071481

Petropoulos GP, Srivastava PK, Feredinos KP, Hristopoulos D (2018b) Evaluating the capabilities of optical/TIR imagine sensing systems for quantifying soil water content. Geocarto International. https://doi.org/10.1080/10106049.2018.1520926

Prasad AK, Chai L, Singh RP, Kafatos M (2006) Crop yield estimation model for Iowa using remote sensing and surface parameters. Int J Appl Earth Obs Geoinf 8(1):26–33

Price JC (1990) Using spatial context in satellite data to infer regional scale evapotranspiration. IEEE Geosci Remote Sens Lett 28:940–948

Rao NH, Sarma PBS, Chander SA (1988) Simple dated water-production function for use in irrigated agriculture. Agric Water Manag 13:25–32

Safa M, Samarasinghe S, Nejat M (2015) Prediction of wheat production using artificial neural networks and investigating indirect factors affecting it: case study in Canterbury province, New Zealand. J Agric Sci Technol 17(4):791–803

Silva AL, Roveratti R, Reichardt K, Bacchi OOS, Timm LC, Bruno IP, Oliveira JCM, Dourado-Neto D (2006) Variability of water balance components in a coffee crop grown in Brazil. Sci Agric 63:105–114

Silva-Fuzzo DF, Rocha JV (2016) Simplified triangle method for estimating evaporative fraction (EF) over soybean crops. J Appl Remote Sens 10(4):046027

Sims DA, Rahman AF, Cordova VD, El-Masri BZ, Baldocchi DD, Bolstad PV, Flanagan LB, Goldstein AH, Hollinger DY, Misson L, Monson RK, Oechel WC, Schmid HP, Wofsy SC, Xu LA (2008) New model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote Sens Environ 112: 1633–1646. https://doi.org/10.1016/j.rse.2007.08.004

Jong SG, Van Lier Q, Dourado-Neto D (2007) Computer assisted Köppen climate classification for Brazil. Int J Climatol 27:257–266

Srivastava PK, Pandley PC, Petropoulos GP, Kourgialas NK, Pandley S, Singh U (2019) GIS and remote sensing aided information for soil moisture estimation: a comparative study of interpolation technique. Resources MDPI 8(2):70. https://doi.org/10.3390/resources8020070

Stein A, Bastiaanssen WGM, De Bruin S, Cracknell AP, Curran PJ, Fabbri AG, Gorte BGH, Van Groenigen JW, Van Der Meer FD, Saldana A (1998) Integrating spatial statistics and remote sensing. Int J Remote Sens 19(9):1793–1814. https://doi.org/10.1080/014311698215252

Stisen S, Sandholt I, Norgaard A, Fensholt R, Eklundh L (2007) Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens Environ 110:262–274

Tao F, Yokozawa M, Liu J, Zhang Z (2008) Climate-crop yield relationships at provincial scales in China and the impacts of recent climate trends. Clim Res 38:83–94. https://doi.org/10.3354/cr00771

Thornthwaite CW, Mather JR (1995) The water balance. In: Centerton, N. J. (ed), 104p. (Publ. in Climatology, v. 8, n. 1

Tian J, Hongbo S, Xiaomin S, Shaohui C, He H, Linjun Z (2013) Impact of the spatial domain size on the performance of the Ts-VI triangle method in terrestrial evapotranspiration estimation. Remote Sens 5:1998–2013

Willmott CJ (1981) On the validation of models. Phys Geogr 2:184–194

Willmott CJ, Ackleson SG, Davis JJ, Feddema KM, Klink DR (1985) Statistics for the evaluation and comparison of models. J Geophys Res 90:8995–9005

Yang H, Luo P, Wang J, Mou C, Mo L, Wang Z, Fu Y, Lin H, Yang Y, Bhatta LD (2015) Ecosystem evapotranspiration as a response to climate and vegetation coverage changes in Northwest Yunnan, China. PLoS One 10(8):e0134795. https://doi.org/10.1371/journal.pone.0134795

Zhang D, Zhou G (2016) Estimation of soil moisture from optical and thermal remote sensing: a review. Sensors 6:1308

Zhu W, Shaofeng J, Aifeng L (2017) A universal Ts-VI triangle method for the continuous retrieval of evaporative fraction from MODIS products. J Geophys Res-Atmos 122(10):206–227

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.