Climate classification for Northeast Brazil using reanalysis data and the Absolute Aridity Index

Eliton Sancler Gomes Sales¹; Wendy Kaori Matsunaga²; Célia Campos Braga³; Meiry Sayuri Sakamoto⁴; Daisy Beserra Lucena⁵; José Ivaldo Barbosa de Brito⁶; Josefina Moraes Arraut⁷

¹Doctor in Meteorology, Academic Unit of Atmospheric Sciences, Federal University of Campina Grande (UFCG) Campina Grande - Paraíba, Brazil, elitonsancler@hotmail.com; ²Doctoral student in Meteorology, Academic Unit of Atmospheric Sciences, Federal University of Campina Grande (UFCG) Campina Grande - Paraíba, Brazil, wmatsunaga@gmail.com; ³Professor at the Academic Unit of Atmospheric Sciences, Federal University of Campina Grande (UFCG) Campina Grande - Paraíba, Brazil, celia@dca.ufcg.edu.br; ⁴Researcher at the Ceará Meteorology and Water Resources Foundation (FUNCME), Brazil, meiry@funceme.br; ⁵Professor at the Geosciences Department of the Federal University of Paraíba (UFPB), Brazil, daisylucena@yahoo.com.br; ⁶Professor, Academic Unit of Atmospheric Sciences, Universidade Federal de Campina Grande (UFCG) Campina Grande - Paraíba, Brazil, josefinaarraut@gmail.com.

Artigo recebido em 14/03/2024 e aceito em 06/04/2024

R E S U M O
O Clima da Região Nordeste do Brasil (NEB) tem sido intensamente estudado e analisado para a classificação climática, onde tem sido utilizado o índice de aridez do Programa das Nações Unidas para o Meio Ambiente (PNUMA) (IAUNEP), porém, sem resultados totalmente satisfatórios. As variáveis de entrada necessárias para seu cálculo são a precipitação e evapotranspiração potencial de referência (ET0). Todavia, embora as estações pluviométricas estejam bem distribuídas no NEB nos registros das medições rotineiras de precipitação, estas estações não fornecem as variáveis necessárias para a estimativa de ET0, então para o seu cálculo é utilizado a interpolação, mas isso pode gerar erros. Outra abordagem objetiva para a classificação do clima é o método de Thornthwaite, baseado na determinação do índice de umidade (Im), cujo cálculo também requer dados de estações meteorológicas. Assim, buscando solucionar a escassez de estações e melhorar a distribuição espacial das informações sobre as variáveis meteorológicas no NEB, o trabalho teve como um dos objetivos validar os dados da reanálise ERA5-Land do European Centre for Medium-range Weather Forecast (ECMWF) e a análise baseada em medidores unificados do projeto de precipitação diária global do Center for Climate Prediction/National Oceanic and Atmospheric Administration (CPC/NOAA). Após a validação, foram desenvolvidas classificações climáticas para o NEB usando os índices IAUNEP e Im. Observou-se que a classificação climática de Thornthwaite superestimou a aridez do NEB, enquanto o IAUNEP tendeu a subestimá-la. Por esta razão, um novo índice de classificação climática foi sugerido, denominado índice de aridez absoluto (IaA), apresentando resultados satisfatórios. Palavras-Chave: Terras Secas, Semiárido, Climas Úmidos, ERA5-Land.

Classificação do Clima para o Nordeste do Brasil utilizando dados de Reanálises e Índice de Aridez Absoluto

A B S T R A C T
The climate of the Northeast Region of Brazil (NEB) has been intensively studied and analyzed for climate classification. The aridity index of the United Nations Environment Programme (UNEP) (AIUNEP) has been used for this purpose but without satisfactory results. The input variables needed for its calculation are precipitation and reference potential evapotranspiration (ET0). However, although rainfall stations recording routine measurements of precipitation are well distributed in the NEB, they do not provide the necessary variables for estimating ET0. Therefore, an interpolation is used to calculate ET0, but this can lead to errors. Another objective approach to climate classification is the Thornthwaite method, which is based on the determination of the humidity index (Im), which also requires data from weather stations. To overcome the problem of the lack of weather stations and to improve the spatial distribution of information on meteorological variables in the NEB, one of the objectives of the present work was to validate the ERA5 reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF) and the Climate Prediction Center/National Oceanic and Atmospheric Administration (CPC/NOAA) global daily precipitation uniform gage-based analysis project. After validation, climate classifications for the NEB were developed using AIUNEP and Im. It was found that the Thornthwaite climate classification overestimated drought in the NEB, while the IAUNEP and IaA. It was observed that the consequently, a new climate classification index, called the absolute aridity index (IaA), was suggested, yielding satisfactory results. Keywords: dry lands; semiarid; humid climates; ERA5-Land.
Introduction

The Northeast region of Brazil (NEB) is located between latitudes 1° S and 18° S, which could be a factor of great influence favoring an adequate distribution and volume of rainfall. Yet, the climatic conditions throughout its territory are characterized by low rainfall rates, high spatio-temporal variability in rainfall distribution, high potential evapotranspiration rates arising from the high number of sunshine hours (approximately 2700 hours/year), high mean annual air temperatures (22-30°C), and prevalence of the semiarid climate over a large area, corresponding to 65% of the NEB, which contributes to climatic vulnerability (Martins et al., 2019; Comin et al., 2020; Da Silva et al., 2020; Marques et al., 2020; Jardim et al., 2021; Costa et al., 2021). It is worth noting that in addition to these climatic characteristics, interannual climatic variability is further enhanced in the semiarid area of the NEB by the strong variability in precipitation, with years of severe droughts and others of abundant rains (Silva, 2004; Hastenrath, 2012; Campos, 2015; Silva et al., 2020).

Due to the local climatic conditions, a significant part of north-eastern Brazil is known as the Brazilian Semi-arid region (SAB). It covers an area of 1.128 million km² (13.2% of the total area) and it is home to a population of 27 million people (13.3% of the Brazilian population) living in 2,262 municipalities in ten states: Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe and Bahia, as well as the northern region of the state of Minas Gerais (Brazil, 2017).

The Brazilian Semi-Arid is a region delimited by the Superintendência de Desenvolvimento do Nordeste (SUDENE), taking into account the prevailing climatic conditions of the semi-desert, in particular, the irregular rainfall. The criteria adopted by SUDENE are (1) average annual rainfall equal to or less than 800 mm; (2) aridity index from the United Nations Environment Programme (UNEP) of up to 0.5, calculated using the water balance, which relates precipitation and potential evapotranspiration; (3) daily percentage of water deficit equal to or more, taking into account all days of the year (Brasil, 2017; IBGE, 2021).

Many studies have sought to understand the causes of the high interannual climatic variability in the NEB. It was observed that this variability was related to that of the El Niño-Southern Oscillation (ENSO) and to sea surface temperature (SST) anomalies in the Tropical Atlantic Ocean (Hastenrath and Heller, 1977; Moura and Shukla, 1981; Andreoli and Kayano, 2006; Alves et al., 2009; Marengo et al., 2017; Lyra et al., 2017; Erfanian et al., 2017; Correia Filho et al., 2019; Costa et al., 2021), as well as to Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO) (Kayano & Capistrano, 2014; Kayano et al., 2019; Figliuolo et al., 2020).

The climatic characteristics of the NEB, namely, the long dry periods, the arid and semiarid climate, and the difficult management of water resources, are not decisive but contribute to the social vulnerability and the worst Human Development indices in the country (Sadeghi et al., 2016; Lyra et al., 2017; Marengo et al., 2018; Feindouno et al., 2020).

In this context, the Report of the Brazilian Panel on Climate Change states that the high evaporation rates and large year-to-year variability in runoff cause a significant oscillation in surface water availability in the NEB, creating a trend towards greater degree of aridity and intense interannual variability in the future climate of this region (Ribeiro & Santos, 2016). This raises concerns about the social situation of the population. Similar results were described in the fifth report of the Intergovernmental Panel on Climate Change (IPCC); the volume of rainfall in Northeast Brazil is expected to decrease by about 20% in 2100, but the IPCC is not categorical due to the low confidence levels generated by the current discrepancies between climate models for much of Brazil (IPCC, 2014).

According to Martins et al. (2017), the drought that started in 2012 in the NEB and lasted until 2017 was the period that represented the most critical quadrennium in terms of annual totals. It is emphasized that Martins et al. (2017) only had data available until 2017, but the drought continued until 2018 (ANA, 2018). Since 1911, there have been two droughts of three years' duration (1930-32; 1941-43), two of four years' duration (1951-54 and 2012-2015) and one of five years' duration (1979-83). As a result, the effects of drought have worsened during this period.

According to Santana and Santos (2020), the prolonged drought that affected the Northeast region between 2012 and 2018 had a significant impact on the performance of agricultural activities, especially in the semi-arid part. There was a sharp drop in production (and therefore productivity) of most crops typical of family economies.
farming. In some municipalities, negative deviations of more than 90% were observed at the beginning of the drought despite maintenance and alterations in part of the productive activities. It is worth mentioning that Santana and Santos (2020) analyzed the impact of the drought in the period from 2012 to 2017. However, the drought continued until 2018 (ANA, 2018).

Large-scale disasters have been recorded as a result of climate change and instability all over the planet, including changes in water resources and temperature, heavy rains, and intense droughts that affect agriculture over the years (Marengo & Bernasconi, 2014; Campos, 2015; Soares et al., 2021). Recently, Marengo et al. (2020) analyzed the projections of vegetative stress conditions based on the Vegetation Health Index (VHI) and indicated that semi-desert and arid conditions will replace the Caatinga in 2100 and the NEB may be one of the Brazilian regions most impacted by climatic variations. Considering the current socioeconomic scenario, it is anticipated that vulnerable rural populations living in the semiarid region will be more intensely affected, as it was the case in the last extensive drought in the region (2012-2018), the most extreme in the last 50 years. This event caused several problems to local populations, such as loss of crops and animals, reduced income, among others (Marengo et al., 2017; Marengo et al., 2020; Pontes Filho et al., 2020; Brito et al., 2021).

Despite the real need for knowledge on the climatology of the NEB, the absence of long-term, high quality and flawless meteorological observations and the low density of weather stations pose an obstacle to this type of studies (De Pauw et al., 2000). To compensate for the lack of spatio-temporal data, other meteorological data sources have been developed and constantly used, such as satellite-based data, global and regional numerical forecast models, and atmospheric reanalysis, whose potential has already been explored in several studies (Pelosi et al., 2016; Negm et al., 2017; Chirico et al., 2018; Medina et al., 2018; Jiang et al., 2019; Gleixner et al., 2020; Longo-Minollo et al., 2020; Vanella et al., 2020; McNicholl et al., 2021; Minnolo et al., 2022; Wu et al., 2022).

Matsunaga et al. (2023) compared precipitation data from CPC/NOAA with those from meteorological stations in Bahia, affirming that CPC/NOAA data represent station observations well. Sales et al. (2023) validated ERA5-Land and CPC/NOAA reanalysis data with meteorological station data to perform climate classification of the Northeast region of Brazil. The results showed satisfactory statistical outcomes and adequately represented the observed data, obtaining spatial classifications according to the physical and climatic characteristics of the region concluding that ERA5-Land and CPC/NOAA precipitation data are reliable and can be used in the absence of observed or doubtful data, for climatic and environmental studies and analyses in the NEB.

Atmospheric reanalysis has attracted growing interest in the last decade due to its potential to provide comprehensive information and consistent time series (Tarek et al., 2020). The ERA5 – the fifth-generation reanalysis product of the European Center for Medium-range Weather Forecast (ECMWF) – is one of the most used reanalysis dataset. ERA5 assimilates a broad range of measured and remote sensing atmospheric and oceanic information within a physical-dynamic environment of a coupled numerical model (Poli et al., 2016). One of the main advantages of using reanalysis is that the data do not depend directly on the density of terrestrial observational networks, offering the possibility to obtain variables in areas with little and/or no surface coverage, in addition to being an efficient data source for studies aimed at the planning and design of management of water resources and energy (Tarek et al., 2020; Pelosi et al., 2020; Ruiz et al., 2021; Wu et al., 2022).

Thus, ERA5-LAND reanalysis data represent a powerful tool for climatic studies in Brazil, offering a wide range of high-resolution information on meteorological and climatic variables. These datasets provide a solid foundation for detailed analyses and modeling of climatic processes at regional and local scales. Authors such as Silva et al. (2023), Santos et al. (2023), Oliveira et al. (2023), and Costa et al. (2023) have demonstrated the potential of these data to investigate specific climatic phenomena such as precipitation patterns, extreme events, and seasonal changes in Brazil. Additionally, recent publications by authors such as Pereira et al. (2024) and Lima et al. (2024) suggest that the continuous use of these reanalysis data is essential for advancing our understanding of regional climate patterns and their impacts.

The global atmospheric model is used in a data assimilation system in which information from various meteorological sources in the world, such as weather radars and satellites, is gathered. ECMWF data, for example, including information on several meteorological variables, are provided by the meteorological bank of the Joint Research...
Another widely used dataset is the one from the CPC unified gauge-based analysis of global daily precipitation project, which is ongoing at CPC/NOAA. It should be noted that this dataset has a consistent quantity and improved quality, combining all information sources available at CPC/NOAA, covering the entire globe with horizontal resolution of 0.5° generated through objective interpolation analysis techniques (Chen et al., 2008).

One of the greater challenges in studies of the climate of the NEB is the delimitation of arid and semiarid areas. They are areas with a high degree of aridity located between two humid regions: the Amazon Forest to the west and the Atlantic Forest to the east (Sobral-Souza et al., 2015; Castro et al., 2019). Brazilian government agencies have used a set of criteria to demarcate the boundaries of arid and semiarid lands (Jesus et al., 2019; Jesus, 2021; Oliveira and Castro, 2021). The aridity index of the United Nations Environment Programme (UNEP) (AI_{UNEP}) is one of these criteria. The input variables needed for calculation of the AI_{UNEP} are precipitation and reference potential evapotranspiration (ET\textsubscript{0}). However, although rainfall stations that record routine measurements of precipitation data are well distributed in the NEB, weather stations that provide the set of variables necessary for the estimation of ET\textsubscript{0} are scarce. When weather stations are absent at a given location, the ET\textsubscript{0} value is interpolated, which can lead to errors. Another objective approach of climate classification is the Thornthwaite method, which is based on the determination of the moisture index (I\textsubscript{m}), whose calculation also requires weather stations data. Therefore, reanalysis information represents a plausible alternative to overcome the scarcity of weather station data. However, the reliability of reanalysis information depends on its validation.

Thus, this work aims to validate ERA5-Land and the CPC/NOAA reanalysis data with weather stations data and then provide a climate classification for the NEB using the effective moisture index proposed by Thornthwaite (1948) and the UNEP aridity index (Middleton & Thomas, 1992; 1997), as well as propose the use of a new index called “absolute aridity index”. The ET\textsubscript{0} values used for the calculation of these indices were estimated by the full-form Penman-Monteith-FAO mathematical model developed by Allen et al. (1998) and recommended by the Food and Agriculture Organization (FAO).

It is important to highlight that Xavier et al. (2022) conducted a fruitful work of reanalysis data of maximum and minimum air temperature, precipitation, solar radiation, relative humidity and wind speed at 2 m height over the entire territory of Brazil for the period from 1961 to 2020 using a grid of 0.1° x 0.1°, constituting an excellent dataset for climate and agroclimatic studies. However, ERA5-Land and CPC reanalyses are updated in near real time, while those provided by Xavier et al. (2022) are not routinely updated. Therefore, the use of ERA5-Land and CPC reanalyses is very useful in environmental studies on the territory of Brazil.

**Materials and methods**

The work is divided into the following stages: definition of the study area, data collection, validation with statistical analysis, and climate classification of the entire study area using ERA5-Land and CPC/NOAA atmospheric reanalysis data.

**Study Area**

The study area is the NEB, which includes the states of Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, and Bahia. In order to validate the reanalysis data, 88 weather stations of the National Institute of Meteorology (INMET) distributed in the NEB and recording hourly data (1981-2019) were surveyed. After a one-by-one examination, data from only five of the 88 stations were suitable for use. They were the stations from the municipalities of Cruzeta – RN (6.43°S, 36.79°W, 226m), Palmeira dos Índios – AL (9.42°S, 36.62°W, 278m), Piripiri – PI (4.28°S, 41.79°W, 158m), Morro do Chapéu – BA (11.55°S, 41.15°W, 1002m), and Recife – PE (8.06°S, 34.96°W, 7m). The spatial distribution of these stations is shown in red dots in Figure 1. The other stations had a large number of missing data, making them inconvenient for validating reanalysis data.
The five selected weather stations were the ones that presented the longest period of collected data (2000-2016) with the fewest possible flaws, and thus can be used without compromising the validation of reanalysis data. Thus, the study period was 2000-2016.

Data

Daily precipitation (mm), total insolation (hours), and minimum (°C) and maximum (°C) temperature data from the INMET weather stations, as well as data at synoptic hours, were used. From these data, we calculated the mean daily temperature (°C), relative humidity (%), and wind speed (m/s). ERA5-Land air temperature (°C), dew point temperature (°C), surface solar radiation balance (J/m²), surface longwave radiation balance (J/m²), wind speed (m/s), and rainfall reanalysis data with a spatial resolution of 0.1 x 0.1 degrees, for the same period mentioned above, were retrieved. CPC/NOAA rainfall data with a spatial resolution of 0.5 x 0.5 degrees were used (Chen et al., 2008).

In a first analysis, it was observed that ERA5-Land reanalysis data were not able to capture the rainfall values of the wettest period in the semiarid region of the NEB. In turn, CPC/NOAA rainfall data successfully captured the observed data. The monthly rainfall for the Cruzeta – RN station (6.43° S, 36.58° W) in the period 2000-2016 according to the three data sources (weather station, ERA5-Land, CPC/NOAA) is shown in Figure 2. It can be observed that the monthly climatology from CPC/NOAA data in the wettest period follows that of the weather station, while ERA5-Land data show lower values. Lavers et al. (2022) analyzed the ability of ERA5-Land data to capture observed precipitation across the globe and found relatively good efficiency in extratropical zones, but low efficiency in the tropics. Therefore, our results in the present study for the NEB agree with those of Lavers et al. (2022).
For the ERA5-Land data, it was necessary to calculate the wind speed ($w_s$, in m/s) from the components $u$ (longitudinal component parallel to the x-axis) and $v$ (latitudinal component parallel to the y-axis) using the equation $w_s = \sqrt{u^2 + v^2}$. Wind speed data of the reanalysis products at 10 m were converted to 2 m according to the equation proposed by Allen et al. (1998).

Climatological means were calculated for all variables from the three different sources for the period 2000-2016. Microsoft Excel 2019 and the Python programming language were used for the calculations. As the ERA5-Land and CPC/NOOA dataset had, respectively, a 0.1 and 0.5-degree spatial resolution, mean values at the scale of 0.5º in latitude and longitude were obtained through the Python 3.8 programming language during the manipulation of reanalysis data, covering the whole territory of the study area. Thus, a total of 505 grid points were generated, as illustrated in Figure 3, representing 5.74 times the number of INMET stations distributed in the study area.

Figure 2. Climatological normal precipitation (mm) (2000-2016) from ERA5-Land, CPC/NOAA, and Cruzeta - RN (INMET) weather station data.

Figure 3. Map of the 505 ERA5-Land and CPC/NOAA reanalysis data points used for climate classification.
The climatology was estimated for the period 2000-2016 at each point and then the climatic classification was calculated. Figures were prepared by the interpolation of data using the Surfer 8.1 software.

Calculation of reference evapotranspiration - \( ET_0 \)

After manipulation of the data for climatology, reference evapotranspiration (\( ET_0 \)) values were estimated using the full-form Penman-Monteith-FAO mathematical model proposed by Allen et al. (1998) and recommended by FAO.

Climate classification

After estimating \( ET_0 \), the following climatic indices were calculated: Humidity Index (Ih), Aridity Index (Ia), \( I_m \) proposed by Thornthwaite (1948), and AIUNEP (Middleton & Thomas, 1992; 1997).

These indices were calculated using serial water balance values according to the model proposed by Thornthwaite & Mather (1955) and developed in a Microsoft Excel spreadsheet by Rolim et al. (1998).

Table 1 shows the climate classes according to the AIUNEP limit values. It is noteworthy that this index has been used to detect areas subject to desertification processes, according to the definition of the United Nations.

<table>
<thead>
<tr>
<th>Aridity index</th>
<th>Climate classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AI_{UNEP} \geq 1.00 )</td>
<td>Humid</td>
</tr>
<tr>
<td>( 0.65 &lt; AI_{UNEP} &lt; 1.00 )</td>
<td>Moist subhumid</td>
</tr>
<tr>
<td>( 0.50 &lt; AI_{UNEP} \leq 0.65 )</td>
<td>Dry subhumid</td>
</tr>
<tr>
<td>( 0.20 &lt; AI_{UNEP} \leq 0.50 )</td>
<td>Semiarid</td>
</tr>
<tr>
<td>( 0.05 &lt; AI_{UNEP} \leq 0.20 )</td>
<td>Arid</td>
</tr>
<tr>
<td>( AI_{UNEP} \leq 0.05 )</td>
<td>Hyperarid</td>
</tr>
</tbody>
</table>

Source: Adapted from Middleton & Thomas (1992).

Table 2 shows the classification of the climate type according to the \( I_m \).

<table>
<thead>
<tr>
<th>Moisture index</th>
<th>Climate type</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_m \geq 100 )</td>
<td>Perhumid</td>
</tr>
<tr>
<td>( 80 \leq I_m &lt; 100 )</td>
<td>Humid 4</td>
</tr>
<tr>
<td>( 60 \leq I_m &lt; 80 )</td>
<td>Humid 3</td>
</tr>
<tr>
<td>( 40 \leq I_m &lt; 60 )</td>
<td>Humid 2</td>
</tr>
<tr>
<td>( 20 \leq I_m &lt; 40 )</td>
<td>Humid 1</td>
</tr>
<tr>
<td>( 0 \leq I_m &lt; 20 )</td>
<td>Moist subhumid</td>
</tr>
<tr>
<td>( -20 \leq I_m &lt; 0 )</td>
<td>Dry subhumid</td>
</tr>
<tr>
<td>( -40 \leq I_m &lt; -20 )</td>
<td>Semiarid</td>
</tr>
<tr>
<td>( -60 \leq I_m &lt; -40 )</td>
<td>Arid</td>
</tr>
</tbody>
</table>

Source: Adapted from Thornthwaite (1948).

In addition to the climatic classifications based on the AIUNEP and \( I_m \), a classification was made using a new index called ‘absolute aridity index’ \( (I_{ab}) \), which is the ratio between \( ET_0 \) and total annual precipitation. We propose a climatic classification based on this index as presented in Table 3.

<table>
<thead>
<tr>
<th>Absolute aridity index</th>
<th>Climate type</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I_{ab} &gt; 12.00 )</td>
<td>Hyperarid</td>
</tr>
<tr>
<td>( 3.50 &lt; I_{ab} \leq 12.00 )</td>
<td>Arid</td>
</tr>
<tr>
<td>( 1.80 &lt; I_{ab} \leq 3.50 )</td>
<td>Semiarid</td>
</tr>
<tr>
<td>( 1.35 &lt; I_{ab} \leq 1.80 )</td>
<td>Dry subhumid</td>
</tr>
<tr>
<td>( 1.00 &lt; I_{ab} \leq 1.35 )</td>
<td>Mois subhumid</td>
</tr>
<tr>
<td>( 0.40 &lt; I_{ab} \leq 1.00 )</td>
<td>Humid</td>
</tr>
<tr>
<td>( I_{ab} \leq 0.40 )</td>
<td>Hyperhumid</td>
</tr>
</tbody>
</table>

The threshold \( I_{ab} \) values discriminating climate types in the newly proposed classification were chosen based on a solid justification. Values equal to 1.00 indicate that the precipitation is equal to \( ET_0 \) and, thus, areas with \( I_{ab} \) equal to or lesser than 1.00 present a climate classified as humid, since the precipitation is greater than the evaporative demand of the atmosphere. However, when precipitation is equal to or 2.5 times higher than \( ET_0 \), there is a moisture surplus of at least 1.5 times the evaporative demand of the atmosphere over a year and, thus, areas with \( I_{ab} \) values lower or equal to 0.40 present a climate classified as hyperhumid. On the other hand, \( I_{ab} \) values greater than 1.00 indicate that annual precipitation is lower than \( ET_0 \), that is, the area presents a non-humid climate. However, when \( ET_0 \) is greater than the precipitation but the \( I_{ab} \) does not exceed 1.35, the

Sales, E. S. G.; Matsunaga, W. K.; Braga, C. C.; Sakamoto, M. S.; Lucena, D. B.; Brito, J. I. B.; Arraut, J. M
climate is classified as moist subhumid, since the annual precipitation is at least 74% of ET₀ and there are rainy and dry seasons throughout the year. Therefore, it is likely that there is a water surplus during the rainy season, which results in a dry season without major water deficit, making the climate not dry, but rather moist subhumid. In this context, it is better to classify dry climates as those in which Iₐₑ values are greater than 1.35. However, as recommended by UNEP (Middleton & Thomas, 1992; 1997), it is convenient to present four types of dry climates: dry subhumid, semiarid, arid, and hyperarid.

In cases in which the total annual precipitation is 55% lower than ET₀, with Iₐₑ greater than 1.80, the climate has a high degree of aridity, with little or no water surplus throughout the year, which does not characterize a subhumid climate, not even a dry subhumid climate. Therefore, the dry subhumid climate is one in which Iₐₑ values are between 1.35 and 1.80 (Table 3). Following this conjuncture, a semiarid climate is one in which Iₐₑ values are greater than 1.80 but lower than or equal to 3.5, because when ET₀ is 3.5 times higher than precipitation, the degree of aridity is quite high and water deficit is likely to occur throughout the twelve months of the year, which leads to the classification of arid climate. Finally, in areas where Iₐₑ is greater than 12, there is an extremely high degree of aridity, which produces a hyperarid climate (Table 3).

We used available water capacity (AWC) values specific for each type of soil in the municipalities of the NEB provided by ANA (2021), whose spatial configuration is shown in Figure 4.

Figure 4. Soil available water capacity (mm) in Northeast Brazil.
Validation

Statistical parameters, mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), and standard error of estimate (SES), and Pearson correlation coefficient (r) were used to validate the data and analyze the results. Temperature, precipitation and wind speed data were validated from the five INMET weather stations that presented the lowest number of flaws in the period 2000-2016 were analyzed in order to validate the ERA5-Land and CPC/NOAA data.

Precipitation from CPC/NOAA was the variable that presented the highest MAE, RMSE and SES values, precisely because it was the one with the highest seasonal, interseasonal and interannual variability. However, it was observed that precipitation presented good correlation indices in all municipalities, with $r \geq 0.95$, indicating that CPC/NOAA reanalysis can be used in the absence of data.

One of the studies developed by Sena et al. (2012) compared rainfall data from the CPC/NOAA project with observed rainfall data for the Cariri region of Paraiba during the period 1979-2010 and the results showed a good correlation between the series, with coefficients varying from 0.58 to 0.89, all significant at 95% confidence. The CPC/NOAA data were also able to reproduce well the rainiest trimester, between the months of February and April in the study area, with a margin of error of less than 20%, which can be considered relatively small considering the great variability found in precipitation.

Cardoso & Quadro (2017) analyzed the performance of new-generation CPC precipitation data for the Southern region of Brazil, comparing them with observational data from National Water Agency (ANA) and INMET weather stations. The CPC data showed good accuracy when compared to INMET and ANA observational data, and regarding seasonality, the CPC data showed better performance in all statistical parameters evaluated.

Wind speed and temperature presented relatively low MAE, MAPE, RMSE, SES values and high correlation coefficients, indicating that ERA5-Land reanalysis can be used to estimate these variables in the NE. Araújo et al. (2022) statistically analyzed ERA5-Land reanalysis air temperature estimates with surface data for the state of Pernambuco and concluded that ERA5-Land reanalysis estimates agree well with weather station-based data in almost the entire state, showing accuracy with $r^2 = 0.98$ and RMSE = 0.60 °C.

Lompar et al. (2019) tested the use of temperature data from ERA5 reanalysis to fill gaps in serially meteorological data for different landscapes, latitudes and altitudes, including tropical and mid-latitudes. An evaluation of the results was performed in terms of RMSE obtained using hourly and daily data. The study showed very low mean RMSE values, ranging from 1.1 °C (Montecristo, Italy) to 1.9 °C (Gumpenstein, Austria), what indicates that ERA5 data can be used to fill in temperature gaps in case of lack of temperature data.

Siefert et al. (2021) also evaluated the performance of 3 reanalysis products (ERA5, GLDAS 2.1, and MERRA-2) for surface wind speed data on a daily scale based on observational data from 521 weather stations for the period 2000-2018 in Brazil. Among the three products, ERA5 was more accurate for the country’s climate zones in terms of mean trends and seasonality. Fernandes et al. (2021) compared ERA5 atmospheric reanalysis wind speed data with wind observations from three coastal regions of Brazil: Maranhão, Santa Catarina, and Santos Basin. The results demonstrated that ERA5-Land is well suited for daily to monthly scale analysis of wind speeds, with $r \geq 0.74$, but the resolution of the current model precludes a close representation of the diurnal variability in places where the sea breeze is an important component of the circulation.

Jiang et al. (2019) analyzed the deviations of ERA5-Land hourly radiation data when compared to in situ measurements from 98 sites in China and showed that the reanalysis estimates correlated well with the ground observations and fully reflected regional and daily variations at individual sites.

Therefore, in view of the statistics found in our study and the data presented in similar previous studies, reanalysis data can be used to supply missing data from weather stations, emerging as an alternative to carry out and improve studies on climate change that depend on long-term data series, as for example in the NE.
Results and discussion

Evapotranspiration and Precipitation

The mean monthly Penman-Monteith-FAO ET₀ estimates (mm/month) for the period 2000-2016 obtained using ERA5-Land and station-based data are presented in Figure 5a and 5b, respectively.

Figure 5. Monthly Penman-Monteith-FAO ET₀ estimates (mm) calculated with ERA5-Land (A) and weather station (B) data.

In general, the values obtained were very close, presenting the same behavior throughout the months of the year. The highest and lowest values in the different months could be identified and represented. A strong correlation was found, with $r \geq 0.95$, for the five locations, confirming the efficiency of ERA5-Land reanalysis data when observational data for ET₀ calculation are absent.

Ismael Filho et al. (2015) proved that temperature and radiation are the two variables with the greatest direct effect on evapotranspiration estimates, in line with the works of Lompar et al. (2019) and Jiang et al. (2019) who demonstrated the reliability of temperature and radiation data from ERA5-Land. Furthermore, the behavior of ET₀ in Figure 5a and 5b allows us to conclude that ERA5-Land data can be reliably used in the absence of observational data.

Similar research carried out by Paredes et al. (2021) evaluated the accuracy of daily Penman-Monteith-FAO ET₀ estimates using shortwave radiation data ($R_s$) and ERA5-Land temperature provided by ECMWF when station data were not available. Paredes et al. (2021) used data from 37 weather stations distributed on the mainland of Portugal, where climatic conditions vary from semiarid to humid, and 12 weather stations located on the Azores islands, characterized by humid, windy and often cloudy conditions, were used for validation. In general the results showed a good accuracy when ET₀ was calculated using ERA5-Land variables, with acceptable RMSE values and $r \geq 0.8$ in most locations, allowing the authors to conclude that the use of this product was a good alternative when observed meteorological data were not available; however, despite the good usability of the ERA5-Land product, further research on its application is still needed.

Vanella et al. (2022) statistically assessed the reliability and consistency of the global ERA5 single levels and ERA5-Land reanalysis datasets to calculate ET₀ estimates by comparing them with agrometeorological data from 66 weather stations for the period 2008-2020 under different climates and topographies in Italy. A good general agreement was obtained between ET₀ estimates and station data on a daily and seasonal time scale, especially under temperate climate conditions, with slightly higher accuracy values for ET₀ estimates using the ERA5-Land product. This confirms the potential usefulness of reanalysis datasets as an alternative data source to estimate ET₀, overcoming the unavailability of observational data.

Figure 6a and 6b show the mean annual spatial configurations of ET₀ (mm/year) and precipitation (mm/year) in the NEB, respectively, using ERA5-Land data (ET₀) to estimate ET₀ and CPC/NOAA data to estimate precipitation.
As shown in Figure 6a, maximum ET₀ values were found in part of the hinterland of the states of Rio Grande do Norte, Paraíba, Pernambuco, Ceará, Piauí, and Bahia, consequently associated with high levels of solar radiation, low relative humidity and low level of precipitation (Figure 6b), creating specific conditions of semiarid and even arid climates. The ET₀ values found here are similar to those found in other works. For example, Júnior & Bezerra (2018) found a total mean annual ET₀ estimate in Northeast Brazil of up to 2098.0 mm for the western region of the state of Rio Grande do Norte, Paraíba, Pernambuco, southern Ceará, eastern Piauí, and part of northern Bahia.

CPC/NOAA data were able to represent well the spatial configuration of the precipitation data (Figure 6b), following the pattern presented by INMET and researchers such as Nobre and Molion (1988) and Marengo et al. (2011). With this dataset, it was possible to identify specific points of higher precipitation in some locations whose surrounding areas present lower precipitation, such as central Bahia and southern Ceará State, corresponding to the location of the Chapada Diamantina in the former and Chapada do Araripe in the latter, which are two high-altitude mountain regions.

Climate Classification

After the validation of the reanalysis data, the climatic indices Iₙ, Iₐ, and Iₘ and the AIᵪᵢᵥₑₑₑₑ were calculated for the study area, the latter being the one currently used for the climatic classification of the Brazilian semiarid region.

The climate classification using AIᵪᵠₑₑₑₑ is shown in Figure 7. This index was apparently able to represent well the transition between climate types of the coastal region and the hinterland, that is, from humid to semiarid. The largest highlighted area corresponds to the semiarid region, with 834,448 km², representing 53.8% of the total area of the NEB (1,552,175 km²). Similar results were found by Sales et al. (2021), who carried out a climate classification for Northeast Brazil using INMET 1981-2010 climatological data and the AIᵪᵠₑₑₑₑ calculated using ET₀ estimates by the Penman-Monteith-FAO equation. They found a total area of 812,026.9 km² of semiarid climate, a value very close to that obtained in the present study.
A small arid area of 3,800 km² can be observed in the map, inserted in the Submedium mesoregion of the São Francisco River (Figure 7). This region has specific characteristics of high temperature and evapotranspiration and irregular precipitation, with an annual mean of less than 500 mm (Figure 6b). When comparing Figs. 6a and 6b with Figure 7, it appears that the area classified as presenting arid climate is very small and possibly does not represent the regional reality, as in Figure 6b a large area on the border between Pernambuco and Bahia is observed, extending from Piauí to the border of Bahia with Alagoas and Sergipe, where a high reference potential evapotranspiration is observed (Figure 6a). Therefore, the arid area along the Pernambuco-Bahia border likely extends from Piauí to the Bahia-Sergipe border, and not in an isolated core as shown in Figure 7. Thus, the arid area in the NEB is greater than that depicted in Figure 7. The climate classification based on AI\textsubscript{UNEP} values in Figure 7 for the central area of the NEB led to an underestimate of the arid climate in relation to reality. However, in the vicinity of Salvador, in the central part of the coast of Bahia, there is a moist subhumid climate (Figure 7), but the mean annual rainfall in this area is greater than 2000 mm/year (Simões, 2017) and the climate is, thus, humid. On the other hand, it is still possible to observe that the calculation of AI\textsubscript{UNEP} with ERA5-Land and CPC/NOAA data allowed to detect areas with a dry subhumid climate in central Bahia and southern Ceará, precisely where the Chapada Diamantina and Chapada do Araripe are located, two mountainous regions with high altitudes and mean annual precipitation higher than the surrounding areas.

Lopes et al. (2017) found similar results shown in Figure 1.5. They performed the calculation of the AI\textsubscript{UNEP} and analyzed climate trends towards desertification in the semiarid region of the NEB from 1961 to 2015 and detected statistically significant trends of increasing aridity, leading to the conclusion that this region of Brazil may become highly prone to desertification.

The climate classification based on I\textsubscript{im} is presented in Figure 8. In this classification, the area...
with arid climate (363,919 km²) was 95.8 times larger than that found with AI\textsubscript{UNEP} (3,800 km²). The largest highlighted area (692,385 km²) still corresponds to the semiarid region, representing 44.6% of the total area of the NEB, but 17% smaller than the area found with AI\textsubscript{UNEP} (834,448 km²). Further, in relation to the classification based on AI\textsubscript{UNEP}, there is an increase in the semiarid region in the state of Maranhão and the coast of the state of Ceará, and a decrease in the area with dry subhumid climate (Figure 7 and 8). Similar results of those shown in Figure 8 were obtained by other researchers such as Marcos Junior (2018), Jesus et al. (2019), Sales et al. (2021), and Oliveira et al. (2021).

This increase in the arid region according to I\textsubscript{m} (Figure 8) in relation to AI\textsubscript{UNEP} (Figure 7) is due precisely to the high levels of ET\textsubscript{0} and low precipitation in this region (see Figure 6a and 6b) and consequent higher water deficit. However, in the central part of Ceará, in part of the border between Ceará and Piauí, and on the western border of Paraíba with Pernambuco, rainfall is higher than that of the Pernambuco-Bahia border, and in these same areas the reference potential evapotranspiration is lower than that of the Pernambuco-Bahia border. Evidently, these areas do not have the same climate. Thus, Thornthwaite climate classification produced an overestimation of the arid climate in relation to reality. Similarly, according to this classification, the climate in the southeastern coast of Bahia fell into the moist subhumid category, but this area is actually known to have a humid climate (Sambuchi & Haridasan, 2007; Simões et al., 2017; Mencia et al., 2017; Mencia et al., 2021). Overall, the classifications based on AI\textsubscript{UNEP} and I\textsubscript{m} generated different climates in many areas of the NEB. However, comparing the configurations of these two climate classifications (Figure 7 and Figure 8) with that shown in Figure 6a and 6b, it is not possible to determine which of the two best represents the climate of the NEB, especially concerning the extent of the arid area, which is large according to I\textsubscript{m} but very small according to AI\textsubscript{UNEP}. Therefore, in the present work, a new index, the I\textsubscript{ab}, is proposed.

**Figure 8.** Climate classification for the NEB according to I\textsubscript{m}.

<table>
<thead>
<tr>
<th>Climate</th>
<th>Area (Km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humid 3</td>
<td>6,618</td>
</tr>
<tr>
<td>Humid 2</td>
<td>16,797</td>
</tr>
<tr>
<td>Humid 1</td>
<td>42,531</td>
</tr>
<tr>
<td>Moist subhumid</td>
<td>129,184</td>
</tr>
<tr>
<td>Dry subhumid</td>
<td>301,741</td>
</tr>
<tr>
<td>Semiarid</td>
<td>692,385</td>
</tr>
<tr>
<td>Arid</td>
<td>362,919</td>
</tr>
<tr>
<td><strong>Total Area</strong></td>
<td><strong>1,552,175</strong></td>
</tr>
</tbody>
</table>

Sales, E. S. G.; Matsunaga, W. K.; Braga, C.C.; Sakamoto, M. S.; Lucena, D. B.; Brito, J. I. B.; Arraut, J. M.
The climate classification based on $I_{ab}$ is presented in Figure 9. It is observed that the classification with this index was able to represent very well the climate types of the NEB, respecting the climatic transition from the coast (humid) to the central part (arid), as well as, from the central part to the northwest, in the border with the Amazon Forest, describing with good reliability the transition from arid to humid climates.

Two areas classified with arid climate are observed in Figure 9: a small area in the center-north region of Rio Grande do Norte and other in the Submedium mesoregion of the São Francisco River and its surroundings, covering totaling a total of 128,940 km$^2$ in areas of the states of Bahia, Piauí, and Pernambuco, which represents 8.3% of the territory of the NEB. In Piauí the arid area is found in the high and medium Canindé microregion; in Pernambuco, in the Submedium mesoregion of the São Francisco River; and in Bahia, in the region known as Raso da Catarina. Comparing Figs. 6a and 6b with Figure 9, it is observed that the degree of aridity – which leads to the classification of the climate as arid – presented in Figure 9 is consistent with the reference evapotranspiration (Figure 6a) and precipitation (Figure 6b) fields. It is noteworthy that these areas are known to be very dry and present high degree of aridity, especially the Raso da Catarina (Conti, 2005; Lucena et al., 2016; Lopes et al., 2017). The center-north region of Rio Grande do Norte, which corresponds to the Angicos microregion, is also known for its high degree of aridity, with rainfall below 500 mm/year and reference evapotranspiration above 2000. These characteristics were also observed by Diniz & Pereira (2015). Thus, important differences are seen in the extent of the arid climate obtained by the three methods. When using $AI_{UNEP}$ and $I_{ab}$, arid areas cover 0.25% and 23.4% of the total area of the NEB, respectively, while this percentage is found to be 8.3% when using the $I_{ab}$. In their analysis of areas of the NEB that have the highest

---

**Table 9. Climate Classification for the NEB according to $I_{ab}$**

<table>
<thead>
<tr>
<th>Climate</th>
<th>Area (km$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperarid</td>
<td>0</td>
</tr>
<tr>
<td>Arid</td>
<td>128,940</td>
</tr>
<tr>
<td>Semi-arid</td>
<td>823,032</td>
</tr>
<tr>
<td>Dry sub-humid</td>
<td>282,759</td>
</tr>
<tr>
<td>Moist Sub-humid</td>
<td>218,044</td>
</tr>
<tr>
<td>Humid</td>
<td>99,400</td>
</tr>
<tr>
<td>Hyper-humid</td>
<td>0</td>
</tr>
<tr>
<td>Total Area</td>
<td>1,552,175</td>
</tr>
</tbody>
</table>

---

Sales, E. S. G.; Matsunaga, W. K.; Braga, C.C.; Sakamoto, M. S.; Lucena, D. B.; Brito, J. I. B.; Arraut, J. M
degree of susceptibility to desertification, Lopes et al. (2017) found an area that is greater in relation to the aridity indicated by $I_{m}$ and lower than that indicated by $I_{UNEP}$. Therefore, it is observed that $AI_{UNEP}$ underestimated and $I_{m}$ overestimated the size of arid areas in the NEB.

The climate classifications with $AI_{UNEP}$ (Figure 7) and $I_{lab}$ (Figure 9) detected very similar areas with semiarid climate, namely, 833,448 km$^2$ and 823,032 km$^2$, representing 53.7% and 53% of the total area of the NEB, respectively, corresponding to a difference of only 0.7% between the two indices. In turn, the semiarid area obtained with $I_{m}$ represented 44.6% of the area of the NEB, since part of the areas with semiarid climate was estimated to have arid climate. Regarding the dry subhumid climate type, the areas obtained with the three methods, $I_{lab}$, $AI_{UNEP}$ and $I_{m}$, were very close, representing 18.2% (282,759 km$^2$), 17.3% (268,063 km$^2$) and 19.4% (301,741 km$^2$) of the total area of the NEB, respectively. On the other hand, the estimated areas with moist subhumid climate varied: 218,044 km$^2$ (14.0% of the NEB) with $I_{lab}$, 346,483 km$^2$ (22.3% of the NEB) with $AI_{UNEP}$ and 129,184 km$^2$ (8.3% of the NEB) with $I_{m}$. The areas classified as presenting humid climate presented very similar values according to $I_{lab}$ (99,400 km$^2$) and $AI_{UNEP}$ (100,381 km$^2$), representing 6.4% and 6.5% of the total area of the NEB, respectively. In turn, according to $I_{m}$, the humid climate covered 65,946 km$^2$, which corresponds to 4.2% of the area of the NEB.

An interesting result is the classification of the climate on the coast of the border between the states of Alagoas and Sergipe as dry subhumid observed with the use of the three indices (Figure 7, 8 and 9). Marengo et al. (2019) described remnants of savanna vegetation near the coast of the border between the states of Alagoas and Sergipe, and Cantidio and Souza (2019), in their study on Atlantic Forest, described areas of Caatinga in that region too. Another commonality among the three indices is that the semiarid climate type occupied the largest area compared to the other climate types, covering 53.8%, 44.6% and 53.0% of the total area of the NEB according to the $AI_{UNEP}$, $I_{m}$ and $I_{lab}$, respectively.

Thus, our results showed that climate systems based on $I_{m}$ and $AI_{UNEP}$ presented a tendency towards more arid and more humid climates, respectively, in relation to reality, while the $I_{lab}$ proved to be more robust for a more accurate classification of the climate.

reanalysis data when observed data are not available for the calculation of ET$_{0}$

The results obtained in the statistical analysis indicate that ERA5-Land and CPC/NOAA data can be used in the absence of reliable observational data, emerging as an alternative to solve problems related to the terms of temporal and spatial coverage of data in the NEB. Comparisons with observed data are fundamental for the identification of uncertainties in their use in studies addressing agricultural, climatological and hydrological simulations on the Brazilian territory.

Regarding the climate classification, both $AI_{UNEP}$ and $I_{m}$ represented well the transition of climate types from the coastal region to the hinterland, from humid to arid. However, in general, in transitional areas between climate types, the classification based on $AI_{UNEP}$ showed a trend towards more humid climates, while the one with $I_{m}$ showed a trend towards more arid climates.

In turn, the use of the $I_{lab}$ is safe and indicated for climatic classifications mainly of dry lands, as it was able to clearly represent the different types of climates of the NEB, especially the arid, semiarid and dry subhumid climates.
Acknowledgments

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for funding this research.

References
Chirico, G. B., Pelosi, A., De Michele, C, Bolognesi, S. F., D’Urso, G. 2018. Forecasting potential evapotranspiration by combining numerical weather predictions and visible and near-infrared satellite images: an application in
southern Italy. The Journal of Agricultural Science 156, 702-717.


Fontana, D.C., Alves, G.M., Roberti, D., Moraes, O. L.L., Gerhardt, A., 2012. Estimativa da radiação fotossinteticamente ativa absorvida...
pela cultura da soja através de dados do sensor Modis. Revista Bragantia 71, 563-571.


Sales, E. S. G.; Matsunaga, W. K.; Braga, C.C.; Sakamoto, M. S.; Lucena, D. B.; Brito, J. I. B.; Arraut, J. M
evapotranspiration over two temperate semiarid grasslands in North America. Agricultural and Forest Meteorology 152, 31-44.


evapotranspiration forecasts for the contiguous United States based on multi-model numerical weather predictions. Journal of Hydrology. 562, 502-517.


Sales, E. S. G.; Matsunaga, W. K.; Braga, C.C.; Sakamoto, M. S.; Lucena, D. B.; Brito, J. I. B.; Arraut, J. M


Wu, G., Qin, S., Mao, Y., Ma, Z., & Shi, C. 2022. Validation of precipitation events in ERA5 to gauge observations during warm seasons over eastern China. Journal of Hydrometeorology 23, 807-822.
