Research Article

Improving Regional Dynamic Downscaling with Multiple Linear Regression Model Using Components Principal Analysis: Precipitation over Amazon and Northeast Brazil

Aline Gomes da Silva[1,2] and Claudio Moises Santos e Silva[2,3]

[1] *Instituto Federal do Rio Grande do Norte, Brazil*
[2] *Programa de Pós-Graduação em Ciências Climáticas, Rio Grande do Norte, Brazil*
[3] *Universidade Federal do Rio Grande do Norte, Brazil*

Correspondence should be addressed to Aline Gomes da Silva; aline.gomes@ifrn.edu.br

In the current context of climate change discussions, predictions of future scenarios of weather and climate are crucial for the generation of information of interest to the global community. Due to the atmosphere being a chaotic system, errors in predictions of future scenarios are systematically observed. Therefore, numerous techniques have been tested in order to generate more reliable predictions, and two techniques have excelled in science: dynamic downscaling, through regional models, and ensemble prediction, combining different outputs of climate models through the arithmetic average, in other words, a postprocessing of the output data species. Thus, this paper proposes a method of postprocessing outputs of regional climate models. This method consists in using the statistical tool multiple linear regression by principal components for combining different simulations obtained by dynamic downscaling with the regional climate model (RegCM4). Tests for the Amazon and Northeast region of Brazil (South America) showed that the method provided a more realistic prediction in terms of average daily rainfall for the analyzed period prescribed, after comparing with the prediction made by set through the arithmetic averages of the simulations. This method photographed the extreme events (outlier) that the prediction by averaging failed. Data from the Tropical Rainfall Measuring Mission (TRMM) were used to evaluate the method.

1. Introduction

General circulation models (GCM) have been used for climate prediction over Brazil [1, 2]. Although these models represent the influence of synoptic scale weather systems and aspects of the general circulation that have limitations in representing mesoscale processes, such as squall lines, meteorological systems formed by complex topography, watershed, and others [3]. Thus, due to the large size and complexity of terrain and biomes covering the Brazilian territory, the GCM are limited to representation of regional aspects of the climate on Brazil.

A solution to this problem is use of downscaling technique through regional climate models (MCR). Over South America several studies [1, 4–7] showed the RCM skill relatively to GCM. The effectiveness and suitability of this technique are due to the possibility of using more appropriate physical parameterizations for mesoscale due the increasing of the spatial resolution. These characteristics are important because in regions such as Brazil, forcing mesoscale regulates the spatial and temporal distribution of atmospheric variables, reducing errors in GCMs that are performed with low spatial resolution [8].

An important regional model is the regional climate model (RegCM), which was originally developed at the National Center for Atmospheric Research (NCAR) during the 80s decade [9, 10]. Due to the contribution of many researchers to the RegCM there are six versions: RegCM1,

RegCM2, RegCM2.5, RegCM3, RegCM4, and RegCM4.1. It is widely used because it is public and open source code, moreover, it has a good computational performance.

Despite progress achieved in modeling regional in recent years, there are still many aspects to be explored, evaluated, and improved to a substantial improvement of climate representation through the RCM compared to the representations through the GCMs.

Systematic errors that regional models exhibit, including the RegCM in different regions, especially over tropical region, are due to a lack of fit in the physical parameterizations, especially in parameterizations of convective cumulus and precipitation in grid scale [11, 12].

The most used technique to overcome the lack of adjustment in the parameterization and reduce forecasting errors is called ensemble prediction, which consists of combination of the multiple simulations, performed with different initial conditions or parameterization, for the same period and region. Studies have shown that this method produces more consistent results with observation.

For South America (SA), consequently to Brazil, the situation is no different, there is a need of studies that aim to enhance regional and technical treatment models to output models. However, several studies have focused on simulation with the standard model configuration [3, 8, 15–17], and for the ensemble prediction technique using the arithmetic mean. Therefore, this paper investigates the possibility of improvement of regional model simulations RegCM4 through proper adjustment of physical parameterizations and using appropriate statistical methods to combine multiple simulations. In this sense, we use the technique of multiple linear regression using principal components in order to combine different simulations with the RegCM4. To test the method we apply this technique of combination in the period from February to June 1998. This year was an atypical year, El Nino.

The work is organized as follows. Section 2 will show a brief presentation of the physical parameters of the regional climate model RegCM4 that most influence the simulated rainfall, together with the input data of this model and the data that are used to verify the method proposed in this work. In Section 3, the method multiple linear regression using principal components to combine different simulations will be presented. In Section 4, the results are present. Finally, in Section 5, the conclusions and discussions are drawn.

2. Model Used and Numerical Experiments Performed

2.1. Regional Climate Model (RegCM4). The RegCM4 [18] is the fifth version of RegCM, originally developed by the National Center for Atmospheric Research NCAR [19] and based on mesoscale model (MM5). This is a model of limited area discretized into grid points (Arakawa B). In the vertical system sigma coordinates are used. The primitive equations, which correspond to the core of the dynamic model, are for a compressible hydrostatic fluid [20].

The physical processes are represented in the model by a series of parameterizations. The radiative transfer scheme is the same used in the global model Community Climate Model version 3 (CCM3). This scheme calculates the interaction of gases (H_2O, CO_2, O_3, CH_4, N_2O, and CFC) and aerosols with radiation in the infrared and ultraviolet. For the soil-vegetation-atmosphere interaction, RegCM4 uses the biosphere-atmosphere transfer scheme (BATS) and community land model (CLM: version 3.5). The full description of the model as well as the parameterization options are shown in [18].

The model has three options of convective schemes: (i) Kuo scheme, the most simplified and that is activated when the moisture convergence exceeds a threshold value, and (ii) the convective schemes of Grell [20], which considers the cloud as a plume entrainment model composed by a downdraft and updraft. The interaction, via entrainment of air, with the atmosphere occurs only in the top and in the base of the cloud. The convective activity is activated when the updraft reaches the moist adiabatic. This scheme is more sensitive to precipitation efficiency (PEFF) parameter. This parameter quantifies the portion of precipitation that will evaporate before reaching the ground. Therefore, high PEFF values decrease precipitation. Two types of closures can be used: Arakawa and Schubert (all potential energy available for convection is adjusted for each time step [13]) and Fritsch-Chappell (1980) (scale convective adjustment in the order of 30 minutes [14]); (iii) parameterization MIT-emanuel [21], which characterizes the convection trigger when the level of free convection is higher than the cloud base.

For stratiform precipitation the RegCM4 use the subgrid explicit moisture (SUBEX), which was developed by [22]. The formulation for the auto conversion of cloud water into precipitation is as follows:

$$P = C_{ppt} \left(\frac{Q_c}{FC} - Q_c^{th} \right) FC. \tag{1}$$

C_{ppt} is the conversion rate, Q_c is the amount of water present, Q_c^{th} is the minimum amount of water that must remain in the cloud, and FC is the conversion factor of water present in precipitation. FC value depends on the minimum relative humidity (RH_{min}) for cloud formation, according to the equation:

$$FC = \sqrt{\frac{RH - RH_{min}}{RH_{max} - RH_{min}}}. \tag{2}$$

The RH_{max} value is 101%, RH_{min} may vary between 1 and 100%, and RH is local relative humidity. The threshold amount of water in the cloud is given by:

$$Q_c^{th} = C_{acs} 10^{-0.49 + 0.013T}. \tag{3}$$

T is the temperature in degrees Celsius and C_{acs} is scaling factor.

2.2. Numerical Experiments

2.2.1. Data. The initial and boundary conditions of the atmosphere (wind, temperature, surface pressure, and water
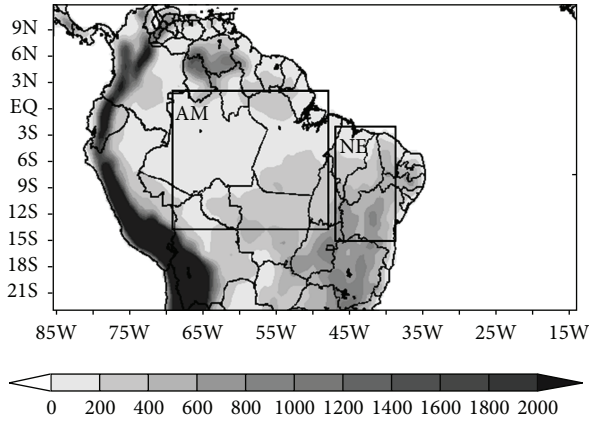
Figure 1: Topography in meters (m) of the domain used to run the simulations. Amazon region (AM) and Northeast (NE) of Brazil indicated by black boxes.

vapor) used in the simulation conditions are of the ERA-Interim reanalysis. The ERA-Interim is a global dataset of the atmosphere produced by the European Centre for Medium-Range Weather Forecast (ECMWF) with a horizontal grid spacing of $1.5°$ by $1.5°$ and frequency of six hours (00:00, 06:00, 12:00, and 18:00 UTC) [23]. The topography and ground cover are from the United States Geological Survey (USGS) and Global Land Cover Characterization (GLCC), with 60 minutes of horizontal grid spacing [24].

The dataset of sea surface temperature (SST) used were produced by the National Oceanic and Atmospheric Administration (NOAA) using in situ data and satellite, through optimal interpolation (OI) [25]. The data are weekly and available from 1989 to the present day, centered on Wednesday, with a resolution of $1.0°$ by $1.0°$.

The simulated precipitation data will be compared with data Tropical Rainfall Measuring Mission (TRMM) product 3B42-V7. These data are obtained by using satellite infrared channels with $0.25°$ by $0.25°$ resolution, latitude versus longitude [26].

2.2.2. Configuration of the Experiments. Seven simulation tests were performed during the Austral autumn, beginning at 00:00 UTC on February 15th, 1998, and ending at 00:00 UTC on June 1th in the same year. February was discarded because this is the time adjustment (spin-up) of the model.

The model grid spacing is 50 km and 18 vertical levels, with the top at 5 hPa. The domain and the topography are shown in Figure 1. Two regions will be analyzed: Amazon (AM) and Northeast (NE) region of Brazil as indicated in Figure 1. Table 1 summarizes the settings of the experiments that varied according to the convective scheme (Grell and MIT-Emanuel), minimum relative humidity for formation of cloud in scale grid (RH_{min}), and the dynamic control (closure) of Grell model (Arakawa-Schubert or Fritsch-Chappell); in addition different PEFF are used if the scheme was the Grell convective.

Table 1: Configuration of the seven simulations.

Simulation	Precipitation Efficiency (PEFF)	SUBEX (RHmin)	Dynamic control (closure)
GR_PD_SD	0.25–1.00	0.65	Arakawa and Schubert (1974) [13]
GR_PD_SW	0.25–1.00	0.90	Arakawa and Schubert (1974) [13]
EM_SD		0.65	
EM_SW		0.90	
GR_PW_SD	0.25–0.50	0.65	Arakawa and Schubert (1974) [13]
GR_PW_SW	0.25–0.50	0.90	Arakawa and Schubert (1974) [13]
GR_PW_SW_FC	0.25–0.50	0.90	Fritsch and Chappell (1980) [14]

GR: parameterization of cumulus Grell; MS: parameterization of cumulus MIT-Emanuel; PD: PEFF dry (high evaporation rate of the raindrop, therefore decreases precipitation); PW: PEFF wet (low evaporation rate of the raindrop, therefore increases the precipitation); SD: SUBEX dry (low minimum relative humidity for cloud formation); SW: SUBEX wet (high minimum relative humidity for cloud formation); FC: closure cloud Fritsch and Chappell (1980) [14].

3. Multiple Linear Regression Using Principal Components

To minimize the error in climate forecasts, predictions with several different configurations are generated and combined. This method is called ensemble prediction [27]. Usually the ensemble prediction is made via a simple arithmetic average (AA) from different simulations or models, or weighting by measures of dispersion.

In this paper we will compare the usual method with the method of multiple linear regression using principal components (here we call PCR method), to produce a combination of the seven experiments described in Section 2.2.2.

The method of multiple linear regression is a multivariate technique that consists in finding a linear relationship between a dependent variable (response variable), in this case, the observed data, and more than one independent variable (predictors variables) that describe the system; here, these are output of the climate model RegCM4.

The following equation shows this relationship, where Y_i is the variable to be estimated, X_{mi} are the predictors variables, α_0 the intercept, and α_m the coefficients of multiple linear regression to be estimated by least squares method [28]. This method consists in finding a solution that minimizes the sum of squared residuals, which is the difference between the observed and predicted (estimated):

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_m X_{mi} + \epsilon_i. \tag{4}$$

The problem of multiple linear regression is to find the α_m coefficients that relate the independent variables and the dependent variable; this step can be called calibration of the regression model. To find this solution we rewrite (4) in matrix form, taking the Y-matrix with the dependent variable, the X-matrix with the independent variables, the A-matrix with α_m coefficients, and the E-matrix with errors ϵ_i:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_i \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{m1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1i} & \cdots & X_{mi} \end{bmatrix},$$

$$A = \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_m \end{bmatrix}, \qquad E = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \end{bmatrix}. \tag{5}$$

Rewriting the problem in matrix form, we have

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_i \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{m1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1i} & \cdots & X_{mi} \end{bmatrix} \cdot \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \end{bmatrix}. \tag{6}$$

Multiplying the A-matrix by X-matrix and adding the ϵ-matrix, we obtain the equation below, but in matrix form:

$$Y = XA + \epsilon. \tag{7}$$

The least squares method is used to find the coefficients of multiple linear regression with the condition that the sum of the squares of the errors be minimum. For this, isolate the error in (4), getting

$$Y_i - (\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_m X_{mi}) = \epsilon_i. \tag{8}$$

Then the sum of squared errors (SSE: shown in (9) and in matrix form in (10)) is minimized through the derivative with respect to A-matrix equaling to zero, as shown in (11). By isolating A-matrix (step not shown), we have as the solution of the multiple linear regression in (12). Consider the following:

$$\text{SSE} = \sum \left(y_i - (\alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_m X_{mi})\right)^2, \tag{9}$$

$$\text{SSE} = \epsilon^T \epsilon = (Y - XA)^T (Y - XA), \tag{10}$$

$$\frac{\partial (\text{SSE})}{\partial A} = 0, \tag{11}$$

$$A = \left[X^T X\right]^{-1} X^T Y. \tag{12}$$

A possible obstacle to find the solution of (12) is that the matrix $X^T X$ cannot be inverted. In other words, it can be a singular matrix, where some predictors variables are linear combinations of other, so there is a correlation between the independent variables. When this occurs, there is multicollinearity and there is no single least squares estimators for the parameters. For climate ensemble prediction, the simulations with different configurations from a single climate model are correlated. Thus, to avoid multicollinearity

we will use the principal components of the simulations. This technique aims to explain the structure of variance and covariance of a random vector by constructing linear combinations of the original variables, which are, for this problem, the predictors variables of multiple linear regression. These linear combinations are called principal components and are not correlated [29]. Therefore, the principal components of the explanatory variables are a new set of variables with the same information of the original variables, but uncorrelated, eliminating multicollinearity. The use of principal components to fit a multiple linear regression model was proposed initially by [30]. This technique is called multiple linear regression using principal components.

The first step is to find the principal components (PCs), Z-matrix, of the matrix of predictors variables X, where the relationship between them is given by

$$Z = P'X. \tag{13}$$

P is the orthogonal matrix of dimension $m \times m$ (m is the number of predictors variables) consisting of eigenvectors of the covariance matrix or correlation matrix, X. Thus, (7) and (12) can be rewritten in the forms

$$Y = ZA + \epsilon,$$
$$A = \left[Z^T Z\right]^{-1} Z^T Y. \tag{14}$$

Finding P-matrix, with the weights of each simulation, and the A-matrix, with the regression coefficients, the regression model is calibrated, this matrix should be used as setting for new ensemble prediction. The eigenvectors of the P-matrix that provides the weights of each predictors variable are used to find the new matrix of principal components Z_{NEW} of new simulations X_{NEW}, given by

$$Z_{\text{nova}} = P' X_{\text{nova}}. \tag{15}$$

After, to find the principal components, using the coefficient A-matrix, the ensemble prediction Z_{PRED} is obtained by the relation

$$Y_{\text{prev}} = Z_{\text{nova}} A. \tag{16}$$

The multiple linear regression using principal components can work with all PC obtained from the original data, or only to work with components that have higher correlation with the response variable [31]. In the latter case, the errors can be minimized.

For the analysis the results were calculated Bias, mean error (ME), mean absolute error (MAE), and root mean square error (RMSE), according to (17), (18), (19), and (20), respectively. P_{oi} is the observed precipitation, P_{Pi} is the precipitation predicted, and n is the number of data:

$$\text{Bias} = P_{oi} - P_{Pi}, \tag{17}$$

$$\text{ME} = \frac{\sum_{i=1}^n (P_{oi} - P_{Pi})}{n}, \tag{18}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |P_{oi} - P_{Pi}|}{n}, \quad (19)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (P_{oi} - P_{Pi})^2}{n}}. \quad (20)$$

4. Results

4.1. The Regression Model via Principal Component. The TRMM data, which will be the dependent variable Y, was obtained through average daily precipitation, from March 01 to May 31, for the Amazon region and Northeast region of Figure 1. Similarly independent variables were obtained, which is simulated precipitation (X-matrix) of the seven experiments. Preliminary tests indicated that the larger the number of simulations improves the ensemble prediction.

First step was to find the seven principal components of the X-matrix, which composes the Z-matrix (Section 3). Despite the cumulative variance explained to be equal to 86% e 96% in the fourth principal component (see Tables 2 and 3), for AM e NE region, respectively, the implementation of PCR method were considered all PCs (seven PCs not shown here) because each one captures a different parameter of the configuration of the model RegCM4, except for the first component which is a measure of the intensity of the rain. The PC[2] split the effect of the different parameterizations of cumulus used, Grell and MIT-Emanuel; PC[3] differentiates PEF/Wet and PEF/Dry associated with the Grell scheme; PC[4] captures the difference SUBEX/Dry and SUBEX/Wet associated with Grell scheme; PC[5] distinguishes different PEFF associated with different closure of the clouds; PC[6] differentiates closure of the cloud used for parameterization of Grell; and PC[7] captures the difference between the association of Emanuel parameterization with SUBEX/Dry and SUBEX/Wet. Finally, to run the PCR, the regression equations (21) show the regression coefficients that associate each component principal (PC) with the precipitation (Prec), for the Amazon (Prec$_{AM}$) and Northeast (Prec$_{NE}$) of the Brazil, respectively. This equation allows us to estimate the average daily precipitation for the period analyzed, using with the same coefficients:

$$\begin{aligned}\text{Prec}_{AM} = {} & 7.15 - 1.00 * \text{PC}[1] - 0.80 * \text{PC}[2] \\ & + 0.17 * \text{PC}[3] - 0.81 * \text{PC}[4] - 0.66 * \text{PC}[5] \\ & - 0.57 * \text{PC}[6] + 1.6 * \text{PC}[7], \\ \text{Prec}_{NE} = {} & 2.03 - 0.52 * \text{PC}[1] - 0.26 * \text{PC}[2] \\ & - 0.20 * \text{PC}[3] + 0.24 * \text{PC}[4] - 0.46 * \text{PC}[5] \\ & + 0.08 * \text{PC}[6] + 1.6 * \text{PC}[7].\end{aligned}$$
$$(21)$$

For the regression model to be appropriate, one must satisfy three requirements: (i) the residues must to present random distribution around the mean zero, (ii) the residues must have a normal distribution, and (iii) the variance must

TABLE 2: Proportion of variance and cumulative proportion for Amazon region for each principal component (PC).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Proportion of variance	0.48	0.17	0.12	0.09	0.07	0.06	0.01
Cumulative proportion	0.48	0.65	0.77	0.86	0.93	0.99	1.00

TABLE 3: Proportion of variance and cumulative proportion for Northeast region for each principal component (PC).

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Proportion of variance	0.67	0.18	0.08	0.03	0.02	0.01	0.01
Cumulative proportion	0.67	0.85	0.93	0.96	0.98	0.99	1.00

to be homogeneous. The residues in the graphs of Figures 2(a) and 2(d), to Amazon and Northeast of the Brazil, respectively, apparently do not present any particular pattern or trend indication.

The plots in Figures 2(b) and 2(e) show the quantiles of the residuals versus the quantiles of the normal distribution, called QQ-plot for the Amazon and Northeast, respectively. This is necessary to verify the assumption of normality of residuals. The closer to a line, the residues are close to a normal distribution. Figures 2(c) and 2(f) show the square root of the normalized residual versus predicted values randomly distributed, indicating the homogeneity of variance. Therefore, we conclude that model satisfies the three conditions.

4.2. The Performance of Regression Model. For the AA method, we calculated the arithmetic mean of the seven simulations. With the purpose of comparing the performance of the PCR and AA methods to represent the daily rainfall, the graphs in Figure 3 present data from TRMM versus simulated for both methods and regions. The results were compared for the Amazon region and Northeast in Figure 3, with the PCR method in Figures 3(a) and 3(c), and AA method in Figures 3(b) and 3(d). We concluded that the simulation through the ensemble PCR shows a better correlation with the TRMM data relatively to the AA ensemble, especially in the Amazon region.

Despite the North and Northeast of Brazil being located in the tropical region, one has different responses in simulations in climate models. Overall, the simulations for the Northeast converge to the observed, presenting a smaller bias compared to the northern region bias. This is due to the variation of topography, distance to the ocean, the diversity of vegetation types, and forms of land use and other factors. Therefore, the efficiency of the PCR method is sharper in the region with the largest bias.

From the boxplot of TRMM data, PCR, and AA ensembles in Figure 4, we find that the median and interquartile range of the data obtained by AA ensemble diverges significantly from the TRMM data. For the model obtained
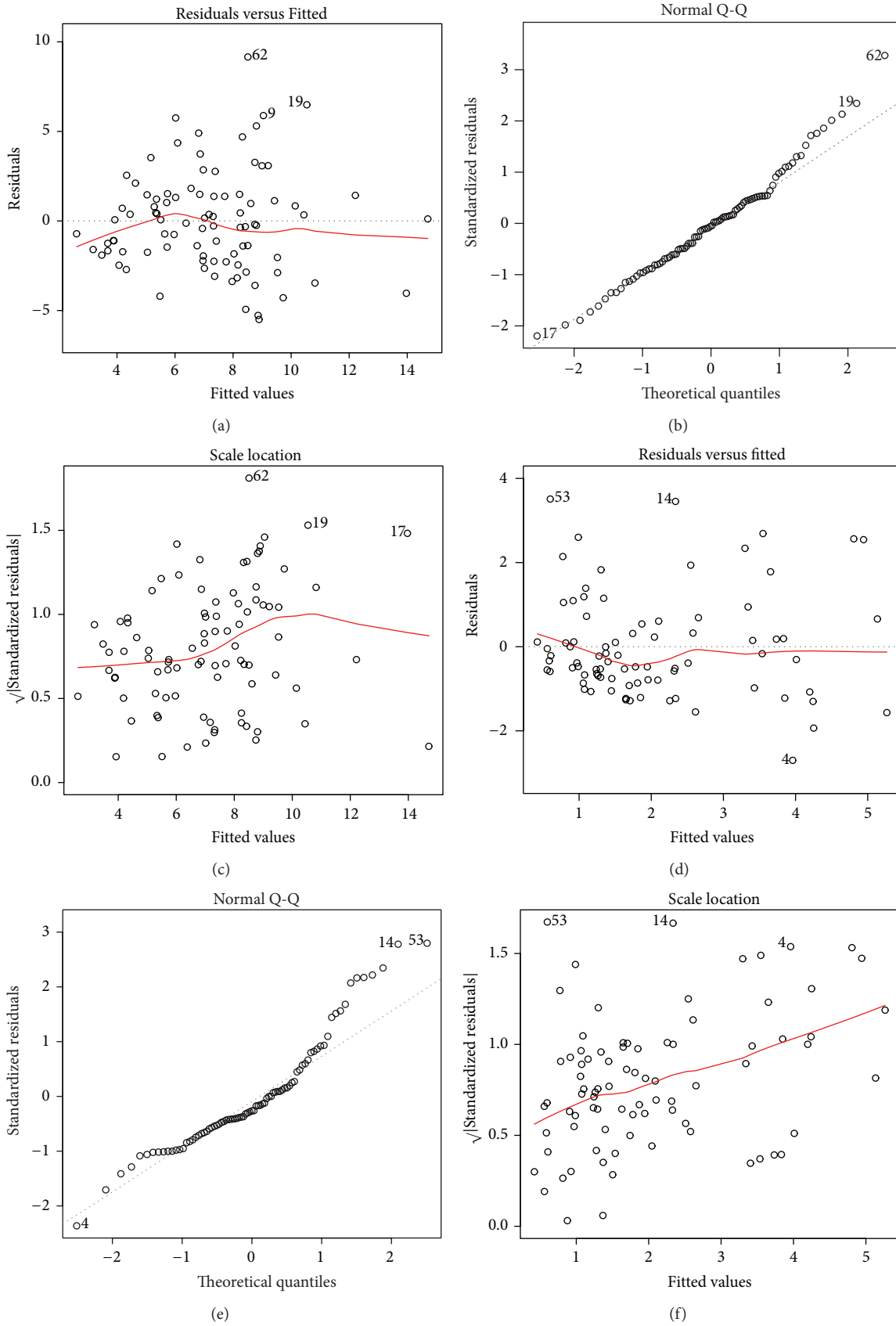
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 2: (a) and (d) residues (mm/day) with zero mean, (b) and (e) residues with normal distribution, and (c) and (f) homogeneity of variance of the residue. (a), (b), and (c) the Amazon region. (d), (e), and (f) Northeast region.
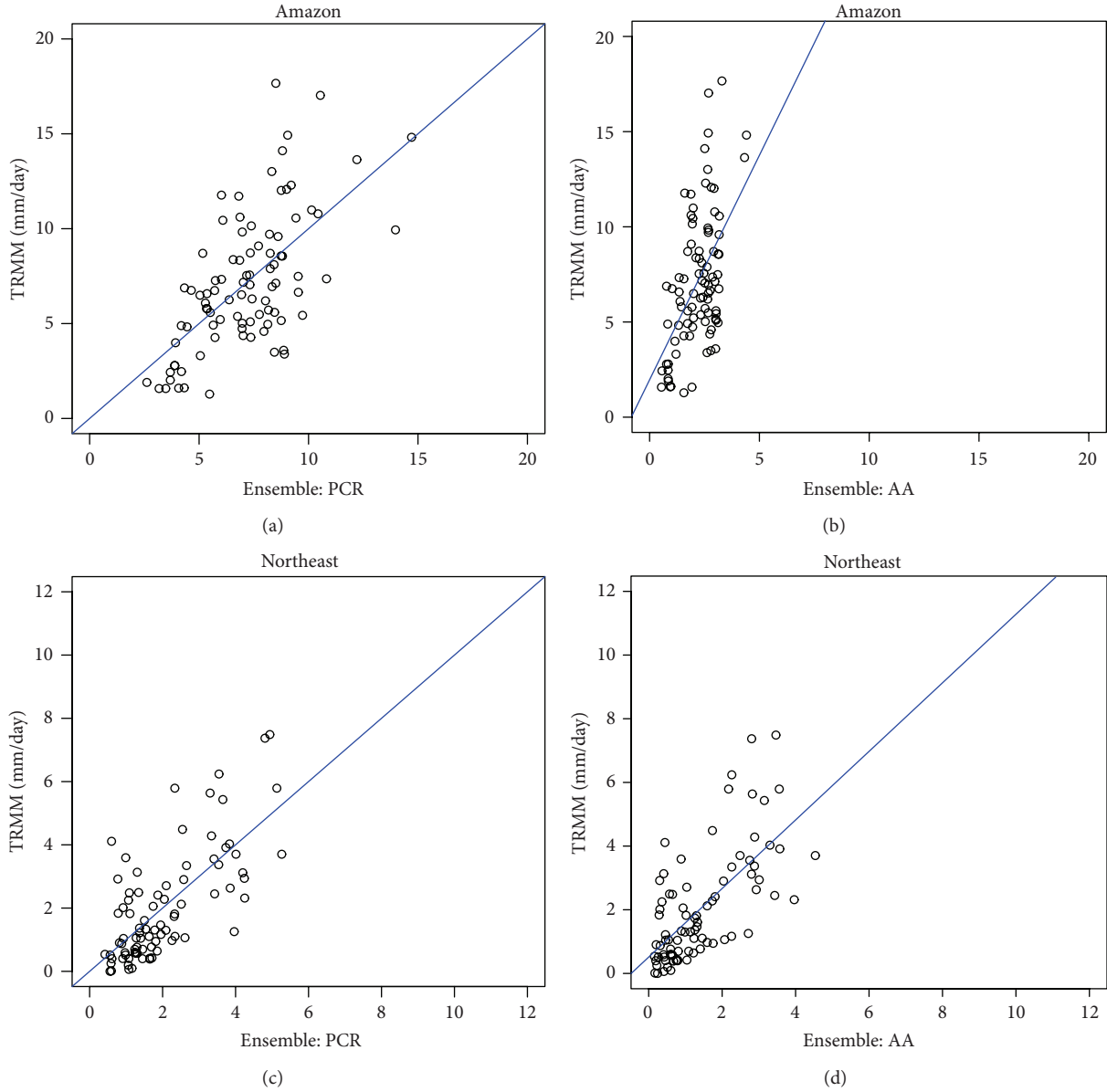
FIGURE 3: (a) and (c) average daily precipitation data for the PCR ensemble versus TRMM data for the Amazon region and Northeast, respectively. (b) and (d) AA ensemble versus TRMM data for the Amazon region and Northeast, respectively.

with the PCR ensemble, compared to data from TRMM, there is the similarity in median precipitation and variance. Regarding the PCR method, there is a slight underestimation of the intense events and overestimation of the weak events. Moreover, this method is able to capture two extremes events (outliers) in accordance with the data TRMM.

The variability of observation explained by simulations R^2 for the PCR ensemble was approximately 40%. This value is higher than that obtained with the AA method, which was 28% (see Table 4). The F-test, also shown in Table 4, is higher than the tabulated F-value, which for a confidence level of 95% is 2.214. The probability of obtaining this result is measured by the P value, which showed low values, of the order of 10^{-7} for the PCR method.

TABLE 4: F-test, P value, and R^2 for PCR and AA methods, to the North and Northeast domain.

Region	Method	Estatística-F	Valor-P	R^2
Amazon	PCR	7.795	$3.279 \cdot 10^{-7}$	0.399
Amazon	AA	35.51	$5.13 \cdot 10^{-8}$	0.287
Northeast	PCR	10.78	$2.552 \cdot 10^{-9}$	0.501
Northeast	AA	67.05	$3.195 \cdot 10^{-12}$	0.452

Table 5 shows the mean error (ME), mean absolute error (MAE), and root mean square error (RMSE), calculated according to (16), (17), and (18), respectively, for PCR and AA ensembles. As expected, the ME was approximately zero
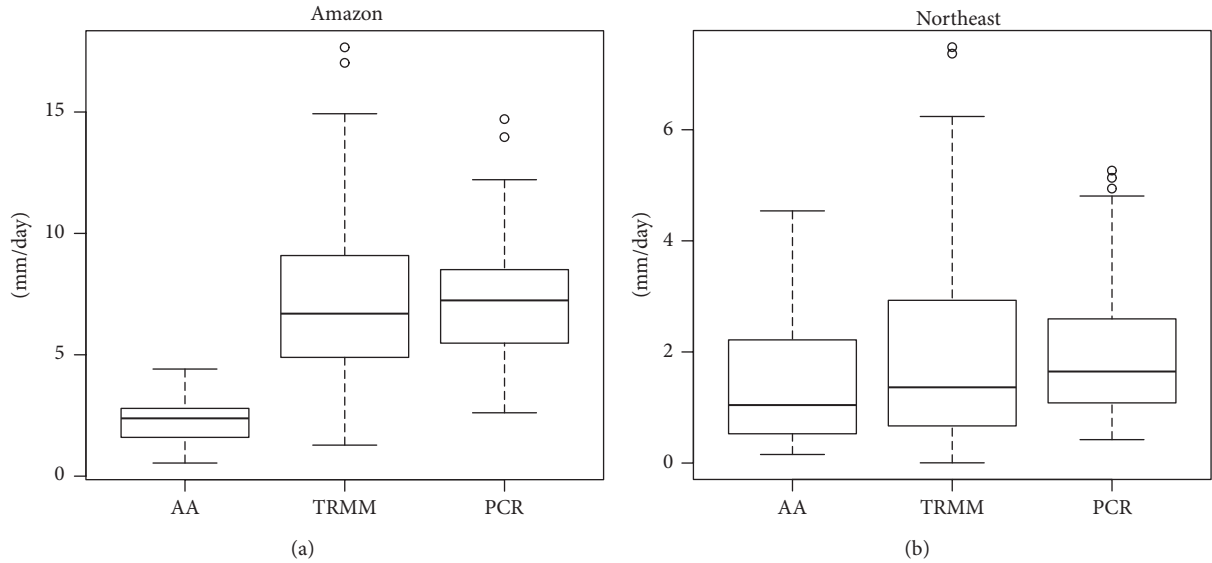
FIGURE 4: Boxplot of average daily precipitation (mm/day) for the AA ensemble, TRMM data, and PCR ensemble in the (a) Amazon region and (b) Northeast region.

TABLE 5: ME (mm/day), MAE (mm/day), and RMSE (mm/day) for PCR and AA methods, to the North and Northeast domain.

Region	Method	ME	MAE	RMSE
Amazon	PCR	$-1.1 \cdot 10^{-4}$	2.14	2.76
Amazon	AA	4.94	4.95	5.89
Northeast	PCR	$-3.6 \cdot 10^{-5}$	0.94	1.22
Northeast	AA	0.62	0.98	1.43

for the PCR method to the two regions, once the graph of Figures 3(a) and 3(d) shows the uniform distribution of residue around zero. This shows that there is a trend of underestimation or overestimation of the method. The MAE indicates the magnitude of the error. The MAE for the AA method was approximately twice the value obtained by the PCR method for Amazon region. For Northeast region, the value MAE was 8% less with PCR method. The RMSE had results similar to MAE.

5. Final Comments

Errors and uncertainties in weather and climate forecasting will always exist due to several sources of errors present in a simulation and can be classified into two classes: incomplete or erroneous atmospheric initial conditions and inadequacy of the numerical model.

These errors in the initial conditions are due to instrumental limitations for data collection, discretized observations, and irregularly spaced, increasing the difficulty of interpolation to the grid structure. In the case of models of limited area, the artificial boundary condition increases the errors and uncertainties.

Inadequacy of the numerical model consists in difficulty to represent the influence of all physical-chemical-biological factors in the state of the atmosphere and its evolution in time.

With the ensemble prediction method by varying the physical parameterization, the error due to the inadequacy of the model is minimized since several possibilities of representing the state of the atmosphere are reproduced, and a solution is generated from these. Thus, decreases in the probability of observing extremes surprise that a particular setting or parameter could not represent the forecast.

By comparing the prediction method routinely performed (AA) together with the method presented here, we found that combination of simulations that are correlated, in other words, simulations that bring the same information, or contribution to the final solution does not improve the prediction. A treatment is needed to remove redundant information from simulations, that is, a principal component analysis. And from this, assign specific weights to this new set of variables using multiple linear regression.

The PCR method performed better in the Amazon region, where individual forecasts more diverged from the observations. For the Northeast region, where the bias was close to zero, the result was comparable to the average of the simulations. A significant advantage of the PCR method was the ability to capture extreme events (outlier) for both regions, since the prediction of these events is of interest to the community.
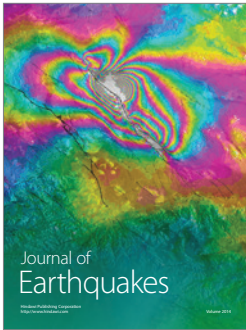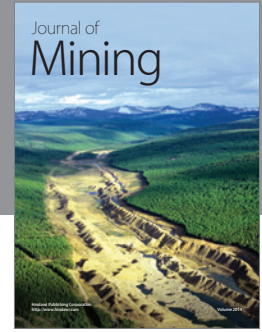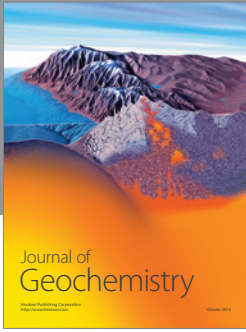
Studies are still needed. Besides, to check the effectiveness of the methods to other regions and periods, it is necessary to take point to point of grid to obtain a spatial distribution of precipitation, refining the process, instead of using the average of a region, as performed here with the purpose a preliminary analysis.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Nobre, A. D. Moura, and L. Sun, "Dynamical downscaling of seasonal climate prediction over Nordeste Brazil with ECHAM3 and NCEP's regional spectral models at IRI," *Bulletin of the American Meteorological Society*, vol. 82, no. 12, pp. 2787–2796, 2001.

[2] B. Liebmann, S. J. Camargo, A. Seth et al., "Onset and end of the rainy season in South America in observations and the ECHAM 4.5 atmospheric general circulation model," *Journal of Climate*, vol. 20, no. 10, pp. 2037–2050, 2007.

[3] J. P. R. Fernandez, S. H. Franchito, and V. B. Rao, "Simulation of the summer circulation over South America by two regional climate models. Part I: mean climatology," *Theoretical and Applied Climatology*, vol. 86, no. 1–4, pp. 247–260, 2006.

[4] L. Sun, F. H. M. Semazzi, F. Giorgi, and L. A. Ogallo, "Application of the NCAR regional climate model to eastern Africa—1. Simulation of the short rains in 1988," *Journal of Geophysical Research*, vol. 104, pp. 6529–6548, 1999.

[5] F. Giorgi and L. O. Mearns, "Introduction to especial section: regional climate modeling revisited," *Journal of Geophysical Research*, vol. 104, no. D6, pp. 6335–6352, 1999.

[6] V. Misra, P. A. Dirmeyer, and B. P. Kirtman, "Dynamic downscaling of seasonal simulations over South America," *Journal of Climate*, vol. 16, no. 1, pp. 103–117, 2003.

[7] L. Sun, D. F. Moncunill, H. Li, A. D. Moura, F. D. A. D. S. Filho, and S. E. Zebiak, "An operational dynamical downscaling prediction system for Nordeste Brazil and the 2002–04 real-time forecast evaluation," *Journal of Climate*, vol. 19, no. 10, pp. 1990–2007, 2006.

[8] R. D. Machado and R. P. Rocha, "Previsões climáticas sazonais sobre o Brasil: Avaliação do RegCM3 aninhado no modelo global CPTEC/COLA," *Revista Brasileira de Meteorologia*, vol. 26, no. 1, pp. 121–136, 2011.

[9] R. E. Dickinson, R. M. Errico, F. Giorgi, and G. T. Bates, "A regional climate model for the Western United States," *Climatic Change*, vol. 15, no. 3, pp. 383–422, 1989.

[10] F. Giorgi and G. T. Bates, "The climatological skill of a regional model over complex terrain," *Monthly Weather Review*, vol. 117, no. 11, pp. 2325–2347, 1989.

[11] F. Giorgi, X. Bi, and J. S. Pal, "Mean, interannual variability and trends in a regional climate change experiment over Europe. I. Present-day climate (1961-1990)," *Climate Dynamics*, vol. 22, no. 6-7, pp. 733–756, 2004.

[12] E. B. Souza, M. N. Lopes, E. G. Rocha et al., "recipitação sazonal sobre Amazônia Oriental no período chuvoso: Observações e simulações regionais com o RegCM3," *Revista Brasileira de Meteorologia*, vol. 24, no. 2, pp. 111–124, 2009.

[13] A. Arakawa and W. A. Schubert, "Interaction of a cumulus cloud ensemble with the large scale environment, part I," *Journal of the Atmospheric Sciences*, vol. 31, pp. 674–701, 1974.

[14] J. M. Fritsch and C. F. Chappell, "Numerical prediction of convectively driven mesoscale pressure systems. Part I: convective parameterization," *Journal of the Atmospheric Sciences*, vol. 37, no. 8, pp. 1722–1733, 1980.

[15] A. Seth and M. Rojas, "Simulation and sensitivity in a nested modeling system for South American—part I: reanalyses boundary forcing," *Jornal of Climate*, vol. 16, pp. 2437–2453, 2003.

[16] A. Seth, S. A. Rauscher, S. J. Camargo, J.-H. Qian, and J. S. Pal, "RegCM3 regional climatologies for South America using reanalysis and ECHAM global model driving fields," *Climate Dynamics*, vol. 28, no. 5, pp. 461–480, 2007.

[17] C. M. Santos e Silva, A. Silva, P. Oliveira, and K. C. Lima, "Dynamical downscaling of the precipitation in Northeast Brazil with a regional climate model during contrasting years," *Atmospheric Science Letters*, vol. 15, no. 1, pp. 50–57, 2014.

[18] F. Giorgi, E. Coppola, F. Solmon et al., "RegCM4: Model description and preliminary tests over multiple CORDEX domains," *Climate Research*, vol. 52, no. 1, pp. 7–29, 2012.

[19] F. Giorgi, "Simulation of regional climate using a limited area model nested in a general circulation model," *Journal of Climate*, vol. 3, pp. 941–963, 1990.

[20] G. A. Grell, "Prognostic evaluation of assumptions used by cumulus parameterizations," *Monthly Weather Review*, vol. 121, no. 3, pp. 764–787, 1993.

[21] K. A. Emanuel and M. Živković-Rothman, "Development and evaluation of a convection scheme for use in climate models," *Journal of the Atmospheric Sciences*, vol. 56, no. 11, pp. 1766–1782, 1999.

[22] J. S. Pal, E. E. Small, and E. A. B. Eltahir, "Simulation of regional-scale water and energy budgets: representation of subgrid cloud and precipitation processes within RegCM," *Journal of Geophysical Research: Atmospheres*, vol. 105, no. D24, pp. 29579–29594, 2000.

[23] D. P. Dee, S. M. Uppala, A. J. Simmons et al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011.

[24] T. R. Loveland, B. C. Reed, J. F. Brown et al., "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1303–1330, 2000.

[25] R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, "An improved in situ and satellite SST analysis for climate," *Journal of Climate*, vol. 15, no. 13, pp. 1609–1625, 2002.

[26] G. J. Huffman, R. F. Adler, D. T. Bolvin et al., "The TRMM Multi-satellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *Journal of Hydrometeorology*, vol. 8, no. 1, pp. 38–55, 2007.

[27] E. Kalnay, "Atmospheric modelling, data assimilation and predictability," *Quarterly Journal of the Royal Meteorological Society*, vol. 129, no. 592, p. 2442, 2003.

[28] D. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press, 1995.

[29] S. A. Mingoti, "Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada," Belo Horizonte, Editora da UFMG, 2005.

[30] M. G. Kendall, *A Course in Multivariate Analysis*, Griffin, London, UK, 1957.

[31] R. Mo and D. M. Straus, "Statistical-dynamical seasonal prediction based on principal component regression of GCM ensemble integrations," *Monthly Weather Review*, vol. 130, no. 9, pp. 2167–2187, 2002.