International Journal of Biometeorology (2020) 64:319–329
https://doi.org/10.1007/s00484-019-01810-1

ORIGINAL PAPER

NASA/POWER and DailyGridded weather datasets—how good they are for estimating maize yields in Brazil?

Yury C. N. Duarte[1] · Paulo C. Sentelhas[1]

Received: 5 March 2019 / Revised: 24 July 2019 / Accepted: 27 September 2019 / Published online: 24 October 2019

Abstract

The low availability of high-quality meteorological data resulted in the development of synthetic meteorological data generated by satellite or data interpolation, which are available in grids with varying spatio-temporal resolution. Among these different data sources, NASA/POWER and DailyGridded databases have been applied for crop yield simulations. The objective of this study was to evaluate the performance of these two datasets, in different time scales (daily, 10-day, monthly, and annual), as input data for estimating potential (YP) and attainable (YA) maize yields, using the FAO Agroecological Zone crop simulation model (FAO-AEZ), properly calibrated and validated. For that, daily weather data from ten Brazilian locations were collected and compared to the data extracted from NASA/POWER and DailyGridded systems and later applied to estimate the potential and attainable maize yields. DailyGridded data showed a better performance than NASA/POWER for all weather variables and time scales, with confidence index (C) ranging from 0.52 to 0.99 for the former and from 0.09 and 0.99 for the latter. As a consequence of that, DailyGridded data was better than NASA/POWER to estimate maize yields with estimates close to those obtained with observed data, with a lower mean absolute errors (< 30 kg ha^{-1}) and a higher confidence index (C = 0.99).

Keywords Crop simulation model · Gridded weather data · Maize crop · Virtual weather station

Introduction

High-quality weather data are essential to be used as inputs in the crop simulation models. The quality and reliability of these data depend on their origin and may be a source of uncertainties when simulating crop growth, development, and yield (Aggarwal 1995; Rivington et al. 2006; Bai et al. 2010). When the objective is to simulate crop yield in large areas, in addition to the temporal variability of the weather data, it is necessary to consider their spatial variability which may bring even more uncertainties to the simulations (de Wit and van Diepen 2008).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00484-019-01810-1) contains supplementary material, which is available to authorized users.

✉ Paulo C. Sentelhas
pcsentel.esalq@usp.br

[1] Department of Biosystems Engineering, Agricultural College Luiz de Queiroz (ESALQ), University of São Paulo, Avenida Pádua Dias, 11, Piracicaba, SP 13418-900, Brazil

Currently, there are several sources of weather data that can be used in agroclimatic studies with crop simulation models. Among them, there are (1) series from surface weather stations (SWS), which provide data measured with high precision; however, the spatial coverage of SWS are normally insufficient and the series present frequent data gaps; (2) series from the interpolation of data of several SWS networks, generating historical series by grid point, with spatial resolution varying from 0.25° to 1°; (3) series generated from numerical models (Voyant et al. 2012; Jimenes et al. 2016; Lorenc et al. 2015; Bengtsson et al. 2017), which usually present low resolution (2°); and (4) series based on satellite data, such as NOAA-AVHRR, MeteoSat, and NASA Prediction of Worldwide Energy Resources (NASA/POWER-NP), which provides daily data of all variables required by crop simulation models in a 0.5 to 1° resolution (Monteiro 2015). Thus, the treatment of these data is important to make them valid for studies with crop simulation models (Bates et al. 1998; Charles et al. 2004).

In Brazil, most of the weather stations belongs to National Institute of Meteorology (INMET), which provides access to maximum, minimum, and average air temperature; solar radiation or sunshine hours; rainfall; relative humidity; and wind speed. INMET offers, through the Meteorological Database for Teaching and Research (BDMEP), access to the data of 265 conventional weather stations, with data available since 1961. In the 2000s, INMET began to operate an automated weather station network, providing hourly and daily data from more than 530 locations in Brazil. Despite this increased number of stations, this network still presents low density, with only 0.001 stations per 100 km^2, when the recommended by the World Meteorological Organization (WMO) is 6.3 stations per 100 km^2 (Plummer et al. 2003). Besides BDMEP, the National Water Agency (ANA) has a rain gauge network in the country, which provides rainfall data in higher spatial resolution (0.07 rainfall stations per 100 km^2).

The NP system provides a global gridded meteorological dataset, with a resolution of 0.5°. This dataset is updated weekly, with most weather variables available since July 1983. The system synthesizes and analyzes meteorological data on a global scale, from several sources, such as Nasa World Climate Research Program (WCRP), Global Energy and Water Cycle Experiment (GEWEX), Surface Radiation Budget (SRB), Global Modeling and Assimilation Office (GMAO), and Goddard Earth Observing System (GEOS), which provide data in different time scales (Stackhouse et al. 2015).

In addition to NP, there are other gridded point data sources such as the Brazilian interpolated data system, also called DailyGridded (DG) (Xavier et al. 2016), providing a grid point database with a resolution of 0.25° for the period from 1980 to 2013. For this interpolation, Xavier et al. (2016) considered 3625 rainfall and 735 meteorological stations. Several methods of interpolation were tested and used for each meteorological variable, allowing the choice of the best one for each case. The authors used a cross-validation system, giving even more consistency to the obtained results. Several studies have been done for testing grid point data around the world (White et al. 2008; Bai et al. 2010; van Wart et al. 2013; van Wart et al. 2015) and in Brazil (Battisti et al. 2018; Bender and Sentelhas 2018; Monteiro et al. 2018) and their impact on crop yield simulations for several crops, such as soybean (Battisti et al. 2018) and sugarcane (Monteiro et al. 2018) in Brazil.

Considering the poor coverage of weather stations in Brazil and that their historical data series often present missing data, the aim of this study was to evaluate two gridded weather data bases (NP and DG) to fill the missing data on historical weather stations series in Brazil and also to create virtual weather stations to be used as input data for a previously calibrated and validated crop model for simulating maize yield.

Material and methods

The accuracy of two gridded weather datasets, namely NASA/POWER (NP) (Stackhouse et al. 2015) and DailyGridded (DG) (Xavier et al. 2016), was assessed by comparison to observed data from ten Brazilian locations, where there are INMET weather stations (Table 1). The variables considered were incoming solar radiation (SRAD, MJ m^{-2} day^{-1}); maximum air temperature (TMAX, °C); minimum air temperature (TMIN, °C); rainfall (RAIN, mm); wind speed at 2 m height (WIND, m s^{-1}); and relative humidity (RHUM, %).

Daily weather data were collected for the period between 1990 and 2013 and also aggregated at the following time scales: 10-day, monthly, and annual. In order to compare SRAD between the databases, the sunshine hours, provided by the INMET stations, were converted to global solar radiation using the Angströn-Prescott method with the coefficients suggested by Allen et al. (1998) ($a = 0.25$ and $b = 0.50$). Using the data provided by INMET as reference, the performances of weather gridded data were evaluated using the following statistical metrics: mean error (ME); mean absolute error (MAE); root mean square error (RMSE); agreement index (d) (Willmott 1981); coefficient of determination (R^2); and confidence index (C) (Camargo and Sentelhas 1997).

In order to confirm the quality of the weather gridded data, the Agroecological Zone (AEZ-FAO) crop simulation model, previously calibrated by Duarte (2018) (Tables S1, S2 and S3), was used to simulate potential (YP) and attainable (YA) maize yields. Following the same procedure described by Monteiro (2015), rainfall data was replaced by those collected from ANA to simulate YA, as this is the variable that most affects attainable yield and with the largest spatial variability (Bender 2017). The same statistical metrics described above were also used to assess model performance.

To estimate maize potential (YP) and attainable (YA) yields, the following weather variables were required as inputs: average air temperature (Tmean), extraterrestrial solar radiation (Qo), sunshine hours (n), and photoperiod (N). N and Qo were estimated according to Allen et al. (1998), while n data were estimated using the relationships described by Angströn-Prescott (Allen et al. 1998) equation. The AEZ-FAO model was programmed in the R software in a daily time step, according to the following equations:

$$\mathrm{YP} = \sum_{i=1}^{m} \left(\mathrm{YPAp}_i{*}\mathrm{c}_{\mathrm{LAI}}{*}\mathrm{c}_{\mathrm{resp}}{*}\mathrm{c}_{\mathrm{har}}{*}\mathrm{c}_{\mathrm{wc}}\right) \tag{1}$$

$$YA_m = \prod_{i=1}^{m} \left\{1-\left[ky_i{*}\left(1-\frac{ETa_i}{ETc_i}\right)\right]\right\}{*}YA_{i-1} \tag{2}$$

wherein i ranges from 1 to m, which express the final day of the crop cycle; YPAp is the absolute dry matter potential yield for a hypothetical crop, with leaf area index (LAI) = 5, in kg of DM ha^{-1} day^{-1}; YP is the final potential maize yield, in kg ha^{-1}; c_{LAI} is the correction

Table 1 Brazilian locations from where INMET weather stations were used for the comparative analysis between observed and gridded weather data provided by the NASA/POWER and DailyGridded systems

Location	State	Latitude (°)	Longitude (°)	Elevation (m ASML)
Balsas	MA	− 08.24	− 46.34	283
Barreiras	BA	− 12.02	− 45.52	452
Catalão	GO	− 17.93	− 47.67	835
Correntina	BA	− 13.48	− 45.41	575
Ipameri	GO	− 17.51	− 48.05	800
Jatai	GO	− 17.91	− 51.71	663
Londrina	PR	− 23.51	− 51.11	610
Pelotas	RS	− 31.78	− 52.41	13
Porto Nacional	TO	− 10.54	− 48.50	212
Rio Verde	GO	− 17.74	− 51.04	748

coefficient for the leaf area index (LAI); c_{resp} is the correction coefficient for maintenance respiration; c_{har} is the harvest index; c_{wc} is the coefficient that considers the water content in the grains; ky is the water deficit sensitivity index; ETc and ETa are, respectively, the maximum and actual crop evapotranspiration, in mm day^{-1}; and YA_n is the final attainable maize yield, in kg ha^{-1}. When $i = 1$, YA_{i-1} = YP. ETc was estimated by the product between Penman-Monteith reference evapotranspiration (ETo) (Allen et al. 1998) and crop coefficient (kc) for each crop phase (Duarte 2018).

The crop cycle duration and LAI values of different maize phenological phases considered in this study were obtained from Doorenbos and Kassam (1979), Fancelli and Dourado Neto (2000), and Manfron et al. (2003). Table 2 presents the duration of the maize phenological phases, in days after sowing for vegetative phase and days after pollination for reproductive phase, as well as their respective LAI.

Table 2 Duration of maize phenological phases and their respective leaf area index (LAI) considered for simulating yields

Vegetative phase	Days after sowing	LAI
0—Germination / Emergency	8	0.00
1—4 leaves open	16	0.01
2—8 leaves open	31	1.00
3—12 leaves open	47	3.00
4—Tasseling/Silking	60	4.50
5—Flowering and pollination	78	5.50
Reproductive phase	Days after pollination	LAI
6—Kernel blister	13	5.50
7—Kernel milk	27	5.00
8—Kernel dough	40	2.50
9—Kernel dent	54	2.00
10—Physiological maturation	62	1.00
Total crop cycle and mean LAI	140	3.30

Adapted from: Doorenbos and Kassam (1979); Fancelli and Dourado Neto (2000); Manfron et al. (2003)

The absolute dry matter potential yield of a hypothetical crop, in kg DM ha^{-1} day^{-1}, was calculated under two conditions: clear sky (YPAc) and cloudy (YPAo):

$$YPAc = \left(107.2 + 8.604{*}Q_o\right){*}\,{}^{n}\!/_{N}{*}cTc \quad (3)$$

$$YPAo = \left(31.7 + 5.234{*}Q_o\right){*}\left(1-{}^{n}\!/_{N}\right){*}cTn \quad (4)$$

where n is the number of effective sunshine hours per day; N is the daily photoperiod, in h day^{-1}; Q_o is the extraterrestrial solar radiation, in MJ m^{-2} day^{-1}; cTc and cTo represent the correction factors for Tmean (°C) for C4 metabolism plants under clean and cloudy sky conditions, respectively, as presented by Pereira et al. (2002):

If Tmean ≥ 16.5 °C:

$$cTc = -4.16 + 0.4325{*}Tmean-0.00725{*}Tmean^2 \quad (5)$$

$$cTo = -1.064 + 0.173{*}Tmean-0.0029{*}Tmean2 \quad (6)$$

If Tmean < 16.5 °C:

$$cTc = -9.32 + 0.865{*}Tmean-0.0145{*}Tmean^2 \quad (7)$$

$$cTo = -4.16 + 0.4325{*}Tmean-0.00725{*}Tmean^2 \quad (8)$$

As LAI varies along the crop cycle (Table 2, Eq. 9), the correction in relation to the reference crop (c_{LAI}) is given by Eq. 10:

$$LAI_i = \left(-0.0005{*}DAS_i\right)^2 + 0.0924{*}DAS_i \quad (9)$$

$$c_{LAI} = 2{*}\left(0.0093 + 0.185{*}LAI_i-0.0175{*}LAI_i^2\right) \quad (10)$$

where LAI_i represents the LAI at the day i and DAS is the number of days after sowing, including the reproductive phase.

The correction coefficient for maintenance respiration (c_{resp}) is a function of Tmean, in which two conditions were considered:

$$If\ Tmean \geq 20°C, c_{resp} = 0.5$$

If $Tmean < 20°C, c_{resp} = 0.6$

The harvest index (c_{har}) is the proportion of the total dry matter that will be harvested, which was considered as 0.36 for maize grains. As YP is estimated on the basis of dry matter, it was also necessary to consider in the model a coefficient to incorporate the water content in the grains (c_{mo}), as follows:

$$c_{wc} = [1-0,01*WC(\%)]^{-1} \quad (11)$$

where WC (%) represents the water content in the kernels in percentage. The WC (%) was considered as 13%, according to Doorenbos and Kassam (1979).

To estimate YA, a water deficit penalization module was applied to YP (Doorenbos and Kassam, 1979) and adapted for a 10-day time scale to better represent the impact of water stress on maize yield (Eq. 2). The water deficit during the maize growth cycle for different sowing dates was determined by the crop water balance, using the Thornthwaite and Mather (1955) model. An average soil water holding capacity (SWHC) for all locations was assumed, according to BRASIL (1981), since there is a large variability of soils in each one of these sites. Based on that, a SWHC of 100 mm was considered for a root-depth of 1.0 m. Thirty-six sowing dates per year were simulated in each location for a cultivar with 140 days of cycle (Duarte 2018). The simulated yield data were compared for each year and location using the same statistical metrics previously described (i.e., ME; MAE; RMSE; d; R^2; and C).

Results and discussion

Comparison of meteorological variables

NASA/POWER × INMET (observed) data

Table 3 presents the statistical indices and errors regarding the comparison of the INMET and NP meteorological data, in different time scales, when all locations were assessed together. The results of the overall analysis by location is presented in Table S4 in the Supplementary Material. In general, SRAD presented a better performance for daily and 10-day scales, whereas for TMAX and TMIN the best performances were for monthly and annual scales. For RAIN and RHUM, the better performances were obtained at 10-day and monthly scales, whereas for WIND, the performance of NP data was poor in all time scales. Except for the annual scale, when the weather variables were accumulated (10-day and monthly), the tendency was to have lower errors and data dispersion.

The daily scale presented reasonable precision (R^2) for the variables SRAD (0.77), TMAX (0.73), and TMIN (0.71) and a good accuracy index (d) for them, respectively, 0.91, 0.89, and 0.88. For SRAD, R^2 ranged from 0.29 in Balsas, MA, to 0.88 in Pelotas, RS (Table 3). The main reason for that is the type of meteorological systems that affect the weather conditions in these two locations. Whereas in Pelotas the rainfall nebulosity is caused mainly by cold fronts, being homogeneous, in Balsas, rainfall and nebulosity is mainly caused by convective clouds, which has a high spatial variability, causing more errors in SRAD. White et al. (2011) obtained a $R^2 = 0.86$ and a RMSE between 2 and 3 MJ m^{-2} day^{-1} for daily SRAD when comparing NP and observed data, which is slightly better than what was found in the present study (Table 3), and also by Monteiro et al. (2018) and Bender and Sentelhas (2018), with a R^2 between 0.71 and 0.76 and RMSE = 3.1 MJ m^{-2} day^{-1}. For TMAX, R^2 ranged from 0.39 in Balsas, MA, to 0.80 in Pelotas, RS, and for TMIN, R^2 ranged from 0.07 in Porto Nacional, TO, to 0.88 in Pelotas, RS. In general, ME for air temperature variables ranged between −1.1 (TMAX) and 0.93 °C (TMIN), which is close to what was found by White et al. (2008) when comparing data from 2500 weather stations in the USA with NP data on a daily scale (ME = −2.4 °C for TMAX and ME = 1.1 °C for TMIN).

RAIN, RHUM, and WIND were the weather variables with the worst performance (Table 3), which is mainly due to their high spatial variability, depending also on local factors, such as latitude, altitude, topography, and vegetation. Van Wart et al. (2015), when comparing daily RAIN and RHUM of NP with data collected from 18 weather stations across the world (China, Burkina Faso, Zambia, Kenya, Ethiopia, the USA, Germany, Argentina), found a R^2 always below 0.2 for RAIN and between 0.1 to 0.6 for RHUM, which agrees with the results from Bender and Sentelhas (2018) and also with those obtained here. These authors highlighted that it is important to adjust NP data locally in order to reduce the errors when estimating yields through crop simulation models.

Figure 1 presents the relationship between daily NP and observed weather data for all assessed locations in Brazil. A large dispersion was observed for RAIN, RHUM, and WIND daily data, resulting in the worst R^2 (0.15, 0.12, and 0.52, respectively), d (0.55, 0.77, and 0.52, respectively), and C (0.18, 0.53, and 0.18, respectively). SRAD, TMAX, and TMIN, at the same time scale, showed better agreement with measured data ($R^2 > 0.65$, $d > 0.86$, and $C > 0.66$). With exception for TMAX and RHUM, all other weather variables provided by NP presented a tendency of overestimation, with ME > 0. The percentage MAE, calculated as the ratio between MAE and the average of the variables, ranged between 7.5% for TMAX and 120.6% for RAIN. These results agree with those obtained by other authors around the world (White et al. 2008; Bai et al. 2010; van Wart et al. 2013; van Wart et al. 2015) and in other studies in Brazil (Battisti et al. 2018; Bender and Sentelhas 2018; Monteiro et al. 2018). The dispersion observed in RAIN, WIND, and RHUM directly affect reference evapotranspiration and soil water balance, and, consequently, crop yield simulations. To mitigate the poor

Table 3 Comparison between observed (INMET) and NASA/POWER grid weather data, on a daily, 10-day, monthly, and annual time scales, and their respective errors and performance indices, considering different Brazilian locations

Variable	Scale	ME	MAE	RMSE	*d*	*r*	R^2	*C*
SRAD (MJ m^{-2})	Daily	0.55	2.21	3.06	0.91	0.84	0.72	0.77
	10-day	− 0.07	0.47	0.57	1.00	1.00	0.99	0.99
	Monthly	1.24	1.76	3.26	0.77	0.67	0.45	0.51
	Annual	1.86	2.17	4.03	0.67	0.51	0.26	0.34
RAIN (mm)	Daily	0.12	3.33	9.34	0.61	0.38	0.15	0.24
	10-day	1.50	4.34	6.49	0.75	0.58	0.34	0.43
	Monthly	1.63	2.86	4.18	0.82	0.72	0.52	0.59
	Annual	1.30	1.58	2.12	0.57	0.53	0.28	0.30
TMAX (°C)	Daily	− 1.10	2.28	2.83	0.89	0.82	0.68	0.73
	10-day	− 1.52	1.97	2.37	0.88	0.86	0.73	0.75
	Monthly	− 1.51	1.87	2.24	0.87	0.87	0.75	0.76
	Annual	− 1.42	1.64	1.84	0.89	0.93	0.87	0.83
TMIN (°C)	Daily	0.93	1.95	2.64	0.88	0.81	0.65	0.71
	10-day	0.86	1.49	2.01	0.94	0.84	0.71	0.79
	Monthly	0.86	1.39	1.87	0.94	0.86	0.73	0.80
	Annual	0.81	1.03	1.23	0.94	0.93	0.87	0.88
RHUM (%)	Daily	− 6.96	11.88	15.34	0.80	0.72	0.52	0.57
	10-day	− 1.26	4.55	7.54	0.99	0.94	0.89	0.94
	Monthly	0.69	2.43	4.02	1.00	0.94	0.89	0.94
	Annual	− 9.91	10.28	12.35	0.68	0.47	0.22	0.32
WIND (m s^{-1})	Daily	1.45	1.54	1.82	0.45	0.35	0.12	0.16
	10-day	0.43	0.74	0.98	0.54	0.33	0.11	0.18
	Monthly	0.43	0.72	0.95	0.50	0.30	0.09	0.15
	Annual	0.44	0.69	0.93	0.37	0.25	0.06	0.09

ME - mean error, *MAE* - mean absolute error, *RMSE* - root mean square error, *d* - agreement index, *r* - Pearson coefficient, R^2 - coefficient of determination, *C* - confidence index, *SRAD* - solar radiation, *RAIN* - precipitation, *TMAX* - maximum temperature, *TMIN* - minimum temperature, *RHUM* - relative humidity, *WIND* - wind speed at 2 m above ground

performance of NP to estimate rainfall, data from the ANA database can be used, as it has an extensive rain gauge network in all Brazilian territory (Monteiro 2015).

DailyGridded × INMET (observed) data

Table 4 presents the statistical indices and errors for the comparison between observed (INMET) and gridded (DG) weather variables in the different time scales for the Brazilian locations assessed (Table 1). The results of the overall analysis by location is presented in Table S5 in the Supplementary Material.As observed for the relationship between NP and INMET data, the performance of DG data varied between time scales (Table 4). For SRAD, RHUM, and WIND, the best performance of DG data was for daily scale, with *C* of 0.95, 0.94, and 0.61, respectively. For TMAX and TMIN, no difference was observed among the time scales, with very good performance for all of them ($C = 0.98$). For RAIN, the best performance of DG data was for monthly scale, with *C* reaching 0.90. As observed for NP rainfall data when compared to measurements, data dispersion for this variable decreased after being accumulated for different time scales, resulting in a better correlation.

The DG data showed a better performance for representing daily weather data than the NP, mainly for TMAX, TMIN, and RHUM, with a high precision (R^2 of 0.99, 0.99, and 0.92, respectively) and accuracy (*d* of 0.99, 0.99, and 0.98, respectively). Such performance was very good not only for daily time scale but also for 10-day, monthly, and annual scales (Table 4). Figure 2 shows the relationship between daily INMET and DG weather data for the studied locations, from where it is possible to observe that both over and underestimations were very small for all weather variables. Tendency of overestimations were observed for SRAD, RAIN, and RHUM and underestimations for TMAX, TMIN, and WIND. MAE percentage variations were also lower than those observed for NP data, ranging from 0.3% for TMAX to 71.0% for RAIN. The confidence index (*C*) for DG data was also higher than those obtained for NP, being classified as "good" (Camargo and Sentelhas 1997) for all assessed weather variables, with exception for WIND, which remained low but much better than those observed for NP.

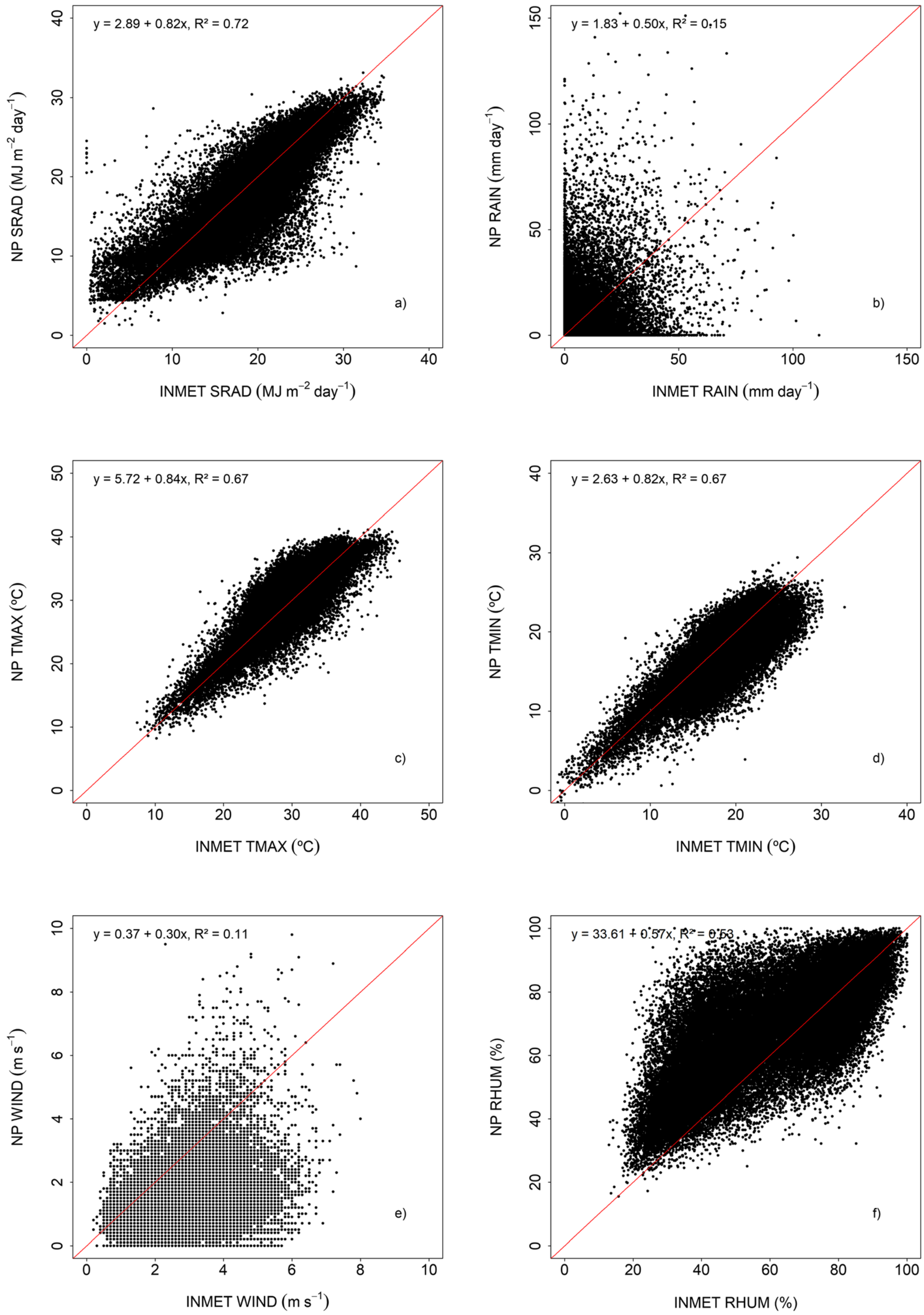

Fig. 1 Relationship between daily solar radiation (SRAD), rainfall (RAIN), maximum temperature (TMAX), minimum temperature (TMIN), wind speed (WIND), and relative humidity (RHUM), obtained from INMET and NASA/POWER grid data in ten Brazilian locations

Among the databases tested, the performance of DG system was always superior, even for weather variables with high spatial variability such as RAIN, RHUM, and even WIND. Since DG is a relatively new weather database, restricted to Brazil, it was very difficult to found other studies comparing it to measured data. The exception is the results presented by Battisti et al. (2018) and Bender and Sentelhas (2018), who highlighted the good agreement between DG and observed data, mainly for TMAX and TMIN.

Beyond the results presented by Xavier et al. (2016), the results presented here lead us to consider the possibility of using this database not only to fill missing data on weather stations series but also to create virtual weather stations for agroclimatic studies. Even with good results shown for RAIN data, DG series still present a high level of uncertainty (Fig. 2), which makes the use of observed rainfall a better option for crop yield estimation in Brazil.

Table 4 Comparison between observed (INMET) and DailyGridded weather data, on a daily, 10-day, monthly, and annual time scales, and their respective errors and performance indices, considering different Brazilian locations

Variable	Scale	ME	MAE	RMSE	*d*	*r*	R^2	*C*
SRAD (MJ m^{-2})	Daily	0.16	0.86	1.36	0.98	0.97	0.93	0.95
	10-day	0.42	0.76	1.93	0.93	0.87	0.76	0.81
	Monthly	0.48	0.77	2.08	0.89	0.82	0.67	0.73
	Annual	0.76	0.95	2.48	0.80	0.71	0.50	0.57
RAIN (mm)	Daily	0.13	2.69	6.80	0.85	0.72	0.52	0.61
	10-day	0.13	1.24	2.23	0.94	0.89	0.80	0.84
	Monthly	0.14	0.82	1.39	0.97	0.93	0.87	0.90
	Annual	0.14	0.38	0.53	0.95	0.88	0.78	0.84
TMAX (°C)	Daily	− 0.06	0.14	0.30	0.99	0.99	0.99	0.98
	10-day	− 0.01	0.07	0.13	0.99	0.99	0.99	0.98
	Monthly	− 0.01	0.06	0.12	0.99	0.99	0.99	0.98
	Annual	− 0.01	0.05	0.10	0.99	0.99	0.99	0.98
TMIN (°C)	Daily	− 0.01	0.03	0.05	0.99	0.99	0.99	0.98
	10-day	− 0.03	0.13	0.35	0.99	0.99	0.99	0.98
	Monthly	− 0.03	0.12	0.34	0.99	0.99	0.99	0.98
	Annual	− 0.03	0.09	0.22	0.99	0.99	0.99	0.98
RHUM (%)	Daily	− 0.04	2.91	4.32	0.98	0.96	0.92	0.94
	10-day	0.60	3.22	4.11	0.97	0.95	0.90	0.92
	Monthly	0.49	3.09	4.02	0.97	0.94	0.89	0.92
	Annual	− 0.12	3.00	4.28	0.90	0.81	0.66	0.73
WIND (m s^{-1})	Daily	0.10	0.40	0.61	0.82	0.75	0.56	0.61
	10-day	− 0.25	0.46	0.61	0.70	0.75	0.56	0.52
	Monthly	− 0.25	0.44	0.59	0.71	0.76	0.57	0.53
	Annual	− 0.24	0.43	0.57	0.71	0.78	0.61	0.55

ME - mean error, *MAE* - mean absolute error, *RMSE* - root mean square error, *d* - agreement index, *r* - Pearson coefficient, R^2 - coefficient of determination, *C* - confidence index, *SRAD* - solar radiation, *RAIN* - precipitation, *TMAX* - maximum temperature, *TMIN* - minimum temperature, *RHUM* - relative humidity, *WIND* - wind speed at 2 m above ground

Grid point weather data impact on maize yield estimation

Maize YP and YA simulated for 36 sowing dates, 10 Brazilian locations, and for 16 growing seasons (1997–1998 to 2012–2013) are presented in Table 1. This period was chosen to avoid the previous period when there was a large amount of missing data in the INMET weather series, especially for rainfall between 1990 and 1996. All the sowing dates were simulated considering the observed weather data as well as those data provided by the NP and DG. Maize yields presented here were the average of all 36 sowing dates simulated along the years.

INMET × NASA/POWER

Table 5 and Fig. 3 present the maize YP and YA when simulated by NP and INMET (observed) data. Table 5 illustrates how promising the NP database is to simulate maize YP and YA in different locations when compared with simulations with observed weather data. For YP, R^2 was 0.91 and *d* reached 0.99, whereas ME was 40.4 kg ha^{-1}, with about 0.3% of overestimation. The RMSE reached 300.1 kg ha^{-1} while the MAE was 68.4 kg ha^{-1}, which represents about 2.12% and 0.5% in relation to the average YP_{INMET}, respectively. The *C* index was also very high (0.95), expressing an optimal confidence of the model to estimate YP. The errors observed in Table 5, even small, could be related to the need of converting SRAD into sunshine hours, through the empirical equation of Angströn-Prescott, described by Allen et al. (1998), as also used out by Monteiro (2015) when simulating sugarcane YP with NP data in Brazil.

By using observed rainfall data from the closest rain gauge from ANA, it was possible to estimate YA satisfactorily (Fig. 3 and Table 5). The *C* index (0.98) was classified as optimal in this comparison, which shows a very high precision and accuracy in the simulations with weather data from NP and rainfall from ANA. With the ME of − 24.5 kg ha^{-1}, the estimates showed a very small tendency of underestimation, whereas MAE (36.0 kg ha^{-1}) and RMSE (210.4 kg ha^{-1}) were even smaller than those obtained for YP estimated from NP data alone.

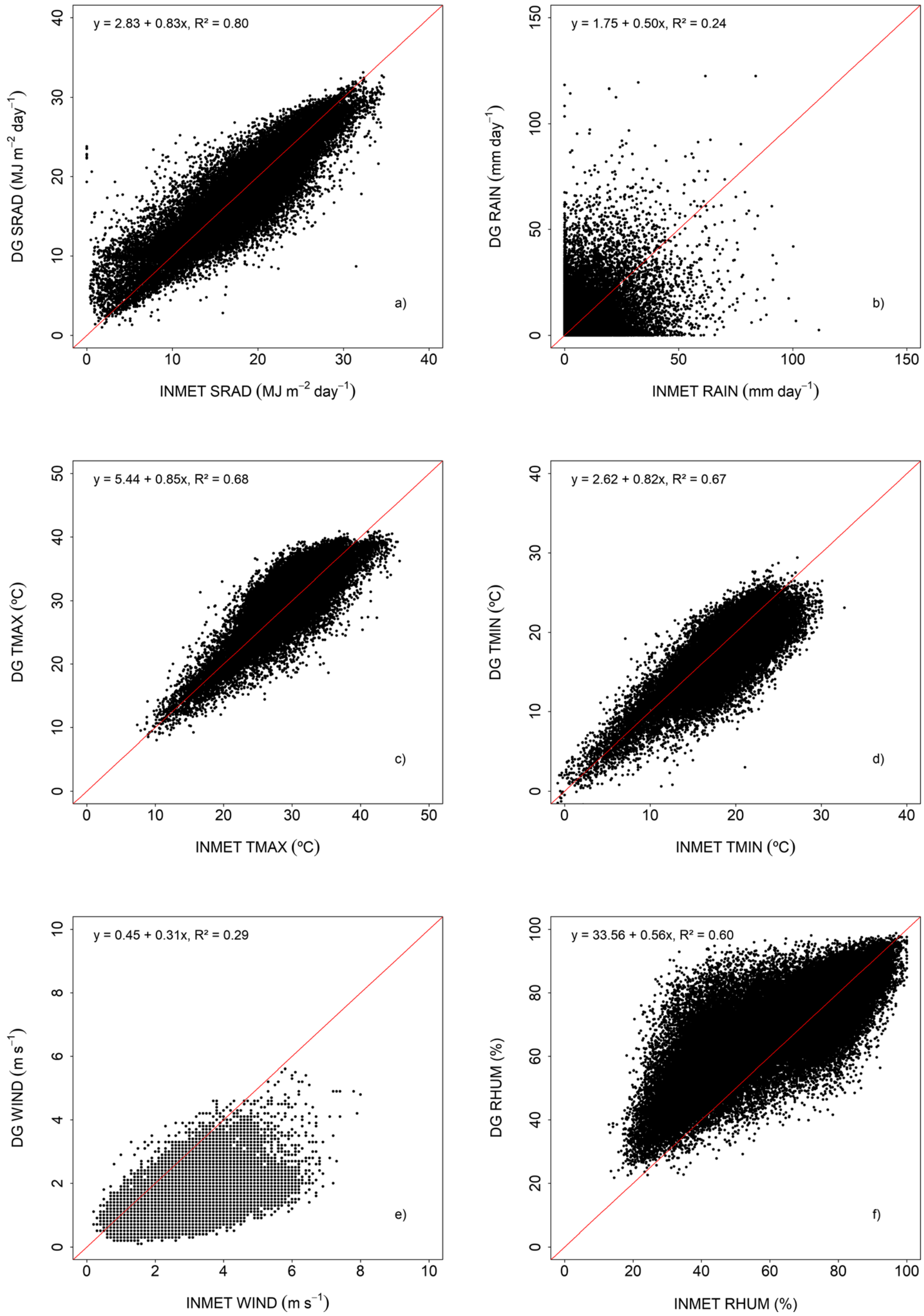

Fig. 2 Relationship between daily solar radiation (SRAD), rainfall (RAIN), maximum temperature (TMAX), minimum temperature (TMIN), wind speed (WIND), and relative humidity (RHUM) obtained from INMET weather stations and DailyGridded data, in ten Brazilian locations

The results obtained in this study (Table 5 and Fig. 3) show that the maize YP and YA values simulated with weather data from NP system and observed rainfall were quite satisfactory and with better performance than when used for simulating sugarcane YA, as showed by Monteiro (2015), who obtained R^2 = 0.82, d = 0.96, and RMSE = 16,700 kg ha^{-1}, which represents a percentage error of 18.5%.

Van Wart et al. (2015), using NP to simulate maize, rice and wheat YP in China, Burkina Faso, Zambia, Kenia, Ethiopia, the USA, Germany, and Argentina, found a ME of about 10% and a RMSE percentage ranging from 19 to 45%. Bai et al. (2010), using the Hybrid-Maize model to simulate maize YP in China with NP data, found a ME = 1400 kg ha^{-1} and a RMSE = 2600 kg ha^{-1}, about 25% in relation to observed yield data. Also, de Wit and van Diepen (2008), using data from the EUROSAT satellite as input to the WOFOST model, found MAE higher than 15% and R^2 ranging from 0.50 to 0.80 when simulating wheat dry matter accumulation in Poland, Spain, and Belgium.

The results obtained in the present study suggest that the NP data can be used as input to simulate maize potential and attainable yields in different Brazilian regions; however,

Table 5 Comparison between maize potential (YP) and attainable (YA) yields when simulated by AEZ-FAO model with observed (INMET) and NASAPOWER (NP) weather data, for different Brazilian locations

Indices/Errors	YP (Mg ha^{-1})	YA (Mg ha^{-1})
YA_{NP}-mean	14567.5 (±2916)	4450.7 (±3587)
YA_{INMET}-mean	14139.5 (±2907)	4823.2 (±3563)
ME	40.4	−24.5
MAE	68.4	36.0
RMSE	300.1	210.4
d	0.99	0.99
r	0.95	0.98
R^2	0.91	0.96
C	0.95	0.98

ME - mean error, *MAE* - mean absolute error, *RMSE* - root mean square error, *d* - agreement index, *r* - Pearson coefficient, R^2 - coefficient of determination, *C* - confidence index, $YA_{NP\text{-mean}}$ - yield attainable

Table 6 Comparison between maize potential (YP) and attainable (YA) yields simulated with observed (INMET) and DailyGridded (DG) weather data, for different locations in Brazil

Indices/Errors	YP (Mg ha^{-1})	YA (Mg ha^{-1})
YA_{DG}-mean	14246.5 (±2917)	4812.0 (±3651)
YA_{INMET}-mean	14139.4 (±2907)	4823.2 (±3563)
ME	10.1	−0.7
MAE	25.4	13.9
RMSE	123.4	104.9
d	0.99	0.99
r	0.99	0.99
R^2	0.98	0.99
C	0.99	0.99

ME mean error, *MAE* mean absolute error, *RMSE* root mean square error, *d* agreement index, *r* Pearson coefficient, R^2 coefficient of determination, *C* confidence index, $YA_{NP\text{-mean}}$ yield attainable estimated with NP weather data, $YA_{INMET\text{-mean}}$ yield attainable estimated with INMET weather data

having measured rainfall as input, which improved the performance of the estimates. When maize YA was simulated also with NASA/POWER daily rainfall data, the results were worst than when using observed rainfall data from ANA ($R^2 = 0.88$; d = 0.96; C = 0.90; ME = 388.9 Mg ha^{-1}; MAE = 993.0 Mg ha^{-1}; and RMSE = 1507.8 Mg ha^{-1}).

INMET × DailyGridded

Table 6 and Fig. 4 present the maize YP and YA when simulated by DG and INMET (observed) data. The results show a notorious superior performance for estimating YP and YA with DG (Table 6 and Fig. 4) when compared to NP (Table 5 and Fig. 3). The errors were reduced as well as data dispersion (Fig. 4), which resulted in better accuracy (d = 0.99) and precision ($R^2 \geq 0.98$), resulting in a C index classified as "very good" ($C \geq 0.98$). MAE was reduced when compared to the maize yield simulated with NP data. The performance for estimating maize yield with DG data surpassed that presented by NP, as well as those obtained using other systems with a similar approach (de Wit and van Depen 2008; Bai et al. 2010; van Wart et al. 2015). Battisti et al. (2018), using DG to estimate soybean YA in 24 Brazilian states, obtained a ME ranging from −170 to 1160 kg ha^{-1} and a RMSE ranging from 187 to 1586 kg ha^{-1}, quite similar to those presented here (Table 6). They also identified DG as a feasible source of data to be used for soybean yield simulations in Brazil. The use of observed rainfall data from ANA when running the AEZ-FAO model for maize YA estimation, performed even better than when using rainfall data from DG. For DG rainfall data, the performance of the estimates when comparing with observed data from INMET was good but with lower R^2 (0.93), d (0.98), C (0.94), and higher errors

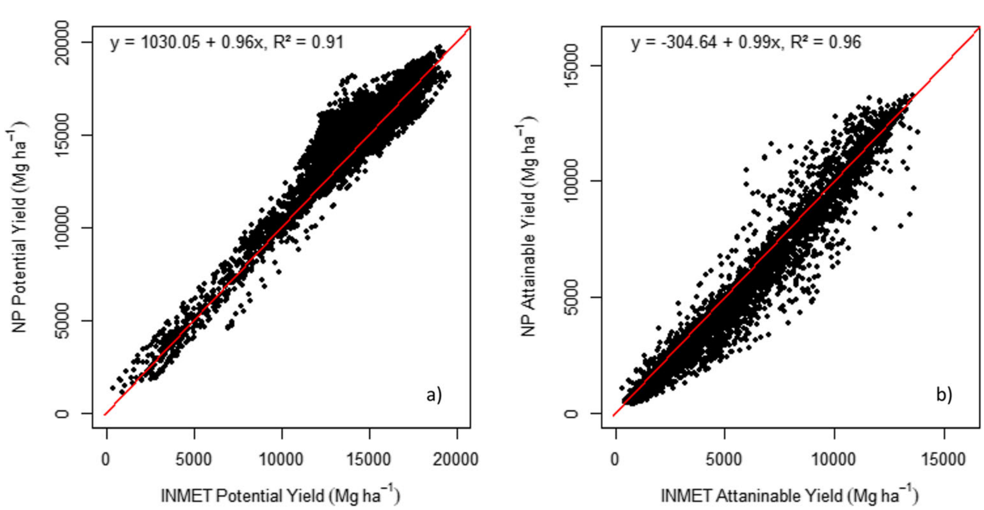

Fig. 3 Relationship between maize potential yield (YP) simulated with weather data from INMET and from NASA/POWER system (**a**), and maize attainable yield (YA) simulated with weather data from INMET and from NASA/POWER system (**b**), for ten Brazilian locations, using the AEZ-FAO model

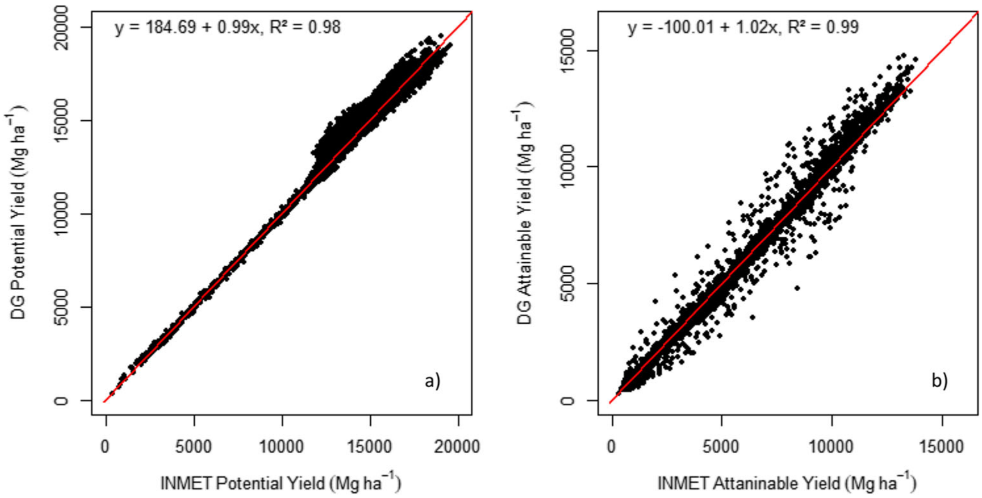

Fig. 4 Relationship between maize potential yield (YP) simulated with weather data from INMET and from DailyGridded system (**a**), and maize attainable yield (YA) simulated with weather data from INMET and from DailyGridded system (**b**), for ten Brazilian locations, using the AEZ-FAO model

(ME = 455.5 Mg ha^{-1}; MAE = 751.6 MG ha^{-1}; and RMSE = 1190.6 Mg ha^{-1}).

Conclusions

Both the NP and DG weather databases were able to provide satisfactory data of TMAX, TMIN, and SRAD, with DG database showing a superior performance than NP, mainly for RAIN and RHUM. Thus, DG was considered the best potential tool to create "virtual" weather stations where actual measurements are absent, as well as a data source to fill in missing data in observed weather series. The errors observed on gridded weather data for the meteorological variables considered in this study were not large enough to cause significant errors on potential and attainable maize yield simulations. The simulated YP and YA with both gridded databases showed excellent performance, with the DG system being superior to NP. These results suggest that both datasets can be used for simulating maize YP and YA at a national scale. The association of observed rainfall data (ANA) with other variables in the DG database proved to be the best option for estimating maize YA, which can be an advantage since there are more than 4000 rainfall stations from ANA and other agencies in Brazil.

References

Aggarwal PK (1995) Uncertainties in crop, soil and weather inputs used in growth models: implications for simulated outputs and their applications. Agric Syst 48:361–384. https://doi.org/10.1016/0308-521X(94)00018-M

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: guidelines for computing crop water requirements. Roma, FAO (Irrigation and Drainage Paper 56).

Bai J, Chen X, Dobermann A, Yang H, Cassman KG, Zhang F (2010) Evaluation of NASA satellite- and model-derived weather data for simulation of maize yield potential in China. Agron J 102:9–16. https://doi.org/10.2134/agronj2009.0085

Bates BC, Charles SP, Hughes JP (1998) Stochastic downscaling of numerical climate model simulations. Environ Model Softw 13:325–331. https://doi.org/10.1016/S1364-8152(98)00037-1

Battisti R, Bender FD, Sentelhas PC (2018) Assessment of different gridded weather data for soybean yields simulations in Brazil. Theor Appl Climatol 135:237–247. https://doi.org/10.1007/s00704-018-2383-y

Bender FD 2017, Climate change and its impact on maize yields and crop management strategies to minimize yield losses in different Brazilian regions. University of São Paulo. PhD. Thesis. (In Portuguese).

Bender FD, Sentelhas PC (2018) Solar radiation models and gridded databases to fill gaps in weather series and to project climate change in Brazil. Adv Meteorol 2018:1–15. https://doi.org/10.1155/2018/6204382

Bengtson L, Andrae U, Aspelien T, Batrak Y, Calvo J, Rooy W, Gleeson E, Hansen-Sass B, Homleid M, Hortal M, Ivarsson K, Lenderink G, Niemela S, Nielsen KP, Onvlee J, Rountu L, Samuelson P, Munoz DS, Subias A, Tijm S, Toll V, Yang X, Koltzow MO (2017) The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System. Monthly Weather Rev 145:1919–1935. https://doi.org/10.1175/MWR-D-16-0417.1

Brasil (1981) Ministério das Minas e Energia. Secretaria Geral. Projeto RADAMBRASIL. Rio de Janeiro: Levantamento de Recursos Naturais, 25, 29, 31.

Camargo AP, Sentelhas PC (1997) Avaliação do desempenho de diferentes modelos de estimativa de evapotranspiração potencial no Estado de São Paulo, Brasil. Rev Bras Agrometeorol 5:89–97

Charles SP, Bates BC, Smith IN, Hughes JP (2004) Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. Hydrol Proc 18:1373–1394. https://doi.org/10.1002/hyp.1418

de Wit AJ, van Depien CA (2008) Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. Int J Appl Earth Observ Geoinf 10:414–425. https://doi.org/10.1016/j.jag.2007.10.004

Doorenbos J, Kassam AH (1979) Yield response do water. Rome, FAO (Irrigation and Drainage Paper 33).

Duarte YCN (2018) Maize simulation models and their use to determine yield gaps and yield forecast in Brazil. University of São Paulo. MSc. Thesis. (In Portuguese).

Fancelli AL, Dourado Neto D (2000) Produção de milho. Guaíba: Editora Agropecuária.

Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Appl Eng Agric 1:96–99. https://doi.org/10.13031/2013.26773

Jimenes PA, Hacker JP, Dudhia J, Haupt SE, Ruiz-Arias JA, Gueymard CA, Thompson G, Eidhammer T, Deng A (2016) WRF-Solar: An augmented NWP model for solar power prediction. Model description and clear sky assessment. BAMS, 2016:1249–1264. https://doi.org/10.1175/BAMSD-14-00279.1

Lorenc AC, Bowler NE, Clayton AM, Pring SR (2015) Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP. Monthly Weather Rev 143:212–229. https://doi.org/10.1175/MWR-D-14-00195.1

Manfron PA, Dourado Neto D, Pereira AR, Bonnecarrère RAG, Medeiros SLM, Pilau FG (2003) Modelo do índice de área foliar da cultura do milho. Rev Bras Agrometeorol 11:333–342

Monteiro LA (2015) Sugarcane yield in Brazil: a crop modelling approach. University of São Paulo. Ph.D. Dissertation. 130p. 2015.

Monteiro LA, Sentelhas PC, Pedra GU (2018) Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation. Int J Climatol 38:1571–1581. https://doi.org/10.1002/joc.5282

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

Plummer N, Allsopp T, Lopez JA (2003) Guidelines on Climate Observation Networks and Systems. WMO/TD N. 1185

Rivington M, Matthews KB, Bellocchi G, Buchan K (2006) Evaluating uncertainty introduced to process-based simulation model estimates by alternative sources of meteorological data. Agric Syst 88:451–471. https://doi.org/10.1016/j.agsy.2005.07.004

Stackhouse PWJ, Westberg D, Hoell JM, Chandler WS, Zhang T (2015) Prediction of world-wide energy resource (POWER)—Agroclimatology methodology—(1.0° latitude by 1.0° longitude spatial resolution). Hampton, NASA Langely Research Center.

Thornthwaite CW, Mather JR (1955) The water balance. Publications in Climatology. New Jersey: Drexel Institute of Technology.

van Wart J, Grassini P, Cassman KG (2013) Impact of derived global weather data on simulated crop yields. Glob Chang Biol 19:3822–3834. https://doi.org/10.1111/gcb.12302

van Wart J, Grassini P, Yang H, Claessens L, Jarvis A, Cassman KG (2015) Creating long-term weather data from thin air for crop simulation modeling. Agr Forest Meteorol 209-210:49–58. https://doi.org/10.1016/j.agrformet.2015.02.020

Voyant C, Muselli M, Paoli C, Nivet M (2012) Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. Energy 39:341–355. https://doi.org/10.1016/j.energy.2012.01.006

White JW, Hoogenboon G, Stackhouse PWJ, Hoel JM (2008) Evaluation of NASA satellite and assimilation model-derived long-term daily temperature data over the continental US. Agr Forest Meteorol 148:1574–1584. https://doi.org/10.1016/j.agrformet.2008.05.017

White JW, Hoogenboom G, Wilkens PW, Stackhouse PWJ, Hoel JM (2011) Evaluation of satellite-based, model-derived daily solar radiation data for the continental United States. Agron J 103:1242–1251. https://doi.org/10.2134/agronj2011.0038

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

Xavier AC, King CW, Scanlon BR (2016) Daily gridded meteorological variables in Brazil (1980–2013). Int J Climatol 36:2644–2659. https://doi.org/10.1002/joc.4518