Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Evaluation of air temperature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil	
Article Sub-Title		
Article CopyRight	The Author(s), under exclusive licence to Springer Nature Switzerland AG (This will be the copyright line in the final PDF)	
Journal Name	Environmental Monitoring and Assessment	

Corresponding Author	FamilyName	**Almeida**
	Particle	de
	Given Name	**Ceres Duarte Guedes Cabral**
	Suffix	
	Division	Dom Agostinho Ikas Agricultural School
	Organization	Federal Rural University of Pernambuco
	Address	São Lourenço da Mata, Pernambuco, Brazil
	Phone	
	Fax	
	Email	ceres.cabral@ufrpe.br
	URL	
	ORCID	http://orcid.org/0000-0001-6073-3853
Author	FamilyName	**Araújo**
	Particle	de
	Given Name	**Caio Sérgio Pereira**
	Suffix	
	Division	Department of Agricultural Engineering
	Organization	Federal Rural University of Pernambuco
	Address	Recife, Brazil
	Phone	
	Fax	
	Email	caiosergio.ufersa@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-6747-7028
Author	FamilyName	**Silva**
	Particle	e
	Given Name	**Ivis Andrei Campos**
	Suffix	
	Division	Department of Agricultural Engineering
	Organization	Federal Rural University of Pernambuco
	Address	Recife, Brazil
	Phone	
	Fax	
	Email	ivisandrei@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-9555-4238
Author	FamilyName	**Ippolito**
	Particle	
	Given Name	**Matteo**
	Suffix	
	Division	Department Agricultural, Food and Forest Sciences (SAAF)

	Organization	University of Palermo
	Address	Palermo, Italy
	Phone	
	Fax	
	Email	matteo.ippolito@unipa.it
	URL	
	ORCID	http://orcid.org/0000-0001-6919-5000

Schedule	Received	10 Oct 2021
	Revised	
	Accepted	9 Apr 2022

Abstract

This study aimed to analyze the average air temperature data estimated by ERA5-Land reanalysis over 10 years (01/01/2011 to 12/31/2020), based on data from 12 automatic weather stations located in different cities in the state of Pernambuco, northeast of Brazil. For more careful evaluation, the average air temperature data were stratified by mesoregions of the state. ERA5-Land reanalysis data were validated using statistical indices that evaluated the accuracy. The average air temperature estimated by ERA5-Land reanalysis agree well with weather stations in almost the entire state of Pernambuco. The highest accuracy of the average air temperature estimated by ERA5-Land occurred in the city of Ibimirim ($R2 = 0.98$), while the lowest accuracy was measured in the city of Caruaru ($R2 = 0.57$). Root-mean-square error (RMSE) generated by the ERA5-Land reanalysis was lower than 0.60 °C in most Pernambuco. The highest RMSE of ERA5-Land for average air temperature was calculated using data from the city of Caruaru (1.11 °C), whereas the lowest RMSE was obtained with data from the city of Recife (0.41 °C). According to mean bias errors, (MBE) the ERA5-Land underestimated the average air temperature data compared to automatic weather stations data, especially in Ouricuri, which had the highest underestimation (−0.80 °C). On the other hand, among the municipalities where ERA5-Land overestimated the temperature values, the highest overestimation was identified in Garanhuns (0.35 °C). Therefore, based on the results of this study, ERA5-Land reanalysis successfully estimated the average air temperature for the state of Pernambuco.

Keywords (separated by '-') Climatology - ECMWF - Automatic weather stations - Semiarid

Footnote Information

Environ Monit Assess _####################_
https://doi.org/10.1007/s10661-022-10047-2

Check for updates

Evaluation of air temperature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil

Caio Sérgio Pereira de Araújo · **Ivis Andrei Campos e Silva** · **Matteo Ippolito** · **Ceres Duarte Guedes Cabral de Almeida**

Received: 10 October 2021 / Accepted: 9 April 2022

Abstract This study aimed to analyze the average air temperature data estimated by ERA5-Land reanalysis over 10 years (01/01/2011 to 12/31/2020), based on data from 12 automatic weather stations located in different cities in the state of Pernambuco, northeast of Brazil. For more careful evaluation, the average air temperature data were stratified by mesoregions of the state. ERA5-Land reanalysis data were validated using statistical indices that evaluated the accuracy. The average air temperature estimated by ERA5-Land reanalysis agree well with weather stations in almost the entire state of Pernambuco. The highest accuracy of the average air temperature estimated by ERA5-Land occurred in the city of Ibimirim ($R^2 = 0.98$), while the lowest accuracy was measured in the city of Caruaru ($R^2 = 0.57$). Root-mean-square error (RMSE) generated by the ERA5-Land reanalysis was lower than 0.60 °C in most Pernambuco. The highest RMSE of ERA5-Land for average air temperature was calculated using data from the city of Caruaru (1.11 °C), whereas the lowest RMSE was obtained with data from the city of Recife (0.41 °C). According to mean bias errors, (MBE) the ERA5-Land underestimated the average air temperature data compared to automatic weather stations data, especially in Ouricuri, which had the highest underestimation (−0.80 °C). On the other hand, among the municipalities where ERA5-Land overestimated the temperature values, the highest overestimation was identified in Garanhuns (0.35 °C). Therefore, based on the results of this study, ERA5-Land reanalysis successfully estimated the average air temperature for the state of Pernambuco.

Keywords Climatology · ECMWF · Automatic weather stations · Semiarid

C. S. P. de Araújo · I. A. C. e Silva
Department of Agricultural Engineering, Federal Rural University of Pernambuco, Recife, Brazil
e-mail: caiosergio.ufersa@gmail.com

I. A. C. e Silva
e-mail: ivisandrei@gmail.com

M. Ippolito
Department Agricultural, Food and Forest Sciences (SAAF), University of Palermo, Palermo, Italy
e-mail: matteo.ippolito@unipa.it

C. D. G. C. de Almeida (✉)
Dom Agostinho Ikas Agricultural School, Federal Rural University of Pernambuco, São Lourenço da Mata, Pernambuco, Brazil
e-mail: ceres.cabral@ufrpe.br

Introduction AQ1

The use of meteorological data is characterized as an important information database for society in general with applications in different sectors (de Oliveira Castro AQ2 et al., 2019). An example of this is the study carried out by de Gois et al. (2019) in which the annual and monthly AQ3 variability of rainfall and air temperature were evaluated, as well as quantified the components of the climatic water balance in the municipality of Resende, state of Rio de Janeiro, in 22 years of meteorological data. In addition, Ventura et al. (2016) with computational support applied four methodologies of genetic algorithms derived from artificial intelligence and neural networks to be used to fill gaps in climatological databases. Another example is the study carried out by dos Santos et al. (2017) in which used the weather stations databases from the National Institute of Meteorology (INMET) and the National Institute of Space Research (INPE), both in Brazil, to investigate the relationship between the variables land use and occupation, rainfall, maximum temperature, and relative air humidity in southwestern of the state of Piaui in the 1984–2015 period, supported by images of Landsat 5 and Landsat 8 satellites. Franco et al. (2019) constructed a rainfall zoning for cacao cultivation in the state of Bahia, Brazil, using the fuzzy logic based on historical series of 519 meteorological stations.

According to da Silva et al. (2017), the atmosphere has been suffering from high temperatures in recent decades, with a significant increase in this variable from the 1980s due to global warming. Automatic weather stations are considered one of the main tools for monitoring and assessing environmental conditions of a given locality, consisting mainly of providing measurements of parameters that indicate the instantaneous, or average, state of the atmosphere (Lopes et al., 2021). However, the standardization of observation schedule, the weather station location, preventive maintenance operations, calibration procedures and measuring instruments testing, and the cost for maintaining these weather stations are some of the reasons that restrict their expansion in Brazil (de Oliveira Castro et al., 2019). The quality of this dataset provides assessing the consequences of natural resource management, such as the soil-crop-water relationship in an irrigated area.

To date, there are about 786 automatic and conventional weather stations in Brazil belonging to the INMET network, and 219 of which are in the northeast region (146 automatic and 73 conventional). The state of Pernambuco has 13 automatic weather stations and seven conventional weather stations that provide hourly and monthly data of numerous climatic variables (INMET, 2021). Weather stations have fast data processing of various climate variables that are generated by electronic sensors and analog measurement equipment. However, these stations have some limitations in data acquisition, including reading errors, caused by equipment failures in automatic weather stations and human observation errors in conventional weather stations (Bier & Ferraz, 2017; Lopes et al., 2021). Consequently, the user should consider the impact of these occurrences on historical series, which are more evident in daily data and may mask the variability of climatic data, resulting in unreliable data. Therefore, the identification and remotion of the discrepancies of the historical series are essential before running the analysis, through homogenization procedures, which are extremely important for studies of climatic variability (Azorin-Molina et al., 2016).

In recent years, the use of climatological data from global atmospheric models has become an important alternative for assessment and environmental monitoring conditions. This database provides measurements of several variables that can help minimize the scarcity of observations and historical data series in surface weather stations. Thus, the reanalysis technique of climatological data is an effective alternative for many scientific and operational studies in recent decades (de Oliveira Castro et al., 2019; Moraes et al., 2012).

Reanalysis is obtained through a process that involves data assimilation, mainly based on data observed in global weather stations and on numerical models of forecasting aiming to estimate the weather conditions with greater consistency (Parker, 2016). Consequently, reanalysis is among the most used data sets in climatological and environmental studies at different scales for the most distinct purposes (de Oliveira Castro et al., 2019). However, this data set suffers interference from climate and local topography due to the spatial resolution, which is considered a disadvantage for reanalysis (Chen et al., 2021).

The Tiled European Centre for Medium-Range Weather Forecasts (ECMWF) Scheme for Surface Exchanges over Land (TESSEL) is used operationally in the Integrated Forecast System (IFS) for explaining the development of soil, vegetation, and snow at various spatial resolutions over Earth. A revised land surface hydrology (HTESSEL) is introduced in the ECMWF operational model, to report failings of the land surface scheme, specifically the lack of surface runoff and the choice of a global uniform soil texture (Hersbach et al., 2020). ERA5 produced by the Copernicus Climate Change Service

Journal : Medium 10661	Article No : 10047	Pages : 13	MS Code : 10047	Dispatch : 13-4-2022

(C3S) is the fifth generation ECMWF atmospheric reanalysis of the global climate covering the period from January 1950 to the present, with spatial resolution 0.25°×0.25° and the high temporal scan hourly available also aggregate at monthly scale.

ERA5-Land released in July 2019 is an improved version of ERA5 that was created by forcing the terrestrial component of ERA5 reanalysis, covering the same period. The ERA5-Land product is delivered at the same time resolution as ERA5 but with a spatial resolution of 0.1°×0.1°against 0.25°×0.25° of ERA5 (C3S, 2021) and without data available for the oceans. Thus, the reanalysis technique has become a monitoring system designed to estimate many atmospheric and land climate variables.

The air temperature estimated by ERA5-Land reanalysis has been widely studied. Liu et al. (2021) analyzed the spatial–temporal consistency of Earth surface temperature trends derived from satellites (AIRS and MODIS) and ERA5-Land reanalysis. Cao et al. (2020) used the ERA5-Land reanalysis database to assess the accuracy of soil temperature in permafrost regions. Graham et al. (2019) evaluated five different types of atmospheric reanalysis in an Arctic portal and concluded that all products well estimated air temperature, but ERA5 reanalysis provided best performance.

The most common climatic variable measured is air temperature due to being easy to obtain and the importance for hydrology and agricultural applications, besides for research in meteorology and climatology and the society in general (Pohlmann & Lazzari, 2018). Air temperature data from the atmospheric reanalysis model, such as the ECMWF, can be used to overcome the data absence of historical series from surface weather stations. Thus, the objective of this study was to analyze the accuracy of the monthly air temperature data estimated by ERA5-Land reanalysis over 10 years (01/01/2011 to 12/31/2020) in the state of Pernambuco, northeastern Brazil, based on the data set observed in automatic weather stations. The state of Pernambuco is an interesting case study region for the evaluation of the accuracy of the air temperature dataset estimated by ERA5-Land reanalysis since there are three mesoregions with different climate units; as a result, this study could provide valorous information to ERA5-Land application in other regions.

Material and methods

The study was conducted in the state of Pernambuco (8°20′24″ S, 37°48′36″ W), located in northeastern Brazil (Fig. 1a) in a tropical zone, which occupies an area of approximately 98,146 km^2 and has three different climatic classes (Tropical monsoon; Tropical savannah and Arid), according to Köppen's classification (Alvares et al., 2013). This region stands out in agricultural production, especially in the sugarcane cultivation and, in recent decades, with the continuous increase in the irrigated area cultivated with tropical fruit, for example, grape, mango, guava, and banana. The data set used as a reference to validation of ERA5-Land was provided from automatic weather stations (AWS) distributed throughout the state of Pernambuco, in different mesoregions. These stations are situated at the municipalities of Arcoverde, Cabrobó, Caruaru, Floresta, Garanhuns, Ibimirim, Ouricuri, Petrolina, Serra Talhada, Palmares, Recife, and Surubim, totaling 12 municipalities (Fig. 1b). The data set from the weather station located in the municipality of Salgueiro was not used due to the absence of data in the period defined for this study. In fact, the absence of climatic data was identified at various periods and in several municipalities in the state, with different intensities. Thus, the use of reanalysis data can be an alternative to address this lack of public information. Hourly air temperature data (°C) were acquired through automatic weather stations belonging to the network of the National Institute of Meteorology (INMET), covering 10 years (01/01/2011 to 12/31/2020).

The second database of air temperature adopted in this study was provided by ERA5-Land atmospheric reanalysis produced by the European Center for Medium-Range Weather Forecast (ECMWF) through virtual stations corresponding to the same cities and the same period as automatic weather stations, on a monthly scale with a spatial resolution of 0.10° and measured in Kelvin (Fig. 1c). The monthly data of air temperature were converted from Kelvin to degrees Celsius to unify both databases and enable the global atmospheric model validation. The geographic information of the weather station is presented in Table 1, in addition, to the percentage of missing data in each AWS.

AQ7

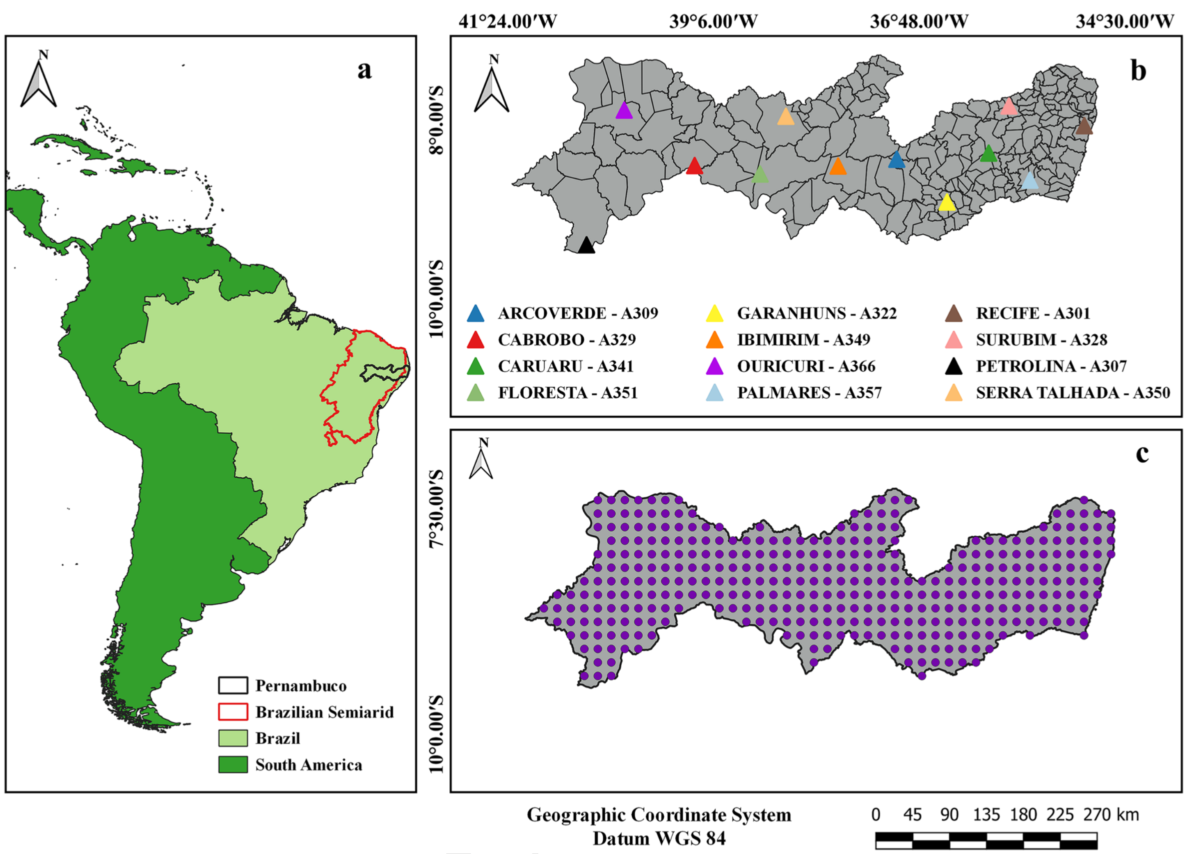

Fig. 1 Location of the Pernambuco state in northeastern Brazil (**a**); Spatial distribution of automatic weather stations belonging to National Institute of Meteorology (INMET) network in the state of Pernambuco, identified by the city and the respective World Meteorological Organization, WMO code (**b**); Spatial distribution of ECMWF virtual stations (**c**)

The absence of data observed in some periods and locations in the historical series analyzed required the application of a gap-filling methodology known as the simple average method. For this purpose, hourly measurements of automatic weather stations from other geographically close locations, which had similar climatic characteristics, were used to obtain the arithmetic mean and estimate the missing value according to respective hour (Kashani & Dinpashoh, 2012). The monthly average air temperature was calculated from the hourly data, using the arithmetic mean for each month of the year.

The performance of ERA5-Land reanalysis in estimating air temperature data was evaluated using statistical indices of accuracy, adopting the data from automatic weather stations as standard series. For more careful evaluation, the data set of average air temperature were compared by city and grouped by climatic mesoregions. Thus, the root-mean-square error (RMSE) (Eq. (1)) and the mean bias error (MBE) (Eq. (2)) were calculated. In addition, the graphical analysis of the data dispersion around the 1:1 line was performed, and the corresponding coefficient of determination (R^2) was obtained (Eq. (3)), which indicates the data set accuracy degree, ranging from 0 to 1; the closer to 1, the better the fit of the evaluated model. The temperature observed is that measured by the electronic sensor at the automatic weather station and the estimated temperature is that derived from the ERA5-Land reanalysis.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(T_{obsi} - T_{esti}\right)^2}{N}} \quad (1)$$

Table 1 Geographical characteristics of the automatic weather stations used in this study

Station code[a]	City	Elevation (m)	Mesoregion	Climatic classification[b]	Latitude (S)	Longitude (W)	Missing observations (%)
A307	Petrolina	372.5	Semiarid	BSh	9.39°	40.52°	2.74
A309	Arcoverde	683.9	Semiarid	Aw	8.43°	37.05°	3.66
A329	Cabrobó	342.8	Semiarid	BSh	8.50°	39.31°	4.20
A349	Ibimirim	434.2	Semiarid	Aw	8.51°	37.71°	2.52
A350	Serra Talhada	499.0	Semiarid	BSh	7.95°	38.29°	3.18
A351	Floresta	327.4	Semiarid	BSh	8.60°	38.60°	4.20
A366	Ouricuri	462.0	Semiarid	BSh	7.88°	40.10°	11.66
A322	Garanhuns	827.8	Agreste	Aw	8.91°	36.49°	5.17
A328	Surubim	421.4	Agreste	Aw	8.84°	35.80°	6.67
A341	Caruaru	852.0	Agreste	Aw	8.36°	36.03°	7.52
A301	Recife	11.3	Forest/Coastal	Am	8.06°	34.96°	2.04
A357	Palmares	164.0	Forest/Coastal	Aw	8.67°	35.57°	9.30

Am tropical, monsoon, *Aw*, tropical, savannah, *BSh*, arid, desert, hot

[a]World Meteorological Organization (WMO)

[b]Alvares et al. (2013)

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(T_{esti} - T_{obsi}\right) \quad (2)$$

$$R^2 = \left[1 - \frac{\left(T_{obsi} - T_{esti}\right)^2}{\left(T_{obsi} - \overline{T}\right)^2}\right] \quad (3)$$

where T_{obsi} represents the monthly observed temperature (°C); T_{esti} is the monthly estimate temperature (°C); and N is the number of the observations available for analysis ($i = 1, 2, \ldots, N$); $\overline{T}$ is the observed average values (°C).

RMSE expresses error data the magnitude of the error between estimated values by ERA5-Land and measured/observed, ranging from zero to ∞, and the closer to zero, the better the model (Janssen & Heuberger, 1995). The mean bias error (MBE) index expresses the mean deviation of the estimated values relative to the observed values, indicating the model tends to overestimate or underestimate; thus, the closer to zero, the lower the magnitude of the model systematic error (Leite & de Andrade, 2002).

Statistical analyses were performed by the R software (R CoreTeam, 2020), which uses a multi-paradigm programming language to provide, manipulate, analyze, and visualize data.

Results

The average monthly air temperature was calculated from the hourly data observed in each automatic weather station, during the period from 2011 to 2020 (Fig. 2). The average monthly air temperature amplitude in the Semiarid mesoregion ranged from 20.6 to 29.2 °C in the municipalities of Arcoverde in July and Cabrobó in November, respectively. While in the Agreste mesoregion, the average air temperature values ranged from 19.0 to 25.6 °C in Garanhuns in August and Surubim in March, respectively. Finally, in the Forest/Coastal mesoregion, the average monthly air temperature ranged from 22.6 to 26.9 °C, in Palmares in July and Recife in December, respectively.

The average monthly air temperature in the state of Pernambuco varied according to the geographical location of the municipalities (longitude), distance from water bodies (rivers and lakes), and elevation. These data show that there has been a slight decrease in the monthly air temperatures in the municipalities located at higher elevations, whereas municipalities in the Semiarid mesoregion have high average monthly air temperatures.

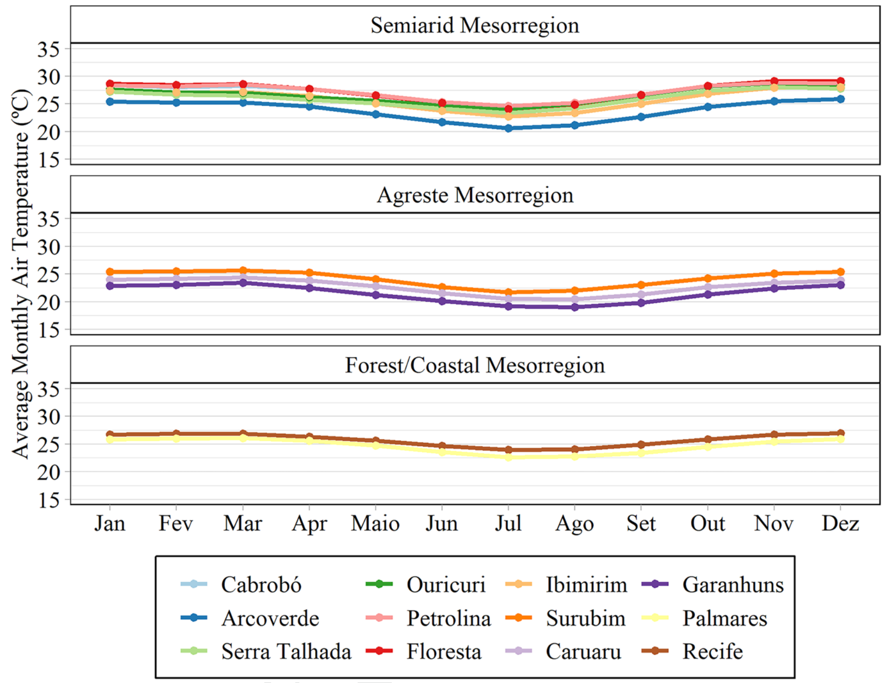

Fig. 2 Monthly averages of air temperature (°C) measured by automatic weather stations in the state of Pernambuco, from 2011 to 2020

The average monthly air temperature values estimated by ERA5-Land, in general, were overestimated, whatever of the mesoregion in the state of Pernambuco. Figure 3 shows the error (E) caused by the difference between the average monthly air temperature estimated in the ERA5-Land reanalysis

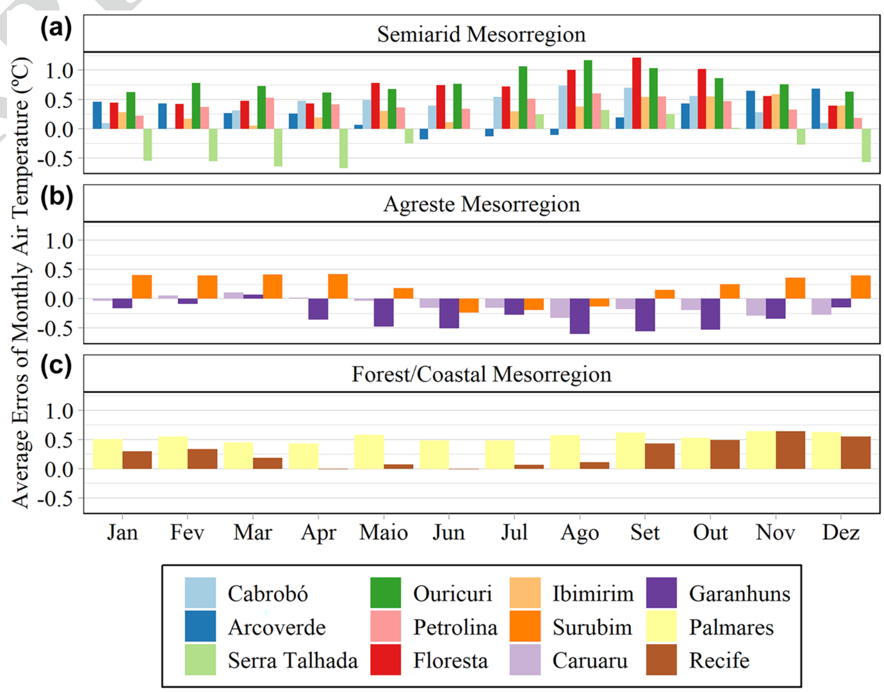

Fig. 3 Average errors of monthly air temperature (°C) estimated by ERA5-Land subdivided by mesoregions of Pernambuco: **a** Semiarid, **b** Agreste, and **c** Forest/Coastal

Table 2 Performance statistical indices of ERA5-Land reanalysis in the estimation of air temperature compared to the automatic weather stations in the state of Pernambuco, for the period from 2011 to 2020

Statistical indices	Municipalities					
	Ouricuri	Floresta	Palmares	Petrolina	Cabrobó	Ibimirim
R^2	0.92	0.96	0.96	0.96	0.96	0.98
RMSE (°C)	0.96	0.80	0.60	0.53	0.54	0.43
MBE (°C)	−0.80	−0.67	−0.53	−0.40	−0.37	−0.32
	Arcoverde	Caruaru	Garanhuns	Serra Talhada	Recife	Surubim
R^2	0.94	0.57	0.93	0.93	0.93	0.94
RMSE (°C)	0.58	1.11	0.57	0.57	0.41	0.47
MBE (°C)	−0.23	0.13	0.35	0.22	−0.27	−0.22

RMSE root-mean-square error, *MBE* mean bias error

and the average monthly air temperature calculated using the data recorded by automatic weather stations over 10 years. The largest underestimation and overestimation generated by the ERA5-Land reanalysis in the state of Pernambuco occurred in the Semiarid mesoregion, in the municipalities of Serra Talhada (−0.68 °C) in April and Floresta (1.22 °C) in September, respectively (Fig. 3a). In the Agreste mesoregion (Fig. 3b), the largest underestimation of ERA5-Land was −0.60 °C, in August, in Garanhuns, while the highest overestimation occurred in the municipality of Surubim, in April (0.42 °C). In the Forest/Coastal mesoregion, ERA5-Land data overestimated monthly air temperature by up to 0.64 °C in the cities of Recife and Palmares in November (Fig. 3c).

ERA5-Land successfully simulates the average air temperature throughout the state of Pernambuco, according to performance statistical indices (Table 2). In general, the highest values of R^2 correspond to the lowest values of RMSE, and the opposite is equally true. The estimate of average air temperature by ERA5-Land reanalysis obtained accuracy higher than 90% in almost the entire state of Pernambuco, especially for the city of Ibimirim ($R^2 = 0.98$) located in the Semiarid mesoregion (Table 2) while the lowest accuracy of ERA5-Land was identified in Caruaru ($R^2 = 0.57$) in the Agreste mesoregion.

According to RMSE values, lower than 0.60 °C in most of the state of Pernambuco and with a minimum value of 0.41 °C (Table 2), the ERA5-Land reanalysis presents suitability to estimated monthly air temperature. Again, the city of Caruaru (Agreste mesoregion) resulted in the largest error (RMSE = 1.11 °C) followed by Floresta and Ouricuri (0.80 °C and 0.96 °C, respectively), both municipalities located in the Semiarid mesoregion.

The mean bias error (MBE) indicates that ERA5-Land reanalysis generated underestimated data of average air temperature compared to the data observed in the AWS in the Semiarid and Forest/Coastal mesoregions of Pernambuco (Table 2). This result is confirmed in the data graphical analysis (Fig. 4), which shows that of the 12 automatic weather stations used in this study, ERA5-Land underestimated the average air temperature values in nine of them, corresponding to 75% of the stations. The highest underestimation occurred in Ouricuri (−0.80 °C), in the Semiarid region of Pernambuco, while the highest overestimation of ERA5-Land occurred in Garanhuns (0.35 °C), in the Agreste mesoregion.

In the graphical analysis of the municipality of Caruaru (Fig. 4), it can be noticed that there was a change in the relationship between data observed in the AWS and those estimated by the ERA5-Land, overestimating in a certain period of monitored time and underestimating in another period. This finding encouraged the monthly analysis of each year, over the 10 years, and thus, it was identified that this change in trend occurred from July/2018 (Fig. 5, dashed line). According to information obtained from INMET, this AWS was initially installed in the central part of the city, with 570 m (−08.23 S, −35.98 W); however, at 07/18/2018, this AWS was transferred to an area with preserved vegetation with 852 m (−08.35S, −36.02 W) a little far from the city. This place transfer absolutely justifies the change mentioned above, since lower temperature data started to be recorded, distancing a little further from those estimated by ERA-5-Land, underestimating around 2 °C from July 2018.

The means of R^2, RMSE, and MBE for all stations corresponded to 0.92, 0.63 °C, and −0.26 °C, respectively, indicating the quality of fit and agreement

Journal : Medium 10661 | Article No : 10047 | Pages : 13 | MS Code : 10047 | Dispatch : 13-4-2022

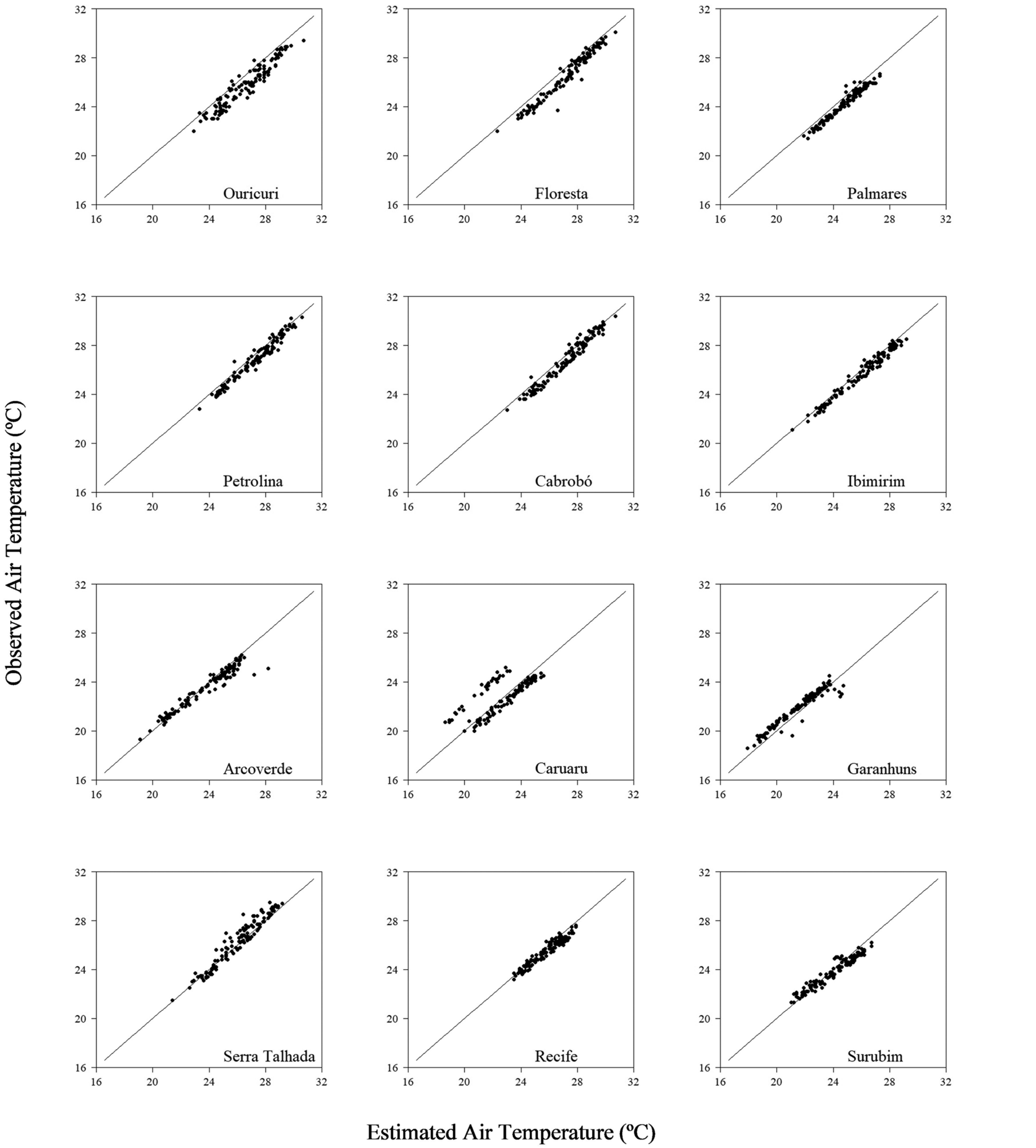

Fig. 4 Monthly averages of air temperature measured by automatic weather stations (observed) and estimated by ERA5-Land in municipalities in the state of Pernambuco, in the period from 2011 to 2020

Journal : **Medium 10661**	Article No : **10047**	Pages : **13**	MS Code : **10047**	Dispatch : **13-4-2022**

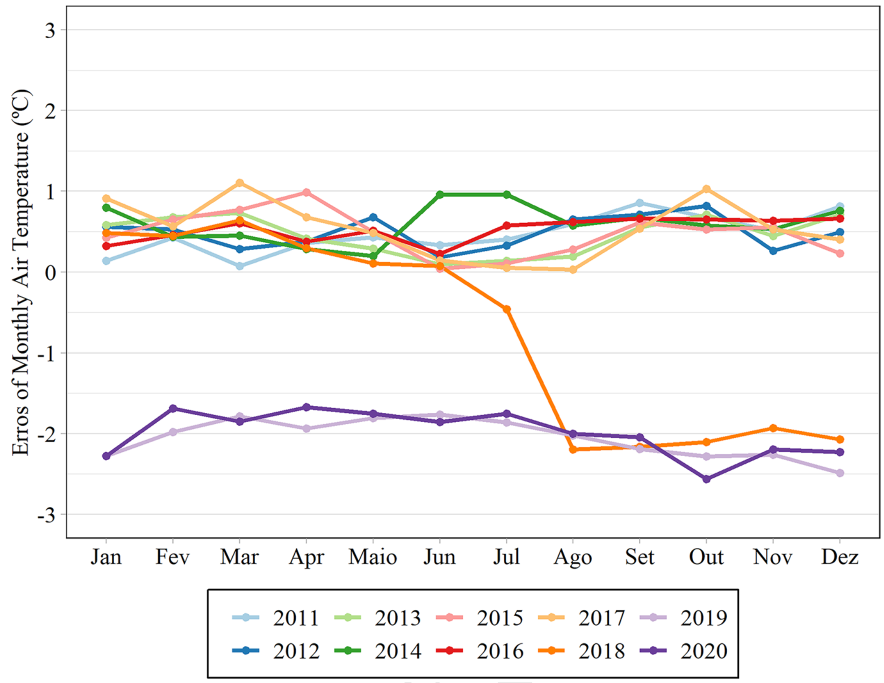

Fig. 5 Errors of monthly air temperature (°C) estimated by ERA5-Land at the municipality of Caruaru, state of Pernambuco over the 10 years

between ERA5-Land reanalysis and most automatic weather stations. The association between elevation and temperature estimation errors is identified (Fig. 6). Specifically, Fig. 6 shows the digital terrain model (DTM) of Pernambuco, as well as the RMSE (Fig. 6a) and MBE (Fig. 6b) values associated to air temperature, computed in each weather station. In terms of RMSE and MBE, the best performance was obtained in weathers station located in proximity of the coastal region, with RMSE minimum and MBE close to zero. On the other hand, both statistical indicators are worse for inland weather stations, probably due to the orographic characteristics of the terrain.

Discussion

This study analyzed the air temperature estimated by ERA5-Land based on data from 12 automatic weather stations, belonging to the National Institute of Meteorology (INMET) network, in the state of Pernambuco (98,146 km^2), northeast of Brazil. However, the density of this network is relatively low comparing to earlier studies, such as Valeriano et al. (2019) analyzing data from 10 stations in southeastern Brazil (10,014 km^2), Moraes et al. (2012) studying simulated decennial data of precipitation, maximum temperature, and minimum air in São Paulo (248,209 km^2) with 33 weather stations, Pelosi et al. (2020) with data from 18 stations in the Campania region (14,000 km^2), in Italy, and Paredes et al. (2018) with data from 24 stations in Portugal (92,200 km^2) and 42 agrometeorological stations distributed over Sicily (25,711 km^2), Italy (Negm et al., 2017). Therefore, one of the limitations is the low availability of specialized meteorological data, measured on the surface, which becomes one of the main motivations for increasing the application of reanalysis data and this study seeks to obtain data that will help to address these research gaps.

The ERA5-Land showed, in general, good performance in estimating the average air temperature throughout the state of Pernambuco. According to Pelosi et al. (2020), this performance can be explained by the fact that air temperature is one of the variables predicted by mathematical modeling since it can be directly estimated by the numerical integration of the laws describing atmosphere dynamics. Therefore, it can be better predicted than diagnostic variables, which are determined from observations in surface weather stations.

The RMSE values ranged from 0.41 to 1.11 °C, which was considered slightly lower than those found in the previous studies. Valeriano et al. (2019), studying

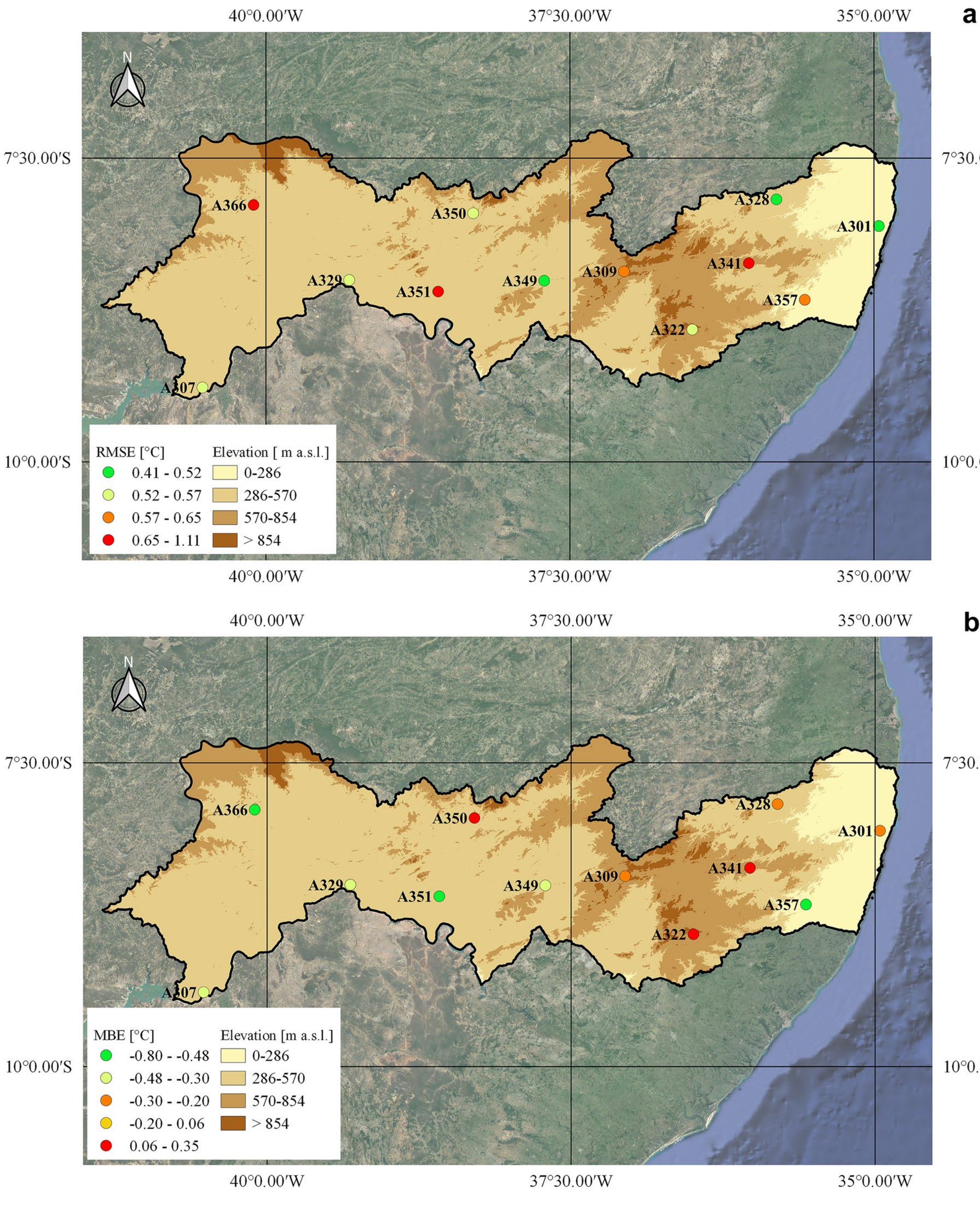

Fig. 6 Digital terrain model (DTM) of Pernambuco with the location of weather stations used to evaluate air temperature with the range of RMSE (**a**) and MBE (**b**) values associated with air temperature

different global atmospheric models in southeastern Brazil, observed RMSE of up to 4.59 °C. Aboelkhair et al. (2019), in a study conducted in India, using the NASA POWER reanalysis, found good performance to estimate the average air temperature, with RMSE ranging between 0.96 and 2.54 °C. Aparecido et al. (2019) reported that the air temperature estimated by ERA-Interim proved to be accurate and feasible in the replacement of surface data in the state of Paraná, Brazil, with RMSE ranging from 0.37 °C in summer to 1.11 °C in spring. These results are similar to those reported by Moraes et al. (2012) that found RMSE below 1 °C for almost the entire state of São Paulo to the minimum temperature and RMSE values lower than 4.9 °C for the maximum temperature variable, corroborating with this work the feasibility of using ECMWF global model data for air temperature.

The accuracy of the average air temperature estimated by ERA5-Land was higher than 90% in a great portion of the Pernambuco state. This result corroborates the findings of a great deal of the previous work. This is exemplified in the work undertaken by Aparecido et al. (2019), in a study in the state of Paraná, Brazil, using ERA-Interim reanalysis, which also found accuracy higher than 90% for the average air temperature throughout the state. Another example of what is meant by (Martins et al., 2017), in a study in the Iberian Peninsula, observed accuracy of data from reanalysis higher than 80% for the maximum air temperature; however, for the minimum temperature, the accuracy was lower. On the other hand, Silva et al. (2020), in a study also in the state of Pernambuco, observed ERA5-Land accuracy for the average air temperature higher than 79% with data only of 2019. As well as Monteiro et al. (2018) reported good accuracy in the maximum, minimum, and average temperature data from the NASA POWER reanalysis for Brazil, with R^2 of 0.72, 0.75, and 0.73, respectively.

Overall, these results indicate that ERA5-Land underestimated the average air temperature values in nine of the 12 stations used, which corresponds to 75%. This finding was also reported by Aboelkhair et al. (2019) with NASA POWER reanalysis data, which underestimated the average air temperature in 45% of the stations used in India. In this context, Paredes et al. (2018) also identified trends of underestimation of maximum and minimum air temperatures, using ERA-Interim.

Conclusions

The main goal of the current study was to determine the accuracy of the air temperature estimated by ERA5-Land reanalysis in the state of Pernambuco, northeast of Brazil, based on the automatic weather stations data set. Overall, the results, according to the statistical indices analyzed, indicate that the ERA5-Land reanalysis successfully simulated the average air temperature of the state of Pernambuco.

For more careful evaluation, the average air temperature database was stratified in three climatic mesoregions of the state of Pernambuco (Forest/Coastal; Agreste and Semiarid). Thus, ERA5-Land reanalysis obtained greater accuracy of the average air temperature in the Semiarid, followed by the Forest/Coastal and Agreste. The average air temperature data were slightly underestimated in the Semiarid and Forest/Coastal mesoregions and slightly overestimated in the Agreste.

This approach will prove useful in expanding our understanding of how the use of climatological data from global atmospheric models has become an important alternative for understanding and monitoring environmental conditions. In general, therefore, it seems that the average air temperature database by ERA5-Land reanalysis can replace the respective data from surface weather stations, in case of missing or scarce observations. However, ERA5-Land still needs improvement, especially in climatic regions similar to the Agreste mesoregion of Pernambuco, which showed the lowest coefficient of determination and the highest RMSE.

Funding The study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES), Finance Code 001 and the Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE IBPG-0433–5.03/20).

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aboelkhair, H., Morsy, M., & El Afandi, G. (2019). Assessment of agroclimatology NASA POWER reanalysis datasets for temperature types and relative humidity at 2 m against ground observations over Egypt. *Advances in Space Research, 64*(1), 129–142. https://doi.org/10.1016/j.asr.2019.03.032

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Aparecido, L. E. D. O., Rolim, G. D. S., Moraes, J. R. D. S. C., Torsoni, G. B., Meneses, K. C. D., & Costa, C. T. S. (2019). Accuracy of ECMWF ERA-interim reanalysis and its application in the estimation of the water deficieny in paraná, Brazil. *Revista Brasileira De Meteorologia, 34*(4), 515–528. https://doi.org/10.1590/0102-7786344066

Azorin-Molina, C., Guijarro, J.-A., McVicar, T. R., Vicente-Serrano, S. M., Chen, D., Jerez, S., & Espírito-Santo, F. (2016). Trends of daily peak wind gusts in Spain and Portugal, 1961–2014. *Journal of Geophysical Research: Biogeosciences, 121*, 2652–2666. https://doi.org/10.1002/2015JD024485.Received

Bier, A. A., & Ferraz, S. E. T. (2017). Comparação de Metodologias de Preenchimento de Falhas em Dados Meteorológicos para Estações no Sul do Brasil. *Revista Brasileira De Meteorologia, 32*(2), 215–226. https://doi.org/10.1590/0102-77863220008

Cao, B., Gruber, S., Zheng, D., & Li, X. (2020). The ERA5-Land soil temperature bias in permafrost regions. *The Cryosphere, 14*, 2581–2595. https://doi.org/10.5194/tc-14-2581-2020

Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., & Khadka, N. (2021). Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. *Atmospheric Research, 250*, 105365. https://doi.org/10.1016/j.atmosres.2020.105365

da Silva, R. O. B., Montenegro, S. M. G. L., & de Souza, W. M. (2017). Tendências de mudanças climáticas na precipitação pluviométrica nas bacias hidrográficas do estado de pernambuco. *Engenharia Sanitaria e Ambiental, 22*(3), 579–589. https://doi.org/10.1590/S1413-41522017142481

de Gois, G., De Freitas, W. K., de Bodas Terassi, P. M., De Oliveira-Júnior, J. F., & Portz, A. (2019). Variabilidade Anual e Mensal da Chuva e da Temperatura do ar no Município de Resende, Rio de Janeiro. *Revista Brasileira De Climatologia, 24*, 67–88. https://doi.org/10.5380/abclima.v24i0.61611

de Oliveira Castro, B. V., Silva, F. P., & Pires, G. D. (2019). Avaliação de parâmetros meteorológicos gerados pela reanálise ERA 5 para a cidade de Seropédica, região da Baixada Fluminense-RJ. *Revista Engenharia, Meio Ambiente e Inovação, 2*(1), 7–16.

dos Santos, J. C., dos Santos Lisboa, G., de Jesus França, L. C., Stepka, T. F., da Silva, J. B. L., de Miranda, D. L. C., & Cerqueira, C. L. (2017). Relação entre variáveis meteorológicas e o uso e ocupação do solo no Sudoeste do Piauí, Brasil. *Nativa, 5*(6), 414–420. https://doi.org/10.5935/2318-7670.v05n06a06

Franco, L. B., De Almeida, C. D. G. C., Freire, M. M., Franco, G. B., & Silva, S. D. A. (2019). Rainfall zoning for cocoa growing in Bahia state (Brazil) using fuzzy logic land suitability; Cacao is a species of great economic and social importance. Expanding the area grown with this crop has been limited by. *Engenharia Agricola, 4430*, 48–55.

Graham, R., Hudson, S. R., & Maturilli, M. (2019). Improved performance of ERA5 in Arctic gatway relative to four global atmospheric reanalyses. *Geophysical Research Letters, 46*, 6138–6147. https://doi.org/10.1029/2019GL082781

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Janssen, P. H. M., & Heuberger, P. S. C. (1995). Calibration of process-oriented models. *Ecological Modelling, 83*(1–2), 55–66. https://doi.org/10.1016/0304-3800(95)00084-9

Kashani, M. H., & Dinpashoh, Y. (2012). Evaluation of efficiency of different estimation methods for missing climatological data. *Stochastic Environmental Research and Risk Assessment, 26*(1), 59–71. https://doi.org/10.1007/s00477-011-0536-y

Leite, H. G., & de Andrade, V. C. L. (2002). Um Método Para Condução De Inventários Florestais Sem O Uso De Equações Volumétricas a Method for Conducting Forest Inventories Without Using Volumetric Equations. *Revista Árvore, 26*(3), 321–328. http://www.scielo.br/pdf/rarv/v26n3/a07v26n3.pdf AQ9

Liu, J., Hagan, D. F. T., & Liu, Y. (2021). Global land surface temperature change (2003–2017) and its relationship with climate drivers: AIRS, MODISM and ERA5-Land based analysis. *Remote Sensing, 13*(1), 44. https://doi.org/10.3390/rs13010044

Lopes, I., Guimarães, M. J. M., de Melo, J. M. M., de Almeida, C. D. G. C., Lopes, B., & Leal, B. G. (2021). Comparison of meteorological data, related to reference evapotranspiration, from conventional and automatic stations in the sertão and agreste regions of pernambuco, brazil. *DYNA (Colombia), 88*(216), 176–183. https://doi.org/10.15446/dyna.v88n216.86372

Martins, D. S., Paredes, P., Raziei, T., Pires, C., Cadima, J., & Pereira, L. S. (2017). Assessing reference evapotranspiration estimation from reanalysis weather products. An application to the Iberian Peninsula. *International Journal of Climatology, 37*(5), 2378–2397. https://doi.org/10.1002/joc.4852

Monteiro, L. A., Sentelhas, P. C., & Pedra, G. U. (2018). Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation. *International Journal of Climatology, 38*(3), 1571–1581. https://doi.org/10.1002/joc.5282

Moraes, R. A., Rocha, J. V., de Souza Rolim, G., Lamparelli, R. A. C., & Martins, M. M. (2012). Avaliação dos dados decendiais de precipitação e temperatura máxima e mínima do ar simulados pelo modelo ECMWF para o estado de São Paulo. *Irriga, 17*(3), 397–407.

Negm, A., Jabro, J., & Provenzano, G. (2017). Assessing the suitability of American National Aeronautics and Space Administration (NASA) agro-climatology archive to predict daily meteorological variables and reference evapotranspiration in Sicily, Italy. *Agricultural and Forest Meteorology, 244–245*(May), 111–121. https://doi.org/10.1016/j.agrformet.2017.05.022

Paredes, P., Martins, D. S., Pereira, L. S., Cadima, J., & Pires, C. (2018). Accuracy of daily estimation of grass reference evapotranspiration using ERA-Interim reanalysis products

Journal : Medium 10661	Article No : 10047	Pages : 13	MS Code : 10047	Dispatch : 13-4-2022

with assessment of alternative bias correction schemes. *Agricultural Water Management, 210*(May), 340–353. https://doi.org/10.1016/j.agwat.2018.08.003

Parker, W. S. (2016). Reanalyses and observations: What's the Difference? *Bulletin of the American Meteorological Society, 97*(9), 1565–1572. https://doi.org/10.1175/BAMS-D-14-00226.1

Pelosi, A., Terribile, F., D'Urso, G., & Chirico, G. B. (2020). Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water (switzerland)*. https://doi.org/10.3390/W12061669

Pohlmann, V., & Lazzari, M. (2018). Climatic characterization of Cachoeira do Sul (Rio Grande do Sul) and adjacent municipalities regarding air temperature. *Revista Eletrônica Científica Da UERGS, 4*(5), 643–650.

Silva, L. M. C., Ippolito, M., Silva, J. L. B., Araújo, C. S. P., Provenzano, G., & de Almeida, C. D. G. C. (2020). Validation of air temperature data obtained from the ERA5- Land reanalysis in the state of Pernambuco, Brazil. *Inovagri International Meeting*, 148–162.

Valeriano, T. T. B., de Souza Rolim, G., Bispo, R. C., da Silva Cabral de Moraes, J. R., & Aparecido, L. E. D. O. (2019). Evaluation of air temperature and rainfall from ECMWF and NASA gridded data for southeastern Brazil. *Theoretical and Applied Climatology, 137*(3–4), 1925–1938. https://doi.org/10.1007/s00704-018-2706-z

Ventura, T. M., Santana, L. L. R., Martins, C. A., & de Figueiredo, J. M. (2016). Análise Da Aplicabilidade De Métodos Estatísticos Para Preenchimento De Falhas Em Dados Meteorológicos. *Revista Brasileira De Climatologia, 12*(19), 168–177.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal:	**10661**
Article:	**10047**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	LE: Please check if the affiliations are presented correctly.	
AQ2	The citation 'Castro et al., 2019' has been changed to 'de Oliveira Castro et al., 2019' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ3	The citation 'Gois et al. (2019)' has been changed to 'de Gois et al. (2019)' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ4	The citation 'Santos et al. (2017)' has been changed to 'dos Santos et al. (2017)' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ5	The citation 'Silva et al. (2017)' has been changed to 'da Silva et al. (2017)' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ6	References INMET, 2021, C3S, 2021, and R CoreTeam, 2020 were mentioned in the manuscript; however, these were not included in the reference list. As a rule, all mentioned references should be present in the reference list. Please provide the reference details to be inserted in the reference list.	
AQ7	The citation 'Alvares et al. (2014)' found in Table 1 footnote has been changed to 'Alvares et al. (2013)' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ8	The citation 'Leite and Andrade, 2002' has been changed to 'Leite and de Andrade, 2002' to match the author name/date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.	
AQ9	Kindly check Reference Leite and de Andrade (2002) if presented correctly and please provide access date as well.	