The impact of gridded weather database on soil water availability in rice crop modeling

Caio Teodoro Menezes1 · Derblai Casaroli2 · Alexandre Bryan Heinemann3 · Vinicius Cintra Moschetti4 · Rafael Battisti2

Received: 25 March 2021 / Accepted: 19 December 2021 / Published online: 6 January 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract
In recent years, there has been an increase in studies suggesting that gridded weather database (GWD) is a suitable source for simulating crop yield. Brazil has low geospatial coverage by measured weather database (MWD). Based on that, this study aimed to compare two different GWD sources, Daily Gridded (DG) and NASA/POWER (NP), on the simulated yield of upland rice (UR) against the MWD input. The GWD and MWD were obtained for seven locations across UR Brazilian region, considering a period ranging from 1984 to 2016. GWD and MWD were used to estimate rice potential (Yp) and attainable yield (Ya), in clay soil and sandy soil, using the ORYZA (v3) model. DG had the best performance for all variables. GWD-based yields had a reasonable performance. However, DG had a slightly better performance than NP in all conditions, DG-based yields showed RMSE values of 0.57, 0.71, and 0.52 for Yp and Ya in clay and sandy soil, and NP showed RMSE values of 0.86, 0.91, and 0.64. DG also showed higher R² and d values for yields assessed. Both GWD overestimated Ya; these overestimations in DG-based yield were 3.54, 9.61, and 21.35% for Yp and Ya in clay and sandy soil, respectively, in NP-based yield were 13.67, 18.45, 29.11%, showing that for both GWD-based yield increased as the soil type texture as well as water storage decreased. As a consequence, we do not recommend the use of precipitation data in daily time-step crop modeling.

Keywords ORYZA (v3) · Daily Gridded · Virtual weather station · NASA/POWER · Crop modeling

1 Introduction
Mathematical models that simulate growth and yield of agricultural crops are important tools for decision making (Caetano and Casaroli 2017; Justino et al. 2019; Battisti et al. 2020; Caetano et al. 2021). The use of estimated weather data has been widely used, improving its degree of accuracy more and more (Jha et al. 2019; Dubrovsky et al. 2020). Thus, these data can be used in different applications, such as, to obtain yield gap in agrometeorological models (Santos et al. 2021; Paixão et al. 2021).

Useful and reliable weather datasets are a time-saving tool for crop modeling purposes. The absence of measured weather data (MWD) and a proper spatial resolution are impediments to predicting both current and future effects of climate on crop yields. Therefore, gridded weather data (GWD) has become a welcomed alternative weather data source for agricultural purposes. GWD has the advantage of complete geospatial cover, thus becoming an option for regions where the MWD has a lower spatial distribution or incomplete data (Paixão et al. 2021).

There are two main sources of MWD in Brazil, the Brazilian National Institute of Meteorology (INMET), which provides MWD from automatic and conventional weather stations, and the Brazilian Water Agency (ANA), which has a rainfall station network across Brazilian territory. Both
sources are available in low spatial resolution (Xavier et al. 2015). Central Brazil, the main region for grains and pulses farming (CONAB 2020), is situated in three river basins: Tocantins, Amazon, and Paraná. MWD resolutions are lower for Amazon and Tocantins river basins when compared with other parts of Brazil (Xavier et al. 2015). Currently, this is the greatest limitation for crop modeling in upland rice region (Heinemann et al. 2008).

Crop model simulation for application and agrometeorological purposes is made easier by GWD. It has been applied worldwide to simulate crop yields (van Wart et al. 2013a; Ruane et al. 2015; Mourtzinis et al. 2017; Jha et al. 2019; Müller et al. 2019; Paixão et al. 2021). Amongst worldwide available GWD for crop modeling purposes, the NASA/POWER (NP) can be pointed out. This dataset has a 0.5° × 0.5° spatial resolution at a daily temporal resolution comprising most variables used in crop modeling. This dataset is updated weekly with a time series available since 1983.

In Brazil, there is an interpolated weather database denominated Daily Gridded (DG) (Xavier et al. 2015), which provides an agrometeorological daily and monthly grid data with two spatial resolution for the years 1980 until 2017. The highest resolution available – 0.1° × 0.1° contains data for maximum and minimum air temperature, and the lowest resolution – 0.25° × 0.25° contains data for the other variables: reference evapotranspiration, global solar radiation, rainfall, relative humidity, and wind speed. DG has been applied for wide variety of crops, such as soybean (Battisti and Sentelhas 2019; Junior and Sentelhas 2019; Teixeira et al. 2019), sugarcane (Monteiro et al. 2018; Cesconetto et al. 2019; Perin et al. 2019; Paixão et al. 2020), maize (Andrea et al. 2019), and bamboo (Battisti et al. 2019).

Nevertheless, there are still uncertainties related to the accuracy for its application in crop model simulation. GWD has been accurate in applications related to temperature, but uncertain in estimates related to precipitation. Those uncertainties may generate misleading water deficit and crop yield outcomes (Mourtzinis et al. 2017).

For this study, we selected ORYZA (v3) crop model (Li et al. 2017). The model predicts rice growth and yield as influenced by local environmental conditions, agronomic practices, and cultivar traits. Its strong ability to quantify the influence of soil water on rice growth and yield (Bouman and Laar 2006; Feng et al. 2007) allows the evaluation of rice cultivars’ response under drought stress (Li et al. 2013). ORYZA (v3) has been used to quantify rice yield gaps at global, national, and regional scales (Boling et al. 2010; Laborte et al. 2012; Espe et al. 2016), to provide a cost-effective method for drought-tolerance screening (Li et al. 2013), and for determining drought-stress profiles for upland rice in Brazil under current (Heinemann et al. 2015) and future climates (Ramirez-Villegas et al. 2018) and to assess the performance of the model across lowland tropical rice region in Brazil (Menezes 2021).

Based on the current literature, we hypothesize that GWD can be used for simulating rice yield using crop growth models in the areas of upland rice. This way, the study aims: (1) to compare measured weather data with two sources of gridded weather data (Daily Gridded and NASA/POWER) and (2) to evaluate the impact of measured and gridded weather data on the potential (YP) and attainable (Ya) in upland rice yield, considering two soil types—clay and sandy.

### 2 Material and methods

#### 2.1 Locations and weather databases

The measured weather data was obtained from seven locations (Table 1) across upland rice production region in Brazil. The data for six weather stations were obtained from the Brazilian Meteorological Service (INMET) and one from the Brazilian Agricultural Research Corporation (EMBRAPA). Missing data in the MWD were filled out using the day of year from the climate normal. The percentage of missing data for each weather station is given in Table 1. The locations are classified as tropical by Köppen–Geiger method.

Table 1 Brazilian measured weather data (MWD) locations

| ID* | Location     | State | Latitude (°) | Longitude (°) | Elevation (m ASML) | Weather station | **Missing data (%) |
|-----|--------------|-------|--------------|---------------|-------------------|-----------------|-------------------|
| s1  | Altamira     | PA    | -321         | -5221         | 74                | INMET           | 168%              |
| s2  | Alto Parnaíba| MA    | -91          | -4593         | 285               | INMET           | 300%              |
| s3  | Jataí        | GO    | -1791        | -5171         | 662               | INMET           | 087%              |
| s4  | Paracatu     | MG    | -1724        | -4688         | 712               | INMET           | 201%              |
| s5  | Porto Nacional| TO  | -1071        | -4841         | 239               | INMET           | 910%              |
| s6  | Rio Branco   | AC    | -996         | -678          | 160               | INMET           | 405%              |
| s7  | Santo Antônio de Goiás | GO | -1628 | -4917 | 823 | EMBRAPA | 905% |

*ID—weather station identification; **Percentage of missing data set considering maximum and minimum air temperature, precipitation, and solar radiation from 1984 to 2016

C. T. Menezes et al.
where Porto Nacional and Altamira are classified as Am climate type, and the other five locations as Aw (Alvares et al. 2013).

The two gridded weather databases (GWD) and its variables applied in this study are: i) Daily Gridded (Xavier et al. 2015), which has daily variables at 0.1°×0.1° for maximum and minimum air temperature, and 0.25°×0.25° spatial resolution for rainfall and solar radiation, and ii) NASA/POWER (National Aeronautics and Space Administration’s POWER, 2020) which has daily variables, including maximum and minimum air temperature, rainfall and solar radiation at scale of 0.5°×0.5° horizontal resolution.

A historical daily weather data of maximum and minimum air temperature, incoming solar radiation (MJ m\(^{-2}\) d\(^{-1}\)), and precipitation (mm d\(^{-1}\)) were obtained for the years 1984 until 2016 from both gridded datasets. The solar radiation of measured dataset was estimated by Angstrom-Prescott (Angstrom 1924; Prescott 1940) from sunshine hours because only weather station S7 contains a radiometer (Table 1).

### 2.2 Crop model simulations

The rice crop modeling ORYZA (v3) (Bouman et al. 2001) was used to assess the potential (Yp) and attainable (Ya) upland rice yield. ORYZA (v3) operates on a daily time step; hence, it requires as input daily weather data such as rainfall, global solar radiation, and maximum and minimum air temperature. ORYZA (v3) is a process-based rice simulation model developed for a wide variety of applications in rice research (Bouman and Laar 2006; Li et al. 2013), e.g., puddled or non-puddled, with free or impeded drainage at a given depth in the soil profile. The ‘PADDY’ module is described in detail by Boling et al. (2007). The same study demonstrated that it is suitable for rainfed conditions.

The cultivar crop model parameters were taken from upland rice BR5 Primavera cultivar. This is a short-cycle cultivar, with cycle 100 days from emergence to physiological maturity, and it is still used as cultivar check in the upland rice breeding program. Its parameters were obtained from a previous study (Heinemann et al. 2015), and they have been applied in many upland rice crop modeling applications in Brazil (Heinemann et al. 2015, 2019; Ramirez-Villegas et al. 2018). For all locations (Table 1), the simulation start date was set 225 days prior to the sowing date in order to establish realistic soil water profiles and the sowing date was set in November 1st. This sowing date is considered the ideal for upland rice in central Brazil (Heinemann et al. 2015). The sowing date window for all upland rice regions is available in the Brazilian agricultural climate risk zoning (ZARC, 2019). Seed density was set at 200 seeds m\(^{-2}\). A sandy and a clay soil type were used in the simulation. Classified as oxisol on the American classification system, both are predominant in the study region (Heinemann et al. 2015). Further parameters are described in Table S1.

### 2.3 Statistical analysis

The statistical indices used were mean error (ME), mean absolute error (MAE), root-mean-square error (RMSE), agreement index (d) (Willmot 1981), Pearson’s correlation coefficient (r), and coefficient of determination (R\(^2\)). This was applied to compare minimum and maximum air temperature, rainfall, solar radiation in the crop cycle as well as potential (Yp) and attainable (Ya) yield.

The sequence of graphics used in this study was empirical cumulative probability function (ECDF), scatter plot, regression lines, violin plot and line chart.

### 3 Results

#### 3.1 Descriptive statistics to measured and gridded weather data

An empirical cumulative probability function (ECDF) is plotted (Fig. 1) to assess the four variables from the datasets in the crop cycle. ECDF for daily solar radiation (Sr) and maximum air temperature (Tmax) in Daily Gridded (DG) data showed a higher similarity to the measured weather data (MWD) than NASA/POWER (NP) (Fig. 1a and c). NP only exhibits a higher similarity to MWD regarding minimum air temperature (Tmin) (Fig. 1d). Both gridded weather data (GWD) had similar distributions to each other but with a higher degree of distinction to MWD (Fig. 1b). NP underestimates Tmax in the range of 24.5 °C to 35.8 °C in about 95% of the times, which represents 6% of underestimation in this range; meanwhile, DG Tmax data are more similar to MWD for the ECDF (Fig. 1a). NP overestimated Sr in the range of 11 to 32 MJ m\(^{-2}\) d\(^{-1}\) in about 89% of the cases (Fig. 1d); it represents 5% of overestimation.

DG overestimates MWD rainfall in 73% for the cumulative events in the range of 0.1 to 13.36 mm; from 8 to 81% of measured events (Fig. 1c), it represents 84.6% of overestimation in this range. On average, this overestimation is about 163 mm cycle\(^{-1}\). NP overestimates MWD rainfall in 73.7% of cumulative events in the range of 0.1 to 11.8 mm—from 5.4 to 79.1% of measured events (Fig. 1c), which results in a 95% of overestimation in this range. On average, this overestimation is about 160 mm cycle\(^{-1}\). This trend takes place in both GWD from all assessed municipalities (Fig. 2).

In comparison, MWD precipitation starts after about 38% of the cumulative events. From 80 and 79% of the events onwards, for DG and NP, respectively, GWD underestimates extreme rainfall events (Fig. 1c). On average, it
represents 27 and 31% of underestimation. This underestimation is about 165 and 196 mm cycle$^{-1}$ for DG and NP, respectively. DG underestimates total rainfall by 1.6 mm cycle$^{-1}$ (0.2%), whereas NP underestimates it by 36 mm cycle$^{-1}$ (4.5%).

Scatter plot with commonly employed indexes was also employed for comparison to other studies. DG data show superior performance against NP data for all weather variables assessed (Fig. 3). Tmax is the variable with best agreement and fit to MWD in DG dataset, showing coefficient of determination ($R^2$), root-mean-square error (RMSE) and agreement index (d) of 0.77, 1.34, and 0.93 against 0.46, 2.87, and 0.74 from NP, respectively (Fig. 3a and b). NP shows a tendency to underestimate Tmax, thus exhibiting higher ME value than DG, -1.84 against -0.3, respectively (Figs. 1a and 3b).

Tmin shows overall lower agreement with MWD from both GWD when compared to Tmax. In this case, DG also shows better agreement and fit than NP, with $R^2$, RMSE and d values of 0.6, 1.19, and 0.86 against 0.47, 1.49, and 0.83 from NP (Fig. 3c and d). Rainfall has the worst performance amongst all variables in both GWD when compared with MWD. DG shows $R^2$, RMSE, and d values of 0.42, 10.7, and 0.75, whereas NP showed 0.11, 13.67, and 0.53, respectively.

### 3.2 Crop modeling outputs

All locations are assessed as an aggregate for statistical analysis (Figs. 4 and 5), and each location is assessed individually in Table 2. GWD show a better performance simulating Yp than Ya, but NP only exhibited better performance when all locations were aggregated (Table 2). Both GWD-based physiological maturity simulated days showed good agreement, with $R^2$ of 0.66 and 0.74, and RMSE of 1.93 and 2.85 for DG and NP, respectively (Fig. 5).

Porto Nacional is the location with overall better performance for both GWD, $R^2$ of 0.65, 0.59, and 0.61, RMSE 0.29, 0.73, and 0.57 for each DG condition (Yp, Ya clay and Ya sandy), respectively, and $R^2$ of 0.17, 0.67, and 0.67 and RMSE of 0.63, 0.79, and 0.60 for each NP condition, respectively. The worse agreement in the NP Yp condition in relation to Ya condition also strengthens the fact of lower NP agreement in Tmax, Tmin, and RS (Figs. 1 and 2) since these variables define Yp. Rio Branco is the city with worse performance in both GWD, except for DG Ya clay soil condition.
4 Discussion

4.1 Weather variables

This is the first study to evaluate the rainfall distribution pattern created by gridded weather database (GWD) considering its impact on soil condition in crop modeling. These patterns of GWD rainfall are of critical importance, which are prone to generate mistaken attainable yields. There are studies related to GWD rainfall estimation uncertainties in amount and distribution, such as Ruane et al. (2015), Xavier et al. (2015) and Mourtzinis et al. (2017). Battisti et al. (2019) showed that Daily Gridded (DG) underestimated rainfall by 21.17 mm cycle\(^{-1}\) in the soybean crop; they also found good agreements between DG and MWD for Tmax and Tmin, reporting an \(r\) of 0.92 and RMSE of 0.72 for Tmax, and \(r\) of 0.87 and RMSE of 1.05 for Tmin.

The RMSE observed in DG rainfall data (Fig. 3) agrees with the ones reported by Xavier et al. (2015) for Amazon and Tocantins river basins, with values of 13.17 and 10.54, respectively. Lower ME values from DG agree with previous statements, given that the dataset has a lower underestimation of total rainfall than NASA/POWER (NP). Both GWD have a good RS, agreement with MWD; DG has a slightly better agreement, showing \(R^2\), RMSE, and \(d\) of 0.65, 3.09, and 0.9, respectively, against 0.52, 4.09, and 0.84 from NP.

4.2 Simulated yield and phenology

Battisti et al. (2019) found a better performance at simulating attainable (Ya) and potential (Yp) yield using DG datasets for simulating soybean crop modeling. In our study, we found that DG has a better performance when simulating Yp and Ya than NP (Table 2, Figs. 4 and 5).
Since potential yield depends only on maximum (Tmax) and minimum (Tmin) air temperature, as well as solar radiation (Sr), DG also showed better agreement than NP. Because Ya has rainfall as input, the most uncertain variable, it is expected a reduction in its agreement to MWD ones.

The ORYZA (v3) crop model phenology is defined by air temperature and photoperiod (Bouman et al. 2001). Errors in GWD-based simulations for these variables are low enough not to cause high-magnitude errors for this estimation (Figs. 2 and 5). In this case, the phenology is the same for all soil conditions, since drought do not change rice phenology in the simulation.

Weather variables play an important role in growth and crop yield (Sridevi and Chellamuthu 2015). Tropical rice has an optimal temperature range between 25 and 35 °C. High temperatures, above 35 °C (Hussain et al. 2019) for most cultivars or 36.6 °C for the studied cultivar (Heinemann et al. 2015), have a negative impact on growth and pollination, leading to spikelet sterility. Likewise, low temperature,
below 25 °C, may cause delays in phenological stages (Hussain et al. 2019). Both situations decrease productivity.

### 4.3 Soil water availability effect on attainable yield

In our study, it is observed that GWD-based simulations have a closer result to Yp; meanwhile, GWD overestimate Ya, when compared to MWD-based simulations (Fig. 4). GWD overestimate Ya as a consequence of misestimating the daily rainfall amount and distribution (Figs. 1 and 3) (van Wart et al. 2013a; Mourtzinis et al. 2015; Xavier et al. 2015).

In general, this overestimation took place mainly in the distribution range of 0.1 to 12 mm, which accounts for around 70% of the times observed (Fig. 3). On the other hand, extreme rainfall events (> 50 mm d⁻¹) are underestimated by GWD (Figs. 1, 2 and 3). Therefore, gridded weather data end up simulating an irrigation-sheet-like
rainfall pattern in the range of 0.1 to 12 mm, an overestimation by GWD-based rainfall. This rainfall pattern is incompatible with measured weather data, resulting in increased productivity for all soils. As a consequence of soil characteristics, its overestimations increase as soil textures become coarser. It is possible to visualize this difference in the cumulative evapotranspiration (ETCUM) (Fig. 6).

Both regression lines from GWD and MWD showed that in both soil conditions, GWD overestimates MWD ETCUM, with higher overestimations in the sandy soil condition (Fig. 4e and f). It is noticeable that cumulative evapotranspiration (ETCUM) from GWD tends to have increased overestimation in sandy soils, even though the accumulated rain is similar (Fig. 6e, f, i, and j). The violin plot (Fig. 4i and j) shows MWD ETCUM values range at lower values than GWD ones. The opposite pattern is seen in accumulated drainage (DRAIN) (Fig. 6k and l). In this regard, GWD DRAIN values are ranging in values lower than MWD ones, showing that irrigation-sheet-like rain pattern generated by GWD reduces total drainage, a consequence of an underestimation of high-precipitation events and an overestimation of precipitation values in the range of 0.1 to 12 mm.

According to Bouman et al. (2001), when drought occurs in the ORYZA (v3) crop model, solar radiation absorption is negatively affected due to leaf rolling, reduced leaf expansion rate, and changed assimilate partitioning. Hence, the reduced PARCUM in sandy soil conditions. Li et al. (2017) stated that photosynthesis for non-drought-tolerant cultivars linearly decreases as the soil water content decreases; the yield for drought-tolerant cultivars would be mildly impacted in a mild drought.

It is observed that DG- and NP-based crop simulations overestimated YP in 3.5% and 13.7%, respectively. Meanwhile, for Ya in clay soil condition, there were 9.6% and 18.4% yield overestimation, with 67% and 75% of total events overestimating MWD, for DG and NP, respectively. The widest differences between GWD- and MWD-based simulations are observed for Ya in sandy soil condition, where there are 21.3% and 29.1% yield overestimation, with 84% and 85% of total events overestimating MWD for DG and NP datasets. These overestimation percentages are based on the mean errors and the average yield (Table 2, Figs. 4 and 7).

In Yp conditions, NP-based simulations overestimate MWD-based simulations in a wider productivity range, from 2.5 to 6.5 Mg ha⁻¹ (Fig. 7). NP overestimates MWD in about 20% of the simulations before DG, overestimating MWD ones in the range of 2.5 to 6.5 Mg ha⁻¹, whereas DG overestimates in the range of 3 to 5 Mg ha⁻¹ in Fig. 7a. In Ya clay soil condition, both GWD overestimate MWD around the same range, from 2 to 4 Mg ha⁻¹ in Fig. 7b. In Ya sandy soil condition as in occurred before, both GWD overestimate in about the same range of conditions, ranging from 1 to about 2.5 Mg ha⁻¹ in Fig. 7c. It is noteworthy to point out that the graph curve is further away in relation to the last case (Fig. 7a, b, and c). Overestimation worsens as soil water storage capacity decreases.

In the clay soil condition, only two DG GWD municipalities show higher values of attainable yield than MWD (Paracatu and Rio Branco), NP did not show any underestimation at clay soil condition, whereas both GWD did not show underestimation at sandy soil conditions (Table 2), which are less prone to happen. We also divided the 32 years’ time series in two in order to assess the temporal impact on crop model (Table 3), it is possible to notice that DG overestimates even more for all conditions in the second part of time series, in this period DG used more rain gauge in interpolations than the first period assessed (Xavier et al. 2015). NP overestimated more in the second period for Yp condition and underestimated for Ya in sandy soil condition (Table 3).

Similar yield trends are observed for all simulated conditions along 32 seasons (Fig. 7d, e, and f). Soybean simulations by Battisti et al. (2019), based on GWD and MWD, also found the same trends along sowing years in simulating attainable yield. These similar systematic errors may indicate model sensitivity on capturing climate effects on yield variability (Pirtiojja et al. 2015).
Ya was overestimated by the model running the GWD, similar to what has been found in other studies. Battisti et al. (2019), simulations of soybean yield with DG dataset, found the ME for Yp and Ya of 69 kg ha\(^{-1}\) and 178 kg ha\(^{-1}\), respectively. It was concluded that the DG data estimate soybean yield withinacceptable error boundaries. However, this study used a soil with 0.162 cm\(^3\) cm\(^{-3}\) soil water storage, classified as high soil water
storage (Jensen and Allen 2016), and set the root growth to hit maximum values of 120 cm, possibly smoothing the errors related to GWD precipitation rain pattern. Van Wart et al. (2015) found a ME for Ya of about 10% on maize, wheat, and rice simulations in several countries using the NP dataset. There is no study that analyzed the interaction between GWD rain pattern and soil water storage.

### 4.4 Agro-climatic zones and crop models

The agro-climatic zones are a technic used to classify climatic homogeneity areas for multiple proposes, defining the area that needs to be covered to simulate or represent the region (van Wart et al. 2013b). For example, sowing date is one of the crop managements that plays a major role in reducing climate risk during critical crop phases (Li et al. 2013). In Brazil, Agricultural Zoning of Climatic Risk (ZARC, 2019) has been used to establish the best sowing dates, using an approach that includes homogeneous agro-climatic zones and crop modeling. The agro-climatic zones define areas where climatic risk is low for the crop and has been used since the 1995/1996 crop season for reimbursement of agricultural activity insurance program and as an instrument of public by the Ministry of Agriculture, Livestock and Food Supply-MAPA (Rossetti 2001).

The pioneer studies of sowing periods, which gave methodological support for the implementation of the ZARC, were developed with the upland rice in the 1980s (Steinmetz and Silva 2017).
Fig. 7 Empirical cumulative probability function (ECDF) (a, b and c) and line chart (d, e and f) showing the performance of measured data (MWD) in purple and gridded data (GWD), Daily Gridded (DG) in green and NASA/POWER (NP) in orange, considering three conditions, potential yield (Yp) (a and d), attainable yield in clay soil (b and e) and attainable yield in sandy soil (c and f). Line charts show the average yield values considering all locations along 32 crop seasons, whereas ECDF considers all values.

### Table 3

| Time series | Condition           | Datasets Yield Mg ha⁻¹ | Datasets Overestimation |
|-------------|---------------------|------------------------|-------------------------|
|             |                     | MWD        | DG         | NP         | DG         | NP         |
| 1984–2015   | Yp                  | 395        | 409        | 449        | 354        | 1367       |
|             | Ya clay soil        | 260        | 285        | 308        | 961        | 1845       |
|             | Ya sandy soil       | 155        | 188        | 199        | 2135       | 2911       |
|             | Yp                  | 402        | 409        | 446        | 174        | 1095       |
| 1984–1999   | Ya clay soil        | 271        | 287        | 322        | 590        | 1882       |
|             | Ya sandy soil       | 160        | 193        | 213        | 2063       | 3313       |
|             | Yp                  | 387        | 408        | 451        | 543        | 1654       |
| 2000–2015   | Ya clay soil        | 248        | 282        | 293        | 1371       | 1815       |
|             | Ya sandy soil       | 148        | 182        | 185        | 2297       | 2500       |

MWD—measured weather database; DG—Daily Gridded; NP—NASA/POWER; ME—mean error
Climatic homogeneity areas can be defined using basic climatic data (van Wart et al. 2013b) or process-based crop models (Heinemann et al. 2015). Crop models are tools that can be used, for example, to define climatic homogeneity areas (Heinemann et al. 2015; Battisti and Sentelhas 2019), sowing dates Paixão et al. 2021; Sampaio et al. 2021), irrigation (Justino et al. 2019), and impact of future climate scenarios (Battisti et al. 2018). However, the crop model calibration and validation are required for a given region (Challnor et al. 2015; Iizumi et al. 2021). In this context, Heinemann et al. (2015) used the crop model Oryza2000, calibrated for upland rice in Brazil, to classify similar environment for sowing dates and soil types across central Brazil. Further, crop models can quantify yield potential (Yp) and attainable yield (Ya), as well as rice yield gaps and their causes in upland rice (UR) (Heinemann et al. 2019).

4.5 Limitations and future work

Although current findings point out an acceptable correspondence between GWD and MWD on crop model for rice, several limitations must be considered. GWD are known to be less accurate in regions where MWD are less available or do not exist (van Wart et al. 2013a; Xavier et al. 2015). These uncertainties increase as virtual weather stations get further away from measured weather stations (van Wart et al. 2015). Second, GWD do not estimate daily MWD rainfall distribution with accuracy. Therefore, we suggest caution when using GWD as an input in crop models, especially in sandy soils, since they tend to have lower water storage, which is prone to worse attainable yield overestimation in upland rice.

We assume this may happen in other crops as well, but there are no studies supporting this hypothesis. Third, studies comparing crop performance among different regions with distinct soil profiles may generate misleading results. Fourth, the Brazilian agricultural climate risk zoning (ZARC, 2019) considers different types of soils for its results. GWD data must be used parsimoniously, given that the rainfall pattern may result in a misleading lower risk profile for both soils. Lastly, according to Li et al. (2017), ORYZA (v3) crop model has coupled nitrogen uptake with water uptake, hence affecting nitrogen studies with ORYZA (v3) that relies on GWD.

5 Conclusion

This study has shown that both Daily Gridded and NASA/POWER may be suitable weather data sources for simulating upland rice (UR) potential yield (Yp). Despite this accuracy, attainable yield (Ya) modeling should be avoided, especially short-cycle cultures on daily-step model simulations. These uncertainties are even higher at sandy soil conditions since it has lower soil water storage. Further studies using the GWD dataset for other crops in a wide range of locations and soils could improve its estimation capabilities while defining its acceptable use range.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00704-021-03906-4.

Acknowledgements The authors would like to thank: The National Council for Scientific and Technological Development (CNPq) for the support to this study through a research fellowship for the first author.

Author’s contribution All authors contributed equally to all sections of the manuscript. All authors read and approved the final manuscript.

Data availability Data can be made available if necessary.

Code availability Not applicable.

Declarations section

Ethics approval Not applicable.

Consent to participate All authors agreed to participate in the manuscript.

Consent for publication All authors agree with the publication of the manuscript in Theoretical and Applied Climatology.

Conflicts of interest The authors declare no conflicts of interest.

References

Alvares CA, Stape JL, Sentelhas PC, de Moraes Gonçalves JL, Sparovek G (2013) Köppen’s climate classification map for Brazil. Meteorol Z 22:711–728. https://doi.org/10.1127/0941-2948/2013/0507
Andrea MSC, Dallacort R, Barbieri JD, Tieppo RC (2019) Impactos de Predições Climáticas Futuras sobre o Milho de Segunda Safra em Agrossistema em Região de Transição de Bioma no Estado do Mato Grosso. Rev Bras Meteorol [online] 34:335–347. https://doi.org/10.1590/0102-77863340241
Angstrom A (1924) Solar and terrestrial radiation. Q J R Meteorol Soc 50:121–125. https://doi.org/10.1002/qj.49705021008
Battisti R, Sentelhas PC, Parker PS, Nendel C, Câmara GMS, Farias JRB, Basso CJ (2018) Assessment of crop-management strategies to improve soybean resilience to climate change in Southern Brazil Crop & Pasture. Science 69:154–162. https://doi.org/10.1007/s00704-018-2383-y
Battisti R, Bender FD, Sentelhas PC (2019a) Assessment of different gridded weather data for soybean yield simulations in Brazil. Theor Appl Climatol 135:237–247. https://doi.org/10.1007/s00704-018-2383-y
Battisti R, Casaroli D, Júnior JA, Evangelista AWP, Mesquita M (2019) Agro-climatic zoning of bamboo as a support for crop farming in the central-north region of the Brazilian Savannah. Pesqui Agropecu Trop 49. https://doi.org/10.1590/1983-40632019v4952794
Battisti R, Ferreira MDP, Tavares EB, Knapp FM, Bender F, Casaroli D, Alves Junior J (2020) Rules for grown soybean-maize cropping system in Midwestern Brazil: Food production and economic profits. Agric Syst 182:102850. https://doi.org/10.1016/j.agsy.2020.102850

Battisti R, Sentelhas PC (2019) Characterizing Brazilian soybean-growing regions by water deficit patterns. Field Crop Rev 24:95–105. https://doi.org/10.1016/j.fcr.2019.06.007

Bouman BAM, Kropff MJ, Tuong TP, Wopereis MCS, Ten Berge HFM, Van Laar HH (2001) Oryza2000: Modeling Lowland Rice. Los Baños (Philippines): International Rice Research Institute, and Wageningen, Netherlands: Wageningen University and Research Centre. 235 pp

Bouman BAM, Van Laar HH (2006) Description and evaluation of rice growth model Oryza 2000 under nitrogen limited conditions. Agric Syst 87:249–273. https://doi.org/10.1016/j.agsy.2004.09.011

Boling AA, Bouman BAM, Tuong TP, Murty MV, Jatmiko SY (2007) Modeling the effect of groundwater depth on yield-increasing interventions in rainfed lowland rice in Central Java, Indonesia. Agric Syst 92:115–139. https://doi.org/10.1016/j.agsy.2006.05.003

Boling AA, Tuong TP, Keulen H, Bouman BAM, Suganda H, Spiertz JHJ (2010) Yield gap of rainfed rice in farmers’ fields in Central Java, Indonesia. Agric Syst 103:307–315. https://doi.org/10.1016/j.agsy.2010.02.003

Caetano JM, Alves Júnior J, Casaroli D, Evangelista AWP (2021) Estimated productivity of sugarcane through the Agro-Ecological Zone method. Revista Ceres 68:1–9. https://doi.org/10.1590/0034-737X202168010001

Caetano JM, Casaroli D (2017) Sugarcane yield estimation for climatic conditions in the state of Goiás. Revista Ceres 64:298–306. https://doi.org/10.1590/0034-737X201764030011

Santos TG, Battisti R, Casaroli D, Alves Júnior J, Evangelista AWP (2021) Assessment of agricultural efficiency and yield gap for soybean in the Brazilian Central Cerrado biome. Bragantia 80:e1821. https://doi.org/10.1590/1678-4499.20200352

Cesconetto LB, Pruski FF, Rodriguez RDG, Marcaggi GE (2019) Potentiality of sugarcane expansion under irrigation conditions considering natural and potential water availability. Agric Water Manag 203:162–171. https://doi.org/10.1016/j.agwat.2018.03.008

Challinor AJ, Parkes B, Ramirez-Villegas J (2015) Crop yield response to climate change varies with cropping intensity Glob. Change Biol 21:1679–1688. https://doi.org/10.1111/gcb.12808

CONAB (2019) Survey of crop season; rice. http://www.conab.gov.br/conteudos.php?area=1253&. Accessed 20 Jun 2019

Dubrovsky M, Huth R, Dabihi H (2020) Parametric gridded weather generator for use in present and future climates: focus on spatial temperature characteristics. Theor Appl Climatol 139:1031–1044. https://doi.org/10.1007/s00704-019-03027-z

Espe MB, Cassman KG, Yang H, Guipilart N, Grassini P, van Wart J, Anders M, Beighley D, Harrell D, Linscombe S, McKenzie K, Mutters R, Wilson LT, Linquist BA (2016) Yield gap analysis by Using a Central Pivot System to Irrigate Common Beans during the Rainfed Season in Central Brazil. Agric Water Manag 164:276–283. https://doi.org/10.1016/j.-agwat.2016.07.011

Feng L, Bouman B, Tuong T, Cabangon R, Li Y, Lu G, Feng Y (2007) Exploring options to grow rice using less water in northern China using a modeling approach I Field experiments and model evaluation. Agric Water Manag 88(3):1–13. https://doi.org/10.1016/j.agwat.2006.10.006

Heinemann AB, Dingkuhn M, Luquet D, Combes J, Chapmann S (2007) Characterization of drought stress environments for upland rice and maize in central Brazil. Euphytica 162(3):395–410. https://doi.org/10.1007/s10681-007-9579-z

Heinemann AB, Barrios-Perez C, Ramirez-Villegas J, Arango-Loango D, Bonilla-Findji O, Medeiros JC, Jarvis A (2015) Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. J Exp Bot 66:3625–3638. https://doi.org/10.1093/jxberv126

Heinemann AB, Ramirez-Villegas J, Rebolledo MC, Neto GMFC, Castro AP (2019) Upland rice breeding led to increased drought sensitivity in Brazil. F Crop Res 231:57–67. https://doi.org/10.1016/j.fcr.2018.11.009

Heinemann AB, Ramirez-Villegas J, Stone LF, Silva AP, Matta DH, Diaz MEP (2020) The impact of El Niño Southern Oscillation on cropping season rainfall variability across Central Brazil. Int J Climatol 1:1–22. https://doi.org/10.1002/joc.6684/joc.6684

Hussain S, Khalqia A, Ali B, Hussain S (2019) Temperature extremes: impact on rice growth and development. In: Plant abiotic stress tolerance agronomic, molecular and biotechnological approaches. Springer, Berlin, pp 153–171. https://doi.org/10.9787/PBB.2018.6.1.1

Iizumi T, Yoshimitsu Masaki Y, Takahiro Takimoto T, Yuji Masutomi Y (2021) Aligning the harvesting year in global gridded crop model simulations with that in census reports is pivotal to national-level model performance evaluations for rice. Eur J Agron 130:126367. https://doi.org/10.1016/j.eja.2021.126367

INMET. The Brazilian National Institute of Meteorology (2019) Available at: <http://www.inmet.gov.br/portal/ Accessed 20 may 2019

Jha PK, Athanasiaidis P, Gualdi R, Trabucco A, Shelia V, Hoogenboom G (2019) Evaluating the applicability of using daily forecasts from seasonal prediction systems (SPSs) for agriculture: a case study of Nepal’s Terai with the NCEP CFSv2. Theor Appl Climatol 135:1143–1156. https://doi.org/10.1007/s00704-018-2433-5

Justino LF, Alves Júnior J, Battisti R, Heinemann AB, Leite CV, Evangelista AWP, Casaroli D (2019) Assessment of Economic Returns to using a Central Pivot System to Irrigate Common Beans during the Rainfed Season in Central Brazil. Agric Water Manag 224:105749. https://doi.org/10.1016/j.agwat.2019.105749

Laborte AG, de Bie KC, Smaling EM, Moya PF, Boling AA, Van Itersum MK (2012) Rice yields and yield gaps in Southeast Asia: past trends and future outlook. Eur J Agron 36:9–20. https://doi.org/10.1016/j.eja.2011.08.005

Li T, Raman AK, Marcada M III, Kumar A, Angeles O, Radanielson M (2013) Simulation of genotype performances across a larger number of environments for rice breeding using ORYZA2000. Field Crops Res 149:312–321. https://doi.org/10.1016/j.fcr.2013.05.006

Li T, Angeles O, Marcada M III, Manalo E, Manalili MP, Radanielson A, Mohanty S (2017) From ORYZA2000 to ORYZA (v3): an improved simulation model for rice in drought and nitrogen-deficient environments. Agric for Meteorol 237:246–256. https://doi.org/10.1016/j.agrmet.2017.02.025

Menezes CT (2021) Avaliação de estações climáticas virtuais e do modelo ORYZAv3 para a região de arroz tropical irrigado, Master Thesis, Universidade Federal de Goiás, Goiânia, Brazil.

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

Mourtzinis S, Edreira JIR, Conley SP, Grassini P (2017) From grid to field: assessing quality of gridded weather data for agricultural applications. Eur J Agron 82:163–172. https://doi.org/10.1016/j.eja.201610013

Paixão JS, Casaroli D, Battisti R, Evangelista AWP, Alves Júnior J, Mesquita M (2020) Characterizing Sugarcane Production Areas Using Actual Yield and Edaphoclimatic Condition Data for the State of Goiás Brazil. Int J Plant Prod 14:511–520. https://doi.org/10.1007/s42106-020-00101-9
Paixão JS, Casaroli D, dos Anjos JCR, Alves Júnior J, Evangelista AWP, Dias HB, Battisti R (2021) Optimizing Sugarcane Planting Windows Using a Crop Simulation Model at the State Level. Int J Plant Prod (2021) 101007/s42106–021–00134–8

Perin V, Sentelhas PC, Dias HB, Santos EA (2019) Sugarcane irrigation potential in Northwestern São Paulo, Brazil, by integrating Agrometeorological and GIS tools. Agric Water Manag 220:30–36. https://doi.org/10.1016/j.agwat.201904012

Pirrittoja N, Carter TR, Fronzek S et al (2015) Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. Clim Res 65:87–105. https://doi.org/10.3354/cr01322

Prescott JA (1940) Evaporation from a water surface in relation to solar radiation. Trans Roy Soc Sci Austr 64:114–118. https://doi.org/10.1155/2013/168048

Ramirez-Villegas J, Heinemann AB, Pereira de Castro A, Breseghello F, Navarro-Racines C, Li T, Rebollo MC, Challinor AJ (2018) Breeding implications of drought stress under future climate for upland rice in Brazil. Glob Chang Biol 24:2035–2050. https://doi.org/10.1111/gcb14071

Rossetti LA (2001) Zoneamento agrícola em aplicações de crédito e securidade rural no Brasil: aspectos atuariais e de política agrícola. Revista Brasileira De Agrometeorologia 9(3):386–399

Ruane AC, Goldberg RC, Chryssanthacopoulos J (2015) Climate forcing datasets for agricultural modeling: merged products for gap-filling and historical climate series estimation. Agric For Meteorol 200:233–248. https://doi.org/10.1016/j.agrformet.201409016

Sampaio LS, Battisti R, Lana MA, Boote KJ (2021) Assessment of sowing dates and plant densities using CSM-CROPGR-Oilseed soybean maturity groups in low latitude. J Agric Sci 1:1–14. https://doi.org/10.1017/S0021859621000204

Sparks A (2018) nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R J Open Source Softw 3(30):1035 https://doi.org/10.21105/joss.01035

Sridevi V, Chellamuthu V (2015) Impact of Weather on Rice-a Review. Int J Appl Res 1:825–831

Steinmetz S and Silva SC (2017) Início dos estudos sobre zoneamento agrícola de risco climático (ZARC) no Brasil. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 27 p.

Teixeira WR, Battisti R, Sentelhas PC, Moraes MF, Oliveira Junior A (2019) Uncertainty assessment of soybean yield gaps using DSSAT-CSM-CROPGR-Oilseed soybean calibrated by cultivar maturity groups. J Agron Crop Sci 205:533–544. https://doi.org/10.1111/jac.12343

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/gcb12302

van Wart J, van Bussel LGi, Wolf J, Licker R, Grassini P, Nelson A, Boogaard H, Gerber J, Mueller ND, Claessens L, Ittersum MK, Cassman KG (2013) Use of agro-climatic zones to upscale simulated crop yield potential. Field Crop Res 143:44–55. https://doi.org/10.1016/j.fcr.2012.11.023

van Wart J, Grassini P, Yang H, Claessens L, Jarvis A, Cassman KG (2015) Creating long-term weather data from thin air for crop simulations modeling. Agr Forest Meteorol 209–210:49–58. https://doi.org/10.1016/jagrfmet.201502020

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

Xavier AC, King CW, Scanlon BR (2015) Daily gridded meteorological variables in Brazil (1980–2013). Int J Climatol 36(6):2644–2659. https://doi.org/10.1002/joc.4518

ZARC - Zoneamento Agrícola de Risco Climático, MAPA - Ministério da Agricultura, Pecuária e Abastecimento (2019) https://indicadoresagriculturagovbr/zarc/index.htm

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