Calibration and validation of the AquaCrop model to estimate soybean production in the Campos Gerais, Parana State, Brazil

Stefanie Lais Kreutz Rosa[1*], Jorge Luiz Moretti de Souza[1], Cibelle Tamiris de Oliveira[1], Rodrigo Yoiti Tsukahara[2]

(1. *Federal University of Paraná, Soil and Agricultural Engeneering Department, Curitiba, Paraná, Brazil;*
2. ABC Foundation Agricultural Research and Development, Castro, Paraná, Brazil.)

Abstract: The Campos Gerais region, located in the Subtropical Zone in Southern Brazil, stands out for presenting agricultural yields above the national average, especially for soybean. This study aimed to calibrate and validate the AquaCrop model for soybean crop in the edaphoclimatic conditions of Campos Gerais region. The data used were from the experimental stations of ABC Foundation in Arapoti, Castro and Ponta Grossa, Parana State, and Itaberá, São Paulo State. The input data (climate, crop, soil and soil management) from the 2006/07 to 2015/16 harvests were collected at the respective experimental stations and entered in the AquaCrop model for yield simulations. The data for model calibration were different from those used in the validation. The observed and simulated productivities were evaluated by simple linear regression analysis, mean absolute and relative errors, Pearson correlation coefficient (r), agreement (d) and performance (c) indexes. The model calibration was satisfactory in the studied localities, with agreement indices ranging from $0.87 < d < 0.99$. In the validation, the model performance index ranged from “Terrible” to “Excellent”, with agreement ranging from $0.59 < d < 1.00$. The results showed a good relationship between the observed and simulated yields, indicating that the AquaCrop model is an option to plan and investigate alternatives that improve soybean crop productivity in the Campos Gerais region.

Keywords: crop models, agricultural production, modeling, productivity simulation.

Citation: Rosa, S. L. K., J. L. M. de Souza, C. T. de Oliveira and R. Y. Tsukahara. 2023. Calibration and validation of the AquaCrop model to estimate soybean production in the Campos Gerais, Parana State, Brazil. Agricultural Engineering International: CIGR Journal, 25(4): 1-13.

1 Introduction

The Campos Gerais region, located in Parana and São Paulo States, stands out for presenting agricultural yields above the national average, especially for soybean. According to the IBGE (2020), at 2017 harvest, the cities that being part of the region had a soybean average yield of 4060 kg ha^{-1}, while national agricultural production in the same period remained at 3772 kg ha^{-1}.

Technology innovation in the Campos Gerais region enables high agricultural productivity. In this perspective, the use of simulation models are interesting tools to be tested in the region, as they can contribute to further increase yields (Souza et al., 2020).

Crop simulation models are commonly used to simulate growth, development, final grain yield and other characteristics of field crops (Kostková et al., 2021), being a tool to guide decision-making in the field (Souza et al., 2020). Depending on the structure,

---

Received date: 2022-02-14 **Accepted date:** 2022-05-26

***Corresponding author: Stefanie Lais Kreutz Rosa,** Federal University of Paraná, Soil and Agricultural Engeneering Department, Curitiba, Paraná, Brazil. E-Mail: skreutzrosa@gmail.com.

this system allow the estimation of plant response to interactions with edaphoclimatic factors, aiding on crop management, harvest monitoring and yield forecasting (Martín et al., 2014; Morell et al., 2014; Franke et al., 2020), as well as assessing the impact of climate change on agricultural production (Zhao et al., 2019; Huynh et al., 2020).

It is appropriate to use models that are complex and precise enough to describe what happens in the field (Amiri, 2016). Among the established systems presented in the literature (APSIM, CROPSYST, DSSAT-CSM, STICS, among others), the model developed by the Food and Agriculture Organization of the United Nations (FAO), termed AquaCrop has stood out (FAO, 2022).

AquaCrop is a precise, simple, and robust model that requires a relatively small number of parameters, which are mostly intuitive (Foster et al., 2017; FAO, 2022). Their routines allow estimating biomass and crop yield using water as a determining factor for the simulation of rainfed and irrigated production systems (FAO, 2022).

In general, agricultural simulation models are developed basically for specific study conditions. The models applicability under different cultivation sites and water regimes only becomes viable after adjusting their parameters through the calibration process, since the variability of production depends on climatic conditions, soil, and genetic characteristics of the plant used (Yin et al., 2018; Lecerf et al., 2019; Olanrewaju et al., 2021).

AquaCrop consists of two-parameter groups: conservative and non-conservative. Conservative parameters are those that do not change over time, such as management practices and geographical location, which were previously calibrated with data from crops grown under favorable and non-limiting conditions for growth and development. Non-conservative parameters are affected by planting, field management, soil, and climate conditions, might require adjustments for the environmental conditions of the inserted cultivar (Raes et al., 2018a).

The calibration is an important step before applying a model (Nguyen et al., 2022) and aims to minimize the errors between observed data in the field and simulated data in the program. The process is necessary as many parameters are difficult to measure directly in the field land (He et al., 2017; Rackl and Hanley, 2017; Shen et al., 2022). After careful calibration, the model should be validated to verify its robustness in the studied environment. The data used in model validation should be distinct from those previously used in calibration to represent the complete array of environments and crops to which the model will be applied (Jones et al., 2003; Souza et al., 2020).

Given the context presented, the present study aimed to calibrate and validate the AquaCrop model for soybean crop under Campos Gerais region edaphoclimatic conditions in the Parana and São Paulo States.

2 Material and methods

2.1 Study site

The present study was carried out in the Campos Gerais region, in the Subtropical Zone of Southern Brazil. The data were collected at the ABC Foundation Experimental Stations, located in Arapoti, Castro and Ponta Grossa cities, in Parana State, and in Itaberá, São Paulo State. The relief of the region varies from flat to gently undulating. The tillage system is no-tillage with homogeneous vegetation mulching, with crop rotation in winter (wheat and black oats) and summer (soybean and maize). According to Köppen's climate classification for Brazil (Alvares et al., 2013), Cfa is classified as a humid subtropical, oceanic climate without dry season with hot summer and Cfb is humid subtropical, oceanic climate without dry season with temperate summer. Table 1 shows the soil and climate classification, geographic coordinates, and altitude of the evaluated sites.

The model used in the analysis was AquaCrop, version 5.0, developed by the Food and Agriculture Organization of the United Nations (FAO, 2016).

The input data inserted in the AquaCrop model, required in both calibration and validation process, consider aspects of climate, crop, soil, and soil management (Raes et al., 2009; Raes et al., 2018b). Data were entered regarding:

a) Climate: Maximum, minimum, and medium air temperatures of the day (°C); precipitation (mm day^{-1}); incident solar radiation (MJ m^{-2} day^{-1}); relative humidity (%); and wind speed (km h^{-1}). Available data from ABC Foundation, obtained from a 10-year historical series (September/2006 to April/2016), measured at the automatic agrometeorological stations installed in each experimental station analyzed. Reference evapotranspiration (*ETo*; mm day^{-1}) was estimated with the Penman-Monteith method (Allen et al., 1998) for the same historical series. Average atmospheric CO_2 concentration (ppm) values were provided by the AquaCrop, based on data obtained from the Mauna Loa Observatory, Hawaii (Raes et al., 2009; Raes et al., 2018b);

b) Crop: Sowing and harvest dates, duration of crop phenological cycle (day; emergence, beginning and duration of flowering, senescence and physiological maturity), and plant population (plants ha^{-1}), obtained from experiments realized at ABC Foundation stations in 2006/07 to 2015/16 harvests. The following parameters were calibrated in the model: canopy decline coefficient (*CDC*); water productivity index (*WP**); reference harvest index (HI_o); and the proportional factor of crop transpiration coefficient ($Kc_{TR,x}$). The other parameters values required in the model were obtained from Raes et al. (2018c). Salinity was not considered;

c) Soil management: The soil fertility was considered near optimal and the soil cover by mulches was fixed in 100% of unincorporated plant residues in all plots. Phytosanitary control and fertilization were carried out as required by the crop. Irrigation was not considered since there was no adoption of the practice at the sites;

d) Soil: Soil data entered in AquaCrop were soil texture, volumetric water content at permanent wilting point (θ_{PMP}; m^3 m^{-3}), field capacity (θ_{FC}; m^3 m^{-3}), saturation (θ_{SAT}; m^3 m^{-3}), and saturated hydraulic conductivity (K_{SAT}; mm day^{-1}). Three soil layers (0-0.10 m; 0.10-0.25 m and 0.25-0.40 m depth) were considered for physical-water attributes insertion.

Table 1 Edaphoclimatic characteristics of ABC Foundation experimental stations in Arapoti, Castro, Itaberá, and Ponta Grossa(1)

Site	Soil	Climate classification(2)	Latitude	Longitude	Altitude (m)
			------ (degrees) ------		
Arapoti	Oxisol	Cfa/Cfb(3)	24.18° S	49.85° W	902
Castro	Inceptisol	Cfb	24.85° S	49.93° W	1001
Itaberá	Alfisol	Cfa	24.07° S	49.15° W	735
Ponta Grossa	Oxisol	Cfb	25.01° S	50.15° W	1000

Note: (1)Adapted from Souza et al. (2017); (2)Adapted from Alvares et al. (2013); (3)Climate transition site.

2.2 Model calibration and validation

The AquaCrop calibration process for soybean crop was carried out for the four evaluated localities, totaling 19 experiments. Once calibrated, the model is no longer adjusted to suit changes in environmental conditions, so the validation process is responsible for identifying how the adjustment is reflected in the final productivity of the model.

Conservative and non-conservative crop parameters used as a starting point in the AquaCrop calibration process for soybean were entered into each experiment, as presented by the AquaCrop Reference Manual (Table 2; Raes et al., 2018c).

After entering values of the parameters in the AquaCrop, the productivities simulations were performed. The parameters were modified in the calibration until the absolute and relative errors of the observed yield concerning the simulated yield were minimal and the "*d*" indexes of each experiment were high.

The AquaCrop validation process for soybean was performed after the calibration of the model parameters. The harvests used in the validation process were different from those used in the calibration process, totaling 26 experiments. Table 3 shows the localities, cultivars used, planting and harvesting dates, plant population, and the number of experiments used for each locality in the model calibration and validation processes.

Table 2 Conservative and non-conservative crop parameters used as a starting point in the AquaCrop calibration process for soybean in the Campos Gerais region, Subtropical Zone of Southern Brazil

Symbol	Description	Type (1), (2), (3), (4)	Values/ranges
	------ Crop phenology ------		
	Threshold air temperatures		
T_{base}	Base temperature (°C)	Conservative (1)	5.0
T_{upper}	Upper temperature (°C)	Conservative (1)	30
	Development of green canopy cover		
CC_o	Soil surface covered by seedling at 90% emergence (cm^2 plant^{-1})	Conservative (2)	5.0
	Number of plants per hectare	Management (3)	250– 450
	Time from sowing to emergence (GDD, growing degree days)	Management (3)	150 – 300
CGC	Canopy growth coefficient (fraction per GDD)	Conservative (1)	0.004 – 0.005
CC_x	Maximum canopy cover (%)	Management (3)	Almost entirely covered
	Time from sowing to start senescence (GDD)	Cultivar (4)	Time to emergence + 1600 – 2400
CDC	Canopy decline coefficient (fraction per GDD)	Conservative (1)	0.015
	Time from sowing to maturity (GDD)	Cultivar (4)	Time to emergence + 2000 – 3000
	Flowering		
	Time from sowing to flowering (GDD)	Cultivar (4)	Time to emergence + 1000 – 1500
	Length of the flowering stage (GDD)	Cultivar (4)	400 – 800
	Crop determinacy linked with flowering	Conservative (1)	Yes
	Development of root zone		
Z_n	Minimum effective rooting depth (m)	Management (3)	0.30
Z_x	Maximum effective rooting depth (m)	Management (3)	Up to 2.40
	Shape factor describing root zone expansion	Conservative (1)	1.50
	------ Crop transpiration ------		
$Kc_{TR,x}$	Crop coefficient when canopy is complete	Conservative (1)	1.10
	Decline of crop coefficient (% day^{-1})	Conservative (1)	0.30
	Effect of canopy cover on reducing soil evaporation in late season stage	Conservative (1)	25
	------ Biomass production and yield formation ------		
	Crop water productivity		
$WP*$	Water productivity normalized for *ETo* and CO_2	Conservative (1)	15.0
	Water productivity normalized for *ETo* and CO_2 during yield formation (as percent *WP** before yield formation)	Conservative (1)	60
	Harvest Index		
	Reference harvest index (%)	Cultivar (4)	40
	Possible increase (%) of *HI* due to water stress before flowering	Conservative (1)	Small
	Excess of potential fruits (%)	Conservative (2)	Medium
	Coefficient describing positive impact of restricted vegetative growth during yield formation on *HI*	Conservative (1)	None
	Coefficient describing negative impact of stomatal closure during yield formation on *HI*	Conservative (1)	Strong
	Allowable maximum increase (%) of specified *HI*	Conservative (1)	10
	------ Stresses ------		
	Soil water stresses		
$p_{exp,lower}$	Soil water depletion for canopy expansion - Upper threshold	Conservative (1)	0.15
$p_{exp,upper}$	Soil water depletion for canopy expansion - Lower threshold	Conservative (1)	0.65
	Shape factor for Water stress coefficient for canopy expansion	Conservative (1)	3.0
p_{sto}	Soil water depletion for stomatal control - Upper threshold	Conservative (1)	0.50

Symbol	Description	Type $^{(1), (2), (3), (4)}$	Values/ranges
	Shape factor for Water stress coefficient for stomatal control	Conservative $^{(1)}$	3.0
p_{sen}	Soil water depletion for canopy senescence - Upper threshold	Conservative $^{(1)}$	0.70
	Shape factor for Water stress coefficient for canopy senescence	Conservative $^{(1)}$	3.0
P_{pol}	Soil water depletion for failure of pollination - Upper threshold	Conservative $^{(1)}$	0.85 (Estimate)
	Vol% at anaerobiotic point (with reference to saturation)	Cultivar$^{(4)}$ Environment $^{(3)}$	Moderately tolerant to water logging
	Air temperature stress		
	Minimum air temperature below which pollination starts to fail (cold stress)	Conservative $^{(1)}$	8.0 (Estimate)
	Maximum air temperature above which pollination starts to fail (heat stress)	Conservative $^{(1)}$	40.0 (Estimate)
	Minimum GDD required for full biomass production (°C - day)	Conservative $^{(1)}$	10.0 (Estimate)

Note: $^{(1)}$ Conservative generally applicable; $^{(2)}$ Conservative for a given specie but can or may be cultivar specific; $^{(3)}$ Dependent on environment and/or management; $^{(4)}$ Cultivar specific

Table 3 Harvests used in the AquaCrop calibration and validation process for soybean crop, considering the ABC Foundation experiments localities, cultivars, planting and harvesting dates, plant population (*p*; plants ha^{-1}), and the number of experiments (*n*)

Locality	Cultivar	Planting date	Harvesting date	*p*	*n*
	--- Calibration ---				
Arapoti	NK3363	Oct 21, 2010	Mar 21, 2011	307031	3
	M5917IPRO	Oct 15, 2015	Mar 11, 2016	301562	
	NA5909RG	Oct 15, 2015	Mar 11, 2016	296875	
Castro	NK3363	Nov 13, 2007	Apr 08, 2008	309998	9
	NK3363	Nov 25, 2010	Apr 06, 2011	303926	
	NA5909RG	Nov 21, 2011	Apr 23, 2012	309998	
	NA5909RG	Dec 03, 2011	Apr 30, 2012	297855	
	NA5909RG	Nov 26, 2012	Apr 05, 2013	271875	
	NA5909RG	Nov 21, 2012	Apr 23, 2013	309998	
	BMX Apolo RR D.Mario 5.8	Nov 06, 2014	Mar 23, 2015	291406	
	M5917IPRO	Oct 21, 2015	Mar 11, 2016	286718	
	M5917IPRO	Nov 19, 2015	Mar 28, 2016	230469	
Itaberá	NA5909RG	Nov 05, 2012	Mar 19, 2013	239063	3
	NA5909RG	Nov 19, 2013	Mar 04, 2014	194531	
	BMX Apolo RR D.Mario 5.8	Oct 27, 2014	Feb 27, 2015	311719	
Ponta Grossa	NK3363	Nov 16, 2006	Mar 21, 2007	406250	4
	NA5909RG	Nov 28, 2012	Apr 03, 2013	314062	
	NA5909RG	Nov 13, 2014	Mar 24, 2015	313281	
	M5917IPRO	Oct 16, 2015	Mar 17, 2016	286718	
	--- Validation ---				
Arapoti	NK3363	Nov 03, 2010	Mar 25, 2011	321093	3
	NA5909RG	Nov 10, 2015	Mar 24, 2016	297656	
	M5917IPRO	Nov 10, 2015	Mar 24, 2016	340625	
Castro	NK3363	Nov 25, 2010	Apr 06, 2011	291037	14
	NK3363	Nov 25, 2010	Apr 06, 2011	296817	
	NK3363	Nov 27, 2010	May 02, 2011	286427	
	NA5909RG	Dec 03, 2012	Apr 17, 2013	306561	
	NA5909RG	Dec 03, 2012	Apr 30, 2013	303125	
	NA5909RG	Nov 06, 2014	Mar 23, 2015	276562	
	NA5909RG	Nov 27, 2014	Mar 27, 2015	326562	
	BMX Apolo RR D.Mario 5.8	Nov 27, 2014	Mar 27, 2015	328906	
	NA5909RG	Oct 21, 2015	Mar 11, 2016	300000	
	M5917IPRO	Nov 09, 2015	Mar 21, 2016	208593	
	NA5909RG	Nov 09, 2015	Mar 21, 2016	277344	
	NA5909RG	Nov 19, 2015	Mar 28, 2016	284375	
	M5917IPRO	Dec 08, 2015	Apr 05, 2016	237500	
	NA5909RG	Dec 08, 2015	Apr 05, 2016	275000	

Locality	Cultivar	Planting date	Harvesting date	*p*	*n*
Itaberá	NA5909RG	Nov 18, 2013	Mar 04, 2014	291406	4
	NA5909RG	Dec 12, 2013	Mar 14, 2014	259375	
	NA5909RG	Nov 25, 2014	May 23, 2015	274219	
	BMX Apolo RR D.Mario 5.8	Oct 27, 2014	Feb 27, 2015	311719	
Ponta Grossa	NK3363	Oct 17, 2010	Mar 29, 2011	317968	5
	NA5909RG	Nov 13, 2014	Mar 24, 2015	313281	
	NA5909RG	Dec 02, 2014	Apr 02, 2015	313281	
	M5917IPRO	Nov 20, 2015	Mar 31, 2016	261718	
	NA5909RG	Nov 20, 2015	Mar 31, 2016	328906	

In the validation process, it was used the same series of ABC Foundation climate and management data, and soil data performed by Souza et al. (2017). Initially, data related to the crop were modified, such as planting date (days), plant population (plants ha^{-1}), and duration of phenological cycles, being this latter dependent on the plant and cultivation condition. Subsequently, the calibrated parameters were inserted and the simulations were performed.

In the AquaCrop calibration and validation process, the simulated yields in the model (*Ys*, kg ha^{-1}) were compared with the real yields observed in the field (*Yr*, kg ha^{-1}) in simple linear regression analysis. The mean absolute (*MAE*; Equation 1) and relative errors (*MRE*; Equation 2), root mean square error (*RMSE*; Equation 3; Jacovides and Kontoyiannis, 1995), Pearson correlation coefficient (*r*; Equation 4), and "*d*" index concordance (Equation 5; Willmott, 1982) were also used to compare simulated to real data. The model validation process had the performance calculated with the "*c*" index (Equation 6; Camargo and Sentelhas, 1997). The interpretation criteria of "*c*" performance was classified by "Excellent" ("*c*" > 0.85); "Very good" (0.75 < "*c*" ≤ 0.85); "Good" (0.65 < "*c*" ≤ 0.75); "Medium" (0.60 < "*c*" ≤ 0.65); "Tolerable" (0.50 < "*c*" ≤ 0.60); "Bad" (0.40 < "*c*" ≤ 0.50); and, "Terrible" ("*c*" ≤ 0.40). The statistical analyzes were generated in an electronic spreadsheet and the *p*-value in the *software* RStudio, with *ggplot2* package (Wickham et al., 2022).

$$MAE = \frac{\sum_{i=1}^{n} |Yr_i - Ys_i|}{n} \tag{1}$$

$$MRE = \frac{\sum_{i=1}^{n} |Yr_i - Ys_i|}{\sum_{i=1}^{n} Ys_i} \times 100 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (Yr_i - Ys_i)^2} \tag{3}$$

$$r = \frac{\sum_{i=1}^{n} [(Yr_i - \bar{Y}r) \cdot (Ys_i - \bar{Y}s)]}{\sqrt{\sum_{i=1}^{n} (Yr_i - \bar{Y}r)^2 \cdot \sum_{i=1}^{n} (Ys_i - \bar{Y}s)^2}} \tag{4}$$

$$d = 1 - \frac{\sum_{i=1}^{n} (Ys_i - Yr_i)^2}{\sqrt{\sum_{i=1}^{n} (|Ys_i - \bar{Y}r| \cdot |Yr_i - \bar{Y}r|)^2}} \tag{5}$$

$$c = \mathrm{d} \cdot \mathrm{r} \tag{6}$$

Where: *MAE* is the mean absolute error (kg ha^{-1}); *MRE* is the mean relative error (%); *RMSE* is the root mean square error (kg ha^{-1}); *r* is the Pearson correlation coefficient (unitless); *d* is the "*d*" index (unitless); Yr_i is the real yield observed in the field at each *i*-experiment (kg ha^{-1}); $\bar{Y}r$ is the real average yield from all cultivars observed in the field (kg ha^{-1}); Ys_i is the estimated yield observed in the model at each *i*-experiment (kg ha^{-1}); $\bar{Y}s$ is the observed average yields from all cultivars estimated in the model (kg ha^{-1}); *n* is the number of harvests in the localities (unitless); *c* is the "*c*" index (unitless).

3 Results and discussion

3.1 AquaCrop model calibration

The stability of the model concerning conservative and non-conservative parameters is assessed in the sensitivity analysis. Considering the values recommended in the AquaCrop Reference Manual (Raes et al., 2018c), sensitivity analysis was performed to identify the parameters that are most sensitive to the potential crop productivity. In the sensitivity analysis of all model parameters (Raes et al., 2018b; Raes et al., 2018c), it was observed that AquaCrop did not show significant sensitivity to the cultivation practices and water regime in the region, which are related to soil management and the total water available in the soil, respectively. Thus, it was identified as more sensitive, and consequently calibrated (Table 4): the *CDC*; $Kc_{TR,x}$; *WP** and HI_o.

Table 4 Final calibrated parameters considering 19 experiments with soybean crop, in the Campos Gerais region, Parana and São Paulo States

Symbol	Description	Values
WP^*	Water productivity normalized for ETo and CO_2 (g m^{-2})	16 to 20
$Kc_{TR,x}$	Crop coefficient when canopy is complete but prior to senescence	1.10 to 1.15
CDC	Canopy decline coefficient (% GDD^{-1})	4.8 to 14.5
HIo	Reference harvest index (%)	26 to 47

During the calibration of the parameters, it was observed limiting aspects in the process. It was not possible to calibrate the parameters for each cultivar due to the low number of experiments with the same cultivar in each locality. The use of more than one cultivar in the calibration process caused difficulties, such as the different degree-days required by the cultivars at each phenological stage, different plant populations throughout the cycles, and climatic conditions that resulted in low productivity observed in the field.

The sensitivity of CDC, WP^*, $Kc_{TR,x}$, and HI_o parameters is directly related to its participation in two main equations that compose AquaCrop. Rosa et al. (2020) also observed sensitivity of these parameters for the wheat crop in the same region. Equation 7 determines the above-ground dry matter, which includes WP^* and $Kc_{TR,x}$, the latter being part of the crop transpiration equation (Tr). CDC is very sensitive in the program as it describes the declining phase due to leaf senescence. HI_o is part of Equation 8 and also showed high sensitivity as it determines the final yield.

$$B = \mathrm{WP} \cdot \sum_{i=1}^{n} Tr_i \qquad (7)$$

$$Y = \mathrm{B} \cdot HI \qquad (8)$$

Where: B is the dry above-ground biomass (kg ha^{-1}); WP is the water productivity parameter (kg m^{-2}); Tr_i is the crop transpiration at each i-period range (mm); n is the period considered (unit); Y is the crop productivity (kg ha^{-1}); HI is the crop harvest index (dimensionless) (Raes et al., 2018a).

The WP* water productivity parameter varied according to the range recommended for C_3 crops cycle, between 15 and 20 g m^{-2} (Raes et al., 2018a). Silva et al. (2018) analyzing two soybean cultivars, MSOY 9144 and TMG 1244, in Matopiba region, Tocantins State in Brazil, observed good AquaCrop calibration results for soybean crop, considering WP* = 15.5 g m^{-2}. Adeboye et al. (2017) calibrating and validating the AquaCrop model for soybean under different water regimes in Nigeria, found high variability of the results analyzed. Considering WP* = 17.6 g m^{-2} for the five treatments evaluated, the authors observed errors between 18.2% and 24.5% in calibration and between 0% and 135% in the validation. Lievens (2014) evaluating soybean cultivation scenarios in northeastern Thailand obtained low crop productivity under rainfed conditions (0.44 ± 0.16 ton ha^{-1}) when compared to the irrigated scenario (2.59 ± 0.03 ton ha^{-1}) while considering WP* = 15 g m^{-2} in calibration.

The maximum $Kc_{TR,x}$ varied between 1.10 and 1.15. The values obtained are close to those adopted by Adeboye et al. (2017) and suggested by Raes et al. (2018c). Adeboye et al. (2017) comment that the $Kc_{TR,x}$ = 1.10 showed a tendency to underestimate the canopy cover, with high estimation error (NRMSE > 43%). Lievens (2014) obtained $Kc_{TR,x}$ = 1.05 in calibration, and with this value, the result was not compatible for different simulation scenarios (rainfed and irrigated) since in the dry condition the author observed much lower yield than in the irrigated condition. Paredes et al. (2015) evaluating the Zhonghuang N°.13 soybean variety, cultivated in a conventional planting system, for 4 years in Daxin, China, obtained good simulation results considering $Kc_{TR,x}$ = 1.12.

The CDC had considerable variation in the calibration analysis of the present study (Table 4). The CDC controls the time required for the canopy to mature until the end of the crop cycle (Abedinpour et al., 2012), depending fundamentally on the duration of each phenological stage. The longer the interval between onset of senescence and physiological maturity, the longer the AquaCrop accounts for the influence of CDC on final crop yield. Thus, as the CDC was calibrated for different genotypes, values varied. Lievens (2014) observed moderate CDC sensitivity to soybean cultivation, with yield estimates influenced by up to 10%. Adeboye et al. (2017) performed several iterations with the trial and error method in adjusting the CDC value to obtain a good simulation. The authors adjusted values ranging from 12.0 to 29.1 (% GDD^{-1}).

The HI$_o$ may vary between genotypes, also depending on the edaphoclimatic condition in which the plant is inserted. Therefore, there was a large variation of values (26% to 47%; Table 4). Battisti et al. (2017) obtained HI$_o$ = 45% for soybean in similar climatic conditions as the present study. Paredes et al. (2015) observed errors of approximately 302 kg ha^{-1} in final yield considering HI$_o$ = 38%. The authors comment that the good results observed were due to the good WP* calibration, fixed at WP* = 17, with the trial and error procedure. Lievens (2014) obtained HI$_o$ = 38% in calibration and, according to the author, as the HI$_o$ is cultivar specific (Raes et al., 2018c), the validation of its value is necessary for the crop development condition. The author points out that comparing the parameter value with other studies involving AquaCrop would only be useful if the same cultivar were used.

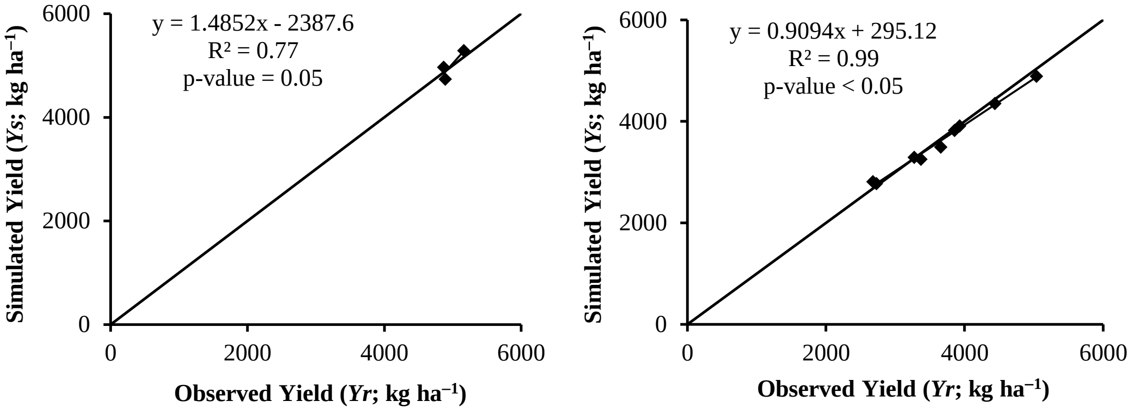
In general, the errors obtained in the calibration analysis for soybean were small, with the highest value observed in Ponta Grossa (Table 5). The mean absolute (MAE) and relative errors (MRE) observed for Arapoti, Castro, Itaberá and Ponta Grossa were very similar. On average, considering all harvests, MAE = 97.21 kg ha^{-1} and MRE = 2.52% were obtained. Silva et al. (2018) obtained MAE ranging from 100 to 330 kg ha^{-1}. Paredes et al. (2015) found higher MRE, which ranged from 7.2% to 10.2% in the calibration process, over the four harvests analyzed.

Table 5 *MAE* and *MRE*, *RMSE*, Pearson correlation coefficient (*r*), "*d*" index and yield averages ($\bar{Y}r$ and $\bar{Y}s$) obtained in the calibration process between observed (*Yr*) and simulated (*Ys*) yields in the AquaCrop model

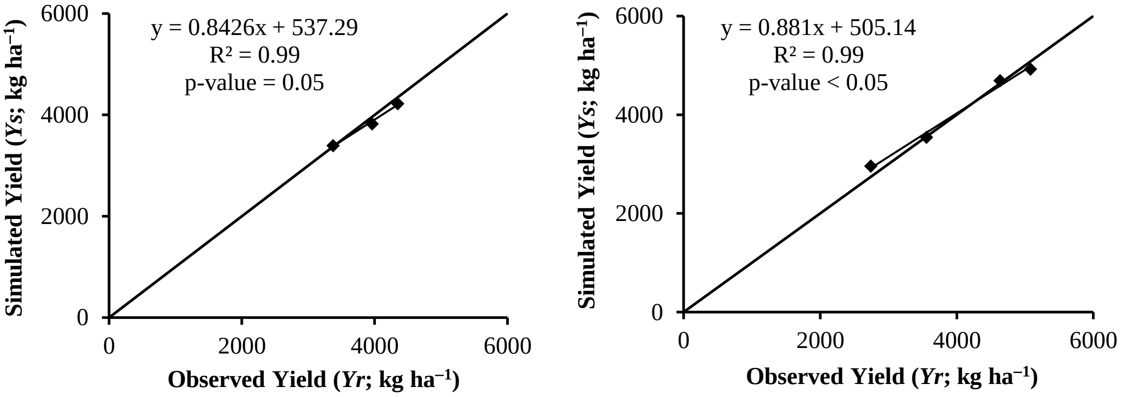
Localitie	*MAE*	*MRE*	*RMSE*	*r*	"*d*"	$\bar{Y}r$	$\bar{Y}s$	*n**
	kg ha^{-1}	%	kg ha^{-1}	--- unitless ---		---- kg ha^{-1} ----		
Arapoti	126.33	2.54	128.04	0.88	0.87	4974	5000	3
Castro	82.44	2.34	98.53	1.00	0.99	3662	3626	9
Itaberá	92.33	2.35	103.80	0.99	0.98	3890	3815	3
Ponta Grossa	112.25	3.06	137.96	1.00	0.99	4003	4031	4
All localities	97.21	2.52	113.56	0.99	0.99	3977	3958	19

Note: *n** – Total number of harvests (*Yr* and *Ys*) evaluated in the calibration process; $\bar{Y}r$ – Average yield observed in the field; $\bar{Y}s$ – Average yield simulated in the AquaCrop model; MAE- Mean absolute; MRE -relative errors-; *RMSE* -root mean square error;

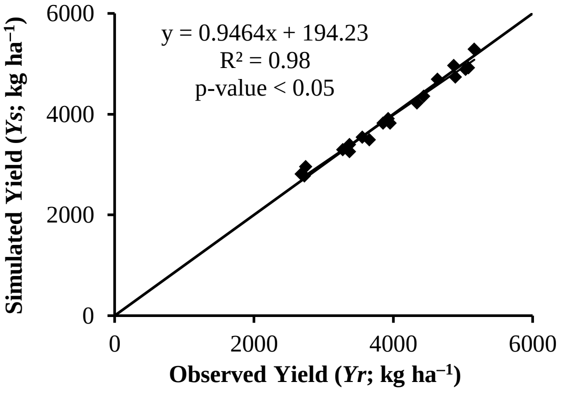
Except for Arapoti, the determination coefficients and "*d*" index obtained in the localities had excellent results (r = 0.99; d > 0.98), indicating a perfect association between the estimated and observed yield values. Arapoti and Itaberá had the inconvenience of having only three harvests for linear regression analysis, which made it difficult to describe the results obtained (Table 5 and Figure 1). However, even with the limitations, the r = 0.88 and d = 0.87 values obtained in Arapoti can be considered good. Based on the average of the 19 observed harvests, evaluated in the calibration process (3977 kg ha^{-1}), the *RMSE* ≤ 138.00 kg ha^{-1} values obtained between the yields observed in the field and estimated in the program for all localities showed excellent results in association with the calibrated parameters analyzed in AquaCrop, representing almost 3.47% of the observed productivity average (Figure 1e).



(a) Arapoti (b) Castro



(c) Itaberá (d) Ponta Grossa



(e) All locations together

Figure 1 Linear regression analysis and respective determination coefficients (R^2) obtained in the calibration, between soybean yield observed in the field and simulated in the AquaCrop model

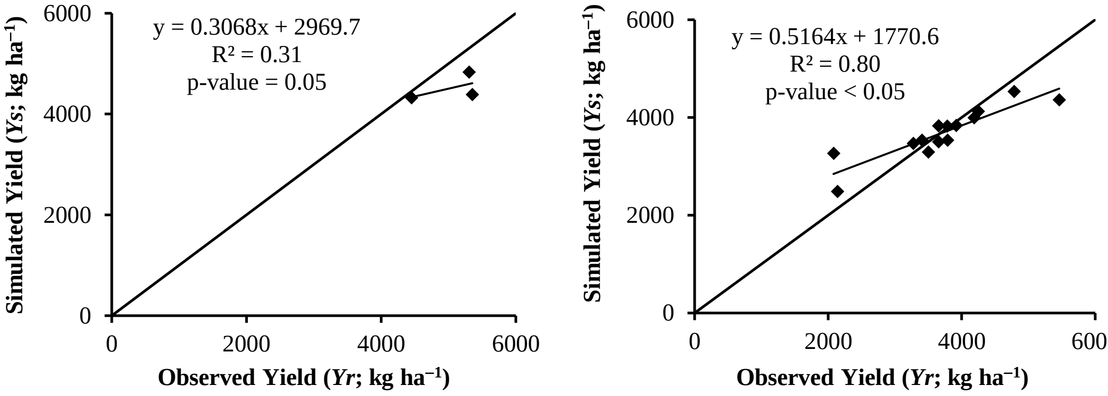
3.2 AquaCrop model validations

Even obtaining great adjustments in the calibration, there was an increase in errors in the validation analysis. The largest errors occur in Arapoti, with a relatively average error higher than 10%. AquaCrop underestimated yields with higher intensity in this locality, and all simulated crops presented yields lower than those observed in the field (Table 6 and Figure 2). However, on average, considering all crops and localities analyzed, *MAE* = 277.69 kg ha^{-1} and *MRE* = 7.12% were found between yields observed in the field and simulated in the program, which corresponded to an increase of only 4.6% compared to *MRE* = 2.52% verified in the calibration. The main limitation in Arapoti city was due to the low number of crops available in the calibration process, which made it difficult to obtain consistent parameters. Battisti and Sentelhas (2014) found a similar mean absolute error (*MAE* = 284 kg ha^{-1}) for soybean in several locations in southern Brazil.

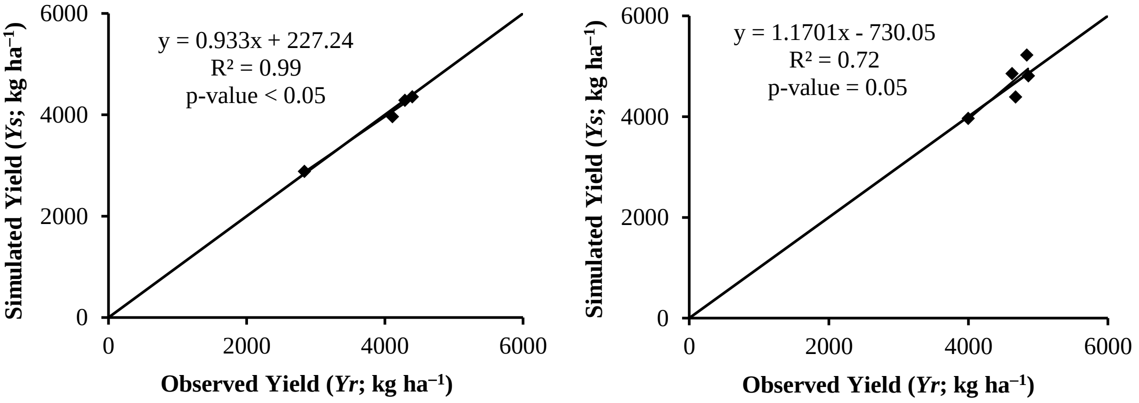
Table 6 *MAE* and *MRE*, *RMSE*, Pearson correlation coefficient (*r*), “*d*” and “*c*” indexes and yield average ($\bar{Y}r$ and $\bar{Y}s$) obtained in the validation process between observed (*Yr*) and simulated (*Ys*)

Localitie	*MAE*	*MRE*	*RMSE*	*r*	“*d*”	“*c*”	Performance	$\bar{Y}r$	$\bar{Y}s$	*n**
	kg ha^{-1}	%	kg ha^{-1}	----- unitless -----				-- kg ha^{-1} --		
Arapoti	522.67	11.59	625.26	0.56	0.59	0.33	“Terrible”	5038	4516	3
Castro	316.36	8.82	469.05	0.90	0.87	0.78	“Very good”	3706	3684	14
Itaberá	59.75	1.60	78.02	1.00	1.00	0.99	“Excellent”	3906	3871	4
Ponta Grossa	196.80	4.09	240.86	0.85	0.89	0.76	“Very good”	4598	4650	5
All localities	277.69	7.12	419.13	0.89	0.92	0.81	“Very good”	4062	3995	26

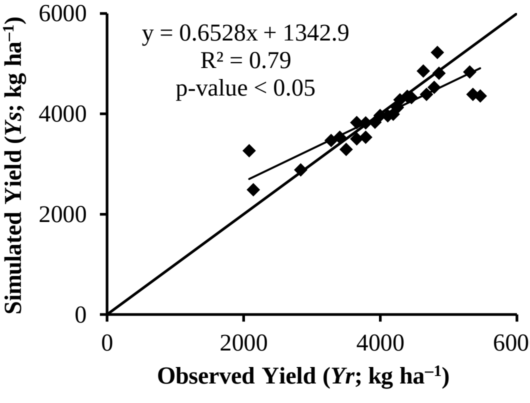
Note: *n** – Total number of harvests (*Yr* and *Ys*) evaluated in the validation process; $\bar{Y}r$ – Average yield observed in the field; $\bar{Y}s$ – Average yield simulated in the AquaCrop model.



(a) Arapoti (b) Castro



(c) Itaberá (d) Ponta Grossa



(e) All locations together

Figure 2 Linear regression analysis and respective determination coefficients (R^2) obtained in the validation, between soybean production observed in the field and simulated in the AquaCrop model

In Castro, two harvests presented extreme productivity. In the 2014/2015 harvest, the “BMX Apolo” cultivar presented an observed yield of 4361 kg ha^{-1} and a simulated yield of 5462 kg ha^{-1}, resulting in *MRE* = 25.25%. In the 2015/2016 harvest, a water deficit was observed, resulting in 2080 kg ha^{-}

^{1} of observed yield and 3267 kg ha^{-1} of simulated yield, with *MRE* = 36.33%. Both harvests contributed considerably to the increase in errors observed in the validation for this locality (Figure 2b). However, even with both occurrences, it was observed "Very Good" performance in the analysis. The correlation coefficient was very similar to that obtained by Silva et al. (2018), being $r = 0.95$ and $r = 0.93$ for soybean varieties "MSOY 9144" and "TMG 1244", respectively. The authors pointed out that the good adjustment obtained was due to the irrigation treatments adopted, and the high precipitation observed in the treatment without irrigation. As irrigation practice was not adopted in the present study, it can be ensured that the calibration process was adequate, as it was adjusted to the natural climatic conditions of the Campos Gerais region.

Akumaga et al. (2017) and Battisti et al. (2017) state that models can overestimate simulated yields, since they were developed to estimate the productivity that can be theoretically achieved, not accounting factors such as pests and diseases, and the photoperiod effect on soybean development, which can lead to increase or decrease on yield estimates for early or late planting dates. In the present study, AquaCrop underestimated productivity in most locations. Only in Ponta Grossa there was an overestimation, which can be considered negligible due to the productivities amplitudes (Table 6). Adeboye et al. (2017) also observed good adjustments in soybean validation, with estimated yield ranging from 0 to 3% (percentage deviation) of observed grain yield.

The "*d*" indexes obtained in the analysis showed results similar to those found for correlation coefficients (*r*), being higher than those obtained by Battisti et al. (2017), which obtained "*d*" = 0.68 for the soybean crop in southern Brazil.

In general, the validation analysis performed for Castro, Itaberá, and Ponta Grossa cities showed great performances, ranging from "Very good" to "Excellent" (Table 6). The result obtained in Arapoti points out the importance of data quantity and consistency to perform the calibration and validation analysis. The result of all crops analyzed together showed $R^2 = 0.79$, "*d*" = 0.92 and "*c*" = 0.81 in the validation process, indicating "Very good" performance.

The results indicated that the calibration of the AquaCrop model was adequate to the edaphoclimatic conditions of Campos Gerais region, Subtropical Zone in Southern Brazil, since only the parameters WP^*, $Kc_{TR,x}$, CDC, and HI_o had their values adjusted, according to the importance identified in the model sensitivity analysis. Thus, it is possible in future studies to use this computational tool to assist in crop forecasting and to identify better soybean planting alternatives, aiming to further increase yields in the region.

To make the AquaCrop model suitable to be used with reasonable accuracy to estimate soybean productivity, in regions near to Campos Gerais or with edaphoclimatic conditions similar to the ones in the present study, the values attributed to the most sensitive parameters of the model (CDC, $Kc_{TR,x}$, WP^*, and HI_o) should be used as recommended by the present study.

4 Conclusions

Sensitivity and calibration analysis with AquaCrop model identified the parameters CDC (4.8% to 14.5% GDD^{-1}), $Kc_{TR,x}$ (1.10 to 1.15), WP^* for ETo and CO_2 (16 to 20 g m^{-2}) and reference HI_o (26% to 47%) as the most sensitive to estimate soybean production in the Campos Gerais region, Subtropical Zone in Southern Brazil.

AquaCrop showed performances between "Very good" and "Excellent" in the validation analysis for the soybean crop in Castro, Itaberá, and Ponta Grossa. The worst result obtained in Arapoti was due to the few harvests available for calibration, resulting in "terrible" performance in the model validation process, which highlights the importance of data quantity and consistency to perform analysis with AquaCrop.

AquaCrop is adequately able to estimate soybean production in the Campos Gerais region and can be used as a computational tool in decision making in search of better soybean planting and management alternatives in the region.

References

Abedinpour, M., A. Sarangi, T. B. S. Rajput, M. Singh, H. Pathak, and T. Ahmad. 2012. Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. *Agricultural Water Management*, 110: 55-66.

Adeboye, O. B., B. Schultz, K. O. Adekalu, and K. Prasad. 2017. Modelling of response of the growth and yield of soybean to full and deficit irrigation by using Aquacrop. *Irrigation and Drainage*, 66(2): 192-205.

Akumaga, U., A. Tarhule, and A. A. Yusuf. 2017. Validation and testing of the FAO AquaCrop model under different levels of nitrogen fertilizer on rainfed maize in Nigeria, West Africa. *Agricultural and Forest Meteorology*, 232: 225-234.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration. Guidelines for computing crop water requirements. Irrigation and Drainage Paper 56. Rome, Italy: FAO.

Alvares, C. A., J. L. Stape, P. C. Sentelhas, J. L. M. Gonçalves, and G. Sparovek. 2013. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6): 711-728.

Amiri, E. 2016. Calibration and testing of the AquaCrop model for rice under water and nitrogen management. *Communications in Soil Science and Plant Analysis*, 47(3): 387-403.

Battisti, R., and P. C. Sentelhas. 2014. New agroclimatic approach for soybean sowing dates recommendation: a case study. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18: 1149-1156.

Battisti, R., P. C. Sentelhas, and K. J. Boote. 2017. Inter-comparison of performance of soybean crop simulation models and their ensemble in Southern Brazil. *Field Crops Research*, 200: 28-37.

Camargo, A. D., and P. C. Sentelhas. 1997. Avaliação do desempenho de diferentes métodos de estimativas da evapotranspiração potencial no Estado de São Paulo, Brasil. *Revista Brasileira de Agrometeorologia*, 5(1): 89-97.

FAO (Food and Agriculture Organization of the United Nations). 2016. AquaCrop Software.

FAO (Food and Agriculture Organization of the United Nations). 2022. Land & Water. AquaCrop. Available at: www.fao.org/land-water/databases-and-software/aquacrop. Accessed 03 24 2022.

Foster, T., N. Brozivić, A. P. Butler, C. M. U. Neale, D. Raes, P. Steduto, E. Fereres, and T. C. Hsiao. 2017. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agricultural Water Management*, 181: 18-22.

Franke, J. A., C. Müller, J. Elliott, A. C. Ruane, J. Jägermeyr, J. Balkovic, P. Ciais, M. Dury, P. D. Falloon, C. Folberth, L. François, T. Hank, M. Hoffmann, R. C. Izaurralde, I. Jacquemin, C. Jones, N, Khabarov, M. Koch, M. Li, W. Liu, S. Olin, M. Phillips, T. A. M. Pugh, A. Reddy, X. Wang, K. William, F. Zabel, and E. J. Moyer. 2020. The GGCMI Phase 2 experiment: global gridded crop model simulations under uniform changes in CO2, temperature, water, and nitrogen levels (protocol version 1.0). *Geoscientific Model Development*, 13(5): 2315-2336.

He, D., E. Wang, J. Wang, and M. J. Robertson. 2017. Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234-235: 136-148.

Huynh, H. T. L., L. N. Thi, and N. D. Hoang. 2020. Assessing the impact of climate change on agriculture in Quang Nam Province, Viet Nam using modeling approach. *International Journal of Climate Change Strategies and Management*, 12(5): 757-771.

IBGE (Instituto Brasileiro de Geografia e Estatística). 2020. Produção Agrícola - Lavoura Temporária (2020).

Jacovides, C. P., and H. Kontoyiannis. 1995. Statistical procedures for the evaluation of evapotranspiration computing models. *Agricultural Water Management*, 27(3-4): 365-371.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4): 235-265.

Kostková, M., P. Hlavinka, E. Pohanková, K. C. Kersebaum, C. Nendel, A. Gobin, J. E. Olesen, R. Ferrise, C. Dibari, J. Takáč, A. Topaj, S. Medvedev, M. P. Hoffmann, T. Stella, J. Balek, M. Ruiz-Ramos, A. Rodríguez, G. Hoogenboom, V. Shelia, D. Ventrella, L. Giglio, B. Sharif, I. Oztürk, R. P. Rötter, J. Balkovič, R. Skalský, M. Moriondo, S. Thaler, Z. Žalud, and M. Trnka. 2021. Performance of 13 crop simulation models and their ensemble for simulating four field crops in Central Europe. *The Journal of Agricultural Science*, 159(1-2): 69-89.

Lecerf, R., A. Ceglar, R. López-Lozano, M. Van Der Velde, and B. Baruth. 2019. Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. *Agricultural Systems*, 168: 191-202.

Lievens, E. 2014. Parameterization and testing of the FAO Aquacrop model to simulate yield response to water in North-eastern Thailand. Ph.D. diss., Universiteit Gent. Faculty of Bioscience Engineering.

Martín, M. M. S., J. E. Olesen, and J. R. Porter. 2014. A genotype, environment and management (GxExM) analysis of adaptation in winter wheat to climate change in Denmark. *Agricultural and Forest Meteorology*, 187: 1-13.

Morell, F. J., H. S. Yang, K. G. Cassman, J. Van Wart, R. W. Elmore, M. Licht, J. A. Coulter, I. A. Ciampitti, C. M. Pittelkow, S. M. Brouder, P. Thomison, J. Lauer, C. Graham, R. Massey, and P. Grassini. 2014. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? *Field Crops Research*, 192: 1-12.

Nguyen, H. T., Y. Abu-Zidan, G. Zhang, and K. T. Q. Nguyen. 2022. Machine learning-based surrogate model for calibrating fire source properties in FDS models of façade fire tests. *Fire Safety Journal*, 130: 103591.

Olanrewaju, O. S., O. Oyatomi, O. O. Babalola, and M. Abberton. 2021. Genetic diversity and environmental influence on growth and yield parameters of Bambara groundnut. *Frontiers in Plant Science*, 12: 796352.

Paredes, P., Z. Wei, Y. Liu, D. Xu, Y. Xin, B. Zhang, and L. S. Pereira. 2015. Performance assessment of the FAO AquaCrop model for soil water, soil evaporation, biomass and yield of soybeans in North China Plain. *Agricultural Water Management*, 152: 57-71.

Rackl, M., and K. J. Hanley. 2017. A methodical calibration procedure for discrete element models. *Powder Technology*, 307: 73-83.

Raes, D., P. Steduto, T. C. Hsiao, and E. Fereres. 2009. AquaCrop – The FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, 101(3): 438-447.

Raes, D., P. Steduto, T. C. Hsiao, and E. Fereres. 2018a. Reference Manual, Chapter 1 - FAO crop-water productivity model to simulate yield response to water. AquaCrop version 6.0 – 6.1. Rome, Italy: FAO Land and Water Division.

Raes, D., P. Steduto, T. C. Hsiao, and E. Fereres. 2018b. Reference manual, Chapter 2 - Users guide. AquaCrop version 6.0 – 6.1. Rome, Italy: FAO Land and Water Division.

Raes, D., P. Steduto, T. C. Hsiao, and E. Fereres. 2018c. Reference Manual: Annexes. AquaCrop version 6.0 – 6.1. Rome, Italy: FAO Land and Water Division.

Rosa, S. L. K., J. L. M. Souza, and R. Y. Tsukahara. 2020. Performance of the AquaCrop model for the wheat crop in the subtropical zone in Southern Brazil. *Pesquisa Agropecuária Brasileira*, 54: e01238.

Shen, H., B. A. Tolson, and J. Mai. 2022. Time to update the split-sample approach in hydrological model calibration. *Water Resources Research*, 58(3): e2021WR031523.

Silva, V. P. R., R. A. Silva, G. F. Maciel, C. C. Braga, J. L. C. Silva Júnior, E. P. Souza, R. S. R. Almeida, M. T. Silva, and R. M. Holanda. 2018. Calibration and validation of the AquaCrop model for the soybean crop grown under different levels of irrigation in the Motopiba region, Brazil. *Ciência Rural*, 48(1): e20161118.

Souza, J. L. M., K. R. Piekarski, R. Y. Tsukahara, and B. C. Gurski. 2017. Atributos físico-hídricos de solos no sistema de plantio direto, região dos Campos Gerais. In: *Convibra Congresses Conference*. Available at: https://docplayer.com.br/78640095-Atributos-fisico-hidricos-de-solos-no-sistema-de-plantio-direto-regiao-dos-campos-gerais-resumo.html. Accessed 03 24 2022.

Souza, J. L. M., C. T. De Oliveira, S. L. K. Rosa, and R. Y. Tsukahara. 2020. Calibration and Validation of the AquaCrop Model to Estimate Maize Production in Campos Gerais, Paraná State, Brazil. *Revista Brasileira de Meteorologia*, 35(2): 243-253.

Wickham, H., W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, and D. Dunnington. 2022. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Version: 3.3.6. Available at: https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf.1. Accessed 03 24 2022.

Willmott, C. J. 1982. Some comments on the evaluation of model performance. *Bulletin American Meteorology Society*, 63(11): 1309-1313.

Yin, X., C. G. Van Der Linden, and P. C. Struik. 2018. Bringing genetics and biochemistry to crop modelling, and vice versa. *European Journal of Agronomy*, 100: 132-140.

Zhao, C., B. Liu, L. Xiao, G. Hoogenboom, K. J. Boote, B. T. Kassie, W. Pavan, V. Shelia, K. S. Kim, I. M. Hernandez-Ochoa, D. Wallach, C. H. Porter, C. O. Stockle, Y. Zhu, and S. Asseng. 2019. A SIMPLE crop model. *European Journal of Agronomy*, 104: 97-106.