Alfaro, Marcela Daniela Mollericona; Lopes, Iug; Montenegro,
Abelardo Antônio Assunção; Leal, Brauliro Gonçalves
CFSR- NCEP Performance for weather data forecasting in the Pernambuco Semiarid, Brazil

CFSR- NCEP Performance for weather data forecasting in the Pernambuco Semiarid, Brazil

Marcela Daniela Mollericona Alfaro [a], Iug Lopes [a], Abelardo Antônio Assunção Montenegro [a] & Brauliro Gonçalves Leal [b]

[a] Departamento de Engenharia Agrícola, Universidade Federal Rural de Pernambuco, Pernambuco, Brasil. marcela.mollericonaalfaro@yahoo.com, iuglopes@hotmail.com, montenegro.ufrpe@gmail.com,
[b] Universidade Federal do Vale do São Francisco, Pernambuco, Brasil. brauliro.leal@univasf.edu.br

Abstract
The present study aims to evaluate meteorological data -with real time actualization- from the *Climate Forecast System Reanalysis* (CFRS) of the *National Centers for Environmental Prediction* (NCEP), comparing them with data from local stations in two mesoregions: Sertão de Pernambuco (SP) and Sertão do São Francisco (SFF), semi-arid region of Pernambuco, Brazil. Statistical performance indicators were used for the period since 1979 to 2014 and the variables: precipitation (P), average, minimum and maximum temperature (Tm, Tn, Tx respectively), relative humidity (HR), wind speed (Vv), solar radiation (RS) and potential evapotranspiration (ETo). Tn, Tm and Tx showed the best results for the determination coefficient (R^2), Willmott concordance index (*d*), Nash-Sutcliffe efficiency index (NSE) and percentage bias (PBIAS). ETo, P and HR obtained acceptable values for R^2, *d* and NSE. CFSR data shows good performance with *d* values between 0.63 and 0.94.

Keywords: CFSR; NCEP; weather stations; climate forecast system reanalysis.

Desempeño del CFSR- NCEP en la predicción de datos meteorológicos en el Semiárido de Pernambuco, Brasil

Resumen
El presente estudio tiene como objetivo evaluar datos meteorológicos –con actualización en tiempo real- del *Reanálisis del Sistema de Predicción del Clima* (CFRS) de los *Centros Nacionales de Predicción Ambiental* (NCEP), comparándolos con datos de estaciones locales en dos mesoregiones: Sertão de Pernambuco (SP) y Sertão do São Francisco (SFF), región Semiárido de Pernambuco, Brasil. Se emplearon indicadores estadísticos de desempeño, para el período de 1979 a 2014 y las variables: precipitación (P), temperatura media, mínima y máxima (Tm, Tn, Tx respectivamente), humedad relativa (HR), velocidad del viento (Vv), radiación solar (RS) y evapotranspiración potencial (ETo). La Tn, Tm y Tx demostraron los mejores resultados para el coeficiente de determinación (R^2), índice de concordancia de Willmott (*d*), índice de eficiencia de Nash-Sutcliffe (NSE) y sesgo porcentual (PBIAS). La ETo, P y HR obtuvieron valores aceptables para R^2, *d* y NSE. Datos CFSR muestran buen desempeño con valores de *d* entre 0.63 a 0.94.

Palabras clave: CFSR; NCEP; estaciones meteorológicas; reanálisis del sistema de predicción del clima.

1. Introduction

Development requires efficient management of water resources and regions such as the semi-arid in northeast Brazil present important water management problems. The region has very specific climatic characteristics, such as being prone to drought, and therefore has a high exposure to agricultural losses that lead to food insecurity [2]. Studies of climatic conditions, especially those related to the water

balance, are of great importance [1] and could help overcome these challenges.

The scarcity of climate data - in quantity and quality - has been a problem in water resources modeling, as conventional weather stations with adequate spatial and temporary distribution are not always available; especially in developing countries. Also, when data exist, they may not be reliable due to gaps or random errors, and may not represent the climate of a river basin [3-5]. Therefore, alternative data sources are needed to better simulate hydrological processes [6].

Much of the current climate knowledge was obtained from global reanalysis data. The reanalysis, or retrospective analysis, consists of forecasting models and a data assimilation routine [7]. The reanalysis data set of the Global Forecasting System of the *Climate Forecast System Reanalysis (*CFRS) of the *National Centers for Environmental Prediction (*NCEP) can be a valuable option for forecasting, where conventional measurements are not available [7,8]. The use of CFSR data in basin modeling could be reliably applied and offers new opportunities in real-time modeling [4].

CFSR data set contains historical precipitation and temperatures for each hour anywhere in the world; being produced using state-of-the-art techniques (conventional meteorological observations and satellite irradiations) [10,4]. It is based on a fully coupled ocean-atmosphere model, which uses numerical weather forecasting techniques to assimilate and forecast atmospheric conditions with a resolution of 0.3125 (~ 38 km); also, the forecast models are restarted every 6 hours using information from the global weather stations network and satellite products [10,4]. The production involves various spatial, and temporal interpolations on meteorological data, other conventional observations and satellite products [3].

The advantages of CFSR over conventional data are that it provides complete climate data sets and has useful parameters for the use of Penman Montieth and Priestley-Taylor equations [11]. Data can be obtained at http://globalweather.tamu.edu/ from SWAT (Soil and Water Assessment Tool) - Texas A&M (TAMU) in the SWAT input format [5]. In addition to global spatial coverage, the CFSR offers a complete, continuous, and consistent record from 1979 to the present, providing a record of estimates of variables of limited availability such as solar radiation, air humidity, and wind speed [8]. This could allow a comprehensive modeling of watersheds in regions with non or missing data [12].

According to [13], compared to the previous NCEP reanalysis (R1 and R2), there are three major differences with the CFSR, such as 1) higher horizontal and vertical resolution (horizontal spectrum T382, ± 35 km); 2) a forecast generated from a coupled ocean-ice land-atmosphere system, and 3) historical assimilations of satellite radiation. Different studies demonstrated the applicability and satisfactory performance of the hydrological SWAT model with CFSR data in regions with scarce data [18]. As [8] in Ethiopia; [4] in the United States, using precipitation and temperature data,

obtaining discharge simulations as good or better than models using traditional meteorological measurements, when CFSR data are calculated over an area comparable to watershed areas; [7] concluded that the CFSR provided the best correlation in three regions of South America compared to other sets of reanalysis; [14] used CFSR data as input variables to the WXGEN data simulator in a basin with limited data, obtaining a satisfactory fit for the simulation of agricultural practice scenarios; [9] determined that the CFSR simulation was able to generate acceptable accuracy in China, making a previous validation with terrestrial MS data to confirm sufficient accuracy.

In Brazil, different investigations used CFSR data with the SWAT model, such as [15], demonstrated the possibility of using observed data and reanalysis jointly, where there was a deficiency of information or stations; [5] evaluated precipitation data, obtaining better performance in flow simulation (best statistical values) compared to other data sets, recommending that the use of CFSR data for variables other than precipitation, can provide reasonable hydrological responses; [6] also used data from local stations jointly with CFSR data, obtaining satisfactory results; [2] they evaluated CFSR data to reproduce temperature and rainfall extremes using climate indices and [13] observed remarkable improvements in large-scale precipitation patterns compared to previous reanalysis.

It is therefore noted that different authors highlight the importance of testing with reanalysis data to verify whether climatic characteristics represent specific local realities, especially in dry and data-poor areas [28]. In that sense, with the present study, the objective is to evaluate the performance of NCEP-CFSR data in the prediction of climate data in two mesoregions in the Pernambuco Semiarid, comparing them with observed data from local meteorological stations.

2. Materials and Methods

2.1. Geographical location

The study was developed with information from local weather stations (EM) belonging to the observation network of the National Institute of Meteorology (INMET). The 4 municipalities under study are located in the State of Pernambuco in Brazil, between the parallels 7°18'17" and 9°28'43" south latitude; and the meridians 34°48'15" and 41°21'22" west longitude, in the northeast of Brazil. The characteristic climate of the region is classified as BSh - hot semi-arid (steppe) climate - according to the Köppen Climate Classification - (average annual temperature > 18 °C) [13].

The stations are located in for municipalities, in the São Francisco River Basin and distributed in two mesoregions: Sertão of Pernambuco (SP) and Sertão of São Francisco Pernambucano (SSF), Table 1 and Fig. 1. The data sets were: minimum, average, and maximum temperature (Tn, Tm, Tx, °C), relative air humidity (HR, %), wind speed (Vv, m s^{-1}), global solar radiation ((RS, M J m^{-2}), precipitation (P, mm) and potential evapotranspiration (ETo, mm).
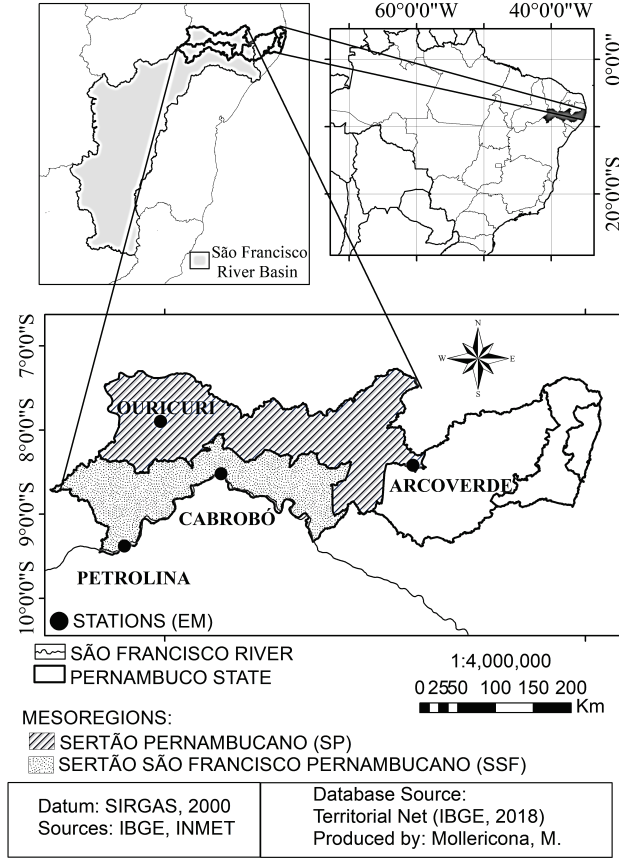
Figure 1. Location of weather stations (EM) in the 2 mesoregions: Sertão of Pernambuco, SP (Ouricuri and Arcoverde) and Sertão of São Francisco Pernambucano, SSF (Petrolina and Cabrobó)
Source: The Authors.

Table 1.
Weather stations (EM) under study in the two mesoregions: SP: Sertão Pernambucano, SFF: Sertão do São Francisco Pernambucano

Station	Mesoregion	Longitude (°)	Latitude (°)	Altitude (m)
Arcoverde	SP	37°05'W	08°25'S	683,91
Ouricuri	SP	40°03' W	07°54' S	462,01
Cabrobó	SSF	39°18' W	08°30' S	342,78
Petrolina	SSF	40°28'W	09°22'S	372,54

Source: The Authors.

EM data were compared with (global) reanalysis data (NCEP-CFSR), selected over the same areas, obtained for 35 years (1979 to 2014) through the portal: Global Weather Data for SWAT (https://globalweather.tamu.edu/) available from Texas A&M University. Data is provided in the SWAT file format, with a horizontal resolution of ~38 km (0.3125°) and global coverage [10,15]. Reanalysis data are obtained from state-of-the-art data assimilation techniques (observations from conventional weather stations such as satellite irradiations) as advanced components of atmospheric, oceanic, and surface modeling [9]. They include daily precipitation, temperatures, air humidity, solar radiation, and wind speed, provided on a Gaussian grid defined by the NCEP (designated T382), longitudes are equally spaced, but latitudes are not [10].

2.2. Methods

In the initial evaluation, box plotting techniques (boxplots), 1:1 dispersion, time series were used, which allowed a visual comparison and are essential for proper evaluation of models [16]. The analysis also included a comparison with central tendency and variability statistics. A regression analysis was performed with the correlation (r) and determination coefficients (R^2); an analysis with non-dimensional indices such as the Willmott Concordance Index (d) and the Nash-Sutcliffe Efficiency Index (NSE); along with error rates such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Percent Bias (PBIAS), Standard Deviation Rate (RSR), eq. (1)-(7). Regression coefficients determine the strength of the relationship between two databases, dimensionless techniques provide an assessment of the goodness of the relative adjustment, and error rates quantify the deviation of units from the data of interest [17].

$$d = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}[|P_i - \bar{O}| + |O_i - \bar{O}|]^2} \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \quad (3)$$

$$RMSE = \sqrt{\left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}\right]} \quad (4)$$

$$PBIAS = \left[\frac{\sum_{i=1}^{n}(Oi - P_i)*(100)}{\sum_{i=1}^{n}(Oi)}\right] \quad (5)$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}} \quad (6)$$

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\left[\sqrt{\sum_{i=1}^{n}(Oi - Pi)^2}\right]}{\left[\sqrt{\sum_{i=1}^{n}(Oi - \bar{O})^2}\right]} \quad (7)$$

Where n is the number of observations, Pi refers to the values of the meteorological variable obtained in the NCEP-CFSR database, and Oi is the data observed in the EM.

R^2 and r describe the degree of collinearity between the simulated and measured data [17]. The values of r vary from -1 to 1 and is an indicator of the degree of the linear relationship between two data series; R^2 varies from 0 to 1, with higher values indicating less error variation, and generally values higher than 0.50 is considered acceptable [17]. The r values were classified according to [18]. The NSE [19] widely used in climate forecasting, varies

from -∞ to 1, is more rigorous than R^2 [20], and determines the relative magnitude of the residual variance, indicating how well the observed and simulated data graph fits on the 1:1 line 1:1 [17,20,21] propose the classification: NSE=1 as "perfect fit"; NSE > 0.50 "satisfactory". Although the same authors consider that between 0 and 1 are generally seen as "acceptable" and <0 as "unacceptable".

Willmott's concordance index (d) [23] measures the degree of error-free prediction, ranging from 0 to 1 as "perfect concordance". To support the analysis, the performance index (c) [24,25] was calculated, which is the product of the correlation coefficient r and d, classified according to [24].

PBIAS (expressed as a percentage) measures the average trend of simulated data to be higher or lower than its observed counterparts, with the optimal value being 0 and low values indicating accurate simulation; positive values indicate underestimation bias and negative values indicate overestimation [17]. "Good performance" values are for 10% < PBIAS < 15% and "unsatisfactory" when the PBIAS ≥ 25%. The RSR varies from 0 for "perfect simulation" (RMSE=0) and the lower the value, the lower the RMSE, so the performance of the model is considered better [17]. Values of MAE and RMSE equal to 0 indicate "perfect fit". The degree to which RMSE exceeds MAE is an indicator of the extent to which outliers exist in the data [26].

3. Results and Discussions

Initially, results for precipitation are presented and discussed. Later, the other variables are analyzed.

In Fig. 2 it can be seen that NCEP-CFSR reanalysis data tend to underestimate EM precipitation in the SP mesoregion; similarly, to Petrolina (SSF mesoregion). The PBIAS (Table 2) agrees with the graphic observations, being that, for all areas, their values are positive, indicating an underestimation bias [17]. The highest PBIAS values correspond to SP (50.97).

Similar results were obtained by [3] who obtained that the CFSR showed more than 50% underestimation of P, in 37% of the sub-basins of the area under study, demonstrating that the CFSR does not represent P in most areas of a basin.

In Fig. 3, using the boxplot diagrams of the P in the EM and the NCEP-CFSR, located in the 4 areas studied, the presence of discrepant values (outliers) for all the data sets can be seen, as well as large differences in the maximum values between the two databases. It is also noted that the interquartile interval is lower for NCEP-CFSR data in all areas, which means a lower degree of data dispersion compared to EM data. In Table 2 the statistical metrics such as correlation r were close, being lower for Arcoverde and Petrolina (0.67 and 0.72), the other areas coincided equally with the value of 0.74; being classified as "high" and "very high". These results are higher compared to results obtained in other regions of South America such as Bolivia, according to the study of [30] that obtained values of $r < 0.3$.

On the other hand, the performance according to the

NSE was very low for Arcoverde (0.04) and "satisfactory" for the other 3 areas (0.48; 0.47 and 0.41). However, all these results were higher than those obtained by [28] (NSE= -2.02) in a region inserted in the Mata Atlántica biome.
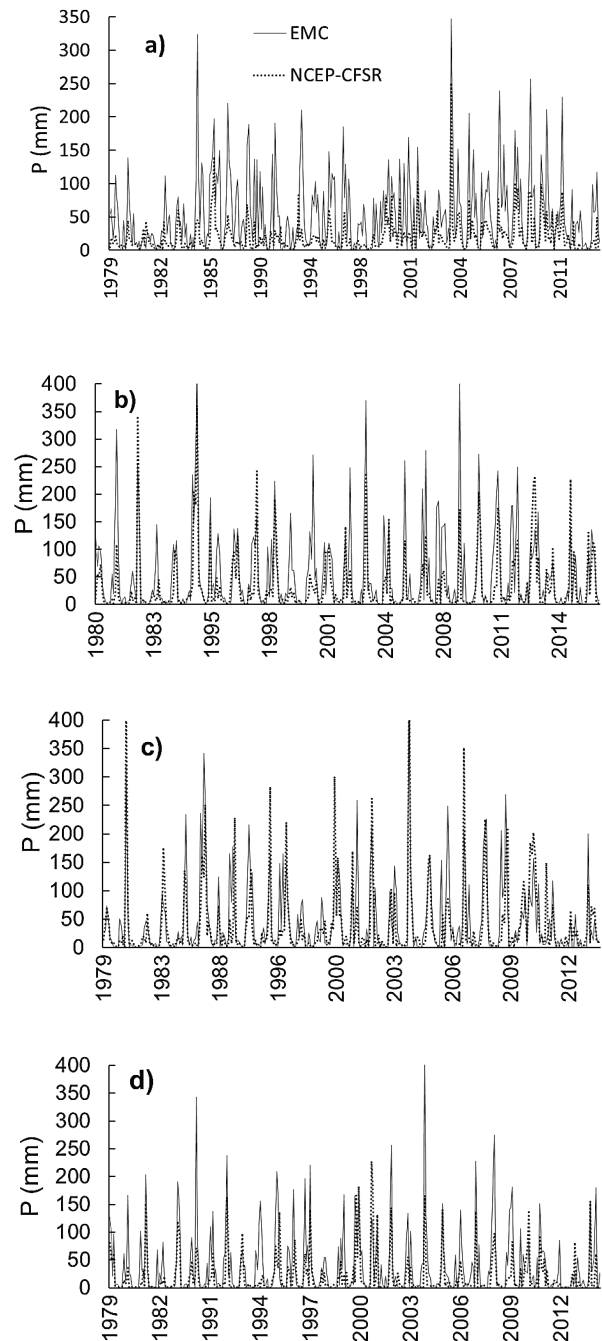

Figure 2. Time trend of monthly precipitation (mm) derived from EM and NCEP-CFSR for a) Arcoverde, b) Ouricuri, c) Cabrobó and d) Petrolina. Period 1979 - 2014
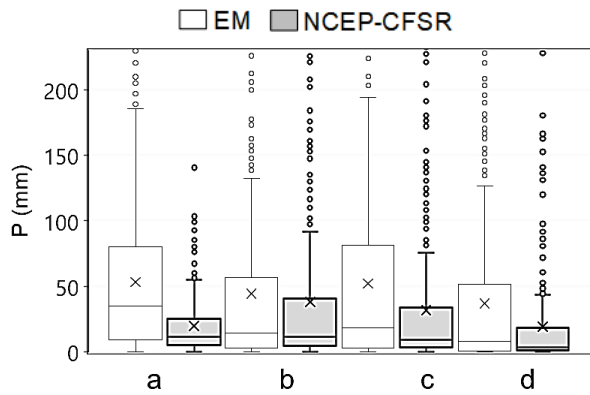Source: The Authors.

Figure 3. Precipitation (P) boxplots (mm) in the 4 areas understudy: a) Arcoverde, b) Cabrobó, c) Ouricuri and d) Petrolina with data from EM and NCEP-CFSR
Source: The Authors.

The determination coefficient R^2 (Fig. 4) obtained was qualified as "acceptable" according to [17] for Ouricuri (0.55), Cabrobó (0.54) and Petrolina (0.51). At mesoregional level (Table 2) it is observed that the values of R^2 are also "acceptable" for SP (0.50) and SFF (0.53), therefore, the value obtained for R^2 for the total area (0.51) is also considered "acceptable". [29] also analyzed CFSR precipitation data in a region characterized as semi-arid, in Paraiba State, and obtained R^2 values between 0.51 and 0.99 and NSE between 0.40 and 0.99, observing that the applicability of CFSR data for hydrological studies is demonstrated.

For the MAE and RMSE (Table 2), which indicate the error in the units of the parameter of interest - in this case the P - it is observed that SP presents the highest values (33.61 and 54.61 respectively) than SSF. These results coincide with those proposed by [31] on which stations likely affected by convective rainfall, have a better correlation coefficient and a smaller RMSE, being that the SSF mesoregion is at lower altitudes than SP. According to Willmott's concordance index d, related to the distance of observed with respect to the estimated values, which varies from 0 to 1 for no concordance and perfect concordance, respectively; it was obtained that the best result was for SSF (0.81) in comparison with SP (0.72).

Regarding the performance index c, for SP, the category "bad" and for SFF "low" was obtained.
Regarding the distribution effect -CFSR (predicted precipitation by grid) and stations (point ground observation)- [31] observed that the performance of the precipitation estimates from the MPEG and CFSR satellites for both point to point and areal comparisons (interpolated observed rainfall stations) was better than that of the precipitation amounts from the TRMM satellite. The study [5] compared inputs from the SWAT climate generator (G1), local stations (G2), NOAA's CFSR (G3), and NOAA's CFSR + local rain gauges (PL), obtaining similar values in the simulation of the SWAT water balance components for mean annual precipitation and potential evapotranspiration.
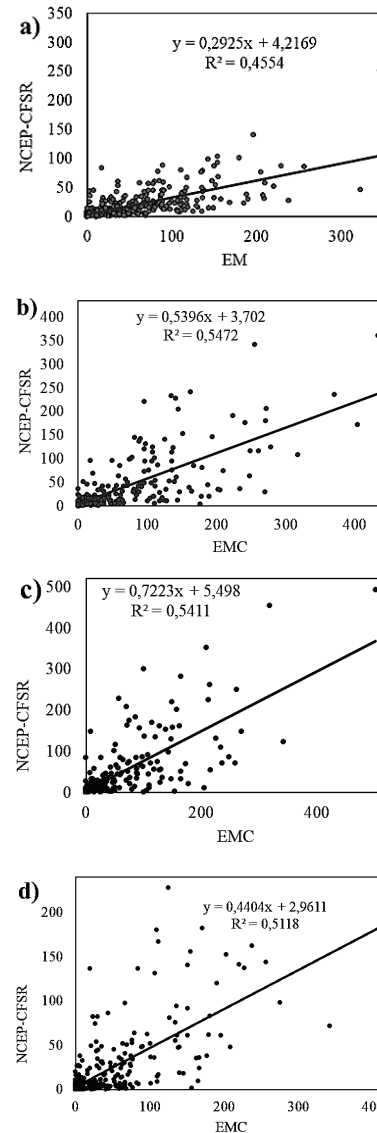

Figure 4. Comparison of monthly P (mm) between EM and NCEP-CFSR for a) Arcoverde, b) Ouricuri, c) Cabrobó and d) Petrolina. Period 1979 to 2014
Source: The Authors.

In that study, the performance of potential evapotranspiration (ETo) simulation with Penman-Monteith (PM), Prestley-Taylor, and Hargreaves methods, with CFSR + PM data obtained values of NSE > 0.75, classified as "satisfactory" to "very good". [6] they also found that the use of CFSR + data from local stations improved statistics in sub-basins with few river stations or with substantial missing data. The authors indicate that the differences found could have occurred due to the semi-arid climate of the region, with strong seasonal and interannual variability in rainfall, which could result in poorly calibrated CFSR data at local stations. In that sense, the authors suggest the use of CFSR data for climate parameters other than P (which are generally less reliable in quantity, and spatial distribution), together with P data from local rain gauges to provide reasonable simulations of

hydrological response in the semi-arid region. These results can be supported by [4], which indicates that in basins that are relatively arid or dry areas, hydrological modeling is more difficult; possibly because large runoff events are triggered by small, localized P events. Which are not represented by coarse-scale CFSR or EM data. [13] also supports that CFSR data exhibit a dry bias along the South American coast and the east coast of northeast Brazil. On the other hand, [12] suggest carefully checking the CFSR data against conventionally measured data climate stations.

The respective results and discussions for the other variables (Tn, Tm, Tx, HR, Vv, RS and ETo) are presented below.

Table 2 and Fig. 6 shows that the R^2 average for Tn, Tm, Tx, HR, Vv and ETo of the 2 mesoregions reached the R^2 value higher than 0.50 and recommended by [17]. But in the comparison at the mesoregion level, it can be seen that HR and Vv also achieve $R^2 > 0.50$ in the SFF mesoregion. The R^2 value of the RS was less than 0.50 in the two mesoregions. The correlation coefficient r in the SP reached classification values of "moderate" for RS; "high" for HR and Vv; "very high" for Tn, Tm, Tx and ETo. In SSF, "moderate" for RS; "very high" Tn, Tx, HR, Vv and ETo, and "almost perfect" for Tm. For the total area. the classification was "moderate" for Tn, Tm, Tx, Vv and "very high" for ETo. Both at the mesoregion level and the total area, the RS presented the lowest values of r and R^2.

The values obtained for the performance index c – which combines r and d indexes -with its respective classification are presented in Table 3, where it is highlighted that RS obtained the lowest performance classification (awful) for both mesoregions and the total area.

Table 2.
Statistical performance indicators of the comparison of variables from weather stations (EM) and NCEP-CFSR reanalysis data, in the mesoregions Sertão Pernambucano (SP) and Sertão do São Francisco (SFF)

	R^2	r	NSE	d	RMSE	PBIAS	RSR	MAE
				SP				
Tn	0.62	0.79	0.52	0.85	1.26	3.23	0.69	1.01
Tm	0.78	0.88	0.68	0.92	0.99	-1.78	0.55	0.75
Tx	0.68	0.82	0.59	0.90	1.53	-1.14	0.64	1.15
HR	0.39	0.63	0.00	0.73	10.48	9.09	1.00	8.07
Vv	0.49	0.70	-1.98	0.61	1.12	-30.93	1.67	1.00
RS	0.22	0.46	-0.13	0.59	219.95	-23.24	1.06	133.20
P	0.50	0.71	0.25	0.72	54.61	50.97	0.85	33.61
ETo	0.72	0.85	0.07	0.83	0.97	-15.09	0.90	0.80
				SSF				
Tn	0.58	0.76	0.35	0.83	1.45	3.93	0.80	1.18
Tm	0.84	0.92	0.77	0.95	0.80	-1.12	0.47	0.61
Tx	0.69	0.83	0.43	0.87	1.58	-2.74	0.75	1.23
HR	0.61	0.78	0.49	0.86	6.45	4.89	0.72	4.86
Vv	0.64	0.80	-2.48	0.68	1.00	-19.70	1.56	0.89
RS	0.28	0.53	0.02	0.66	193.64	-17.56	0.99	114.91
P	0.53	0.73	0.45	0.81	48.23	31.24	0.74	26.92
ETo	0.75	0.87	0.11	0.84	0.91	-9.63	0.90	0.74
				Total area				
Tn	0.60	0.77	0.44	0.84	1.35	3.58	0.75	1.09
Tm	0.81	0.90	0.73	0.94	0.90	-1.45	0.51	0.68
Tx	0.68	0.83	0.51	0.88	1.56	-1.94	0.69	1.19
HR	0.50	0.70	0.24	0.79	8.47	6.99	0.86	6.47
Vv	0.56	0.75	-2.23	0.65	1.06	-25.32	1.62	0.94
RS	0.25	0.49	-0.06	0.63	206.79	-20.40	1.03	124.05
P	0.51	0.72	0.35	0.77	51.42	41.10	0.80	30.27
ETo	0.74	0.86	0.09	0.83	0.94	-12.36	0.90	0.77

Source: The Authors.

Table 3.
Values of performance indicator c and classification for both mesoregions: Sertão Pernambucano (SP) and Sertão do São Francisco (SFF) and the total area

	SP		**SSF**		**Total area**	
	c	Category	c	Category	c	Category
Tn	0.67	Good	0.63	Medium	0.65	Medium
Tm	0.81	Very good	0.87	Very good	0.84	Very good
Tx	0.74	Good	0.72	Good	0.73	Good
HR	0.46	Bad	0.67	Good	0.56	Low
Vv	0.42	Bad	0.56	Low	0.49	Bad
RS	0.27	Awful	0.35	Awful	0.31	Awful
P	0.51	Bad	0.59	Low	0.55	Low
ETo	0.71	Good	0.72	Good	0.72	Good

Source: The Authors

The Tm presented the best performance with the classification of "very good" for both mesoregions, compared to the other variables. The Tx and ETo were classified as "good" for all cases. The results obtained are supported by [20] which indicated that temperature and ETo data from all reanalysis data sets are better than expected from rainfall, since this parameter is more spatially and temporally variable than temperature.

Therefore, the results obtained in this study are consistent with the indices obtained for P. Using the concordance index, Tm (0.94) and Tx (0.88) stood out in both mesoregions as having the highest performance, followed by Tn (0.84) and ETo (0.83). The lowest performances corresponded to Vv (0.65) and RS (0.63).

As a function of the PBIAS (see Table 2), the Tn and HR presented positive values (underestimation bias) in the two mesoregions. together with the P; with the difference that the P exceeded the PBIAS ≥ 25% its performance being considered "unsatisfactory" its performance for this indicator. Negative PBIAS values (overestimation) within the range qualified as "good performance" corresponded to the Tm, Tx y la ETo. Considering that the optimal value of PBIAS is 0.0. and "good performance" is between 10% < PBIAS < 15%. the closest variables were Tm, Tx, Tn, HR and ETo in ascending order. As a result, the worst performances (PBIAS ≥ 25%) were shown by the Vv and P. The standards were similar for both mesoregion and total area.

NSE obtained for variables from reanalysis datasets range from -2.48 to 0.77, with the NSE of Vv and RS being generally less than zero. NSE > 0.50 was obtained for Tn, Tm and Tx in the SP area and for Tm in SSF, being considered as "satisfactory" [17]. In the analysis for the total area, the Tm and Tx showed the best performances for this index. However, it should be noted that the same authors indicate that NSE values between 0 and 1 are generally considered "acceptable". Tn (0.44); HR (0.24); P (0.35) and ETo (0.09) are within this range. On the other hand, values of NSE<0.0, which are considered "unacceptable". were presented for Vv (-2.23) and RS (-0.06). The results from [2] concluded that CFSR data performed well in calculating TNx and TNn indices (maximum daily minimum temperature of each month and minimum monthly value of daily minimum

temperatures. respectively) and support the results presented.

Regarding the RSR, the Tm obtained < 0.70 "satisfactory" values in the two mesoregions, along with Tx and Tn for the mesoregion SP. All other variables exceeded the limit recommended by [17]. The worst performance according to RSR was for Vv and RS. Few studies have been conducted on wind variability using reanalysis data. [27] evaluated the performance of the *European Centre for Medium-Range Weather Forecasts* (ECMWF), obtaining the following results: $r=0.48$, $R^2=0.23$, RMSE=2.81 and for Petrolina $r=0.74$, $R^2=0.54$, RMSE=1.44. By comparing the MAE and RMSE for Tn, Tm and Tx (°C) it was observed that Tm presents lower error values than Tn and Tx in the two mesoregions. By comparing RMSE and MAE between both mesoregions, it is observed that RMSE and MAE are higher in SP for Tm, HR, Vv, RS, P and ETo. RMSE and MAE are higher for SFF only in the case of Tn and Tm.

Fig. 5 compares the box plots of the variables Tn, Tm, Tx, HR, RS, Vv and ETo by comparing the EM and CFSR datasets, which generally shows that the medium of Tn, Tm, Tx and ETo agree better with the observed data. In each box plot, hollow dots represent extreme values. The box plots presented allow us to compare the minimum, maximum, first quartile, second quartile or internal line (representing the median) and third quartile values. The dispersion of the data is represented by the interquartile interval, which is the difference between the third quartile and the first quartile. The proximity between both sets is observed, with the presence of outliers in the case of the EM data for Tn in Ouricuri, Cabrobó and Petrolina; Tx in Cabrobó; Vv and RS in all areas and for ETo in Ouricuri. In the case of reanalysis data, outliers were only observed for Tx in Ouricuri and Cabrobó, and HR in Ouricuri, but for RS, all areas presented outliers.
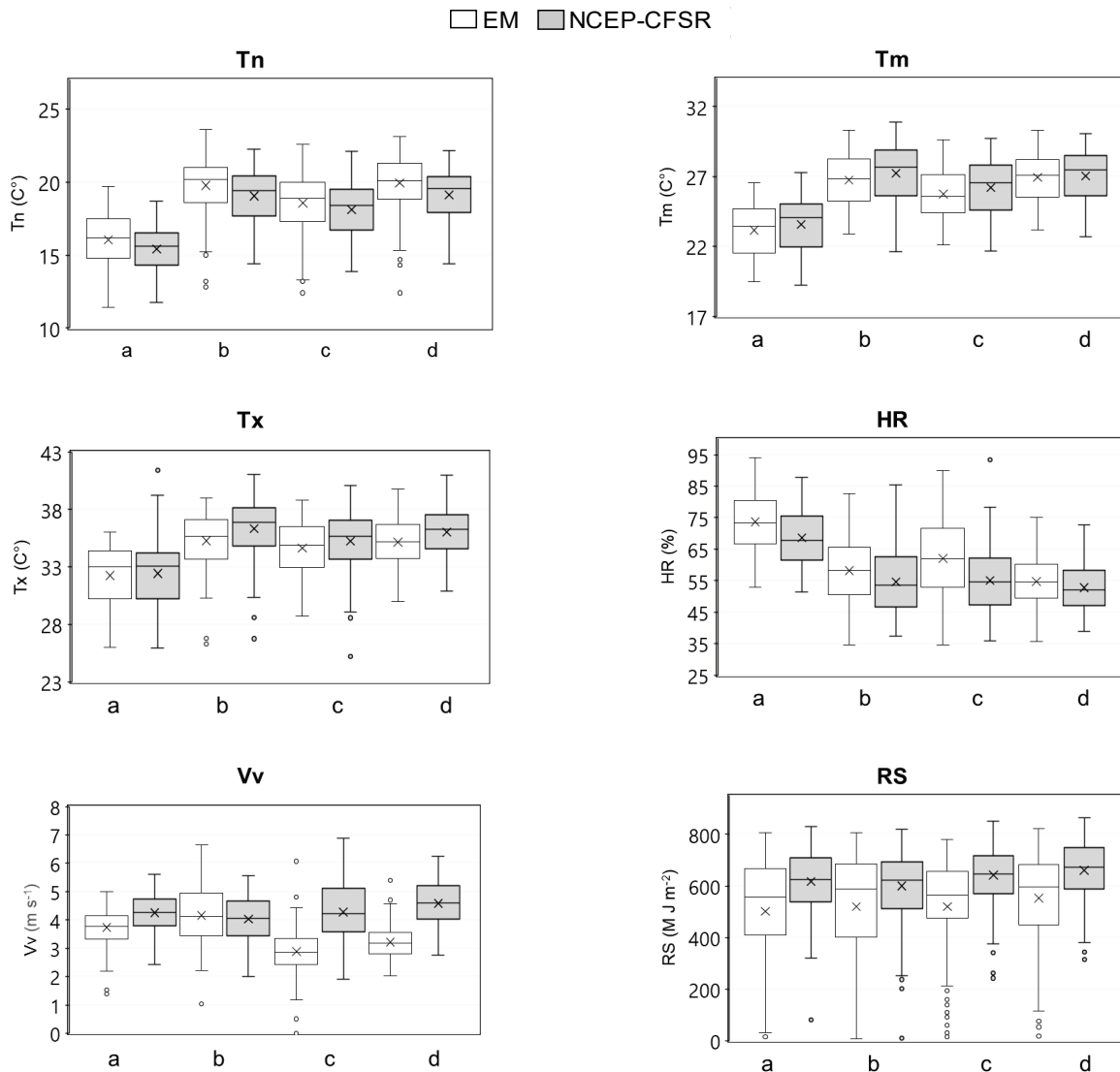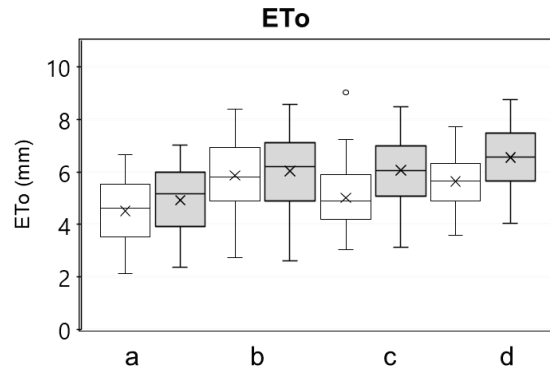
Figure 5. Box plots for the monthly time series (1979 – 2014) of Tn, Tx, Tm, Vv, HR, RS and ETo, of local station (EM) weather variables compared with NCEP-CFSR for a) Arcoverde, b) Cabrobó, c) Ouricuri and d) Petrolina
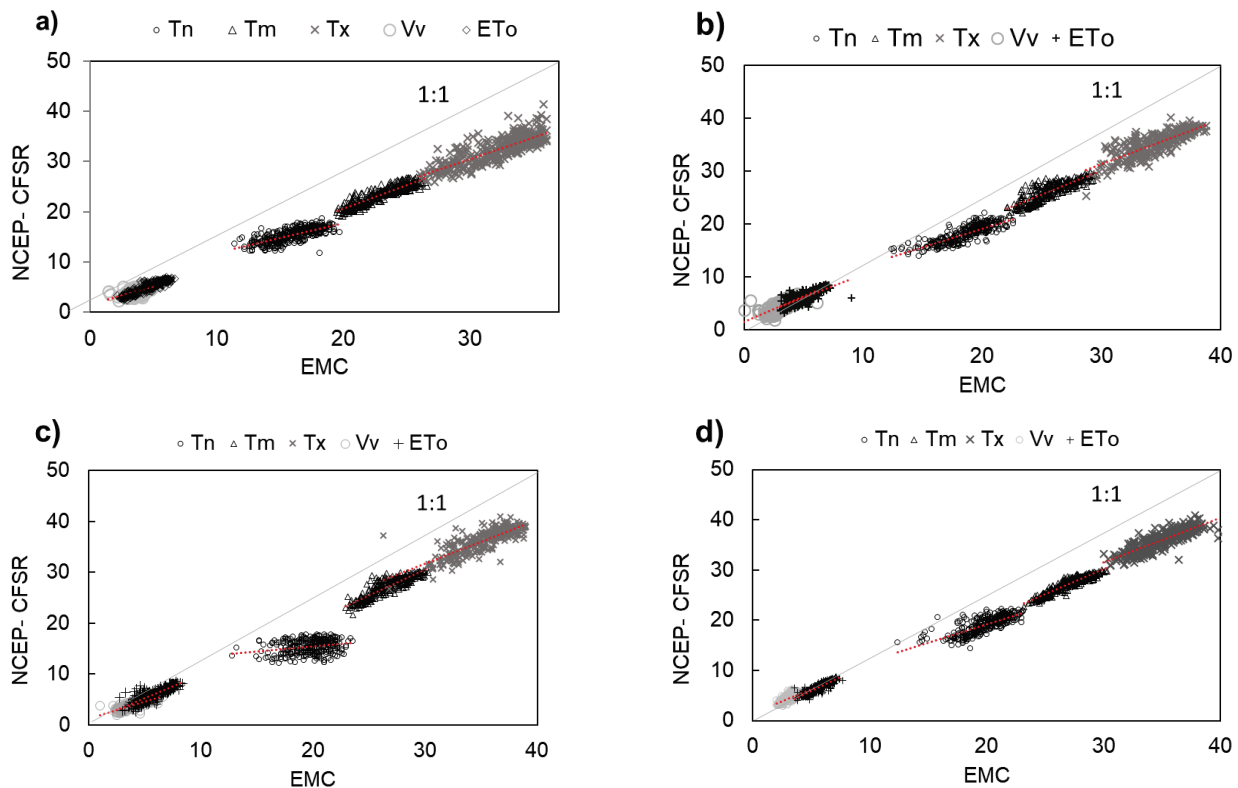Source: The Authors.

Figure 6. Correlation between NCEP-CFSR data and EM for Tn, Tm, Tx, Vv and ETo for a) Arcoverde; b) Ouricuri; c) Cabrobó. and d) Petrolina
Source: The Authors.

4. Conclusions

This study covers the period 1979 to 2014 and evaluates reanalysis meteorological data provided by CFSR-NCEP for two mesoregions: Sertão de Pernambuco (SP) and Sertão do Sao Francisco Pernambucano (SFF), of the semi-arid region of the State of Pernambuco, Brazil. Different statistical performance indicators were used in the evaluation and the main conclusions obtained are presented below:

The mean, minimum and maximum temperature (Tm, Tn, Tx) data provided by the CFSR-NCEP were the best performers according to the correlation and regression coefficients (r and R^2), the Willmott concordance index (d) and the Nash-Sutcliffe Efficiency index (NSE) in both mesoregions, compared to the other meteorological variables.

The potential evapotranspiration (ETo) also showed good performances for r, R^2 and d with the exception of NSE. Precipitation (P) presented acceptable performances for r, R^2, d and NSE, although lower than those obtained from temperature data, since precipitation is more spatially and temporally variable than temperature. The same applies to

relative humidity (HR), which shows acceptable performance.

No major differences were observed between the variables studied for both mesoregions using the root mean square error (RMSE), percent bias (PBIAS), standard deviation rate (RSR) and mean absolute error (MAE) indicators. CFSR-NCEP data do not represent the characteristics of solar radiation (RS) and wind speed (Vv) in the region, as both variables obtained the lowest performances for all indicators, except for R^2 and r in the SSF mesoregion.

Tn, Tm, Tx and ETo were the meteorological variables best represented by the CFSR-NCEP reanalysis data, followed by the P and HR.

This type of study contributes to the evaluation of the representativeness of reanalysis data in different mesoregions, and to the study and management of water resources in regions with low availability of meteorological data, such as the case of the Semi-Arid of Pernambuco. Future research to be developed includes the use of CFSR-NCEP reanalysis data in hydrological simulation models for basins in the region.

Acknowledgments

References

[1] Souza, R., Souza, E., Antonino, A. e Lima, J., Balanço hídrico em área de pastagem no semiárido pernambucano. Revista Brasileira de Engenharia Agrícola e Ambiental, 19(5), pp. 449-455, 2015. DOI: 10.1590/18071929/agriambi.v19n5p449-455

[2] De Lima, J. and; Alcântara, C., Comparison between ERA Interim/ECMWF. CFSR, NCEP/NCAR reanalysis, and observational datasets over the eastern part of the Brazilian Northeast Region. Theoretical and Applied Climatology, 138(3-4), pp. 2021-2041, 2019. DOI: 10.1007/s00704-019-02921-w.

[3] Dile, Y. and Srinivasan, R., Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin. JAWRA Journal of the American Water Resources Association, 50(5), pp. 1226-1241, 2014. DOI: 10.1111/jawr.12182

[4] Fuka, D. et al., Using the climate forecast system reanalysis as weather input data for watershed models. Hydrological Processes, 28(22), pp. 5613-5623, 2014. DOI: 10.1002/hyp.10073.

[5] Bressiani, D. et al., Effects of spatial and temporal weather data resolutions on streamflow modeling of a semi-arid basin. Northeast Brazil. International Journal of Agricultural and Biological Engineering. 8(3), pp. 125-139, 2015. DOI: 10.3965/j.ijabe.20150803.970

[6] Paz, Y. et al., Sensitivity analysis and calibration of the SWAT model for a basin in northeastern Brazil using observed and reanalysis climatic data. Revista Brasileira de Geografia Física. [Online]. 11(1), pp. 371-389, 2018. [date of reference March 12th of 2020]. Avaliable at: https://www.semanticscholar.org/paper/Sensitivity-analysis-and-calibration-of-the-SWAT-a-Paz-Galv%C3%ADncio/ea02995d908cd d91305c7b981307469cbfea4980

[7] Quadro, M. et al., The atmospheric water cycle over South America as seen in the new generation of global reanalyses. AIP Conference Proceedings, American Institute of Physics, 2013. pp. 732-735. DOI:

[8] 10.1063/1.4804874.

[8] Auerbach, D.A. et al., Evaluating weather observations and the climate forecast system reanalysis as inputs for hydrologic modelling in the tropics. Hydrological Processes, 30(19), pp. 3466-3477, 2016. DOI: 10.1002/hyp.10860.

[9] Lu, J.Z. et al., Assessing the climate forecast system reanalysis weather data driven hydrological model for the Yangtze River Basin in China. Applied Ecology and Environmental Research. 17(2), pp. 3615-3632, 2019. DOI: 10.15666/aeer/1702_36153632.

[10] Saha, S. et al., The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, 91(8), pp. 1015-1058, 2010. DOI: 10.1175/2010BAMS3001.1

[11] Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R., Large area hydrologic modeling and assessment part I: model development. Journal of the American Water Resources Association, 34(1), pp. 73-89, 1998. DOI: 10.1111/j.1752-1688.1998.tb05961.x.

[12] Roth, V.T. and Lemann. T., Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands. Hydrology and Earth System Sciences, 20(2), pp. 921-934, 2016. DOI: 10.5194/hess-20-921-2016.

[13] Silva, V.B., Kousky, V.E. and Higgins. R.W., Daily precipitation statistics for South America: an intercomparison between NCEP reanalyses and observations. Journal of Hydrometeorology, 12(1), pp. 101-117, 2011. DOI: 10.1175/2010JHM1303.1.

[14] Romagnoli, M. et al., Assessment of the SWAT model to simulate a watershed with limited available data in the Pampas region. Argentina. Science of the Total Environment, 596, pp. 437-450, 2017. DOI: 10.1016/j.scitotenv.2017.01.041.

[15] Rocha, N.C.V. and Mendes. A.C., Analysis of sensitivity and calibration of the SWAT model applied in a basin in northeast of Brazil using observed and reanalysis climatic data. Revista Brasileira de Geografia Física, 11(1), pp. 234-313, 2018. DOI: 10.26848/rbgf.v11.1.p234-313

[16] Harmel, R.D. et al., Evaluating. interpreting. and communicating performance of hydrologic/water quality models considering intended use: a review and recommendations. Environmental Modelling & Software, 57, pp. 40-51, 2014. DOI: 10.1016/j.envsoft.2014.02.013

[17] Moriasi, D.N. et al., Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), pp. 885-900, 2007. DOI: 10.13031/2013.23153.

[18] Hopkins, W.G., Correlation coefficient: a new view of statistics. [Online]. 2000. [data of reference March 18th of 2020]. Avaliable at: http://www.sportsci.org/resource/stats/

[19] Nash, J.E. and Sutcliffe, J.V., River flow forecasting through conceptual models' part I—A discussion of principles. Journal of hydrology, 10(3), pp. 282-290, 1970. DOI: 10.1016/0022-1694(70)90255-6Get rights and content.

[20] Chen, S. et al., Assessment of CFSR. ERA-Interim. JRA-55. MERRA-2. NCEP-2 reanalysis data for drought analysis over China. Climate Dynamics, 53(1-2), pp. 737-757, 2019. DOI: 10.1007/s00382-018-04611-1.

[21] Gassman, P.W. et al., The soil and water assessment tool: historical development, applications, and future research directions. Transactions of the ASABE, 50(4), pp. 1211-1250, 2007. DOI: 10.13031/2013.23637.

[22] Arnold, J.G. et al., SWAT: model use, calibration, and validation. Transactions of the ASABE, 55(4), pp. 1491-1508.,2012. DOI: 10.13031/2013.42256.

[23] Willmott, C.J., Some comments on the evaluation of model performance. Bulletin American Meteorological Society, 63(11), pp. 1309-1313, 198. DOI: 10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.

[24] Camargo, A.P. de e Sentelhas, P.C., Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo. Brasil. Revista Brasileira de Agrometeorologia, 5(1), pp. 89-97, 1997.

[25] Da Cunha, P.C.R. et al., Eficiência de métodos para o cálculo de coeficientes do tanque classe A na estimativa da evapotranspiração de referência. Embrapa Arroz e Feijão-Artigo em periódico indexado (ALICE), 43(2), 2013. DOI: 10.1590/S1983-40632013000200005.

[26] Legates, D.R. and Mccabe Jr, G.J., Evaluating the use of "goodness-of-fit" measures in hydrologic and hydro climatic model validation. Water Resources Research, 35(1), pp. 233-241, 1999. DOI: 10.1029/1998WR900018.

[27] Chaves Filho, J.B. et al., Variabilidade da velocidade do vento a 10 metros para a região Nordeste do Brasil. MSc. Dissertation (Mestrado em Metereologia) – Programa de Pós-Graduação em Meteorologia. Centro de Tecnologia e Recursos Naturais. Universidade Federal de Campina Grande, Paraíba, Brasil, 2016.

[28] Menezes, D., Moreira, L. e Kobiyama, M., É sempre adequado o uso dos dados de reanálise (CFSR)?. Estudo de caso na bacia do Rio Maquiné/RS. XXIII Brazilian Symposium on Water Resources, [online]. 2019. Available at: http://anais.abrh.org.br/works/5219

[29] Gadelha, C.T. et al., Comparação dos dados de precipitação observada versus reanálise CFSR: o caso da Bacia do Rio do Peixe, estado da Paraíba. XXIII Brazilian Symposium on Water Resources, [online]. 2019. Available at: http://anais.abrh.org.br/works/5231

[30] Blacutt, L. et al., Precipitation comparison for the CFSR, MERRA, TRMM3B42 and Combined Scheme datasets in Bolivia. Atmospheric Research, 163, pp. 117-131, 2015. DOI: 10.1016/j.atmosres.2015.02.002

[31] Worqlul, A.W. et al., Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia, Hidrology and Earth System Sciences, 18(12), pp. 4871-4881, 2014. DOI:10.5194/hess-18-4871-2014

M.D. Mollericona Alfaro, is BSc. Eng. in Agronomist Engineering in 2016, from the Mayor de San Andrés University, Bolivia, currently pursuing for a Master's degree in Agricultural Engineering at the Federal Rural University of Pernambuco, Brazil.
ORCID: 0000-0001-8243-781X

I. Lopes, is BSc. Eng. in Agricultural and Environmental Engineering in 2014, and with a MSc. in Agricultural Engineering in 2016, all of them from the Federal do Vale do São Francisco University, Brazil, currently pursuing a PhD in Agricultural Engineering at the Federal Rural University of Pernambuco, Brazil. Professor of Agricultural Engineering at the Federal Institute of Education, Science and Technology of Bahia, Bom Jesus da Lapa campus, Bahia-Brazil.
ORCID: 0000-0003-0592-4774

A.A. de Assunção Montenegro, is a BSc. Eng in Civil Engineering (1985), from the Federal University of Pernambuco, MSc. in Hydraulics and Sanitation (1989), from the São Carlos, University of São Paulo and a PhD in Water Resources, (1997) from the University of Newcastle Upon Tyne. He has a post-doctoral degree in hydrological modeling (2008), from the Center for Ecology and Hydrology in Wallingford, England. Full professor at the Federal Rural University of Pernambuco and permanent member of the Postgraduate Program in Agricultural Engineering at UFRPE. Experience in the area of water resources and agricultural and environmental engineering, with emphasis on irrigation and drainage, mainly in the topics: semi-arid, watershed management, experimental and numerical hydrology, hydraulics, soil moisture dynamics, environmental services, sanitation and reuse of lower quality water in agriculture, water and soil conservation, geostatistics, and salinity. He was General Research Coordinator at UFRPE between 2014 and 2016.
ORCID: 0000-0002-5746-8574

B. Gonçalves Leal, works at the Federal University of Vale do São Francisco, Brazil, as a professor. He has a BSc. in Physics, a MSc. in Agronomy (Applied Meteorology) and a Dr. in Agricultural Engineering, all of them from the Federal University of Viçosa, Brazil. He has experience in the field of software development for management of irrigation in agricultural systems, acting in modelling, simulation, irrigation and climatology.
ORCID: 0000-0003-4449-6995