



## GEOSCIENCES

# Performance assessment of spatio-temporal regression kriging with GAMLSS models as trends

ELIAS S. DE MEDEIROS, RENATO R. DE LIMA, RICARDO A. DE OLINDA, LEYDSON G. DANTAS & CARLOS A.C. DOS SANTOS

**Abstract:** The main objective of this study is to propose different probabilistic models for adjusting the trend component, since it significantly influences the quality of the spatio-temporal interpolation of rainfalls. We used the monthly total precipitation data of the São Francisco River Basin (SFRB) for the period of 31 years, 1989–2019. The SFRB occupies 8% of the whole Brazilian territory, mostly located in the Northeast Brazilian region. For the trend component, we propose the fitted GAMLSS models by comparing different probability distribution families, which in most cases include the characteristics of these data. The results indicate the existence of a spatio-temporal pattern of the residues obtained from the adjustment of the trend with zero adjusted Gamma distribution for the accumulated monthly precipitation. The adjustment revealed a spatial dependence of up to 873 km between the pluviometric stations and temporal autocorrelation of approximately 1.6 months. The methodology used in this study enabled us to create rainfall maps, interpolating unobserved locations in differences years. The projection of these maps to the SFRB is considered extremely important for planning and implementing activities related to water resources across the river basin.

**Key words:** Brazil, GAMLSS regression, geostatistics, rainfall.

## INTRODUCTION

In recent years, the increasing frequency and intensity of droughts have been the focus of several studies associated with climate risk, most of them provoked by the unprecedented water crisis. Some experts believe that the water crisis in the 21<sup>st</sup> century is more a matter of management than the real crisis of scarcity and stress (Rogers et al. 2005). However, for other experts, this is the result of a series of problems that are aggravated by other issues related to economic and social development (Gleick & Cain 2004). In one of the most recent studies carried out by the Intergovernmental Panel on Climate Change (IPCC), published on August 9, 2021, confirms that the water cycle has been intensifying and will continue to intensify the

planet warms, causing more intense rains and prolonged droughts all around the world (IPCC 2021). In the last 2,000 years, climate change, mainly caused by man, has intensified at an exponential rate, causing serious problems for humans and the planet, such as storms, droughts and heat waves (IPCC 2021). Therefore, there is a need to investigate how climate change has impacted the climate, especially in the spatial and temporal distribution of precipitation in the region.

Several studies have examined the spatial and temporal variability in rainfall precipitation (Foehn et al. 2018, Achite et al. 2017). Gathering precipitation information with measuring instruments placed at specific locations can be an easy task, but what we need is to obtain

the precipitation estimates for locations not sampled. Several methodologies can be used to obtain these estimates, such as regression analysis (Stauffer et al. 2016) and geostatistical models that consider spatio-temporal interaction (Kilibarda et al. 2015, Martínez et al. 2017).

Once we find that the phenomenon under study reveals spatial and temporal variability, we can obtain better results with models to consider the space-time relationship as well as its spatial, temporal, and spatio-temporal component structure (Hu et al. 2017, Monteiro et al. 2017).

In order to apply the methodology presented in this paper, we use the precipitation data set for the São Francisco River Basin (SFRB) evaluated in space and time. The SFRB covers a drainage area of 639,19  $km^2$ , corresponding to 8% of Brazil and an average flow of 2,850  $m^3/s$ , representing 2% of the entire Brazilian territory. Since 2012, the SFRB faces several adverse hydrological conditions, with below-average flows and precipitations reducing the volume of storage in their reservoirs. The 2014-2016 period recorded the lowest annual average natural outflows in the Sobradinho reservoir since 1931, the year the Operator of the National Electricity System (ONS) began recording observations (ANA 2017). The São Francisco River is of great importance to the semiarid Northeast Brazilian (NEB) region for the supply of water for human consumption and irrigation, generation of energy by the Paulo Afonso and Sobradinho hydroelectric power plant, tourism and fishing (Araujo & Celeste 2019, Silveira et al. 2016).

The frequency and intensity that climatic extremes increase also generates an increase in global warming and will continue to increase in a scenario of medium and high greenhouse gases emissions causing droughts and climate changes in rainy regions in the coming centuries,

changing basins and slopes. One of the latest IPCC reports (IPCC 2014) predicts a drastic reduction of up to 22% in precipitation levels in NEB, impacting the distribution of hydraulic power, mainly due to an estimated decrease of 24.6% in São Francisco River flow.

The main objective of this study is to propose different probabilistic models for adjusting the trend component, since it significantly influences the quality of the spatio-temporal interpolation of precipitation. Precipitation maps in the SFRB is considered of extreme importance for the planning and execution of activities related to water resources, identifying priority regions for executing governmental actions in order to minimize the disasters caused by the scarcity of rain in the region, especially, the desertification process that has caused situations of socio-environmental vulnerability in the region.

## MATERIALS AND METHODS

### Study area and data set

In this study, we used the monthly total precipitation data of the SFRB provided by the Agência Nacional das Águas (ANA) for the period of 31 years, 1989-2019 (Figure 1). We considered only the pluviometric stations that presented 10 years or more of information, resulting in 341 stations. Consequently, several pluviometric stations kept of this study presented incomplete data, however, the methodology proposed in this research, spatio-temporal geostatistics, incorporates this type of question, taking into account the neighboring stations for precipitation interpolation in non-sampled locations.

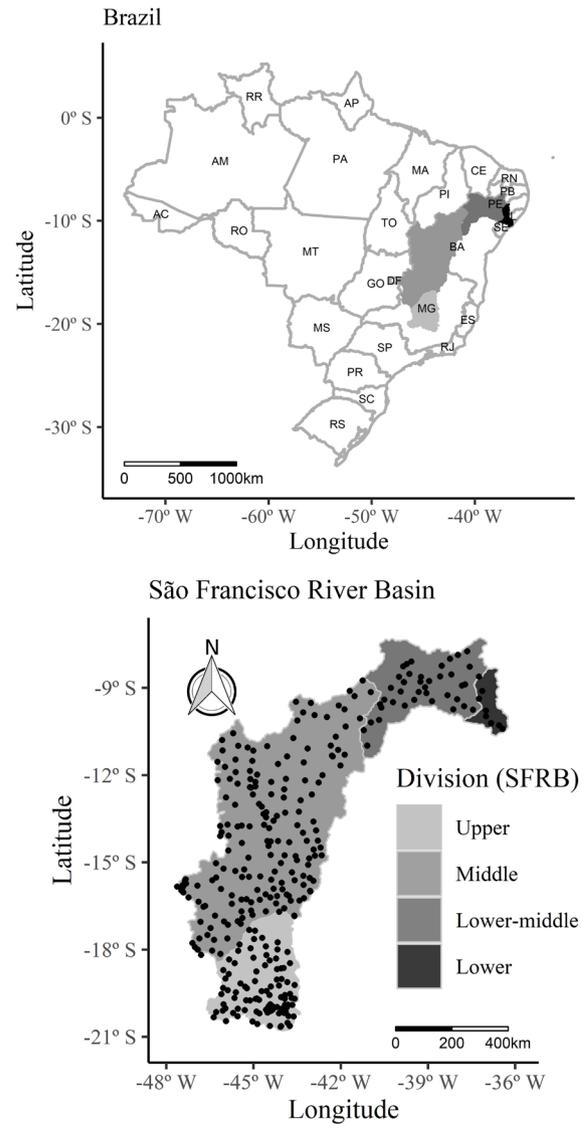
The main river in the SFRB is the São Francisco River. This river originates in Serra da Canastra, the State of Minas Gerais, and covers a drainage area of 639,219  $km^2$  in an extension of 2,863  $km$ , corresponding to 8% of Brazil. On its natural

course to the ocean, the river passes through the states of Minas Gerais, Bahia, Pernambuco, Sergipe, and Alagoas, before flowing into the Atlantic. For planning, the SFRB is divided into four physiographic regions: Upper, Middle, Lower-Middle, and Lower São Francisco. Figure 1 shows the SFRB representation identifying these four regions.

According to the classification of Köppen-Geiger (Alvares et al. 2013), the SFRB includes five different climatic regions: tropical climate (Aw - with a summer rainy season; As - with a winter rainy season), semi-arid climate (BSH - low latitude and altitude), and humid subtropical (Cwa - with dry winter and hot summer; Cwb - with dry winter and temperate summer). Consequently, the different ecosystems throughout the region comprise the Atlantic Rainforest (SFRB Headland), Cerrado (Upper and Middle São Francisco), Caatinga (East and Middle-Lower São Francisco), Caatinga, Atlantic forest, mangroves, and coastal vegetation (Lower São Francisco) (Bezerra et al. 2019).

Because of its extension, several atmospheric systems act in the distribution of rainfall on the SFRB: the Intertropical Convergence Zone (ITCZ) and High-Level Cyclonic Vortices (HLCV) act in Lower São Francisco mainly in the first months of the year; the South Atlantic Convergence Zone (SACZ) and frontal systems (cold fronts) act in the Upper, Middle, and Lower-Middle São Francisco during the austral summer; the East Undulatory Disorders (DOL), Instability Lines, and breezes act throughout the year (Cavalcanti 2016).

The El Niño phenomenon, which designates the significant changes in temperature on the surface of waters in the Pacific Ocean, has negatively influenced the rainfall regime in regions of Northeastern Brazil. Costa (2012) found that the strongest El Niño events imply longer droughts in northeastern Brazil, resulting in cycles that alter weather and climate conditions



**Figure 1.** Map of São Francisco River Basin (SFRB) and distribution of the 341 pluviometric stations.

and negatively impact rainfall distribution across the region. Gurjão et al. (2012) found that El Niño and the action of the HLCV inhibited the development of clouds and contributed to the reduction of rainfall in several locations in the SFRB. Bezerra et al. (2019) stated that the Lower-Middle region of São Francisco is the most vulnerable to severe droughts during periods of strong El Niño.

### Spatio-temporal model

Observations of rainfall precipitation were modeled following the spatio-temporal stochastic process, considering the random function  $\{Z(\mathbf{s},t):(\mathbf{s},t)\in D\subseteq\mathbb{R}^2\times\mathbb{R}_+\}$ ; this is  $d=2$ , corresponding to the spatial domain dimension where the geographical coordinates (latitude and longitude) of each sampled point was used. The  $Z(\mathbf{s},t)$  is decomposed into the sum of the trend component  $m(\mathbf{s},t)$  and stochastic residual  $\varepsilon(\mathbf{s},t)$ :

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

In Equation 1,  $m(\mathbf{s},t) = E[Z(\mathbf{s},t)](\mathbf{s},t)$  gives the expected value of the process  $Z(\mathbf{s},t)$ , representing the large-scale variation in the process.

### Trend component

In most of the observed phenomena, component cannot be modeled by a deterministic function because of the randomness in the process. Thus, we need to model this component through stochastic functions designed to model data patterns observed in space and time (Monteiro et al. 2017). The modeling of this component included the covariates collected along with the response variable. In this study, we used each sample point's geographical coordinates and annual effects inserted through a time index. To model this regression, we considered a Generalized Additive Models for Location, Scale, and Shape (GAMLSS) since it allows for greater flexibility in the regression fitted model. For comparison purposes, for the fitted model, we used the normal Linear Model (LM) and the GAMLSS with Normal (NO), Zero-Adjusted Gamma distribution (ZAGA) and Generalized Gama (GG) models.

Let  $\mathbf{z}^T = (z_1, z_2, \dots, z_n)$  be a vector for response variable with  $n$  observations. Also, assume that observations  $z_i (i=1,2,\dots,n)$  are independent, conditioned to  $\theta_k (k=1,2,\dots,p)$ , where  $p$  is the

number of parameters. Now, function  $\mathbf{g}_k(\cdot)$  is defined as a monotonous link relating  $\theta_k$  to the explanatory variables and random effects through an additive model expressed by:

$$\mathbf{g}_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} Y_{jk} \phi_{jk} \quad (2)$$

In Equation (2)  $\theta_k$  and  $\eta$  are vectors of size  $n$ ;  $\beta_k$  is a vector of size parameters  $J'_k$ ;  $X_k$  and  $Y_{jk}$  are design matrices of dimension  $n \times J'_k$  and  $n \times q_{ij}$ , respectively; and  $\phi_{jk}$  is a  $q_{ij}$ -dimensional random. Model 2 is called GAMLSS (Stasinopoulos et al. 2017).

If  $Y_{jk} = I_n$ , where  $I_n$  is an identity matrix of dimensions  $n \times n$  e  $\phi_{jk} = h_{jk}(\mathbf{x}_{jk})$  for all combinations of  $j$  and  $k$ , Equation (2) takes the form:

$$\mathbf{g}_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}) \quad (3)$$

where  $\mathbf{x}_{jk}$  are vectors of size  $n$ ,  $h_{jk}(\mathbf{x}_{jk})$  is a vector function of covariates  $x_{jk}$  or evaluating the function  $h_{jk}$  by  $\mathbf{x}_{jk}$ . Equation (3) is called a semi-parametric GAMLSS model, because it contains parametric and semi-parametric terms. In this study, when GAMLSS models were fitted, we used the most current version of Smoothing Cubic Spline functions with singular value decomposition (Stasinopoulos et al. 2017).

In the regression fitted with covariates latitude ( $x_{11}$ ), longitude ( $x_{12}$ ) and temporal index ( $x_{13}$ ). The temporal index was constructed as follows: number 1 was assigned to Jan/1989, number 2 to Feb/1989, and so on. Thus, the models fitted for comparison with the trend component are expressed as:

$$g_1(\boldsymbol{\mu}) = \ln(\boldsymbol{\mu}) = \beta_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13}$$

$$g_2(\boldsymbol{\sigma}) = \ln(\boldsymbol{\sigma}) = \beta_{20} \quad (4)$$

$$g_3(\mathbf{v}) = \ln(\mathbf{v}) = \beta_{30}$$

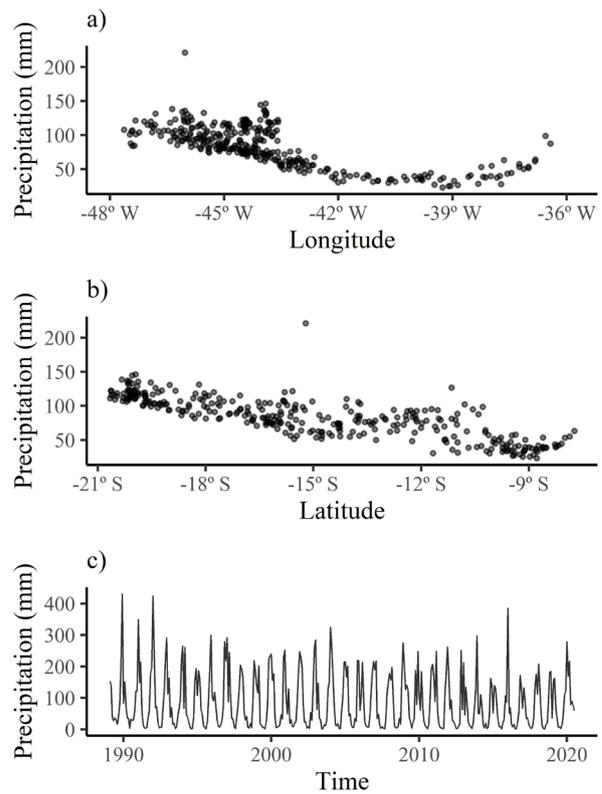
In Equation 4,  $\beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13}$ ,  $\beta_{20}$  and  $\beta_{30}$  represent the parameters to be estimated in the model;  $\mu$ ,  $\sigma$  and  $\nu$  represent the probability distribution parameters related to location, scale, and shape in the GAMLSS models, respectively. The function  $g_k(\cdot)$  is defined as a monotonous link function relating  $\theta_k$  to the explanatory variables and random effects through an additive model. Note that with a Normal distribution in the GAMLSS regression fitted model, we need not include the  $\nu$  parameter because this distribution includes only location and scale parameters. A major advantage of using GAMLSS regression in the trend component is that it allows to insert different link functions into the linear predictor, and it is not necessary to apply Box-Cox transformations to the response variable.

We used in the temporal index the *Smoothing Cubic Spline* functions. These functions are extremely useful for modeling the relationships between linear predictors and the response variable, especially when these relationships are of high degree or are non-linear polynomials. Figure 2 shows the behavior of precipitation between the covariates longitude, latitude and the temporal index in the Basin. There is an effect of precipitation as a function of geographic coordinates and a seasonal behavior, justifying the use of spline functions in the modeling.

In this paper, the fit for Equation 1 is performed separately, taking a two-step approach. First, we estimate the trend component and then model the spatio-temporal residuals having spatio-temporal dependence. In the next section, we give the details of modeling these residuals.

**Stochastic residual**

For a regression fitted for the trend component, we use the maximum likelihood method. It was assumed that the errors are uncorrelated.



**Figure 2.** Relation of precipitation observed between longitude a), latitude b) and time c) covariates in the SFRB for the 1989–2019 period.

However, we know that the residuals produced by this fitted model present a spatio-temporal autocorrelation. Thus, to model these residuals, we proposed the fitted model for a valid space-time variogram, considering the spatial, temporal, and spatio-temporal dependence structures of the residuals. The theoretical variogram model was fitted to the empirical variogram model as follows:

$$\hat{\gamma}_{st}(\mathbf{r}_s, r_t) = \frac{1}{2|L(\mathbf{r}_s, r_t)|} \sum_{L(\mathbf{r}_s, r_t)} [\varepsilon(\mathbf{s} + \mathbf{r}_s, t + r_t) - \varepsilon(\mathbf{s}, t)]^2 \quad (5)$$

In Equation (5),  $|L(\mathbf{r}_s, r_t)|$  is the cardinality for the set  $L(\mathbf{r}_s, r_t)$ , and  $\mathbf{r}_s$  and  $r_t$  are the spatial and temporal distances, respectively.

Several methods are presented in the literature for incorporating a spatio-temporal

dependence structure of the stochastic residual (Cressie & Huang 1999, Iaco et al. 2001, Gneiting 2002, Snepvangers et al. 2003). The main difference between space and time covariance models is in forming their structures since some consider space and time to act separately, while others consider them non-separable (Iaco 2010). In the separable models class, the space-time covariance functions can be written as the sum or product of purely spatial and purely temporal components. Although such models present facilities in the computational aspect, it does not allow for space-time interaction, which is the main disadvantage of these models (Montero et al. 2015). Thus, this study proposes a non-separable type of model approach using the generalized sum-product type spatio-temporal covariance model (Gräler et al. 2016, Medeiros et al. 2019). The generalized sum-product model ( $\gamma_{st}(\mathbf{h}_s, h_t)$ ) is expressed as follows:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_{st}(\mathbf{h}_s, 0) + \gamma_{st}(0, h_t) - k\gamma_{st}(\mathbf{h}_s, 0)\gamma_{st}(0, h_t). \quad (6)$$

In Equation (6),  $\gamma_{st}(\mathbf{h}_s, 0)$  and  $\gamma_{st}(0, h_t)$  represent the marginal variogram models of the spatial and temporal effects, respectively. The parameter  $k$  represents the magnitude of the effect of the spatio-temporal interaction.

### Regression kriging

The choice of the variogram model for each model component in spatio-temporal kriging is of fundamental importance to obtain the best unbiased predictor. This predictor is given by:

$$\hat{\varepsilon}(s_0, t_0) = \mathbf{c}_0^T \mathbf{C}_n^{-1} \bar{\varepsilon}, \quad (7)$$

where  $\hat{\varepsilon}(s_0, t_0)$  is the value obtained using the kriging method for the space-time point  $(s_0, t_0)$ ;  $\mathbf{c}_0^T$  is the vector for size  $n$  of the covariance between the residuals of the estimated and observed points;  $\mathbf{C}_n$  is the covariance matrix for the residuals of dimension  $n \times n$ ; and  $\bar{\varepsilon}$  is the vector for the residuals of fitting Model (Equation 7). The final

predictor (Equation 8) of the variable  $Z$  for location  $(s_0, t_0)$  is given by:

$$z(s_0, t_0) = \hat{m}(s_0, t_0) + \hat{\varepsilon}(s_0, t_0), \quad (8)$$

where  $\hat{m}(s_0, t_0)$  is the trend components estimated value for location  $(s_0, t_0)$ .

Note that regression kriging consists of the separately fitted trend component and residuals. By fitting a regression model, we obtain the predicted values for the response variable and the residuals for all the points sampled in space-time. We also obtain the empirical variogram for these residuals and for the theoretical fitted model. The residuals are then interpolated and added to the estimated trend. In recent years, regression kriging has made theoretical advances, with several applications in real phenomena (Hengl et al. 2012, Heuvelink et al. 2012, Kilibarda et al. 2014).

### Validation test

The best structure was determined using the validation test. The method consisted of randomly removing 10 pluviometric stations and then making the prediction of the entire time series of these 10 stations that had 2315 observations. Finally, the data set will have the observed and predicted values. For select the best spatio-temporal variogram model we calculate the Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), Nash-Sutcliffe Efficiency (NSE) and Determination Coefficient ( $R^2$ ). After selecting the best spatio-temporal variogram model, we built scatter plots between observed and predicted precipitation values, considering the trend component, spatio-temporal regression kriging and data estimated by the NASA POWER project (Sparks 2018).

**Statistical software**

We performed all analyses in this research using the software R (R Core Team 2020). For the fitted trend, we used the library *GAMLSS* (Stasinopoulos et al. 2017). For spatio-temporal regression kriging we used the library *gstat* (Gräler et al. 2016) and for the construction of the maps the *sp* library was used (Bivand et al. 2013).

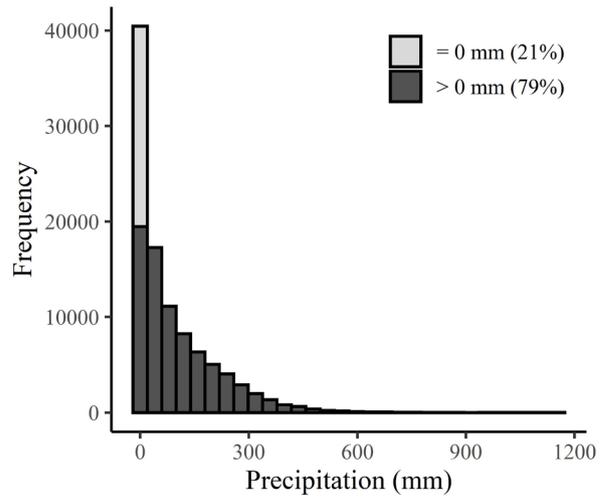
**RESULTS**

**Exploratory analysis**

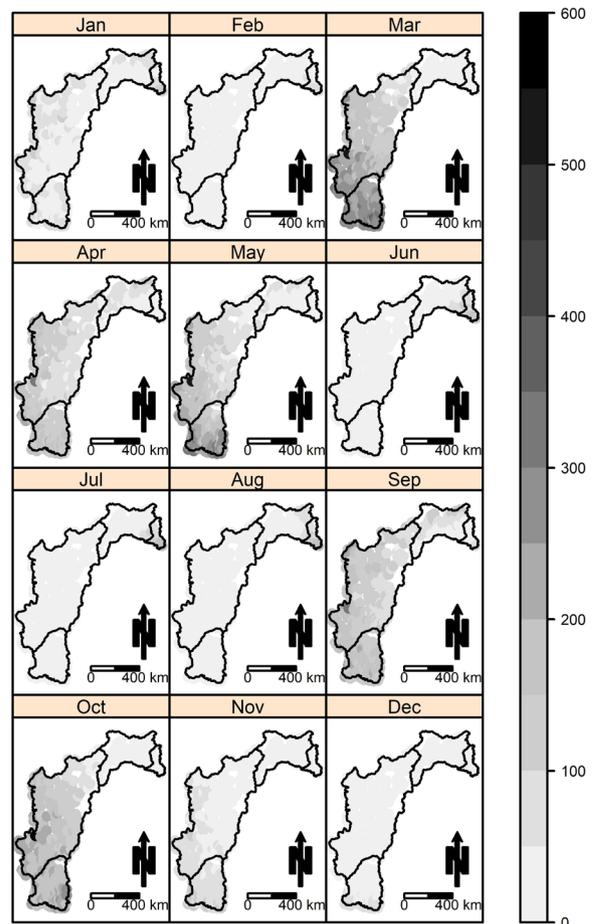
Figure 3 shows the histogram for the variable monthly total precipitation observed in all the 341 points sampled over 31 years. This variable showed a positive skewness, with 21% of the observations equal to zero. The median value of precipitation was 41 mm/month with a standard deviation of 110 mm/month. Additionally, 25% of the observations showed less than 2 mm/month of precipitation. These climatic characteristics of SFRB are associated with the performance and behavior of many atmospheric teleconnection systems and standards influencing the annual average precipitation ranging from 350 mm (Lower-Middle SFRB) to 1500 mm (Upper SFRB), characterizing high spatial climatic variability of the region (Bezerra et al. 2019), as may be seen in Figure 4.

Figure 4 highlights the maps of the median values obtained from the 1989–2019 historical precipitation series for all the 12 months. Generally, the highest rainfall indices are found in the Upper São Francisco region, followed by the Middle region, with lower rainfall indices in the Lower and Lower-Middle São Francisco regions.

The hydrographic regions that divide the basin show different behavior of the spatial variability of precipitation. One justification for this is the climatic differences, especially in



**Figure 3. Histogram of precipitation. The colors in the figure represent the percentages of values equal to or greater than zero in the data set.**



**Figure 4. Map of the median value of precipitation in the SFRB for the 1989–2019 period.**

the region of the mouth of the São Francisco River, due to its proximity to the Atlantic Ocean. Upper São Francisco, mainly in the region of the state of Minas Gerais, has the highest altitudes and a climate classified as humid. Middle São Francisco, having the largest area among the hydrographic regions, is located in the semiarid region, presenting lower precipitation indices in relation to the Alto São Francisco and predominating the semiarid climate. Lower-middle is known for having the lowest precipitation indices and with an average annual temperature of 27 °C, the climate in this region is classified as semi-arid and arid. With its proximity to the Atlantic Ocean, the Lower São Francisco has a milder climate, ranging from semi-arid to humid tropical (Hermuche 2002).

In addition to spatial variability, the SFRB indicates high temporal variability, justifying the application of a model considering spatial, temporal, and space-time interactions. Furthermore, the distribution of rainfall in certain SFRB areas shows monthly variability.

### Spatio-temporal geostatistics

Cross-validation was used to choose the best structure, combining the different regression models for the trend with the different theoretical variogram models for the stochastic residual (Table I). In all, regression kriging fitted nine spatio-temporal structures for the residual component and using four probability density functions in the trend component, resulting in 36 adjusted spatio-temporal models.

According to the values of the RMSE, NRMSE(%), NSE and R<sup>2</sup> statistics, we have the choice of the ZAGA probability distribution for the trend component. For the stochastic residue, the selected adjustment to the generalized product-sum variogram occurred with the Spherical models for the spatial and temporal components.

Table II shows the estimates of the parameters of the generalized product-sum models adjusted to the residues of the GAMLSS model with ZAGA distribution.

The results indicate the existence of a spatio-temporal pattern of the residues obtained from the adjustment of the trend with ZAGA distribution for the accumulated monthly precipitation (Table II). It is noted through the sill parameters that the temporal component is responsible for greater variability compared to the spatial component. In addition, the adjustment revealed a spatial dependence of up to 873 km between the pluviometric stations and temporal autocorrelation of approximately 1.6 months.

The spatio-temporal variogram equation based on Table II is:

$$\hat{\gamma}_{st}(h_s, h_t) = \hat{\gamma}_{st}(h_s, 0) + \hat{\gamma}_{st}(0, h_t) - k\hat{\gamma}_{st}(h_s, 0)\hat{\gamma}_{st}(0, h_t).$$

With

$$\hat{\gamma}_{st}(h_s, 0) = 1043.109 + (3987.192 - 1043109) \left[ (1.5h/873) - 0.5(h/873)^3 \right]$$

and

$$\hat{\gamma}_{st}(0, h_t) = 0.000 + (5607.506 - 0) \left[ 1.5h/1.6 - 0.5(h/1.6)^3 \right].$$

Figure 5 shows the empirical variogram (“sample”) and the adjustment of the generalized product-sum model (“productSum”) for the monthly accumulated precipitation after removing the trend. It is noticed that the residues present a correlation in space and also in time, justifying once again the use of the methodology proposed in this research.

In Figure 6 we can visualize the marginal behavior in relation to space (Figure 6a) and time (Figure 6b) for the empirical spatio-temporal variogram (Figure 5a). Note that the highest values of variances are present in the temporal structure.

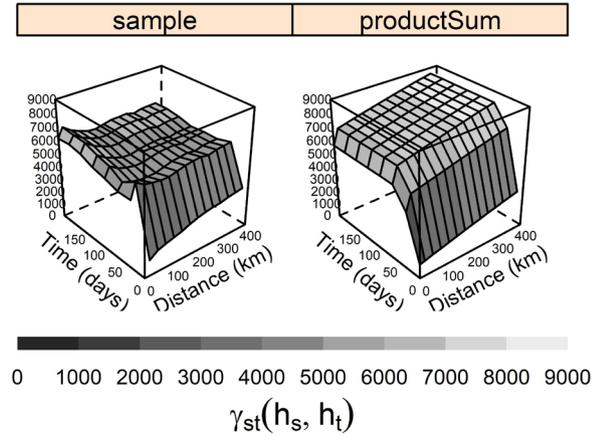
**Table I. Results of cross-validation considering the theoretical models of Spherical, Gaussian and Exponential variogram for the Spatial and Temporal components, and the different probability density (PDF) functions for the trend component, being them Generalized Gamma (GG), Zero Adjusted Gamma (ZAGA), Normal GAMLSS (NO) and Linear Model Normal (LM). Metrics to evaluate the error were Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), Nash-Sutcliffe Efficiency (NSE) and Determination Coefficient ( $R^2$ ).**

PDF	Spatial (S)	Temporal (T)	RMSE	NRMSE(%)	NSE	$R^2$
GG	Exponential	Exponential	42.0077	46.50	0.7834	80.66
GG	Exponential	Spherical	41.5636	45.80	0.7901	81.13
GG	Exponential	Gaussian	42.0899	46.70	0.7823	80.58
GG	Spherical	Exponential	41.5677	45.70	0.7911	81.16
GG	Spherical	Spherical	41.5327	45.70	0.7915	81.18
GG	Spherical	Gaussian	41.5819	45.70	0.7909	81.15
GG	Gaussian	Exponential	44.5708	49.20	0.7578	78.32
GG	Gaussian	Spherical	44.5928	49.30	0.7573	78.30
GG	Gaussian	Gaussian	44.5501	49.20	0.7582	78.35
ZAGA	Exponential	Exponential	41.8600	46.40	0.7847	80.73
ZAGA	Exponential	Spherical	41.8765	46.40	0.7845	80.71
ZAGA	Exponential	Gaussian	41.8489	46.40	0.7849	80.74
ZAGA	Spherical	Exponential	41.3061	45.40	0.7938	81.31
ZAGA	Spherical	Spherical	41.2633	45.30	0.7943	81.35
ZAGA	Spherical	Gaussian	41.3368	45.40	0.7934	81.28
ZAGA	Gaussian	Exponential	44.6650	49.40	0.7558	78.20
ZAGA	Gaussian	Spherical	43.8441	47.60	0.7733	79.21
ZAGA	Gaussian	Gaussian	45.2330	50.50	0.7446	77.56
NO	Exponential	Exponential	41.9194	46.70	0.7822	80.62
NO	Exponential	Spherical	41.9559	46.50	0.7837	80.63
NO	Exponential	Gaussian	41.9016	46.40	0.7845	80.68
NO	Spherical	Exponential	41.3223	45.40	0.7937	81.28
NO	Spherical	Spherical	42.1298	46.60	0.7824	80.49
NO	Spherical	Gaussian	45.6921	49.50	0.7551	77.57
NO	Gaussian	Exponential	45.0989	50.10	0.7486	77.75
NO	Gaussian	Spherical	44.6500	49.30	0.7569	78.25
NO	Gaussian	Gaussian	46.0591	51.80	0.7317	76.71
LM	Exponential	Exponential	41.9293	46.70	0.7819	80.61
LM	Exponential	Spherical	41.9506	46.70	0.7817	80.59
LM	Exponential	Gaussian	41.9359	46.80	0.7806	80.58
LM	Spherical	Exponential	41.3988	45.90	0.7889	81.12
LM	Spherical	Spherical	41.2977	45.80	0.7901	81.21
LM	Spherical	Gaussian	41.8709	46.60	0.7828	80.67
LM	Gaussian	Exponential	44.6093	49.70	0.7532	78.18
LM	Gaussian	Spherical	44.6728	49.70	0.7530	78.13
LM	Gaussian	Gaussian	44.8354	49.50	0.7545	78.06

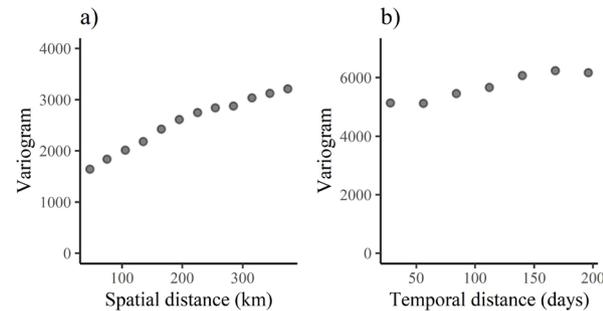
**Table II.** Estimates of parameters (Nugget, Sill, Range and k) in the generalized product-sum model and the spatial and temporal dependency index (DI) for the residuals of accumulated monthly rainfall of the São Francisco River Basin, Brazil. The parameter k represents the effect of the space-time interaction and contemplates the global sill.

Model	Variogram	Nugget	Sill	Range	k	DI(%)
Spatial	Spherical	1043.109	3987.192	873 km	$1.490 \times 10^{-8}$	26.160%
Temporal	Spherical	0.000	5607.506	1.6 months		0.000%

DI = Nugget/Sill. DI < 25%: strong dependency; 25% ≤ DI < 75%: medium dependency; DI ≥ 75% weak dependency (Cambardella et al. 1994).



**Figure 5.** Empirical spatiotemporal variogram (“sample”) and the generalized product-sum adjusted variogram (“productSum”) for the precipitation residuals in the SFRB.



**Figure 6.** Empirical marginal variograms for space a) and time b) of precipitation residues in SFRB.

**Model performance**

Figure 7 shows the difference between the observed and estimated precipitation with the trend component (Figure 7a), spatio-temporal regression kriging (Figure 7b) and data estimated by the NASA POWER project (Figure 7c). Note the absence of a good fit when only the trend component is used to predict precipitation.

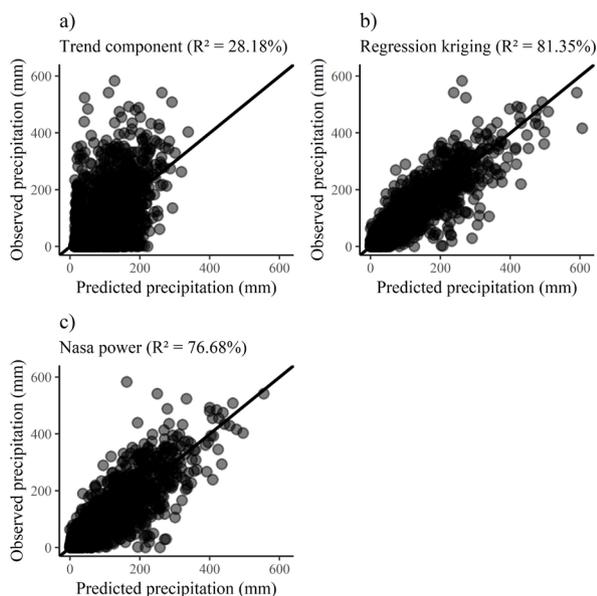
As expected, when regression kriging is used, the predicted values are close to those observed, providing evidence that the proposed methodology performed well, even surpassing the data estimated by NASA POWER.

**Spatio-temporal prediction**

As an illustration of regression kriging, Figure 8 shows the predictions of SFRB precipitation values for all months of 2019. This type of approach is essential to detect the possible changes in rainfall distribution in this region, enabling preventive and mitigation activities (Araujo & Celeste 2019).

Figure 8. Maps of estimated precipitation in SFRB for all months of the year 2019.

In this study, it was not intended to go deeper into the results of spatio-temporal interpolation, but rather to present an alternative to trend component modeling through GAMLSS models, which presented good results for Regression Kriging. Thus, for future work, it is necessary to present an analysis of constructions of precipitation scenarios in the SFRB basin, since there are studies showing a reduction in precipitation in the basin in the coming years. Marengo et al. (2011) show decreasing rainfall projections, with a 15% of annual precipitation by 2040 in the SFRB, which could go up to 20% in the rainy season (Bezerra et al. 2019, Silveira et al. 2016, Santos et al. 2018). This means a further decrease in volume of recharge of the São Francisco River, leading to climate risks very

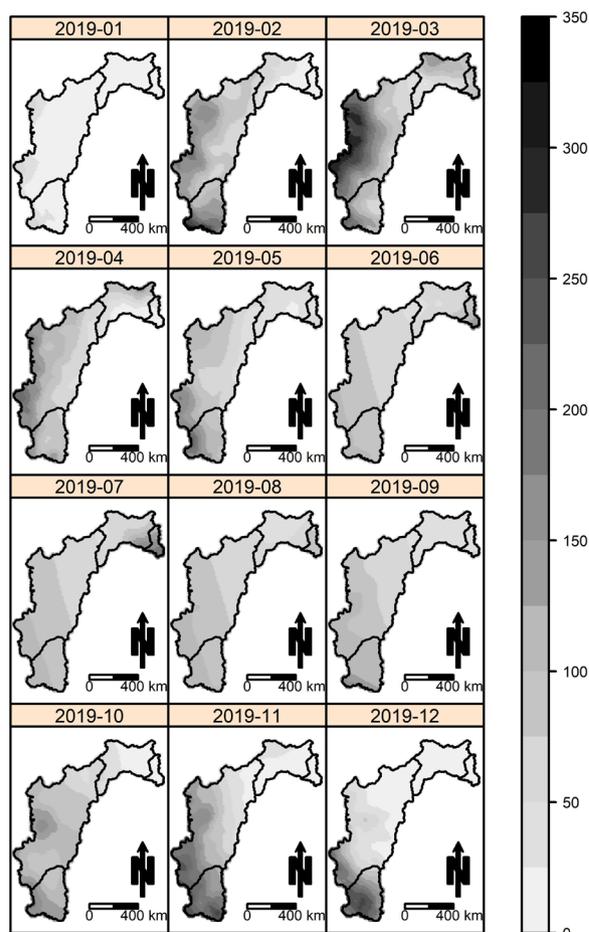


**Figure 7.** Comparison between the observed and predicted precipitation with the trend component a), regression kriging b) and NASA POWER c), in the randomly removing 10 pluviometric stations.

shortly, with even more problems in the country’s economy due to lack of water resources, mainly for human consumption, agriculture, and energy production in the NEB.

**DISCUSSION**

Knowledge about the distribution of precipitation in space and time is of utmost importance for scientific knowledge, which is used to understand regional and global changes to processes better involving water and its associated materials. Efficient monitoring of a region’s water resources, which takes into account rainfall distribution, can deal with natural disasters such as floods, prolonged periods of drought. In addition, rainfall estimation is of fundamental importance during hydrological modeling, models that have a high level of reliability of these estimates are required, as these can be used to prevent flooding and to assess runoff of water quality and ecosystem health in urbanized



**Figure 8.** Maps of estimated precipitation in SFRB for all months of the year 2019.

areas (Knight et al. 2005). So that we can be aware of the distribution of rainfall throughout an area, it is necessary to use tools capable of estimating values with high quality and resolution, mainly for urban interpolation and in catchment-scale (Hu et al. 2019). Additionally, the spatial variability of rainfall can significantly influences the flow generation mechanisms of urban basins (Segond et al. 2007).

The geostatistical methodology presented in this study can, in principle, be used for urban and catchment interpolation. However, when using analysis for the spatial distribution of precipitation data on a daily scale, geostatistical methods present several difficulties, mainly

in the presence of large temporal variability (Javari 2017) and spatial variability, as it increases the complexity of estimates on the uptake scale (Knight et al. 2005). However, once considered that the variable presents an irregular behavior, these difficulties can be considered in the spatio-temporal geostatistics, because this methodology, it is possible to enter autocorrelation structures space-time capable, such as those used in this manuscript, to incorporate modeling the occurrence of high rainfall variability.

In this study, we present a novel method to model the trend component in the space-time stochastic process. GAMLSS regression models (Stasinopoulos et al. 2017) have become a powerful tool in modeling this trend since they allow for the presence of non-parametric terms and use of different probability distribution families in regression adjustment. The modeling of this component is of paramount importance because it is responsible for the large variation in the phenomenon studied. The results demonstrated that GAMLSS regression models are suitable and preferable in regression-fitted models. Additionally, the approach taken in this manuscript for precipitation value interpolation allows for flexible modeling for the trend component, since it is possible to determine different link functions, in particular to the location parameter (mean), no longer need to apply the boxcox type changes in the response variable. In the literature, some studies have used methods involving GLM for trend adjustment, such as (Pebesma et al. 2005) performed a purely spatial interpolation, modeling the trend with Poisson regression as a function of water depth and distance to the coast and residues by ordinary kriging.

Recent studies have suggested the adjustment of these models in phenomena having characteristics that do not follow a normal

distribution, such as discrete quantitative data, censored data, and data with excessive zeros in continuous quantitative variables (Stauffer et al. 2016, Sá et al. 2018). However, when modeling the trend component with GAMLSS, independence was assumed between the observations of the response variable for the different spatio-temporal points, this assumption was necessary to obtain the estimates by the maximum penalized likelihood method. However, this drawback was circumvented by using regression kriging, which by adjusting the co-variance function took into account for the autocorrelation of the residuals. It is noteworthy that the regression kriging interpolated only for the location parameter and that in the modeling of the residuals, it was considered a Gaussian random field. A solution to these weaknesses pointed out in our modeling would be to adjust in one-step a GAMLSS regression and consider in the covariance matrix spatio-temporal autocorrelation structures. This type of solution has already been presented for area data, considering regression model with spatial effects (Bastiani et al. 2018).

To apply the methodology proposed in this paper, we used the set of monthly total precipitation data evaluated in the SFRB region space-time. A descriptive analysis of these data found that the precipitation presented a positive skewness distribution, with excessive zeros and irregular distribution of the monthly median value in space and time, indicating the need for spatio-temporal modeling for these data. For these reasons, it was hypothesized that the normal regression model might not be the most adequate one for the trend component in the multiple regression fitted model, thus suggesting the GAMLSS regression modeling with ZAGA, GG, and Normal distributions. The normal regression model has been widely used in the trend component fitted models in

indexed space-time phenomena (Gräler et al. 2012, Kilibarda et al. 2014, Hu et al. 2015, Iaco et al. 2015). However, recent studies have suggested adjustment of generalized linear models to evaluate the variables with skewed distributions (Menezes et al. 2016, Monteiro et al. 2017). Thus, this paper's proposal is to consider GAMLSS models with different probability distributions in the trend component fitted model present in the spatio-temporal stochastic process.

## CONCLUSIONS

Spatial-temporal regression kriging was used to analyze the SFRB precipitation data, combining the kriging on the residuals with GAMLSS regression, generating better results in comparison with the regression with Normal distribution. A weakness of this research consisted in the non-construction of precipitation projections in the SFRB, which is considered a possibility for works that will follow this research.

## REFERENCES

ACHITE M, BUTTAUFUOCO G, TOUBAL KA & LUCA F. 2017. Precipitation spatial variability and dry areas temporal stability for different elevation classes in the Macta basin (Algeria). *Environ Earth Sci* 76(13): 1-13.

ALVARES CA, STAPE JL, SENTELHAS PC, GONÇALVES JDM & SPAROVEK G. 2013. Köppen's climate classification map for Brazil. *Meteorol Zeitschrift* 22(6): 711-728.

ANA – AGÊNCIA NACIONAL DE ÁGUAS. 2017. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno/Agência Nacional de Águas (ANA). Brasília, 169 p.

ARAUJO MVO & CELESTE AB. 2019. Rescaled range analysis of streamflow records in the São Francisco River Basin, Brazil. *Theor Appl Climatol* 135(1): 249-260.

BASTIANI F, RIGBY RA, STASINOPOULOUS DM, CYSNEIROS AH & URIBE-OPAZO MA. 2018. Gaussian Markov random field spatial models in GAMLSS. *J Appl Stat* 45(1): 168-186.

BEZERRA BG, SILVA LL, SILVA CMS & CARVALHO GG. 2019. Changes of precipitation extremes indices in São

Francisco River Basin, Brazil from 1947 to 2012. *Theor Appl Climatol* 135(1): 565-576.

BIVAND RS, PEBESMA E & GÓMEZ-RUBIO V. 2013. Interpolation and geostatistics. In: *Applied spatial data analysis with R*, Springer, New York, NY, p. 213-261.

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): 1501-1511.

CAVALCANTI IFA. 2016. *Tempo e clima no Brasil*. São Paulo, SP: Oficina de textos, 463 p.

COSTA JA. 2012. O fenômeno El Niño e as secas no Nordeste do Brasil. *Rev Cient IFAL* 1(4): 71-82.

CRESSIE N & HUANG H-C. 1999. Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. *J Am Stat Assoc* 94(448): 1330-1339.

FOEHN A, HERNÁNDEZ JG, SCHAEFLI B & CESARE G. 2018. Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *J Hydrol* 563: 1092-1110.

GLEICK PH & CAIN NL. 2004. *The world's water 2004-2005: the biennial report on freshwater resources*. Island Press, 262 p.

GNEITING T. 2002. Nonseparable, Stationary Covariance Functions for Space-Time Data. *J Am Stat Assoc* 97(458): 590-600.

GRÄLER B, PEBESMA E & HEUVELINK G. 2016. Spatio-temporal interpolation using gstat. *R J* 8(1): 204-218.

GRÄLER B, REHR M, GERHARZ L & PEBESMA E. 2012. Spatio-temporal analysis and interpolation of PM10 measurements in Europe for 2009. *ETC/ACM Tech Pap* 8: 1-29.

GURJÃO CDS, CORREIA MF, CHAVES FILHO JB & ARAGÃO MRS. 2012. Influência do Enos (El Niño-Oscilação Sul) no Regime Hidrológico do Rio São Francisco: uma Análise em Regiões com Fortes Pressões Antrópicas. *Rev Bras de Geogr Fis* 5(4): 774-790.

HENGL T, HEUVELINK GB, TADIĆ MP & PEBESMA EJ. 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor Appl Climatol* 107(1-2): 265-277.

HERMUCHE PM. 2002. *O Rio São Francisco*. Companhia de Desenvolvimento dos Vales do São Francisco e Parnaíba – CODEVASF, 58 p. Available in <https://cdn.agenciapeixe vivo.org.br/media/2019/06/Cartilha-sobre-o-Rio-S%C3%A3o-Francisco.pdf>.

- HEUVELINK GB, GRIFFITH DA, HENGL T & MELLES SJ. 2012. Sampling Design Optimization for Space-Time Kriging. In: *Spatio-Temporal Design*. Hoboken, Nova Jersey, EUA, Wiley-Blackwell 9: 207-230.
- HU D, SHU H, HU H & XU J. 2017. Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Cluster Comput* 20(1): 347-357.
- HU Q, LI Z, WANG L, HUANG Y, WANG Y & LI L. 2019. Rainfall Spatial Estimations: A Review from Spatial Interpolation to Multi-Source Data Merging. *Water* 11(3): 579.
- HU Y ET AL. 2015. Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. *PLOS Neglect Trop D* 9(2): e0003470.
- IACO SD. 2010. Space-time correlation analysis: a comparative study. *J Appl Stat* 37(6): 1027-1041.
- IACO SD, MYERS DE & POSA D. 2001. Space-time analysis using a general product-sum model. *Stat Probabil Lett* 52(1): 21-28.
- IACO SD, PALMA M & POSA D. 2015. Spatio-temporal geostatistical modeling for French fertility predictions. *Spat Stat-Neth* 14: 546-562.
- IPCC. 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. In: Barros VR et al. (Eds), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 688 p.
- IPCC. 2021. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. In: Masson-Delmotte VP et al. (Eds), Cambridge University Press, In Press, 2338 p.
- JAVARI M. 2017. Geostatistical modeling to simulate daily rainfall variability in Iran. *Cogent Geosci* 3(1): 1416877.
- KILIBARDA M, HENGL T, HEUVELINK GB, GRÄLER B, PEBESMA E, TADIĆ PM & BAJAT B. 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *J Geophys Res-Atmos* 119(5): 2294-2313.
- KILIBARDA M, TADIĆ PM, HENGL T, LUKOVIĆ J & BAJAT B. 2015. Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. *Spat Stat-Neth* 14: 22-38.
- KNIGHT Y, YU B, JENKINS G & MORRIS K. 2005. Comparing rainfall interpolation techniques for small subtropical urban catchments. In: *Proceedings of Scientific Conference, Modelling and Simulation*, p. 1674-1680.
- MARENGO JA ET AL. 2011. Development of regional future climate change scenarios in South America using the Eta CPTec/HadCM3 climate change projections: climatology and regional analyses for the Amazon, São Francisco and the Paraná River basins. *Clim Dynam* 38(9-10): 1829-1848.
- MARTÍNEZ WA, MELO CE & MELO OO. 2017. Median Polish Kriging for space-time analysis of precipitation. *Spat Stat-Neth* 19: 1-20.
- MEDEIROS ES, LIMA RR, OLINDA RA, DANTAS LG & SANTOS CAC. 2019. Space-time kriging of precipitation: Modeling the large-scale variation with model GAMLSS. *Water* 11(11): 2368.
- MENEZES R, PIAIRO H, GARCIA SOIDAN P & SOUSA I. 2016. Spatial-temporal modellization of the NO<sub>2</sub> concentration data through geostatistical tools. *Stat Method Appl* 25(1): 107-124.
- MONTEIRO A, MENEZES R & SILVA ME. 2017. Modelling spatio-temporal data with multiple seasonalities: The NO<sub>2</sub> Portuguese case. *Spat Stat-Neth* 22: 371-387.
- MONTERO JM, FERNANDEZ-AVILES G & MATEU J. 2015. *Spatial and spatio-temporal geostatistical modeling and kriging*. Hoboken, Nova Jersey, EUA, J Wiley & Sons, 400 p.
- PEBESMA EJ, DUIN RNM & BURROUGH PA. 2005. Mapping sea bird densities over the North Sea: spatially aggregated estimates and temporal changes. *Environmetrics* 16(6): 573-587.
- R CORE TEAM. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- ROGERS PP, LLAMAS MR & CORTINA LM. 2005. *Water crisis: myth or reality?* CRC Press, 344 p.
- SÁ ACL, TURKMAN MAA & PEREIRA JMC. 2018. Exploring fire incidence in Portugal using generalized additive models for location, scale and shape (GAMLSS). *Model Earth Syst Environ* 4(1): 199-220.
- SANTOS CAG, NETO RMB, SILVA RM & PASSOS JSA. 2018. Integrated spatiotemporal trends using TRMM 3B42 data for the Upper São Francisco River basin, Brazil. *Environ Monit Assess* 190(3): 1-20.
- SEGOND ML, WHEATER HS & ONOF C. 2007. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *J Hydrol* 37(1-2): 116-131.
- SILVEIRA C ET AL. 2016. Climate change in the São Francisco river basin: analysis of precipitation and temperature. *RBRH* 21(2): 416-428.

SNEPVANGERS JJJC, HEUVELINK GBM & HUISMAN JA. 2003. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma* 112(3-4): 253-271.

SPARKS AH. 2018. nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R. *J Open Source Softw* 3(30): 1035.

STASINOPOULOS MD ET AL. 2017. Flexible Regression and Smoothing: Using GAMLSS in R. Boca Raton, Florida, EUA, CRC Press, 572 p.

STAUFFER R, MAYR GJ, MESSNER JW, UMLAUF N & ZEILEIS A. 2016. Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. *Int J Climatol* 37(7): 3264-3275.

#### How to cite

DE MEDEIROS ES, DE LIMA RR, DE OLINDA RA, DANTAS LG & DOS SANTOS CAC. 2022. Performance assessment of spatio-temporal regression kriging with GAMLSS models as trends. *An Acad Bras Cienc* 94: e20211241. DOI 10.1590/0001-376520220211241.

*Manuscript received on September 14, 2021;  
accepted for publication on January 22, 2022*

#### ELIAS S. DE MEDEIROS<sup>1</sup>

<https://orcid.org/0000-0002-9694-4019>

#### RENATO R. DE LIMA<sup>2</sup>

<https://orcid.org/0000-0003-4607-4964>

#### RICARDO A. DE OLINDA<sup>3</sup>

<https://orcid.org/0000-0002-0509-8428>

#### LEYDSON G. DANTAS<sup>4</sup>

<https://orcid.org/0000-0003-4175-0951>

#### CARLOS A.C. DOS SANTOS<sup>4</sup>

<https://orcid.org/0000-0003-2414-2911>

<sup>1</sup>Universidade Federal da Grande Dourados, Faculdade de Ciências Exatas e Tecnologia, Rodovia Dourados/Itahum, Km 12, Cidade Universitária, Caixa Postal 364, 79804-970 Dourados, MS, Brazil

<sup>2</sup>Universidade Federal da Lavras, Departamento de Estatística, Rotatória Professor Edmir Sá Santos, s/n, Campus Universitário, Caixa Postal 3037, 37200-900 Lavras, MG, Brazil

<sup>3</sup>Universidade Estadual da Paraíba, Departamento de Estatística, Rua Baraúnas, 351, Universitário, 58429-500 Campina Grande, PB, Brazil

<sup>4</sup>Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, Rua Aprígio Veloso, 882, Universitário, 58428-830 Campina Grande, PB, Brazil

Correspondence to: **Elías Silva de Medeiros**

E-mail: [eliasmedeiros@ufgd.edu.br](mailto:eliasmedeiros@ufgd.edu.br)

#### Author contributions

Conceptualization: ESM, RRL, RAO, LGD and CACS; Methodology: ESM, RRL and RAO; Software: ESM; Validation: ESM and RRL; Formal analysis: ESM, RRL, RAO, LGD and CACS; Investigation: ESM, RRL, RAO, LGD and CACS; Resources: ESM; Data curation: ESM and CACS; Writing-original draft preparation: ESM, RRL, RAO, LGD and CACS; Writing-review and editing: ESM, RRL, RAO, LGD and CACS; Visualization: ESM; Supervision: ESM, RRL, RAO and CACS.

