

# Modeling of corn yield in Brazil as a function of meteorological conditions and technological level

José Eduardo Boffino de Almeida Monteiro<sup>(1)</sup>, Eduardo Delgado Assad<sup>(1)</sup>,  
Paulo Cesar Sentelhas<sup>(2)</sup> and Letícia da Costa Azevedo<sup>(1)</sup>

<sup>(1)</sup>Embrapa Informática Agropecuária, Avenida André Tosello, nº 209, Campus da Unicamp, Barão Geraldo, Caixa Postal 6041, CEP 13083-886 Campinas, SP, Brazil. E-mail: eduardo.monteiro@embrapa.br, eduardo.assad@embrapa.br, letsdacosta@yahoo.com.br <sup>(2)</sup>Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Departamento de Engenharia de Biossistemas, Avenida Pádua Dias, nº 11, Caixa Postal 9, CEP 13418-900 Piracicaba, SP, Brazil. E-mail: pcsentel.esalq@usp.br

**Abstract** – The objective of this work was to develop and evaluate a method for estimating corn yield using a minimum number of parameters and limited information about crop management. The proposed method estimates potential and attainable yields based on the technological level of the production systems and on relatively simple agrometeorological models. Corn yield was estimated for the crop seasons from 2000/2001 to 2007/2008, considering several locations and regions in Brazil, and was compared with the actual yield data from official surveys. There was a high correlation between the estimated and observed yield ( $0.76 \leq R^2 < 0.92$ ;  $p < 0.01$ ), with model efficiency (E1') ranging from 0.45 to 0.73; mean relative error (MRE) between -0.9 and 2.4%; and mean absolute error (MAE) of less than 70 kg ha<sup>-1</sup>, depending on the technological level adopted. Based on these results, the proposed yield model can be recommended to forecast yields all over the country, contributing to make this process more precise and accurate.

**Index terms:** *Zea mays*, large-area crop modeling, parametrization, risk analysis, technological potential yield, yield forecast.

## Modelagem da produtividade de milho no Brasil em função das condições meteorológicas e do nível tecnológico

**Resumo** – O objetivo deste trabalho foi desenvolver e avaliar um método para estimar a produtividade de milho com uso de um número mínimo de parâmetros e de informações limitadas sobre o manejo da cultura. O método proposto estima rendimentos potenciais e atingíveis com base no nível tecnológico dos sistemas de produção e em modelos agrometeorológicos relativamente simples. A produtividade de milho foi estimada para as safras de 2000/2001 a 2007/2008, tendo-se considerado vários locais e regiões do Brasil, e comparada aos dados de produção reais de levantamentos oficiais. A produtividade estimada apresentou alta correlação com a observada ( $0,76 \leq R^2 < 0,92$ ;  $p < 0,01$ ), com eficiência do modelo (E1') entre 0,45 e 0,73; erro médio relativo (MRE) entre -0,9 e 2,4%; e erro médio absoluto (MAE) inferior a 70 kg ha<sup>-1</sup>, de acordo com o nível tecnológico considerado. Com base nestes resultados, este modelo pode ser recomendado para estimativas de produtividade em todo o País, e contribuir para tornar este processo mais preciso e exato.

**Termos para indexação:** *Zea mays*, modelagem de cultura em grandes áreas, parametrização, análise de risco, produtividade potencial tecnológica, previsão de safra.

### Introduction

Crop yield forecasting is useful at the farm level for crop management and marketing decisions, and at the government level for policy issues and food security actions (Hoogenboom, 2000; Hansen, 2005).

Many models have been developed to estimate crop yield considering different parameters and complexity levels. Challinor et al. (2007) discussed several approaches used, such as crop simulation models, empirical models, and yield transfer functions. Among

these models, those incorporated in the platforms Decision Support System for Agrotechnology Transfer (DSSAT) and Agricultural Production Systems Simulator (APSIM) are widely studied (Tojo Soler et al., 2007; Knörzer et al., 2011). The general large area model, proposed by Challinor et al. (2004) and adapted by Bergamaschi et al. (2013) for corn (GLAM-Maize), is another example. However, most of the methods demand data and parameterization that are not easily available (Zhao et al., 2014), specially for large and heterogeneous areas (Reidsma et al., 2009).

Simpler yield models, as those based mostly on water availability throughout the crop cycle, have also been used to estimate potential and actual yields (Assad et al., 2007; Monteiro et al., 2013), as well as to characterize genotype sensitivity to water deficits (Andrioli & Sentelhas, 2009; Battisti & Sentelhas, 2015). Among these, the FAO method (Doorenbos & Kassam, 1979), which uses a physiological-mathematical approach, has proved to be robust, since it is based on a conservative relationship between biomass and water use (Hsiao, 1993; Morison et al., 2008).

All these approaches have considered different spatial scales, but none have considered very large and heterogeneous areas, like those of corn crops in Brazil. Data from the Brazilian Institute of Geography and Statistics (IBGE, 2013) reveal neighboring municipalities – with similar climatic and soil conditions – with corn yields ranging from 500 to more than 8,000 kg ha<sup>-1</sup>. This heterogeneity has been a real challenge to obtain adequate parameterization of crop models. As farmers have different information access and behavior towards technologies and crop management improvements (Liang et al., 2016), their yields may differ widely. In its periodical surveys of the Brazilian agricultural production, the Companhia Nacional de Abastecimento (Conab) highlights the direct relationships observed between yield gains and changes in the technological packages (Acompanhamento..., 2013). In this context, the yield is directly related to the technological level.

Corn crops are highly demanding in terms of water availability (Gong et al., 2015), which is key aspect for corn yield. Rain amount and distribution are usually the main yield factors influencing variability under rainfed conditions (Bergamaschi et al., 2004; Araus et al., 2012). Ecophysiological studies have demonstrated how drought affects yield differently depending on crop stage. During establishment, drought can reduce plant germination; while water stress during leaf area expansion reduces leaf area and radiation interception (Yan et al., 2016). Later, it will reduce green leaf duration – with accelerated senescence – and reduce radiation-use efficiency. Drought also causes induced infertility and kernel abortion or shriveled grain, which can strongly reduce harvest index. That's why crop stages from tasseling to beginning of grain filling are critical for corn yield (Zinselmeier et al., 1995).

Finally, during the grain filling period, water deficit reduces the supply of assimilates at the source, leading to losses in grain weight (Araus et al., 2012).

The hypothesis of this study is that it is possible to estimate corn yield with a simple agrometeorological model fitted according to yield class with generalized parameters for unknown site-specific yield factors.

The objective of this work was to develop and evaluate a method for estimating corn yield in Brazil using a minimum number of parameters and limited information about crop management.

## Materials and Methods

The model for estimating corn yield in the present study follows the theoretical basis proposed by Doorenbos & Kassam (1979), which is based on the relationship between relative yield loss and relative water deficit throughout the crop growth stages. It comprises two modules: one that estimates potential yield and another that penalizes the potential yield by water deficit.

The original model estimates yield loss and actual yield ( $Y_a$ ) as a relationship between relative water deficit ( $1 - E_{Ta}/E_{Tc}$ ) and relative crop yield loss ( $1 - Y_a/Y_p$ ), weighted by the yield response factor to water deficit, for each growth stage. Yield response factor ( $k_y$ ) represents the slope of the relationship between relative yield reduction and relative water deficit (Doorenbos & Kassam, 1979). Reference values of  $k_y$  for each crop are usually determined based on field experiments. These relationships are described in the equation:  $Y_a/Y_p = 1 - k_y \times [1 - (E_{Ta}/E_{Tc})]$ , where:  $Y_a$  is the actual yield,  $Y_p$  is the potential yield,  $E_{Ta}$  is the actual evapotranspiration, and  $E_{Tc}$  is the crop maximum evapotranspiration. In this equation,  $E_{Ta}/E_{Tc}$  accounts for the weather conditions or, more precisely, for the soil water availability throughout the crop cycle.

Actual crop evapotranspiration ( $E_{Ta}$ ) is usually estimated by the crop water balance, which accounts for daily changes in soil water content. There are several approaches to perform that, and, in this study, the method of Thornthwaite & Mather (1955) was used. Crop maximum evapotranspiration ( $E_{Tc}$ ) is given by the product between reference evapotranspiration ( $E_{To}$ ) and crop coefficient ( $k_c$ ).  $E_{To}$  was estimated with the method of FAO 56 Penman-Monteith (Allen

et al., 1998). Crop coefficient is the ratio between  $ET_c$  and  $ET_o$ , and varies according to crop type and canopy growth (leaf area index). Reference values of  $k_c$  for each corn crop phase will be presented later.

A modification in this original model was proposed here by the introduction of an empirical parameter acting as a correction to the original yield response factor ( $ky$ ). This correction is indicated by  $ky^*$ , inserted in the model as a multiplier of  $ky$ . If  $ky^*$  is greater than 1.0, it results in higher response of the crop to water deficit. Conversely, if it is lower than 1.0, it results in a lower response of the crop to water deficit. Thus, this parameter must be calibrated in order to match the response of the actual yield series in each yield class. Yield class is described latter. When considering  $ky^*$  in the model, it resulted in the equation:  $Y_a/Y_p = 1 - (ky \times ky^*) \times [1 - (ET_a/ET_c)]$ .

Another modification to the original model was the use of a new reference for potential yield, named as technological potential yield ( $TY_p$ ).  $TY_p$  was estimated for each yield class, based on municipality data of eight years of actual yield from IBGE, as follows. In a first step, a simple linear regression between actual yield ( $Y_a$ ) and time (year) was used. The regression line generated by this function defines the actual yield trend over years ( $Y_a'$ ). While  $Y_a$  is the actual yield of a given year,  $Y_a'$  is the actual yield given by the adjusted equation, considering the overall trend. Deviations that may occur above and below this trend ( $Y_a'$ ) are caused by variations in the yield factors, including weather conditions, soil water availability, and soil and crop management. In the second step, a correction factor, named technological potential yield coefficient ( $dt_y$ ) was determined. This coefficient is a dimensionless "delta value" ( $d$ ) used to determine the technological potential yield ( $TY_p$ ) for each year ( $n$ ), when multiplied by  $Y_a'$ , as follows:  $TY_p = Y_a' \times dt_y$ .

Once  $dt_y$  and  $ky^*$  are introduced in the original model, the final model can be expressed by:  $Y_a/TY_p = 1 - (ky \times ky^*) \times [1 - (ET_a/ET_c)]$ .

$TY_p$  is used as the initial reference value, to begin the calculation process in the first day of the crop cycle. Once  $Y_a$  of day one ( $Y_{a_1}$ ) is calculated, it is used as the new reference value replacing  $TY_p$  in day two, in order to calculate  $Y_{a_2}$ , and so on, as presented in the following equation:  $Y_{a_i}/Y_{a_{(i-1)}} = 1 - (ky \times ky^*) \times [1 - (ET_{a_i}/ET_{c_i})]$ ; in which  $i$  is the day of the crop cycle. This equation is used

as a recurrent equation at daily time step, which also enables the present model to be used for monitoring expected yield with daily updates. At this point,  $dt_y$  and  $ky^*$  are still not known, and these are the two parameters to be set and calibrated for each yield class, as it will be described latter.

The proposed model was developed and calibrated based on corn yield data from the series of crop production and harvested area surveyed by IBGE, between 2000 and 2008 (IBGE, 2013). Out of the 5,570 Brazilian municipalities, 4,965 had complete corn yield data for the eight crop seasons. The municipalities were classified into five categories according to the average yield of the last four seasons. The cut-off limits for each yield class were as follows: class 1, from 0 to 2,000 kg ha<sup>-1</sup>; class 2, from 2,001 to 4,000 kg ha<sup>-1</sup>; class 3, from 4,001 to 6,000 kg ha<sup>-1</sup>; class 4, from 6,001 to 8,000 kg ha<sup>-1</sup>; and class 5, from 8,001 to 10,000 kg ha<sup>-1</sup>. The definition of yield classes was based on a preliminary analysis, in which the objective was to set a representative yield trend for each group of municipalities. Narrower classes would allow better fit of the yield trend, but they would generate a greater number of classes to be parameterized. The classification of municipalities in each yield class was also based on the average yield obtained in the last four years of the series, from 2005 to 2008. All 4,965 municipalities with complete yield series were used to determine the yield trend for each yield class.

Daily weather data was obtained from weather stations spread throughout the Brazilian states. Rain data and estimated reference evapotranspiration ( $ET_o$ ) were used as one of the inputs for the model. These stations are managed by governmental and private institutions, and are integrated at Embrapa's Agrometeorological Monitoring System (Agritempo). Out of the total of municipalities with yield and weather data series, only those that presented complete data series were selected. Municipalities whose weather data sets had missing or erroneous values were disregarded. The composition of samples evenly distributed for each yield class also reduced the number of useful paired series (yield  $\times$  weather), as there was a much smaller number of municipalities in the higher yield classes. The highest yield class, class 5, comprehended a small number of municipalities and was not considered as representative for this study. When this selection was carried out, the period between 2000 and 2008

comprehended a reasonable number of locations. This criterion resulted in a total of 143 municipalities, among which 72 were used for model calibration, and the other 71 for validating model's performance.

The most common sowing periods were identified for each Brazilian state, according to Conab's annual surveys (Acompanhamento..., 2009). The average sowing season for each state was divided into five similar sub-periods. The central day of each sub-period was used as the representative sowing date, which resulted, therefore, in five simulations and five yield results for each year. The final result of each year was given by the weighted average of these five values, considering the averaged frequency distribution of planted area, typical for each state, as reported by Acompanhamento... (2009). In the state of Goiás, for example, the regular sowing period lasts typically for nearly 60 days, from mid-October to mid-December, with the highest concentration in the center of the period (mid-November). In this example, each sub-period had 12 days, and its representative sowing dates were Oct 20<sup>th</sup>, Nov 01<sup>st</sup>, Nov 14<sup>th</sup>, Nov 26<sup>th</sup> and Dec 9<sup>th</sup>. The yield for each of these dates accounted for 10.0, 22.5, 35.0, 22.5, and 10.0% of the total yield in the evaluated period, respectively. This procedure was used to reduce the main source of possible errors associated with water availability. As the sowing period may last for up to four months, in some states, this may result in different climatic and soil water conditions in the sites where sowing was in the beginning or in the end of the period.

Crop coefficient ( $k_c$ ) and yield response factor ( $k_y$ ) for corn are available in the literature, and they are normally determined for four crop phases, as follows: crop establishment, vegetative growth, flowering, and maturation. The  $k_c$  values adopted for these crop phases were 0.4, 0.8, 1.2, and 0.9, respectively (Piccinni et al., 2009). The  $k_y$  values adopted for establishment, vegetative growth, flowering and maturation were respectively 0.0, 0.4, 1.50, and 0.35 (Andrioli & Sentelhas, 2009). These parameters were converted into daily values by using a polynomial equation fitted to the original data, distributed in a typical corn cycle of 130 days. At this point, there is no concern about if this produce is accurate or not, since this was just done to generate an appropriate initial curve of  $k_y$ , once final sensitivity to water stress was set later, with the calibration of the correction factor ( $k_y^*$ ).

As it was not possible to determine the most suitable soil water holding capacity for all municipalities, an average value of 50 mm was adopted, considering an effective root system depth of 0.5 m (Hund et al., 2009) and an average water retention of 1 mm cm<sup>-1</sup>, typical of Brazilian soils (Araújo & Assad, 2001). Even considering the limitations of this approach, it was considered more realistic for practical purposes than setting a specific value for each soil type, given the great variety of soils and their spatial variability.

Each yield class was calibrated with its own yield trend and technological potential yield (TYp). Therefore, each yield class was analyzed independently, with a specific calibration for  $dty$  and  $ky^*$ . The following steps were used to determine  $dty$  and  $ky^*$  and to calibrate the model: firstly, an arbitrary value of 1.0 was given to both, which made TYp equal to the actual yield trend, and it did not change the sensitivity of the crop to water deficit; secondly, their values were progressively increased or decreased according to the changes observed in model performance, using an iterative process performed in a programmed electronic spreadsheet. Both parameters were calibrated simultaneously, in order to maximize the coefficient of determination ( $R^2$ ) and to minimize mean absolute error (MAE) between estimated and observed yields. In this process, when the variability of the estimated series is lower than the one of the observed series, that results in low  $R^2$ . Then, in the next step,  $ky^*$  should be increased. This is repeated until there is a productivity series with variability similar to that observed productivity, which results in a greater  $R^2$ . However, higher  $ky^*$  reduce estimated yield, which can cause underestimation of this parameter and, therefore, an increased MAE. In this case,  $dty$  should be gradually increased to obtain the minimum MAE possible.

An independent data set was used for model validation. From the total of 143 municipalities, 72 of them were used for model calibration and the other 71 for evaluations of model performance. Estimated yields were compared to observed yields surveyed by IBGE, as previously described.

A regional test was performed to verify general calibration in regionalized subsamples. The municipalities used for validation were divided in subgroups, one for each of the five macro-regions across the country. Then,  $R^2$ , MRE, and  $E1'$  were

calculated on these subgroups. Local testing was performed in two municipalities to illustrate the differences between results from general calibration and specific calibration for each municipality.

Model performance was assessed by regression analysis, in which significance of the coefficient of determination ( $R^2$ ) was determined by T test at 5 and 1% probability. Mean absolute error (MAE) and mean relative error (MRE) were also determined.

Model accuracy was evaluated according to Legates' modified index of model efficiency ( $E_1'$ ) (Legates & McCabe Jr, 1999), as follows:  $E_1' = 1 - \Sigma(|O_i - P_i|) / \Sigma(|O_i - O|)$ , where  $O_i$  and  $P_i$  stand, respectively, for observed and estimated  $Y_a$  for a given year of the yield series; and  $O$ , for the average observed  $Y_a$ . Thus,  $E_1' = 0$  indicates that observed mean is as good predictor as the model;  $E_1' < 0$  indicates that the observed mean is a better predictor than the model; and  $E_1' > 0$  indicates that the model is a better predictor than the observed mean.

## Results and Discussion

In Brazil, approximately 5,000 municipalities contribute to annual corn production (IBGE, 2013), with 43.6% of them exhibiting mean actual yield of less than 2,000 kg ha<sup>-1</sup>, corresponding to only 5.1% of the total production. Municipalities with actual yields

between 4,000 and 6,000 kg ha<sup>-1</sup> account for 48.4% of Brazilian production. Only 5.9% of municipalities present mean yield above 6,000 kg ha<sup>-1</sup>, with only a few of them reaching 10,000 kg ha<sup>-1</sup> (Table 1). Such data reveals the diversity of corn production systems operating in Brazil, with yield gaps of different magnitudes, which makes the parameterization of a process-based crop simulation model very difficult.

The proposed model was precise and accurate (Table 2 and Figure 1). The coefficients of determination for the comparison between observed and estimated yields ranged from 0.76 to 0.92 ( $p < 0.01$ ). Even in the higher yield classes, such as 3 and 4, in which the effect of water stress and its inter-annual variability were lower,  $R^2$  was significant at 1% probability. Model accuracy, evaluated with the Legates' model efficiency ( $E_1'$ ), ranged from medium (0.45) to high (0.73). The yield classes used proved appropriated for the purposes of this study. The use of narrower classes allows better fit of the yield trend, but it would generate a greater number of classes to be parameterized. Nonetheless, the use of broader classes increases the absolute error derived from the potential yield.

The evaluation of  $E_1'$  values revealed absence of systematic errors in the estimates, which would not be possible by analyzing  $R^2$  alone. The relative error between the estimated and actual yields ranged from -0.9 to 2.4%, depending on the yield class

**Table 1.** Corn production and number of maize-producing municipalities per yield class, averaged between 2005 and 2008 growing seasons.

Class	Yield class (kg ha <sup>-1</sup> )	Production		Crop area (10 <sup>3</sup> ha)	Crop area (%)	Municipalities	
		(10 <sup>9</sup> kg)	(%)			Total	(%)
1	≤2,000	3.0	5.1	3,194	21.7	2,318	43.6
2	2,001–4,000	11.3	19.2	3,472	23.5	1,536	28.9
3	4,001–6,000	28.5	48.4	5,825	39.5	1,158	21.8
4	6,001–8,000	13.7	23.3	1,977	13.4	280	5.3
5	8,001–10,000	2.4	4.0	278	1.9	30	0.6

Source: IBGE (2013).

**Table 2.** Technological potential yield coefficient (dty), correction for yield response factor to water deficit (ky\*), and performance of the proposed model according to: mean absolute error (MAE); mean relative error (MRE); coefficient of determination ( $R^2$ ); and modified index of model efficiency ( $E_1'$ ) for each yield class.

Class	Yield class (kg ha <sup>-1</sup> )	dty	ky*	MAE (kg ha <sup>-1</sup> )	MRE	$R^2$	$E_1'$
1	0–2,000	1.60	2.0	9	1.0%	0.76**	0.45
2	2,001–4,000	1.35	2.0	69	2.4%	0.90**	0.54
3	4,001–6,000	1.40	1.8	38	0.8%	0.83**	0.54
4	6,001–8,000	1.45	1.6	64	-0.9%	0.92**	0.73
5	8,001–10,000	- <sup>(1)</sup>	-	-	-	-	-

<sup>(1)</sup>Data not available, representing only 1.9% of the corn crop area and 0.6% of the municipalities. \*\*Significant at 1% probability.

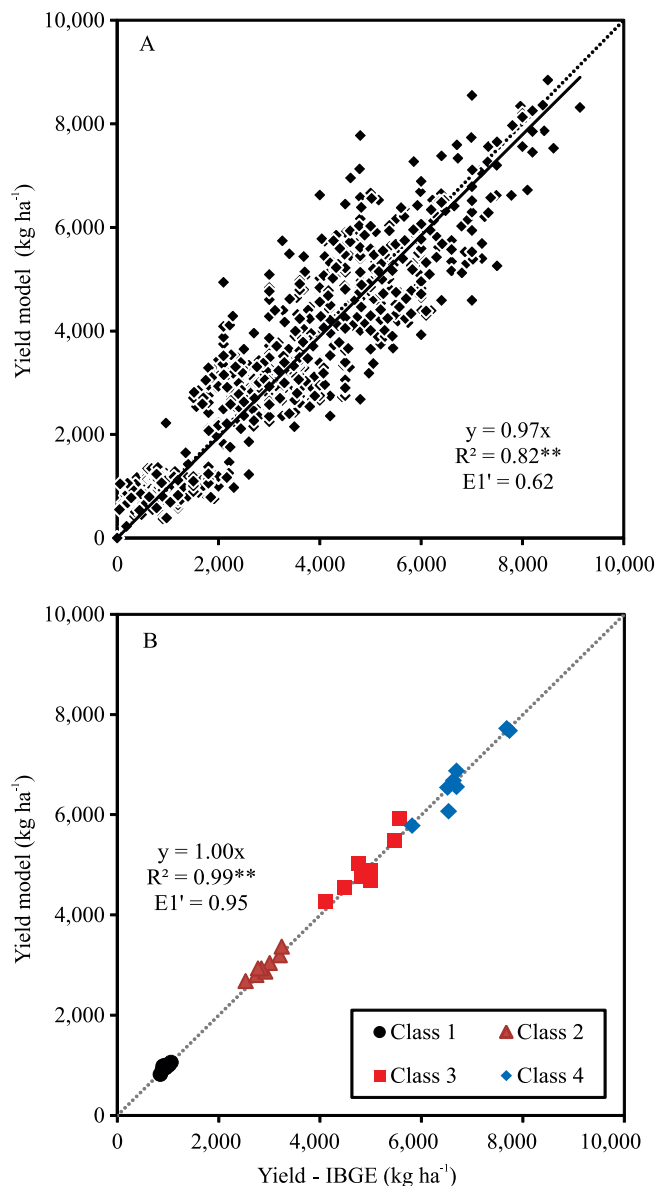
(Table 2). Andrioli & Sentelhas (2009) tested a simple agrometeorological yield model for 26 corn genotypes, under experimental conditions, and reported mean errors between 3.9 and 15%, which are acceptable for yield modeling. However, these authors analyzed the MRE per field trial, whereas in the present study MRE was determined for each yield class. Tojo Soler et al. (2007), using the DSSAT Ceres-Maize model to

estimate rainfed and irrigated actual yields for a single location, obtained MRE between -10.7 and 11.3%.

In most of the crop seasons, the estimated corn yields were very close to those reported by IBGE, with very small deviations from year to year (Figure 2). Since systematic errors can be fixed via calibration, the most important factor in this kind of modeling is to observe if the model is able to quantify the inter-annual yield variability. Thus, more attention was paid to cases when observed yields exhibited increasing or decreasing trend, and the estimated yields presented opposite behavior. In such cases, the error source was possibly related to changes in the crop management practices, i.e. cultivar used, soil management, fertilization, etc. Changes in management practices may occur from one year to another (Acompanhamento..., 2013) and, moreover, there are some uncontrollable factors – such as pests and diseases – whose intensity may vary from year to year. These factors were not computed by the proposed model since they are not consistently available in all studied areas.

Therefore, unpredictable variations may occur and compromise model performance, as indicated by  $R^2$ , MAE, MRE, and  $E_i'$  (Table 2). The presence of irrigated areas in a given municipality is another error source that may add to unpredictability. As the proposed model was developed for rainfed conditions, it cannot account for the yield increments caused by supplementary water application. However, such aspect can also be considered in the proposed model by calibrating  $dty$  and  $ky^*$ , considering the proportion of irrigated areas in a single municipality, or in a group of them. Thus, classifying municipalities by yield class and calibrating  $dty$  and  $ky^*$  individually would allow identifying the areas where irrigation is predominant and, consequently,  $dty$  and  $ky^*$  is lower. In the areas where rainfed conditions are predominant,  $dty$  and  $ky^*$  values are higher. If the corn crop of an entire area of a given municipality is irrigated,  $dty$  tends to 1, whereas  $ky^*$  tends to zero. Such procedure was used by Monteiro et al. (2013), for distinguishing lowland and upland rice yields in Brazil, and it was decisive for improving yield estimation performance for this crop.

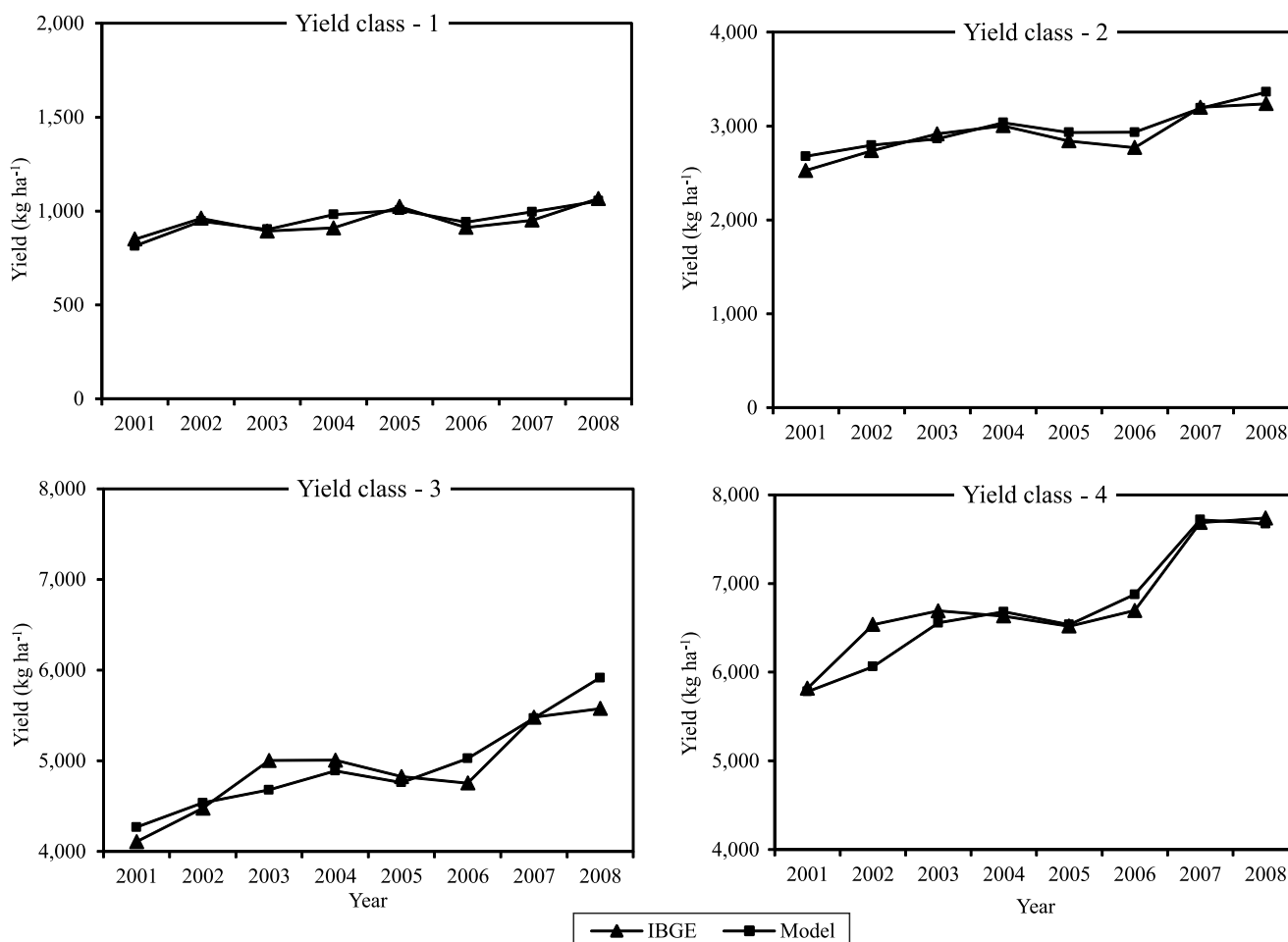
When the model was tested for subsamples, grouped into the five Brazilian macro-regions, the performance was worse, with lower  $R^2$  (0.29–0.87), and  $E_i'$  (-1.20–0.51), compared to the whole country (Table 3, Figure 3). Bergamaschi et al. (2013), using the GLAM-Maize



**Figure 1.** Relationship between observed and estimated corn yields, for each crop season at municipality level (A); and average values for each yield class (B).  $E_i'$ , index of model efficiency. \*\*Significant at 1% probability.

model, also obtained a better correlation ( $r > 0.8$ ;  $p < 0.01$ ) with larger spatial scales than with small ones ( $r$  varied from 0.65 to 0.87;  $p < 0.1$ ). However, in the present study, the cause for this result may be the unbalanced distribution of yield classes per region,

since the number of municipalities in each yield class differed from one region, or state, to another. The use of yield classes, regardless of the evaluated region, was a better approach, since great yield variability can be observed even in neighboring municipalities.



**Figure 2.** Observed and estimated corn yield from 2001 to 2008 crop seasons for yield classes 1, 2, 3, and 4.

**Table 3.** Performance of the corn yield model at regional scale: coefficient of determination ( $R^2$ ); mean relative error (MRE), and modified index of model efficiency ( $E_1'$ ) for the comparison between estimated and observed maize yield for each yield class.

Region	Yield class (kg ha <sup>-1</sup> )											
	0 to 2,000			2,001 to 4,000			4,001 to 6,000			6,001 to 8,000		
	$R^2$	MRE	$E_1'$	$R^2$	MRE	$E_1'$	$R^2$	MRE	$E_1'$	$R^2$	MRE	$E_1'$
South	-( <sup>1</sup> )	-	-	0.87**	-4.4%	0.35	0.87*	-4.6%	0.51	0.71*	-1.8%	0.34
Southeast	-	-	-	0.69**	0.6%	0.40	0.68*	10.2%	-0.21	-	-	-
Central-West	-	-	-	0.48*	1.8%	0.12	0.77*	-4.9%	0.22	0.79*	-5.4%	0.45
North	0.64*	36%	-0.19	-	-	-	-	-	-	-	-	-
Northeast	0.29	19.2%	-1.22	-	-	-	-	-	-	-	-	-

<sup>1</sup>Data not available. \* and \*\*Significant at 5% and 1% probability respectively.

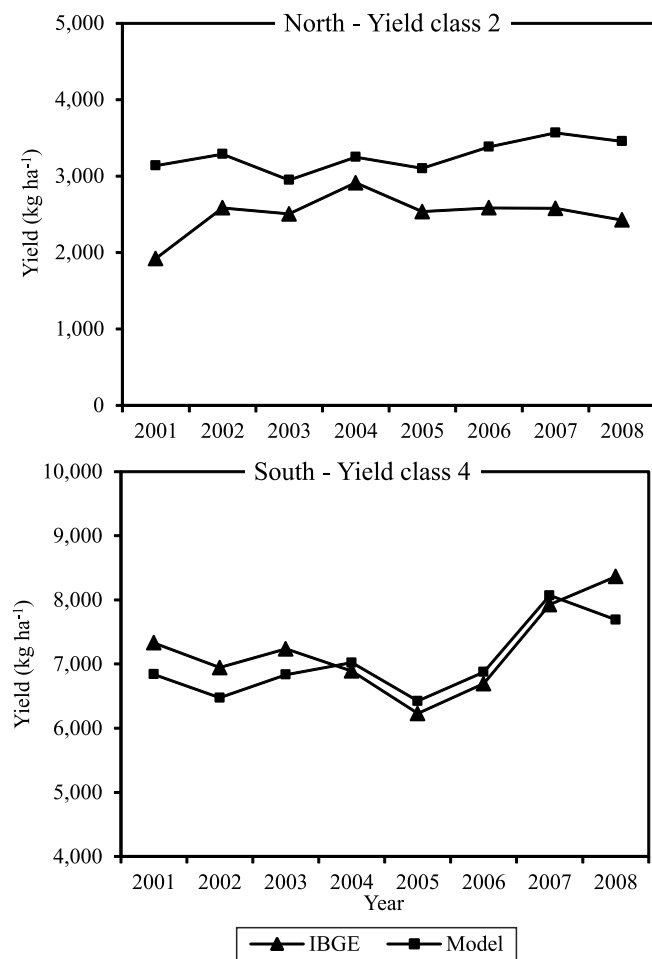
There were no municipalities in class 1 in the South, Southeast, and Central-West regions of Brazil (Table 3). This can be explained by the average higher technological level used for corn production in these regions. Under such conditions, the errors are lower; whereas there were higher systematic errors in the regions where yield classes 1 and 2 predominate, as observed in the Northern region of Brazil (Figure 3). Therefore, the use of narrower yield cutoff ranges, for classes with lower yield averages, should provide better results. Advancing a step further, the best approach most likely considers the data source at the smallest scale possible as municipality data. Therefore, instead of setting an average yield trend for a group or class of

municipalities within a yield class, each municipality should have its own trend. This should result in better correlations between estimated and actual data, lower MRE, and better  $E_1'$ .

By applying this procedure individually to the municipality of Passo Fundo, in the state of Rio Grande do Sul, and municipality of Catanduvas, in the state of Paraná, the model performance was improved. When the yield trend and the  $ky^*$  parameter were optimized for each municipality, and then tested with independent data from 2009 to 2013,  $R^2$  increased from 0.76 to 0.84 in Passo Fundo, and from 0.50 to 0.76 in Catanduvas (Figure 4).  $E_1'$  was also sharply increased, in both municipalities, evidencing the benefits of using local calibration. In this case, model performance was more comparable to those obtained by Andriolli & Sentelhas (2009) and by Tojo Soler et al. (2007), even though these studies were conducted at smaller spatial scales, with field trials.

The procedure used for assigning weights to each of the five sowing dates according to the area of each sowing day was important to assure reasonable model performance. When this distribution was changed to alternative scenarios, with sowing dates concentrated in shorter periods, this sharply reduced model performance. In the yield class 2, for example,  $E_1'$  was 0.54 when the final yield was weighted according to the average distribution of sowed area per date, throughout the sowing period. However, when sowing was concentrated solely in the first, mid or last third of the regular sowing period,  $E_1'$  was reduced to 0.13, 0.03 and 0.23, respectively.

The definition of sowing dates can greatly influence productivity results, as drought affects yield differently depending on crop stage (Zinselmeier et al., 1995; Yan et al., 2016). In the state of Goiás, for example, the regular sowing period lasts typically for nearly 60 days, from mid-October to mid-December, with the highest concentration in the center of the period (mid-November). In this example, each sub-period had 12 days, with the following representative sowing dates: Oct 20<sup>th</sup>, Nov 01<sup>st</sup>, Nov 14<sup>th</sup>, Nov 26<sup>th</sup> and Dec 9<sup>th</sup>. The yield from each of these dates accounted for 10.0, 22.5, 35.0, 22.5, and 10.0% of the total yield in the evaluated period, respectively. According to Bergamaschi et al. (2013), a wide range of sowing dates adopted by farmers is problematic for large-area crop modeling. The authors evaluated the GLAM-Maize model in northwestern Rio Grande do



**Figure 3.** Observed and estimated corn yields in the Northern and Southern regions of Brazil, from 2001 to 2008 crop seasons, for yield classes 2 (4,001–6,000 kg ha<sup>-1</sup>) and 4 (6,001–8,000 kg ha<sup>-1</sup>).

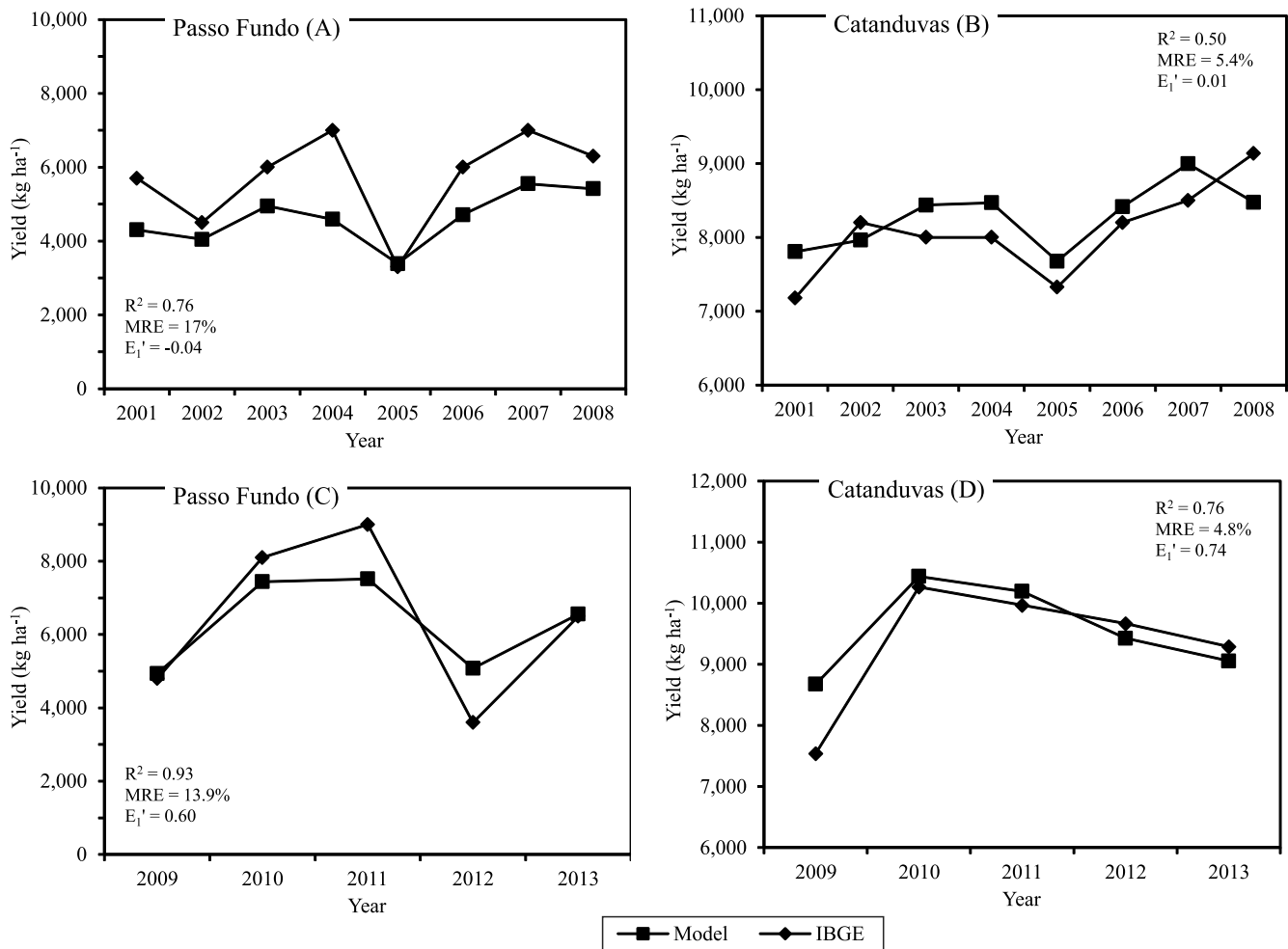


Sul, Brazil, considering only one sowing date for each municipality. Implementing an operational system for early estimation of crop yield would require updated information about sowing dates and crop areas of each site, every season. For such purpose, remote sensing tools could be useful and should be considered for further developments (Prasad et al., 2006).

Comparing potential and actual yields also allowed detecting when the effect of other production factors may have higher or lower impact on crop yield, depending on the crop season. An example of year-to-year variations can be seen in Catanduvas (Figure 4), where an overestimation of 1,135 kg ha<sup>-1</sup> (15%) was observed in 2009. However, the mean relative error for

the period, from 2009 to 2013, was only 4.8%. This is an example of eventual rather than systematic error.

In addition to the effect of water deficit on yield, the use of technological potential yield for correcting the yield trends over time was necessary to improve the accuracy and precision of the estimates, both in the low and in the high yield classes. This new reference yield used in our model represents the portion of the potential yield that can actually be obtained by the production system used in each site. This is also a way to detrend yield data sets in order to minimize the effects of technological advances, such as better crop management practices. The proposed approach proved to be feasible for calibrating dty and ky\* per



**Figure 4.** Observed and estimated corn yields for municipality of Passo Fundo, in the state of Rio Grande do Sul (A), and for municipality of Catanduvas (B), in the state of Paraná, Brazil, with general parameterization (C) and after parameterization for each municipality (D).

yield classes (Figure 1, Table 3) or per municipality (Figure 4).

Several studies have attempted to simplify crop models for application at larger scales, such as Challinor et al. (2004) and Jagtap & Jones (2002). To account for the effect of stresses not included in the calculations, these studies mainly focused on the simplification of biophysical processes, while the deviation between simulated and observed yields was reduced by applying a single yield correction factor (or “management factor” or “yield gap parameter”). In Cropgro-Soybean model, Jagtap & Jones (2002) used the correction factor after crop simulation, reducing the bias from 57 to 11%. In Glam model, Bergamaschi et al. (2013) and Challinor et al. (2004) applied a correction factor during the simulation process for reducing LAI, in order to account for unknown stresses.

Brazilian yield records of corn reveal huge differences between neighboring municipalities, with corn yields higher than 8,000 kg ha<sup>-1</sup> in one location and less than 500 kg ha<sup>-1</sup> in others. This substantial spatial variability occurs frequently within the same agricultural zone, with the same climate, similar landform, similar soil properties, and similar specific range of potentialities and constraints and, consequently, with the same theoretical potential yield. This fact reveals the diversity of technological levels and crop management practices used throughout Brazil.

Due to the diversity of production systems used in the country, it was clearly necessary to introduce another reference yield to start the estimation process. The concept of technological potential yield (TY<sub>p</sub>) proposed in the present study was suitable for corn Ya ranging from 500 to 8,000 kg ha<sup>-1</sup>, and possibly for higher yield levels. In most cases, other more complex models would not have enough information to perform estimations across all Brazilian production regions with different cultivars, soil and crop management practices. For those cases, this model may be a possible option to estimate grain yield. Based on technological level, this method summarizes all yield factors except soil water in a given production system. Therefore, yields from two sites under the same or similar soil and climate conditions, but with different actual yields, are more related to the management practices used at each site than to the potential yield of the agricultural zone itself.

As simple as it is, this method can shorten the time and efforts required to create an operational system for regions where water deficit is the main yield-driving factor.

## Conclusions

1. The proposed model presented consistent relationship between corn yield and water deficit per yield class, with acceptable estimates of actual yield, showing potential to be used as yield forecaster and for risk analysis under different production systems in different Brazilian locations.

2. Model parameterization according to corn yield classes is essential for differentiating the effect of technological level, in order to calibrate sensitivity to water deficit and to define the technological potential yield.

3. Iterative calibration of the coefficient of technological potential yield (dty) and the correction for yield response factor to water deficit (ky\*) allow isolating the effect of water deficit from other factors, in each yield class.

## Acknowledgments

To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for an award for excellence in research to the second and third authors.

## References

- ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS: safra 2008/2009: décimo segundo levantamento, set. 2009.
- ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS: safra 2013/2014: terceiro levantamento, v.1, n.3, dez. 2013. 72p.
- ALLEN, R.G.; PEREIRA, L.S.; RAES, D.; SMITH, M. **Crop evapotranspiration**: guidelines for computing crop water requirements. Rome: FAO, 1998. 300p. (FAO Irrigation and Drainage Paper, 56).
- ANDRIOLI, K.G.; SENTELHAS, P.C. Brazilian maize genotypes sensitivity to water deficit estimated through a simple crop yield model. **Pesquisa Agropecuária Brasileira**, v.44, p.653–660, 2009. DOI: 10.1590/S0100-204X2009000700001.
- ARAÚJO, A.G.; ASSAD, M.L.L. Zoneamento pedoclimático por cultura a partir de levantamento de solos de baixa intensidade. **Revista Brasileira de Ciência do Solo**, v.25, p.103–111, 2001. DOI: 10.1590/S0100-06832001000100011.

- ARAUS, J.L.; SERRET, M.D.; EDMEADES, G.O. Phenotyping maize for adaptation to drought. **Frontiers in Physiology**, v.3, p.305, 2012. DOI: 10.3389/fphys.2012.00305.
- ASSAD, E.D.; MARIN, F.R.; EVANGELISTA, S.R.; PILAU, F.G.; FARIAS, J.R.B.; PINTO, H.S.; ZULLO JÚNIOR, J. Sistema de previsão da safra de soja para o Brasil. **Pesquisa Agropecuária Brasileira**, v.42, p.615-625, 2007. DOI: 10.1590/S0100-204X2007000500002.
- BATTISTI, R.; SENTELHAS, P.C. Drought tolerance of Brazilian soybean cultivars simulated by a simple agrometeorological yield model. **Experimental Agriculture**, v.51, p.285-298, 2015. DOI: 10.1017/S0014479714000283.
- BERGAMASCHI, H.; COSTA, S.M.S. da; WHEELER, T.R.; CHALLINOR, A.J. Simulating maize yield in sub-tropical conditions of southern Brazil using Glam model. **Pesquisa Agropecuária Brasileira**, v.48, p.132-140, 2013. DOI: 10.1590/S0100-204X2013000200002.
- BERGAMASCHI, H.; DALMAGO, G.A.; BERGONCI, J.I.; BIANCHI, C.A.M.; MÜLLER, A.G.; COMIRAN, F.; HECKLER, B.M.M. Distribuição hídrica no período crítico do milho e produção de grãos. **Pesquisa Agropecuária Brasileira**, v.39, p.831-839, 2004. DOI: 10.1590/S0100-204X2004000900001.
- CHALLINOR, A.; WHEELER, T.; GARFORTH, C.; CRAUFURD, P.; KASSAM, A. Assessing the vulnerability of food crop systems in Africa to climate change. **Climatic Change**, v.83, p.381-399, 2007. DOI: 10.1007/s10584-007-9249-0.
- CHALLINOR, A.J.; WHEELER, T.R.; CRAUFURD, P.Q.; SLINGO, J.M.; GRIMES, D.I.F. Design and optimisation of a large-area process-based model for annual crops. **Agricultural and Forest Meteorology**, v.124, p.99-120, 2004. DOI: 10.1016/j.agrformet.2004.01.002.
- DOORENBOS, J.; KASSAM, A.H. **Yield response to water**. Rome: FAO, 1979. (FAO Irrigation and Drainage Paper, 33).
- GONG, F.; WU, X.; ZHANG, H.; CHEN, Y.; WANG, W. Making better maize plants for sustainable grain production in a changing climate. **Frontiers in Plant Science**, v.6, article 836, p.1-6, 2015. DOI: 10.3389/fpls.2015.00835.
- HANSEN, J.W. Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v.360, p.2037-2047, 2005.
- HOOGENBOOM, G. Contribution of agrometeorology to the simulation of crop production and its applications. **Agricultural and Forest Meteorology**, v.103, p.137-157, 2000. DOI: 10.1016/S0168-1923(00)00108-8.
- HSIAO, T.C. Effects of drought and elevated CO<sub>2</sub> on plant water use efficiency and productivity. In: JACKSON, M.B.; BLACK, C.R. (Ed.). **Interacting stresses on plants in a changing climate**. Berlin: Springer-Verlag, 1993. p.435-465. (NATO ASI series. Series I. Global environmental change; v.16). DOI: 10.1007/978-3-642-78533-7\_28.
- HUND, A.; RUTA, N.; LIEDGENS, M. Rooting depth and water use efficiency of tropical maize inbred lines, differing in drought tolerance. **Plant and Soil**, v.318, p.311-325, 2009. DOI: 10.1007/s11104-008-9843-6.
- IBGE. Instituto Brasileiro de Geografia e Estatística. Banco de Dados Agregados. **Sistema IBGE de Recuperação Automática – SIDRA**. 2013. Available at: <http://www.sidra.ibge.gov.br>. Accessed on: Dec. 10 2013.
- JAGTAP, S.S.; JONES, J.W. Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. **Agriculture, Ecosystems and Environment**, v.93, p.73-85, 2002. DOI: 10.1016/S0167-8809(01)00358-9.
- KNÖRZER, H.; LAWES, R.; ROBERTSON, M.; GRAEFF-HÖNNINGER, S.; CLAUPEIN, W. Evaluation and performance of the APSIM crop growth model for german winter wheat, maize and fieldpea varieties in monocropping and intercropping systems. **Journal of Agricultural Science and Technology B**, v.9, p.698-717, 2011. DOI: 10.17265/2161-6264/2011.09B.011.
- LEGATES, D.R.; MCCABE JR, G.J. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. **Water Resources Research**, v.35, p.233-241, 1999. DOI: 10.1029/1998WR900018.
- LIANG, H.; CHU, Z.; LYU, Y.; YANG, W. Analysis of multi-farmers’ technology adoption behavior. **International Journal of Multimedia and Ubiquitous Engineering**, v.11, p.267-274, 2016. DOI: 10.14257/ijmue.2016.11.5.24.
- MONTEIRO, J.E.B. de A.; AZEVEDO, L. da C.; ASSAD, E.D.; SENTELHAS, P.C. Rice yield estimation based on weather conditions and on technological level of production systems in Brazil. **Pesquisa Agropecuária Brasileira**, v.48, p.123-131, 2013. DOI: 10.1590/S0100-204X2013000200001.
- MORISON, J.I.L.; BAKER, N.R.; MULLINEAUX, P.M.; DAVIES, W.J. Improving water use in crop production. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v.363, p.639-658, 2008. DOI: 10.1098/rstb.2007.2175.
- PICCINNI, G.; KO, J.; MAREK, T.; HOWELL, T. Determination of growth-stage-specific crop coefficients (K<sub>c</sub>) of maize and sorghum. **Agricultural Water Management**, v.96, p.1698-1704, 2009. DOI: 10.1016/j.agwat.2009.06.024.
- PRASAD, A.K.; CHAI, L.; SINGH, R.P.; KAFATOS, M. Crop yield estimation model for Iowa using remote sensing and surface parameters. **International Journal of Applied Earth Observation and Geoinformation**, v.8, p.26-33, 2006. DOI: 10.1016/j.jag.2005.06.002.
- REIDSMA, P.; EWERT, F.; BOOGAARD, H.; DIEPEN, K.van. Regional crop modelling in Europe: the impact of climatic conditions and farm characteristics on maize yields. **Agricultural Systems**, v.100, p.51-60, 2009. DOI: 10.1016/j.agsy.2008.12.009.
- THORNTHWAITE, C.W.; MATHER, J.R. **The water balance**. Centerton: Drexel Institute of Technology, 1955. 104p. (Publications in Climatology, v.8, n.1).
- TOJO SOLER, C.M.; SENTELHAS, P.C.; HOOGENBOOM, G. Application of the CSM-CERES-Maize model for planting date

evaluation and yield forecasting for maize grown off-season in a subtropical environment. **European Journal of Agronomy**, v.27, p.165-177, 2007. DOI: 10.1016/j.eja.2007.03.002.

YAN, W.; ZHONG, Y.; SHANGGUAN, Z. Evaluation of physiological traits of summer maize under drought stress. **Acta Agriculturae Scandinavica, Section B — Soil and Plant Science**, v.66, p.133-140, 2016. DOI: 10.1080/09064710.2015.1083610.

ZHAO, G.; BRYAN, B.A.; SONG, X. Sensitivity and uncertainty analysis of the APSIM-wheat model: interactions between cultivar, environmental, and management parameters. **Ecological Modeling**, v.279, p.1-11, 2014. DOI: 10.1016/j.ecolmodel.2014.02.003.

ZINSELMEIER, C.; WESTGATE, W.E.; JONES, R.J. Kernel set at low water potential does not vary with source/sink ratio in maize. **Crop Science**, v.35, p.158-163, 1995. DOI: 10.2135/cropsci1995.0011183X003500010029x.

---

Received on March 11, 2016 and accepted on November 11, 2016