Article

Modeling Spatiotemporal Rainfall Variability in Paraíba, Brazil

Elias Silva de Medeiros [1],*[iD], **Renato Ribeiro de Lima** [2], **Ricardo Alves de Olinda** [3][iD] **and Carlos Antonio Costa dos Santos** [4,5][iD]

[1] Faculty of Exact Sciences and Technology, Federal University of Grande Dourados, Dourados, Mato Grosso do Sul 79.804-970, Brazil

[2] Department of Statistics, Federal University of Lavras—UFLA, Lavras 37.200-000, Brazil

[3] Department of Statistics, State University of Paraíba, Campina Grande, Paraíba 58429-500, Brazil

[4] Academic Unity of Atmospheric Sciences, Center for Technology and Natural Resources, Federal University of Campina Grande, Campina Grande 58.109-970, Brazil

[5] Daugherty Water for Food Global Institute, University of Nebraska, Lincoln, NE 68588-6203, USA

* Correspondence: eliasmedeiros@ufgd.edu.br; Tel.: +55-67-3410-2094

check for updates

Abstract: The purpose of this study was to provide a detailed framework to use the spatiotemporal kriging to model the space-time variability of precipitation data in Paraíba, which is located in the northeastern region of Brazil (NEB). The NEB is characterized by an irregular, highly variable distribution of rainfall in space and time. In this region, it is common to find high rates of rainfall at locations adjacent to those with no record of rain. Paraíba experiences localized periods of drought within rainy seasons and distinct precipitation patterns among the state's mesoregions. The mean precipitation values observed at several irregularly spaced rain gauge stations from 1994 to 2014 showed remarkable variations among the mesoregions in Paraíba throughout the year. As a consequence of this behavior, there is a need to model the rainfall distribution jointly with space and time. A spatiotemporal geostatistical methodology was applied to monthly total rainfall data from the state of Paraíba. The rainfall data indicate intense spatial and temporal variabilities that directly affect the water resources of the entire region. The results provide a detailed spatial analysis of sectors experiencing precipitation conditions ranging from a scarcity to an excess of rainfall. The present study should help drive future research into spatiotemporal rainfall patterns across all of NEB.

Keywords: spatiotemporal kriging; Brazil; precipitation

1. Introduction

The spatiotemporal behavior of rainfall is of great importance in the regional management of water resources and has a direct influence on human activities such as livestock management, agriculture, and commerce. One of the statistical techniques responsible for analyzing this type of behavior is spatiotemporal geostatistical analysis. The interpolation of spatiotemporal observations presents benefits compared to purely spatial predictions. One of these benefits is that interpolation can be applied to georeferencing positions over space-time [1,2].

In this study, a spatiotemporal geostatistical methodology was applied to monthly total rainfall data from the state of Paraíba, which is located in the northeastern region of Brazil (NEB). The NEB is characterized by an irregular, highly variable distribution of rainfall in space and time. In this region, it is common to find high rainfall rates at locations adjacent to those with no record of rain [3]. Paraíba also experiences localized periods of drought within rainy seasons and distinct precipitation patterns among the state's mesoregions [4,5]. As a consequence of this behavior, there is a need to model the rainfall distribution jointly with space and time.

In the literature, several studies have been carried out on the spatial interpolation of precipitation data using geostatistical techniques [6–9]. Martínez-Cob [10] used three methods of geostatistical interpolation (ordinary kriging, cokriging, and modified residual kriging) to interpolate the mean total annual precipitation in a mountainous region of Aragon, located in northeastern Spain. Goovaerts [11] constructed maps of the spatial prediction of rainfall erosivity in the Algarve region of Portugal. To calculate these predictions, the author used three geostatistical methods (simple kriging with varying local means, kriging with external drift, and collocated cokriging), in addition to linear regression, concluding that linear regression, not considering the spatial dependence among the stations, produced the greatest predictive errors. Haberlandt [12] carried out a study on daily precipitation at 281 stations located in the southeast region of Germany using the methods of kriging with external drift and indicator kriging with external drift for spatial interpolation of precipitation. Also, an increasing number of studies in recent years have used geostatistical methods to model not only the spatial, but also the temporal, dynamics of precipitation [13,14].

In recent years, studies have appeared analyzing the regional spatiotemporal distribution of precipitation. Raja et al. [15] applied the spatiotemporal kriging technique to rainfall data from Turkey. They consequently observed that Turkey is located in a region with a higher risk of possible effects from climate change, similar to the state of Paraíba, Brazil. Thus, they reported that an understanding of the temporal and spatial characteristics of rainfall is crucial for risk management, as well as for the planning, management, and utilization of water resources, which depend heavily on rainfall. Similarly, Hu et al. [13] performed spatiotemporal regression using kriging in the Xinjiang Uygur Autonomous region, located in northwestern China, to model monthly rainfall data from 2004 to 2013. The authors identified a strong spatiotemporal correlation in the rainfall distribution in this region, thereby justifying the use of their proposed methodology. Furthermore, Martínez et al. [14] performed a spatiotemporal analysis of rainfall data from Colombia and identified regions that exhibit irregular rainfall distributions.

The novelty of this study is the detailed presentation of a tool for spatiotemporal interpolation of precipitation in the region under study. It is worth noting that no research has applied this type of methodology to rainfall data from this region until now. The present study should help drive future research into spatiotemporal rainfall patterns across all of NEB. The purpose of this study was to provide a detailed framework to use the spatiotemporal kriging to model the space-time variability of precipitation data in Paraíba, Brazil.

2. Materials and Methods

The datasets used in this study were obtained from the Executive Agency of Water Management in the State of Paraíba (AESA), which is responsible for rainfall information in the region. The state of Paraíba spans an area of approximately 56,585 km^2 in northeastern Brazil between the 6° and 8° parallels of south latitude and the 34° and 39° meridians of west longitude. The state of Paraíba is situated in a tropical region, and it is divided into the following four mesoregions: Zona da Mata, Agreste, Borborema, and Sertão.

The dataset utilized in the present study includes the time series of the total monthly rainfall recorded at 269 rain gauge stations from 1994 to 2014. About 80% of the stations have missing data, represented as white tones (Figure 1). Thus, to adjust the trend of the total monthly rainfall data and then to adjust the residual variogram, 54 stations with records for all of the months in the study period were considered because these locations adequately represent the variability in the data. The other stations excluded from the adjustment of the variogram were used to perform spatiotemporal kriging.

The spatial distribution of the 269 rainfall stations is shown in Figure 2. The different color tones in the map identify the four mesoregions—namely, Zona da Mata, Agreste, Borborema, and Sertão—of Paraíba. The smaller points within the map represent rain gauge stations, whereas the larger points are the 54 rainfall stations that were used to adjust the trend and the variogram.
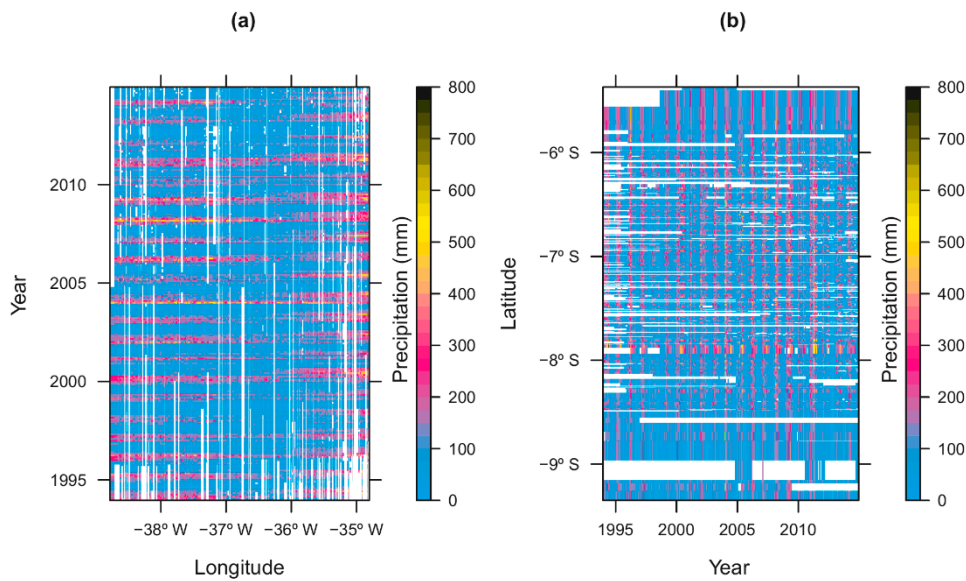
Figure 1. Monthly rainfall of the 269 stations in the period of 1994–2014. Missing data is shown as white tones.
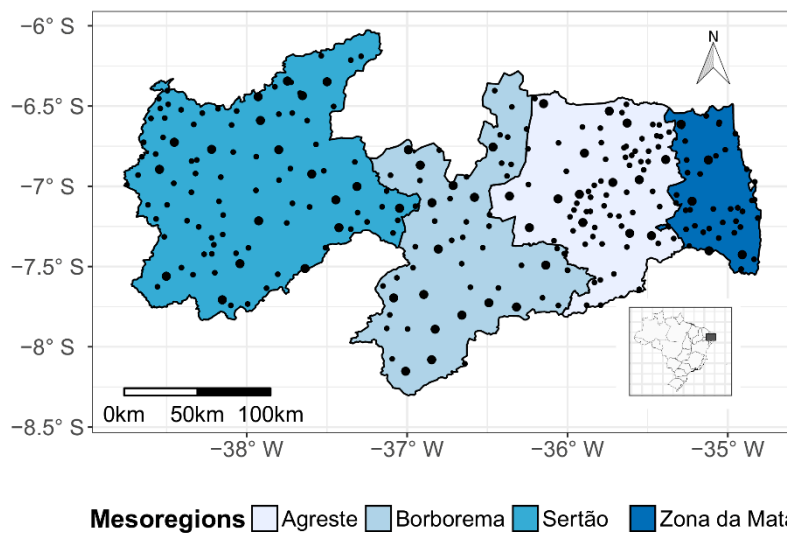
Figure 2. Spatial distribution of 269 rainfall stations sampled (black dots) in the State of Paraíba, divided into four mesoregions. The larger points are the 54 rainfall stations used to adjust the trend and the variogram.

2.1. Spatiotemporal Model

In this study, a spatiotemporal process $\left\{ Z(\mathbf{s},t) : (\mathbf{s},t) \in D \subseteq \mathbb{R}^d \times \mathbb{R} \right\}$ was assumed, where $\mathbf{s} \in \mathbb{R}^d$, $d = 2$ represented the two dimensions in space (latitude and longitude), and $t \in \mathbb{R}$ represented only one dimension, namely, time. The spatiotemporal variation in Z was established by the decomposition of the deterministic m and stochastic residual ε as follows:

$$Z(\mathbf{s},t) = m(\mathbf{s},t) + \varepsilon(\mathbf{s},t) \tag{1}$$

In Equation (1), it was assumed that Z had both first- and second-order moments. The component of the tendency $m(\mathbf{s},t) = E[Z(\mathbf{s},t)]$ is the expectation variable Z. This component was determined here by an ordinary least squares (OLS) regression model. The residual ε included the following three components: spatial, temporal, and interaction-based components [1]. For modeling purposes,

it was assumed that these three components were second-order stationary, mutually independent, and spatially isotropic.

The observed rainfall values were considered in the spatiotemporal random field defined in Equation (1), where the deterministic component $m(\mathbf{s}, t)$, which is spatiotemporally non-constant, represented a large-scale variation, while the residual $\varepsilon(\mathbf{s}, t)$ described random fluctuations at a small scale [16].

2.2. Components of the Trend

The deterministic component $m(\mathbf{s}, t)$ was estimated using multiple linear regression. Several studies have used this type of regression method to model the trend component in spatiotemporal geostatistics [1,13,17,18]. In this study, the geographic coordinates (latitude and longitude) and a temporal index employed to contour the effect of the annual seasonality on the precipitation were considered covariates. The adjusted trend model is given by the following:

$$\hat{m}(\mathbf{s}, t) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_5, \tag{2}$$

where $x_1 = cosx = cos((month/12) \cdot 2\pi)$ and $x_2 = senx = sin((month/12) \cdot 2\pi)$. The *cosx* and *senx* functions were used as covariates in the multiple regression model to explain the seasonal variation. The variables $x_3, x_4,$ and x_5 represent the latitude, longitude, and the quadratic effect of the longitude, respectively. Although the covariates exhibit spatial, temporal, and spatiotemporal variations, the regression model alone cannot account for all the variations. Therefore, the residuals of this model may exhibit spatiotemporal dependencies, indicating that a spatiotemporal variogram can be estimated and later used to interpolate the residues with kriging [19].

2.3. Spatiotemporal Variogram

After specifying the trend component and obtaining the parameter estimates, the results were then subtracted from Z, leaving the spatiotemporal stochastic residue ε. The residuals produced by this regression were then used for the construction of a sample variogram. For a more detailed model, the monthly rainfall (Z) was modeled as a sum of the trend and the spatiotemporal residue $\varepsilon(\mathbf{s}, t)$, as defined by the following:

$$Z(\mathbf{s}, t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon(\mathbf{s}, t) \tag{3}$$

After analyzing and removing the spatial and temporal trends in the data, the spatiotemporal interest is shown in terms of an empirical evaluation of the covariance function [14]. The spatiotemporal covariance was described using a spatiotemporal variogram, γ_{st}, which measured the mean difference between the separated data in the spatiotemporal domain using the distance vector $(\mathbf{h_s}, h_t)$. The spatiotemporal variogram was defined as follows:

$$\gamma_{st}(\mathbf{h_s}, h_t) = 0.5 \times E[\varepsilon(\mathbf{s}, t) - \varepsilon(\mathbf{s} + \mathbf{h_s}, t + h_t)]^2 \tag{4}$$

In Equation (4), $\mathbf{h_s} = \mathbf{s} - \mathbf{s}\prime$ and $h_t = t - t\prime$ for any $(\mathbf{s}, t)$ and $(\mathbf{s}\prime, t\prime)$ in the spatiotemporal domain. An estimation of the variogram $\gamma_{st}(\mathbf{h_s}, h_t)$ was established as follows:

$$\hat{\gamma}_{st}(\mathbf{r_s}, \mathbf{r_t}) = \left(2\left|L(\mathbf{r_s}, \mathbf{r_t})\right|\right)^{-1} \Sigma_{L(\mathbf{r_s}, \mathbf{r_t})} [\varepsilon(\mathbf{s} + \mathbf{h_s}, t + \mathbf{h_t}) - \varepsilon(\mathbf{s}, t)]^2 \tag{5}$$

According to the sample variogram presented in Equation (5), $\left|L(r_s, r_t)\right|$ is the cardinality for the set $L(r_s, r_t)$, that is, $L(\mathbf{r_s}, r_t) = \{(\mathbf{s} + \mathbf{h_s}, t + h_t) \in A, (\mathbf{s}, t) \in A : \mathbf{h_s} \in \mathrm{Tol}(\mathbf{r_s}) \text{ and } h_t \in \mathrm{Tol}(r_t)\}$, where $\mathrm{Tol}(\mathbf{r_s})$ and $\mathrm{Tol}(r_t)$ are the vectors of the spatial and temporal lags of $\mathbf{r_s}$ and r_t, respectively.

2.4. Generalized Product-Sum Model

In the next step of the analysis, a theoretical model was used to adjust the data based on the covariance functions of the spatiotemporal sample variogram. In this study, a non-separable model of a generalized product-sum [1,20] was used to model the spatiotemporal rainfall structure in the state of Paraíba. This model was written as follows:

$$C_{st}(\mathbf{h_s}, h_t) = kC_\mathbf{s}(\mathbf{h_s})C_t(h_t) + C_\mathbf{s}(\mathbf{h_s}) + C_t(h_t) \tag{6}$$

In Equation (6), C_s and C_t are the spatial and temporal covariance functions, respectively. The spatiotemporal variogram ($\gamma_{st}(\mathbf{h_s}, h_t)$) was expressed as follows:

$$\gamma_{st}(\mathbf{h_s}, h_t) = \gamma_{st}(\mathbf{h_s}, 0) + \gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h_s}, 0)\gamma_{st}(\mathbf{0}, h_t) \tag{7}$$

In Equation (7), $\gamma_{st}(\mathbf{h_s}, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$ were the spatial and temporal marginal variograms, respectively. The generalized product-sum model can be seen as a surface with six parameters: two parameters for each variogram (sill and range), and a joint spatiotemporal sill and nugget [2]. Also, the parameter k is positive and has the following identity that involves the global sill ($sill_{st}$), together with the spatial and temporal sills, namely, $sill_s$ and $sill_t$, respectively:

$$sill_{st} = k \cdot sill_s \cdot sill_t + sill_s + sill_t \tag{8}$$

In Equation (8), k was estimated by using the information of the sill. This approach constitutes an alternative that Graler et al. [20] described for the formulation of the product-sum model proposed in Equation (7). In this study, the authors proposed the following restructured Equation (7), which was used in this study:

$$\begin{aligned}\gamma_{st}(\mathbf{h_s}, h_t) = (k \cdot sill_t + 1)\gamma_{st}(\mathbf{h_s}, 0) + \\ (k \cdot sill_s + 1)\gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h_s}, 0)\gamma_{st}(\mathbf{0}, h_t)\end{aligned} \tag{9}$$

To fit a spatiotemporal theoretical variogram to a sample variogram, for the weighting of squared residuals in the least-squares estimation, we used the number of pairs in the spatiotemporal bin divided by the square of the current variogram model's value: $N[j]/\gamma_{st}(\mathbf{h_s}[j], h_t[j])^2$. The marginal variograms in this study were outlined based on the sample spatiotemporal variogram $\hat{\gamma}_{st}(\mathbf{r_s}, r_t)$. Through visual analysis, it was possible to define the theoretical models as being Gaussian, exponential, and spherical, which were components of Equation (9). In this way, it was possible to direct the structures for $\gamma_{st}(\mathbf{h_s}, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$.

2.5. Spatiotemporal Kriging

After establishing the trend function and the covariance for the residual, interpolation was performed using spatiotemporal regression kriging [16,18,21]. According to Kilibarda et al. [21], the variogram model is crucial in spatiotemporal kriging for calculating the best non-biased linear predictor, which was given by the following:

$$\hat{\varepsilon}(\mathbf{s_0}, t_0) = \mathbf{c_0^T C_n}^{-1}\overline{\varepsilon} \tag{10}$$

In Equation (10), $\hat{\varepsilon}(\mathbf{s_0}, t_0)$ was the linear predictor obtained using kriging for the location $(\mathbf{s_0}, t_0)$, $\mathbf{C_n}$ was the covariance matrix of order $n \times n$ of the residues for the n points observed in space-time as derived from the spatiotemporal variogram, $\mathbf{c_0}$ was a covariance vector of residues for the observed and predicted points, and $\overline{\varepsilon}$ was a vector of residues in the n observed points.

The final predictor ($\hat{z}(\mathbf{s_0}, t_0)$) for the rainfall variable (Z) at the location $(\mathbf{s_0}, t_0)$ was defined as follows:

$$\hat{z}(\mathbf{s_0}, t_0) = \hat{m}(\mathbf{s_0}, t_0) + \hat{\varepsilon}(\mathbf{s_0}, t_0) \tag{11}$$

In Equation (11), $\hat{m}(\mathbf{s_0}, t_0)$ is the estimated amount for the location $(\mathbf{s_0}, t_0)$ obtained by Equation (2), and $\hat{\varepsilon}(\mathbf{s_0}, t_0)$ is as defined in Equation (10). The error variance of the prediction in spatiotemporal kriging was defined as:

$$\sigma^2(\mathbf{s_0}, t_0) = C(0,0) - \mathbf{c_0^T C_n^{-1} c_0} + \left(\mathbf{m_0 - M^T C_n^{-1} c_0}\right)^{\mathbf{T}}\left(\mathbf{M^T C_n^{-1} M}\right)^{-1}\left(\mathbf{m_0 - M^T C_n^{-1} c_0}\right)$$

(12)

In Equation (12), $C(\mathbf{0}, 0)$ is the variance for Z, $\mathbf{M}$ is the matrix of order $n \times p$ of variable predictors at the observed points, and $\mathbf{m_0}$ is the vector of predictors at the prediction point.

Thus, kriging by regression was performed through the following steps:

Step 1: the values estimated using Equation (2) were extracted, followed by the extraction of the residues for all of the observations.

Step 2: After obtaining these residues, the values of the sample spatiotemporal variogram were calculated as defined in Equation (5), followed by the theoretical model fitting defined in Equation (9).

Step 3: The residues were then interpolated using Equation (10), after which, the trend component was added, resulting in the final predictor defined in Equation (11).

2.6. Selection and Performance of the Model

To determine the best theoretical variogram model that fit the sample variogram, the root mean square error (*RMSE*) and Nash–Sutcliffe efficiency (*NSE*) between the sampling variogram and the theoretical variogram were calculated [22]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[z_i(\boldsymbol{s}, \mathbf{t}) -- \hat{z}_i(\boldsymbol{s}, \mathbf{t})]^2}$$

(13)

$$NSE = \frac{\sum_{i=1}^{n}[z_i(\boldsymbol{s}, \mathbf{t}) - \overline{z}(\boldsymbol{s}, \mathbf{t})]^2 - \sum_{i=1}^{n}[z_i(\boldsymbol{s}, \mathbf{t}) -- \hat{z}_i(\boldsymbol{s}, \mathbf{t})]^2}{\sum_{i=1}^{n}[z_i(\boldsymbol{s}, \mathbf{t}) - \overline{z}(\boldsymbol{s}, \mathbf{t})]^2}$$

(14)

where, n is the number of observations, $z_i(\boldsymbol{s}, \mathbf{t})$ is the i-th observed value, $\hat{z}_i(\boldsymbol{s}, \mathbf{t})$ is the i-th predicted value, and $\overline{z}(\boldsymbol{s}, \mathbf{t})$ is the mean observed value.

After selecting the model that best fit the cloud of points from the sample variogram, spatiotemporal kriging was conducted to measure the performance of the adjusted model. Thus, interpolation was implemented by accounting for the leave-one-out method [14,18,23], as applied to all 269 rain gauge stations. This method consisted of removing an observed point from the rainfall in space-time $(z(s_0, \mathbf{t}))$ and then predicting its value $(\hat{z}(s_0, \mathbf{t}))$. This process was then repeated for all remaining points and residues of this process, after which the difference between the observed and predicted values is taken at each location $(z(s_0, \mathbf{t}) - \hat{z}(s_0, \mathbf{t}))$.

After the leave-one-out method was applied to measure the kriging performance, a scatter plot of the observed values versus the values predicted by spatiotemporal kriging was constructed, and the coefficient of determination was computed. It is worth noting that the application of the methodology proposed in this study allows for missing data to be entered. One of the advantages of spatiotemporal kriging relative to purely spatial kriging is that it not only interpolates to non-sampled locations but also predicts values for a time in which there was no data collection [1,24].

All statistical analyses were performed in software R [25], using the packages *gstat* [20] and *sp* [26].

3. Results

3.1. Exploratory Analysis

The descriptive measures and boxplots of rainfall values in millimeters (mm) for the 1994–2014 period are shown in Table 1 and Figure 3.

Table 1. Mean values, standard deviation (SD), and coefficient of variation (CV) of rainfall (mm) for the period 1994–2014 in the mesoregions of Zona da Mata, Agreste, Borborema, and Sertão.

Months	Zona da Mata			Agreste			Borborema			Sertão		
	Mean	**SD**	**CV (%)**	**Mean**	**SD**	**CV (%)**	**Mean**	**SD**	**CV (%)**	**Mean**	**SD**	**CV (%)**
January	86.6	79.8	92	62.9	83.8	133	61.5	85.9	140	124.7	116.4	93
February	99.4	79.1	80	66.2	62.8	95	64.9	61.6	95	133.2	87.7	66
March	121.1	85.2	70	85.8	62.3	73	106.6	86.5	81	187.6	115.1	61
April	173.0	111.7	65	95.5	70.1	73	84.2	76.1	90	156.1	106.9	69
May	191.1	121.1	63	100.2	77.2	77	67.8	66.7	98	92.9	76.8	83
June	273.3	146.2	54	131.0	80.2	61	44.9	41.5	93	34.0	36.5	107
July	190.3	114.8	60	106.0	76.2	72	26.3	26.3	100	17.1	19.6	114
August	116.3	67.2	58	64.8	48.7	75	14.4	18.7	130	6.8	14.3	210
September	59.9	68.9	115	29.5	41.7	141	4.6	9.5	207	2.1	7.7	359
October	20.8	19.2	92	9.8	13.3	135	7.8	23.6	302	13.0	30.3	233
November	17.7	18.1	103	12.0	18.9	158	6.1	16.0	260	13.5	25.8	192
December	31.7	30.8	97	20.6	23.8	116	20.0	29.7	149	41.5	45.6	110

Based on the average values and coefficient of variation, differences in rainfall distribution were observed between the analyzed months (Table 1). Moreover, within each month, the monthly rainfall exhibited a distinct behavior in each mesoregion. Similar results were reported in References [27,28]. Figure 3 demonstrates that each mesoregion had a distinct rainy period and a different mechanism for the rainfall distribution. For example, in the mesoregion of Sertão, the highest rainfall volume occurred from January to April, whereas the rainfall in Zona da Mata was well distributed from January to August. For the analyzed period, September presented the greatest variability in all the mesoregions. Through these results, the spatial and temporal variability in the rainfall throughout the state is well known.
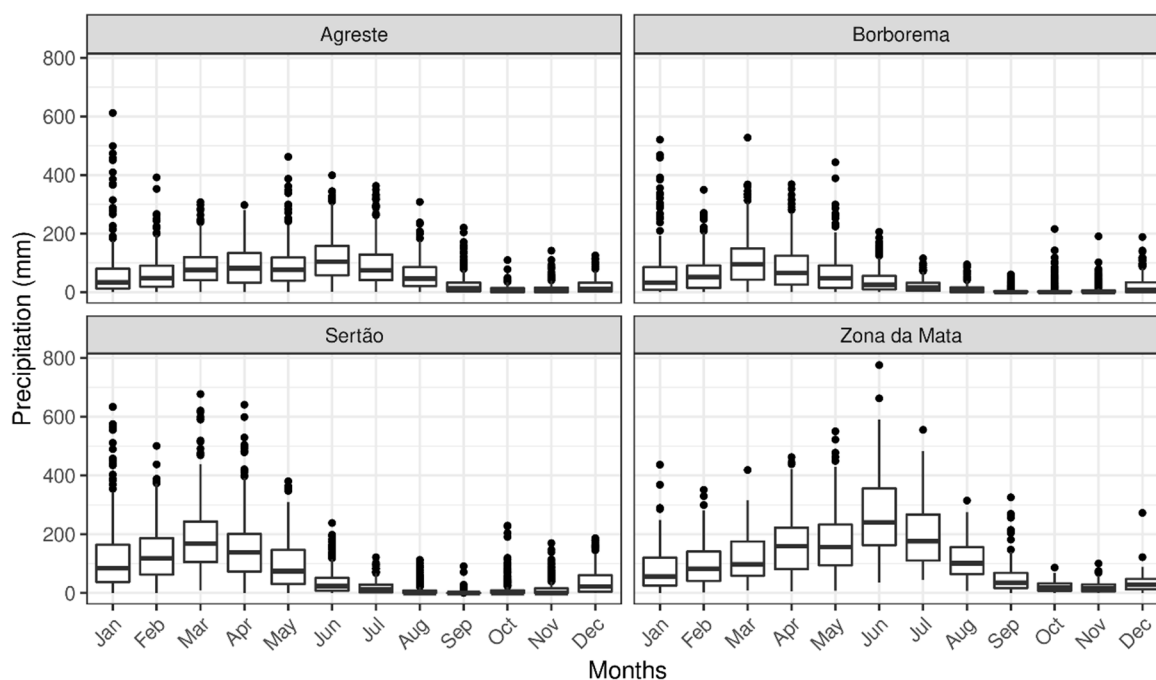

Figure 3. Boxplot for monthly rainfall for the period 1994–2014 in the mesoregions of Zona da Mata, Agreste, Borborema, and Sertão.

3.2. Regression Analysis for the Trend Component

The estimated parameters, standard error, statistical *t*, *p*-value, and adjusted coefficient of determination (R^2) for the monthly total rainfall for the 1994–2014 period are presented in Table 2.

Considering the results presented in Table 2, the covariates *cosx*, *senx*, latitude, and longitude, in addition to the quadratic effect of the longitude, are associated with precipitation at the 0.01 level of significance. The R^2 estimated by the model shows that this trend model explained 29% of the variation in the total monthly precipitation. This low R^2 strongly indicates that the trend component was unable to explain the spatiotemporal variability in the precipitation. Moreover, it was challenging to provide a precise estimate for climatic variables, especially precipitation, because the spatial and temporal distributions of such variables exhibited large variations, causing the trend model to present a low R^2 [15]. It is worth noting that the methodology proposed in this paper to estimate the precipitation accounts not only for the trend but also for the dependence of spatiotemporal data. Then, the residues produced by this trend component were modeled.

Table 2. Estimation of the parameters of the model described in the Equation (2) of the rainfall in the region analyzed from 1994 to 2014.

Variable	Estimation	Standard Error	*t*-Value	*p*-Value	R^2
Intercept	323.193	124.843	2.589	<0.01	
cosx	−19.981	0.854	−23.392	<0.01	
senx	52.554	0.854	61.525	<0.01	0.290
Latitude	0.061	0.013	4.884	<0.01	
Longitude	−2.308	0.079	−29.330	<0.01	
Longitude^2	0.002	0.001	30.082	<0.01	

A histogram of the residues in the multiple linear regression is shown in Figure 4. This is suitable for visual analysis of the residue behavior, which appears to have a positive asymmetry. In Figure 4, it is also possible to perceive the large scale of the variation in the residues, which is explained by the high variability in the precipitation.
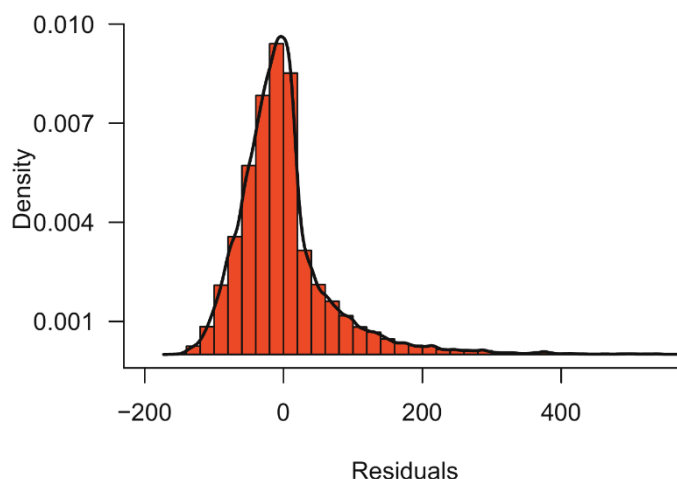

Figure 4. Histogram of the residues of the adjusted model.

Next, the residuals from this regression were used to analyze the spatiotemporal autocorrelation using the variogram described in Equation (5).

3.3. Spatiotemporal Variogram of the Residuals

After removing the spatiotemporal rainfall trend, the variogram was calculated using the residuals. The empirical spatiotemporal variogram model (Figure 5a) and the fitted product-sum model (Figure 5b) are shown in Figure 5. As shown, the residues indicated a clear correlation between both space and time, and spatial and temporal components explained the total variation in these residues. Note that

the spatial structure became weaker as the time differences increased, and the temporal structure became weaker as the spatial differences increased.
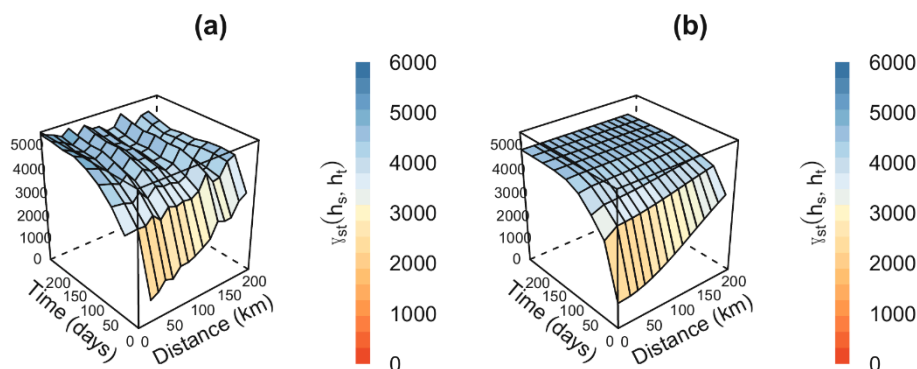

Figure 5. Sample spatiotemporal variogram (**a**) and the adjusted product-sum model (**b**) obtained from the residuals of the multiple linear regression.

The sample variogram can be visualized in terms of its marginal variograms, with one for space and another for time. The increasing trends in the spatial and temporal dimensions in the sample variogram indicated the presence of a strong spatiotemporal correlation. Hu et al. [2] stated that the characteristics present in Figure 5a further reinforced the need to apply spatiotemporal kriging to residues. In this sample spatiotemporal variogram, different structures were fitted for the generalized product-sum model class. Figure 5b shows one of these results when using a Gaussian model for space and a Gaussian model for time.

Subsequently, an exploratory analysis was performed for the empirical spatiotemporal variogram to construct the marginal variograms (Figure 6).
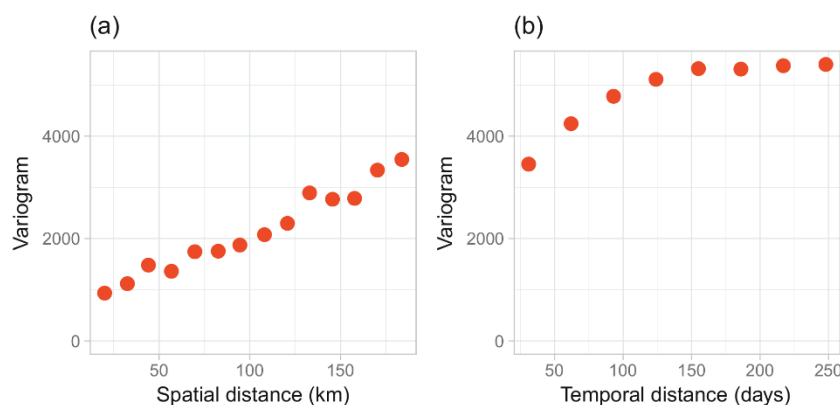

Figure 6. Marginal variograms of the residues: purely spatial (**a**) and purely temporal (**b**).

For the analyzed data, the spatial variability exhibited by the spatial marginal variogram (Figure 6a) was different from the temporal variability shown by the temporal marginal variogram (Figure 6b) because the two variograms present different sill values. Iaco et al. [1] observed characteristics similar to those presented in Figure 6a,b and stated that, in this situation, the spatiotemporal correlation of the residues could be adequately fitted using the generalized product-sum variogram model defined in Equation (9). In this case, as mentioned above, the structures of the spatial and temporal correlation were noticeable. However, some studies on rainfall distribution did not present this significant correlation; for example, Muthusamy et al. [29] developed a spatial variogram for the region of Bradford, England, but did not report a spatial correlation of the rainfall data.

Different variogram models were evaluated during the adjustment of the generalized product-sum model. For both space and time, the following models were used: Gaussian, exponential, and spherical.

3.4. Selection and Validation of the Model

The estimated parameters for the generalized product-sum model are presented in Table 3, were obtained by considering the different structures for the spatial and temporal components. The *RMSE* and *NSE* were used to compare the models. Some studies, such as References [1,2,21,24], used only a single spatiotemporal correlation structure. In this study, a more comprehensive analysis was proposed by using different structures in the model formation.

Table 3. Estimations of parameters (sill, range, and nugget) of the generalized product-sum variogram model fitted for the residuals and root mean square error (*RMSE*) and Nash–Sutcliffe efficiency (*NSE*). The parameter *k* involves the global sill.

	Model	Sill	Range	Nugget	*k*	*RMSE*	*NSE*
Space	Gaussian	7.648	181 km	1.713	23.480	10.207	0.903
Time	Exponential	26.683	43 days	8.446			
Space	Gaussian	4.140	178 km	0.930	25.129	10.155	0.905
Time	Spherical	45.686	143 days	24.184			
Space	Gaussian	8.591	180 km	1.933	19.861	10.150	0.905
Time	Gaussian	27.963	71 days	17.075			
Space	Exponential	6.889	192 km	0.616	23.486	10.646	0.899
Time	Gaussian	29.500	71 days	18.056			
Space	Exponential	4.380	192 km	0.400	23.614	10.667	0.900
Time	Spherical	46.065	147 days	24.232			
Space	Exponential	6.280	195 km	0.574	24.823	10.686	0.897
Time	Exponential	30.775	45 days	10.401			
Space	Spherical	1.804	388 km	0.237	15.857	10.341	0.901
Time	Exponential	162.584	42 days	49.179			
Space	Spherical	5.967	391 km	0.758	9.251	10.288	0.904
Time	Gaussian	85.384	70 days	51.693			
Space	Spherical	3.047	392 km	0.385	23.750	10.294	0.904
Time	Spherical	65.444	143 days	34.775			

For the generalized product-sum variogram model, the best structure, having the lowest *RMSE* and higher *NSE*, was found when considering the Gaussian model for both the spatial component and the temporal component. According to the selected model, the parameters related to the range showed spatial and temporal variations in the rainfall distribution over the state of Paraíba. The spatial range of 180 km suggested that the spatial correlation became insignificant after this sill, whereas the time interval of 71 days (approximately two months) indicated that the temporal correlation became insignificant after this interval.

After the best model was selected by adjusting the variogram, the cross-validation leave-one-out method was applied, and the results are presented below.

The spatiotemporal distribution of the mean *RMSE* values based on the 1994–2014 period is presented for each month in Table 4, in which high values indicate a high variability between the estimated and observed amounts in the rainfall distribution.

The highest values of the *RMSE* are shown between October and December. In the mesoregion of Sertão, smaller values of *RMSE* were noted between January and May, because it was during this period that the lowest variability of precipitation occurred (Table 1). These results are similar to those found in References [15,27]. The rainy seasons presented high spatial variability due to the actions of different weather systems, such as the Intertropical Convergence Zone (ITCZ) and Upper-Level Cyclonic Vortex (ULCV), as described by Macedo et al. [28].
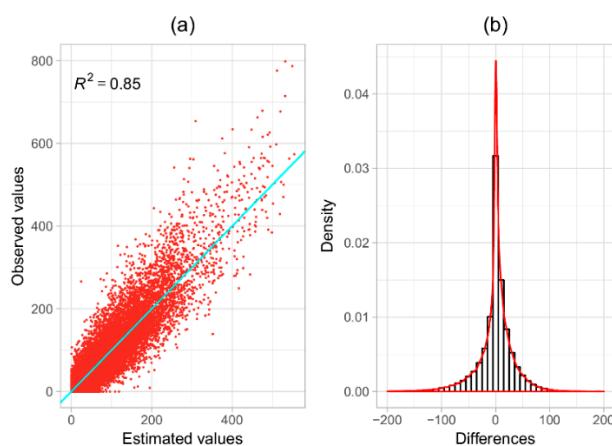
Table 4. Mean values of *RMSE* of rainfall (mm) for 1994–2014 in the mesoregions of Zona da Mata, Agreste, Borborema, and Sertão.

Month	Zona da Mata	Agreste	Borborema	Sertão
January	30.4	32.9	30.4	30.9
February	31.5	30.2	29.2	31.0
March	30.9	30.0	32.1	29.7
April	30.5	31.9	32.9	31.4
May	31.6	35.1	32.1	30.6
June	31.2	32.6	35.8	33.7
July	33.3	31.9	32.7	33.4
August	33.4	33.4	31.5	31.4
September	28.4	31.2	31.5	31.3
October	34.5	33.4	33.1	33.3
November	33.6	34.6	33.7	33.0
December	33.1	35.2	33.4	34.2

The months from January to March provided the lowest *RMSE* throughout most of the study region, indicating that the predicted values were close to the observed rainfall values. These results are consistent with Raja et al. [15], who reported that the lowest variability in Paraíba occurred during these months. Accordingly, it was also possible to observe a substantial monthly statistical variation.

To evaluate the quality of the fit for the model that was selected earlier, graphics were created using all of the values predicted using spatiotemporal kriging according to their actual values and the histogram of the difference between these values (Figure 7) considering all 269 rain gauge stations.

Figure 7 shows a strong linear relationship of the form $y = x$ between the estimated and observed values, which was supported by an R^2 value of 0.85, which indicates that 85% of the total precipitation variability could be explained using spatiotemporal kriging. Additionally, a visual analysis of the differences between these values presented in Figure 7a shows that the differences were distributed symmetrically with a frequency near zero. In other words, the scatter plot of the observed versus the estimated values during the 1994–2014 period, as well as the histogram of the differences between these values (Figure 7b), showed that the predictions were not significantly biased.

Figure 7. Comparison between the observed and estimated values (**a**) and the histogram of the difference between these values (**b**).

3.5. Spatiotemporal Kriging

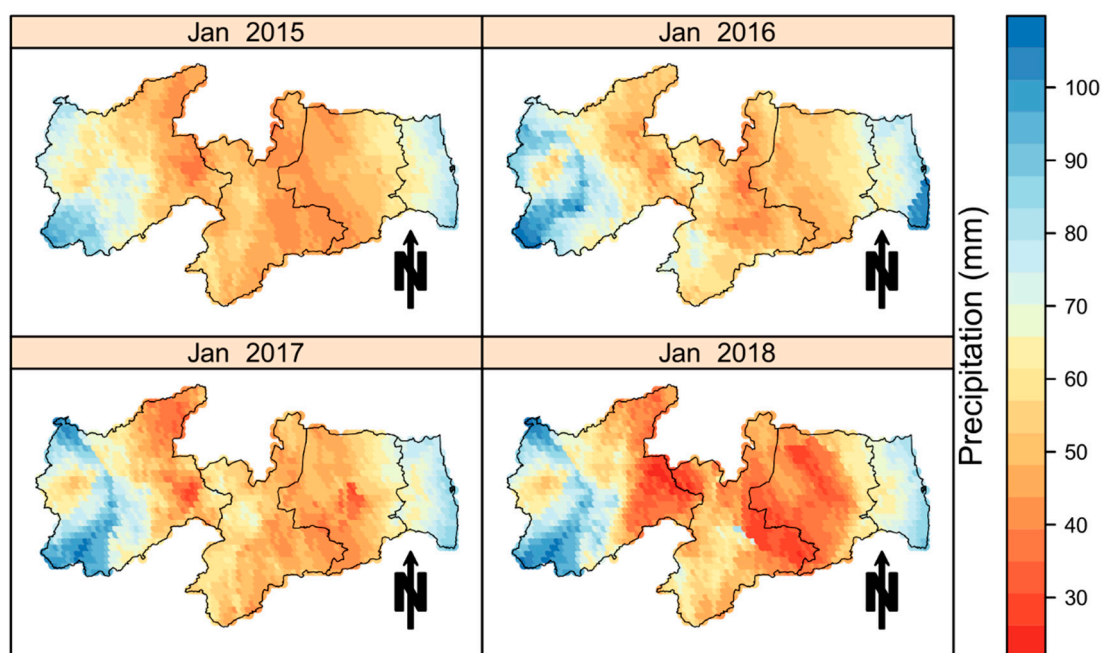Maps of the estimated rainfall for January in the years of 2015–2018 are shown in Figure 8.

Figure 8. Prediction maps of precipitation in January 2015–2018.

The application of the methodology proposed in this study presents a great advantage because it enables the use of spatiotemporal kriging to make predictions at any point in space-time within the chosen region. Thus, it enables an evaluation of how rainfall distribution patterns change with regard to both space and time, in addition to presenting values in places that are not monitored by weather stations. This application is illustrated in Figure 8, which shows the predicted amount of spatiotemporal rainfall in the state of Paraíba for January in the years of 2015–2018. January was chosen because it presented the highest rainfall values throughout the entire region, and it showed high variabilities in both space and time. The predicted rainfall presented strong spatial and temporal patterns in the study region, evidencing the success achieved when applying the proposed geostatistical spatiotemporal methodology.

Regardless of the selected period, the highest rainfall level was distributed in the Zona da Mata mesoregion (Figure 8). There was little variation in precipitation in the north–south direction. However, high variability was observed among predictions in the east–west direction. Finally, the greatest variation in rainfall was found in the Sertão mesoregion of Paraíba. In Figure S1, which is available in the Supplementary Material, the predicted rainfall values are shown for all months within the 2015–2020 period.

In addition to the spatiotemporal kriging-related predictions for January from 2015 to 2018, Figure 9 shows maps of the standard error representing the uncertainty in the rainfall predictions in the chosen period. The error showed variations in the east–west direction. Moreover, within each mesoregion, there were variations in the estimates. It is also worth noting that the Borborema mesoregion presented the highest indexes for the standard error of the prediction; this behavior was attributed to the more intense rainfall in this area.
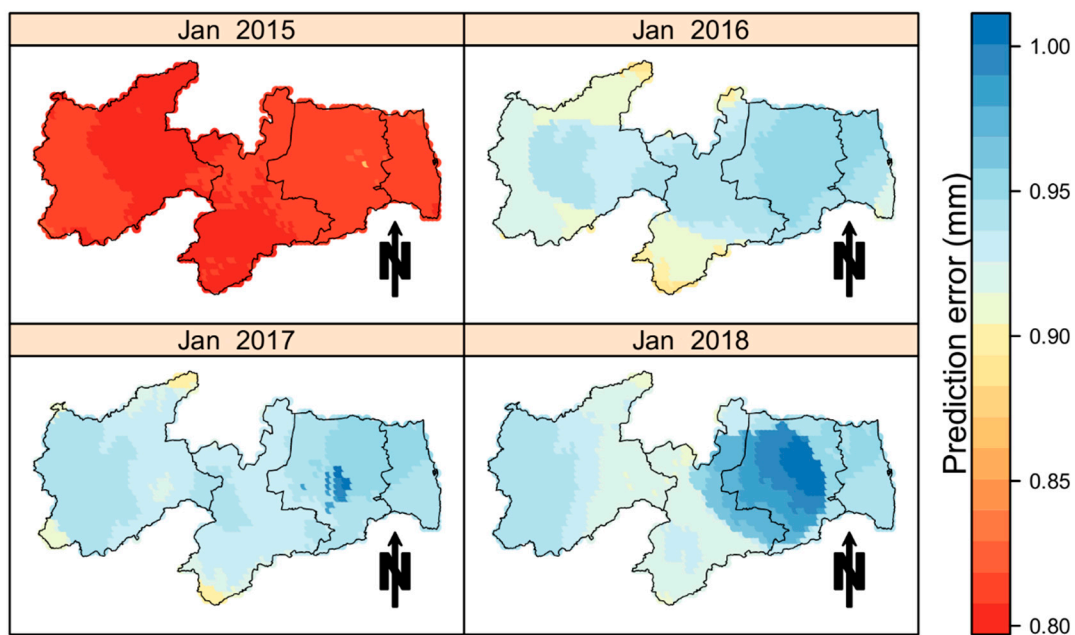
Figure 9. Maps of the standard error of precipitation prediction in January 2015–2018.

4. Discussion

In this study, rainfall in the state of Paraíba was spatiotemporally modeled using well known geostatistical techniques, such as kriging. Similar studies involving spatial [30] or temporal [31] analyses of rainfall distributions have already been performed for the state of Paraíba, although they did not take into account the combined spatiotemporal variability in this phenomenon. Thus, this research presents a joint spatiotemporal study of the rainfall distribution in this state.

Here, the trend component represented 29% of the temporal variability in the rainfall. Studies in References [13,17,18] indicated that modeling this trend component influenced the fitting of the spatiotemporal variogram of the residues. Accordingly, this study could be improved by considering environmental factors, such as the temperature, in the adequate description of the trend component; this would likely result in an expected increase in the phenomenon variability percentage being modeled. Thus, the model for the trend only somewhat explained the variability in the data because it accounted for less than half of the variability of the spatiotemporal precipitation in the study area. Several studies have shown that the integration of climatic variables into precipitation predictions generates models with greater accuracy, thereby reducing error metrics and increasing R^2 values [15,32]. However, there is a lack of meteorological variables for the study region. For the state of Paraíba, only eight weather stations provided recorded data, and it was impossible to use a few of those stations in this work because their data could negatively affect the predictions in the model. For future studies, the results obtained in this study could be compared with those derived from the application of data obtained using simulations. Another important topic addressed in this research is the ability to explain the variability in the data beyond just the trend component; this analytical ability was achieved by adjusting the covariance functions for the residuals obtained by adjusting the trend because they had space-time dependence [14,21].

The rainfall distribution in the state of Paraíba showed a clear correlation for both space and time, thereby justifying the use of spatiotemporal kriging. When constructing the spatiotemporal variogram of the residues, several researchers found that spatial dependence became weaker with an increase in time, while the temporal dependence became insignificant over vast spatial distances [19,21,27,33]. In the studies mentioned above, the modeled phenomena did not present common characteristics in the spatiotemporal structure during the construction of the variogram. After the removal of the trend component, the rainfall data showed a clear spatial correlation structure over different temporal

distances and a temporal correlation structure over different spatial distances. From a theoretical perspective, some studies fitted models to the marginal sample variograms obtained from an empirical spatiotemporal variogram and then inserted the individual fittings into the product-sum model [1,24]. Nevertheless, this study proposes a method to fit the spatiotemporal theoretical model because spatiotemporal interactions directly impact the parameter estimations; when theoretical models are fitted to the marginal variograms, there is no interference in the estimation from this interaction. However, this type of procedure results in problems during the optimization and determination of the parameter estimations. Examples of studies that employed this type of fitting can be found in References [2,34].

This research represents a breakthrough for climatological studies because it employs a technique that is more accurate than traditional statistical methods. The application of the statistical methodology proposed in this paper constitutes a significant contribution toward understanding the variability in precipitation and water resources of a region. The proposed approach considers that the spatiotemporal variability is decomposed by the trend component and the stochastic residue. However, the method presents some limitations, one of which is the difficulty in capturing the variabilities between the locations separated by great distances in the dataset. Additionally, for future work involving this type of modeling, it will be necessary to include auxiliary variables, such as the humidity and temperature, for the trend component to present better metrics.

The interpolation of climate data is widely used for research and management purposes in several disciplines, including social science [1], hydrology [14], and forestry [33]. The importance of both spatial and temporal modeling of the rainfall distribution in the state of Paraíba is crucial for the management of high-priority areas that may experience a scarcity of rainfall caused by periods of drought or instances of flooding caused by excess rainfall, thereby inducing environmental and socioeconomic damages [35]. The application of spatiotemporal kriging to obtain rainfall records from unobserved sites should provide favorable conditions for the planning and management of water resources because it will be possible to analyze regions that have a high impact on the state economy. For example, Biondi [36] performed spatiotemporal rainfall kriging and presented results that directly addressed the planning needs for the distribution of water and natural resources once the occurrence of drought in a river basin has been diagnosed because it was possible to analyze which geographic areas were more likely to be affected by severe droughts.

5. Conclusions

The rainfall mechanisms throughout the analyzed region presented variability over both space and time. Furthermore, the spatiotemporal kriging of precipitation clarified the dramatic spatial and temporal variability of water resources among the mesoregions throughout the state, thereby providing a detailed visual analysis of sectors experiencing precipitation conditions ranging from water scarcity to an excess of rainfall.

In this study, a rainfall data analysis methodology was proposed through spatiotemporal geostatistics and was applied to irregularly spaced weather stations by considering the rainfall distribution over both space and time. Notably, the rainfall variability may be related to spatial and weather covariates, such as the temperature, humidity, and altitude, among others, suggesting the need to include such covariates for future studies to model the effect of the rainfall trend. Moreover, the monthly rainfall in the analyzed region presented a series of records with null values, which affected the application of a standard regression model to a certain extent. Exploring these issues in detail will require the development of a non-normal model that enables a better explanation regarding the variability in the rainfall distribution throughout the region. This model will constitute the next investigative step and it will be reported soon.

The results achieved in this research should motivate further research, including studies throughout the northeastern region of Brazil, to make predictions to diagnose which sites might experience droughts or floods.

Author Contributions: Conceptualization, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.; Methodology, E.S.d.M., R.R.d.L., and R.A.d.O.; Software, E.S.d.M.; Validation, E.S.d.M. and R.R.d.L.; Formal analysis, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.; Investigation, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.; Resources, E.S.d.M.; Data curation, E.S.d.M. and C.A.C.d.S.; Writing—original draft preparation, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.; Writing—review and editing, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.; Visualization, E.S.d.M.; supervision, E.S.d.M., R.R.d.L., R.A.d.O., and C.A.C.d.S.

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

References

1. Iaco, S.D.; Palma, M.; Posa, D. Spatio-temporal geostatistical modeling for French fertility predictions. *Spat. Stat.* **2015**, *14*, 546–562. [CrossRef]
2. Hu, Y.; Li, R.; Bergquist, R.; Lynn, H.; Gao, F.; Wang, Q.; Zhang, S.; Sun, L.; Zhang, Z.; Jiang, Q. Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003470. [CrossRef]
3. Lopes, I.; Santos, S.M.; Leal, B.G.; Melo, J.M.M. Variation of aridity index and climatic trend to desertification for the semi-arid region of the Brazilian Northeast. *Rev. Bras. Geogr. Física* **2017**, *10*, 1014. [CrossRef]
4. de Almeida, H.A.; Medeiros, E.A. Variabilidade no regime pluvial em duas mesorregiões da Paraíba e sua relação com o fenômeno EL Niño Oscilação Sul. *J. Environ. Anal. Prog.* **2017**, *2*, 177. [CrossRef]
5. Gomes, O.M.; dos Santos, C.A.C.; Souza, F.; de Paiva, W.; de Olinda, R.A. Análise Comparativa da Precipitação no Estado da Paraíba Utilizando Modelos de Regressão Polinomial. *Rev. Bras. Meteorol.* **2015**, *30*, 47–58. [CrossRef]
6. Tabios, G.Q.; Salas, J.D. A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation. *J. Am. Water Resour. Assoc.* **1985**, *21*, 365–380. [CrossRef]
7. Phillips, D.L.; Dolph, J.; Marks, D. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agric. For. Meteorol.* **1992**, *58*, 119–141. [CrossRef]
8. Hevesi, J.A.; Flint, A.L.; Istok, J.D. Precipitation Estimation in Mountainous Terrain Using Multivariate Geostatistics. Part II: Isohyetal Maps. *J. Appl. Meteorol.* **1992**, *31*, 677–688. [CrossRef]
9. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129. [CrossRef]
10. Martínez-Cob, A. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *J. Hydrol.* **1996**, *174*, 19–35. [CrossRef]
11. Goovaerts, P. Using elevation to aid the geostatistical mapping of rainfall erosivity. *Catena* **1999**, *34*, 227–242. [CrossRef]
12. Haberlandt, U. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.* **2007**, *332*, 144–157. [CrossRef]
13. Hu, D.; Shu, H.; Hu, H.; Xu, J. Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Clust. Comput.* **2017**, *20*, 347–357. [CrossRef]
14. Martínez, W.A.; Melo, C.E.; Melo, O.O. Median Polish Kriging for space–time analysis of precipitation. *Spat. Stat.* **2017**, *19*, 1–20. [CrossRef]
15. Raja, N.B.; Aydin, O.; Türkoğlu, N.; Çiçek, I. Space-time kriging of precipitation variability in Turkey for the period 1976–2010. *Theor. Appl. Climatol.* **2017**, *129*, 293–304. [CrossRef]

16. Heuvelink, G.B.M.; Griffith, D.A.; Hengl, T.; Melles, S.J. Sampling Design Optimization for Space-Time Kriging. In *Spatio-Temporal Design*; John Wiley & Sons, Ltd.: Chichester, UK, 2012; pp. 207–230. ISBN 9781118441862.

17. Gräler, B.; Gerharz, L.; Pebesma, E. Spatio-temporal analysis and interpolation of PM10 measurements in Europe. *ETC ACM Tech. Pap. 2011/10* **2012**, *8*, 1–29.

18. Hengl, T.; Heuvelink, G.B.M.; Perčec Tadić, M.; Pebesma, E.J. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor. Appl. Climatol.* **2012**, *107*, 265–277. [CrossRef]

19. Monteiro, A.; Menezes, R.; Silva, M.E. Modelling spatio-temporal data with multiple seasonalities: The NO2 Portuguese case. *Spat. Stat.* **2017**, *22*, 371–387. [CrossRef]

20. Graler, B.; Pebesma, E.; Heuvelink, G. Spatio-Temporal Interpolation using gstat. *WP* **2016**, *8*, 1–20. [CrossRef]

21. Kilibarda, M.; Tadić, M.P.; Hengl, T.; Luković, J.; Bajat, B. Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. *Spat. Stat.* **2015**, *14*, 22–38. [CrossRef]

22. Pushpalatha, R.; Perrin, C.; Le Moine, N.; Andréassian, V. A review of efficiency criteria suitable for evaluating low-flow simulations. *J. Hydrol.* **2012**, *420–421*, 171–182. [CrossRef]

23. Kilibarda, M.; Hengl, T.; Heuvelink, G.B.M.; Gräler, B.; Pebesma, E.; Perčec Tadić, M.; Bajat, B. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *J. Geophys. Res. Atmos.* **2014**, *119*, 2294–2313. [CrossRef]

24. Bahrami Jovein, E.; Hosseini, S.M. Predicting saltwater intrusion into aquifers in vicinity of deserts using spatio-temporal kriging. *Environ. Monit. Assess.* **2017**, *189*, 81. [CrossRef] [PubMed]

25. Core Team, R. R: A language and environment for statistical computing. *R A Lang. Environ. Stat. Comput.* **2018**. [CrossRef]

26. Bivand, R.S.; Pebesma, E.; Gómez-Rubio, V. Hello World: Introducing Spatial Data. In *Applied Spatial Data Analysis with R*; Springer: New York, NY, USA, 2013; pp. 1–16.

27. Menezes, H.E.A.; De Brito, J.I.B.; Lima, R.A.F.D.A. Veranico e a produção agrícola no Estado da Paraíba, Brasil. *Rev. Bras. Eng. Agrícola E Ambient.* **2010**, *14*, 181–186. [CrossRef]

28. Macedo, M.J.; Guedes, R.; de Sousa, F.A.; Dantas, F. Analysis of the standardized precipitation index for the Paraíba state, Brazil. *Ambiente E Agua Interdiscip. J. Appl. Sci.* **2010**, *5*, 204–214. [CrossRef]

29. Muthusamy, M.; Schellart, A.; Tait, S.; Heuvelink, G.B.M. Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1077–1091. [CrossRef]

30. dos Santos, C.A.C.; Gomes, O.M.; de Souza, F.A.S.; Paiva, W. De Análise Geoestatística da Precipitação Pluvial do Estado da Paraíba (Geostatistical Analysis of Precipitation in Paraiba State). *Rev. Bras. Geogr. Física* **2012**, *4*, 692. [CrossRef]

31. Dantas, L.G.; Santos, C.A.C.S.; Olinda, R.A. de Resampling series rainfall in the state of Paraiba. *Rev. Bras. Geogr. Física* **2016**, *9*, 997–1006.

32. Apaydin, H.; Anli, A.S.; Ozturk, F. Evaluation of topographical and geographical effects on some climatic parameters in the Central Anatolia Region of Turkey. *Int. J. Climatol.* **2011**, *31*, 1264–1279. [CrossRef]

33. Tonini, F.; Dillon, W.W.; Money, E.S.; Meentemeyer, R.K. Spatio-temporal reconstruction of missing forest microclimate measurements. *Agric. For. Meteorol.* **2016**, *218–219*, 1–10. [CrossRef]

34. Pebesma, E.J. Multivariable geostatistics in S: The gstat package. *Comput. Geosci.* **2004**, *30*, 683–691. [CrossRef]

35. Tavares Diniz, J.M. Variabilidade da precipitação e do número de dias com chuvas de duas cidades distintas da Paraíba. *HOLOS* **2013**, *3*, 171. [CrossRef]

36. Biondi, F. Space-time kriging extension of precipitation variability at 12 km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrol. Earth Syst. Sci. Discuss.* **2013**, *10*, 4301–4335. [CrossRef]