

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v41n2p196-203/2021>

CROP MODELING WITH LESS DATA: THE FAO MODEL FOR SOYBEAN YIELD ESTIMATION

Jonathan Richetti^{1,2*}, Jerry Adriani Johann¹, Miguel Angel Uribe-Opazo³

^{1*}Corresponding author. Geotechnology and Data Science Research Center (GeoScience), Western Paraná State University (UNIOESTE), Cascavel - PR, Brazil.

E-mail: jonathan.richetti@csiro.au | ORCID ID: <https://orcid.org/0000-0003-2675-3425>

KEYWORDS

water balance, evapotranspiration, Thornthwaite and Mather.

ABSTRACT

Crop growth simulation models such as WOFOST and DSSAT are useful, but require several inputs that sometimes are not available, especially in developing areas. In addition, measured data is usually time and labor-intensive. In search of faster and easier methods for soybean estimates, this study presents a lower input requiring methodology for yield estimation. This study combines the FAO-33 yield model with the agro-ecological zone approach for soybean yield estimations using mostly indirect data. Sowing and harvest dates and yield were collected from 74 soybean commercial farms. Agrometeorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF) were used. Fifty farms (66%) were used to calibrate the model and 24 farm areas (33%) were used for evaluation purposes. Two methodologies (FAO-56 and Thornthwaite and Mather) for water balance and actual evapotranspiration (ET_a) estimations were used. The comparison of yield estimations and observations showed that the use of low data input to obtain reasonable accuracy, with a mean error of -310 kg ha^{-1} and a mean absolute percentage error of 23.3%.

INTRODUCTION

Global food issues are frequently raised due to the growing demand for food in the world market and fluctuations in commodity prices (Sakamoto et al., 2014), pointing to the importance of studies that reduce market speculation and assist in food management. In this sense, local and regional estimates of crop yield are important for macro-and microeconomic management (Johnson, 2014). Thus, proposing ways to obtain agricultural statistics reliably and quickly is essential (Johnson, 2014), and one way to obtain such information is through the remote sensing technique. There are different methods to do so. Moges et al. (2007) used a hand-held optical sensor to estimate sorghum yield in Oklahoma – USA, obtaining an in-season estimated yield derived from green Normalized Difference Vegetation Index (NDVI) correlated with final grain yield ($r = 0.71$). Gusso et al. (2017) developed a satellite remote sensing-based procedure to estimate soybean production before crop harvest, with coefficients of determination ranging from 0.91 to 0.98. Richetti et al. (2018) used machine learning algorithms with Enhanced Vegetation Index (EVI) for soybean in Brazil

and obtained a mean error of 3.5 kg ha^{-1} , RMSD of 373 kg ha^{-1} , and Willmott's d of 0.85.

Another approach consists of crop growth mechanistic models such as WOFOST (World Food Studies) (Boogaard et al., 2014) or DSSAT (Decision Support System for Agrotechnology Transfer) (Jones et al., 2003), which are reliable and present accurate results. However, the minimum data requirements are sometimes high for developing or less technologically advanced regions. Antle et al. (2017) presented future strategies for agricultural systems and modeling, indicating that data are the foundation for the science and the analysis of agricultural systems. However, better data may be the greatest need and challenge to achieve the next generation of research and modeling for agricultural systems. In addition, the lack of information for model inputs is still a challenge in many developing areas. There is little information on agrometeorological and soil data, and more complex inputs, such as the soil fertility level, are very difficult to obtain in some cases. Thus, this study aimed to assess whether a low input model is capable of acceptable accuracy results.

² CSIRO Agriculture & Food, Floreat - WA, Australia (current address)

³ Spatial Statistics Laboratory (LEE), Western Paraná State University (UNIOESTE), Cascavel - PR, Brazil

Area Editor: Jefferson Vieira José

Received in: 12-3-2018

Accepted in: 1-12-2021



MATERIAL AND METHODS

Study area

The study area comprised of 74 monitored farm areas with an average size of 102 ha, in the state of Paraná (southern Brazil), located between parallels 22°29' S and 26°43' S and the meridians 48°2' W and 54°38' W (Figure 1), with an area of 199,308 km². The climate in the state of Paraná is mainly a humid subtropical climate (Aparecido et al., 2016), with a predominance of clay-textured Oxisols (EMBRAPA, 2009). The state of Paraná accounts for almost 18% of the Brazilian production (CONAB, 2018), which means a soybean production higher than that of China, the world's fourth-largest soybean producer (FAOSTAT, 2018). The observed

yield (Y_{obs}) and sowing and harvest dates from 74 farms (Figure 1) from 2007/2008 to 2013/2014 growing seasons were used. In the state of Paraná, soybean is usually sown from September to November and harvested from January to April. A total of 46 farms were used in the 2013/2014 growing season, 13 in 2012/2013, four in 2011/2012, three in 2010/2011, three in 2008/2009, three in 2007/2008, and two in 2009/2010. Some farms belong to partners, but most of the data (2012/2013 and 2013/2014 growing seasons) were obtained from field research developed during the Embrapa project named MAPAGRI (Mapping Agricultural Activity in Brazil), developed and completed in 2014 by several universities and research institutions in Brazil and financed by Embrapa Agricultural Informatics.

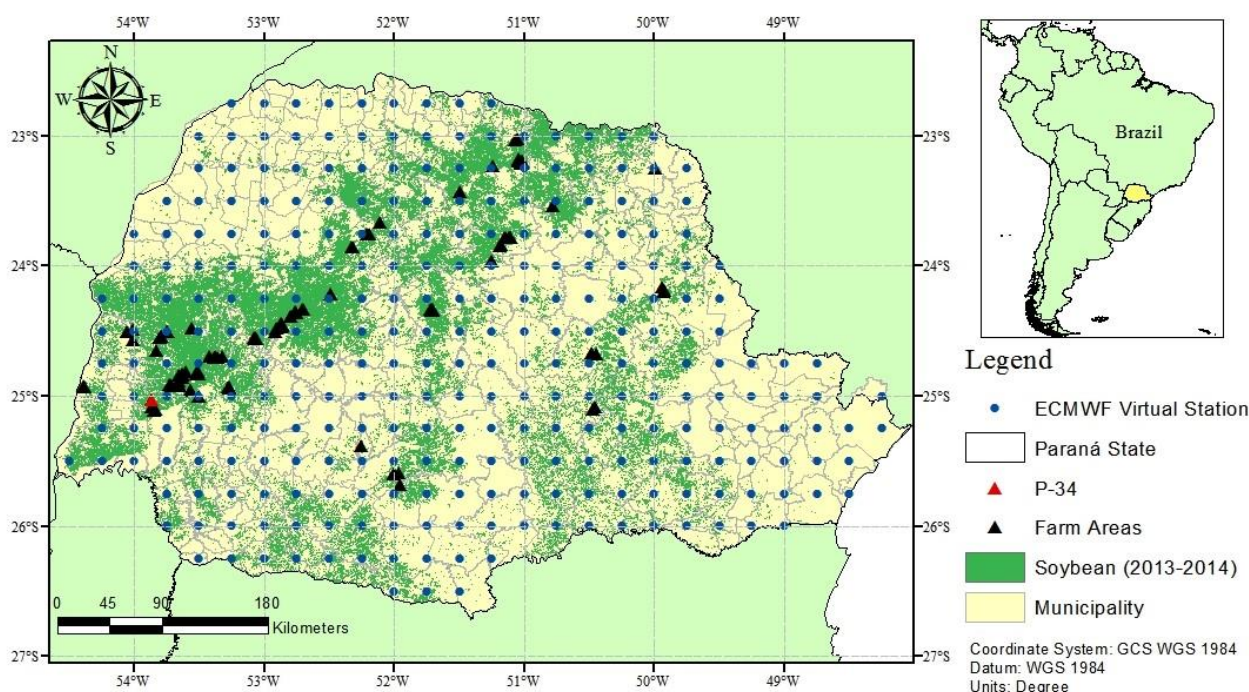


FIGURE 1. Location map of the analyzed farm areas and soybean map for crop-year 2013-2014.

Data

Agrometeorological data, quality, and reliability are essential in the estimation of agricultural yield. Thus, climate data from the European Centre for Medium-Range Weather Forecasts (ECMWF) were used considering the poor network of weather stations in the state, with only 59 weather stations in operation, the red tape for obtaining data, the irregular spatial distribution of the stations, and the lack of real-time data availability (IAPAR, 2015). Johann et al. (2016) conducted a study on ECMWF climate elements in the state of Paraná and concluded that the precipitation and ET_0 data had a significant uncertainty when compared to meteorological station data. However, Hagedorn et al. (2008) pointed out that raw ensemble forecasts of surface temperatures were biased, but approximately 70% of the improvement in ECMWF could be attributed to a simple correction of the mean bias. Therefore, the calibration of the model by HI should correct the ECMWF uncertainty.

The used agrometeorological data from ECMWF are available in a grid with a 10-day time resolution and a 0.25-degree spatial resolution ($\pm 25 \times 25$ km). This grid format was called virtual station (VS) (Figure 1). The data of rainfall (mm; 2007 to 2014), radiation sum ($J\ m^{-2}$; 2007 to 2014),

average air temperature ($^{\circ}C$; 2007 to 2014), and evapotranspiration (Penman-Monteith, mm; 2007 to 2014) from ECMWF were used (Dee et al., 2011). Initially, the 10-day agrometeorological data were interpolated by inverse distance weighting (IDW) at a 250×250 -m spatial resolution. Then, the average values of each agrometeorological data were extracted for each farm area.

The information of the total available soil water (TAW) for the state of Paraná was generated from the total water-holding capacity, as presented by Farias et al. (2007), from soil types obtained from the New Soil Map of Brazil (EMBRAPA, 2011). It is the only used soil information and input for both water balance methods.

The soybean crop was considered as described and characterized by Allen et al. (1998). The crop water stress coefficient ($k_y = 0.8$) and the p -depletion factor (0.5) were fixed for the entire growth cycle. The crop index (k_c) and root depth (Z_r) values were used according to the soybean phenological stage, as presented by FAO (Allen et al., 1998).

Water balance and yield estimation

The FAO-33 model requires four inputs to estimate the yield (Y_{est}) of a crop, namely: maximum yield (Y_x), crop

reference evapotranspiration (ET_c), actual evapotranspiration (ET_a), and crop water stress coefficient (k_s). Y_x can be determined differently, from historical or experimental data. However, the use of historical data requires reliable historical information, and experimental data are costly and time-consuming. Therefore, Y_x was estimated for this study using the agro-ecological zone approach (Kassam & Higgins, 1981), which considers the average temperature to estimate potential production. The crop evapotranspiration (ET_c) and actual evapotranspiration (ET_a) were obtained from two methods: the FAO-56 water balance (FAOWB) (Allen et al., 1998) and the Thornthwaite and Mather water balanced (TMWB) (Black, 2007). The calibration of the model was

performed through a bias adjustment factor, consisting of comparison between field data and the data estimated by the model. All the procedures (Figure 2) were conducted for each farm. The harvest index (HI) was used as the bias adjustment factor, minimizing to zero the difference between the observed yield (Y_{obs} from farms) and the estimated yield (Y_{est} from the FAO-33 model). For this, the Simplex LP method of the Solver tool in Excel was used. Fifty out of the 74 farm areas were used to determine HI and the other 24 farm areas were used to validate the results. The HI was used to determine the potential yield (Y_x) and as a calibration factor for the model, making the process as simple as possible.

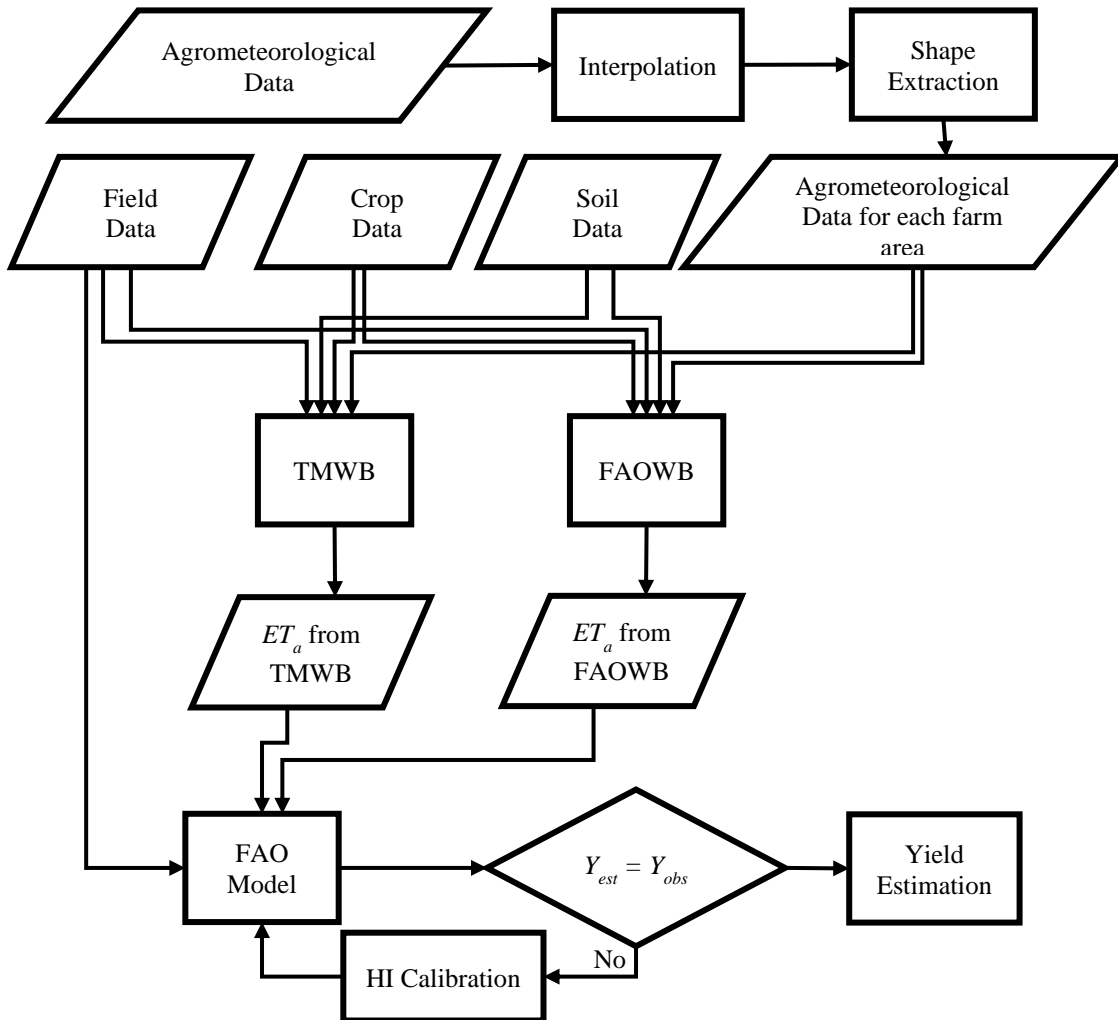


FIGURE 2. Flowchart of the study.

FAO-56 water balance (FAOWB)

The FAOWB uses depletion in the root zone (Equation 1) to assess the crop water stress.

$$D_{r,t} = D_{r,t-1} - P_t + RO_t - I_t - CR_t + ET_{c,t} + DP_t, \quad (1)$$

where:

$D_{r,t}$ is the depletion in the root zone at the end of time t (mm);

$D_{r,t-1}$ is the depletion in the root zone at previous time $t-1$ (mm);

P_t is precipitation at time t (mm);

RO_t is the runoff at time t (mm);

I_t is irrigation at time t (mm);

CR_t is the capillary rise at time t (mm);

$ET_{c,t}$ is the crop evapotranspiration at time t (mm), and

DP_t is the deep percolation at time t (mm).

The values capillary rise, deep percolation, runoff, and irrigation were not considered in this study, as they are extremely hard to be determined and the state of Paraná has no significant irrigation areas. One way of calculating $ET_{a,t}$ is through the $k_{s,t}$ factor, that is, the crop water stress coefficient at time t (Equation 2). The $k_{s,t}$ coefficient (Equation 3)

quantifies the reduction factor relative to the evaporation of soil water availability.

$$ET_{a,t} = k_{s,t} * ET_{c,t}, \quad (2)$$

where:

$ET_{a,t}$ is the actual evapotranspiration at time t (mm), and

$k_{s,t}$ is the water stress coefficient at time t .

$$k_{s,t} = \begin{cases} 1, & \text{if } D_{r,t} < RAW_t \\ (TAW_t - D_{r,t}) / (TAW_t - RAW_t), & \text{if } D_{r,t} \geq RAW_t \end{cases} \quad (3)$$

where:

TAW_t is the total water available at a certain soil depth at time t (mm), and

RAW_t is the readily available water for the crop at time t (mm).

The readily available water at time t (RAW_t , Equation 4) represents the amount of water that the plant needs no effort to use. It is calculated by multiplying the TAW_t value by the p coefficient ($p = 0.5$) (Allen et al., 1998), which is the depletion coefficient representing the difficulty that the plant undergoes because the amount of water in the soil is not readily available.

$$RAW_t = p * TAW_t * Z_{r,t}. \quad (4)$$

The amount of depletion ($D_{r,t}$, Equation 5) is dependent on the TAW_t value and $D_{r,t-1}$.

$$D_{r,t} = \begin{cases} TAW_t, & \text{if } D_{r,t-1} > TAW_t \\ D_{r,t-1}, & \text{if } D_{r,t-1} \leq TAW_t \end{cases} \quad (5)$$

Thorntwaite and Mather water balance (TMWB)

The *TMWB* method calculates the water balance differently using the same data. First, it determines the difference between precipitation (P_t) and ET_c , preserving the positive or negative signs. Subsequently, it estimates both the accumulated negative at time t (AN_t , mm; Equation 6) and the water stored in the soil at time t (WSS_t , mm; Equation 7). The accumulated negative is the sum of the sequence of negative values of the months in which DIF_t has a negative value, showing that rained less than the soil lost water during the period, in this case dekady, indicating the potential for soil dryness.

$$AN_t = \begin{cases} P_t - ET_{c,t}, & \text{if } P_t - ET_{c,t} < 0 \\ TAW_t * \ln(WSS_t / TAW_t), & \text{if } P_t - ET_{c,t} \geq 0 \end{cases} \quad (6)$$

$$WSS_t = \begin{cases} WSS_{t-1} + (P_t - ET_{c,t}), & \text{if } P_t - ET_{c,t} \geq 0 \\ TAW_t * e^{(AN_t / TAW_t)}, & \text{if } P_t - ET_{c,t} < 0. \end{cases} \quad (7)$$

After determining WSS_t , it is possible to verify alterations in soil water at time t (ALT_t , mm; Equation 8). It refers to the difference between WSS_t values while

maintaining the positive and negative signs, showing whether the amount of water in the soil increased.

$$ALT_t = WSS_t - WSS_{t-1}. \quad (8)$$

The $ET_{a,t}$ value (mm; Equation 9) is calculated considering water alterations in the soil, never with negative values.

$$ET_{a,t} = \begin{cases} ET_{c,t}, & \text{if } ALT_t \geq 0 \\ (P_t + |ALT_t|), & \text{if } ALT_t < 0 \end{cases} \quad (9)$$

The $ET_{a,t}$ value allows verifying if there was a water deficit at time t (DEF_t , mm; Equation 10), that is, DEF_t indicates the lack of crop evapotranspiration.

$$DEF_t = ET_{c,t} - ET_{a,t}. \quad (10)$$

The method also introduces the excess water at time t (EXC_t , mm; Equation 11). EXC is the amount of water that is lost during rainy periods and can be determined when WSS_t is lower than TAW_t , showing the water supply in the soil at time t , and when WSS_t is equal to TAW_t , showing that the soil is storing its maximum water capacity.

$$EXC_t = \begin{cases} 0, & \text{if } WSS_t > TAW_t \\ (P_t - ET_{c,t}) - ALT_t, & \text{if } WSS_t \leq TAW_t. \end{cases} \quad (11)$$

Thus, water balance is the difference between water excess and deficit at a given period.

Ecological zone approach for potential yield (Y_x) estimation

The Y_x value is the sum of potential gross yield at time t (PPR_t , kg ha⁻¹; Equation 12).

$$Y_x = \sum_{t=1}^n PPR_t. \quad (12)$$

The crop gross yield is determined by the potential gross yield of the standard crop at time t (PPB_t , kg ha⁻¹; Equation 13) and the harvest index (HI , dimensionless), responsible for the model calibration in the study.

$$PPR_t = 0.265455 * HI * PPB_t. \quad (13)$$

The PPB_t value is the potential gross yield for cloudy days at time t ($PPB_{n,t}$; Equation 14) plus the potential gross yield for clear days at time t ($PPB_{c,t}$; Equation 15).

$$PPB_{n,t} = (31.7 + 0.219 * Q_{0,t}) * ctn_t * 0.6. \quad (14)$$

$$PPB_{c,t} = (107.2 + 0.36 * Q_{0,t}) * ctc_t * 0.6. \quad (15)$$

Potential yields are dependent on temperature corrections for cloudy (ctn_t ; Equation 16) and clear days at time t (ctc_t ; Equation 17), and the radiation at the top of the atmosphere ($Q_{0,t}$; Equation 18).

$$ctn_t = \begin{cases} 0.583 + 0.014 * T_t + 0.0013 * T_t^2 - 0.000037 * T_t^3, & \text{if } 16.5 \leq T_t \leq 37 \\ -0.0425 + 0.035 * T_t + 0.00325 * T_t^2 - 0.0000925 * T_t^3, & \text{if } T_t < 16.5 \text{ or } T_t > 37 \end{cases} \quad (16)$$

$$ctc_t = \begin{cases} -0.0425 + 0.035 * T_t + 0.00325 * T_t^2 - 0.0000925 * T_t^3, & \text{if } 16.5 \leq T_t \leq 37 \\ -1.085 + 0.07 * T_t + 0.0065 * T_t^2 - 0.000185 * T_t^3, & \text{if } T_t < 16.5 \text{ or } T_t > 37 \end{cases} \quad (17)$$

$$Q_{0,t} = 899.521344 * DR_t * \left(\frac{\pi}{180} * hn + \sin(lat) * \sin(\delta_t) \right) + (\cos(\delta_t) + \cos(lat)) * \cos(\delta_t) * \sin(hn), \quad (18)$$

where:

T is the average temperature ($^{\circ}\text{C}$, from ECMWF);

$Q_{0,t}$ is the radiation at the top of the atmosphere at time t (Cal cm^{-2});

DR_t is the Sun-Earth relative distance at time t (UA; Equation 19);

lat is the local latitude (Rad);

hn is the time of sunrise (Rad), and

δ_t is the solar declination at time t (Rad; Equation 20).

$$DR_t = 1 + 0.033 * \cos\left(\frac{360}{365 * \delta_t}\right), \quad (19)$$

$$\delta_t = 23.45 * \sin\left(\frac{360}{365 * (DJ - 80)}\right), \quad (20)$$

where:

DJ is the Julian day.

The HI value was determined by minimizing the difference between the actual yield (Y_{act} , from farm areas) and the estimated yield (Y_a , from the FAO-33 model) to zero, using the Simplex LP method of the Solver tool in Excel. The methods FAO-56 water balance (FAOWB) and Thornthwaite and Mather water balance (TMWB) were used to estimate ET_a . Fifty out of the 74 farm areas were used to determine HI and the other 24 farm areas were used to validate the results. The HI was used to determine the potential yield (Y_x), and the higher the HI value, the higher the potential yield. The HI value was used as a calibration factor for the model, making the process as simple as possible.

FAO-33 yield model

The traditional FAO-33 model (Equation 21) shows that the relative reduction in yield is related to the corresponding relative reduction in evapotranspiration (Steduto et al., 2012). It was operated in a 10-day stage, and the final yield was obtained by the sum of estimated yields for the growing season.

$$\left(1 - \frac{Y_a}{Y_x}\right) = k_y \left(1 - \frac{ET_a}{ET_c}\right), \quad (21)$$

where:

Y_x and Y_a are the potential and estimated yield, respectively (t ha^{-1});

ET_c and ET_a are the crop and actual evapotranspiration, respectively (mm), and

k_y is the yield coefficient that translates the crop yield sensitivity to water stress (dimensionless).

Statistical analysis

The metrics consisted of the mean absolute error (MAE; Equation 22), which measures the magnitude of the errors in a set of estimates, the mean error (ME; Equation 23), the root mean square error (RMSE; Equation 24), which

indicates the size of the error generated by the model, that is, the model performance assessment criteria, and the mean absolute percentage error (MAPE; Equation 25), which shows, in relative terms, the errors of the estimates.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{esti} - Y_{obsi}|, \quad (22)$$

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_{esti} - Y_{obsi}), \quad (23)$$

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (Y_{esti} - Y_{obsi})^2\right)}, \quad (24)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|Y_{esti} - Y_{obsi}|}{Y_{obsi}} * 100\right), \quad (25)$$

where:

n is the number of farms;

Y_{esti} is the estimated yield, and

Y_{obsi} is the observed yield from farm areas.

RESULTS AND DISCUSSION

Model calibration

The model calibration was performed based on 50 farms using HI . The results show that the average HI values of 0.2625 and 0.2093 for FAOWB and TMWB, respectively, are recommended (Table 1). Both HI values presented a homogeneous coefficient of variance, and the results of the t-test showed 5% significance, indicating a significant difference between the two index values. Therefore, a single index cannot be used for both methods. However, the results are the same as those shown in Section "Water balance", regardless of the used water balance method.

TABLE 1. Descriptive Statistics from the obtained Harvest Indexes (HI).

Statistics	FAOHI	TMHI
Count	50	50
Minimum	0.129	0.108
1 st Quartile	0.225	0.175
Median	0.252	0.208
Mean	0.263	0.209
3 rd Quartile	0.305	0.239
Maximum	0.435	0.382
Standard Deviation	0.067	0.059
Coefficient of Variation	25.41%	28.33%
Pearson	0.869*	
Willmott	0.406	

*: significant at 5%

Additionally, the indices found in the literature do not allow obtaining statistically satisfactory yield estimates using agrometeorological models (Araújo et al., 2011). Steduto et al. (2012) recommended that HI should be calibrated for each region. Furthermore, the authors found HI values from 0.25 to 0.40 for oilseed crops using a diverse methodology to estimate Y_r . The results of this study corroborate with Steduto et al. (2012) and the pilot project (Richetti et al., 2015) conducted at a regional level. Kemanian et al. (2007) proposed a linear model to estimate HI values for barley (0.40 to 0.60), spring wheat (0.38 to 0.46), and sorghum (0.45 to 0.53).

Water balance

The soil water balance and ET_a for each farm area were estimated by the FAOWB and TMWB methods. No significant differences were observed between water balances. Therefore, both methods could monitor water stress during the soybean cycle. An example is the farm chosen at

random (Figure 3) and designated as No. 34 (P-34 in Figure 1) for the 2013-2014 growing season. However, Katerji et al. (2011) used the FAOWB method under saline environment conditions for potato and broad beans and concluded that k_s calculations underestimated the evapotranspiration measured during the crop cycle by 12%, on average. Rojas (2007) concluded that the FAOWB method together with data from remote sensing can be used to conduct operational predictions of corn yield. Schwantes et al. (2016) compared the TMWB model with data from a polymer tensiometer in a field experiment with soybean and concluded that the time course of the soil water storage profiles calculated with water balances based on the evapotranspiration models measured with the polymer tensiometer was very similar. Thus, the TMWB method can lead to an immediate response in yield. An immediate decrease in yield is observed with water stress, while the FAOWB method leads to a higher reduction in the relative yield, but not immediate.

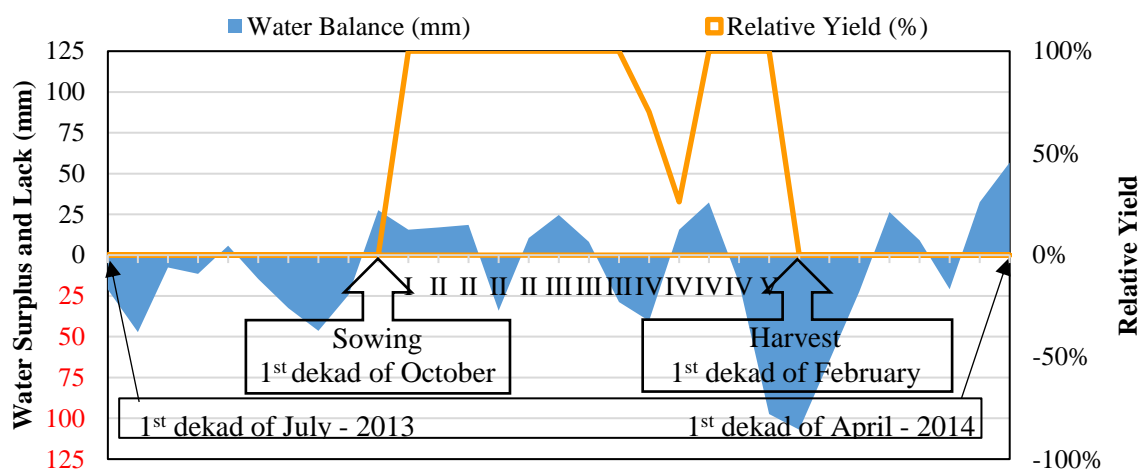


FIGURE 3. Soil Water Balance and relative yield by FAOWB method for the soybean crop in crop-season 2013-2014 at P-34 farm area.

Sowing in the example farm (P-34) was carried out at the first dekad of October, with harvest at the first dekad of February. ET_a presents its first decrease at the first dekad of stage IV (Figure 4), leading to a decrease in the relative yield. The next reduction in ET_a occurred when the cycle was ending, and the water requirements were lower than in other stages of the crop. Therefore, there is no relative yield loss.

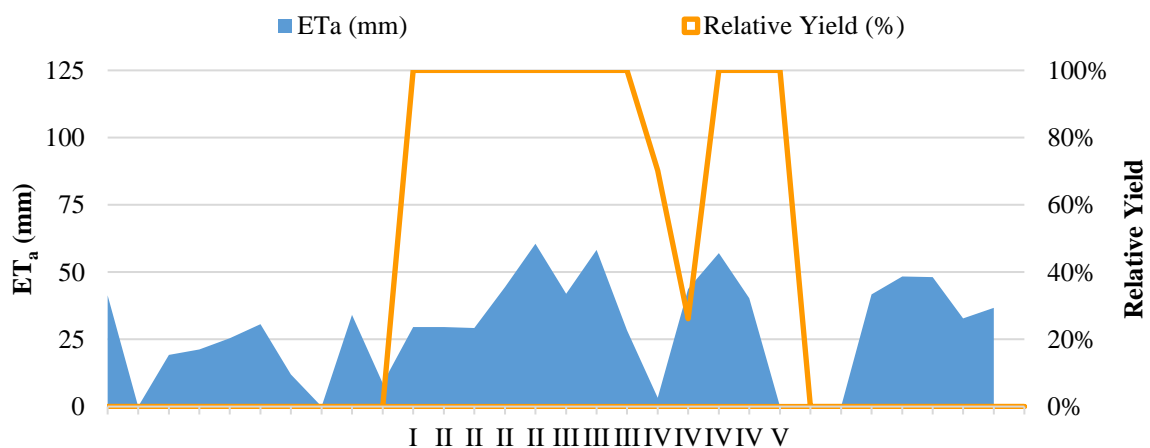


FIGURE 4. ET_a and relative yield by FAOWB method for the soybean crop in crop-season 2013-2014 at P-34 farm area.

Yield estimations

Sowing and harvest dates varied according to the location of each area, but most of the farms carried sowing in the first 10-day of October and harvest in the second 10-day of February. The ME value showed an underestimation of 310 kg ha⁻¹ and the MAPE value of 23%, regardless of the chosen water balance method. Therefore, both methods can be used equally to estimate soybean yield (Table 2). These results are slightly lower than those obtained by Araújo et al. (2011), who found an estimated yield of 3567 kg ha⁻¹, but for only one region and growing season in the state of Paraná. The average estimated yield was 12% higher than the average yield officially reported in Brazil and 6% higher than the average observed in the state (CONAB, 2018). Most farms presented observed yields of around 3100 to 3700 kg ha⁻¹, with a coefficient of variation of 21%.

TABLE 2. Descriptive statistics for observed and estimated yields and statistical metrics between them.

Statistics	Y _{obs} (kg ha ⁻¹)	Y _{est} FAOWB (kg ha ⁻¹)	Y _{est} TMWB (kg ha ⁻¹)
Count	24	24	24
Minimum	994	2307	2027
1 st Quartile	3112	2902	2870
Median	3402	3219	3142
Mean	3339	3090	3090
3 rd Quartile	3732	3312	3375
Maximum	4512	3779	3945
Standard Deviation	689.9	413.6	501.2
Coefficient of Variation	20.66%	13.38%	16.22%
	MAE	667.6	682.9
	ME	-310.9	-311.1
	RMSE	862.5	847.6
	MAPE	23.3%	23.3%

An *HI* value of 0.26 was used as a calibration index in this study. Moreover, the indices obtained in the literature do not allow obtaining statistically satisfactory yield estimates with the agrometeorological models (Araújo et al., 2011). Steduto et al. (2012) recommended that *HI* values should be calibrated for each region. Furthermore, these authors used a diverse methodology for estimating Y_x and found that the *HI* value for oilseed crops ranged from 0.25 to 0.40. The results obtained using the FAO-56 method corroborated with Steduto et al. (2012) and a pilot project (Richetti et al., 2015) conducted at the regional level. Kemanian et al. (2007) proposed a linear model to estimate *HI* values for barley (0.40 to 0.60), spring wheat (0.38 to 0.46), and sorghum (0.45 to 0.53).

CONCLUSIONS

This study estimated soybean yield using only rainfall, ET_o , average temperature, *TAW*, sowing and harvest dates, and yield as inputs. The use of low inputs, in which field data are almost unnecessary (sowing and harvest dates and yield from farms), is a quick way to evaluate agricultural production. The method underestimated the observed values, with a mean error of -310 kg ha⁻¹. Field verifications are necessary, although costly, as there are many displacement

difficulties, some producers are closed and mistrusted, and others have no record of previous growing seasons. Soybean yield estimations using *HI* values of 0.2625 (using FAOWB) and 0.2093 (using TMWB) are recommended. Furthermore, an improvement in the estimates can be obtained using the dual crop coefficient (Allen et al., 2005). Another improvement is the use of an equation/function for *HI* rather than a single value for all areas, as presented by Kemanian et al. (2007). This study has been implemented in simple software, in which the calculations will be assessed on a refined spatial-temporal scale and daily based.

ACKNOWLEDGMENTS

Special thanks to the farmers who provided the data, the support of the Coordination for the Improvement of Higher Education Personnel – Financing Code 001, the Brazilian National Council for Scientific and Technological Development (CNPq), and Araucária Foundation for financial support. To the Laboratory of Applied Statistics (LEA), Spatial Statistics Laboratory (LEE-LISA), Geotechnology and Data Science Research Center (GeoScience) of the Western Paraná State University (UNIOESTE, campus of Cascavel) for the facilities to conduct the research and the Graduate Program in Agricultural Engineering (PGEAGRI).

REFERENCES

- Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration - Guidelines for computing crop water requirements - Irrigation and drainage paper FAO-56. Rome, v1
- Allen RG, Pereira LS, Smith M, Raes D, Wright JL (2005) Dual Crop Coefficient Method for Estimating Evaporation from Soil and Application Extensions. *Journal of Irrigation and Drainage Engineering* 131: 2–13. DOI: [http://doi.org/10.1061/\(ASCE\)0733-9437\(2005\)131:1\(2\)](http://doi.org/10.1061/(ASCE)0733-9437(2005)131:1(2))
- Antle JM, Jones JW, Rosenzweig C (2017). Next generation agricultural system models and knowledge products: Synthesis and strategy. *Agricultural Systems* 155: 179–185. DOI: <http://doi.org/10.1016/j.agsy.2017.05.006>
- Aparecido LE de O, Rolim GDS, Richetti J, Souza PS de, Johann, JA (2016) Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. *Ciência e Agrotecnologia* 40(4): 405–417. DOI: <http://doi.org/10.1590/1413-70542016404003916>
- Araújo MA, Souza, JLM de, Tsukahara RY (2011) Modelos agro-meteorológicos na estimativa da produtividade da cultura da soja na região de Ponta Grossa, Estado do Paraná. *Acta Scientiarum. Agronomy* 33(1): 23–31. DOI: <http://doi.org/10.4025/actasciagron.v33i1.5062>
- Black PE (2007) Revisiting the Thornthwaite and Mather Water Balance. *JAWRA Journal of the American Water Resources Association* 43(6): 1604–1605. DOI: <http://doi.org/10.1111/j.1752-1688.2007.00132.x>
- Boogaard HL, Van Diepen CA, Rötter RP, Cabrera JMC, Van Laar HH (2014) WOFOST Control Centre 2.1 and WOFOST 7.1.7 - User's guide for the WOFOST CONTROL CENTRE 2.1 and the crop growth simulation model WOFOST 7.1.7. Wageningen. Available: <http://www.wageningenur.nl/wofost>

- CONAB (2018) Acompanhamento da Safra Brasileira - Grãos. Brasília. Available: https://www.conab.gov.br/info-agro/safra/safra/boletim-da-safra-de-graos/item/download/12569_5b3e0e675171f49a5b1e9215edc1064a
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balsameda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, Van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kállberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, De Rosnay P, Tavolato C, Thépaut J-N, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137(656): 553–597. DOI: <http://doi.org/10.1002/qj.828>
- EMBRAPA (2011) Novo Mapa de Solos do Brasil. Rio de Janeiro. Available: <http://www.cnps.embrapa.br/solosbr/publicacao.html>
- EMBRAPA (2009) Sistema brasileiro de classificação de solos, EMBRAPA. 2 ed. v1.
- FAOSTAT (2018). Production quantities by country. Available: <http://faostat3.fao.org/browse/Q/QC/E> Accessed Aug 11, 2015.
- Farias JRB, Nepomuceno AL, Neumaier N (2007) Ecofisiologia da soja. Embrapa, p9. Circular Técnica 48.
- Gusso A, Arvor D, Ducati JR (2017) Model for soybean production forecast based on prevailing physical conditions. *Pesquisa Agropecuária Brasileira* 52(2): 95–103. DOI: <http://doi.org/10.1590/S0100-204X2017000200003>
- Hagedorn R, Hamill TM, Whitaker JS (2008) Probabilistic forecast calibration using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures. *Monthly Weather Review* 136: 2608–2619. DOI: <http://doi.org/10.1175/2007MWR2410.1>
- IAPAR (2015) Rede de Estações Meteorológicas no Paraná. Available: <http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=890>. Accessed Aug 19, 2015.
- Johann JA, Rocha J, Lamparelli R (2016) Estimativa de área e produtividade de culturas agrícolas: uma aplicação de Sensoriamento Remoto. *Novas Edições Acadêmicas*. Available: <https://www.morebooks.de/store/gb/book/1/isbn/978-3-330-74365-6>
- Johnson DM (2014) An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment* 141: 116–128. DOI: <http://doi.org/10.1016/j.rse.2013.10.027>
- Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singhe U, Gijsmana AJ, Ritchie JT (2003). The DSSAT cropping system model. *Journal Agronomy* 18: 235–265. DOI: [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
- Kassam AH, Higgins GM (1981) The FAO Agro-Ecological Zone Approach to Determination of Land Potential. *Pedologie* 31(2): 147–168. Available: http://www.bbv-sbss.ugent.be/pedologie/pedologie_1981_2.pdf
- Katerji N, Mastrorilli M, Lahmar F (2011) FAO-56 methodology for the stress coefficient evaluation under saline environment conditions: Validation on potato and broad bean crops. *Agricultural Water Management* 98(4): 588–596. DOI: <http://doi.org/10.1016/j.agwat.2010.10.011>
- Kemarian AR, Stöckle CO, Huggins DR, Viegas LM (2007) A simple method to estimate harvest index in grain crops. *Field Crops Research* 103(3): 208–216. DOI: <http://doi.org/10.1016/j.fcr.2007.06.007>
- Moges SM, Girma K, Teal RK, Freeman KW, Zhang H, Arnall DB, Holtz SL, Tubaña BS, Walsh O, Chung B, Raun WR (2007) In-season estimation of grain sorghum yield potential using a hand-held optical sensor. *Archives of Agronomy and Soil Science* 53(6): 617–628. DOI: <http://doi.org/10.1080/03650340701597251>
- Richetti J, Johann J, Uribe-opazo, MA (2015) Calibração do índice de colheita (Cc) para estimativa de produtividade atingível da cultura da soja. In *Simpósio Brasileiro de Sensoriamento Remoto*. João Pessoa, p3495-3502. Available: <http://www.dsr.inpe.br/sbsr2015/files/p0690.pdf>
- Richetti J, Judge J, Boote KJ, Johann JA, Uribe-Opazo MA, Becker WR, Paludo A, Silva LC de A (2018) Using phenology-based enhanced vegetation index and machine learning for soybean yield estimation in Paraná State, Brazil. *Journal of Applied Remote Sensing* 12(02): 1. DOI: <http://doi.org/10.1117/1.JRS.12.026029>
- Rojas O (2007) Operational maize yield model development and validation based on remote sensing and agro-meteorological data in Kenya. *International Journal of Remote Sensing* 28(17): 3774–3793. DOI: <http://doi.org/10.1080/01431160601075608>
- Sakamoto T, Gitelson AA, Arkebauer TJ (2014) Near real-time prediction of U.S. corn yields based on time-series MODIS data. *Remote Sensing of Environment* 147: 219–231. DOI: <http://doi.org/10.1016/j.rse.2014.03.008>
- Schwantes AP, Reichardt K, Neto DD, Durigon A, Pinto VM (2016) Soil water storage in soybean crop measured by polymer tensiometers and estimated by agrometeorological methods. *Journal of Agricultural Science* 8(7): 30–37. DOI: <http://doi.org/doi:10.5539/jas.v8n7p30>
- Steduto P, Hsiao TC, Fereres E, Raes D (2012) *Crop Yield Response to Water - Irrigation and Drainage Paper* FAO-66. Rome.