

# Networks Analysis Of Brazilian Climate Data Based On The DCCA Cross-Correlation Coefficient

Paulo Rodrigues (✉ [paulocanas@gmail.com](mailto:paulocanas@gmail.com))

Federal University of Bahia

Florêncio Oliveira Filho

SENAI CIMATEC University Center

Everaldo Guedes

Federal University of Recôncavo da Bahia

---

## Research Article

### Keywords:

**Posted Date:** February 28th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1330103/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Networks Analysis of Brazilian Climate Data based on the DCCA cross-correlation coefficient

Florêncio Mendes Oliveira Filho<sup>1</sup>, Everaldo Freitas Guedes<sup>2,3</sup>, and Paulo Canas Rodrigues<sup>4,\*</sup>

<sup>1</sup>Senai Cimatec University Center, Computer Engineering, Salvador, Brazil

<sup>2</sup>Federal University of Recôncavo Baiano, Exact and Technological Sciences Center, Cruz das Almas, Brazil

<sup>3</sup>Brazilian Hospital Services Company - Climério de Oliveira Maternity Hospital, Health Regulation and Evaluation Sector, Salvador, Brazil

<sup>4</sup>Federal University of Bahia, Department of Statistics, Salvador, Brazil

## ABSTRACT

Climate change is one of the most relevant challenges that the world has to deal with. A better understanding about the behaviour of environmental and atmospheric variables, their long range cross correlation and the way they relate to each other, can provide helpful insights about how the climate is changing. However, such studies are complex and rarely found in the literature, especially dealing with data from the Brazilian territory.

In this paper we consider four environmental and atmospheric variables, wind speed, radiation, temperature and humidity, measured in 27 stations (the capital of each of the 26 Brazilian states plus the federal district). We use the detrended fluctuation analysis to evaluate the statistical self-affinity of the signal, the cross correlation coefficient  $\rho_{DCCA}$  to quantify the long range cross correlation between stations, and a network analysis that considers the top 10%  $\rho_{DCCA}$  values for each atmospheric variable to better understand the cross correlation between stations.

The methodology used and combined in this paper represents a step forward in the field of time series and network analysis, that can be applied to other regions, other environmental variables and also to other fields of research. The results of the application are of great importance to better understand the behaviour of environmental and atmospheric variables in the Brazilian territory and to provide helpful insights about climate change and renewable energy production.

## Introduction

Due to its territorial immensity (8,511,965Km<sup>2</sup>), with 93% of the territory located in the Southern Hemisphere and 7% in the Northern Hemisphere, with a location in the intertropical zone of the planet, Brazil has different types of climates, such as: equatorial, tropical, highland tropical, humid tropical, semi-arid and subtropical. This distribution of climates, depending on the region, can have well-distributed rainfall (e.g. the South region of Brazil), high temperatures throughout the year, with insufficient and poorly distributed rainfall (e.g. the Northeast region of Brazil), high temperature, high humidity and low thermal amplitude, with high levels of rainfall (e.g. the North region of Brazil, where the Amazon forest is located), mountain and plateau regions (e.g. the Southeast region of Brazil, which is characterized by low thermal amplitude and average temperature oscillating between 17°C and 22°C)<sup>1-4</sup>.

In countries with a vast territory, studies with climatic variables such as radiation, temperature, humidity and wind speed have been studied to locate sandstorms in Asia<sup>5</sup>. This study explored the relationship between dust transported from various sources and oceanic biological activities with different nutrient conditions, and explores signs of global warming. In that study, climatic variables were investigated to explain hotter summers, frequent heavy rains, prolonged droughts, more severe forest fires, violent storms (blizzards), cyclones, as well as the melting of the polar ice caps<sup>6</sup>. In Brazil, the scalable behavior of wind speed along the coast was analyzed in order to determine long-range correlations and acquire more information about cross-sectional behavior at various scales. It was also investigated in Brazil with a focus on sustainable energy, variables such as radiation, temperature and humidity, taking into account cross-correlation and multiple cross-correlation, by considering three stations located in the Brazilian state of Bahia<sup>7</sup>. Research with climate variables has also investigated significant energy sources in terms of their capacity, reliability, cost, safety and effects on the environment<sup>8</sup>. For both studies the concepts of auto-correlation and cross-correlation were applied to understand behavior, trends and markings.

The concept of networks associated with climate variables has gained prominence in recent years. One study combines species distribution models with ecological network analysis to test the potential impacts of climate change on more than 700 plant and animal species in Central European pollination and seed dispersal networks<sup>9</sup>. Networks have also been applied in

studies to assess the impact of climate change on groundwater resources for regions without a pumping well. The network demonstrated the impact of climate change on average spring discharge. Discharge increased by about +8% in spring and summer and decreased by about -7% in autumn and winter<sup>10</sup>. Network theory has also been used for decision making process in agriculture to effectively predict Fujian rice production under typical mountainous region climatic conditions and evaluate the performance of the network model in relation to variations in development parameters, comparing the effectiveness of multiple linear regression models<sup>11</sup>, to study soil water loss by evaporation and plant water loss by transpiration<sup>12</sup>. Regardless of the type of network, application and/or improvement of models expanded around the world and across scientific domains.

Despite its vast territory and the importance of a better understanding of the behaviour that climate variables such as wind speed, radiation, temperature and humidity have in terms of climate change and renewable energy production, to the best of our knowledge, there is no study that provides a portrait of the Brazilian regions using long range cross correlation and network analysis.

In this paper we consider four important environmental variables: wind speed, radiation, temperature and humidity, measured in 27 stations (the capital of each of the 26 Brazilian states plus the federal district). We use the detrended fluctuation analysis<sup>13</sup> to evaluate the statistical self-affinity of the signal, and use the cross correlation coefficient  $\rho_{DCCA}$ <sup>14</sup> to quantify the long range cross correlation between Brazilian regions. Then, a network analysis is made for each environmental variable, where the top 10% values of  $\rho_{DCCA}$  are highlighted and discussed.

The rest of the paper is organized as follows. Next section includes the materials and methods, where the data collection and organization is described, together with the methodological details of the detrended fluctuation analysis, DCCA cross-correlation coefficient and network analysis. Then, the results for the detrended fluctuation analysis, DCCA cross-correlation analysis and network analysis are discussed. The paper ends with some concluding remarks.

## Materials and Methods

### Data collection and data organization

The data about the four environmental and atmospheric variables (global radiation, temperature, humidity and wind speed) for all 27 Brazilian environmental stations (26 state capitals plus the federal district; Table 1) was extracted from the database of the Instituto Nacional de Pesquisas Espaciais - INPE (<https://www.gov.br/inpe/pt-br>).

In this study, 13 hourly observations (daily local time between 6am to 6pm; UTC 9am to 9pm), between January 1, 2010, and December 31, 2020, (4018 days; 52234 observations) were considered. Knowing that the original data is available per hour and per year, for each of the variables studied, we joined the databases to compile the final time series data. The percentages of missing values in all 112 time series varied between 0.19% and 47.77%, being the missing values imputed by the mean of that hour along the days. The complete list of percentages of missing values are available in Table 1.

### Detrended fluctuation analysis

Peng et al. (1994)<sup>13</sup> developed the detrended fluctuation analysis (DFA) to analyse the existence of serial dependence (the statistical self-affinity of a signal), with the advantage of being also possible to be used in non-stationary data. Its main advantage is to avoid spurious detection of long-range dependence due to nonstationary data<sup>15-19</sup>. For a given “Y” signal, the algorithm is described as follows:

1. Define the the cumulative series of the original signal fluctuations around its mean:

$$X_N = \sum_{i=1}^N (Y_i - \bar{Y}) \quad (1)$$

2. Divide  $X_N$  into boxes of equal length  $n$ .
3. In each box, fit the local trend of  $X_n$  by a polynomial of degree  $(n)$ , which represents the local trend of the box.
4. For the given  $n$  box size, compute the root-mean-squared detrended fluctuation of the signal  $X_N$  as:

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - P(n))^2} \quad (2)$$

5. The last step is repeated for each of the available  $\tau$  scales (box size) to obtain the empirical relationship between the overall fluctuation  $F(n)$  and the box size  $n$ :

$$F(n) \propto n^{\alpha_{DFA}} \quad (3)$$

**Table 1.** Region, state, state capital, latitude, longitude, altitude and percentage of missing values for each atmospheric variable, for all 27 stations. Moer details about the geographical location of each station can be found on the map in the Figure 1S of the supplementary material.

Region	State	City (state capital)	Latitude	Longitude	Altitude	Percentage of missing values			
						Radiation	Temperature	Humidity	Wind speed
North	PA	Belém	-1.41	-48.44	24.00	8.65	1.18	1.20	1.21
North	RR	Boa Vista	2.82	-60.69	94.00	23.44	12.13	12.13	12.21
North	AP	Macapá	3.81	-51.86	21.00	36.95	32.19	32.20	32.23
North	AM	Manaus	-3.10	-60.02	61.25	11.98	1.77	1.90	8.58
North	TO	Palmas	-10.18	-48.30	280.00	11.12	5.20	5.20	5.20
North	RO	Porto Velho	-8.76	-63.47	98.00	32.22	24.76	24.77	47.77
North	AC	Rio Branco	-9.96	-68.17	220.00	44.81	37.00	42.19	37.03
Northeast	SE	Aracaju	-10.95	37.05	4.72	1.90	1.33	12.02	12.68
Northeast	CE	Fortaleza	-3.83	-38.54	26.45	24.01	17.83	17.83	18.05
Northeast	PB	João Pessoa	-7.17	-34.82	47.00	11.15	11.04	11.89	10.59
Northeast	AL	Maceió	-9.55	-35.77	80.00	8.88	6.04	6.08	6.05
Northeast	RN	Natal	-5.90	-35.20	48.60	5.38	1.46	1.46	7.59
Northeast	PE	Recife	-8.06	-34.96	10.00	2.57	1.56	4.14	13.58
Northeast	BA	Salvador	-13.02	-38.52	51.41	3.35	1.95	1.95	1.98
Northeast	MA	São Luís	-2.53	-44.21	56.00	11.79	4.76	6.96	4.81
Northeast	PI	Teresina	-5.07	-42.81	74.36	9.52	7.90	7.90	7.91
Central-West	DF	Brasília	-15.79	-47.93	1159.54	6.20	0.35	0.35	0.89
Central-West	MS	Campo Grande	-20.45	-54.60	530.00	18.52	3.95	4.02	3.96
Central-West	MT	Cuiabá	-15.56	-56.73	240.00	26.67	13.14	13.14	16.93
Central-West	GO	Goiania	-16.64	-49.22	770.00	5.74	0.25	0.29	0.25
Southeast	MG	Belo Horizonte	-19.88	-43.97	869.00	4.73	0.21	0.21	0.19
Southeast	RJ	Rio de Janeiro	-22.99	-43.19	42.00	5.89	1.04	4.70	1.47
Southeast	SP	São Paulo	-23.48	-46.62	792.06	6.28	0.65	0.48	0.48
Southeast	ES	Vitória	-20.27	-40.30	90.00	6.86	4.01	5.22	4.02
South	PR	Curitiba	-25.43	-49.27	923.50	8.98	4.59	11.83	4.59
South	SC	Forianópolis	-27.60	-48.62	1.80	8.39	0.30	0.30	0.31
South	RS	Porto Alegre	-30.05	-51.17	46.97	8.10	0.39	0.39	1.33

The scaling exponent  $\alpha_{DFA}$  quantifies the empirical strength of the long-range power-law correlations of the signal. If the signal is not random, it is characterized by long-range correlation features. The  $\alpha_{DFA}$  can be interpreted as shown in Table 2<sup>20,21</sup>.

**Table 2.** Interpretation of the  $\alpha_{DFA}$ , a self-affinity parameter representing the long-range power-law correlation of the signal of a time series.

Exponent	Type of signal
$\alpha_{DFA} < 0.5$	Antipersistent
$\alpha_{DFA} \simeq 0.5$	Uncorrelated, white noise
$\alpha_{DFA} > 0.5$	Long-range correlated persistent
$\alpha_{DFA} \simeq 1.0$	1/f noise
$\alpha_{DFA} > 1.0$	Non-stationary
$\alpha_{DFA} \simeq \frac{3}{2}$	Brownian noise

The advantages of the DFA over methods are that it permits the detection of long-range correlations embedded in seemingly nonstationary time series, and also avoids the spurious detection of apparent long-range correlations, which are an artefact of nonstationarity.

### The DCCA cross-correlation coefficient

The DCCA cross correlation coefficient proposed by Zebende (2011)<sup>14</sup> is centred on the ratio between the detrended covariance function  $F_{DCCA}^2$ <sup>22</sup> and the variance function  $F_{DFA}$ <sup>13</sup>, as showed in the following five steps:

1. Consider two time series,  $\{x_i\}$  and  $\{y_i\}$ ,  $i = 1, 2, \dots, N$ . These time series are integrated to obtain two new time series  $\{X_k\}$  and  $\{Y_k\}$ :

$$X_k = \sum_{i=1}^k (x_i - \langle x \rangle) \quad \text{and} \quad Y_k = \sum_{i=1}^k (y_i - \langle y \rangle), \quad (4)$$

where  $\langle x \rangle$  and  $\langle y \rangle$  represent the average of each time series.

2. These two integrated time series,  $\{X_k\}$  and  $\{Y_k\}$ , are divided into  $(N - n)$  overlapping boxes of equal length  $n$ , in which  $n$  varies between 4 and  $\frac{N}{4}$ .
3. The local trend for each box is calculated by using a least-squares fit of each series,  $\{\tilde{X}_{k,i}\}$  and  $\{\tilde{Y}_{k,i}\}$ , and the covariance of the residuals is calculated for each box:

$$f_{DCCA}^2(n, i) = \frac{1}{(n+1)} \sum_{k=i}^{i+n} (X_k - \tilde{X}_{k,i})(Y_k - \tilde{Y}_{k,i}). \quad (5)$$

The variance of the residuals is also computed by using:

$$f_{DFA_x}^2(n, i) = \frac{1}{(n+1)} \sum_{k=i}^{i+n} (X_k - \tilde{X}_{k,i})^2 \quad (6)$$

$$f_{DFA_y}^2(n, i) = \frac{1}{(n+1)} \sum_{k=i}^{i+n} (Y_k - \tilde{Y}_{k,i})^2 \quad (7)$$

4. The average over all  $(N - n)$  overlapping boxes is calculated to obtain the detrended covariance function:

$$F_{DCCA}^2(n) = \frac{1}{N-n} \sum_{i=1}^{N-n} f_{DCCA}^2(n, i), \quad (8)$$

and the detrended variance functions:

$$F_{DFA_x}(n) = \sqrt{\frac{1}{N-n} \sum_{i=1}^{N-n} f_{DFA_x}^2(n, i)} \quad (9)$$

$$F_{DFA_y}(n) = \sqrt{\frac{1}{N-n} \sum_{i=1}^{N-n} f_{DFA_y}^2(n, i)} \quad (10)$$

5. Finally, the cross-correlation coefficient is computed as:

$$\rho_{(X_i, X_j)}(n) = \frac{F_{DCCA}^2(n)}{F_{DFA_x}(n) F_{DFA_y}(n)}, \quad (11)$$

where  $\rho_{(X_i, X_j)}(n)$  ranges from  $-1 \leq \rho_{(X_i, X_j)}(n) \leq 1$ , in which:

- i.  $\rho_{(X_i, X_j)}(n) = 1$  means a perfect cross-correlation;
- ii.  $\rho_{(X_i, X_j)}(n) = 0$  there is no cross-correlation; and
- iii.  $\rho_{(X_i, X_j)}(n) = -1$  means a perfectly anti cross-correlation.

It is important to notice that  $\rho_{(X_i, X_j)}$  depends on  $n$  (time scale). So, one of the advantages of this detrended cross-correlation coefficient is to measure cross-correlations between two non-stationary time series at different time scales<sup>23–28</sup>.

### Network analysis

A graph is a mathematical structure used to model the pairwise relations between objects. It includes vertices/nodes/points connected by edges/links/lines. There are two main types of graphs: (i) undirected graphs, where edges link two vertices without a specific direction; and (ii) directed graphs, where edges link two vertices in a specific direction. Formally, a graph can be defined as an ordered pair  $G = (V, E)$ , being  $V$  a set of vertices and  $E \subseteq \{\{x, y\} | x, y \in V, x \neq y\}$  a set of edges that connects two distinct edges  $x$  and  $y$ . Network theory is related to the study of graphs in terms of symmetric or asymmetric relations between objects.

In this study, due to the nature of the variables, we consider undirected graphs and network analysis with symmetric relations between objects (in this case, the locations of the environmental variables). To conduct the network representation and analysis, we consider the DCCA cross-correlation coefficient between pairs of the Brazilian state capitals for each of the four environmental variables, temperature, humidity, radiation and wind speed, for a short median and long distance correlations (box values of 13–one day, 1261–97 days, and 3328–256 days). The graph, based on the DCCA cross-correlation coefficients, was created using the function `graph_from_adjacency_matrix` of the R package `igraph`, and the network plot was obtained using the function `ggnet2` of the R package `GGally`. To avoid overcrowded networks, we decided to depict only edges associated with the top 10% DCCA cross-correlation coefficients between objects.

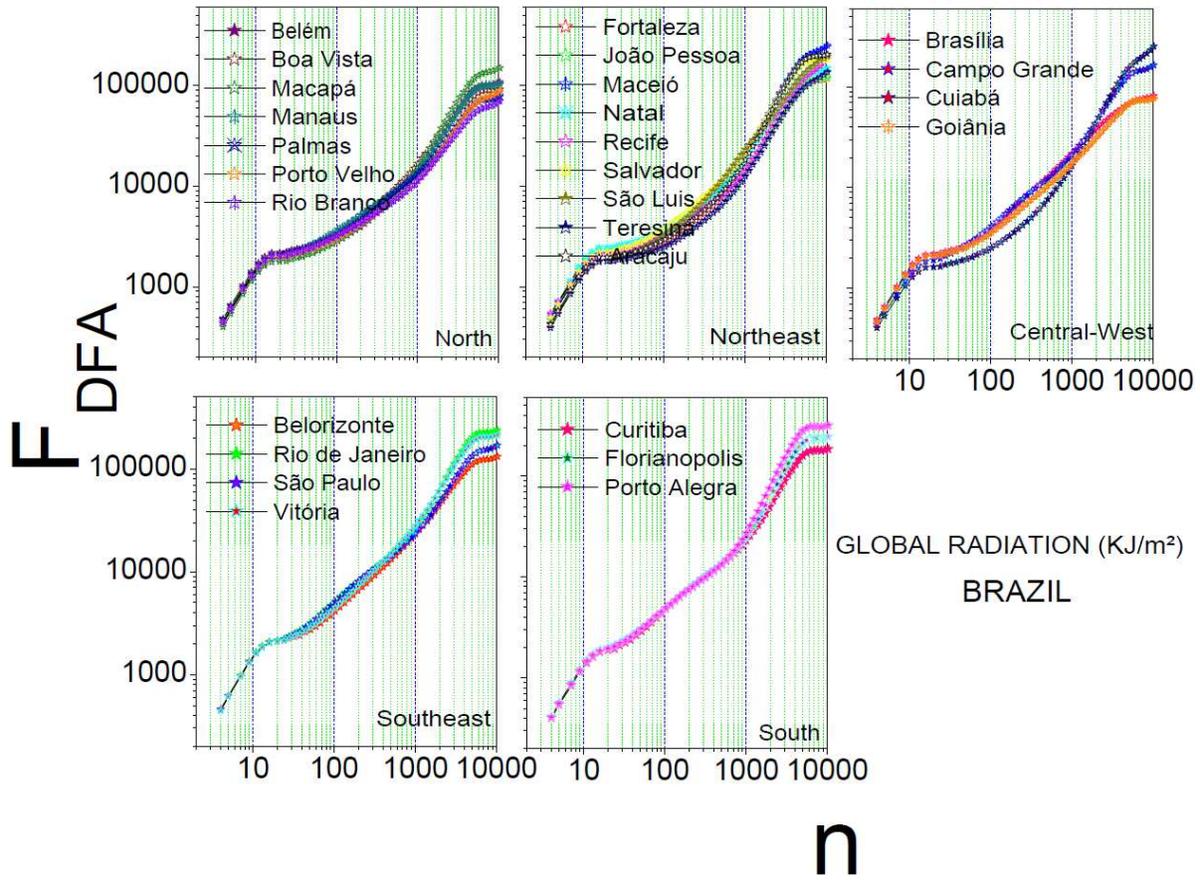
### Results and discussion

In this section we will include some of the main results and their discussion of the analysis of Brazilian climate data per state capital, for the detrended fluctuation analysis, DCCA cross-correlation analysis and network analysis. Further tables and figures can be found in the supplementary material. General descriptive statistics for global radiation, temperature, air humidity and wind speed, can be found in Tables 1S-4S of the supplementary material.

#### Detrended fluctuation analysis (DFA)

We considered time series with 52.234 data points (hours), generated by the observation of 4.018 days (1/01/2010 - 31/12/2020), with 13 observations per day (daily local time between 6am to 6pm; UTC 9am to 9pm), for the variables radiation in  $KJ/m^2$ , maximum air temperature in  $^{\circ}C$  (dry bulb), air humidity in %, and wind speed in  $m/s$ , and for the 26 Brazilian state capitals and the Brazilian capital, for a total of 108 time series. We then calculate the fluctuations of the time series with the fluctuation function  $F_{DFA}(n)$ . In this preliminary phase of the analysis, the objective was to evaluate the self-correlation of the variables in the Brazilian state capitals. The locations of the Brazilian states and regions can be seen in the supplementary material, Figure 1S and the names of the state capitals with their location in Table 1.

Figure 1 shows the behaviour of the fluctuation function  $F_{DFA}(n)$  for global radiation in all state capitals per region. A characteristic pattern is observed for all time scales regardless of the region, with three clearly defined behaviours: (i)  $n \leq 13$  (1 day and 6 scale observations: boxes with length 4, 5, 7, 9, 11, 13); (ii)  $13 < n \leq 5625$  (1 - 430 days, with 48 scale observations: boxes with length between 16 and 5625); and (iii)  $n > 5625$  (430 days and 8 scale observations: boxes with length 6242, 6925, 7680, 8514, 943, 10457, 11585, 12831, 4327). We can also compute the  $\alpha_{DFA}(medium) \cong 1.21(\pm 0.04)$  for up to 1 day of observations, the  $\alpha_{DFA}(medium) \cong 0.87(\pm 0.02)$  for intervals between 1 and 430 days, and the  $\alpha_{DFA}(average) \cong 0.06(\pm 0.01)$  for intervals greater than 430 days.  $\alpha_{DFA}(medium) \cong 1.21$  represents a non-stationary persistent behavior,  $\alpha_{DFA}(medium) \cong 0.87$  represents a long range persistence, and  $\alpha_{DFA}(medium) \cong 0.06$  represents an anti-persistent behavior.



**Figure 1.** Detrended fluctuation analysis for global radiation. The plots show, respectively the curves for all capitals in the regions North, Northeast, Central-West, Southeast and South, respectively. The vertical axis give the  $F_{DFA}$  and the horizontal axis show the box size.

With a behaviour very similar to the global radiation, temperature (see Figure 2S) also presents a specific pattern for  $n \leq 13$ ,  $13 < n \leq 5.625$  and  $n > 5.625$ . Only the state capitals in the Southeast and South regions showed a less intense transition in the trend associated to observations with  $n \leq 13$ , (1 day). Humidity (see Figure 3S) in the state capitals for the Southeast and South regions and wind speed (Figure 4S) in the state capitals for the Central-West, Southeast and South regions did not show the transition to  $n \leq 13$ , showing different characteristics for the state capitals in the North and Northeast regions.

Based on the analysis of the detrended fluctuation analysis of all environmental variables in all Brazilian state capitals, it can be concluded that the non-stationary and persistent behaviour ( $\alpha_{DFA}(medium) \cong 1.21$ ;  $n \leq 13$ ), upward and/or downward trends are time dependent. In the long range persistent behavior ( $\alpha_{DFA}(medium) \cong 0.87$ ;  $13 < n \leq 5.625$ ), it is observed that variables with closer values remain for a longer time influencing the memory. As for the anti-persistent behavior ( $\alpha_{DFA}(medium) \cong 0.06$ ;  $n > 5.625$ ), a more stable behavior is expected as the scale reaches the maximum value.

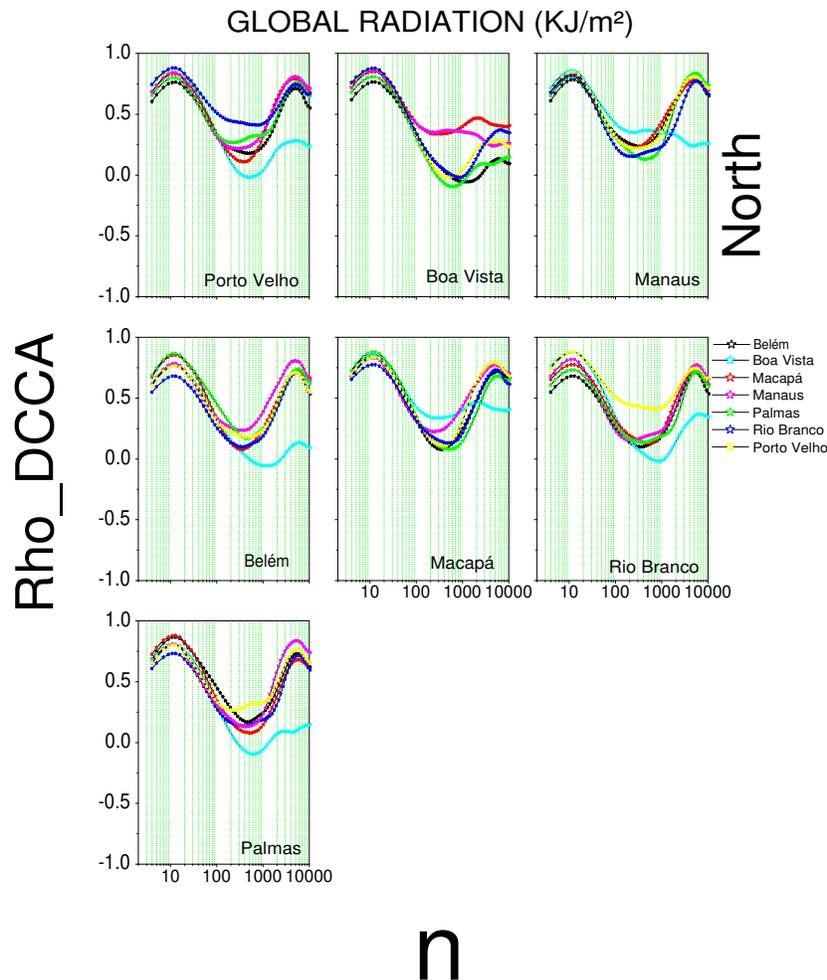
It can be seen that the result of autocorrelation in the state capitals of the Southeast and South regions for humidity (Figure 3S) and wind speed (Figure 4S) have a different pattern, due of their geographic locations (Table 1). The Southeast, with a tropical climate, has the greatest climate variability in Brazil, and its state capitals are influenced by elements such as altitude and latitude (Table 1) and air masses coming from the ocean. It is located in a transition area between an extremely dry region (Northeast region), with the altitude acting as a climate modeler distributing rainfall of different temperatures, and the South region. The South, on the other hand, with temperatures that are usually lower than the rest of Brazil, has a shorter summer with high temperatures and a rigorous winter, being a region that receives winds from the Southeast region, raising the humidity and heat for the region, increasing the temperature and causing rain.

### DCCA cross-correlation coefficient analysis

In this section, to present the analysis of the  $\rho_{DCCA}$  coefficient, the organization is done by Brazilian region (North, Northeast, Central-West, Southeast and South), and by environmental variable (radiation, temperature, humidity and wind speed). A

detailed analysis of radiation is presented in this section, while the plots for the remaining variables are presented in the supplementary material: North (Figures 5S-8S), Northeast (Figures 8S-10S), Central-West (Figures 11S-13S), Southeast (Figures 14S-16S) and South (Figures 17S-19S), for the environmental variables radiation, temperature, humidity and wind speed, respectively.

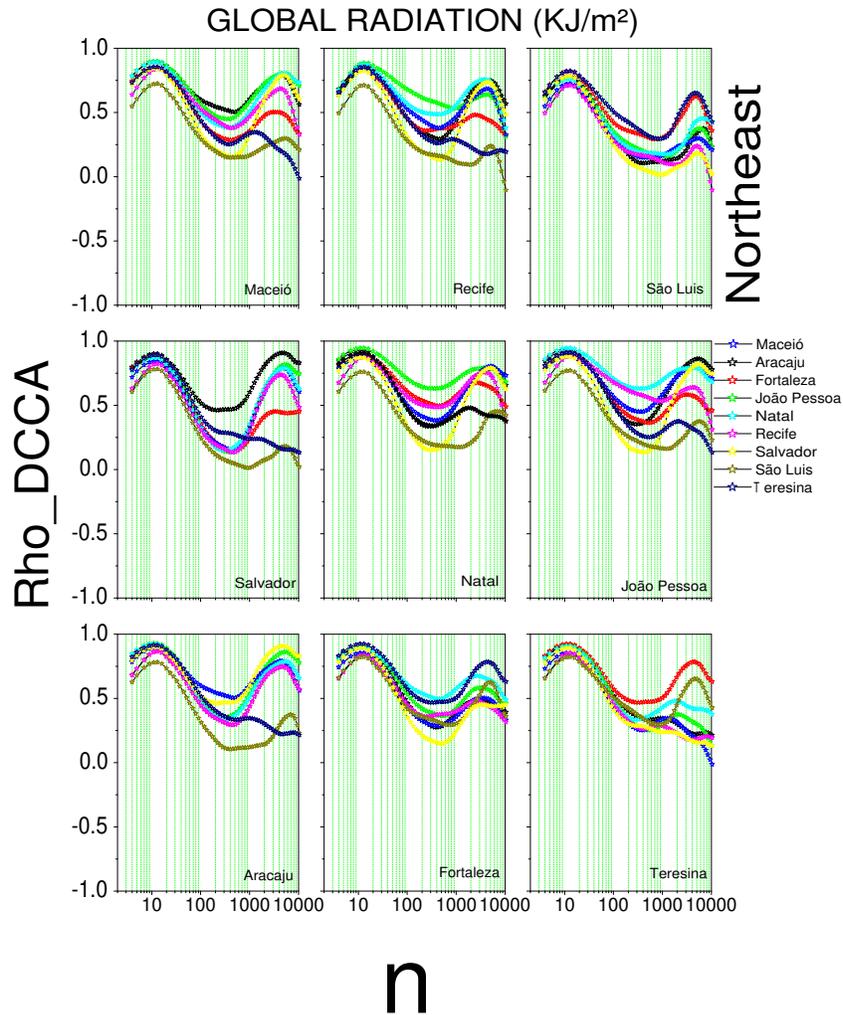
Figure 2 shows the results for global radiation in the North region. We can observe three characteristic behaviors, two maximum points at  $n = 13$  (one day of observations -  $\rho_{DCCA_{Medium}} \sim 0.83$ ) and  $n = 5625$  (430 days of observations -  $\rho_{DCCA_{Medium}} \sim 0.14$ ) and a minimum point at  $n = 457$  (35 days of observations -  $\rho_{DCCA_{Medium}} \sim 0.70$ ), showing a variation in the fluctuation with the increase of the scale. With a minimum approaching zero for all correlations in six capitals in the North region (Belém, Macapá, Manaus, Palmas, Porto Velho and Rio Branco), Boa Vista shows a different behavior with a minimum below zero ( $\rho_{DCCA_{Palmas}} \times \rho_{DCCA_{BoaVista}}$ ), that is, a negative cross correlation ( $\rho_{DCCA} < 0$ ). The temperature (Figure 5S) and humidity (Figure 6S) show a behaviour very similar to radiation. Once again Boa Vista stands out in relation to the other capitals when analysing temperature (Porto Velho, Belém, Macapá and Palmas) and humidity for all capitals. For wind speed (Figure 7S), up to  $n = 13$ , the cross correlation coefficient is very similar, followed by individual variations per capital. The behaviour was very similar for Manaus, Porto Velho and Rio Branco for all scales, whereas Belém, Macapá and Palmas show variations between  $-0.73 < \rho_{DCCA} < 0.61$  for  $n > 457$ .



**Figure 2.** Cross-correlation,  $\rho_{DCCA}(n)$ , for global radiation in the North region of Brazil. The plots show the cross-correlations between the state capital written in the plot and all others in the region.

Figure 3 shows the results for global radiation in the Northeast region. The  $\rho_{DCCA}$  is positive for all capitals with a variation

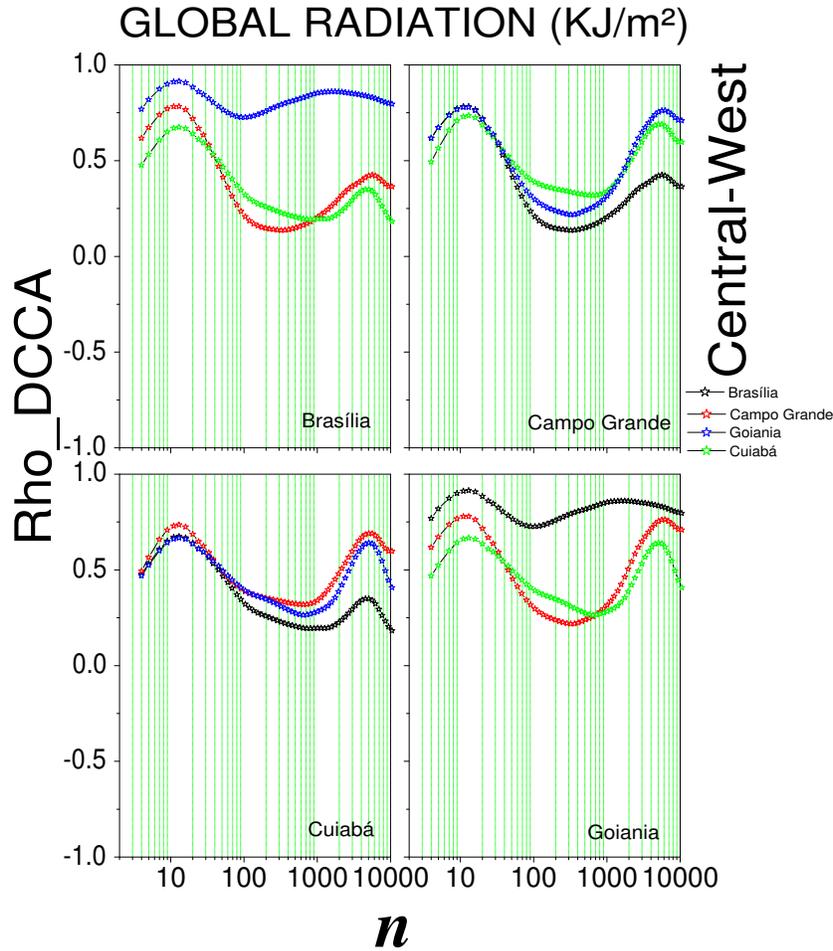
between  $0.21 < \rho_{DCCA} < 0.88$ , and the minimum value between  $(\rho_{DCCA_{Teresina}} \times \rho_{DCCA_{Macei}})$ . Figure 8S shows a positive cross-correlation for temperature between all capitals for  $n \leq 135$ , and with fluctuation between  $-0.30 < \rho_{DCCA} < 0.99$  for all capitals after for values of  $n$  above 135. Humidity (Figure 9S) shows a positive cross-correlation ( $\rho_{DCCA} > 0$ ) for all scales, with the exception of  $(\rho_{DCCA_{Fortaleza}} \times \rho_{DCCA_{Macei}})$  from  $n > 3328$  (256 days) and  $(\rho_{DCCA_{Teresina}} \times \rho_{DCCA_{Macei}})$  to  $n > 1570$  (120 days). The  $\rho_{DCCA}$  coefficient shows a fluctuating between  $-0.06 < \rho_{DCCA} < 0.91$  for all capitals. The wind speed (Figure 10S) showed a positive cross-correlation for all capitals, with the exception of  $(\rho_{DCCA_{Aracaju}} \times \rho_{DCCA_{Recife}})$  and  $(\rho_{DCCA_{Aracaju}} \times \rho_{DCCA_{JoaoPessoa}})$  for  $91 < n < 1261$ , and  $(\rho_{DCCA_{Teresina}} \times \rho_{DCCA_{Aracaju}})$  for  $135 < n < 1261$ . The  $\rho_{DCCA}$  coefficient fluctuates between  $-0.06 < \rho_{DCCA} < 0.92$  for all capitals.



**Figure 3.** Cross-correlation,  $\rho_{DCCA}(n)$ , for global radiation in the Northeast region of Brazil. The plots show the cross-correlations between the state capital written in the plot and all others in the region.

In the Brazilian Central-West region (Figures 4 and 11S-13S, for radiation, temperature, humidity and wind speed, respectively) showed positive cross-correlation ( $\rho_{DCCA} > 0$ ) for all scales, being each variable with a specific fluctuation.

Figure 5 shows the results for global radiation in the Southeast region, with a positive cross-correlation ( $\rho_{DCCA} > 0$ ) in all capitals. Maximum values are observed for  $n = 13$  (one day) and  $n = 5625$  (430 days), and minimum values are observed for  $n = 252$  (19 days). Here the cross-correlation coefficient varies between  $0.19 < \rho_{DCCA} < 0.81$ . The temperature (Figure 14S) also shows a positive cross-correlation coefficient for all capitals. With the increase of the scale for  $n \leq 13$  we can see that the correlations tend to grow, however, for  $13 < n \leq 38$  a decreasing trend in the correlations is visible, followed by growth



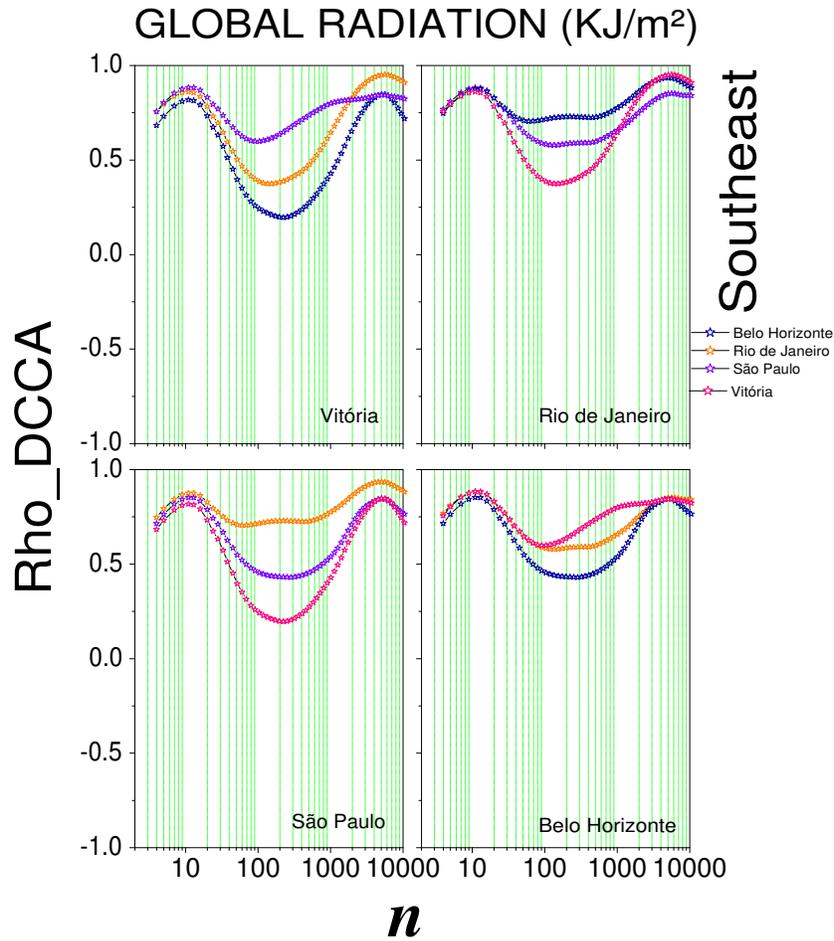
**Figure 4.** Cross-correlation,  $\rho_{DCCA}(n)$ , for global radiation in the Central-West region of Brazil. The plots show the cross-correlations between the state capital written in the plot and all others in the region.

for  $n > 38$ . For humidity (Figure 15S), variations are visible, with a negative cross correlation for large scales ( $n > 4564$ ) for  $(\rho_{DCCA_{Vitoria}} \times \rho_{DCCA_{BeloHorizonte}})$ . For these four capitals, we observed a variation between  $-0.08 < \rho_{DCCA} < 0.62$ . Wind speed (Figure 16S), shows a positive cross-correlation in São Paulo for all scales. Vitoria, Rio de Janeiro and Belo Horizonte in the range between  $45 < n < 1130$  show show negative  $\rho_{DCCA}$  values. In these four capitals, a variation between  $-0.19 < \rho_{DCCA} < 0.69$  is observed.

Figure 6 shows the results for global radiation in the South region. For small scales ( $n \leq 13$ ) an increasing trend is observed, followed by a decrease for  $13 < n \leq 576$ , and by another increase for  $n > 576$ . Here the correlation varies between  $0.18 < \rho_{DCCA} < 0.99$ . For temperature (Figure 17S), after  $n > 38$  (3 days), increasing the scale also increases the cross-correlation. For these three capitals, a variation between  $0.38 < \rho_{DCCA} < 0.99$  is observed. Humidity (Figure 18S) and wind speed (Figure 19S), some stability between  $24 < n < 809$  was observed, being more evident for humidity in Florianópolis ( $0.13 < \rho_{DCCA} < 0.22$ ) and for the wind speed in Curitiba, Florianópolis and Porto Alegre with the cross-correlation coefficient varying between  $0.17 < \rho_{DCCA} < 0.20$ . Differently from other regions, the cross-correlation coefficient trends to fall for wind speed when considering  $n > 5068$ , with a variation between  $0.75 < \rho_{DCCA} < 0.39$ .

### Network analysis

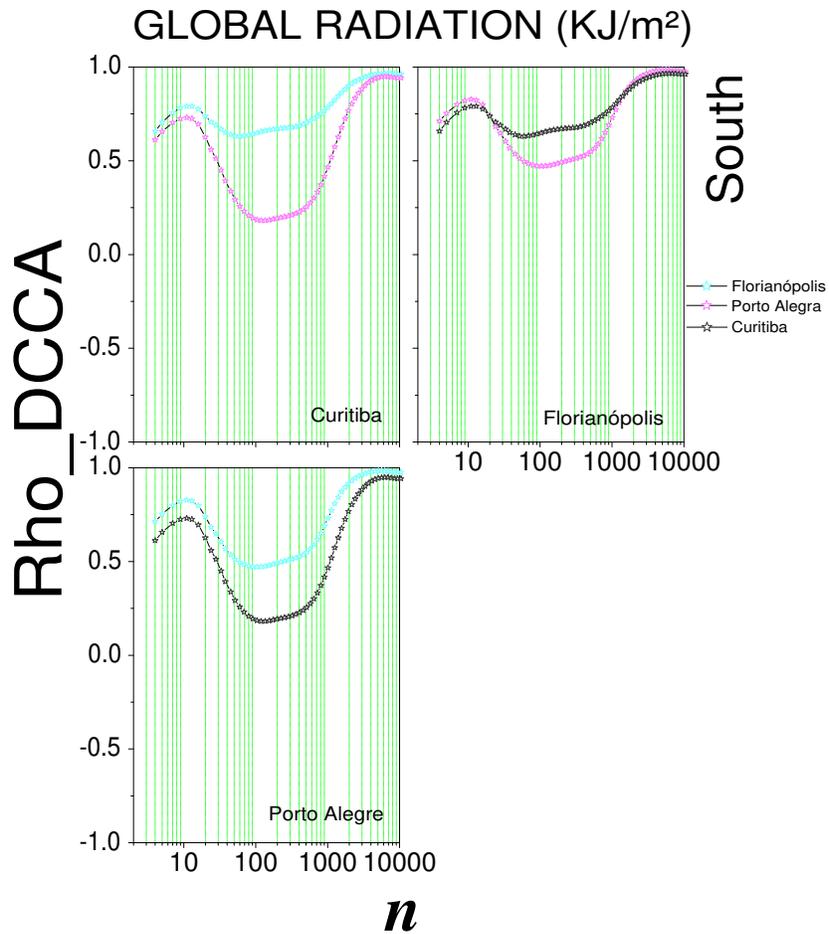
In this third phase of our analysis, we use the concept of network analysis to recognize hidden patterns in the data, to group and to classify them. Our analysis takes into account the stations located in each Brazilian state capital and Federal District



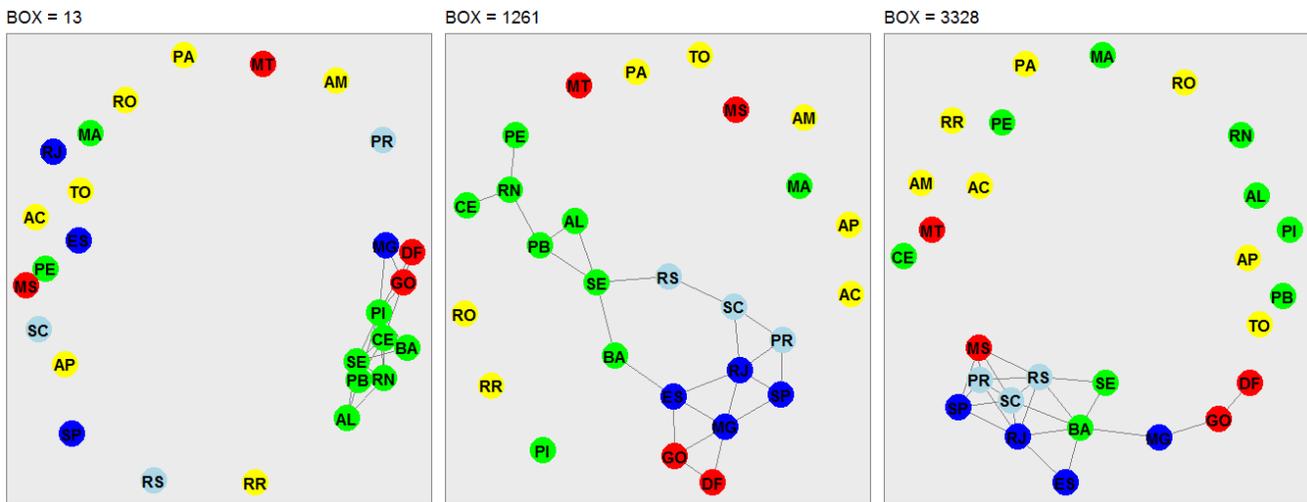
**Figure 5.** Cross-correlation,  $\rho_{DCCA}(n)$ , for global radiation in the Southeast region of Brazil. The plots show the cross-correlations between the state capital written in the plot and all others in the region.

(Table 1). The nodes of the interconnected networks are represented by the capitals and the edges are represented by the strength of the cross-correlation coefficient which satisfy  $\rho_{DCCA} \geq 90\%$ . The choice of a threshold of 90% takes into account the representation of strong correlations and also avoids overcrowding in the representation of networks. The networks showed in Figures 7-10, are associated to the variables radiation, temperature, humidity and wind speed, respectively. The criteria adopted form boxes of sizes  $n = 13$  (one day of observations),  $n = 1261$  (97 days of observations) and  $n = 3328$  (256 days of observations).

Figure 7 shows the network analysis for solar radiation. Regardless of the box size (13, 1261, and 3328), we see subgroups of networks interconnected mostly by nearby regions. For  $n = 13$ , the only interconnected network subgroup shows a link between seven Northeast capitals, two Central-West capitals and one Southeast capital. When considering  $n = 1261$ , this sub-network also includes the three Southeast capitals that border the Atlantic ocean and the three South capitals that also border the ocean. For  $n = 3328$  some of the Northeast capitals closer to the equator are removed from the sub-network.

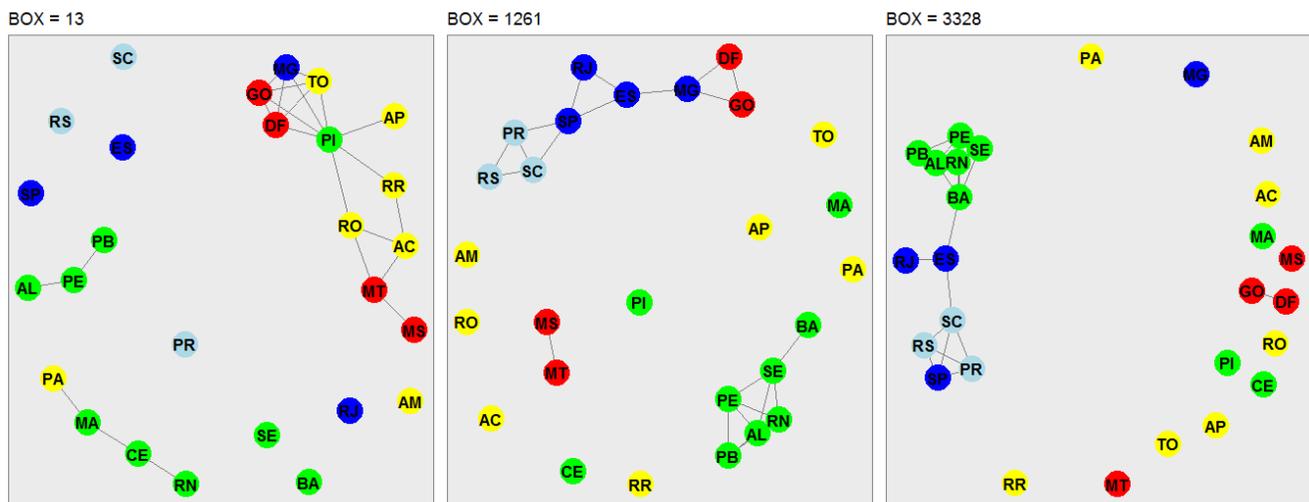


**Figure 6.** Cross-correlation,  $\rho_{DCCA}(n)$ , for global radiation in the South region of Brazil. The plots show the cross-correlations between the state capital written in the plot and all others in the region.



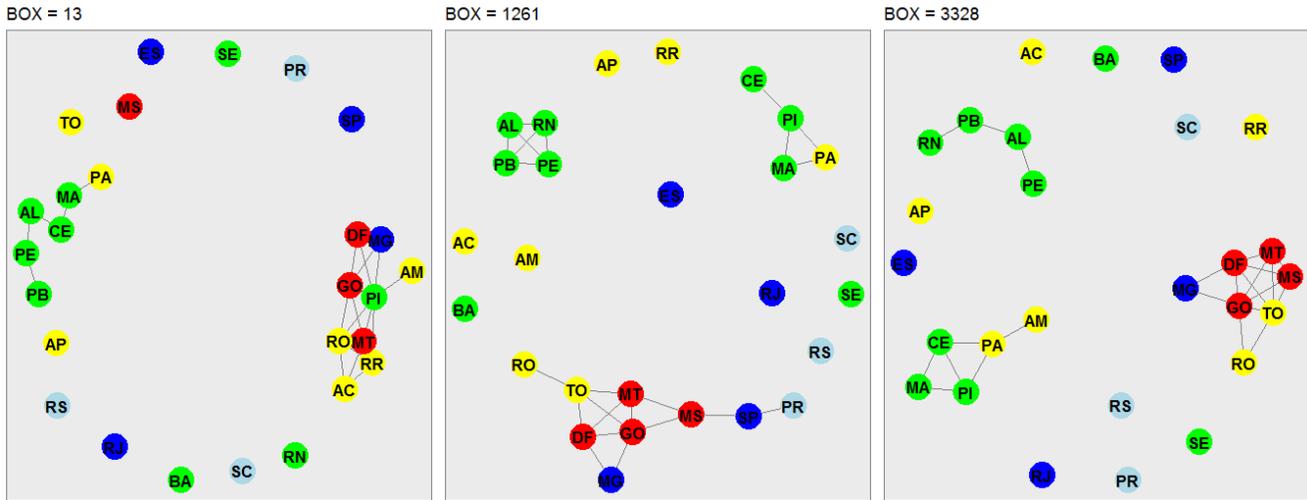
**Figure 7.** Network representation of the environmental variable global solar radiation, considering the 27 Brazilian state capitals as vertices and the edges defined by the DCCA cross-correlation coefficient between pairs of Brazilian state capitals. Only edges associated with the top 10% *DCCA* cross-correlation coefficients between objects were depicted. The five colors in the vertices correspond to the state capitals in each of the five Brazilian regions. Box sizes of 13—one day, 1261–97 days, and 3328–256 days, are depicted, from left to right.

Figure 8 shows the network analysis for the air temperature. When analysing the box size  $n = 13$ , we see a strong cross-correlation, however fragmented into three subgroups of networks, the first represented by the capitals of Rio Grande do Norte, Ceará, Maranhão and Pará, from the North and Northeast regions; the second formed by the capitals of Alagoas, Pernambuco and Paraíba, in the Northeast; and the third by the capitals of Minas Gerais, Tocantins, Goiás, Federal District, Piauí, Amapá, Roraima, Rondônia, Acre, Mato Grosso and Mato Grosso do Sul. For  $n = 1261$ , three subgroups are also formed: (i) the capitals of Mato Grosso and Mato Grosso do Sul, with hot and humid climate; (ii) the capitals of the Northeast states of Bahia, Sergipe, Pernambuco, Rio Grande do Norte, Alagoas and Paraíba, all of them bordering the Atlantic Ocean; and the capitals of Rio Grande do Sul, Santa Catarina, Paraná, São Paulo, Rio de Janeiro, Espírito Santo, Minas Gerais, Distrito Federal and Goiás, states located in a close geographic location between the South, Southeast and Central-West regions. For  $n = 3328$ , two subgroups are formed, the first by the capital of Goiás and the Distrito Federal, which are very near geographically, and the second by the capitals of Pernambuco, Paraíba, Sergipe, Alagoas, Rio Grande do Norte, Bahia, Rio de Janeiro, São Paulo, Espírito Santo, Santa Catarina, Rio Grande do Sul and Paraná, representing almost all Brazilian coast from the Northeast to the South.

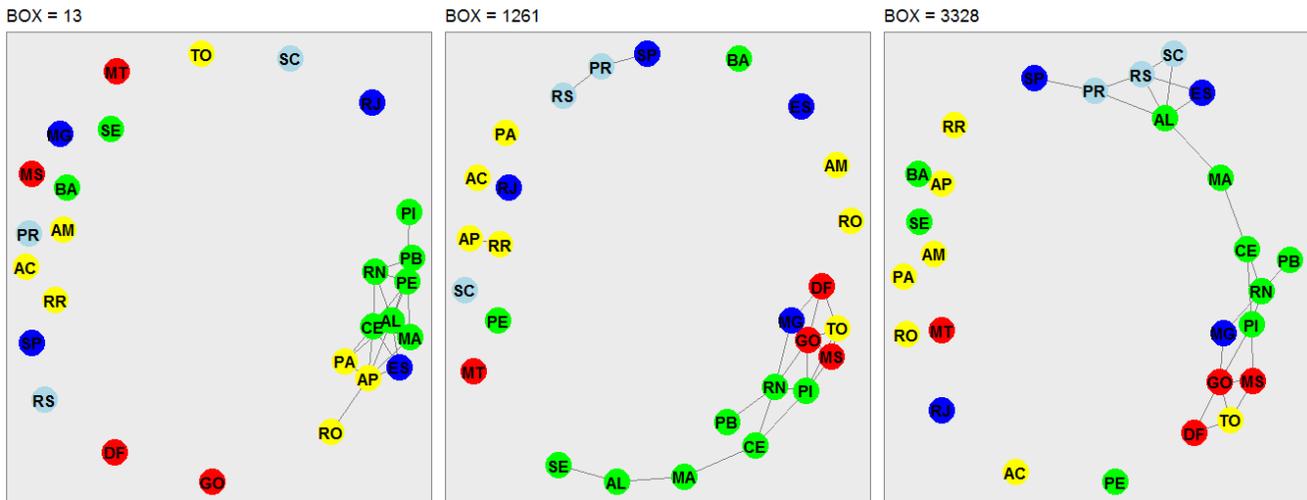


**Figure 8.** Network representation of the environmental variable air temperature, considering the 27 Brazilian state capitals as vertices and the edges defined by the DCCA cross-correlation coefficient between pairs of Brazilian state capitals. Only edges associated with the top 10% DCCA cross-correlation coefficients between objects were depicted. The five colors in the vertices correspond to the state capitals in each of the five Brazilian regions. Box sizes of 13—one day, 1261–97 days, and 3328–256 days, are depicted, from left to right.

Figures 9 and 10 show the network analysis for the relative air humidity and wind speed, respectively, and their interpretation can be done in a similar manner as in Figures 7 and 8. Similarly to radiation and air temperature, the edges are mostly connections between neighbouring regions and regions located in the Brazilian coast.



**Figure 9.** Network representation of the environmental variable relative air humidity, considering the 27 Brazilian state capitals as vertices and the edges defined by the DCCA cross-correlation coefficient between pairs of Brazilian state capitals. Only edges associated with the top 10% DCCA cross-correlation coefficients between objects were depicted. The five colors in the vertices correspond to the state capitals in each of the five Brazilian regions. Box sizes of 13—one day, 1261–97 days, and 3328–256 days, are depicted, from left to right.



**Figure 10.** Network representation of the environmental variable wind speed, considering the 27 Brazilian state capitals as vertices and the edges defined by the DCCA cross-correlation coefficient between pairs of Brazilian state capitals. Only edges associated with the top 10% DCCA cross-correlation coefficients between objects were depicted. The five colors in the vertices correspond to the state capitals in each of the five Brazilian regions. Box sizes of 13—one day, 1261–97 days, and 3328–256 days, are depicted, from left to right.

## Concluding remarks

To better understand the behaviour in Brazilian climate, in this paper, four environmental and atmospheric variables, wind speed, radiation, temperature and humidity, measured in 27 stations (the capital of each of the 26 Brazilian states plus the federal district) were considered. The detrended fluctuation analysis was used to evaluate the statistical self-affinity of the signal being the overall main conclusion that the non-stationary and persistent behaviour ( $\alpha_{DFA}(medium) \cong 1.21$ ;  $n \leq 13$ ) upward and/or downward trends are time dependent. In the long range persistent behaviour ( $\alpha_{DFA}(medium) \cong 0.87$ ;  $13 < n \leq 5.625$ ), it is observed that variables with closer values remain for a longer time influencing the memory. As for the anti-persistent behaviour ( $\alpha_{DFA}(medium) \cong 0.06$ ;  $n > 5.625$ ), a more stable behaviour is expected as the scale reaches the maximum value.

The cross correlation coefficient  $\rho_{DCCA}$  was used to quantify the long range cross correlation between Brazilian state capitals, per geographical region (North, Northeast, Central-West, Southeast and South) and per atmospheric variable. Similar behaviour for the  $\rho_{DCCA}$  were found when considering state capitals near to each other geographically.

The network analysis was based on the DCCA cross-correlation coefficient ( $\rho_{DCCA}$ ) between pairs of the Brazilian state capitals for each of the four environmental variables. Short, median and long distance correlations (box values of 13–one day, 1261–97 days, and 3328–256 days), were considered, and only values of the top 10% DCCA cross-correlation coefficients were depicted in the networks. The edges of the networks for each of the atmospheric variables and each of the box sizes, are mostly connections between neighbouring regions and regions located in the Brazilian coast.

The methodology used and combined in this paper represent a step forward in the field of time series and network analysis, that can be applied to other regions, other environmental variables and also to other fields of research. The results presented are of great importance to better understand the behaviour of environmental and atmospheric variables in Brazil, their long range cross correlation and the way they relate to each other, and provide helpful insights about how the climate is changing and renewable energy production.

## 1 Data availability

The data that support the findings of this study are available from the authors.

## 2 Code availability

Source code for all the analyses included in this paper are available from the authors. The R software was used for the data cleaning, data imputation and network analysis. The results and plots for the DFA and  $\rho_{DCCA}$  were obtained with the software Origin 6.0 and Origin 8.0.

## References

1. Gatti, L. V. *et al.* Amazonia as a carbon source linked to deforestation and climate change. *Nature* **595**, 388–393 (2021).
2. Andrade, L. d. M. B. *et al.* Health-related vulnerability to climate extremes in homoclimatic zones of amazonia and northeast region of brazil. *PloS one* **16**, e0259780 (2021).
3. Menezes, J. A. *et al.* Mapping human vulnerability to climate change in the brazilian amazon: the construction of a municipal vulnerability index. *PloS one* **13**, e0190808 (2018).
4. Sena, A., Ebi, K. L., Freitas, C., Corvalan, C. & Barcellos, C. Indicators to measure risk of disaster associated with drought: Implications for the health sector. *PloS one* **12**, e0181394 (2017).
5. Tan, S.-C., Yao, X., Gao, H.-W., Shi, G.-Y. & Yue, X. Variability in the correlation between asian dust storms and chlorophyll a concentration from the north to equatorial pacific. *PloS one* **8**, e57656 (2013).
6. Srivastava, D. K. & Ramamurthy, V. *Climate Change and Energy Options for a Sustainable Future* (World Scientific, 2021).
7. de Almeida Brito, A., de Araújo, H. A. & Zebende, G. F. Detrended multiple cross-correlation coefficient applied to solar radiation, air temperature and relative humidity. *Sci. reports* **9**, 1–10 (2019).
8. Hodgson, P. E. *Energy, the environment and climate change* (World Scientific, 2010).
9. Schleuning, M. *et al.* Ecological networks are more sensitive to plant than to animal extinction under climate change. *Nat. communications* **7**, 1–9 (2016).
10. Nejat, Z., Reza, N. H., Alireza, S. & Farshad, A. Prediction of the karstic spring flow rates under climate change by climatic variables based on the artificial neural network: a case study of iran. *Environ. monitoring assessment* **192** (2020).

11. Ji, B., Sun, Y., Yang, S. & Wan, J. Artificial neural networks for rice yield prediction in mountainous regions. *The J. Agric. Sci.* **145**, 249–261 (2007).
12. Granata, F. & Di Nunno, F. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric. Water Manag.* **255**, 107040 (2021).
13. Peng, C. K. *et al.* Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
14. Zebende, G. F. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Phys. A* **390**, 614–618 (2011).
15. Stanley, H. E. & Mantegna, R. N. Scaling behavior in the dynamics of an economic index. *Nature* **46** (1995).
16. Stanley, H. *et al.* Anomalous fluctuations in the dynamics of complex systems: from DNA and physiology to econophysics. *Phys. A: Stat. Mech. its Appl.* **224**, 302–321 (1996).
17. Kristoufek, L. Rescaled range analysis and detrended fluctuation analysis: Finite sample properties and confidence intervals. *AUCO Czech Econ. Rev.* **4**, 315–329 (2010).
18. Oliveira-Filho, F. M., Leyva-Cruz, J. A. & Zebende, G. F. Analysis of the EEG bio-signals during the reading task by DFA method. *Phys. A* **525**, 664 – 671 (2019).
19. Mesquita, V. B., Oliveira-Filho, F. M. & Rodrigues, P. C. Detection of crossover points in detrended fluctuation analysis: An application to EEG signals of patients with epilepsy. *Bioinformatics* DOI: [10.1093/bioinformatics/btaa955](https://doi.org/10.1093/bioinformatics/btaa955) (2020).
20. Zebende, G. F., Fernandez, B. F. & Pereira, M. G. Analysis of the variability in the sdB star KIC 10670103: DFA approach. *Mon. Notices Royal Astron. Soc.* **464**, 2638 – 2642 (2017).
21. Walleczek, J. *Self-organized biological dynamics and nonlinear control: toward understanding complexity, chaos and emergent function in living systems* (Cambridge University Press, 2000).
22. Podobnik, B. & Stanley, H. E. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Phys. Rev. Lett.* **100**, 1–4 (2008).
23. Podobnik, B., Horvatic, D., Petersen, A. M. & Stanley, H. E. Cross-correlation between volume change and price change. *Proc. Natl. Acad. Sci.* **106**, 22079–22084 (2009).
24. Podobnik, B. *et al.* Quantifying cross-correlations using local and global detrending approaches. *The Eur. Phys. J. B* **71**, 243–250 (2009).
25. Podobnik, B., Jiang, Z.-Q., Zhou, W.-X. & Stanley, H. E. Statistical tests for power-law cross-correlated processes. *Phys. Rev. E* **84**, 1–8 (2011).
26. Kristoufek, L. Measuring cross-correlation between non-stationary series with DCCA coefficient. *Phys. A* **402**, 291–298 (2014).
27. Kristoufek, L. Detrending moving-average cross-correlation coefficient: Measuring cross-correlations between non-stationary series. *Phys. A* **406**, 169–175 (2014).
28. Zebende, G. F., Brito, A. A. & Castro, A. P. Dcca cross-correlation analysis in time-series with removed parts. *Phys. A* **545**, 123472 (2020).

## Acknowledgements

Paulo Canas Rodrigues acknowledges financial support from the Brazilian National Council for Scientific and Technological (CNPq) grant “bolsa de produtividade PQ-2” 305852/2019-1.

## Author contributions statement

F.M.O.F, E.F.G. and P.C.R. conducted the conceptualization and proposed the methodology, F.M.O.F and E.F.G. conducted the experiments about DFA and  $\rho_{DCCA}$ , P.C.R. conducted the experiments about network analysis. All authors analysed the results, wrote and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Correspondence and requests for materials should be addressed to Paulo Canas Rodrigues ([paulocanas@gmail.com](mailto:paulocanas@gmail.com)).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SMfigures.pdf](#)