







Article

Gap Filling and Quality Control Applied to Meteorological Variables Measured in the Northeast Region of Brazil

Rafaela Lisboa Costa ¹, Heliofábio Barros Gomes ¹, David Duarte Cavalcante Pinto ^{1,2},
Rodrigo Lins da Rocha Júnior ¹, Fabrício Daniel dos Santos Silva ^{1,*}, Helber Barros Gomes ¹,
Maria Cristina Lemos da Silva ¹ and Dirceu Luís Herdies ³

¹ Institute of Atmospheric Sciences, Federal University of Alagoas, Maceió 57072-900, Brazil; rafaelalisboac@gmail.com (R.L.C.); heliofabio@icat.ufal.br (H.B.G.); david.duarte@icat.ufal.br (D.D.C.P.); rodrigo.junior@icat.ufal.br (R.L.d.R.J.); helber.gomes@icat.ufal.br (H.B.G.); cristina.lemos@icat.ufal.br (M.C.L.d.S.)

² Centre of Astronomical Studies of Alagoas, Maceió 57051-090, Brazil

³ National Institute for Space Research, Cachoeira Paulista, São Paulo 12630-000, Brazil; dirceu.herdies@inpe.br

* Correspondence: fabricio.santos@icat.ufal.br

Abstract: In this work, we used the MICE (Multivariate Imputation by Chained Equations) technique to impute missing daily data from six meteorological variables (precipitation, temperature, relative humidity, atmospheric pressure, wind speed and insolation) from 96 stations located in the northeast region of Brazil (NEB) for the period from 1961 to 2014. We then applied tests with a quality control system (QCS) developed for the detection, correction and possible replacement of suspicious data. Both the applied gap filling technique and the QCS showed that it was possible to solve two of the biggest problems found in time series of daily data measured in meteorological stations: the generation of plausible values for each variable of interest, in order to remedy the absence of observations, and how to detect and allow proper correction of suspicious values arising from observations.

Keywords: time series; missing values; data quality control; verification; climate analysis



Citation: Costa, R.L.; Barros Gomes, H.; Cavalcante Pinto, D.D.; da Rocha Júnior, R.L.; dos Santos Silva, F.D.; Barros Gomes, H.; Lemos da Silva, M.C.; Luís Herdies, D. Gap Filling and Quality Control Applied to Meteorological Variables Measured in the Northeast Region of Brazil.

Atmosphere **2021**, *12*, 1278. <https://doi.org/10.3390/atmos12101278>

Academic Editor: Peter Domonkos

Received: 4 August 2021

Accepted: 28 September 2021

Published: 30 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Observational meteorological data are basic elements for climatological analyses [1]. In Brazil, despite the inherent relevance of such observations, the amount of data of this kind has been suffering a significant reduction over the years [2], with several manual weather stations (MWS) being permanently closed, becoming inoperative or functioning precariously, compromising the quality and continuity of the meteorological records.

This situation poses a hindrance to a more detailed, observation-based climate analysis, forcing the use of Reanalyses products, which constitute a synthetic database reconstructed by calibrating a regional climate model to observed historical conditions, in a grid format, with statistical properties, such as means and variances, that are very similar to those of the observations [3]. However, although useful for analysing long-term climate trends and studying future climate change scenarios [4], they are not reliable in terms of extreme events [5].

To overcome such an issue, especially in the case of precipitation data, several methods have been proposed to create gridded products that provide a standardization of the properties of the variable across a spatial field, thus surmounting problems related to sparse and non-uniform rainfall coverage. Some of those methods are based on exploring the surface observations to the maximum [6–9]; others are based on combining observed rainfall data with estimates from remote sensing [10–17].

In spite of all these efforts, a fine-grained surface observation database is essential, whether to aid in composing such grids or to correct and validate the approaches derived from remote sensing techniques. A surface database is not just about collecting and storing

observations; in order to achieve a desirable level of confidence, it needs to undergo three basic steps of processing: quality control checks, gap filling and homogenization. With respect to the daily data, for which accessibility is still very restricted, particularly rigorous Quality Control Systems (QCS) are essential [18,19].

A QCS must combine people and machines. Although QCS software is designed to provide a list of suspicious data, the final decision regarding this should be made by qualified personnel. A QCS should not be expected to detect a suspicious value and automatically remove it from the series; this must be very carefully assessed in order to avoid losses of real extreme values in the time series [20,21]. Therefore, the QCS needs to be based on a variety of consistency tests and provide understandable graphical outputs, as well as summaries containing the list of suspicious data, in order to facilitate decision-making and prevent the rejection of good observations [22,23].

Another problem faced by those who work with time series of meteorological variables is the amount of missing data. Studies on trends and extremes that are based on climate indices recommended by the Expert Team on Climate Change Detection and Indices (ETCCDI) of the World Meteorological Organization (WMO), developed by [24], require flawless daily data [5,25]. In this regard, the aim of this study was to present the results of the application of a QCS and a gap filling technique to time series of meteorological variables collected by manual weather stations in the northeast region of Brazil (NEB) in the period of 1961–2014, corresponding to 54 years of daily data, which were assessed here based on this timescale, as well as on 10-day and monthly averages or accumulated values (depending on the variable). The NEB, which is the most populated dryland region on the planet [26], has a singular biome associated with its semi-arid climate, the Caatinga, and is extremely vulnerable to climatic extremes, especially droughts [27,28]. To fill gaps in the daily data, the technique known as Multivariate Imputation by Chained Equations (MICE) was used, a multiple imputation method that has a number of advantages over other methods for dealing with missing data in time series [29–32].

2. Materials and Methods

2.1. Area of Study and Data

The NEB encloses an area of approximately 1.56 million km², which is equal to 18% of the Brazilian territory, encompassing nine of its states (Figure 1): Alagoas (AL), Piauí (PI), Maranhão (MA), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Sergipe (SE) and Bahia (BA). The Caatinga, which covers more than 750,000 km² of the NEB, exhibits, primarily, xerophytic, arboreal, prickly and deciduous properties. The east coast of the NEB, between RN and BA, hosts an exuberant tropical forest called the Mata Atlântica. The transition zone, of mixed vegetation, between the Mata Atlântica and the Caatinga, is called Agreste. In the westernmost portion, covering most of MA, the predominant vegetation is the Amazon forest, whereas in southern MA and PI, as well as in western BA, the Cerrado vegetation (the Brazilian Savannah) prevails [27,28].

The data come from 96 manual weather stations spatially spread over the NEB, representing part of the National Institute of Meteorology (INMET) station network (Figure 1, blue dots). The period of each series spans from 1 January 1961 to 31 December 2014 [5]. Gap filling and QCS were applied to precipitation (mm), temperature (°C), relative humidity (%), wind speed (m/s), station-level atmospheric pressure (hPa) and insolation (hours).

2.2. Filling in Missing Data

The application of a methodology to fill the gaps in the daily data took place before the evaluation performed by the QCS. The large number of observed gaps was addressed using the MICE technique. There are several fields of the natural sciences in which the MICE technique has been successfully applied [30], including, most notably, biostatistics. In this research, it was adapted to missing data in the time series of meteorological variables across the NEB, following the methodology of [33].

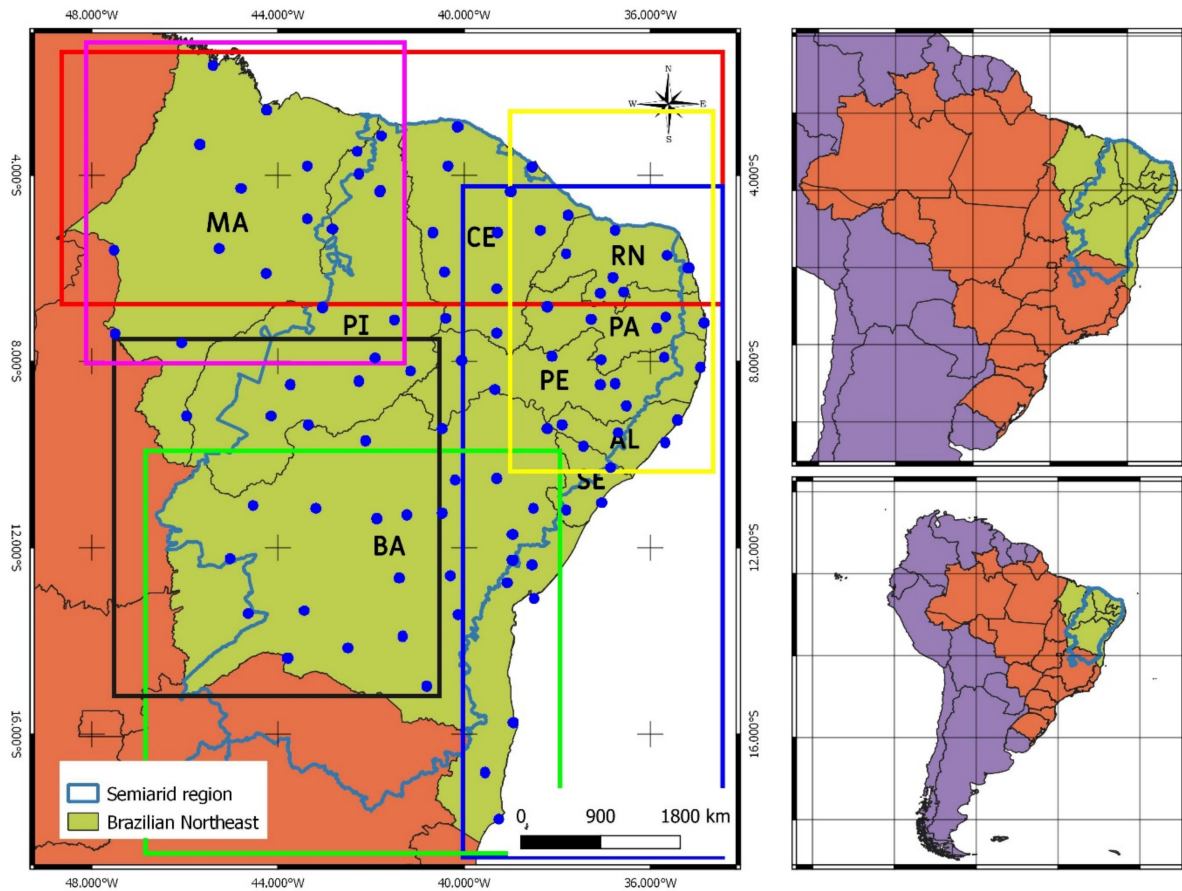


Figure 1. Map of the NEB (left) with the delimitation of the semiarid region and its location relative to Brazil (top right) and South America (bottom right). In the left panel, the blue points represent INMET's weather stations. The climate subregions that are referred to throughout this study are graphically presented in the left panel encompassed by rectangles with the following colour associations: northern NEB, red; northwestern NEB, pink; northeastern NEB, yellow; eastern NEB, blue; southern NEB, green; southwestern NEB, black. Extracted from [28].

For each variable, MICE creates multiple complete datasets based on a variety of methods, which may include linear regressions, logistic regressions, multinomial log-linear models or Poisson regressions. These models have in common the ability to impute the missing data from the known observed values and their relationships with each dataset. Several forecasts are thus created for each missing value, and the one that produces the least uncertainty and fewest errors, when compared with the observed data, is adopted [34].

Of the two modern approaches for multivariate imputation, i.e., joint modelling and fully conditional specification, the MICE technique is of the latter kind [30]. MICE is developed in R language and is made available as a package. A basic description of its approach is presented below.

Let Y_j with $(j = 1, \dots, p)$ be a set of p incomplete variables, where $Y = (Y_1, \dots, Y_p)$. The observed and missing sections of Y_j are denoted as Y_j^{obs} and $Y_j^{missing}$, respectively, so that $Y^{obs} = (Y_1^{obs}, \dots, Y_p^{obs})$ and $Y^{missing} = (Y_1^{missing}, \dots, Y_p^{missing})$ represent the observed and missing data of Y . The number of imputations must always be equal to $m \geq 1$. The imputed dataset h_{th} is given by $Y^{(h)}$, where $h = 1, \dots, m$. Now, let $Y_{-j} = (Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_p)$ be the collection of variables $p - 1$ in Y with the exception of Y_j . Let Q be the amount of missing data to model. In practice, Q is often a multivariate vector representing any model aimed at the imputation of missing data.

Figure 2 illustrates the three main steps in multiple imputation: imputation, analysis and clustering. The software stores the results of each step in specific classes, called *mids*, *mira* and *mipo*, explained in detail below.

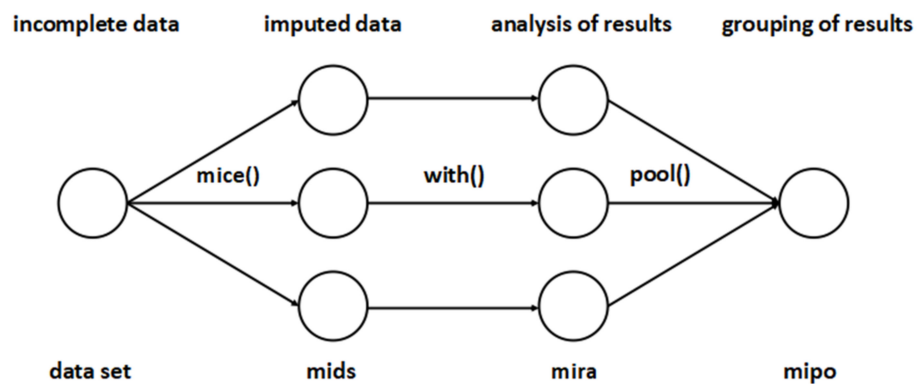


Figure 2. Main steps used in multiple imputation.

The leftmost side in Figure 2 indicates that the analysis starts with a set of observed Y_{obs} data. The problem is that it is not possible to estimate Q from Y_{obs} without making unrealistic assumptions about the unobserved data. Therefore, missing data are randomly generated for the Y_{obs} dataset, and several versions of multiple imputation are generated, with plausible values, according to the nature of the variable, extracted from a distribution specifically modelled for each imputed value for the respective missing value.

In MICE, this task is performed by the *mice()* function. Figure 2 depicts $m = 3$ imputed data, $Y_{(1)}, \dots, Y_{(3)}$. The three imputed sets are identical to the non-missing input datasets regarding their type of distribution. The magnitude of these differences reflects the uncertainty about the values to be imputed.

The second step is to estimate Q in each imputed dataset, just as in a flawless dataset. This becomes easy, since all the sets are complete. The model applied to $Y^{(1)}, \dots, Y^{(m)}$ is generally identical. The estimates $Q^{(1)}, \dots, Q^{(m)}$ are different from each other.

The third and last step is to gather the m estimates $Q^{(1)}, \dots, Q^{(m)}$ into a single Q_{mean} estimate and estimate its variance. For Q quantities that are normally distributed, it is possible to calculate the average of $Q^{(1)}, \dots, Q^{(m)}$ and then add it to the variance of Q_{mean} , according to the method described in [35]. The ideal is to apply this methodology to a column of data containing gaps alongside columns of similar data that do not contain missing values, called predictors, as the relationship established between the datasets will tend to improve the estimates of the data to be imputed to the column that presents missing data [33].

In the adaptation of the MICE technique to fill gaps in precipitation, the series of the gridded precipitation analysis from the Climate Prediction Center of the National Oceanic and Atmospheric Administration (CPC/NOAA) were used as sets of predictors [8,36], based on [37] optimal interpolation method, with a spatial resolution of $0.5^\circ \times 0.5^\circ$. For the other variables, series of gridded analyses provided by the NCEP/NCAR (National Center for Environmental Prediction/National Center for Atmospheric Research), with $1.0^\circ \times 1.0^\circ$ resolution, were used as predictor variables [3]. In order to avoid the influence of dry periods on rainy periods of the year, and vice versa, we applied MICE to independent files organized month by month, that is, to 12 independent files, from January to December, for each variable.

In time series that have missing data, these are characterised by “NA”, using the default number of multiple imputations ($m = 5$ iterations) of the MICE package, version 2.12, of the free statistical software R. The imputations are generated according to the default method, which is, for numerical data, the PMM (Predictive Mean Matching) method. Using precipitation data to exemplify the procedure, the original faulty series of a weather station is placed side by side with data from the four grid points closest to the location (Table 1).

Table 1. Precipitation series with missing data, represented by NA, from the Ouricuri station (WMO code: 82753)—Pernambuco (a) and after the gaps have been filled (imputed values in red; b). The original data of the station (Orig column) are followed by the nearest gridded series, which constitute the set of predictors (G-01, G-02, G-03 and G-04).

(a)								(b)							
Year	Month	Day	Orig	G-01	G-02	G-03	G-04	Year	Month	Day	Orig	G-01	G-02	G-03	G-04
1980	12	15	0	3.2	3.8	2.7	7.9	1980	12	15	0	3.2	3.8	2.7	7.9
1980	12	16	0	1.9	0.7	3.2	0.4	1980	12	16	0	1.9	0.7	3.2	0.4
1980	12	17	NA	8.3	5.5	3.3	3.7	1980	12	17	5.2	8.3	5.5	3.3	3.7
1980	12	18	15.9	14.7	5.5	19.1	5.1	1980	12	18	15.9	14.7	5.5	19.1	5.1
1980	12	19	6.8	3.9	1.4	9.6	1.7	1980	12	19	6.8	3.9	1.4	9.6	1.7
1980	12	20	NA	12.3	4.4	14.1	5.8	1980	12	20	12.6	12.3	4.4	14.1	5.8
1980	12	21	0	0.1	0	0.2	0.2	1980	12	21	0	0.1	0	0.2	0.2
1980	12	22	0	6.2	3.1	4.9	0.8	1980	12	22	0	6.2	3.1	4.9	0.8

After imputing the missing data, at least 5 years without gaps were identified in the original series of observed data. Gaps were then artificially generated for those years and the method was used again, in order to compare the original observed data with the imputed values that replaced them. This allowed us to assess the method's ability. Statistics such as correlations and root mean square error (RMSE) were calculated for these verification periods to validate the methodology. These verification metrics were used for three timescales, namely, daily, 10-day and monthly (the latter two consisted of averages or accumulated values over the period, depending on the variable). The RMSE was selected as the "MICE" dexterity estimation measure because it has, among other advantages, the possibility of expressing the accuracy of the numerical results with error values in the same dimensions as the analysed variables, that is, millimetres for precipitation, degrees Celsius for temperature, percentage for relative humidity, hectoPascals for atmospheric pressure, metres per second for average wind speed and hours for insolation. The daily scale is important for analysing climate extreme indices [38–43]; the 10-day scale is important for application in agrometeorological studies, as this is the time step used in many crop growth simulation models [44–46], whereas the monthly scale is essential for studies that involve analyses of the influence of modes of variability on climate dynamics and also for research in the area of seasonal and subseasonal climate forecasts [47–51].

2.3. QCS

Some QCS techniques have a strong stochastic component that can lead to a high probability of rejecting good observations [20,21,37]. The QCS used here is based on a series of consistency tests, in order to reduce the stochastic dependence and allow the decision-maker to accept or reject data considered doubtful, based on easy-to-understand graphical outputs and reports containing the list of suspicious data.

Consistency tests are an important set of checks for possible errors, as they are expected to explore the temporal and spatial interrelationship of climatological data. The three main kinds of consistency checks are internal, temporal and spatial:

- Internal consistency tests express the physical relationships among different climatological elements. In some cases, they are logical tests based on the following premise: if a certain element exists in a given interval, another must also be contained in another interval [52].
- Temporal consistency tests are based on the persistence over time of climatological elements. Certain selected change thresholds depend on the variable in question, the period of the year and the climatic region to which the elements of the time series belong [53].
- Spatial consistency tests explore the smooth spatial variation of climatological variables. Generally, this type of test involves the estimation of a certain element based on neighbouring observations in the same climatic region [54]. The accepted limit of differences will depend on the type of variable, the climatic region and the distance

between the seasons. Therefore, the effectiveness of this type of test will depend on the availability of neighbouring stations [55].

The QCS used in this study is based on a series of tests, called Test Groups. The flowchart shown in Figure 3 details the step-by-step procedure through which all the datasets were analysed. Every column with information was carefully studied, from the column that contains the station's WMO identification code, to the columns that contain the data collection dates and the respective values of the meteorological variables.

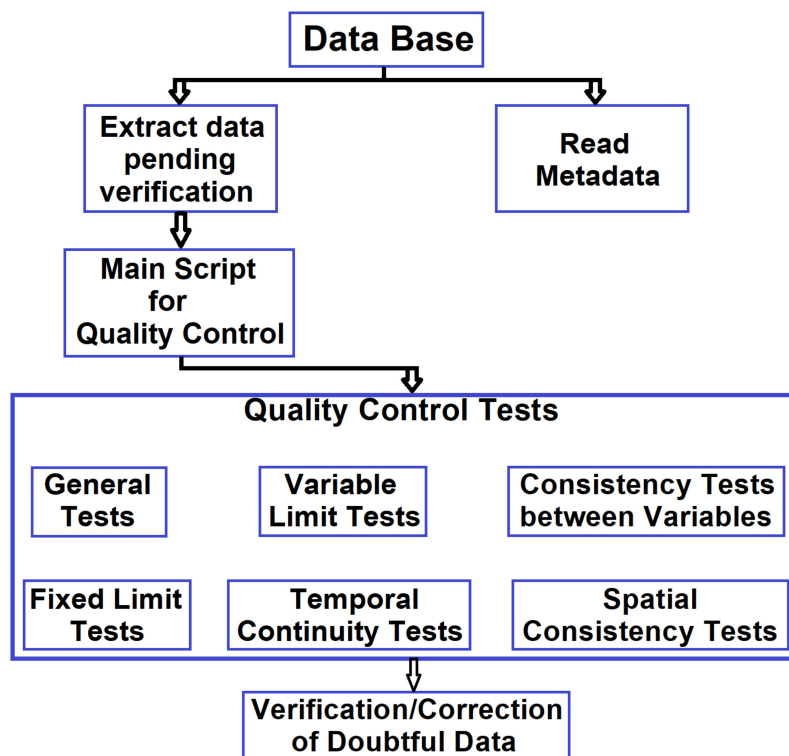


Figure 3. Description of the QCS evaluation routine, from extracting a series from the database to applying the tests and then checking for doubtful data and correcting them.

An important step is the creation of a file called metadata, which contains the station's basic information: the WMO identification code; the station's name; its longitude, latitude and altitude; the country and state to which it belongs; the start and end dates of its operations; the institution to which it belongs and the type of station, whether manual or automatic. In the QCS pre-processing step, this information contained in the metadata was read and served as the basis for some of the general tests shown in Figure 3.

In the final step of verifying and correcting doubtful data, in order to abbreviate the verification process, a routine identified cases of doubtful data that occurred in more than one test. If a given value was classified as suspicious in more than one of the QCS tests, then it was considered as incorrect data and did not need to be evaluated by the specialist, being summarily sent to a correction process.

3. Results and Discussion

3.1. Gap Filling

A high percentage of missing data was observed for the analysed variables. In individual terms, the station with the highest number of gaps had 54% missing data, and the lowest percentage found among the stations was 17%. In relation to a specific variable, the highest prevalence of gaps was found to be 62% and the lowest was 13%. Overall, the average percentage of missing data across all stations was 38%.

Gaps can occur in any segment of the series, so the imputation algorithm assigns "plausible" synthetic values according to the reference data, the predictors. In this way,

MICE keeps the original data in the series, only filling the gaps. The filling is processed on the daily scale, then 5 years of observed data are randomly chosen and have their data removed, and then MICE is applied again, imputing values that are later compared with real observations; this is the validation process. For the filling performed on the daily scale, the correlations ranged from 0.4 to 0.8 for temperature and the RMSE ranged from 0.9 to 1.9 °C; relative humidity showed correlations from 0.5 to 0.8 and a RMSE from 6.7 to 14.6%; atmospheric pressure exhibited correlations from 0.3 to 0.8 and a RMSE from 1 to 5 hPa; the average daily wind speed showed correlations between 0.2 and 0.7 and a RMSE between 0.8 and 1.9 m/s; for precipitation, correlations ranged from 0.5 to 0.9 and the RMSE ranged from 4 to 12 mm; insolation presented correlations from 0.1 to 0.7 and the RMSE ranged from 3 to 5 h.

Figure 4, in its upper panel, shows the daily (4a), 10-day (4b) and monthly (4c) correlations between the observed and imputed rainfall with respect to the aforementioned validation process. Correlations tended to increase with the accumulation interval [56]. The lower panel exhibits the RMSE, which also increased gradually for precipitation, ranging from 4 to 17 mm for daily values, from 12 to 37 mm for 10-day accumulated precipitation and from 22 to 72 mm for the monthly accumulated values. This behaviour is expected for precipitation, due to its accumulative nature, with similar results being observed in [4]. For all analysed variables, correlations increased with larger accumulation intervals, with the opposite being expected for RMSE [57]. This was observed for all variables whose 10-day and monthly values are the averages of these periods, the only exception being precipitation, for which the 10-day and monthly values are the result of the sum of the daily precipitations over these intervals. In this specific case, the RMSE rose in a proportional fashion, building from the highest observed values of the daily rainfall accumulation.

The results of the application of the same statistical techniques (correlation and RMSE) for the temperature variable are shown in Figure 5. Correlations gradually increased from the daily to the monthly timescales, exceeding 0.9 in some areas in central-south BA. Unlike precipitation, for which it was expected that errors would grow proportionally to the period of accumulation, for temperature, a reduction in the errors was expected, which is noticeable here in Figure 5d–f, in which there are areas with a maximum RMSE of almost 2 °C on the daily scale, dropping to 1.2 and 1.4 °C on the 10-day and monthly scales, with a minimum RMSE around 0.5 °C. For relative humidity (Figure 6), a similar behaviour to that of temperature was observed for both the correlations (Figure 6a–c) and the RMSE (Figure 6d–f). The area exhibiting the highest correlations is mostly located in the northern NEB, associated with the lowest RMSE values. In the centre-south of BA, the lowest correlations were observed and, in the west of BA, the highest RMSE values occurred, reaching up to 15% for the daily scale. For atmospheric pressure (Figure 7), moderate correlations were observed on the daily scale (Figure 7a), whereas the 10-day (Figure 7b) and monthly (Figure 7c) scales presented considerably higher values. The highest RMSE numbers were found in the west and south of BA, up to 5 hPa for the daily scale and up to 3.5 hPa for the monthly scale. The central north NEB showed the lowest values of RMSE, ranging from 1.5 to 3.5 hPa on the daily scale (Figure 7d), and from 0.5 to 2 hPa on the monthly scale (Figure 7f).

Wind speed (Figure 8) presented weak to moderate correlations on the daily scale, from 0.2 to 0.75 (Figure 8a), which gradually increased for the 10-day (Figure 8b) and monthly (Figure 8c) scales, with emphasis on the central portion of the NEB, where it exceeded 0.85 for some stations. Errors were greater on the daily scale (Figure 8d), being almost 2 m/s at some points, gradually decreasing, in an inversely proportional manner to the extent of the timescale, reaching minimum RMSE values of the order of 0.4 to 0.6 m/s in the areas of central south MA, eastern PI and mid-west BA. The northern NEB, between CE and RN, is where the highest RMSE values were observed. It should be noted that these are the areas of the NEB where the wind speeds are higher and several wind power plants have been installed.

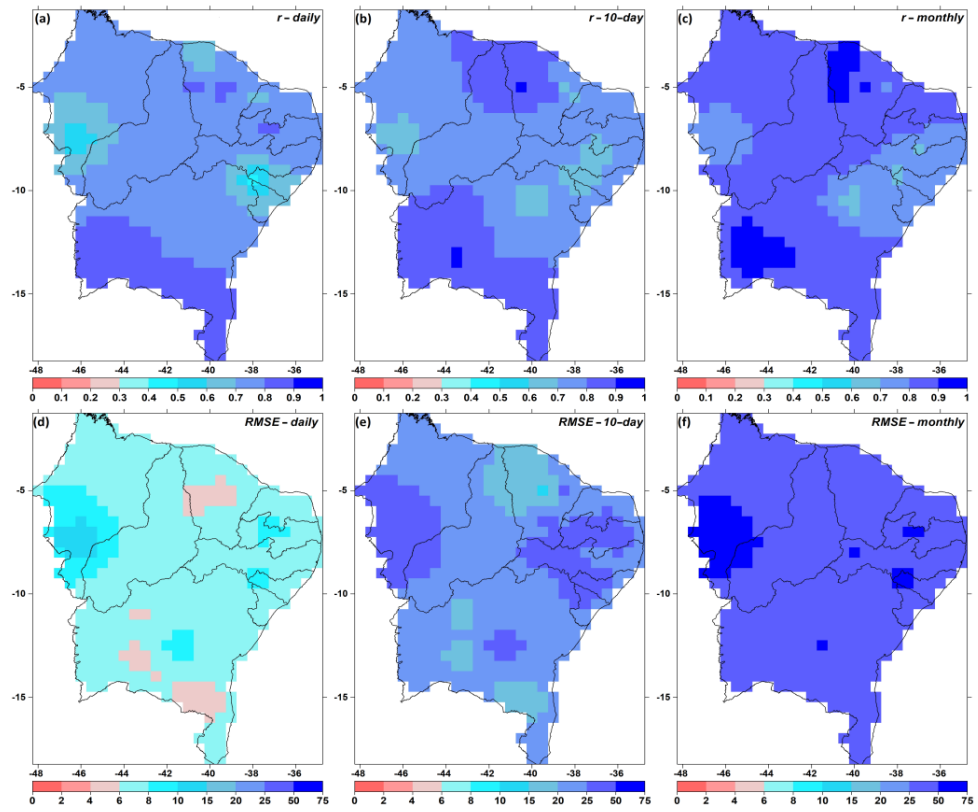


Figure 4. Correlation maps for precipitation between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (mm) on the daily (d), 10-day (e) and monthly (f) timescales.

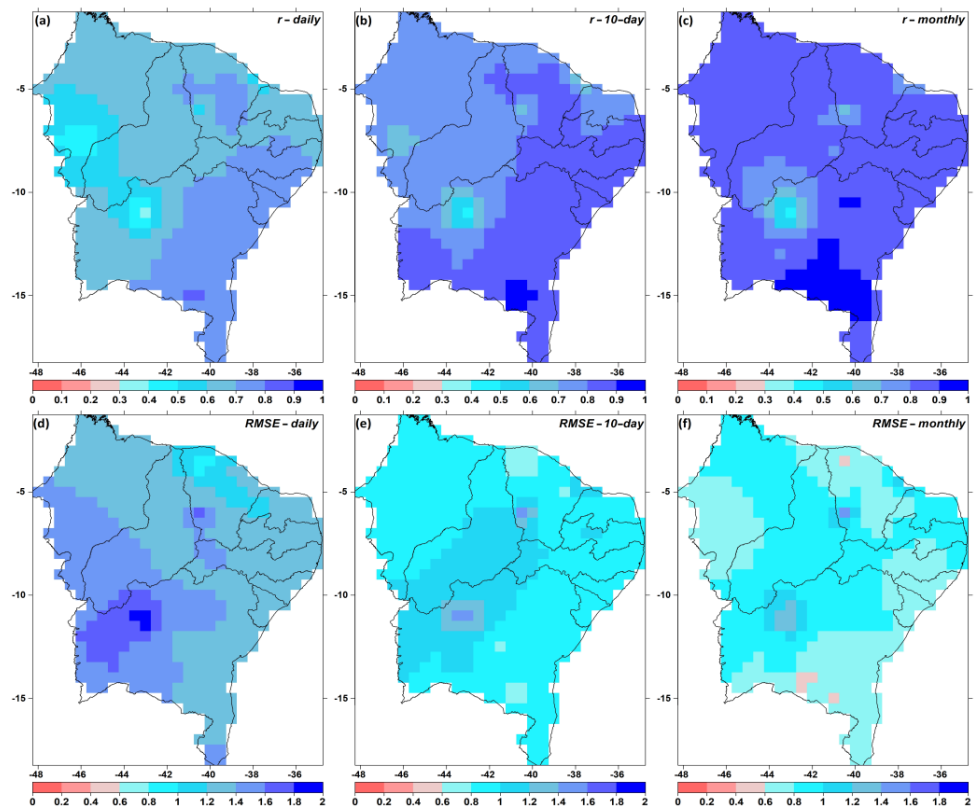


Figure 5. Correlation maps for temperature between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (°C) on the daily (d), 10-day (e) and monthly (f) time-scales.

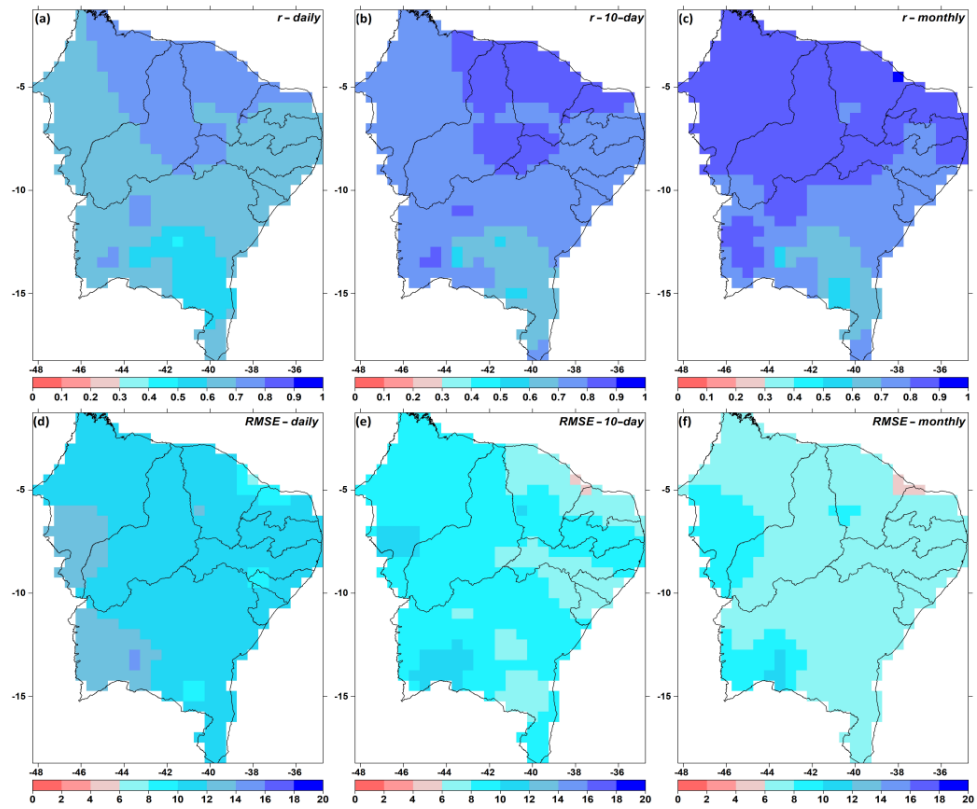


Figure 6. Correlation maps for relative humidity between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (%) on the daily (d), 10-day (e) and monthly (f) time-scales.

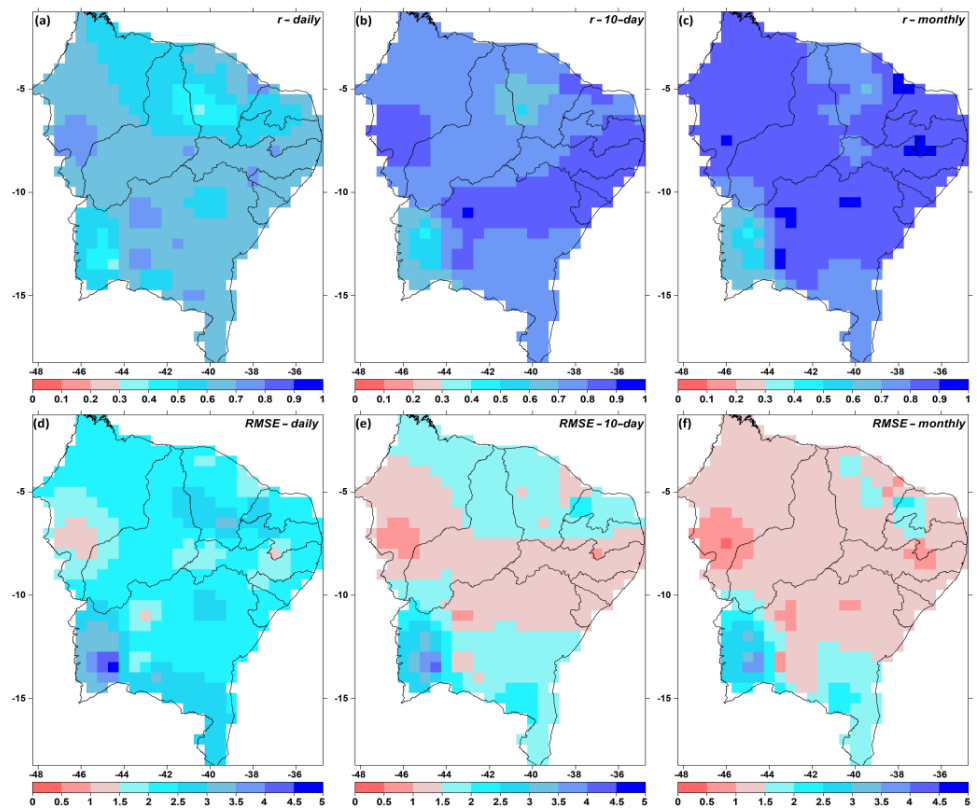


Figure 7. Correlation maps for atmospheric pressure between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (hPa) on the daily (d), 10-day (e) and monthly (f) time-scales.

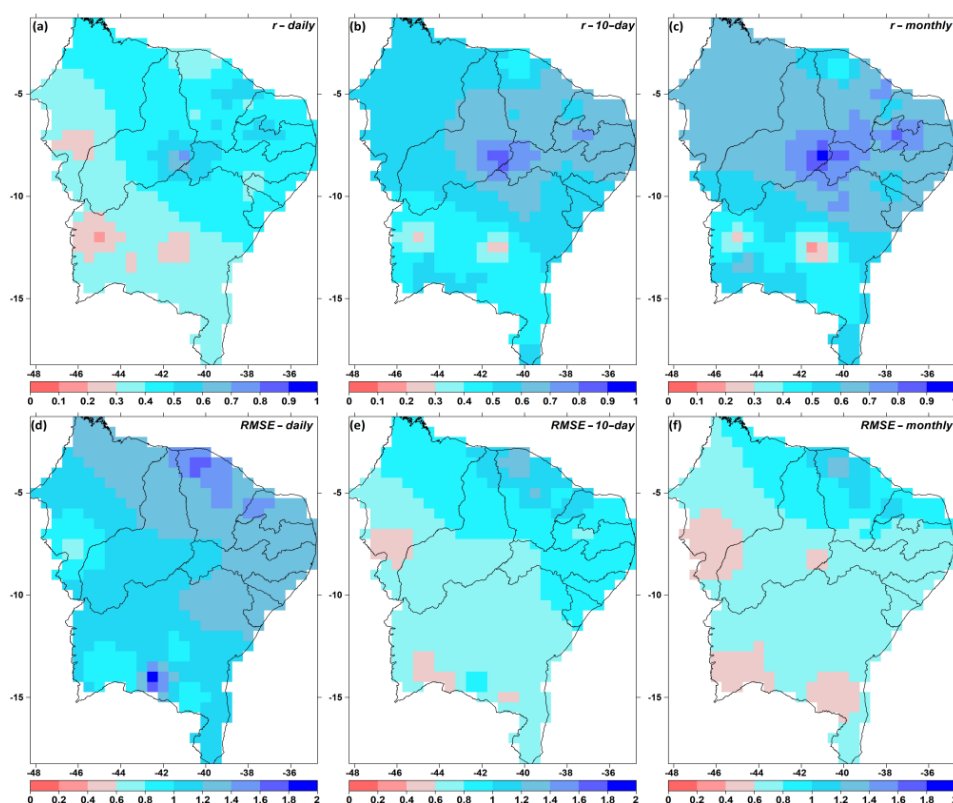


Figure 8. Correlation maps for wind speed between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (m/s) on the daily (d), 10-day (e) and monthly (f) time-scales.

In Figure 9, we show the results of the validation procedure applied to daily insolation data (Figure 9a) and the average over periods of 10 days (Figure 9b) and a month (Figure 9c). For the daily scale, the correlations are as low as 0.3 in most of the NEB, with values exceeding 0.6 in central south MA. However, on the 10-day and monthly scales, the correlations are quite high, especially in the northern and western NEB. The highest associated errors were found on the daily scale (Figure 9d), with a RMSE of up to 5 h in central south BA. The RMSE decreased for the 10-day (Figure 9e) and monthly (Figure 9f) scales, with the lowest values seen for the latter, in the order of 0.5 to 1.5 h, which were observed in central north NEB, up to 2 to 2.5 h in the south of BA.

A flawless and reliable observed database is essential for assessing the accuracy of climate change scenario model estimates, such as verifying monthly average temperature estimates [58], allowing for more accurate regionalized data using different methods of calibration [59]. The successful results obtained here in validation tests corroborate previous findings that regarded MICE as a robust gap filling technique with potential for application to different climatic variables. In recent years, the use of MICE as an efficient alternative for imputing missing data has been growing: [60] used the technique to impute missing solar radiation data under different atmospheric conditions, and [61] used the MICE for multivariate imputation and prediction of missing wind speed data on the decadal scale, whereas [32] showed that filling gaps in daily precipitation data from homogeneous climatic regions in Brazil with MICE was superior to kriging and ordinary cokriging methods [62]. The technique is widely used by healthcare professionals and researchers; in this context, [31] demonstrated its relevance and evolution in terms of available programs and software, which have facilitated its use, for example, by researchers in the field of psychiatry. At a time of continuous worldwide interest in past, present and future climatic conditions, our results, demonstrating the effectiveness of the technique, provide significant assistance in its popularization of its use in climate science.

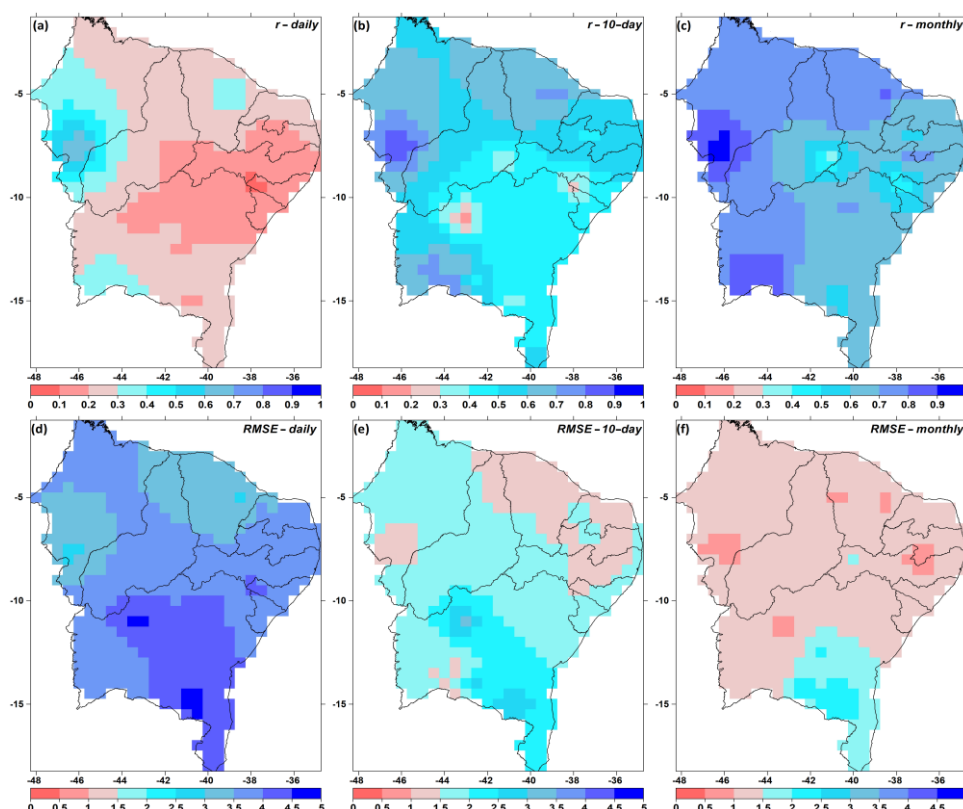


Figure 9. Correlation maps for insolation between imputed and observed data on the daily (a), 10-day (b) and monthly (c) timescales; and RMSE (h) on the daily (d), 10-day (e) and monthly (f) time-scales.

Figure 10 shows an example of filling gaps in daily rainfall data for a station located in the interior of the state of CE. The time series is divided into five 10-year groups, from 1961 to 2010. The original data distribution is shown on the left, with the missing values highlighted in red along the horizontal axis, whereas the right-hand panel exhibits the entire series composed of the preserved original data with the addition of imputed values from the filling technique. This is a good example that shows one of the main types of problems that arise during the analysis of a time series: long periods without any kind of information, as in the first 2 years of the series, 1961 and 1962, as well as between 1971 and 1972, and the large period with no observations from 1985 to 1994. Sparse short gaps are observed in other segments of the series. Obtaining complete series, after a reliable and validated filling scheme, is essential for, among other kinds of research, analysing the temporal evolution of extreme indices. The study of [5] used these complete series to verify the trends and tipping points associated with extreme precipitation and temperature indices in the NEB in order to fill gaps in studies that, due to an excessive occurrence of this issue, had to resort to fewer time series for analysing extremes in the NEB or were limited to the study of specific areas across the region [33,39–41,43,63,64].

3.2. QCS

After filling in the gaps, we present examples of results concerning the implementation of a QCS, in order to provide, for the NEB, daily data that are reliable and unburdened by missing information. The QCS is based on a series of tests, or Test Groups, following the flowchart presented in Figure 3. The examples are for a station bearing the WMO code of 82979, located in a municipality in the state of BA called Remanso, with the following geographic coordinates: -9.63° latitude, -42.10° longitude and an altitude of 400.5 m. The average percentage of filled gaps for this station was 30%.

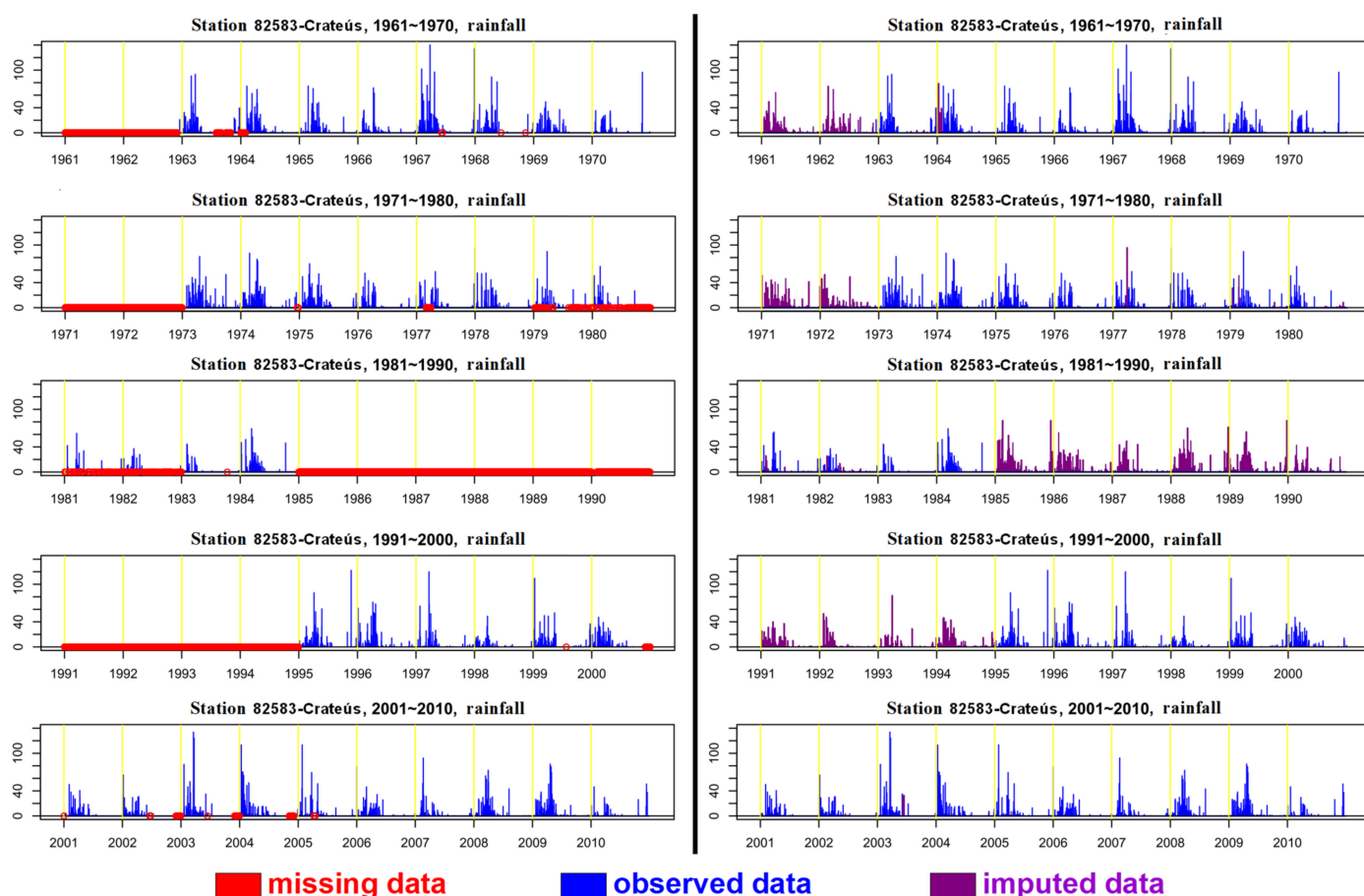


Figure 10. The original series from a station for every 10 years from 1961 to 2010, showing the gaps (left) and after these gaps had been filled (right). The code and name of the station, the subperiods and the variables are indicated in the graph. Originally missing data are indicated by the red colour, while blue implies observed data and violet indicates imputed data.

The first test, belonging to the group of general tests, checks if all the data from a station are associated with a single identification code (WMO code), that is, it certifies that there was no contamination by data from another station. The second group of tests, called fixed limit tests, is applied to precipitation, temperature, relative humidity and insolation. This test establishes the lower and upper limits for reasonable values, generally chosen according to historical records relative to a reference climate normal. In this case, we used the climate normals of Brazil for the period of 1961–1990 [65,66]. For this station, the limits for temperature were 10 °C and 41 °C (the minimum historical value recorded for daily minimum temperature and the maximum recorded value for daily maximum temperature, respectively). For this test, some errors were detected (Table 2) for atmospheric pressure (fixed limits of 955–975 hPa) and for insolation (fixed limits of 0–12 h).

Variable limit tests, which are part of the third group of tests, identify “extreme” residuals relative to a seasonal cycle, adjusted for the considered variable. Two percentile thresholds are defined (lower and higher), for example, 0.01 (i.e., the 1% percentile) and 0.99 (i.e., the 99% percentile). Extreme percentiles can be calculated with respect to each month or to the whole time series. Residuals less than the lower percentile and greater than the upper percentile are considered extremes. Taking the maximum temperature data from the Remanso station as an example, we found that, for the month of January, the 1% and 99% percentiles are, respectively, the values 25.7 °C and 36.6 °C. If we take all the values of the series, with no monthly distinction, the values of the 1% and 99% percentiles correspond, respectively, to 26.3 °C and 37.2 °C. Therefore, values that exceed

these thresholds will be considered extreme and doubtful, and will be reported in a printed output to be analysed manually, as shown in Table 3.

Table 2. Doubtful values found by the QCS through the fixed limit test for atmospheric pressure (in hPa) and insolation (in hours/day).

Year	Month	Day	Atmospheric Pressure (hPa)
1996	1	28	953.8
1996	8	7	977.7
1996	7	27	997.7
1996	3	20	1002.1
1996	3	21	982.5
1996	6	4	322.8
1996	6	5	0
1996	6	6	646.1
1996	6	14	975.8
2002	3	4	953.7
Year	Month	Day	Insolation (hours)
1998	8	25	25

Table 3. An excerpt of the time series of maximum temperature (in degrees Celsius), corresponding to the first 10 days of April 1961, with the doubtful values found by the QCS through the variable limits test highlighted in red. For April, the values of the 1% and 99% percentiles are, respectively, 26.6 °C and 36.4 °C.

Year	Month	Day	Maximum Temperature (°C)
1961	4	1	34.4
1961	4	2	30.6
1961	4	3	36.4
1961	4	4	37.0
1961	4	5	34.8
1961	4	6	37.2
1961	4	7	36.6
1961	4	8	36.0
1961	4	9	35.8
1961	4	10	33.6

The result of this test is also graphically presented in a box-plot graph. Figure 11 shows these results for relative humidity, maximum temperature, average temperature and minimum temperature, where the red dots refer to values that exceeded these variable limits based on the 1% and 99% percentiles. For precipitation, it was estimated that, for each month, the 95%, 97.5% and 99% percentiles would be used as limits to identify potential extremes of daily rainfall. These percentiles were calculated using the parameters of the gamma distribution adjusted for each month of the year.

The fourth group of tests, called temporal continuity tests, investigates two very common problems in daily datasets from manual weather stations: sequences of repeated values, and unjustified extremely high day-to-day discrepancies or “jumps” that occur in the data. The function of detecting sequences of repeated values is intended to find sequences from a defined limit, for example, from 3 days. To detect jumps or extreme discontinuities between values of a variable for consecutive days, a series was assembled with all the daily differences: the difference between 2 January 1961 and 1 January 1961, then the difference between 3 January 1961 and 2 January 1961, and so on. With these series of differences, the percentiles (e.g., 95%, 99% and 99.5%) of the absolute values of the differences were calculated, which were used as limits to define extreme jumps. If the absolute value (that is, a positive or negative difference) was greater than the percentile

used as a threshold, this jump was identified as extreme. In this study, the 99.5% percentile was used as the threshold.

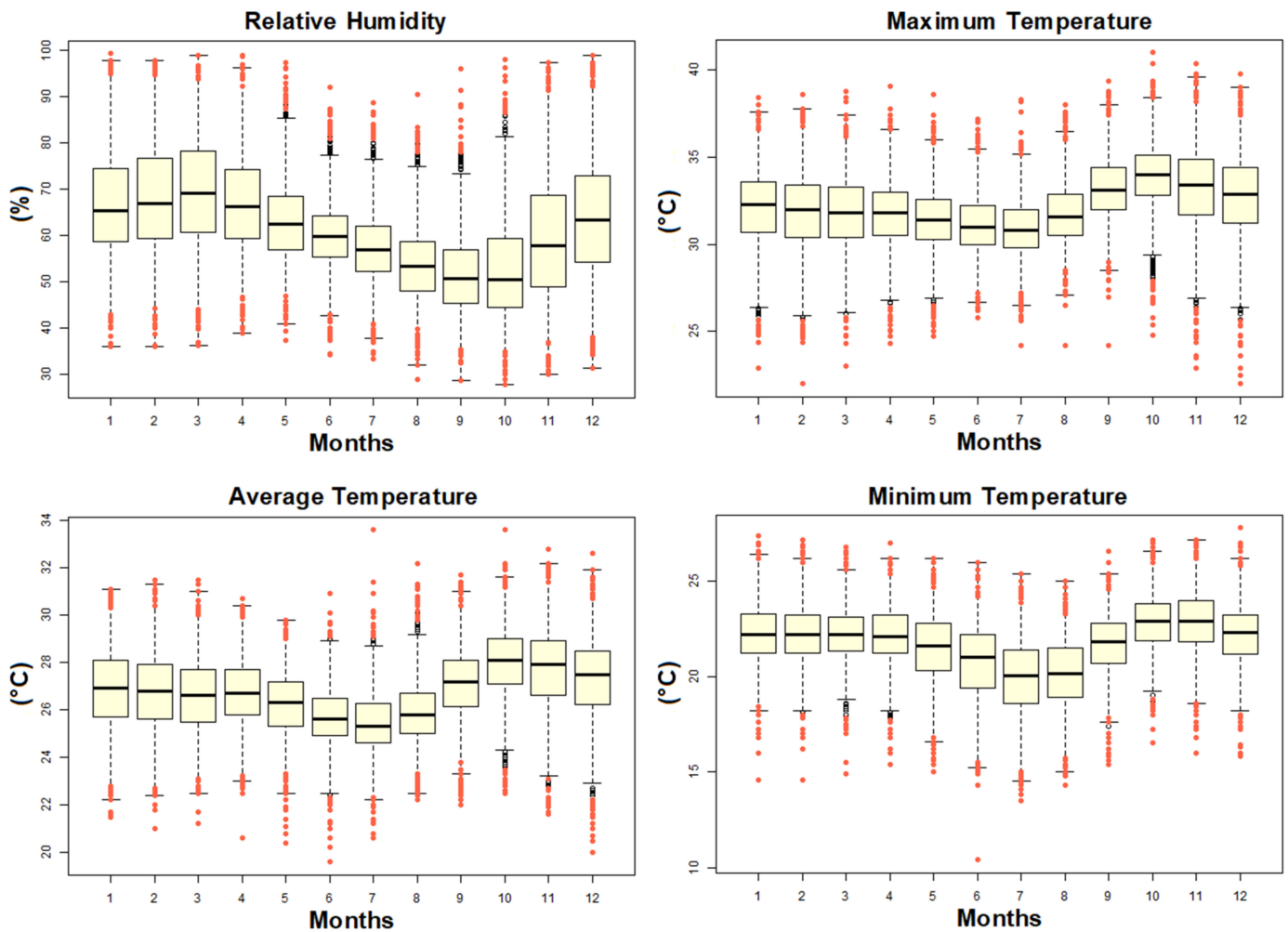


Figure 11. Variable limits test results for the 1% and 99% percentiles for the variables that are identified on the top of each graph. Red dots indicate the values that exceeded these thresholds.

The output of this test is printed and plotted, comprising the day, the variable and the doubtful value. Figure 12 shows these results, in two graphical forms, for relative humidity, maximum temperature, average temperature and minimum temperature, where the dots and/or red lines correspond to values that did not respect the limit of the 99.5% percentile of the differences and were therefore characterized as suspicious data.

In the fifth group of tests, the one concerning consistency tests between variables, the consistency is checked relative to the minimum, average and maximum temperatures. This test starts with the condition that the minimum temperatures must not be higher than the maximum temperatures, and that the average temperatures must lie between the daily minimum and maximum temperatures. Both printed and graphical outputs are generated. Suspicious data are analysed and then corrected, when the database has already incorporated the proper correction or rejected it, in which case the test is rerun until the imputed values satisfy the test conditions. Figure 13 shows the expected ideal condition: the left panel shows that all the maximum temperature values were higher than the average temperature, and the right panel shows that all average temperature values in the series were higher than the minimum temperature values.

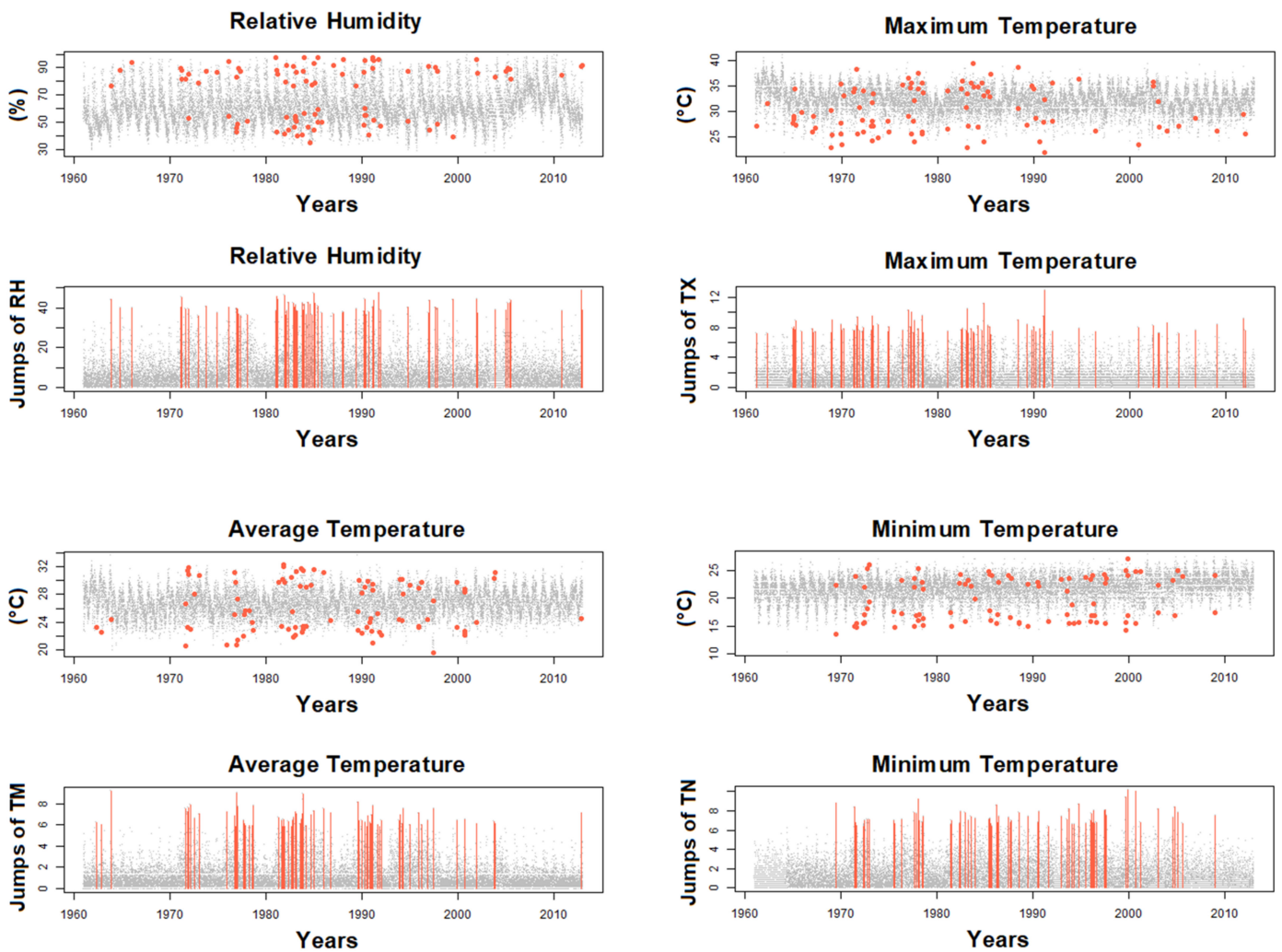


Figure 12. Results of the moving limits test for the 1% and 99% percentiles for relative humidity, and maximum, average and minimum temperatures. Red dots indicate values that exceeded these thresholds.

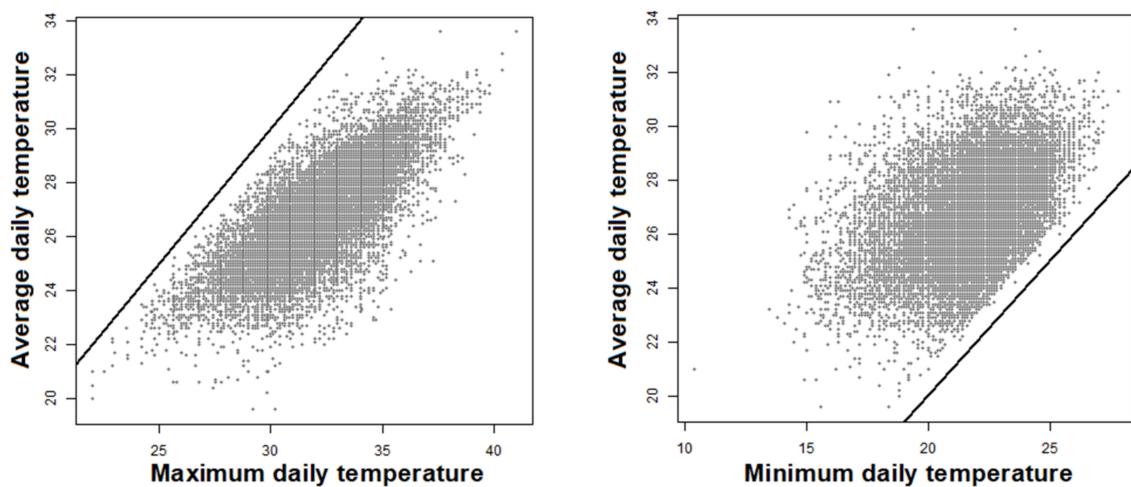


Figure 13. Results of the first consistency test between variables. The figure on the left shows that no average temperature value was higher than the daily maximum temperature, and the figure on the right shows that no average temperature value was lower than the daily minimum temperature. If any of the conditions failed due to suspicious data, these values would be identified as red dots to the left or to the right of the diagonal lines in the graphs.

The daily average temperature obtained for a weather station is called the compensated average temperature, as estimated by the following equation:

$$TM = [T(12 \text{ h}) + 2 \times T(24 \text{ h}) + TMX + TN]/5$$

where TM is the average daily compensated temperature, T(12 h) is the temperature observed at 12 h UTC, T(24 h) is the temperature observed at 24 h UTC, TX is the daily maximum temperature and TN is the daily minimum temperature.

Therefore, after checking with the first test, another way to identify any problems in the daily data relative to average temperatures is to compare the daily value directly with the daily average between the maximum temperature and the minimum temperature. For the execution of this test, a series of differences between the station's daily average temperature and the average obtained between the daily maximum and minimum temperatures is assembled. From this series of differences, the 99% percentile is calculated, which is the maximum tolerance threshold for the difference between the daily average temperature and the daily average between the maximum and minimum temperatures. The output of this test is both in printed and graphical forms. Figure 14 shows the results of this test.

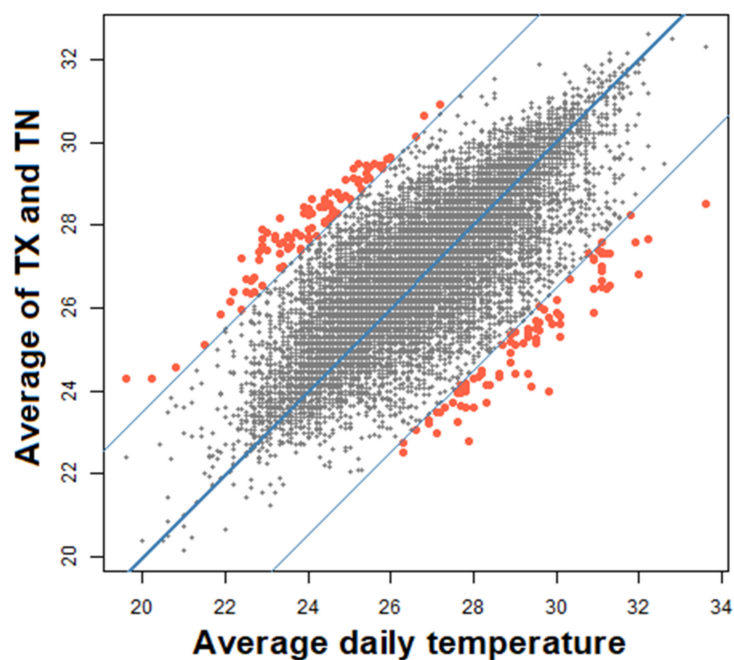


Figure 14. Result of the consistency test for the average daily temperature. The red dots represent values for which the difference between the daily average temperature and the average between maximum and minimum temperatures exceeded the 99% percentile.

4. Conclusions

In this study, the potential of the MICE technique to fill gaps in daily data from time series of meteorological variables collected from multiple MWS over the NEB was presented. The completed data were validated against observations, through correlations and RMSE, on three timescales: daily, 10-day and monthly. For all the variables, correlations increased with the number of days over which the accumulated values (of precipitation) or the average (for the other variables) were calculated.

Precipitation, followed by temperature, relative humidity and atmospheric pressure, were the variables for which the highest correlations were observed among all the compared temporal scales. On the daily scale, wind speed presented moderate correlations, and insolation showed weak correlations. However, the increase in correlations was significant, for all variables, in the 10-day and monthly comparisons. As expected for an accumulative

variable, the errors increased with the period of accumulated precipitation, whereas, for the other variables, the errors become gradually smaller on the 10-day and monthly scales.

The QCS is composed of strict criteria (specific tests) for identifying, in an automated manner, doubtful imputed data, although it allows adjustments by expert users, according to their knowledge of the local climate. Suspicious data can be kept if other tests and verifications allow them, such as in the case of spatial consistency tests that facilitate the comparison of similar suspicious occurrences at nearby stations. Otherwise, if more than one test indicates data inconsistency and the spatial analysis does not indicate close similarities, the doubtful value can be eliminated from the series, which will undergo as many filling procedures as necessary until a plausible synthetic value can occupy the place of the rejected data.

These results showed the efficiency of the technique for filling time series of meteorological variables, as well as that of the QCS. In the case of precipitation and temperature, both the filled and control/comparison datasets from this research were successfully used in studies of analyses of climatic extremes indices [5], and for statistical downscaling of regionalized climate change scenarios [4,45,59]. In the field of seasonal and subseasonal climate forecasting, these series will compose a database of surface observations for the calibration and verification of the Brazilian Global Atmospheric Model (BAM) [67], which is the atmospheric module of the Brazilian Earth System Model (BESM), aiming to achieve a hybrid dynamic–statistic coupling for the observed surface data and to perform adjustments in the BAM’s seasonal forecasting for the NEB.

Author Contributions: Conceptualization, R.L.C., H.B.G. (Heliofábio Barros Gomes) and F.D.d.S.S.; methodology, R.L.C., H.B.G. (Heliofábio Barros Gomes), R.L.d.R.J., F.D.d.S.S., D.D.C.P. and D.L.H.; software, R.L.C., R.L.d.R.J. and F.D.d.S.S.; validation, H.B.G. (Heliofábio Barros Gomes) and F.D.d.S.S.; formal analysis, D.D.C.P., H.B.G. (Heliofábio Barros Gomes), D.L.H., H.B.G. (Helber Barros Gomes), M.C.L.d.S. and F.D.d.S.S.; data curation, R.L.d.R.J. and F.D.d.S.S.; writing—original draft preparation, R.L.C., H.B.G. (Heliofábio Barros Gomes) and F.D.d.S.S.; writing—review and editing, R.L.C., D.D.C.P., H.B.G. (Helber Barros Gomes) and M.C.L.d.S.; visualization, H.B.G. (Heliofábio Barros Gomes), F.D.d.S.S. and D.D.C.P.; funding acquisition, F.D.d.S.S. and D.L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the following project of the Coordination for the Improvement of Higher Education Personnel (CAPES): CAPES/Modelagem#88881.148662/2017-01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The observational data used in this study is made available by INMET through their Meteorological Database (“Banco de Dados Meteorológicos do INMET”), at <https://bdmep.inmet.gov.br/> (accessed on 2 August 2021).

Acknowledgments: The first author thanks the Coordination for the Improvement of Higher Education Personnel (CAPES, acronym in Portuguese) for the financial support during the conception of this study. The authors wish to thank the National Institute of Meteorology (INMET) for making the data from their stations available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saurral, R.I.; Camilloni, I.A.; Barros, V.R. Low-frequency variability and trends in centennial precipitation stations in southern South America. *Int. J. Climatol.* **2016**, *37*, 1774–1793. [[CrossRef](#)]
2. Carvalho, L.M.V. Assessing precipitation trends in the Americas with historical data: A review. *WIREs Clim. Chang.* **2020**, *11*, e627. [[CrossRef](#)]
3. Sheffield, J.; Goteti, G.; Wood, E.F. Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *J. Clim.* **2006**, *19*, 3088–3111. [[CrossRef](#)]
4. Costa, R.L.; de Mello Baptista, G.M.; Barros Gomes, H.; dos Santos Silva, F.D.; da Rocha Júnior, R.L.; Nedel, A.S. Analysis of future climate scenarios for northeastern Brazil and implications for human thermal comfort. *An. Da Acad. Bras. De Ciências* **2021**, *93*, e20190651. [[CrossRef](#)]

5. Costa, R.L.; de Mello Baptista, G.M.; Barros Gomes, H.; dos Santos Silva, F.D.; da Rocha Júnior, R.L.; de Araújo Salvador, M.; Herdies, D.L. Analysis of climate extremes indices over northeast Brazil from 1961 to 2014. *Weather Clim. Extrem.* **2020**, *28*, 100254. [[CrossRef](#)]
6. Liebmann, B.; Allured, D. Daily precipitation grids for South America. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1567–1570. [[CrossRef](#)]
7. New, M.; Hulme, M.; Jones, P. Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **2000**, *13*, 2217–2238. [[CrossRef](#)]
8. Silva, V.B.S.; Kousky, V.E.; Shi, W.; Higgins, R.W. An improved gridded historical daily precipitation analysis for Brazil. *J. Hydrometeorol.* **2007**, *8*, 847–861. [[CrossRef](#)]
9. Xavier, A.C.; King, C.W.; Scanlon, B.R. Daily gridded meteorological variables in Brazil (1980–2013). *Int. J. Climatol.* **2016**, *36*, 2644–2659. [[CrossRef](#)]
10. Xie, P.; Arkin, P.A. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 2539–2558. [[CrossRef](#)]
11. Huffman, G.J.; Adler, R.F.; Morrissey, M.M.; Bolvin, D.T.; Curtis, S.; Joyce, R.; McGavock, B.; Susskind, J. Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.* **2001**, *2*, 36–50. [[CrossRef](#)]
12. Adler, R.F.; Huffman, G.J.; Chang, A.; Ferraro, R.; Xie, P.P.; Janowiak, J.; Rudolf, B.; Schneider, U.; Curtis, S.; Bolvin, D.; et al. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.* **2003**, *4*, 1147–1167. [[CrossRef](#)]
13. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503. [[CrossRef](#)]
14. Levizzani, V.; Bauer, P.; Turk, F.J. *Measuring Precipitation from Space: Eurainsat and the Future*; Springer Science & Business Media: Dordrecht, The Netherlands, 2007; Volume 28.
15. Becker, A.; Finger, P.; Meyer-Christoffer, A.; Rudolf, B.; Schamm, K.; Schneider, U.; Ziese, M. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data* **2013**, *5*, 71–99. [[CrossRef](#)]
16. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations: A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [[CrossRef](#)]
17. Tapiador, F.J. Measuring Precipitation from Space. In *Remote Sensing of Aerosols, Clouds, and Precipitation*; Islam, T., Hu, Y., Kokhanovsky, A., Wang, J., Eds.; Elsevier: New York, NY, USA, 2018; pp. 211–221.
18. Camargo, M.B.P.; Hubbard, K.G. Spatial and temporal variability of daily weather variables in sub-humid and semi-arid areas of the U.S. High Plains. *Agric. For. Meteorol.* **1999**, *93*, 141–148. [[CrossRef](#)]
19. WMO. *Guidelines on Climate Observation Networks and Systems*; WMO Technical Document; WMO: Geneva, Switzerland, 2003.
20. Guttman, N.V.; Quayle, R.G. A review of cooperative temperature data validation. *J. Atmos. Ocean. Technol.* **1990**, *7*, 334–339. [[CrossRef](#)]
21. Meek, D.W.; Hatfield, J.L. Data quality checking for single station meteorological databases. *Agric. For. Meteorol.* **1994**, *69*, 85–109. [[CrossRef](#)]
22. Thorne, P.W.; Allan, R.J.; Ashcroft, L.; Brohan, P.; Dunn, R.J.H.; Menne, M.J.; Pearce, P.R.; Picas, J.; Willett, K.M.; Benoy, M.; et al. Toward an Integrated Set of Surface Meteorological Observations for Climate Science and Applications. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2689–2702. [[CrossRef](#)]
23. Brugnara, Y.; Pfister, L.; Villiger, L.; Rohr, C.; Isotta, F.A.; Brönnimann, S. Early instrumental meteorological observations in Switzerland: 1708–1873. *Earth Syst. Sci. Data* **2020**, *12*, 1179–1190. [[CrossRef](#)]
24. Zhang, X.; Yang, F. *R-ClimDex (1.0) User Guide*; Climate Research Branch Environment Canada: Downsview, ON, Canada, 2004; 22p.
25. Lucas, E.W.M.; de Souza, F.D.A.S.; dos Santos Silva, F.D.; da Rocha Júnior, R.L.; Pinto, D.D.C.; da Silva, V.D.P.R. Trends in climate extreme indices assessed in the Xingu river basin—Brazilian Amazon. *Weather. Clim. Extrem.* **2021**, *31*, 100306. [[CrossRef](#)]
26. Santos, C.A.C.; Mariano, D.A.; Nascimento, F.C.A.; Dantas, F.R.C.; Oliveira, G.; Silva, M.T.; Silva, L.L.; Silba, B.B.; Bezerra, B.G.; Safa, B.; et al. Spatio-temporal patterns of energy exchange and evapotranspiration during an intense drought for drylands in Brazil. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *85*, 101982. [[CrossRef](#)]
27. Júnior, R.L.D.R.; dos Santos Silva, F.D.; Costa, R.L.; Barros Gomes, H.; Herdies, D.L.; Silva, V.D.P.R.D.; Xavier, A.C. Analysis of the Space–Temporal Trends of Wet Conditions in the Different Rainy Seasons of Brazilian Northeast by Quantile Regression and Bootstrap Test. *Geosciences* **2019**, *9*, 457. [[CrossRef](#)]
28. Júnior, R.L.D.R.; dos Santos Silva, F.D.; Costa, R.L.; Barros Gomes, H.; Pinto, D.D.C.; Herdies, D.L. Bivariate Assessment of Drought Return Periods and Frequency in Brazilian Northeast Using Joint Distribution by Copula Method. *Geosciences* **2020**, *10*, 135. [[CrossRef](#)]
29. Schafer, J.L.; Graham, J.W. Missing data: Our view of the state of the art. *Psychol. Methods* **2002**, *7*, 147–177. [[CrossRef](#)] [[PubMed](#)]
30. Van Buuren, S.; Groothuis-Oudshoorn, K. MICE: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [[CrossRef](#)]
31. Azur, M.J.; Stuart, E.A.; Frangakis, C.; Leaf, P.J. Multiple imputation by chained equations: What is it and how does it work? *Int. J. Methods Psychiatr. Res.* **2011**, *20*, 40–49. [[CrossRef](#)]

32. Carvalho, J.R.P.; Monteiro, J.E.B.A.; Nakai, A.M.; Assad, E.D. Model for Multiple Imputation to Estimate Daily Rainfall Data and Filling of Faults. *Rev. Bras. De Meteorol.* **2017**, *32*, 575–583. [[CrossRef](#)]
33. Costa, R.L.; Silva, F.D.S.; Sarmanho, G.F.; Lucio, P.S. Imputação Multivariada de Dados Diários de Precipitação e Análise de Índices de Extremos Climáticos. *Rev. Bras. De Geogr. Física* **2012**, *3*, 661–675. [[CrossRef](#)]
34. Greenland, S.; Finkle, W.D. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am. J. Epidemiol.* **1995**, *142*, 1255–1264. [[CrossRef](#)]
35. Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*; Wiley: New York, NY, USA, 1987.
36. Chen, M.; Shi, W.; Xie, P.; Silva, V.C.B.; Kousky, V.; Higgins, R.W.; Janowiak, J. Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.* **2008**, *113*, D04110. [[CrossRef](#)]
37. Gandin, L.S. *Objective Analysis of Meteorological Fields*; Israel Program for Scientific Translation: Jerusalem, Israel, 1965; 242p.
38. Aguilar, E.; Peterson, T.C.; Ramirez Obando, P.; Frutos, R.; Retana, J.A.; Solera, M.; Soley, J.; Gonzalez Garcia, I.; Araujo, R.M.; Rosa Santos, A.; et al. Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *J. Geophys. Res.* **2005**, *110*, D23107. [[CrossRef](#)]
39. Vincent, L.A.; Peterson, T.C.; Barros, V.R.; Marino, M.B.; Rusticucci, M.; Carrasco, G.; Ramirez, E.; Alves, L.M.; Ambrizzi, T.; Berlato, M.A.; et al. Observed trends in indices of daily temperature extremes in South America 1960–2000. *J. Clim.* **2005**, *18*, 5011–5023. [[CrossRef](#)]
40. Alexander, L.V.; Zhang, X.; Peterson, T.C.; Caesar, J.; Gleason, B.; Klein Tank, A.M.G.; Haylock, M.; Collins, D.; Trewin, B.; Rahimzadeh, F.; et al. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.* **2006**, *111*, D05109. [[CrossRef](#)]
41. Haylock, M.R.; Peterson, T.C.; Alves, L.M.; Ambrizzi, T.; Anunciação, Y.M.T.; Baez, J.; Barros, V.R.; Berlato, M.A.; Bidegain, M.; Coronel, G.; et al. Trends in total and extreme South American rainfall in 1960–2000 and links with sea surface temperature. *J. Clim.* **2006**, *19*, 1490–1512. [[CrossRef](#)]
42. Skansi, M.; Brunet, M.; Sigro, J.; Aguilar, E.; Groening, J.A.A.; Bentancur, O.J.; Geier, Y.R.C.; Amaya, R.L.C.; Jacome, H.; Ramos, A.M.; et al. Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Glob. Planet. Chang.* **2013**, *100*, 295–307. [[CrossRef](#)]
43. Bezerra, B.G.; Silva, L.L.; Santos e Silva, C.M.; Carvalho, G.G. Changes of precipitation extremes indices in Sao Francisco River basin, Brazil from 1947 to 2012. *Theor. Appl. Climatol.* **2019**, *135*, 565–576. [[CrossRef](#)]
44. Lima, C.I.S.; dos Santos Silva, F.D.; Freitas, I.G.F.; Pinto, D.D.C.; Costa, R.L.; Barros Gomes, H.; Silva, E.H.L.; Silva, L.L.; Silva, V.P.R.; Silva, B.K.N. Método Alternativo de Zoneamento Agroclimático do Milho para o Estado de Alagoas. *Rev. Bras. De Meteorol.* **2021**, *35*, 1057–1067. [[CrossRef](#)]
45. Dos Santos Silva, F.D.; Costa, R.L.; da Rocha Júnior, R.L.; Barros Gomes, H.; Vieira de Azevedo, P.; Rodrigues da Silva, V.d.P.; Monteiro, L.A. Cenários Climáticos e Produtividade do Algodão no Nordeste do Brasil. Parte II: Simulação Para 2020 a 2080. *Rev. Bras. De Meteorol.* **2020**, *35*, 913–929. [[CrossRef](#)]
46. Oliveira, L.P.M.; dos Santos Silva, F.D.; Costa, R.L.; da Rocha Júnior, R.L.; Barros Gomes, H.; Pereira, M.P.S.; Monteiro, L.A.; Rodrigues da Silva, V.d.P. Impacto das Mudanças Climáticas na Produtividade da Cana de Açúcar em Maceió. *Rev. Bras. De Meteorol.* **2020**, *35*, 969–980. [[CrossRef](#)]
47. Kane, R.P. Prediction of droughts in Northeast Brazil: Role of ENSO and use of periodicities. *Int. J. Climatol.* **1997**, *17*, 655–665. [[CrossRef](#)]
48. Hastenrath, S. Circulation and teleconnection mechanisms of Northeast Brazil droughts. *Prog. Oceanogr.* **2006**, *70*, 407–415. [[CrossRef](#)]
49. Shimizu, M.H.; Ambrizzi, T.; Liebmann, B. Extreme precipitation events and their relationship with ENSO and MJO phases over northern South America. *Int. J. Climatol.* **2017**, *37*, 2977–2989. [[CrossRef](#)]
50. Marengo, J.A.; Alves, L.M.; Alvalá, R.C.; Cunha, A.P.; Brito, S.; Moraes, O.L. Climatic characteristics of the 2010–2016 drought in the semiarid Northeast Brazil region. *An. Da Acad. Bras. De Cienc.* **2017**, *90*, 1973–1985. [[CrossRef](#)] [[PubMed](#)]
51. da Rocha Júnior, R.L.; Pinto, D.D.C.; dos Santos Silva, F.D.; Gomes, H.B.; Barros Gomes, H.; Costa, R.L.; Santos Pereira, M.P.; Peña, M.; dos Santos Coelho, C.A.; Herdies, D.L. An Empirical Seasonal Rainfall Forecasting Model for the Northeast Region of Brazil. *Water* **2021**, *13*, 1613. [[CrossRef](#)]
52. Gandin, L.S. Complex quality control of meteorological observations. *Mon. Weather. Rev.* **1988**, *116*, 1137–1156. [[CrossRef](#)]
53. Eischeid, J.K.; Baker, C.B.; Karl, T.; Diaz, H.F. The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteorol.* **1995**, *34*, 2787–2795. [[CrossRef](#)]
54. Hubbard, K.G.; Goddard, S.; Sorensen, W.D.; Wells, N.; Osugi, T.T. Performance of Quality Assurance Procedure for an Applied Climate Information System. *J. Atmos. Ocean. Technol.* **2005**, *22*, 105–112. [[CrossRef](#)]
55. You, J.K.; Hubbard, G.; Goddard, S. Comparison of methods for spatially estimating station temperatures in a quality control system. *Int. J. Climatol.* **2007**, *28*, 777–787. [[CrossRef](#)]
56. Silva, F.D.S.; Pereira Filho, A.J.; Hallak, R. Classificação de sistemas meteorológicos e comparação da precipitação estimada pelo radar e medida pela rede telemétrica na bacia hidrográfica do alto Tietê. *Rev. Bras. De Meteorol.* **2009**, *24*, 292–307. [[CrossRef](#)]
57. Hallak, R.; Pereira Filho, A.J. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: Sensibilidade a variações com os esquemas de advecção e assimilação de dados. *Rev. Bras. De Meteorol.* **2011**, *26*, 591–608. [[CrossRef](#)]

58. Carvalho, J.R.P.; Assad, E.D.; Pinto, H.S. Kalman filter and correction of the temperatures estimated by PRECIS model. *Atmos. Res.* **2011**, *102*, 218–226. [[CrossRef](#)]
59. Costa, R.L.; Barros Gomes, H.; dos Santos Silva, F.D.; de Mello Baptista, G.M.; da Rocha Júnior, R.L.; Herdies, D.L.; Rodrigues da Silva, V.d.P. Cenários de Mudanças Climáticas para a Região Nordeste do Brasil por meio da Técnica de Downscaling Estatístico. *Rev. Bras. De Meteorol.* **2020**, *35*, 785–801. [[CrossRef](#)]
60. Turrado, C.C.; López, M.D.C.M.; Lasheras, F.S.; Gómez, B.A.R.; Rollé, J.L.C.; Juez, F.J.D.C. Missing data imputation of solar radiation data under different atmospheric conditions. *Sensors* **2014**, *14*, 20382–20399. [[CrossRef](#)]
61. Wesonga, R. On multivariate imputation and forecasting of decadal wind speed missing data. *SpringerPlus* **2015**, *4*, 1–8. [[CrossRef](#)] [[PubMed](#)]
62. Carvalho, J.R.P.; Nakai, A.M.; Monteiro, J.E.B.A. Spatio-Temporal modeling of data imputation for daily rainfall series in Homogeneous Zones. *Rev. Bras. De Meteorol.* **2016**, *31*, 196–201. [[CrossRef](#)]
63. Silva, V.P.R. On climate variability in Northeast of Brazil. *J. Arid Environ.* **2004**, *58*, 575–596. [[CrossRef](#)]
64. Oliveira, P.T.; Santos e Silva, C.M.; Lima, K.C. Climatology and trend analysis of extreme precipitation in subregions of Northeast Brazil. *Theor. Appl. Climatol.* **2016**, *130*, 77–90. [[CrossRef](#)]
65. Ramos, A.M.; Santos, L.A.R.; Fortes, L.T. *Normais Climatológicas do Brasil 1961–1990*; INMET: Brasília, Brazil, 2009; 465p.
66. Diniz, F.A.; Ramos, A.M.; Rebello, E.R.G. Brazilian climate normals for 1981–2010. *Pesqui. Agropecuária Bras.* **2018**, *53*, 131–143. [[CrossRef](#)]
67. Figueroa, S.N.; Bonatti, J.P.; Kubota, P.Y.; Grell, G.A.; Morrison, H.; Barros, S.R.M.; Fernandez, J.P.R.; Ramirez, E.; Siqueira, L.; Luzia, G.; et al. The Brazilian Global Atmospheric Model (BAM): Performance for Tropical Rainfall Forecasting and Sensitivity to Convective Scheme and Horizontal Resolution. *Weather Forecast.* **2016**, *31*, 1547–1572. [[CrossRef](#)]