



Research Article

The assessment of annual rainfall field by applying different interpolation methods in the state of Rio Grande do Sul, Brazil



Denis Rafael Silveira Ananias¹ · Gilberto Rodrigues Liska² · Luiz Alberto Beijo³ ·
Geraldo José Rodrigues Liska⁴ · Fortunato Silva de Menezes⁵

Received: 15 January 2021 / Accepted: 27 May 2021

Published online: 06 June 2021

© The Author(s) 2021

Abstract

An accurate analysis of spatial rainfall distribution is of great importance for managing watershed water resources, in addition to giving support to meteorological studies and agricultural planning. This work compares the performance of two interpolation methods: Inverse distance weighted (IDW) and Kriging, in the analysis of annual rainfall spatial distribution. We use annual rainfall data for the state of Rio Grande do Sul (Brazil) from 1961 to 2017. To determine which proportion of the sample results in more accurate rainfall distribution maps, we use a certain amount of points close to the estimated point. We use mean squared error (MSE), coefficient of determination (R^2), root mean squared error (RMSE) and modified Willmott's concordance index (md). We conduct random fields simulations study, and the performance of the geostatistics and classic methods for the exposed case was evaluated in terms of precision and accuracy obtained by Monte Carlo simulation to support the results. The results indicate that the co-ordinary Kriging interpolator showed better goodness of fit, assuming altitude as a covariate. We concluded that the geostatistical method of Kriging using nine closer points (50% of nearest neighbors) was the one that better represented annual rainfall spatial distribution in the state of Rio Grande do Sul.

Keywords Environmental planning · Cross-validation · Geostatistics · Inverse distance weighted · Ordinary Kriging

1 Introduction

Rainfall is a measure (indicator) of an ecosystem's water availability and has strong relationships with the productivity of a region [1, 2]. Variables related to precipitation such as average, maximum and annual variability, among others, are important for explaining spatial patterns of anomalies, as well as allowing assessments of climate change by increasing the frequency of extreme events

[2], Al-Yaari et al.[3]). Its monitoring makes it possible to understand the hydrological cycle that influences ecological and environmental dynamics, also affecting economic and social activities (Morales and Araujo [4]).

The analysis of climatic variables essentially consists of two stages: an exploratory one, using descriptive statistics in order to verify the normality of the data and to discard the need for transformation in the set and also to identify the existence of possible outliers; and

✉ Gilberto Rodrigues Liska, gilbertoliska@ufscar.br; Denis Rafael Silveira Ananias, denisrsananias@gmail.com; Luiz Alberto Beijo, luiz.beijo@unifal-mg.edu.br; Geraldo José Rodrigues Liska, geraldo.liska@unifal-mg.edu.br; Fortunato Silva de Menezes, fmenezes@dfi.ufla.br | ¹Federal University of Pampa, campus Itaqui, Rua Luiz Joaquim de Sá Britto, s/n - Bairro Promorar, Itaqui, RS 97650-000, Brazil. ²Department of Agroindustrial Technology and Rural Socioeconomics, Federal University of São Carlos, Araras, Rod. Anhanguera km 174 - SP-330, Araras, SP 13600-970, Brazil. ³Department of Statistics, Federal University of Alfenas, Rua Gabriel Monteiro da Silva, 700, Alfenas, MG 37130-001, Brazil. ⁴Federal University of Alfenas, Rua Gabriel Monteiro da Silva, 700, Alfenas, MG 37130-001, Brazil. ⁵Department of Physics, Federal University of Lavras, Lavras, MG 37200-900, Brazil.



the second, which is based on the adjustment of mathematical statistical models to the data, which can study the phenomena with different approaches, as well as the occurrence of extreme values [5–7], temporal distribution (Pereira Britto, Barletta and Mendonça [8]), spatial distribution [9], intensity of the phenomenon [10], among others.

According to Pereira Britto et al. [8], even though the state of Rio Grande do Sul is industrially developed, its economy is dependent on agriculture, and despite the great advances that have taken place in this sector during the last years, agricultural activity and crop yields depend on the occurrence of rainfall. The state has a large part of its territory in the La Plata Basin, which has a sparse and irregular rainfall network [11]. Therefore, in order to have knowledge about the regime or behavior of precipitation in these locations, it is necessary to spatially distribute the precipitation. In this situation, it is essential to apply a spatial interpolation process where points with known values are used to estimate unknown values at other points.

Geostatistics began in the 1960s with a series of publications by Georges Matheron with theoretical bases on a method of spatial interpolation called Kriging developed by South African Daniel G. Krige [12]. Geostatistics is a set of statistical methods appropriate for analyzing an attribute of a phenomenon with continuous distribution over a geographic area.

Among the various methods of interpolation are the inverse distance weighted (IDW), Kriging, closest neighbor, spline and top-to-raster methods. According to Chirinos & Mallqui [13], the most used interpolators are Kriging, inverse distance and spline. The results obtained in each of the methods may be different for the same set of data. The inverse distance weighted is a univariate weighted interpolator that uses the distance between the points sampled to estimate an attribute of interest and has been highlighted by the ease of its applicability. Several studies prove its applicability. However, simulation studies involving these methods to investigate the influence of the number of points on the accuracy of the estimates are uncommon [14, 15]. The number of points used to obtain estimates through interpolators is directly related to the algorithm's execution time, representing a significant concern from the computational point of view [16], Das et al. [17]).

The Kriging using a continuous covariance function, which explains the behavior of the variable in space, stands out in the literature for presenting non-biased estimates and the minimum variance associated with the estimated value [18].

Choosing interpolators for climate data is not a simple task and the end result does not just depend on the interpolator. The interpolation process requires knowledge about

the nature of the data to be interpolated and the spatial distribution of the samples [19].

Spatial analyzes involving climatic events are rare and require simulation studies [20]. In studies of this type, it is desirable that the random variables have a symmetrical distribution in order to satisfy the first- and second-order stationarities. From a practical point of view, in the meteorological studies, other statistics such as maximum and median are usual, their probability distributions are not necessarily symmetrical, and more complex geostatistical methodologies are required [21, 22].

In the state of Rio Grande do Sul, Brazil, the analysis of spatial events has been based on specific regions [23, 24]. Understanding the spatial distribution of data from phenomena that occur in space today constitutes a major challenge for the elucidation of central issues in several areas of knowledge [9]. In order to analyze the spatial distribution of rain, we usually use interpolators that can be classical or geostatistical. The interpolation methods are used to evaluate the spatial variability of a given attribute based on sample data located in a locality of interest (Gardiman [25]). According to Nashwan & Shahid [26], although there is a certain degree of reliability in the analyzes using grid data, the uncertainty in the results is still a considerable problem and arises mainly due to factors, such as the interpolation of observed data, the model used for data generation, the number of stations used for the development of models, in addition to the quality of the observed data.

Because the state of Rio Grande do Sul is a region of high importance for the economy of Brazil and has most of its activities dependent on the occurrence of rains, it is important to know the spatial behavior of different variables related to annual precipitation. In addition, there are no studies for this region, indicating (analyzing) different techniques used in the spatial analysis of rainfall in order to have more accurate and accurate results. Given these facts, the main objective of the present work is to evaluate the performance of different interpolation methods in the estimation of spatially dependent annual rainfall levels in the state of Rio Grande do Sul. Therefore, it sought to answer the following questions: (i) which interpolation method is best suited to estimate spatially dependent annual rainfall levels in the state of Rio Grande do Sul? (ii) how do the interpolators behave in the use of different statistics (mean, median and maximum)? (iii) to predict spatially, more precisely, rainfall, how many neighboring points are needed?

2 Materials and methods

Rio Grande do Sul is located in the extreme south of Brazil (Fig. 1), with a territorial area of 281,737.888 km² [27]. The relief is represented by plateaus of the Paraná Basin,

Peripheral Depression, Eastern Plateau of Rio Grande do Sul and Patos and Mirim lagoons plains (Ross [28]). According to Sartori [29],

Due to its location in a transition zone, the climate of Rio Grande do Sul reflects the participation of Extratropical Atmospheric Systems (polar masses and fronts) and intertropical (Tropical masses and Disturbed Currents), although the first control the types of weather in 90% of the days of the year, providing also the monthly and annual distribution of rainfall.

According to the Köppen system, Rio Grande do Sul falls into the fundamental temperate zone, denoted by "C", and humid, denoted by "Cf". Due to the altimetric differences, the state's climate split into "Cfa" and "Cfb". The oceanic climate, with mild summers (Cfb), occurs in the mountains of the southeast and northeast, where the average temperatures of the summer months are below 22 °C, and the Cfa type, or humid subtropical climate, occurs in the other regions, where the average temperature is warmer, exceeding 22 °C in the month with high temperatures [30].

The data set was obtained from the meteorological database for teaching and research (BDMEP) provided by National Institute of Meteorology (INMET). Information from 18 conventional monitoring stations were used

and cover the period from 1961 to 2017, totaling 56 years of analysis (Table 1). The variable considered was the monthly cumulative rainfall, the data were separated by year and the maximum, mean and median of each season were extracted, thus forming a subset with the annual averages, medians and maximums values of each weather station. The georeferenced points that form the polygon of Rio Grande do Sul were extracted from a file of the type Shapefile made available by the Brazilian Institute of Geography and Statistics (IBGE). The percentage of failures was calculated by the ratio between the number of monthly records in which there was no reading and the total number of monthly records. We emphasize that the choice for the stations, as mentioned above, was because they are the stations with the most considerable amount of records available in the state according to the BDMEP.

2.1 Inverse Distance Weighted (IDW)

The IDW method, univariate interpolator, is a classical interpolator that uses the distance between the points sampled to determine the value of the points of interest. Thus, to estimate a point not sampled at a given location, we used Eq. (1) described by

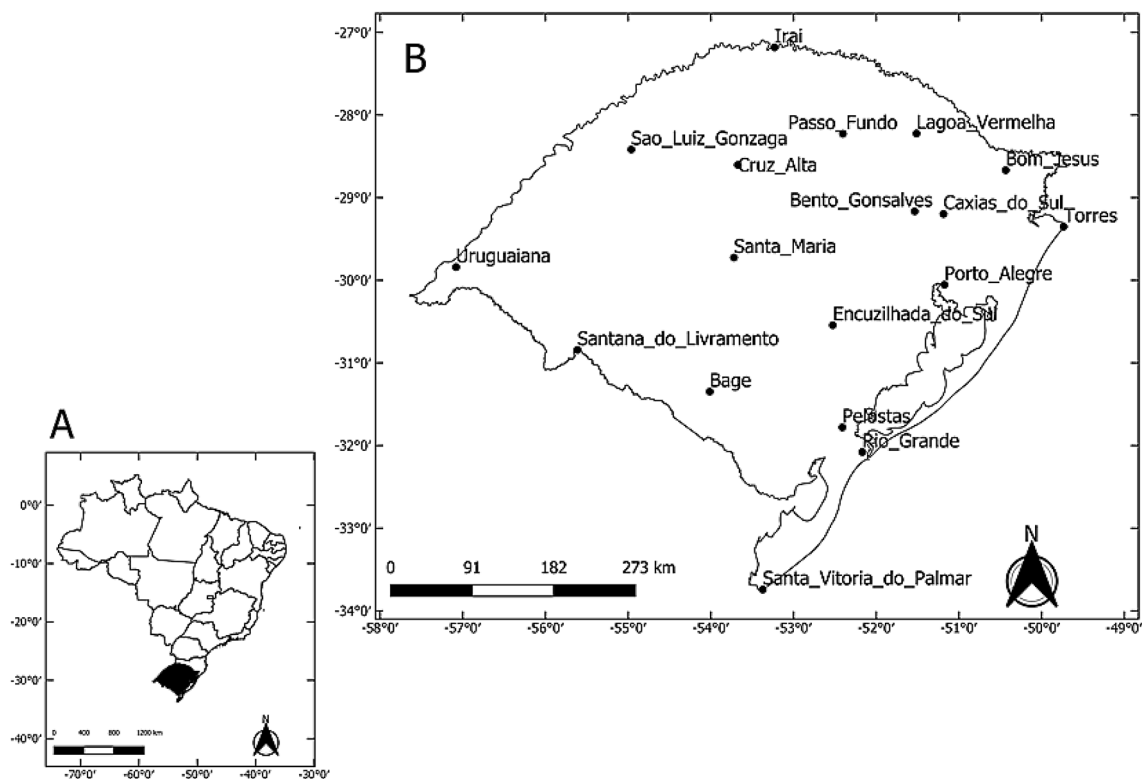


Fig. 1 a Location of the state of Rio Grande do Sul, Brazil; b Location of the 18 weather stations in the state

Table 1 Data from the meteorological stations evaluated in the state of Rio Grande do Sul, Brazil, as well as the beginning and end dates of the series and percentage of failures

Station ID	City	Latitude	Longitude	Altitude (m)	Beginning of the series	End of series	% failure
83,980	Bagé	-31.347801	-54.013292	245.66	01/31/61	12/31/17	9.44
A840	Bento Gonçalves	-29.164581	-51.534202	623.27	01/31/61	12/31/13	22.31
83,919	Bom Jesus	-28.666667	-50.433333	1048.59	01/31/61	12/31/17	23.24
83,942	Caxias do sul	-29.196389	-51.186389	750.16	01/31/61	12/31/17	14.38
83,912	Cruz Alta	-28.60344	-53.673597	475.52	01/31/61	12/31/17	14.57
83,964	Encruzilhada do Sul	-30.543164	-52.524669	428.52	01/31/61	12/31/17	14.19
83,881	Iraí	-27.18	-53.23	247.1	01/31/61	12/31/17	12.69
83,916	Lagoa Vermelha	-28.222381	-51.512845	833.73	01/31/61	12/31/17	44.00
83,914	Passo Fundo	-28.226805	-52.403582	684.05	01/31/61	12/31/17	10.32
83,985	Pelotas	-31.78	-52.41	12.01	01/31/61	12/31/17	65.62
83,967	Porto Alegre	-30.053536	-51.174766	41.18	01/31/61	12/31/17	7.04
83,995	Rio Grande	-32.07878	-52.167738	2.46	01/31/61	12/31/15	16.94
83,936	Santa Maria	-29.72496	-53.720469	103.1	01/31/61	12/31/17	12.13
83,997	Santa Vitória do Palmar	-33.742297	-53.372218	24.01	01/31/61	12/31/17	14.77
83,953	Santana do Livramento	-30.842449	-55.613089	328	01/31/61	12/31/13	76.67
83,907	São Luiz Gonzaga	-28.417113	-53.962403	245.49	01/31/61	12/31/17	11.22
83,948	Torres	-29.350359	-49.733263	8.47	01/31/61	12/31/17	11.95
83,927	Uruguaiana	-29.83987	-57.081899	62.31	01/31/61	12/31/17	19.60

$$Z(u_i) = \frac{\sum_{i=1}^n \frac{Z(u_i)}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \tag{1}$$

where d_i is the distance between the observation pairs (u_i), p is a power parameter generally equal to two and n represents the number of sampled points used for the estimation [10]. In Eq. (1), more significant power parameters emphasize closer points, making the result less smooth. Smaller power parameters emphasize more distant points, making the result smoother but less accurate (Marcuzzo, Andrade and Melo [31]). As the distance increases, weights diminish significantly when the value of the power parameter rises. Nearby stations have a heavier weight more influence on the estimation [32].

2.2 Kriging methods

Kriging is a geostatistical method based on the Theory of Regionalized Variables, which assumes that phenomenon variation depends on its location. A semivariogram quantifies the spatial variation in the Kriging method. The semivariogram is, in turn, calculated by Eq. (2).

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^n [Z(u_i) - Z(u_i + h)] \tag{2}$$

in which, h is the vector distance between pairs of observations, $N(h)$ is the number of ordered pairs and $Z(u_i)$ and $Z(u_i + h)$ are values observed at respective locations.

The empirical semivariogram estimated by $\hat{\gamma}(h)$ provides a graph to which a function fits, thus generating a theoretical semivariogram. This function should best represent the behavior of $\hat{\gamma}(h)$. Omnidirectional semivariograms were used.

Three theoretical semivariogram models were considered: spherical, exponential and Gaussian models. All of these models have as parameters: range (a), distance within which the samples are spatially correlated; partial sill (C) is the maximum value of the semivariogram within its range; effect nugget (C_0) is the value of semivariogram for distance zero.

The adjustment of the theoretical models was done by the methods of ordinary least squares (ols) and weighted least squares (wls), to determine which would best fit the graph generated by $\hat{\gamma}(h)$. The choice of the best semivariogram was aided by the Cambardella criterion, which quantifies the degree of spatial dependence [33]. The degree of dependence is obtained by expressing the nugget effect as a percentage of the partial sill, so semivariograms with a nugget effect $\leq 25\%$ of the partial sill are considered with strong spatial dependence, moderate dependence when the nugget effect is between 25 and 75% and weak dependence nugget effect is $> 75\%$.

The ordinary Kriging method (OK) is used when the regionalized average of the data is unknown. In order to estimate the amount of annual rainfall at a non-sampled location ($\widehat{Z}(u_i)$), the following ordinary Kriging estimator was used,

$$\widehat{Z}(u_i) = \sum_{i=1}^n \lambda^{KM} Z(u_i) + \sum_{i=1}^n \lambda^{OK} \left[Z(u_i) - \sum_{i=1}^n \lambda^{KM} Z(u_i) \right] \quad (3)$$

where $Z(u_i)$ is the observed value, λ^{OK} refers to the weights calculated on the basis of n data $Z(u_i)$ and $\sum_{i=1}^n \lambda^{KM} Z(u_i)$ is the estimator of the average for each region. Further details on the estimation of the mean can be seen in Yamamoto [12].

The universal Kriging (UK) is used to model a trend in rainfall in longitude and/or latitude, when this happens to occur. Instead of considering the regional average $m(u)$ unknown, which can be estimated by an average around a region with the n closest points (ordinary Kriging), a first-order or higher-order polynomial for the trend and/or covariates is set (E. J Pebesma and Bivand [34]). Assuming that we have a first-degree bias in n points, then

$$m(u) = \beta_0 + \beta_1 x(u) \quad (4)$$

where x may be the direction of the trend. Similarly, if a covariate (secondary variable) is analyzed joint with the annual rainfall, the co-ordinary Kriging (COK) estimator can be considered and is given by

$$\widehat{Z}(u_i) = \sum_{i=1}^n \lambda^{OK} Z(u_i) + \sum_{j=1}^m \beta_{j_i}^{OK} S_k(u_{j_i}) \quad (5)$$

where $S_k(u_{j_i})$ is the observed value from secondary variable and $\beta_{j_i}^{OK}$ are the ordinary Kriging weights. We used the $k=1$ case for altitude as secondary variable.

The influence of the neighboring points in the prediction of non-sampled points is verified through the maps of the IDW and Kriging methods constructed considering fractions of points near the ones that will be estimated, given by 10, 20, 30, 50, 70 and 100% of the set of points data. The fractions formed by the 18 points are equivalent to 2, 4, 6, 9, 13 and 18 neighbors.

2.3 Evaluation of the models

The quality of the estimates was evaluated by the leave-one-out method, which is a particular case of k -fold, also known as cross-validation. According to Andriotti [35], the leave-one-out method consists of taking a sample point from the data set and calculating the estimate for that point. Repeating the procedure for all points in the sample set and at the end comparing the known point

with the estimate gives the mean square error (MSE) of these discrepancies. In this sense, five methods were used to evaluate the residuals calculated by the cross-validation, MSE (6), R^2 (7), square root of the MSE (RMSE) and modified Willmott's concordance index (md) (8),

$$MSE = \frac{\sum_{i=1}^n (Z(u_i) - \widehat{Z}(u_i))^2}{n} \quad (6)$$

$$R^2 = 1 - \frac{SQR}{SQT} = 1 - \frac{\sum_{i=1}^n (Z(u_i) - \widehat{Z}(u_i))}{\sum_{i=1}^n (Z(u_i) - \bar{Z}(u_i))^2} \quad (7)$$

$$md = 1 - \frac{\sum_{i=1}^n |Z(u_i) - \widehat{Z}(u_i)|}{\sum_{i=1}^n (|Z(u_i) - \bar{Z}(u_i)| + |\widehat{Z}(u_i) - \bar{Z}(u_i)|)}, \quad (8)$$

where $\bar{Z}(u_i)$ is the mean and $\widehat{Z}(u_i)$ is the fitted value. The md index is less sensitive to the presence of outliers and has a range between 0 and 1, where values close to 1 better indicate the model's performance in the prediction [36].

We conducted a computational simulation study to evaluate the performance of the geostatistical and classical methods used. The Monte Carlo simulation method was used, which consists of making several achievements of a phenomenon according to pre-established parameters. At the end of these simulations, we can calculate the mean and standard deviation of the simulations, and these represent measures of accuracy and precision, respectively [37, 38]. Thus, two scenarios are considered:

1. Random fields generated considering spatial dependence structure according to the best result of the geostatistical method;
2. Random fields generated considering the spatial dependence structure different from the best results obtained by the geostatistical method.

In the proposed methodology we use, the R version 4.0.0 [39], the packages `geoR` [40], `gstat` (E. J. [41]), `sp` (E. J. Pebesma and Bivand [34]), `rgdal` (Bivand and Keitt [42]), `hydroGOF` [43] and `ggplot2` [44].

3 Results

All rainfall series are homogeneous by Levene's test considering 1% as significance level (Fig. 2). The size of the dataset is an essential issue in the analysis of the data. As

shown in Table 1, there is an imbalance in the amount of data among the weather stations. The occurrence of technical problems at the stations has meant that many other stations in the state have a much lower number of records than our work, which is similar in quantity to the work of Alvares et al. [30]. Given these considerations and given the focus of the work, we analyze annual rainfall data. However, we emphasize that this approach constitutes one possibility of analysis, and others such as monthly, seasonal and/or cumulative annual analysis could be considered [11, 32, 45, 46]. The plots of the exploratory analyses are shown in Fig. 3. In the mean and median histograms, it is possible to notice a slight negative asymmetry, -0.231671 and -0.230965 , respectively, while for the maximum, the asymmetry is positive 0.583762 . This result suggests that the data may not follow a normal distribution. However, it is possible to assume the normality of the data before the results of the Shapiro–Wilk test [47], 1% of significance, with the following p values: for a maximum of 0.5466 , for a mean of 0.3167 and a median of 0.06111 .

Figure 4 shows the relation between the rainfall variable with the latitude and longitude coordinates, where it is possible to verify that there is no trend around the longitude for any of the measures under study. For latitude, the mean and median tend as the coordinates increase the observed values. This behavior indicates that the trend can be modeled.

Table 2 shows the semivariogram fitted values, and Fig. 5 shows the plots. We analyzed the degree of spatial dependence calculated by Cambardella and the square root of the mean square error to choose the method and model that best fits the empirical semivariogram. The values of the Cambardella test are all within the same range of less than 25%, indicating strong spatial dependence [33]. For the variables analyzed in OK, the Gaussian model is the one that presented the lowest value of the RMSE for the mean and the median, fitted by the weighted least squares (wls) method. For the maximum, the exponential model fitted by ordinary least squares (ols) was the best. In the UK, the mean was adjusted by

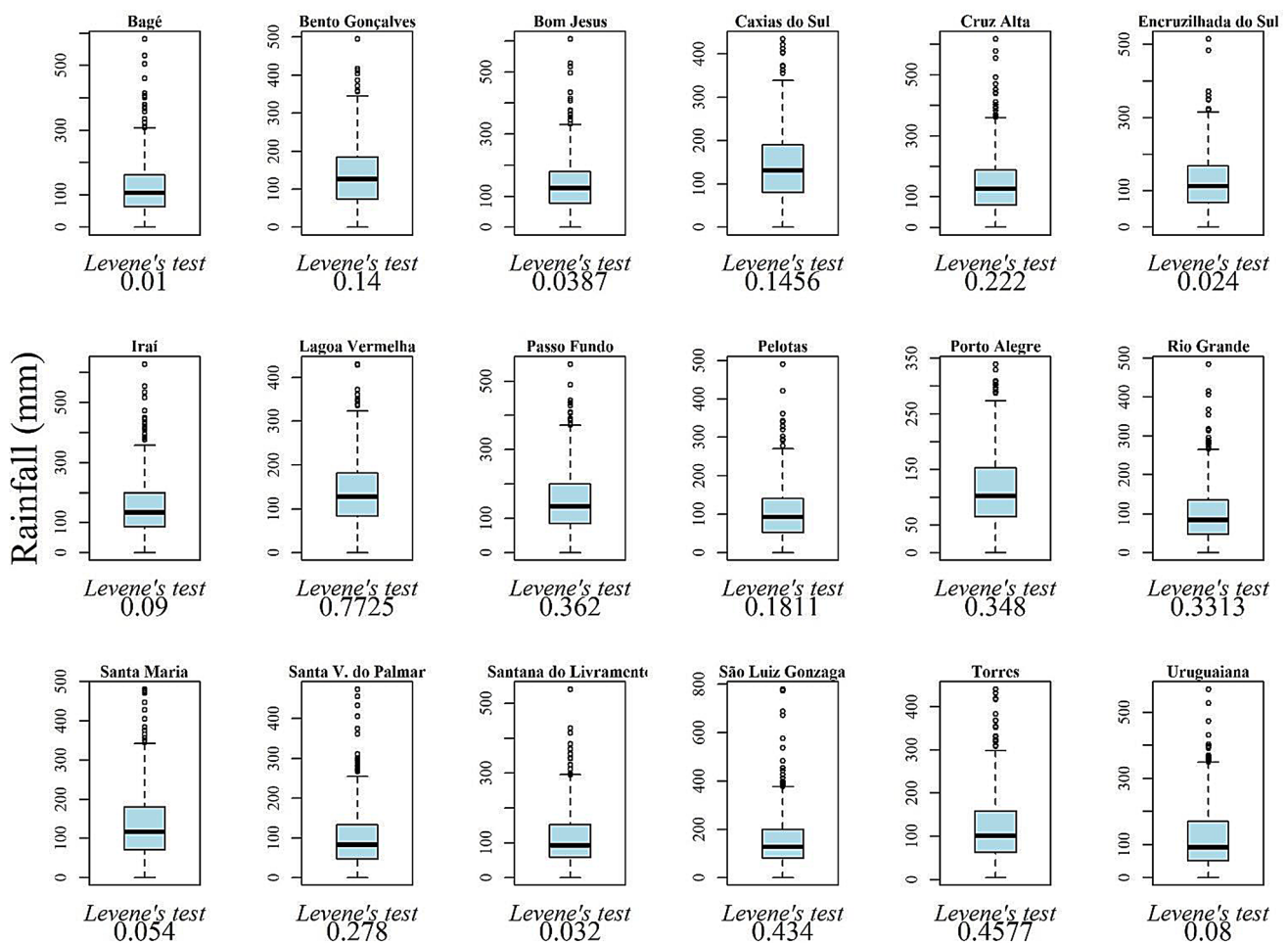


Fig. 2 Box plots of the daily rainfall series in each weather station evaluated according to Table 1 and Levene's homogeneity test (p -value)

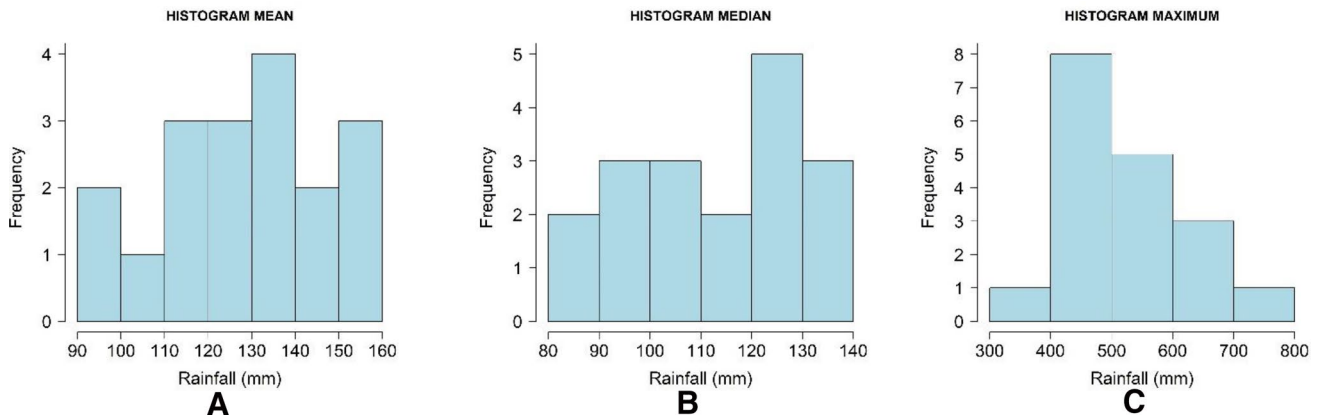


Fig. 3 Histograms of annual rainfall in mm. **a** Histogram of the mean; **b** histogram of the median; **c** histogram of the maximum

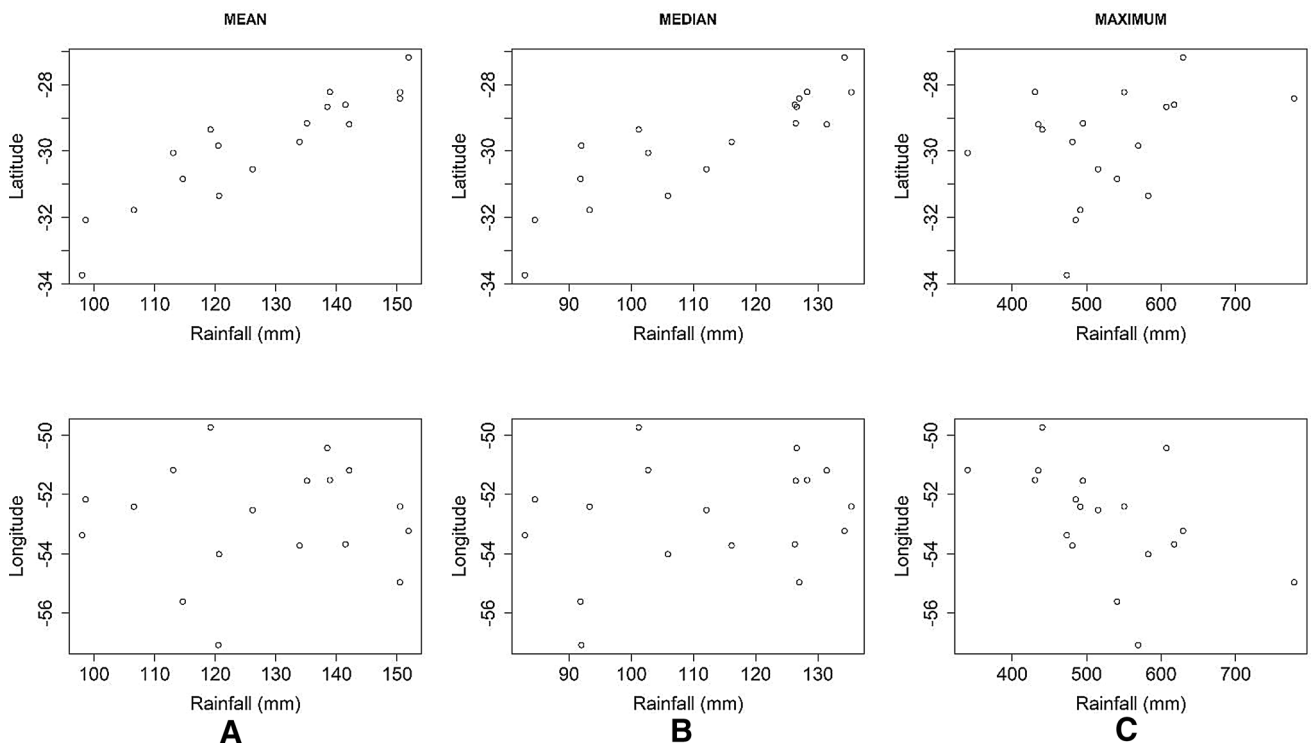


Fig. 4 Plots of mean **a**, median **b** and maximum **c** annual rainfall with latitude and longitude coordinates

the spherical model by wls, and the median with the Gaussian model by ols.

Figure 6 shows the generated maps for IDW and Kriging. It is possible to notice that the estimation surfaces provided by the IDW show some regions of rainfall well delimited generating some islands or points located in the vicinity of the sampling points. The maps provided by Kriging have smoother surfaces and evidence of transition zones between different levels of rainfall.

The rainfall variable has no relation with the longitude coordinates. It was possible to verify that there is no trend

around the longitude for any of the measures under study. For latitude, the mean and median have a trend as the coordinates increase the observed values, indicating that the trend can be modeled. The universal Kriging was used to adjust the first-degree polynomial, explaining this feature in the rainfall estimates.

Table 3 shows the cross-validation results for the mean and median. For the mean, the models presented the lowest MSE with 50% of the data set, which equals nine points. According to Andriotti [35], the number of neighboring samples required depends on the

Table 2 Estimates of the parameters (C_0), contribution (C) and range (a) of the semivariograms by the ordinary (ols) and weighted (wls) least squares methods and results of the Cambardella criterion and root mean squared error (RMSE)

Variable	Method	Model	C_0	C	a	Cambardella	RMSE
Ordinary Kriging							
Mean	wls	Gaus	40.07	2065.70	7.62	1.94	6.95
Median	wls	Gaus	23.56	1023.56	4.76	2.30	7.18
Maximum	wls	Exp	3977.14	12,612.37	5.76	23.97	8.76
Universal Kriging							
Mean	wls	Sph	2.13	48.82	0.83	4.18	5.14
Median	ols	Gaus	14.23	117.72	3.12	10.78	107.23
Maximum	wls	Sph	1179.15	9385.49	4.35	11.16	7.17
Co-Ordinary Kriging*							
Mean	wls	Gaus	43.83	700.94	4.70	5.88	397.67
Median	wls	Gaus	43.16	764.60	4.70	5.34	479.12
Maximum	wls	Exp	1931.82	21,929.13	5.10	8.10	16,492.98

* Results related to the fitted annual rainfall semivariogram. The results of the altitude and cross semivariogram will be omitted because of the work focus

configuration of locations sampled and the degree of anisotropy, and for irregularly spaced samples is 10 samples. OK presented the best result with a 3.1 mm error in relation to IDW, corresponding to 31.62%, lower than the IDW and a determination coefficient of 84.12%.

This study shows a slight difference in the MSE values, indicating OK as the most appropriate for the mean. Analyzing the values of the RMSE it is possible to notice that the difference of the OK and UK errors is of 0.88 mm of rainfall. This result indicates that considering the assumed modeling trend did not bring substantial improvement in Kriging estimates, which leads to consider the equivalence between them. Thus, the trend observed in the data may have more relation with other variables such as altitude. This result is in agreement with Baratto & Wollmann [48] who studied the rainfall profile of state of Rio Grande do Sul. According to the authors, it is possible to trace two profiles in the state, one in the south–north direction and the other in the west–north-east direction, and conclude that rainfall is influenced by orography and this influence is more significant in the north–south direction between Santa Maria and Júlio de Castilho cities. The results for COK were better than OK and UK, which shows that considering the covariate altitude for rainfall has brought substantial improvements in decreasing the prediction error (Table 3).

Concerning the median, the interpolators diverged on the ideal number of neighboring neighbors. The MSE lowest value for OK was obtained with 70% of the points. The best MSE results were obtained for the UK using 100% of the set. In IDW, the lowest RMSE was with 50% of the data set, but for other numbers of points, there were no significant discrepancies between the RMSE, showing that for the median, the IDW showed approximately consistent results (Table 3).

The values of the determination coefficient (R^2) related to the best results of Kriging, ordinary and universal, of the mean and median are more significant than 0.7. This value of R^2 means that the choice of the number of points is correctly determined. The R^2 values of the IDW, 0.6975 for mean and 0.6828 for median, indicate that the estimated values are moderately close to the real ones.

The modified Willmott’s index of agreement shows that the co-ordinary Kriging approach is more appropriate in all predictions evaluated, unless in the case of the maximum, as $md < 0.6$, as pointed out by Da Silva Moraes et al. [49], for all approaches.

Three points not sampled in different regions of the state were chosen to verify if the estimates for these locations are within the mean error margin of the OK and COK models. For the city of Itaquí, located on the western border, considering the mean annual rainfall between January and December 2019, the model underestimated the rainfall by 4.05 and 2.04 mm. In same way, for the city located in the south of the state, Jaguarão, the observed rainfall refers to the period from 2019, was underestimated at 2.59 mm (OK) and 2.49 mm (COK). Federico Westphalen at the north at 7.86 mm (OK) and 10.11 mm (COK). In this case, both models slight overestimate the mean annual rainfall (Table 4).

The annual maximum rainfall analysis presented high values for the MSE (Table 3). The lowest result for MSE was obtained using the IDW method using 50% of the data set equivalent to the nine points closest to the estimated point. The RMSE expresses the error in the same proportion of the variable. Thus, the best fit presented an average error of 85.37 mm of rainfall. The low values for the coefficient of determination (R^2) are direct results of the MSE values and, at most, can explain 0.2295 of the rainfall that occurred.

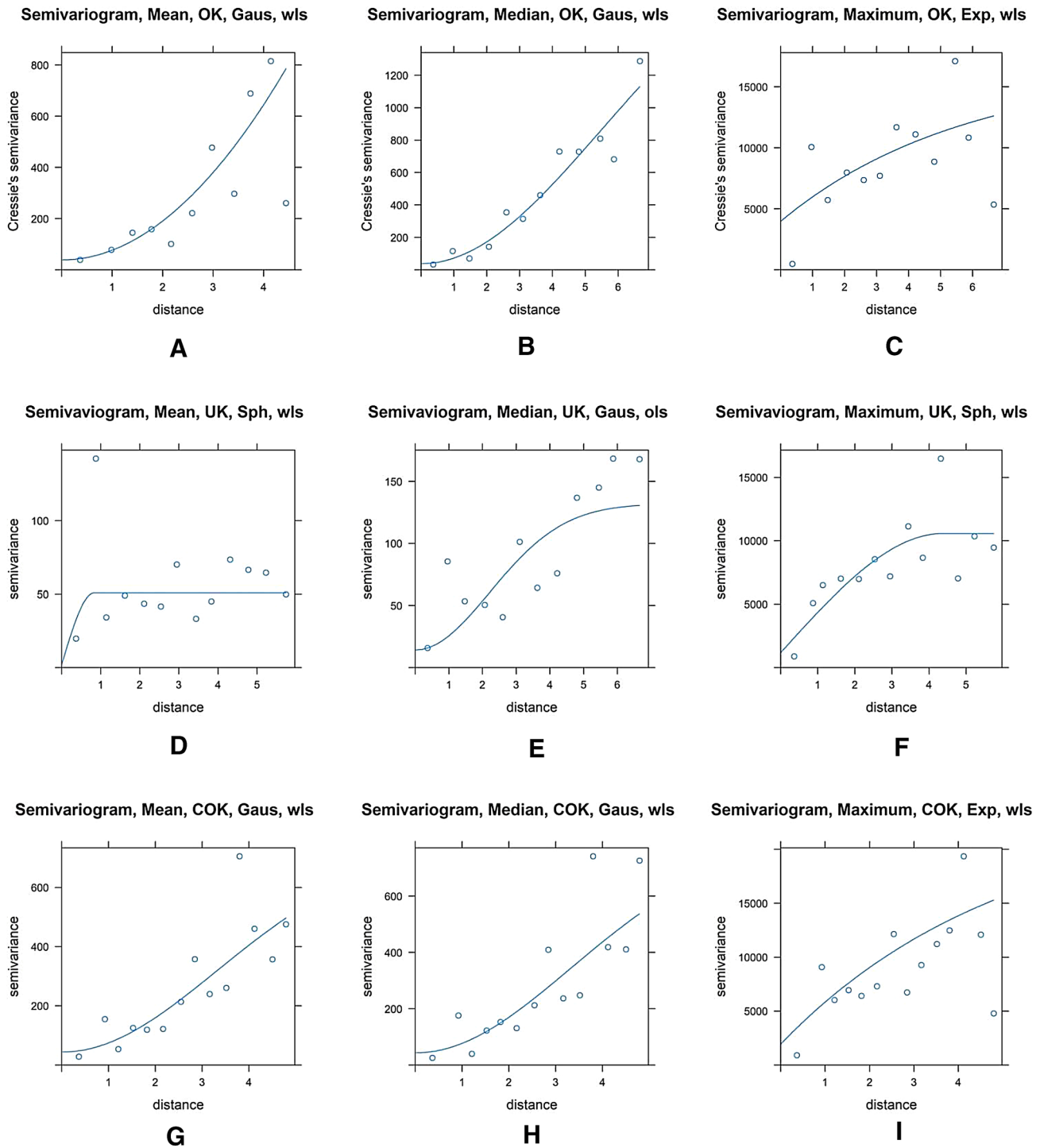


Fig. 5 Fitted theoretical semivariograms to the empiricals. **a** semivariogram of the mean for ordinary Kriging (OK), **b** semivariogram of the median for OK, **c** semivariogram of the maximum for OK, **d** semivariogram of the mean for universal Kriging (UK), **e** semivariogram of the median for UK, **f** semivariogram of the maximum for UK, **g** semivariogram of the mean for co-ordinary Kriging (COK), **h** semivariogram of the median for COK, **i** semivariogram of the maximum for COK

ogram of the median for UK, **f** semivariogram of the maximum for UK, **g** semivariogram of the mean for co-ordinary Kriging (COK), **h** semivariogram of the median for COK, **i** semivariogram of the maximum for COK

Since R^2 values are far below 0.7, it is clear that the models do not apply to maximum values. These values can be approached by other probability distributions

that are not considered in this work (Table 3). It should be noted that since Fig. 4 does not provide an indicative trend for maximum rainfall in the direction of latitude

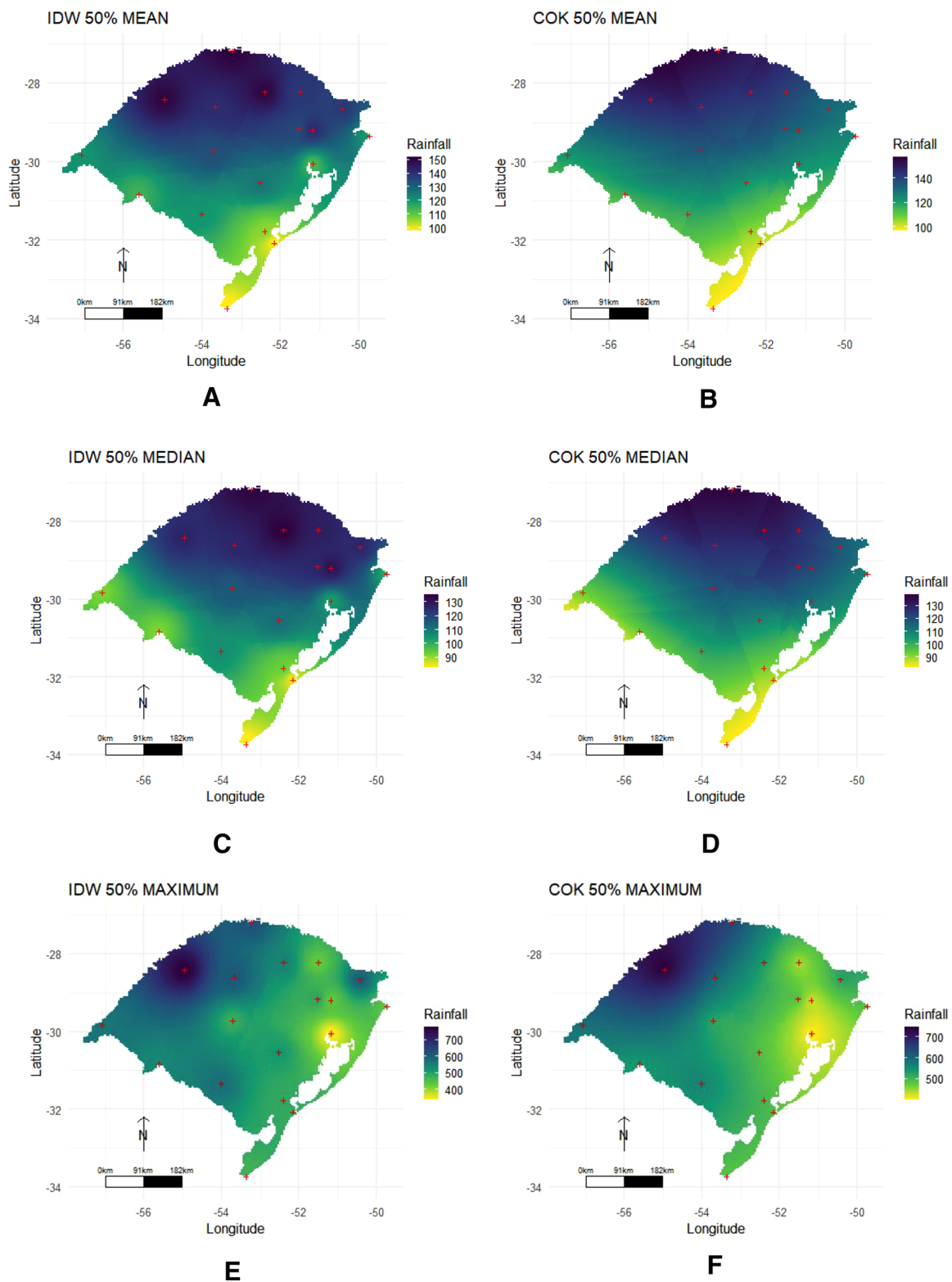


Fig. 6 Prediction maps of annual rainfall levels. **a** Map of the mean generated with the IDW using 50% of the data set; **b** map of the mean generated with co-ordinary Kriging (COK) using 50% of the data set; **c** map of the median generated with the IDW using 50%

of the dataset; **d** map of the median generated with co-ordinary Kriging using 50% of the data set; **e** map of the maximum generated with the IDW using 50% of the data set; **f** map of the maximum generated with co-ordinary Kriging using 50% of the data set

Table 3 Results of cross-validation for mean, median and maximum, mean squared error (MSE), root mean squared error (RMSE), determination coefficient (R^2) and modified Willmott's index of agreement (md) for the evaluated methods (OK—ordinary Kriging, UK—universal Kriging, COK—co-ordinary Kriging, IDW—inverse distance weighted)

N° of points	Method	MSE	RMSE	R^2	md	MSE	RMSE	R^2	md	MSE	RMSE	R^2	md
		Mean				Median				Maximum			
10%	OK	118.43	10.88	0.5869	0,70	122.21	11.06	0.6236	0.72	8653.10	93.02	0.1557	0.44
	UK	3202.61	56.59	0.0578	0.45	2401.90	49.01	0.1717	0.49	141,713.3	376.45	0.2836	0.32
	COK	73.78	8.59	0.7462	0.72	45.90	6.77	0.8526	0.80	7406.13	86.06	0.2608	0.49
	IDW	117.71	10.85	0.5935	0,69	129.42	11.38	0.6054	0.71	8706.01	93.31	0.1342	0.43
20%	OK	81.13	9.01	0.7160	0,71	75.27	8.68	0.7580	0.74	8342.93	91.34	0.1499	0.43
	UK	104.62	10.23	0.7010	0.72	117.67	10.85	0.7098	0.72	9405.63	96.98	0.1016	0.42
	COK	65.01	8.06	0.7805	0.75	47.18	6.87	0.8548	0.80	6302.66	79.39	0.3116	0.50
	IDW	102.28	10.11	0.6486	0,68	118.04	10.87	0.6471	0.70	7837.34	88.53	0.1658	0.42
30%	OK	62.52	7.91	0.7825	0,75	66.54	8.16	0.7876	0.76	7447.98	86.30	0.2080	0.45
	UK	70.62	8.40	0.7733	0.76	81.92	9.05	0.7569	0.75	7579.92	87.06	0.2288	0.48
	COK	54.64	7.39	0.8124	0.77	42.74	6.54	0.8671	0.81	5915.53	76.91	0.3494	0.53
	IDW	105.99	10.30	0.7569	0,66	124.84	11.17	0.6252	0.68	7650.53	87.47	0.1738	0.42
50%	OK	45.05	6.71	0.8412	0,76	55.52	7.45	0.8207	0.78	7298.51	85.43	0.2205	0.46
	UK	58.09	7.62	0.8022	0.77	78.95	8.89	0.7680	0.76	7670.88	87.58	0.2044	0.47
	COK	49.00	7.00	0.8345	0.79	42.11	6.49	0.8691	0.82	5379.89	73.35	0.4082	0.55
	IDW	96.35	9.82	0.6975	0,67	113.89	10.67	0.6828	0.69	7288.61	85.37	0.2085	0.42
70%	OK	46.25	6.80	0.8382	0,73	50.92	7.14	0.8375	0.77	7515.12	86.69	0.2295	0.46
	UK	65.71	8.11	0.7847	0.76	72.05	8.49	0.7901	0.77	7530.31	86.78	0.2220	0.49
	COK	48.24	6.94	0.8359	0.79	43.21	6.57	0.8648	0.82	5365.87	73.25	0.4101	0.54
	IDW	112.36	10.60	0.6591	0,64	127.73	11.30	0.6557	0.65	7783.58	88.23	0.1560	0.37
100%	OK	48.24	6.95	0.8338	0,74	51.49	7.18	0.8341	0.78	7332.41	85.63	0.2251	0.46
	UK	59.09	7.69	0.7956	0.77	65.86	8.12	0.8010	0.77	7003.81	83.69	0.2566	0.51
	COK	47.44	6.89	0.8381	0.79	43.89	6.62	0.8646	0.82	5380.68	73.35	0.4084	0.53
	IDW	124.80	11.17	0.6266	0,60	140.05	11.83	0.6242	0.61	7870.39	88.72	0.1477	0.35

Table 4 Annual mean rainfall estimates by ordinary Kriging (OK) and co-ordinary Kriging (COK) for non-sampled points in Rio Grande do Sul state, Brazil, to the period from January 2018 to December 2020 at Jaguarão, Frederico Westphalen and Itaquí

City	Weather Station (Code)	Observed			Predicted	
		2018	2019	2020	OK	COK
Itaquí	ANA ¹ (2,956,005)	114.17	135.33	101.27	131.28	133.29
Jaguarão	INMET ² (A863)	114.92	107.42	80.58	104.83	104.93
Frederico Westphalen	INMET (A854)	131.77	143.32	105.33	151.18	153.43

¹ Agência Nacional das Águas (National Water Agency); ² Instituto Nacional de Meteorologia (National Institute of Meteorology)

or longitude, we did not perform the universal Kriging for this variable. These results may be associated with the inherent difficulties in the process of forecasting extreme values. Ferro [50] points out that, due to the extreme values being at the tail of the distribution, there is a difficulty in verifying the forecast of these events, as this fact causes only a small number of events to be observed and this generates great variation in the verification measures and, therefore, great uncertainty about

the quality of the forecast. More information must be incorporated into the modeling To reduce this uncertainty, as highlighted by Cox et al. [51]. When discussing some theoretical and practical aspects in the analysis of extreme values related to torrential rainfalls and floods, they concluded that, in order to obtain more accurate and reliable estimates, are needed models that take into account as much information as possible, such as trend, dispersion and temporal and spatial dependence.

The Kriging model with the fitted semivariograms used was satisfactory for the mean and median rainfall. We can then test this model for other variables using algorithms that simulate their behavior in the same sample space of this study, thus determining the level of robustness or evidencing the sensitivity of the model to the different sample situations.

Monte Carlo method is used to simulate different scenarios to evaluate the interpolators' behavior before a new data set is introduced, as mentioned in Sect. 2.3. The simulation results (Table 5) point out that the Gaussian random field for the OK presented greater accuracy and precision for all sets of simulated samples. This greater accuracy and precision provide to be robust when submitted to a different data set from which it was modeled.

Random fields were generated under the hypothesis of an Exponential random field with parameters C_0 , C and chosen randomly to analyze the interpolators' behavior when submitted to scenario 2. The results of Table 5 show that with the lowest number of samples, 100 observations, the IDW method presented better performance, being more accurate, since it presented lower Monte Carlo mean of the MSE, and more accurate, with lower standard deviation Monte Carlo of the MSE, than the ordinary Kriging. For the other sample sizes, this result did not maintain. Therefore, Kriging is the most accurate and precise.

As for computational efficiency, in a laptop computer core i7 CPU, 2.2 GHz, 8 GB RAM, the execution time for

1 simulation by ordinary Kriging using the parameters established in Table 5 for $N=750$ was 48.3 s considering 50% of the data set to perform spatial estimates. Considering 100% of the data set, the time was 99.64 s. We can see that time has practically doubled. Regarding our application, the computation time was approximately 0.2 s considering 50% of the data set and 0.26 s considering 100% of data set.

4 Discussion

For Marcuzzo et al. [31] the formation of islands is a feature of the IDW and are formed because as the distance between the interpolated point and the sampling point tends to zero, the weight attributed to the influence of the sampled point on the interpolated tends to infinity thus points esteemed many close to an observed suffer influence practically only of him. However, among classic interpolators, it cannot be said that this is the best among them, as shown by Lyra et al. [15] with the use of splines. Still, there may be situations in which the IDW may perform better, as pointed out by Qiao et al. [52], in the study of maximum and minimum concentrations of heavy metals in the ground. Lu and Wong [53] when studying rainfall data for Taiwan, and because the data are limited to support Kriging (such as problems with the spatial

Table 5 Monte Carlo simulation results for scenario 1, best results for the methods, ordinary Kriging (OK) and inverse distance weighted (IDW), with random field generated under the assumption of spatial dependence by the Gaussian semivariogram, nugget $C_0=40.07$, contribution $C=2065.703$ and range $a=7.623926$, in different sample sizes N . In similar way, scenario 2, from Monte Carlo simulation for the best results of the methods, ordinary Kriging

(OK) and inverse of the weighted distance (IDW), with random field generated in the assumption of spatial dependence by the exponential semivariogram, nugget $C_0=1$ contribution $C=0$ and range $a=0.1$, in different sample sizes N . The results are in terms of mean squared error (MSE) and square root mean squared error (RMSE). In both scenarios, 50% of the data set was considered for calculating the MSE and RMSE

Scenario (1)									
Method	Measures descriptive	MSE				RMSE			
		$N=100$	$N=200$	$N=500$	$N=750$	$N=100$	$N=200$	$N=500$	$N=750$
OK	Mean	63.13	56.2	44.82	44.14	7.81	7.36	6.69	6.64
	Standard deviation	25.19	26.25	3.62	2.64	1.47	1.39	0.26	0.20
IDW	Mean	117.45	135.65	194.9	224.73	10.67	11.48	13.83	14.87
	Standard deviation	45.02	47.67	56.02	59.13	3.88	1.96	1.90	1.91
Scenario (2)									
Method	Measures descriptive	MSE				RMSE			
		$N=100$	$N=200$	$N=500$	$N=750$	$N=100$	$N=200$	$N=500$	$N=750$
OK	Mean	1.032	1.014	1.005	1.006	1.013	1.006	1.002	1.003
	Standard deviation	0.163	0.108	0.059	0.047	0.080	0.053	0.029	0.023
IDW	Mean	1.028	1.043	1.025	1.027	1.011	1.020	1.012	1.013
	Standard deviation	0.156	1.126	0.068	0.057	0.076	0.062	0.033	0.028

correlation structure), IDW was shown to be a logical and viable alternative.

Interpolators require the number of points as a hyperparameter to calculate spatial estimates. As shown in Table 2, the ideal number of points to estimate the average using ordinary Kriging corresponds to 50% of the data set. In large data sets, reducing data to make estimates is an important concern from a computational perspective. Geostatistical methods based on machine learning are considered in some situations [54–57]. However, in these studies, this parameter has not been explored.

Carvalho et al. [58] concluded in their study that the weighted least squares method allows the estimation of model parameters with greater precision. This greater precision reflects in the smallest sum squared residuals, as occurred in the present study. Hatvani et al. [59] used geostatistics resources to predict isoscape areas, consisting of water isotopes preserved in ice bodies in Antarctica. They concluded that the range parameter of the adjusted Gaussian semivariogram estimated that the area of influence of the referred isotopes is 350 km. In our paper, this semivariogram showed the most negligible sums of squares of the residuals.

Medeiros et al. [60] point out that universal Kriging was the one with the lowest prediction error being the most appropriate for the region of the state of Rio Grande do Norte. It should be noted that these authors make a uniquely visual analysis of the Kriging and Kriging variance maps to arrive at this conclusion.

For Lundgren et al. [61], cross-validation provides estimates of errors consistent with genuine errors in irregularly spaced samples, and when $R^2 \geq 0.7$ can be used as the determinant of the best number of points. Several studies show that IDW performed worse than geostatistical methods [62, 63]. Cross-validation is widely used to assess the quality of the geostatistical and classical method and simulation studies are rare. In the present work, in addition to the one done with cross-validation, we present simulation results and show the consistency of our results.

More complex models of Kriging can be considered when, for example, variables cause some trend in the spatially dependent variable, as reported Collardos-Lara et al. [45], who used regression Kriging and external drift Kriging to relate rainfall to elevation. The latter, also known as universal Kriging, was also used in the work of Lado et al. [64], who found slightly better results than ordinary Kriging in terms of RMSE in the analysis of the concentration of heavy metals in the soil. Kizza [65] used two interpolation methods were used for generating the gridded rainfall dataset and the universal Kriging method performed slightly better than the inverse distance weighting method. However, there is no guarantee that modeling external drift can increase the accuracy of rainfall estimates [66].

Furthermore, for Haberlandt [67], the extra value of a variable's information, such as elevation, can be influenced by the time interval and the type of rainfall considered. In multivariate analysis, data obtained from satellites as an additional variable have greater relevance. Our results indicate an advantage for ordinary and co-ordinary Kriging.

Co-Kriging is a viable method to be applied. However, there is no guarantee that the results in terms of RMSE will be better than those achieved by ordinary Kriging [62, 68]. Many applications have been used to improve the accuracy of expensive-to-measure properties (sparsely sampled) using one or more spatially interdependent, cheaper-to-measure properties. This is because the benefits of co Kriging are maximized when the secondary data are related to the primary data and the amount of secondary data is considerably larger than the primary data. In this respect, remotely sensed data provide cheap, intensively sampled spatial information for use with this technique. According to Cunha et al. [46], the improvement in the prediction accuracy is conditioned to the covariate to be used and they verified that the altitude allows better performance of the interpolation by co-Kriging of the rainfall than the distance from the sea. Slightly better results with co-Kriging were also found by Hoshmand & Delghandi [69], but note that ordinary Kriging can also be used to estimate water salinity parameters. Our results indicate that using altitude as a covariate has brought about a substantial improvement in the annual rainfall accuracy for the mean, median and maximum, corroborating with studies by Ma et al. [70], who, by using altitude as a covariate, reduced errors in the generation of rainfall grid in the Tibetan Plateau (TP). However, even with the best global indicators of fit quality point out that the geostatistical methods considered and classic are not suitable in the study of maximum annual rainfall. Possibly, for situations involving maximums, methods for analyzing extreme events are more suitable [71, 72].

Our application is a small size data set because it is longitudinal records, so computing time is relatively short. In the simulation studies, we observed that larger sample sizes demand more computation time, which was already expected, and that using half of the data provides results as good as the whole. In addition, more complex models and large data sets, such as big data, pose a challenge concerning modeling (Gárate-Escamilla, El Hassani, and Andres [73]; Das et al. [17]).

We can make a point regarding missing data. We decided to work with original data and not to embed artificial data since we have a long series of data for each weather station. Given the objective of the work and part of the results, it is evident the importance of considering statistics when analyzing climatological data. It is possible to verify that the models used managed, in a certain

way, to capture the behavior of the series, since the margin of error was low and variable depending on the year (Table 4). This study may have ramifications such as those carried out by Barrios, Trincado, and Garreaud [74] in monthly precipitation records and Aieb et al. [75] to daily rainfall considering data imputation methods such as an artificial neural network (ANN), multiple linear regression (MLR), among others.

All Kriging techniques use ordinary Kriging as opposed to simple Kriging because of the underlying assumption made that the rainfall process is stationary only within local neighborhoods, and hence, the mean's population is unknown [68]. For this reason, we prefer not to address simple Kriging. On the other hand, we explore the number of neighboring points that must be used to make spatial estimates. We found the surprising result that it is not necessary to use the entire data set to obtain accurate spatial estimates, which in none of the papers presented was discussed. It should be noted that, as presented, there are many other promising methodologies for spatial analysis and that we do not address in this paper. Issues such as simulation of random fields with the presence of external and / or covariable drift are not well established at the moment but constitute possibilities for future work. Thus, the simulation results of this work are restricted to the absence of external or secondary influence, as shown in Table 5.

5 Conclusions

The geostatistical method was the one that best represented the spatial distribution of the annual rainfall in the state of Rio Grande do Sul with the estimates that were closer to those observed. Minimum error for non-sampled point was 2.04 mm.

For the mean, Kriging methods presented the slightest mean square error using only nine points (50% of the dataset) for the calculus, representing a decrease in the computational cost. Our results indicate that using altitude as a covariate has brought about a substantial improvement the mean, median and maximum annual rainfall accuracy. Among the Kriging methods, co-ordinary Kriging resulted in the best goodness of fit indicators, which indicates that the inclusion of the altitude covariate increased the accuracy of the predictions. If this information were not available, ordinary Kriging would still be a viable method to be applied. Our results points out those conventional Kriging methods are not suitable when analyzing maximum rainfall events. For situations that involves maximum events (or extreme events), methods for modeling extreme events could be considered.

The simulation results show that Kriging is the interpolator with the highest degree of reliability in the annual rainfall estimates according to the scenarios evaluated. Sample sizes greater than 500 points and using 50% of the data set indicate that both MSE and RMSE tend to stabilize. The spatial interpolators methods used in this study were not suitable for analysis of maximum annual rainfall data.

Acknowledgements We acknowledge INMET and ANA for providing the dataset of this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Gherardi LA, Sala OE (2019) Effect of interannual precipitation variability on dryland productivity: a global synthesis. *Glob Change Biol* 25(1):269–276. <https://doi.org/10.1111/gcb.14480>
2. Knapp AK, Ciais P, Smith MD (2017) Reconciling inconsistencies in precipitation-productivity relationships: implications for climate change. *New Phytol* 214(1):41–47. <https://doi.org/10.1111/nph.14381>
3. Al-Yaari A, Wigneron J-P, Ciais P, Reichstein M, Ballantyne A, Ogée J, Ducharne A et al (2020) Asymmetric responses of ecosystem productivity to rainfall anomalies vary inversely with mean annual rainfall over the conterminous United States. *Glob Change Biol* 26(12):6959–6973. <https://doi.org/10.1111/gcb.15345>
4. Morales RAV, de Araujo CES (2017) Avaliação e Comparação de Dois Métodos de Qualificação de Dados Diários de Precipitação No Estado de Santa Catarina. *Revista Brasileira de Meteorologia* 32(1):65–75. <https://doi.org/10.1590/0102-778632120150154>
5. Liska, Gilberto Rodrigues, Juliano Bortolini, Thelma Sáfyadi, and Luiz Alberto Beijo (2013) "Estimativas de Velocidade Máxima de Vento Em Piracicaba – SP via Séries Temporais e Teoria de Valores Extremos." *Revista Brasileira de Biometria* 31 (2): 295–309. http://jaguar.fcav.unesp.br/RME/fasciculos/v31/v31_n2/indice_v31_n2.php.
6. Martins ALA, Liska GR, Beijo LA, de Menezes FS, Cirillo MÂ (2020) Generalized pareto distribution applied to the analysis of maximum rainfall events in Uruguaiiana, RS, Brazil. *SN Appl Sci* 2(9):1479. <https://doi.org/10.1007/s42452-020-03199-8>
7. Sansigolo CA (2008) Distribuições de Extremos de Precipitação Diária, Temperatura Máxima e Mínima e Velocidade Do Vento

- Em Piracicaba, SP (1917–2006). *Revista Brasileira de Meteorologia* 23(3):341–346. <https://doi.org/10.1590/S0102-77862008000300009>
8. Britto P, Fabiane RB, Mendonça M (2008) Variabilidade Espacial e Temporal Da Precipitação Pluvial No Rio Grande Do Sul: Influência Do Fenômeno El Niño Oscilação Sul. *Revista Brasileira de Climatologia* 3:37–48. <https://doi.org/10.5380/abclima.v3i0.25408>
 9. Câmara, Gilberto, Antônio Miguel Vieira Monteiro, Suzana Druck, and Marília Sá Carvalho. (2004) “Análise Espacial e Geoprocessamento.” In *Análise Espacial de Dados Geográficos*, 209. Brasília: Embrapa. <http://www.dpi.inpe.br/gilberto/livro/analise/cap1-intro.pdf>
 10. Mello CR, Lima JM, Silva AM, Mello JM, Oliveira MS (2003) Krigagem e Inverso Do Quadrado Da Distância Para Interpolação Dos Parâmetros Da Equação de Chuvas Intensas. *Rev Bras Ciênc Solo* 27(5):925–933. <https://doi.org/10.1590/S0100-06832003000500017>
 11. Cerón WL, Andreoli RV, Kayano MT, Canchala T, Carvajal-Escobar Y, Souza RAF (2021) Comparison of spatial interpolation methods for annual and seasonal rainfall in two hotspots of biodiversity in South America. *An Acad Bras Ciênc* 93(1):1–22. <https://doi.org/10.1590/0001-3765202120190674>
 12. Yamamoto JK (2013) *Geoestatística Conceitos e Aplicações*. 1st ed. Oficina de Textos
 13. Chirinos LR, Mallqui CG (2016) Comparación entre interpoladores espaciales en el estudio de distribución de partículas sedimentables insolubles en la cuenca atmosférica de lima y callao. *Información Tecnológica* 27(4):111–120. <https://doi.org/10.4067/S0718-07642016000400012>
 14. Cagnazzo C, Potente E, Rosato S, Mastronuzzi G (2020) Geostatistics and structure from motion techniques for coastal pollution assessment along the Policoro Coast (Southern Italy). *Geosciences* 10(1):28. <https://doi.org/10.3390/geosciences10010028>
 15. Lyra GB, Correia TP, de Oliveira-Júnior JF, Zeri M (2018) Evaluation of methods of spatial interpolation for monthly rainfall data over the State of Rio de Janeiro Brazil. *Theor Appl Clim* 134(3–4):955–65. <https://doi.org/10.1007/s00704-017-2322-3>
 16. Bühlmann P, Hothorn T (2007) Boosting Algorithms: Regularization, Prediction and Model Fitting. *Stat Sci* 22(4):477–505. <https://doi.org/10.1214/07-STS242>
 17. Das H, Barik RK, Dubey H, Roy DS (2019) Cloud computing for geospatial big data analytics studies in big data. Springer, Cham. <https://doi.org/10.1007/978-3-030-03359-0>
 18. Karami S, Madani H, Katibeh H, Marj AF (2018) Assessment and modeling of the groundwater hydrogeochemical quality parameters via geostatistical approaches. *Appl Water Sci* 8(1):23. <https://doi.org/10.1007/s13201-018-0641-x>
 19. Castro José FS, Pezzopane EM, Cecílio RA, PezzopaneXavier JRMAC (2010) Avaliação Do Desempenho Dos Diferentes Métodos de Interpoladores Para Parâmetros Do Balanço Hídrico Climatológico. *Revista Brasileira de Engenharia Agrícola e Ambiental* 14(8):871–880. <https://doi.org/10.1590/S1415-43662010000800012>
 20. Pires CA, Fonseca Da, Reis JT, Dullius ÁIDS, Priesnitz MC, Pereira MJ (2016) Ordinary kriging applied to the rainfall in the States of Rio Grande Do Sul and Santa Catarina. *Ciência e Natura* 38(3):1303. <https://doi.org/10.5902/2179460X22140>
 21. Davison AC, Gholamrezaee MM (2012) Geostatistics of extremes. *Proc Royal Soc A: Math, Phys Eng Sci* 468(2138):581–608. <https://doi.org/10.1098/rspa.2011.0412>
 22. Youngman BD, Stephenson DB (2016) A geostatistical extreme-value framework for fast simulation of natural hazard events. *Proc Royal Soc A: Math, Phys Eng Sci* 472(2189):20150855. <https://doi.org/10.1098/rspa.2015.0855>
 23. Bitencourt DGB, Timm LC, Guimarães EC, Pinto LFS, Pauletto EA, Penning LH (2015) Spatial variability structure of the surface layer attributes of gleysols from the Coastal Plain of Rio Grande Do Sul. *Biosci J* 31(6):1711–21. <https://doi.org/10.14393/BJ-v31n6a2015-29457>
 24. Siqueira TM, Louzada JAS, Pedrollo OC, dos Nilza M, Castro R, de Oliveira MHC (2019) Soil physical and hydraulic properties in the donato stream basin, RS, Brazil. Part 2: geostatistical simulation. *Revista Brasileira de Engenharia Agrícola e Ambiental* 23(9):675–680. <https://doi.org/10.1590/1807-1929/agriambi.v23n9p675-680>
 25. Junior G, Sirtoli B, Magalhães IAL, de Freitas CAA, Cecílio RA (2012) Análise de Técnicas de Interpolação Para Espacialização Da Precipitação Pluvial Na Bacia Do Rio Itapemirim (ES) / Analysis of Interpolation Techniques for Spatial Rainfall Distribution in River Basin Itapemirim (ES). *Revista Ambiência* 8(1):61–71. <https://doi.org/10.5777/ambiencia.2012.01.05>
 26. Nashwan MS, Shahid S (2019) Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmos Res* 230:104632. <https://doi.org/10.1016/j.atmosres.2019.104632>
 27. IBGE (2019) “Instituto Brasileiro de Geografia e Estatística.” <https://www.ibge.gov.br/estatisticas-novoportal/por-cidade-estado-estatisticas.html?#destaques&c=43>
 28. Ross JLS (1985) Relevo Brasileiro: uma Nova Proposta de Classificação. *Geogr Dep, Univ Sao Paulo*. <https://doi.org/10.7154/RDG.1985.0004.0004>
 29. Sartori MDGB (2003) A Dinâmica Do Clima Do Rio Grande Do Sul: Indução Empírica e Conhecimento Científico. *Terra Livre Mudanças Climáticas Globais e Locais* 1:27–49
 30. Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLDM, Sparovek G (2013) Köppen’s climate classification map for Brazil. *Meteorol Z* 22(6):711–728. <https://doi.org/10.1127/0941-2948/2013/0507>
 31. Marcuzzo FFN, Andrade LR, Melo DCR (2011) Métodos de Interpolação Matemática No Mapeamento de Chuvas No Estado Do Mato Grosso. *Revista Brasileira de Geografia Física* 4:793–804
 32. Borges P, de Amorim J, Franke YM, da Anuniação T, Weiss H, Bernhofer C (2016) Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoret Appl Climatol* 123(1–2):335–348. <https://doi.org/10.1007/s00704-014-1359-9>
 33. Cambardella CA, Moorman TB, Parkin TB, Karlen DL, Novak JM, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in central iowa soils. *Soil Sci Soc Am J* 58(5):1501. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>
 34. Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. *R News* 5(2):9–13
 35. Andriotti JLS (2004) *Fundamentos de Estatística e Geoestatística*, 1st edn. Unisinos, São Leopoldo RS
 36. Pereira HR, Meschiatti MC, de Matos Pires RC, Blain GC (2018) On the performance of three indices of agreement: an easy-to-use r-code for calculating the willmott indices. *Bragantia* 77(2):394–403. <https://doi.org/10.1590/1678-4499.2017054>
 37. Rizzo ML (2007) *Statistical Computing with R*. Chapman and Hall/CRC. <https://www.crcpress.com/Statistical-Computing-with-R/Rizzo/p/book/9781584885450>
 38. Rizzo ML (2019) “Monte Carlo Methods in Inference.” In *Statistical Computing with R*, 1st ed., 183–211. Second edition. | Boca Raton : CRC Press, Taylor & Francis Group, 2019.: Chapman and Hall/CRC. <https://doi.org/10.1201/9780429192760-7>
 39. R Core Team (2018) *R: A language and environment for statistical computing*. Austria, Vienna
 40. Diggle PJ, Ribeiro PJ (2007) *Model-based geostatistics* springer series in statistics. Springer, New York. <https://doi.org/10.1007/978-0-387-48536-2>

41. Pebesma EJ (2004) Multivariable geostatistics in s: the gstat package. *Comput Geosci* 30(7):683–691. <https://doi.org/10.1016/j.cageo.2004.03.012>
42. Bivand, Roger, and Tim e Rowlingson. B Keitt. (2018) "Rgdal: Bindings for the 'Geospatial' Data Abstraction Library"
43. Zambrano-Bigiarini M (2020) "HydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological Time Series
44. Hadley W (2009) *Ggplot2: elegant graphics for data analysis*. Springer, New York
45. Collados-Lara A-J, Pardo-Igúzquiza E, Pulido-Velazquez D, Jiménez-Sánchez J (2018) Precipitation fields in an alpine mediterranean catchment: inversion of precipitation gradient with elevation or undercatch of snowfall? *Int J Climatol* 38(9):3565–3578. <https://doi.org/10.1002/joc.5517>
46. Cunha A, de Mello J, Lani L, Rodrigues G, dos Santos E, Filho IF, Trindade FS, de Souza E (2013) Espacialização Da Precipitação Pluvial Por Meio de Krigagem e Cokrigagem. *Pesq Agrop Brasileira* 48(9):1179–1191. <https://doi.org/10.1590/S0100-204X2013000900001>
47. Royston JP (1982) An extension of shapiro and wilk's w test for normality to large samples. *Appl Stat* 31(2):115. <https://doi.org/10.2307/23479735>
48. Baratto J, Wollmann CA (2015) Perfil Topo-Oro-Pluviométrico Do Estado Do Rio Grande Do Sul, Brasil. *Revista Interface*, 10:40–55
49. da Silva MT, De Oliveira Costa CD, Da Silva LA, Júnior EDF, de Faria Pereira VGM, de Oliveira Costa CD, Cherri U (2018) Comparative performance of empirical methods to estimate the reference evapotranspiration in aquidauana MS Brazil. *Científica* 46(2):143. <https://doi.org/10.15361/1984-5529.2018v46n2p143-150>
50. Ferro CAT (2007) A probability model for verifying deterministic forecasts of extreme events. *Weather Forecast* 22(5):1089–1100. <https://doi.org/10.1175/WAF1036.1>
51. Cox DR, Isham VS, Northrop PJ (2002) "Floods: Some Probabilistic and Statistical Approaches." Edited by David Cox, Julian Hunt, Paul Mason, Howard Wheeler, and Peter Wolf. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360 (1796): 1389–1408. Doi: <https://doi.org/10.1098/rsta.2002.1006>
52. Qiao P, Lei M, Yang S, Yang J, Guo G, Zhou X (2018) Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environ Sci Pollut Res* 25(16):15597–15608. <https://doi.org/10.1007/s11356-018-1552-y>
53. Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci* 34(9):1044–1055. <https://doi.org/10.1016/j.cageo.2007.07.010>
54. Appelhans T, Mwangomo E, Hardy DR, Hemp A, Nauss T (2015) Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. *Spat Stat* 14:91–113. <https://doi.org/10.1016/j.spasta.2015.05.008>
55. Chen L, Ren C, Li L, Wang Y, Zhang B, Wang Z, Li L (2019) A comparative assessment of geostatistical, machine learning, and hybrid approaches for mapping topsoil organic carbon content. *ISPRS Int J Geo Inf* 8(4):174. <https://doi.org/10.3390/ijgi8040174>
56. Fouedjio F, Klump J (2019) Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches. *Environ Earth Sci* 78(1):38. <https://doi.org/10.1007/s12665-018-8032-z>
57. Veronesi F, Schillaci C (2019) Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecol Ind* 101:1032–1044. <https://doi.org/10.1016/j.ecolind.2019.02.026>
58. Carvalho JRP, Vieira SR, Grego CR (2009) Comparação de Métodos Para Ajuste de Modelos de Semivariograma Da Precipitação Pluvial Anual. *Revista Brasileira de Engenharia Agrícola e Ambiental* 13(4):443–448. <https://doi.org/10.1590/S1415-43662009000400011>
59. Hatvani IG, Leuenberger M, Kohán B, Kern Z (2017) Geostatistical analysis and isoscape of ice core derived water stable isotope records in an Antarctic Macro Region. *Polar Sci* 13:23–32. <https://doi.org/10.1016/j.polar.2017.04.001>
60. Medeiros FJ, Lucio PS, Silva HJF (2017) Analysis of kriging methods in the estimation of rainfall on Rio Grande Do Norte State. *Revista Brasileira de Geografia Física* 10(5):1668–76. <https://doi.org/10.26848/rbfg.v10.5.p1668-1676>
61. Lundgren WJC, da Silva JAA, Ferreira RLC (2017) A Precisão Da Estimativa Do Erro Da Krigagem Pela Validação Cruzada. *Floresta e Ambiente*. <https://doi.org/10.1590/2179-8087.124114>
62. Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J Hydrol* 228(1–2):113–129. [https://doi.org/10.1016/S0022-1694\(00\)00144-X](https://doi.org/10.1016/S0022-1694(00)00144-X)
63. Manjarrez-Domínguez C, Jesús Prieto-Amparán M, Valles-Aragón M, Delgado-Caballero M-H, Nevarez-Rodríguez M, Vázquez-Quintero G, Berzoza-Gaytan C (2019) Arsenic distribution assessment in a residential area polluted with mining residues. *Int J Environ Res Pub Health* 16(3):375. <https://doi.org/10.3390/ijerph16030375>
64. Lado LR, Hengl T, Reuter HI (2008) Heavy metals in european soils: a geostatistical analysis of the FOREGS geochemical database. *Geoderma* 148(2):189–199. <https://doi.org/10.1016/j.geoderma.2008.09.020>
65. Kizza M, Westerberg I, Rodhe A, Ntale HK (2012) Estimating areal rainfall over lake Victoria and its Basin using ground-based and satellite data. *J Hydrol* 464–465:401–411. <https://doi.org/10.1016/j.jhydrol.2012.07.024>
66. Lloyd CD (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in great Britain. *J Hydrol* 308(1–4):128–150. <https://doi.org/10.1016/j.jhydrol.2004.10.026>
67. Haberlandt U (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J Hydrol* 332(1–2):144–157. <https://doi.org/10.1016/j.jhydrol.2006.06.028>
68. Chappell A, Renzullo LJ, Raupach TH, Haylock M (2013) Evaluating geostatistical methods of blending satellite and gauge data to estimate near real-time daily rainfall for Australia. *J Hydrol* 493:105–114. <https://doi.org/10.1016/j.jhydrol.2013.04.024>
69. Hooshmand A, Delghandi M (2011) Application of kriging and cokriging in spatial estimation of groundwater quality parameters. *Afr J Agric Res* 6(14):3402–3408. <https://doi.org/10.5897/AJAR11.027>
70. Ma J, Li H, Wang J, Hao X, Shao D, Lei H (2020) Reducing the statistical distribution error in gridded precipitation data for the Tibetan Plateau. *J Hydrometeorol* 21(11):2641–2654. <https://doi.org/10.1175/JHM-D-20-0096.1>
71. Davison AC, Padoan SA, Ribatet M (2012) Statistical modeling of spatial extremes. *Stat Sci* 27(2):161–186. <https://doi.org/10.1214/11-STS376>
72. Tawn J, Shooter R, Towe R, Lamb R (2018) Modelling spatial extreme events with environmental applications. *Spat Stat* 28:39–58. <https://doi.org/10.1016/j.spasta.2018.04.007>
73. Gárate-Escamilla, Anna Karen, Amir Hajjam El Hassani, and Emmanuel Andres. 2019. "Big Data Execution Time Based on Spark Machine Learning Libraries." In *Proceedings of the 2019 3rd International Conference on Cloud and Big Data Computing - ICCBDC 2019*, 78–83. New York, New York, USA: ACM Press. <https://doi.org/10.1145/3358505.3358519>
74. Barrios A, Trincado G, Garraud R (2018) Alternative approaches for estimating missing climate data: application to monthly

- precipitation records in South-Central Chile. For Ecosyst 5(1):1–10. <https://doi.org/10.1186/s40663-018-0147-x>
75. Aieb A, Madani K, Scarpa M, Bonaccorso B, Lefsih K (2019) A new approach for processing climate missing databases applied to daily rainfall data in Soummam Watershed, Algeria. Heliyon 5(2):e01247. <https://doi.org/10.1016/j.heliyon.2019.e01247>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.