Revista Brasileira de Meteorologia, **v. 37, n. 1,** 141−155, 2022 rbmet.org.br
DOI: http://dx.doi.org/10.1590/0102-77863710015

Article

Spatial Interpolation Techniques to Map Rainfall in Southeast Brazil

Lucas Eduardo de Oliveira Aparecido[1], Jose Reinaldo da Silva Cabral de Moraes[2], Rafael Fausto de Lima[2], Guilherme Botega Torsoni[2]

[1]*Departamento de Agrometeorologia, Instituto Federal do Sul de Minas, Muzambinho, MG, Brazil.*
[2]*Departamento de Agrometeorologia, Instituto Federal de Mato Grosso do Sul, Naviraí, MS, Brazil.*

Received: 18 February 2022 - Accepted: 24 April 2022

Abstract

The prediction, as well as the estimation of precipitation, is one of the challenges of the scientific community in the world, due to the high spatial and seasonal variability of this meteorological element. For this purpose, methodologies that allow the accurate interpolation of these elements have fundamental importance. Thus, we seek to evaluate the efficiency of the interpolation methods in the mapping of rainfall and compare it with multiple linear regression in tropical regions. The interpolation methods studied were inverse distance weighted (IDW) and Kriging. Monthly meteorological data rainfall from 1961 to 1990 was obtained from 1505 rainfall stations in the Southeast region of Brazil, provided by the National Institute of Meteorology. The comparison between the interpolated data and the real precipitation data of the surface meteorological stations was performed through the following analyzes: accuracy, presicion and tendency. The mean P_{YEAR}, for summer, autumn, winter, and spring are 596 mm seasons^{-1} (s = ±118 mm), 254 mm seasons^{-1} (s = ±52 mm), 114 mm seasons^{-1} (s = ±54 mm) and 393 (s = ± 58 mm) mm seasons^{-1}, respectively. The Kriging highlight accuracy slightly high in relation to IDW. Since the $MAPE_{KRIGING}$ was of 2% while the $MAPE_{IDW}$ was of 3%. The IDW and Kriging methods were accurate and, with low trends in precipitation estimation. While multiple linear regression showed low accuracy when compared with interpolation methods. Despite the lower accuracy the regression linear is more practical and easy to use, as it estimates the rain with only altitude, latitude and longitude, input variables that commonly known input variables. The largest errors in estimating the spatial distribution of precipitation occurred in Winter for all interpolation methods.

Keywords: spatial prediction, big data, geostatistics, climate modeling.

Técnicas de Interpolação Espacial Para Mapear Precipitações em Regiões Tropicais

Resumo

A previsão, assim como a estimativa de precipitação, é um dos desafios da comunidade científica no mundo, devido à alta variabilidade espacial e sazonal deste elemento meteorológico. Para tanto, metodologias que permitam a interpolação precisa desses elementos são de fundamental importância. Assim, buscamos avaliar a eficiência dos métodos de interpolação no mapeamento de chuvas e compará-la com regressão linear múltipla em regiões tropicais. Os métodos de interpolação estudados foram distância inversa ponderada (IDW) e Krigagem. Dados meteorológicos mensais de chuva de 1961 a 1990 foram obtidos de 1.505 estações pluviométricas da região Sudeste do Brasil, fornecidos pelo Instituto Nacional de Meteorologia. A comparação entre os dados interpolados e os dados reais de precipitação das estações meteorológicas de superfície foi realizada através das seguintes análises: acurácia, precisão e tendência. A média P_{YEAR} para verão, outono, inverno e primavera foram 596 mm estações^{-1} (s = ±118 mm), 254 mm estações^{-1} (s = ±52 mm), 114 mm estações^{-1} (s = ±54 mm) e 393 mm (s = ±58 mm) mm estações^{-1}, respectivamente. A precisão da Kriging é um pouco alta em relação ao IDW. Já o $MAPE_{KRIGING}$ foi de 2% enquanto o $MAPE_{IDW}$ foi de 3%. Os métodos IDW e Krigagem foram precisos e com baixas tendências na estimativa de precipitação. Enquanto a regressão linear múltipla apresentou baixa acurácia quando comparada aos métodos de interpolação. Apesar da menor precisão a regressão linear múltipla é mais prática e fácil de usar, pois estima a chuva apenas com altitude, latitude e longitude, variáveis de entrada que todos conhecem. Os maiores erros na estimativa da distribuição espacial da precipitação ocorreram no inverno para todos os métodos de interpolação.

Palavras-chave: previsão espacial, big data, geoestatística, modelagem climática.

Correspoding author: Lucas Eduardo Oliveira Aparecido, lucas.aparecido@muz.ifsuldeminas.edu.br.

1. Introduction

Rainfall is is one of the most important processes of the hydrological cycle (Alvares *et al.*, 2013), considering that its distribution and spatial variability is the most effective component in the regionalization of climatic conditions and also in vegetation growth (Javari, 2017). Rainfall is the most difficult meteorological element to model (Moraes *et al.*, 2020; Chahine, 1992) and for this reason, it requires more efficient prevision methodologies which allow the inference of a value that represents the rainfall of the area of interest (di Piazza *et al*., 2011; Javari, 2016).

Interpolation is a spatialization technique used to estimate a certain numerical variable (Apaydin *et al.*, 2004) for a particular unstamped geographical position, from nearby sampled areas (Lanza *et al.*, 2001; Tveito *et al.*, 2008; di Piazza *et al.*, 2011; Borges *et al.*, 2016). In interpolation, the estimator methods can be divided into two categories: deterministic and stochastic. The first one is based only on geometric criteria and it does not provide measures of uncertainty, such as the Inverse Distance Weighting Method (IDW). In stochastic methods, the collected values are interpreted as results of random processes and stochastic methods are capable of quantifying the uncertainty to the estimator, as the geostatistical models, such as the Kriging Method (Yamamoto and Landim, 2015).

The choice of the method depends on the objective of the study, on the territorial context of the area in question and the available data set and its correlation (Renard and Comby, 2006; Tveito *et al.*, 2008; Wackernagel, 2013; Borges *et al.*, 2016). Several researches compared different methods (IDW, Kriging, and Cokriging) to monthly precipitation in various parts of the world (di Piazza *et al.*, 2011; Keblouti *et al.*, 2012; Javari, 2016). There are few evidences of which method is more suitable on account of a variety of conditions (Borges *et al.*, 2016).

Some authors point out the Kriging Method as the most accurate (Carvalho and Assad, 2005; Viola *et al.*, 2010), while others show that the IDW Method presents a better performance (Keblouti *et al.*, 2012; Gong *et al.*, 2014). Mello and Oliveira (2016) emphasized that kriging was the method that showed the best results in all validation parameters, generating an annual average rainfall of 2,130.1 mm for Joinville, with no trend and minimal variance (Baú *et al.*, 2006; Carvalho *et al.*, 2012). In the interpolation by IDW, the weight of each point is the inverse of a distance function (Shepard, 1968). The main factor that affects the precision of IDW is the energy parameter value. As the increase of distance, there is a reduction in weight, especially when the energy parameter increases (Borges *et al.*, 2016). Closer stations have greater weight and therefore have a greater impact on the estimate (Isaacs and Srivastava, 1989; Nalder and Wein, 1998).

These methods have as limitations the use only of the observations of the localities and not the covariables (Barbulescu, 2016), since the precipitation is correlated with environmental information, such as longitude, latitude and altitude (Cantet, 2017). However, the use of several variables can make the model complex and hinder the use of the model by most users, so multiple linear regression is generally performed to relate precipitation to physical predictor variables.

The novelty of this research work is highlighted by the following points. Southeast Brazil is one of the main regions of agricultural importance in the country, thus, meteorological elements with precipitation is one of the main limitations in agricultural production. However, there are limited studies evaluating the spatial spread of rain in the region and the most appropriate interpolator for heating the matrix images of precipitation at non-sampling points. As each season has its own climatological characteristics, we need to find out the spatial interpolation method more suitable for making maps. In addition, works in the literature are limited to small areas or with reduced meteorological points, limiting themselves to using the standard power for the IDW method and modeling the variogram only for annual rain. In this work, we evaluated the interpolation of rain with a base of 1,505 rain points at different times of the year. In this way, we reinforce the statement of (Dirks *et al.*, 1998), who found that the results of an interpolation are dependent on the sampling density of meteorological stations and, in some cases, the precision of complex methods such as kriging is not greater than the of simple algorithms like IDW and can even be less than that. Finally, the interpolators used were compared by an in-depth assessment of the results of the cross-validation, with different parameters for modifying the models.

In Brazil, there are few meteorological stations spread across the country and 30% of the installed stations need maintenance for an accurate collection of climatic elements.. These issues make it very difficult, especially the collection of precipitation data, due to the high spatial variability of the climatic element.. One way out is to use data from nearby stations and interpolate the rainfall data using an interpolation model that promotes smaller errors. Thus, we seek to evaluate the efficiency of the interpolation methods in the mapping of rainfall and compare it with multiple linear regression in tropical regions.

2. Material and Methods

2.1. Study area

The monthly precipitation, which is measured in unit millimeters (mm), between 1961 and 1990 were obtained from 1505 pluviometric stations to reach out all the Southeast Brazil (Latitude: -14.215/-25.271, Longitude:

-53.121/-39.674). The database came from the Instituto Nacional de Meteorologia (INMET) and and the spatial distribution of weather stations cover the entire southeast region (Fig. 1). We do not apply any homogenization technique to the station data.

We evaluated the influence of altitude on precipitation. When an air mass approaches a mountain (or group of mountains) it is forced to rise reaching lower temperatures, which causes precipitation. The altitude information for the region was obtained through the TOPODATA project (Valeriano and Rossetti, 2008), which culminates in an extensive march of processing of the original data from the Shuttle Radar Topography Mission (SRTM), available for South America and refined by interpolation models for the entire Brazilian territory, with spatial resolution of 30 m. TOPODATA images are arranged in squares compatible with the articulation in the scale of 1: 250,000 of the Brazilian Cartographic System, being in sheets 1° latitude by 1.5° longitude. After obtaining the squares, the mosaic of the entire area obtained was performed to cut out the shape file of the study area.

2.2. Spatial prediction methods

Spatial interpolation to assess rainfall variability in the southeastern region of Brazil was compared using the deterministic and geostatistical method. The deterministic approach was carried out using the Inverse Distance Weighting (IDW) (Eq. (1)).

$$Z(x) = \frac{\sum_{i=1}^{nx} \omega\mathrm{i}\ Z(xi)}{\sum_{i=1}^{nx} \omega\mathrm{i}} \tag{1}$$

where $Z(x)$ is the value of the point for which the interpolation is desired; nx is the quantity of the closest points used in the interpolation of the point x; Z (xi) is the value of the point xi, and ωi is the weight of xi on the point x.

We determine ωi the following equation is used Eq. (2).

$$\omega\mathrm{i} = \frac{1}{h(x, xi)^{p}} \tag{2}$$

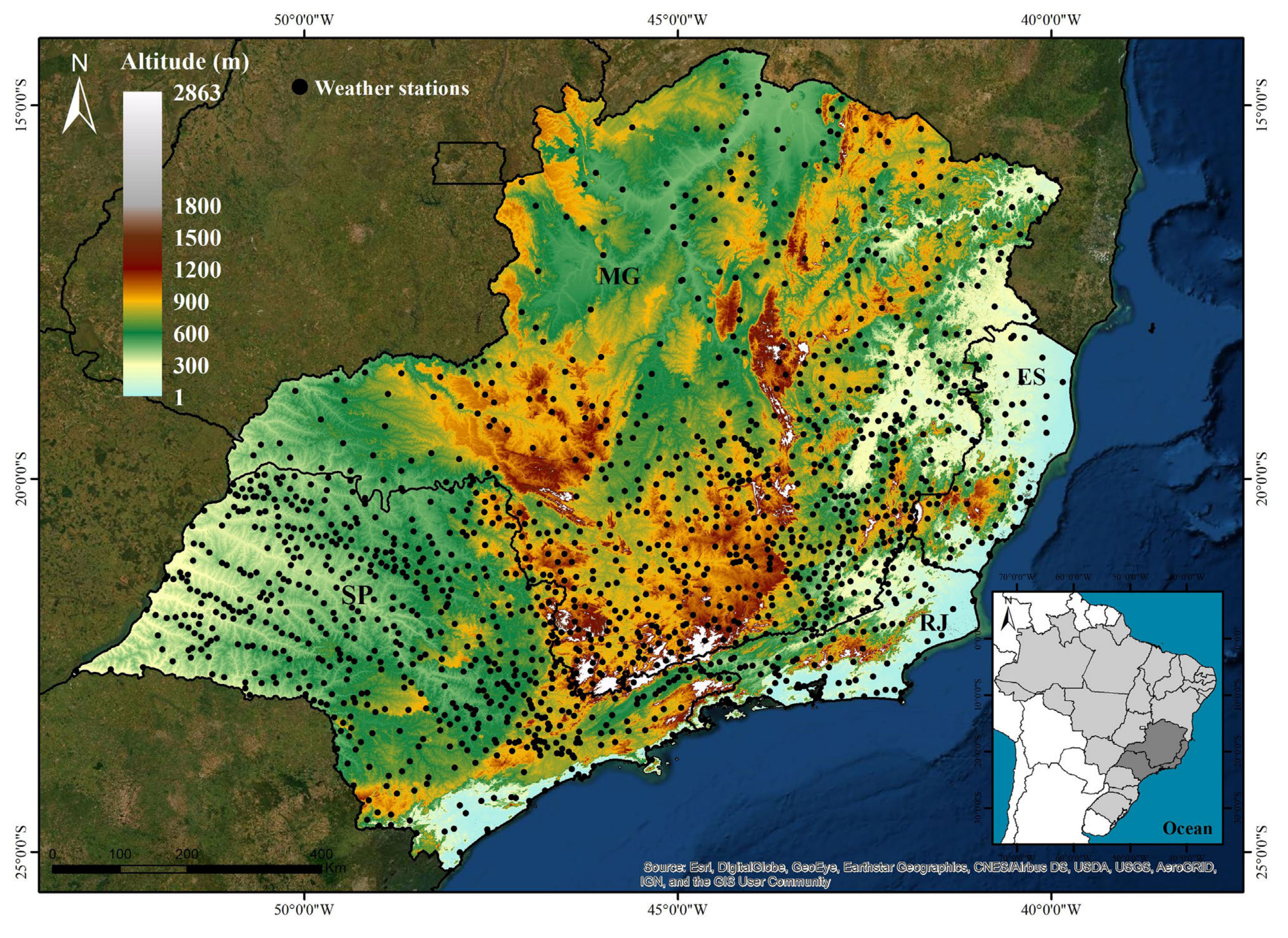

Figure 1 - Spatial distribution of meteorological stations in southeastern Brazil with their respective altitudes. Black dots = weather stations.

where $h(x, xi)$ is the distance between the point x and the point xi; and p is the power parameter, generally equal to two.

This method assumes that the variable being mapped decreases its influence with the distance from its sampled location (Ding *et al.*, 2020; Watson and Philip, 1985). The IDW depends mainly on the inverse of the distance raised to a mathematical power. When defining the power of the point, the influence of the surrounding points is defined, so that as the power increases, the interpolated values begin to approach the value of the nearest sample point (Shi *et al.*, 2020; Li *et al.*, 2012).

The determination of the best power adjustment with the sampled points was made by evaluating the p value equal to 2, 4, 6, 8 and 10. The best adjustment was determined by the RMSE accuracy of the real and measured points.

The IDW assumes that the surface has a local variation, and works best if the sampling points are evenly distributed across the area, without being concentrated in a specific location (Maleika, 2020), so the technique does not evaluate the prediction of errors, as with methods geostatistical, producing small areas that differ from the general smoothing of the variable (Lu and Wong, 2008).

Kriging (Eq. (3)) is a geostatistical technique, generalized least squares regression (Krige, 1951), which takes into account the spatial dependence between observations.

$$\hat{Z}(x) - m(x) = \sum_{i=1}^{n(x)} \lambda i(x)[Z(x_1) - m(xi)] \tag{3}$$

where $\lambda i(x)$ is observation weights $Z(x_1)$; $Z(x_1)$ is interpreted as the realization of VAZ (x); VAZ (x) is Semivariogram modeling $m(x)$, is the expected value of Z (x) at the point x; $n(x)$, is the number of data inside a neighborhood x.

This method assumes that the distance or direction between the sample points reflects a spatial correlation that can be used to explain the variation in the surface, according to the variogram modeling, in the special forecast (Rata *et al.*, 2020; Oliver, 1990). Thus, geostatistical techniques not only have the capacity to produce a forecast surface, but also provide some measure of the certainty or accuracy of the predictions (Ryu *et al.*, 2020, Sen and Sahin, 2001). Modeling the variogram is a fundamental step between the description and the spatial forecast of kriging (Rata *et al.*, 2020). Thus, a theoretical model must be adjusted to this variogram. We adjust different models, selecting spherical, exponential and Gaussian. The best models were determined by the cross-validation obtained by the accuracy of the RMSE.

2.3. Regression linear models

To compare the interpolation methods, a multiple linear regression (RLM) was adjusted to estimate the spatial variability of the rainfall (Eq. (4)). The independent variables used in the construction of the models RLM were altitude (*ALT*, meters), latitude (*LAT*, kilometers) and longitude (*LON*, kilometers) (Cantet, 2017). The dependent data were the rainfall of each season of the year. The applied method was the Ordinary Least Squares (OLS) which seeks to minimize the sum of the squares of the errors of the model (Draper and Smith, 1980), through the optimization system called "Generalized Reduced Gradient" (GRG_2) (Lasdon and Waren, 1982).

$$\begin{aligned} RAINFALL = CL + a \times ALT_{.} + \\ b \times LAT_{.} + c \times LON_{.} + \varepsilon \end{aligned} \tag{4}$$

where, *RAINFALL* is the rainfall of each season of the year (mm $seasons^{-1}$); a, b, and c, are the parameters of the model (weight), *ALT* is altitude (m), *LAT* is latitude (°) and *LON* is longitude (°), *CL* is the linear coefficient (constant term) and ε the random error.

2.4. Criteria for comparison

The differences between the observed and measured values were used to assess the performance of the interpolators through cross-validation. This parameter allows the samples (± 30%) to be excluded temporarily, estimating the value at z from the remaining points. Thus, the real and measured values are obtained in the interpolation. Different numerical indices were used to measure this approximation, including: The Pearson correlation coefficient (r) was used to assess the linearity of the correlated between the interpolated data and the real precipitation data from the surface meteorological stations (Eq. (5)).

$$r = \sum_{i=1}^{n} \left(Yobs_i - \overline{Yobs}\right) \times \frac{\left(Yest_i - \overline{Yest}\right)}{\sqrt{\sum_{i=1}^{n} \left(Yobs_i - \overline{Yobs}\right)^2} \times \sqrt{\sum_{i=1}^{n} \left(Yest_i - \overline{Yest}\right)^2}} \tag{5}$$

where $Yest_i$: interpolated variable; $Yobs_i$: observed variable; n: number of data; $\overline{Yobs}$: mean of the observed variable; $Yest$: mean of the interpolated variable.

The % explained variance derived from the adjusted coefficient of determination ($adjR^2$) allows a realistic comparison of different models as an increased number of parameters are penalized (Eq. (6)). adjR^2 compares the sum of squared prediction errors to the sum of squared deviations of Y about its mean.

$$adjR^2 = \left[1 - \frac{(1 - R^2) \times (n - 1)}{N - k - 1}\right] \quad (6)$$

where R^2: coefficient of determination; n : number of data, and k: number of independent variables in the regression.

The Random Error (Ea) is random variations in measurements from factors that can not be controlled or which, for some reason, have not been controlled (Eq. (7)).

$$Ea = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{est_i} - \overline{Y}\right)^2}{N}} \quad (7)$$

where Y_{est_i}: interpolated variable; $\overline{Y}$: mean of the variable; N : number of data.

The accuracy of interpolated precipitation data performance was analyzed using the following quantitative metrics: The Mean Squared Errors (MSE) metric is defined as the average squared error between interpolated data and the real precipitation data from the surface meteorological stations (Eq. (8)); Root Mean Squared Error (RMSE) is the difference between values predicted by the model and values actually observed from the environment being modeled (Eq. (9)); The Mean Absolute Error (MAE) expresses the accuracy in the same unit as the original data, helping us to conceptualize the amount of error (Eq. (10)); The Mean Absolute Percentage Error (MAPE) is the accuracy as a percentage of the error (Eq. (11)).

$$MSE = \frac{\sum_{i=1}^{n} (Y_{obs_i} - Y_{est_i})^2}{N} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_{obs_i} - Y_{est_i})^2}{N}} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^{n} |Y_{obs_i} - Y_{est_i}|}{N} \quad (10)$$

$$MAPE(\%) = \frac{\sum_{i=1}^{n} \left(\left|\frac{Yest_i - Yobs_i}{Yobs_i}\right| \times 100\right)}{n} \quad (11)$$

where $Yest_i$: interpolated variable; $Yobs_i$: observed variable; n: number of data.

We used the Willmott's Concordance index (d) ranges from 0 to 1, with precision being greater the closer to 1 and less precise when closer to 0. The index d is defined by Eq. (12).

$$d = 1 - \frac{\sum_{i=1}^{n} (Yobs_i - Yest_i)^2}{\sum_{i=1}^{n} \left(|Yest_i - \overline{Y}| + |Yobs_i - \overline{Y}|\right)} \quad (12)$$

where $Yest_i$: interpolated variable; $Yobs_i$: observed variable; n: number of data; $\overline{Y}$: mean of the variable.

The tendency, the degree of deviation, between the estimated average value and the actual values of interpolated precipitation data was analyzed using the following quantitative metrics: The Systematic Error (Es) indicates the tendency of interpolated precipitation values to express results systematically above or below the actual value and what the expected amplitude of this variation (Eq. (13)) and Maximum Absolute Error (EAmax) is the largest forecasted error, expressed in the same units as the dependent series (Eq. (14)).

$$Es = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{obs_i} - \overline{Y}\right)^2}{N}} \quad (13)$$

$$EAmax = max(|Y_{obs_i} - Y_{est_i}|)_{i=1}^{n} \quad (14)$$

where Y_{est_i}: interpolated variable; Y_{obs_i}: observed variable; n: number of data; $\overline{Y}$: mean of the variable.

Reliability was determined by the Confidence Index (C) proposed by Camargo and Sentelhas (1997), it is represented by Eq. (15).

$$C = r.d \quad (15)$$

where r is Pearson correlation coefficient; d is accuracy (Willmott's Concordance index).

The criterion adopted to interpret the performance by the Confidence Index by Camargo and Sentelhas (1997) is represented in Table 1.

The precipitation data were stratified and standardized by seasons of the year for a more detailed analysis (Table 2).

We performed the descriptive statistical analysis whose objective was to identify the variations of the collected data set, in which they were represented by boxplot.

Table 1 - Confidence Index C established by Camargo and Sentelhas (1997).

Value of "C"	Performance
> 0.85	Excellent
0.76 to 0.85	Very good
0.66 to 075	Good
0.61 to 0.65	Median
0.51 to 0.60	Bad
0.41 to 0.50	Very bad
< 0.40	Terrible

2.5. Software

We used Arcgis through the Geostatistical Analyst extension to calculate the values of the experimental variograms and the theoretical models that were adjusted for kriging, as well as the power value for the IDW method. The input of the fields was the precipitation. Through the exploratory analysis provided by the program, it was also verified the normality of the data and the effect of global and anisotropic trend. The maps for the different seasons of the year between the evaluated interpolators were also produced using ArcGIS, by obtaining the adjusted matrix images while the graphics were produced using Python's Matplotlib library.

Table 2 - Precipitation convention for the seasonal period

Season	Period
P>summer	DEC/21 + JAN + FEB + MAR/20
P>autumn	MAR/21 + APR + MAY + JUN/20
P>winter	JUN/21 + JUL + AUG + SEP/23
P>spring	SEP/24 + OCT + NOV + DEC/20

Legend: P_{SUMMER}: Precipitation in summer, P_{AUTUMN}: Precipitation in autumn, P_{WINTER}: Precipitation in winter; P_{SPRING}: Precipitation in spring. DEC/21 is data collection from December 21 onwards

3. Results and Discussion

The Southeast Brazil region showed great spatial variability for annual precipitation (P_{YEAR}) (Fig. 2). The

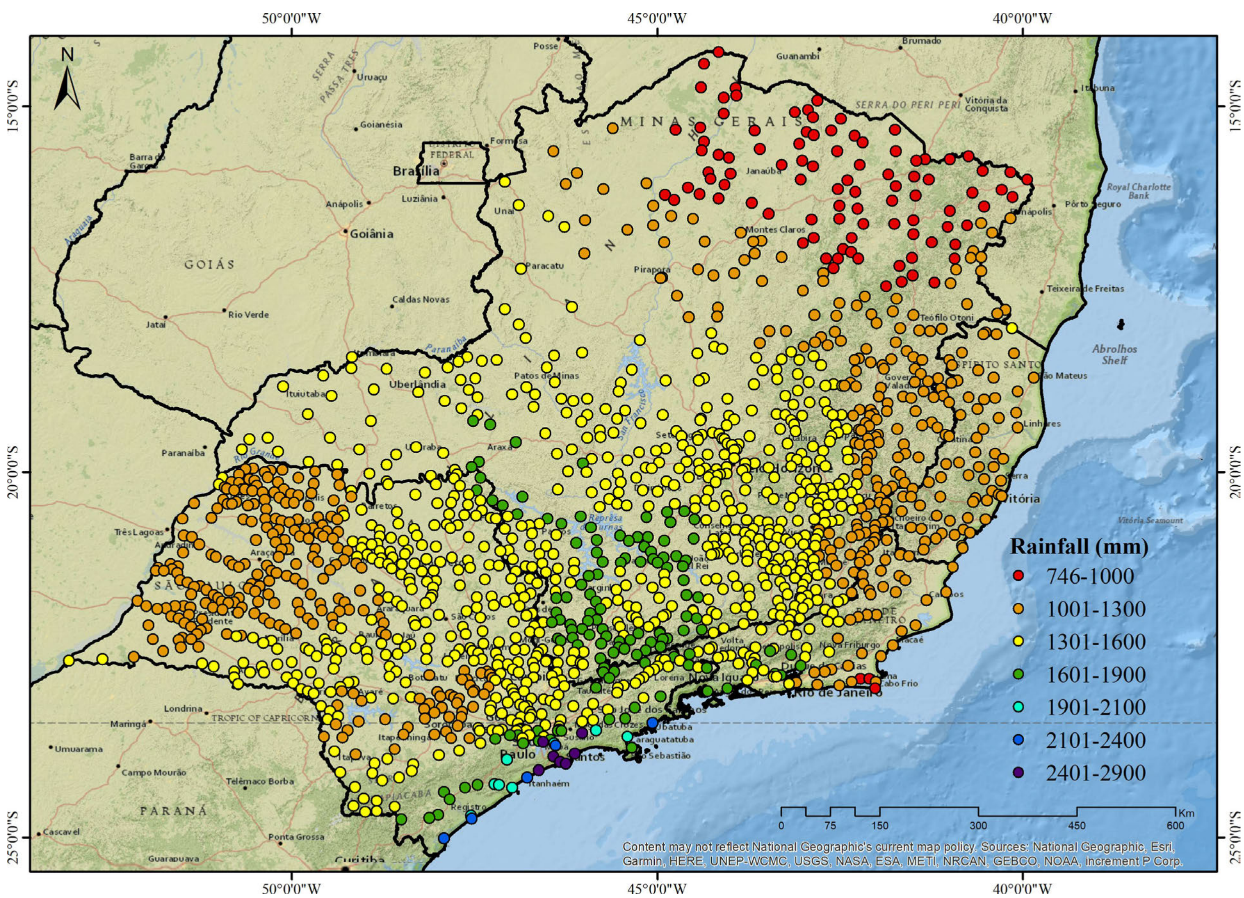

Figure 2 - Spatial variability of annual precipitation of Southeast Brazil. Legend: Red dots = weather stations with rainfall ranging from 746-1,000 mm y^{-1}; Orange dots = weather stations with rainfall ranging from 1,001-1,300 mm y^{-1}; Yellow dots = weather stations with rainfall ranging from 1,301-1,600 mm y^{-1}; Green dots = weather stations with rainfall ranging from 1,601-1,900 mm y^{-1}; Light blue dots = weather stations with rainfall ranging from 1,901-2,100 mm y^{-1}; Dark blue dots = weather stations with rainfall ranging from 2,101-2,400 mm y^{-1} and, Black dots = weather stations with rainfall ranging from 2,401-2,900 mm y^{-1}.

mean P_{YEAR} for the Brazilian Southeast is 1,379 mm $year^{-1}$ with a standard deviation (*s*) of ± 220 mm $year^{-1}$. The smallest P_{YEAR} were of 790 mm $year^{-1}$ and occurred in the North/Northeast of Minas Gerais and the highest P_{YEAR} were of 2,869 mm $year^{-1}$ and occurred mainly in São Paulo coast. In most of the Southeast, the P_{YEAR} shows a variation between 1,200 and 1,600 mm $year^{-1}$. This spatial variability of P_{YEAR} was also described by other authors, such as Alvares *et al.* (2013) and Aparecido *et al.* (2018).

The mean P_{YEAR}, for summer, autumn, winter, and spring are 596 mm $seasons^{-1}$ ($s = \pm 118$ mm), 254 mm $seasons^{-1}$ ($s = \pm 52$ mm), 114 mm $seasons^{-1}$ ($s = \pm 54$ mm) and 393 ($s = \pm 58$ mm) mm $seasons^{-1}$, respectively (Fig. 3). The States of São Paulo (SP), Minas Gerais (MG), Rio de Janeiro (RJ), and Espírito Santo (ES), that compose the Southeast region of Brazil have distinct seasonal precipitations (Fig. 3), since the P_{YEAR} of SP, MG, RJ, and ES for the summer were of 613 mm $seasons^{-1}$; 604 mm $seasons^{-1}$; 542 mm $seasons^{-1}$ and 406 mm $seasons^{-1}$ and for the winter were of 143 mm $seasons^{-1}$; 84 mm $seasons^{-1}$; 143 mm $seasons^{-1}$ and 158 mm $seasons^{-1}$, respectively (Fig. 3).

The relationship between rainfall and altitude in the southeastern region was weak (Table 3). The Pearson correlation with the exception of winter was positive between the seasons, with a higher value in the summer (0.50) and R^2 of 0.25 and a lower value in the autumn, with r of 0.09 and R^2 of 0.01. This relationship, being positive, shows that high altitudes correspond to values of high rain and low altitudes correspond to values of low rain. However, the low correlation coefficients, as well as the determination between the seasons, can limit the use of altitude information in mapping rainfall in this region.

Table 3 - Coefficients of determination (R^2) and correlation (r) between precipitation (mm) in different seasons and altitude (m).

	Summer	Autumn	Winter	Spring
r	0.50	0.09	-0.26	0.35
R^2	0.25	0.01	0.07	0.13

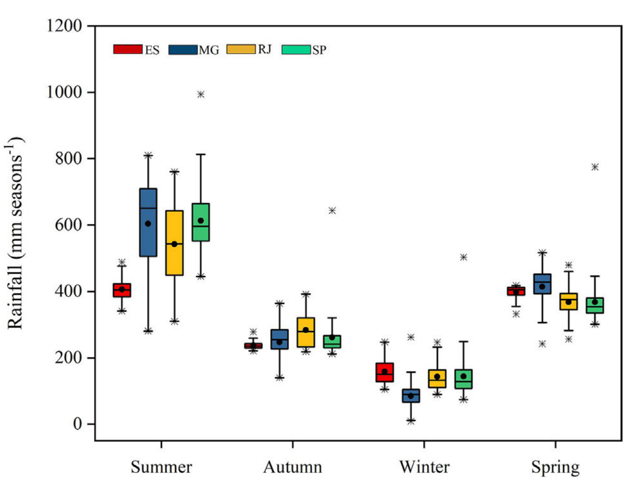

Figure 3 - Boxplot of the rains for each season of the year in southeastern Brazil. Legends: SP is São Paulo; MG is Minas Gerais; RJ is Rio de Janeiro; ES is Espírito Santo states. (Boxplot = ● is average, — is median, ⬚ is 50% of values, I is 90% of values and, * is values extremes).

The power value for interpolation by precipitation IDW according to the RMSE is defined in Fig. 4. The p value that presents the greatest accuracy is p4, with the lowest RMSE in autumn (13.0) and highest spring (19.1). P10 obtained the lowest performance among all parameters. This result can be explained, because as the power value increases, more emphasis can be placed on the nearest points. However, distant points lose weight in the interpolation, which may increase the prediction error. Therefore, due to the accuracy obtained from the RMSE, the ideal power value for rain interpolation in the southeastern region is equal to 4.

The variogram model adjusted for precipitation is observed in Fig. 5. The exponential model showed the highest performance in all seasons, with RMSE of 12.16, 8.56, 12.31 and 14.86 in summer, autumn, winter and spring, respectively. The Gaussian model had the lowest performance, with RMSE values reaching 19.54 in the spring. We also observed that the errors obtained between the actual and measured values in the different models show a trend of greater negative error, that is, with overestimated values as greater precipitation values occur. Based on the experimental variogram, the exponential mode was selected for rain spatialization in southeastern Brazil.

The spatialization of pluvial precipitation for each season of the year in Southeast Brazil, performed by dif-

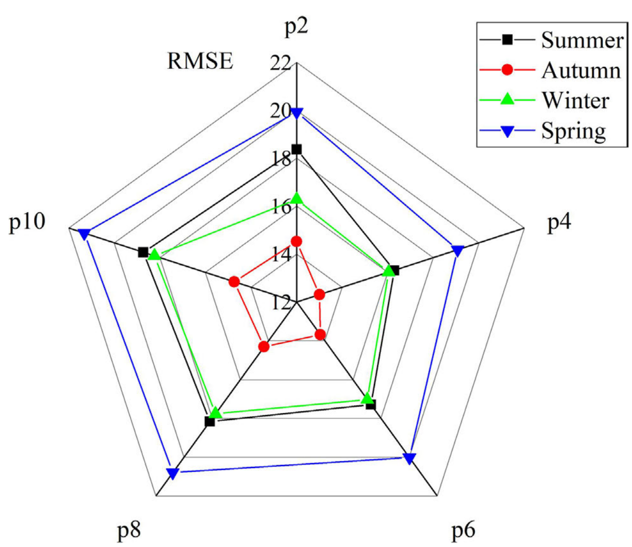

Figure 4 - Determination of the RMSE value (accuracy) of power p using the IDW interpolator for different seasons in the southeastern region of Brazil.

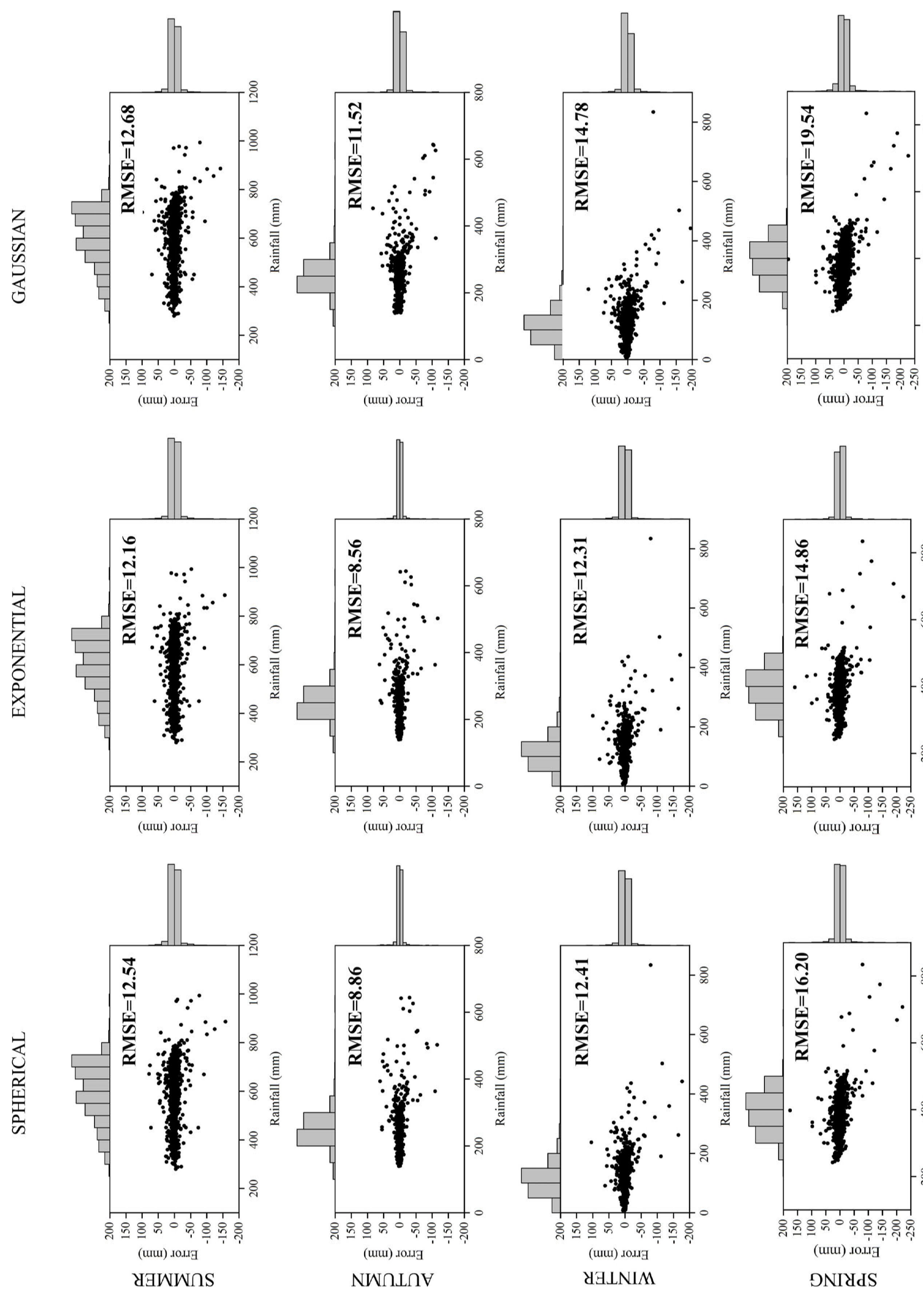

Figura 5 - Accuracy of kriging experimental variograms, between seasons for southeastern Brazil for the variable rainfall in the period from 1961 to 1990. RMSE is Root Mean Squared Error.

ferent methods of interpolation in this study (IDW and Kriging), it was observed that both methods followed the spatial tendency of the real precipitation data. The precipitation in spring (P_{SPRING}) that occurs in the Northeast of MG shows values close to 235-242 mm, and the methods interpolated and estimated values between 243-295 mm (IDW) and 244-296 mm (Kriging), respectively (Fig. 6 D).

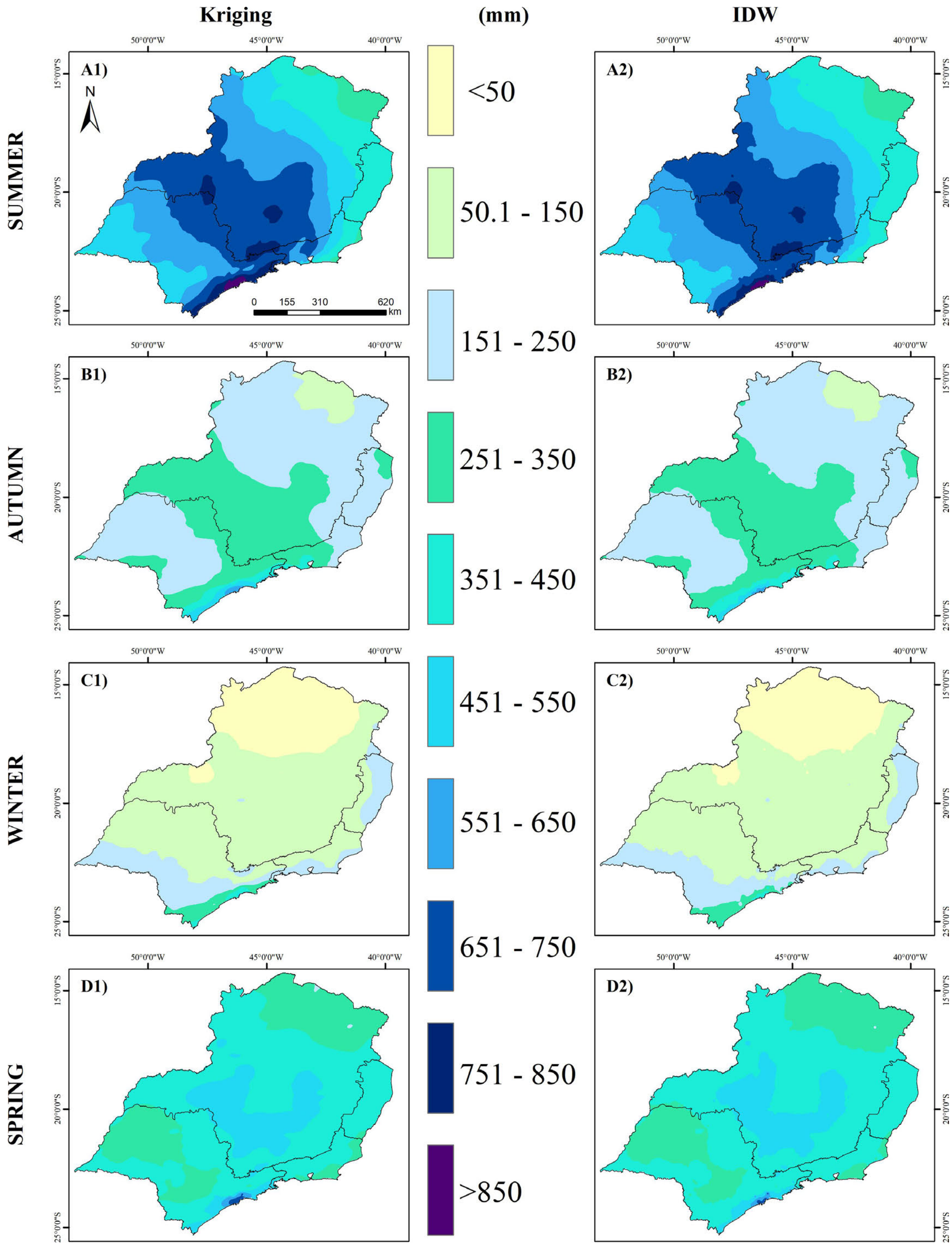

Figure 6 - Rainfall interpolation with IDW and kriging methods for each season of the year in southeastern Brazil in the period from 1961 to 1990. Legend: A_1 is Rain interpolation for summer with IDW method; A_2 is Rain interpolation for summer with Kriging method; B_1 is Rain interpolation for autumn with IDW method; B_2 is Rain interpolation for autumn with Kriging method; C_1 is Rain interpolation for winter with IDW method; C_2 is Rain interpolation for winter with Kriging method; D_1 is Rain interpolation for spring with IDW method and, D_2 is Rain interpolation for spring with Kriging method.

The interpolation methods IDW and Kriging demonstrated high accuracy to estimate the precipitation for all seasons of the year in Southeast region of Brazil (Fig. 7) since the R^2 was above 0.84 and the MAPEs below 6% for all seasons (Table 3). By the Confidence Index C, established by Camargo and Sentelhas (1997), both estimation methods were considered "excellent" for all the seasons of the year, since they show a performance index of 0.85.

The Kriging highlight accuracy slightly high in relation to IDW. Since the $MAPE_{KRIGING}$ was of 2% while the $MAPE_{IDW}$ was of 3%. Considering that the mean P_{YEAR} of Southeast is of 1,379 mm year^{-1}, this difference between the errors ($MAPE_{KRIGING}$ - $MAPE_{IDW}$) of 0.5%, represents a difference in P_{YEAR} of just ± 7 mm. Carvalho and Assad (2005); Viola *et al.*, (2010); Das (2019) also deem the Kriging method more accurate in comparison with the IDW

The methods were more accurate in the interpolation of precipitation in summer (P_{SUMMER}) and less accurate in the interpolation of precipitation in winter (P_{WINTER}). The best accuracy was observed in the interpolation of P_{SUMMER} by the Kriging method, where it was observed the following statistical indices: r = 0.99; R^2 = 0.98; d = 1.00; C = 0.99; Ea = 15; Es = 1; EAmax = 103; MSE = 102; RMSE = 12; MAE = 4, and MAPE = 1%. The low accuracy was the interpolation of P_{WINTER} using the IDW method, since the following statistical indices were revealed: r = 0.86; R^2 = 0.81; d = 0.90; C = 0.91; Ea = 20; Es = 8; EAmax = 295; MSE = 278; RMSE = 17; MAE = 11, and MAPE = 6% (Table 4). Bargaoui and Chebbi (2009) showed a high accuracy in rainfall interpolation for Kriging. Pellicone *et al.* (2018) evidenced the maps obtained with the IDW showed a distribution with punctual areas corresponding to high or low rainfall input data values.

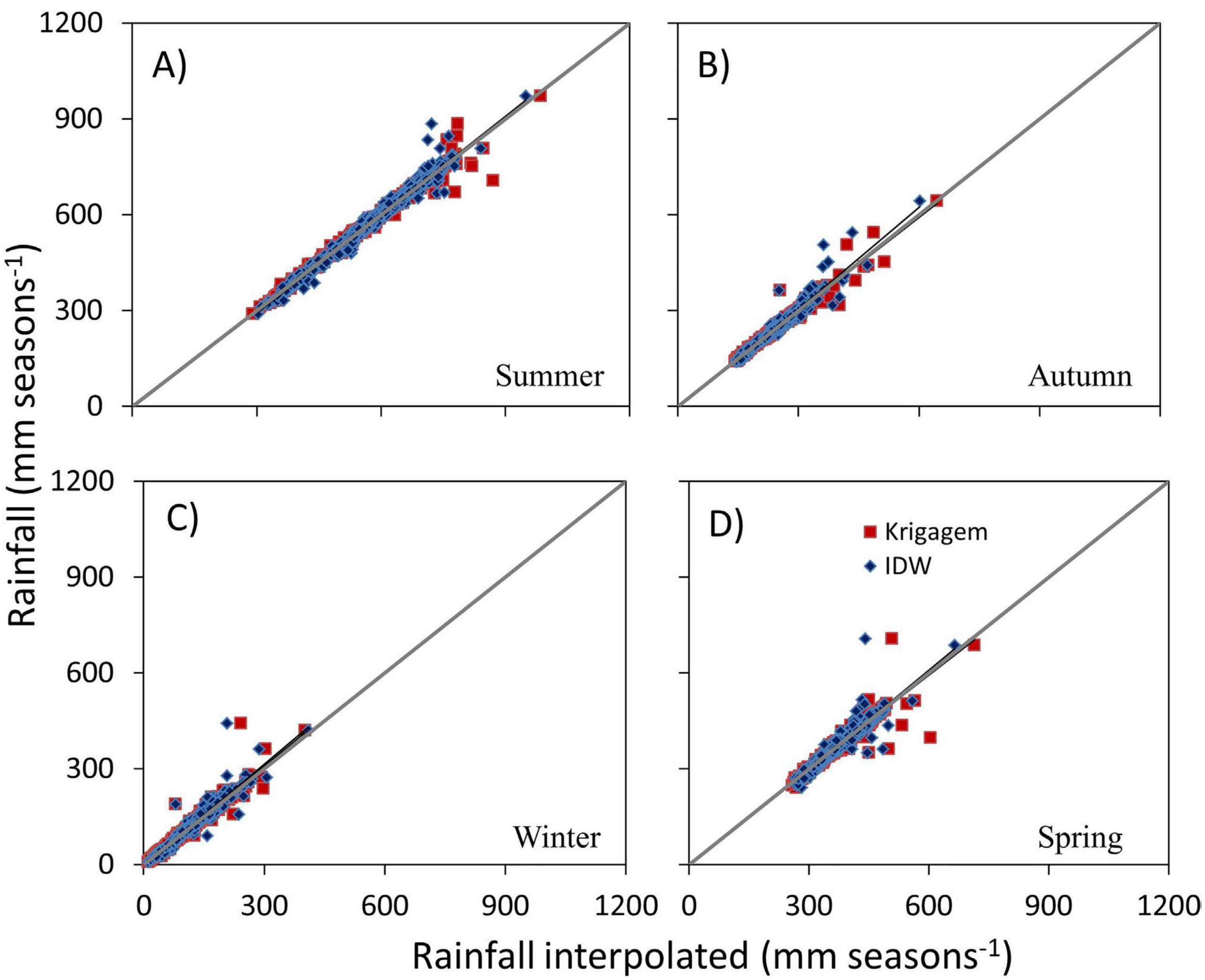

Figure 7 - Performance between real and estimate rainfall by the interpolation methods of IDW and Kriging. A) Summer; B) Autumn; C) Winter; D) Spring.

Table 4 - Statistical indices used to evaluate the accuracy of the interpolation methods of IDW and Kriging for the estimate of rainfall in the period from 1961 to 1990.

Statistical indices	Summer		Autumn		Winter		Spring		Average*	
	IDW	Kriging	IDW	Kriging	IDW	Kriging	IDW	Kriging	IDW	Kriging
r	0.99	0.99	0.96	0.98	0.86	0.97	0.94	0.95	0.94	0.97
R^2	0.98	0.98	0.93	0.96	0.84	0.93	0.89	0.90	0.91	0.94
d	0.99	1.00	0.98	0.99	0.90	0.98	0.97	0.97	0.96	0.99
C	0.99	0.99	0.95	0.97	0.91	0.95	0.91	0.92	0.94	0.96
Ea	16	15	13	10	20	13	18	19	17	14
Es	5	1	7	1	8	6	8	5	7	3
EAmax	103	104	144	110	295	201	267	205	202	155
MSE	218	102	204	1,010	278	208	384	363	271	196
RMSE	17	12	14	11	17	14	20	19	17	14
MAE	9	4	6	4	11	6	9	7	9	5
MAPE	2	1	2	1	6	5	2	2	3	2

*Means of the parameters for all seasons of the year.

According to the distributions of the errors (P_{REAL} - $P_{INTERPOLATED}$) in function of the rainfall variability for each season of the year (Fig. 8). The interpolation of P_{SUMMER} and P_{AUTUMN} obtained the highest errors with high rainfall, above 700 mm for P_{SUMMER} and above 400 mm for P_{AUTUMN}, for both methods (Fig. 6 (A,B)). In P_{WINTER} and P_{SPRING}, the highest deviations occurred for

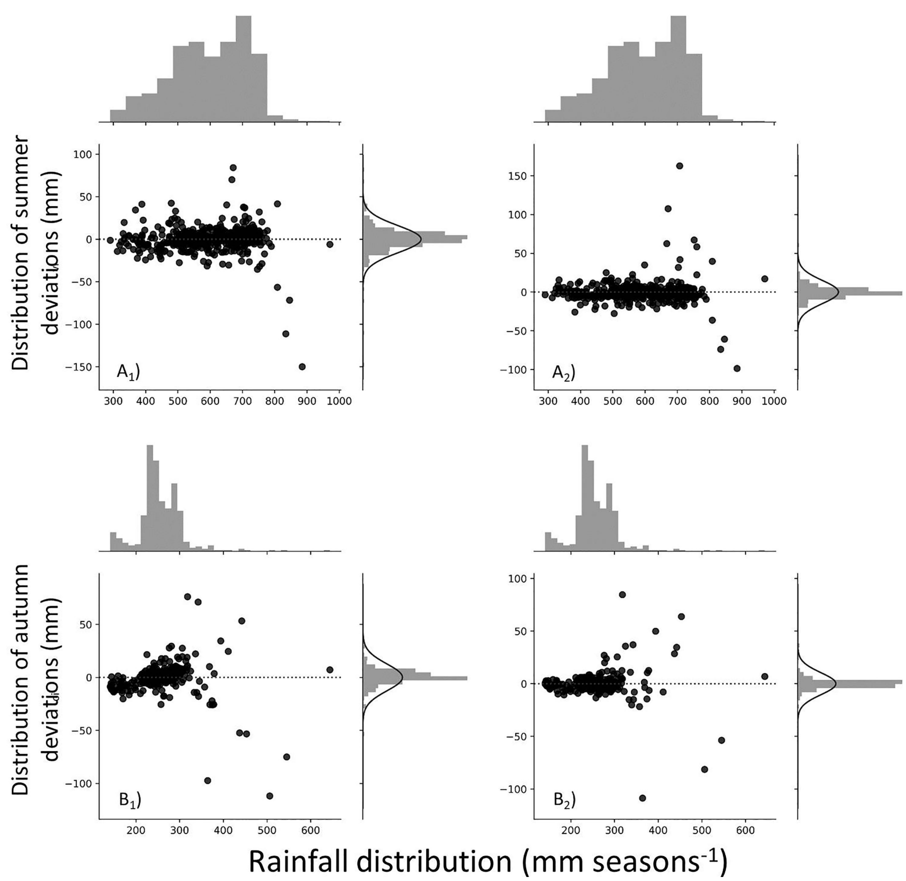

Figure 8 - Distribution of deviations for each interpolation method of the rainfall of the seasons of the year in Southeast Brazil in the period from 1961 to 1990.

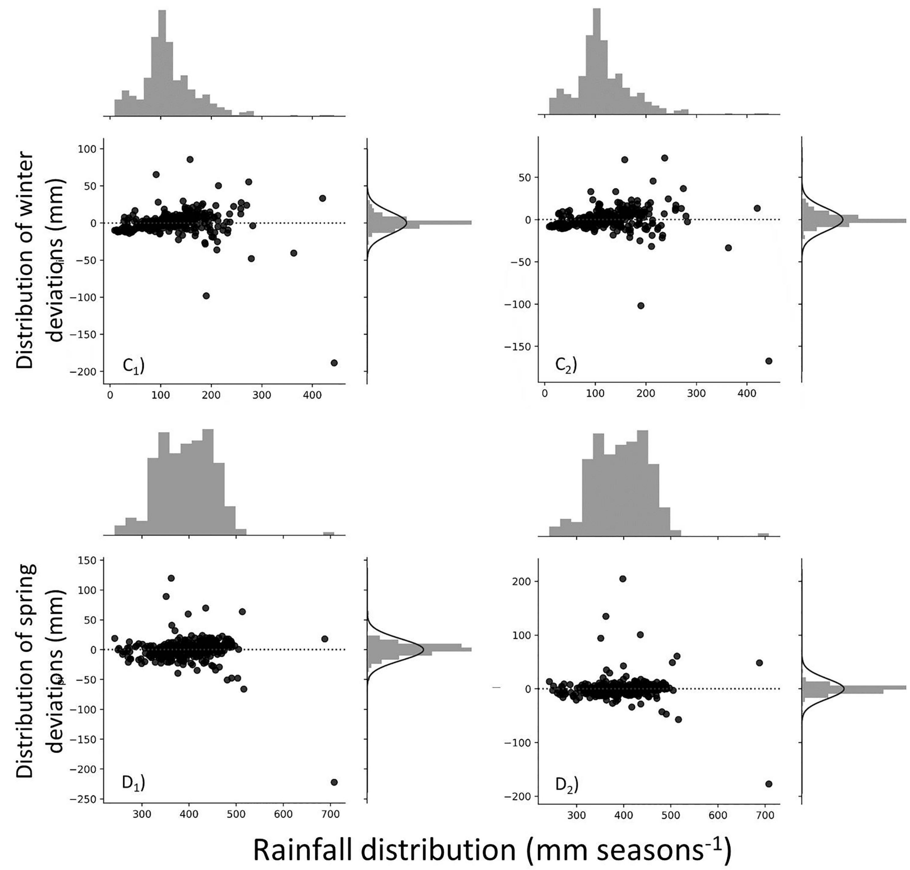

Figure 8 (cont.) - Distribution of deviations for each interpolation method of the rainfall of the seasons of the year in Southeast Brazil in the period from 1961 to 1990.

low rainfall (100 mm seasons^{-1}) and higher rainfall (700 mm seasons^{-1}) (Fig. 8 (C,D)).

The performance of RLM to estimate the rainfall showed a mean accuracy in the rainfall estimates, with MAPEs of 13%, 13%, 30%, and 11% for Summer, Autumn, Winter, and Spring, respectively (Fig. 9). A value of MAPE of 11% as observed in Spring is considered low, taking into account that for average rainfall of 500 mm an error of approximately ± 53 mm can happen. The RLM obtained less efficient results in spatial estimates of precipitation in Southeast Brazil, in comparison with the interpolation methods studied, whereas the mean MAPEs for IDW and Kriging were 3% and 2% (values considered low), while the mean MAPE of RLM was of 17%.

The variable with greater weight in RLM was the latitude, showing inverse relation and coefficients of −14.6; −10.3; −13.4, and −0.3, for Summer, Autumn, Winter, and Spring, respectively (Table 5).

4. Final Considerations

These results are important for the scientific community to know which interpolator to use to spatially estimate the rainfall values for the southeastern region of Brazil.

Comparing the methods of IDW and Kriging, both were accurate, and with low tendencies for precipitation estimate. The accuracy in estimating rainfall level by methods interpolation (IDW and Kriging) in terms of r, R^2, d, C, Ea, Es, EAmax, MSE, RMSE, MAE, and MAPE, varies according to the season of the year. For example, R^2 in the winter were 0.86 and 0.97 mm for IDW and Kriging and summer were 0.99 and 0.99 mm for IDW and Kriging, respectively.

The multiple linear regression (MLR) demonstrated low accuracy in comparison with the interpolation methods IDW and Kriging. For example, the average MAPE for IDW and Kriging were 3 and 2%, respectively and for MLR it was 16.75%. Despite the lower accuracy the regression linear is more practical and easy to use, as it estimates the rain with only altitude, latitude and longitude, input variables that everyone knows.

The biggest errors in the estimate of the spatial distribution of precipitation occurred in Winter for all the interpolation methods (IDW, Kriging, and RLM). For example, MAPE was 6, 5 and 30% for IDW, Kriging, and RLM, respectively. The spatial information about rainfall

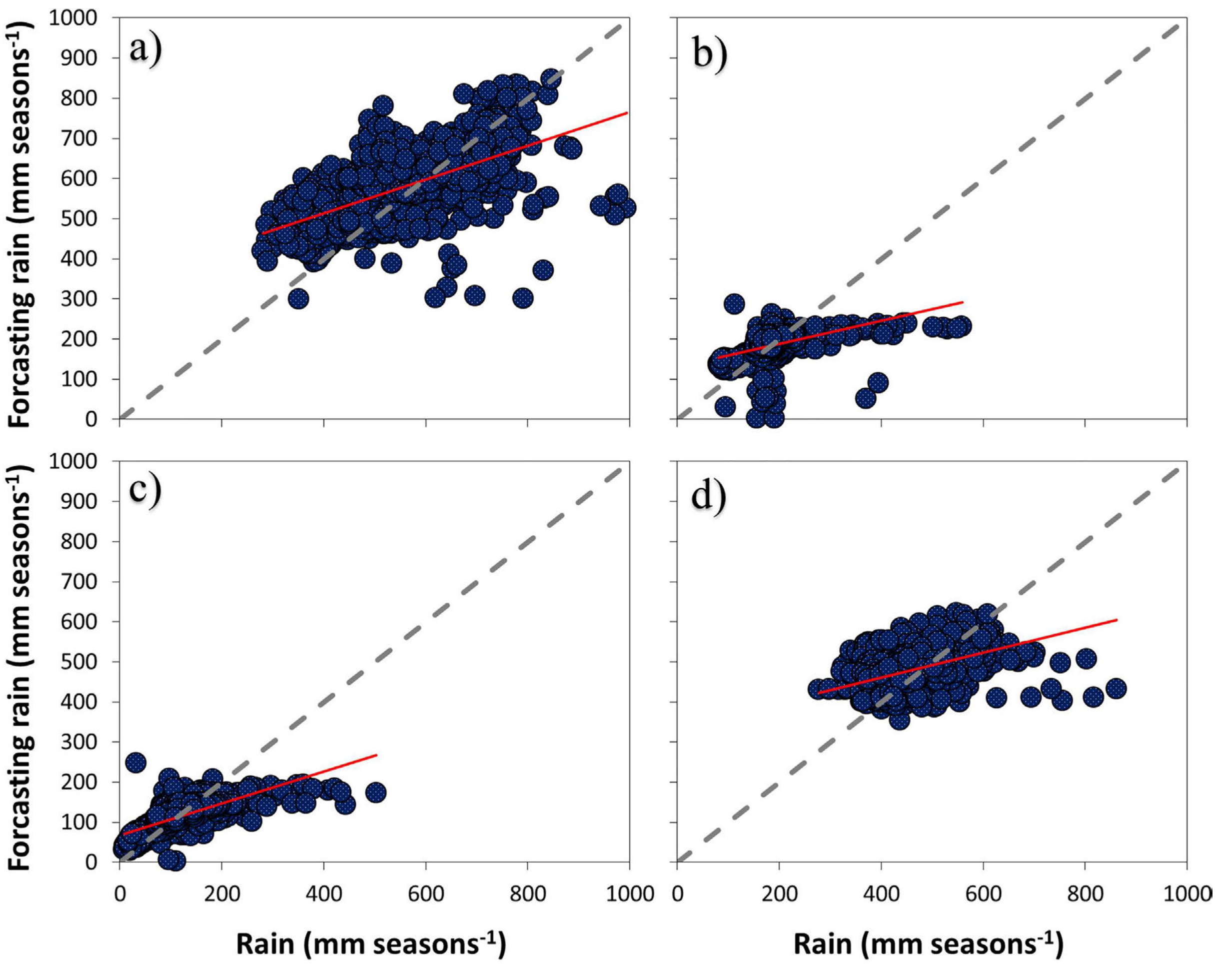

Figure 9 - Performance between real and estimated rainfall in the period from 1961 to 1990 by multiple linear regression for the season's summer (a), autumn (b), winter (c), and spring (d).

Table 5 - Parameters of the multiple linear regression models with their statistical indices for the estimate of rainfall in the period from 1961 to 1990.

	Multiple linear regression (coefficients)			
	Summer	Autumn	Winter	Spring
Intersection	-157.9	34.2	-65.5	741.7
Altitude	0.2	-0.0	-0.1	0.1
Latitude	-14.6	-10.3	-13.4	-0.3
Logitude	-6.7	1.1	1.5	7.6
	Statistical metrics			
	Summer	Autumn	Winter	Spring
p-value	0.05	0.05	0.05	0.05
r	0.65	0.54	0.63	0.56
R^2	0.42	0.29	0.40	0.31
D	0.77	0.65	0.74	0.68
C	0.50	0.35	0.47	0.38
Ea	59	21	26	37
Es	69	32	32	55
EAmax	491	327	364	430

(*continued*)

Table 5 - continued

	Multiple linear regression (coefficients)			
	Summer	Autumn	Winter	Spring
MSE	8198	1495	1690	4310
RMSE	91	39	41	66
MAE	71	23	27	51
MAPE(%)	13	13	30	11

is an important factor in terms of formation of governing character. Southeast of Brazil demonstrated average annual rainfall for summer, autumn, winter, and spring are 596 mm seasons^{-1} ($s = \pm 118$ mm), 254 mm seasons^{-1} ($s = \pm 52$ mm), 114 mm seasons^{-1} ($s = \pm 54$ mm) and 393 ($s = \pm 58$ mm) mm seasons^{-1}, respectively. As future works, we suggest testing the same interpolators throughout Brazil, covering specific regions such as the Pantanal and the Amazon rainforest.

Acknowledgment

This work was done with financial support from Instituto Federal de Mato Grosso do Sul “IFMS”.

References

ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; DE MORAES GONÇALVES, J.L.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 1 2013.

APARECIDO, L.E.O.; ROLIM, G.S.; MORAES, J.R.S.C.; ROCHA, H.G.; LENSE, G.H.E.; *et al.* Agroclimatic zoning for urucum crops in the state of Minas Gerais, Brazil. **Bragantia**, v. 77, n. 1, p. 193-200, 2017.

APAYDIN, H.; SONMEZ, F.; YILDIRIM, Y. Spatial interpolation techniques for climate data in the GAP region in Turkey. **Climate Research**, v. 28, n. 1, p. 31-40, 2004.

BARBULESCU, A. A new method for estimation the regional precipitation. **Water Resources Management**, v. 30, n. 1, p. 33-42, 2016.

BARGAOUI K.; CHEBBI A.Z. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. **Journal of Hydrology**, v. 365, n. 1-2, p. 56-73, 2009.

BAÚ, A.L.; GOMES, B.M.; DE QUEIROZ; M.M.F.; OPAZO, M.A.U.; SAMPAIO, S.C. Comportamento espacial da precipitação pluvial mensal provável da mesoregião oeste do Estado do Paraná. **Irriga**, v. 11, n. 2, p. 150-168, 2006.

BORGES, P.A.; FRANKE, J.; DA ANUNCIAÇÃO, Y.M.T.; WEISS, H.; BERNHOFER, C. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. **Theoretical and Applied Climatology**, v. 123, n. 1-2, p. 335-348, 2016.

CAMARGO, A.; SENTELHAS, P.C. Avaliação do desempenho de diferentes métodos de estimativa da evapotranspiração potencial no Estado de São Paulo, Brasil. **Revista Brasileira de Agrometeorologia**, v. 5, n. 1, p. 89-97, 1997.

CANTET, P. Mapping the mean monthly precipitation of a small island using kriging with external drifts. **Theoretical and Applied Climatology**, v. 127, n. 1-2, p. 31-44, 2017.

CARVALHO, J.R.; ASSAD, E.D. Spatial analysis of precipitation data in São Paulo state: Comparison of interpolation methods. **Engenharia Agrícola**, v. 25, n. 2, p. 377-384, 2005.

CARVALHO, J.R.P.; ASSAD, E.D.; PINTO, H.S. Geostatistical interpolation in the analysis of spatial distribution of annual rainfall and of its relationship to altitude. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1235-1242, 2012.

CHAHINE, M.T. The hydrological cycle and its influence on climate. **Nature**, v. 359, n. 6394, p. 373-380, 1992.

DAS, S. Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique. **International Journal of Climatology**, v. 39, n. 1, p. 407-423, 2019.

DI PIAZZA, A.; CONTI, F.L.; NOTO, L.V.; VIOLA, F.; LA LOGGIA, G. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, n. 3, p. 396-408, 2011.

DING, Z.; MEI, G.; CUOMO, S.; LI, Y.; XU, N. Comparison of estimating missing values in IoT time series data using different interpolation algorithms. **International Journal of Parallel Programming**, v. 48, n. 3, p. 534-548, 2020.

DRAPER, N.R.; SMITH, H. **Applied Regression Analysis**. John Wiley & Sons: Canadá, 326 p., 1998.

GONG, G.; MATTEVADA, S.; O'BRYANT, S.E. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. **Environmental Research**, v. 130, n. 1, p. 59-69, 2014.

ISAACS, E.H., SRIVASTAVA, R.M. **Applied Geostatistics**. Oxford University Press: New York, 561 p., 1989.

JAVARI, M. Geostatistical and spatial statistical modelling of precipitation variations in Iran. **Journal of Civil & Environmental Engineering**, v. 6, n. 3, p.1-30, 2016.

JAVARI, M. Spatial monitoring and variability of daily rainfall in Iran. **International Journal of Applied Environmental Sciences**, v. 12, n. 5, p. 801, 2017.

KEBLOUTI, M.; OUERDACHI, L.; BOUTAGHANE, H. Spatial interpolation of annual precipitation in Annaba-Algeria-comparison and evaluation of methods. **Energy Procedia**, v. 18, n. 1, p. 468-475, 2012.

LANZA, L.G.; RAMIREZ, J.A.; TODINI, E. Stochastic rainfall interpolation and downscaling. **Hydrology and Earth System Sciences**, v. 5, n. 2, p. 139-143, 2001.

LASDON, L.S.; WAREN, A.D. **General Reduced Gradient Software for Linearly and Non-Linearly Contained Problems. GRG2 Users Guide**. Univ. Texas: Austin, 1982.

LI, Z.; WANG,Y.; ZHOU, D.; WU, C. An intelligent method for fault diagnosis in photovoltaic array. In: **Communications in Computer and Information Science**, v. 327, n. 1, p. 10-16, 2012.

LU, G.Y.; WONG, D.W. An adaptive inverse-distance weighting spatial interpolation technique. **Computers & Geosciences**, v. 34, n. 9, p. 1044-1055, 2008.

MALEIK, W. Inverse distance weighting method optimization in the process of digital terrain model creation based on data collected from a multibeam echosounder. **Applied Geomatics**, v. 12, n. 4, p. 397-407, 2020.

MELLO, Y.R.; OLIVEIRA, T.M.N. Statistical and geostatistical analysis of the average rainfall in the municipality of Joinville (SC). **Revista Brasileira de Meteorologia**, v. 31, n. 2, p. 229-239, 2016.

MORAES, J.R.S.C.; ROLIM, G.S.; MARTORANO, L.G.; APARECIDO, L.E.O.; BISPO, R.C.; *et al.* Performance of the ECMWF in air temperature and precipitation estimates in the Brazilian Amazon. **Theoretical and Applied Climatology**, v. 141, n. 3-4, p. 803-816, 2020.

NALDER, I.A.; WEIN, R.W. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. **Agricultural and Forest Meteorology**, v. 92, n. 4, p. 211-225, 1998.

OLIVER, M.A.; WEBSTER, R. Kriging: A method of interpolation for geographical information systems. **International Journal of Geographical Information System**, v. 4, n. 3, p. 313-332, 1990.

PELLICONE, G.; CALOIERO, T.; MODICA, G.; GUAGLIARDI, I. Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy). **International Journal of Climatology**, v. 38, n. 9, p. 3651-3666, 2018.

RATA, M.; DOUAOUI, A.; LARID, M.; DOUAIK, A. Comparison of geostatistical interpolation methods to map annual rainfall in the Chéliff watershed, Algeria. **Theoretical and Applied Climatology**, v. 141, n. 3-4, p. 1009-1024, 2020.

RENARD, F.; COMBY, J. Evaluation de techniques d'interpolation spatiale de la pluie en milieu urbain pour une meilleure gestion d'événements extrêmes: le cas du Grand Lyon. **La Houille Blanche**, v. 92, n. 6, p. 73-78, 2006.

RYU, S.; SONG, J.J.; KIM, Y.; JUNG, S.H.; DO, Y.; *et al.* Spatial interpolation of gauge measured rainfall using compressed sensing. **Asia-Pacific Journal of Atmospheric Sciences**, v. 57, n. 2, p. 331-345, 2021.

SEN, Z.; SAHİN, A.D. Spatial interpolation and estimation of solar irradiation by cumulative semivariograms. **Solar Energy**, v. 71, n. 1, p. 11-21, 2001.

SHEPARD, D.A. two-dimensional interpolation function for irregularly-spaced data. In: **Proceedings of the 1968 23rd ACM National Conference**, p. 517-524, 1968.

SHI, Y.; HE, W.; ZHAO, J.; HU, A.; PAN, J.; *et al.* Expected output calculation based on inverse distance weighting and its application in anomaly detection of distributed photovoltaic power stations. **Journal of Cleaner Production**, v. 253, n. 1, p. 119965, 2020.

TVEITO, O.E.; WEGEHENKEL, M.; VAN DER WEL, F.; DOBESCH, H. **Cost Action 719: The Use of Geographic Information Systems in Climatology and Meteorology: Final Report**. Office for Official Publications of the European Communities: Luxembourg, 2008.

VALERIANO, M.M.; ROSSETTI, D.F. **Topodata: Selection of Geostatistical Coefficients for the Unified Refinement of SRTM Data**. INPE, p. 50, 2008. Available at http://www.dpi.inpe.br/topodata/documentos.php, accessed on 15 March, 2020.

VIOLA, M.R.; MELLO, C.R.; PINTO, D.B.; MELLO, J.M.; ÁVILA, L.F. Spatial interpolation methods for mapping of rainfall. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 9, p. 970-978, 2010.

WACKERNAGEL, H. **Multivariate Geostatistics: An Introduction with Applications**. Springer: Berlin, 2003.

WATSON, D.F.; PHILIP, G.M. Um refinamento da interpolação ponderada por distância inversa. **Geoprocessamento**, v. 2, n. 4, p. 315-327, 1985.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geostatistics: Concepts and Applications**. Oficina de Textos: São Paulo, 2015.