



ELIAS SILVA DE MEDEIROS

**KRIGAGEM COM REGRESSÃO ESPAÇO-TEMPORAL COM
MODELOS GAMLSS**

LAVRAS – MG

2018

ELIAS SILVA DE MEDEIROS

KRIGAGEM COM REGRESSÃO ESPAÇO-TEMPORAL COM MODELOS GAMLSS

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr. Renato Ribeiro de Lima

Orientador

LAVRAS – MG

2018

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Medeiros, Elias Silva de.

Krigagem com regressão espaço-temporal com modelos
GAMLSS / Elias Silva de Medeiros. - 2018.
121 p.

Orientador(a): Renato Ribeiro de Lima.

.
Tese (doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Geoestatística. 2. Variabilidade. 3. Climatologia. I. Lima,
Renato Ribeiro de. . II. Título.

ELIAS SILVA DE MEDEIROS

**KRIGAGEM COM REGRESSÃO ESPAÇO-TEMPORAL COM MODELOS GAMLSS
SPATIO-TEMPORAL REGRESSION KRIGING WITH GAMLSS MODELS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 07 de Junho de 2018.

Dr. Marcelo Silva de Oliveira	DES-UFLA
Dr. João Domingos Scalon	DES-UFLA
Dra. Izabela Regina Cardoso de Oliveira	DES-UFLA
Dr. Carlos Rogério de Mello	DEG-UFLA
Dr. Ricardo Alves de Olinda	DE-UEPB



Prof. Dr. Renato Ribeiro de Lima
R/ Orientador

**LAVRAS – MG
2018**

RESUMO

A variabilidade espaço-temporal de um fenômeno pode ser decomposta pelo componente de tendência e pelo resíduo estocástico. Esta tese tem por objetivo considerar diferentes modelos probabilísticos no ajuste do componente de tendência, uma vez que a estimativa precisa da distribuição espaço-temporal da variável resposta requer uma distribuição de probabilidade adequada. Para o ajuste deste componente foram considerados os modelos de regressão gaussiana e os modelos aditivos generalizados para posição, escala e forma (GAMLSS). Os resíduos produzidos por esta regressão foram modelados por funções de covariâncias espaço-temporal, que levavam em consideração a dependência espacial, temporal e espaço-temporal. A krigagem com regressão espaço-temporal que combinou a krigagem sobre os resíduos com a regressão GAMLSS apresentou bons resultados, fato este comprovado pelas estatísticas obtidas depois de realizada a validação cruzada. A metodologia apresentada nesta tese possibilitou a criação de cenários futuros para a área em estudo, interpolando locais não observados e prevendo estimativas para anos futuros.

Palavras-chave: Geoestatística. Variabilidade. Climatologia.

ABSTRACT

The spatio-temporal variability of a phenomenon can be decomposed by the trend component and the stochastic residue. This thesis aims to consider different probabilistic models in the adjustment of the trend component, since the precise estimation of the spatial-temporal distribution of the response variable requires an adequate probability distribution. Gaussian regression models and the generalized additive models for position, scale and shape (GAMLSS) were considered for the adjustment of this component. The residuals produced by this regression were modeled by spatio-temporal covariance functions, which took into account spatial, temporal and spatial-temporal dependence. The Spatiotemporal regression Kriging that combined kriging on the residues with the GAMLSS regression showed good results, as evidenced by the statistics obtained after cross-validation. The methodology presented in this thesis allowed the creation of future scenarios for the study area, interpolating unobserved locations and predicting estimates in future years.

Keywords: Geostatistics. Variability. Climatology.

SUMÁRIO

	PRIMEIRA PARTE	6
1	INTRODUÇÃO	6
2	REFERENCIAL TEÓRICO	10
2.1	Modelo estocástico espaço-temporal	10
2.2	Componente de tendência	11
2.2.1	Modelos GAMLSS	12
2.2.2	Estimação do modelo	15
2.2.3	Termos Aditivos	17
2.2.3.1	Splines cúbicos	17
2.2.3.2	<i>Splines</i> de penalização	18
2.3	Geoestatística espaço-temporal	18
2.3.1	Variograma espaço-temporal	19
2.4	Modelos de covariâncias	20
2.4.1	Modelo soma	22
2.4.2	Modelo métrico	23
2.4.3	Modelo “product-sum”	24
2.4.4	Modelo “product-sum” generalizado	26
2.4.5	Modelo “sum-metric”	27
2.4.6	Modelos de Cressie-Huang	28
2.5	Mínimos quadrados generalizados	29
2.6	Krigagem espaço-temporal	30
2.6.1	Equações da krigagem espaço-temporal	31
2.6.2	Análise de desempenho	32
3	CONSIDERAÇÕES FINAIS	34
	REFERÊNCIAS	35
	SEGUNDA PARTE- ARTIGOS	37
	ARTIGO 1 - Rainfall spatiotemporal variability analysis covering recent decades in Paraíba, Brazil	37
	ARTIGO 2 - Space-time kriging of precipitation: modeling the large-scale variation with GAMLSS model	67

ARTIGO 3 - Krigagem com regressão espaço-temporal com modelos

GAMLSS 94

PRIMEIRA PARTE

1 INTRODUÇÃO

A análise de dados espaço-temporais ainda tem sido realizada sob duas perspectivas: fixando a variável espacial e estudando o seu comportamento ao longo do tempo ou fixando o tempo e estudando o espaço. Entretanto, nos últimos anos tem surgido o interesse em estudar conjuntamente a dinâmica (espaço-tempo) de uma dada característica de interesse. Assim, a modelagem espaço-temporal está sendo abordada em trabalhos teóricos (IACO; MYERS; POSA, 2001; DIMITRAKOPOULOS; LUO, 1994; SNEPVANGERS; HEUVELINK; HUISMAN, 2003), bem como em estudos aplicados (JOVEIN; HOSSEINI, 2017; MONTEIRO; MENEZES; SILVA, 2017), sendo que esse tipo de abordagem tem ganhado visibilidade nos últimos anos, dado que fenômenos indexados tanto no espaço quanto no tempo precisam ser compreendidos. Em vista disto, o interesse em analisar a dinâmica espaço-temporal tem proporcionado o desenvolvimento de modelos espaço-temporais, para análise de dados em áreas como epidemiologia (HU et al., 2015), demografia (IACO; PALMA; POSA, 2015), ciências ambientais (MONTEIRO; MENEZES; SILVA, 2017) e hidrologia (HU et al., 2017; MARTÍNEZ; MELO; MELO, 2017).

Uma vez identificado que o fenômeno em estudo apresenta uma variabilidade espacial e temporal, os modelos que consideram a relação espaço-tempo têm apresentado melhores resultados por considerar em sua estrutura os componentes espaciais, temporais e espaço-temporais (CESARE; MYERS; POSA, 2001; SNEPVANGERS; HEUVELINK; HUISMAN, 2003; KILIBARDA et al., 2015; MONTEIRO; MENEZES; SILVA, 2017).

Na geoestatística espaço-temporal, os fenômenos indexados no espaço-tempo podem ser decompostos em dois componentes, o de tendência e o de resíduo estocástico. Os resíduos podem ser modelados por funções de covariâncias que englobam estruturas espaciais, temporais e da interação (IACO; MYERS; POSA, 2001; IACO; PALMA; POSA, 2015), inclusive funções que sejam capazes de modelar a anisotropia geométrica (SNEPVANGERS; HEUVELINK; HUISMAN, 2003; KILIBARDA et al., 2015). A grande dificuldade no ajuste destas funções de covariâncias consiste em determinar as estimativas relacionadas á grande quantidade de parâmetros presentes nestes modelos. Por outro lado, o componente de tendência é, na maioria dos trabalhos, ajustado por meio de uma regressão linear normal (GRÄLER et al., 2012; HU et al., 2017; KILIBARDA et al., 2014). Estudos recentes têm levado em consideração que a variável resposta pode apresentar um comportamento de

não-normalidade (MENEZES et al., 2015; MONTEIRO; MENEZES; SILVA, 2017), propondo então a utilização de um modelo linear generalizado.

Neste trabalho serão utilizados dois conjuntos de dados de precipitação pluviométrica. O primeiro conjunto de dados é referente ao Estado da Paraíba. O segundo conjunto refere-se a Bacia do Rio São Francisco (BHSF). Ambos os conjuntos de dados foram avaliados no espaço-tempo. Sabe-se que na meteorologia, a precipitação pluviométrica é um dos principais fenômenos de interesse, influenciando diretamente as condições hídricas de uma determinada região. O comportamento espaço-temporal das chuvas é de grande importância para os recursos hídricos de uma região e tem influência direta sobre as atividades humanas, tais como, pecuária, agricultura e comércio. Nos últimos anos têm surgido estudos que analisaram a distribuição da precipitação em função da interação espaço-tempo (KILIBARDA et al., 2015; MARTÍNEZ; MELO; MELO, 2017).

A região Nordeste do Brasil (NEB) é caracterizada pela distribuição irregular das chuvas no espaço e no tempo, ocasionando assim, alta variabilidade espacial e temporal da precipitação. Nessa região é comum a ocorrência de altos índices pluviométricos num local e nenhum registro de chuvas na sua circunvizinhança (LOPES et al., 2017). O Estado da Paraíba, localizado nessa região, apresenta as mesmas características, com períodos de estiagem dentro das estações chuvosas e comportamento distinto da precipitação entre as mesorregiões que constituem o estado (ALMEIDA; MEDEIROS, 2017). Dessa forma, esta região apresenta variabilidade na distribuição das chuvas tanto no espaço quanto no tempo que necessitam ser compreendidas.

Desde 2012, a Bacia do Rio São Francisco (BHSF) enfrenta condições hidrológicas adversas, com vazões e precipitações abaixo da média, reduzindo o volume armazenado em seus reservatórios. Entre 2014 e 2016 foram registradas as menores vazões naturais médias anuais no reservatório de Sobradinho desde que foram iniciadas as observações pelo Operador Nacional de Sistemas Elétrico (ONS) em 1931 (ANA, 2017). O Rio São Francisco é de suma importância para o semiárido do NEB no abastecimento de água para o consumo humano e irrigação, na geração de energia pelas hidrelétricas de Paulo Afonso e Sobradinho, turismo e pesca (ARAUJO; CELESTE, 2018; SILVEIRA et al., 2016). Desenvolver cenários do comportamento da precipitação na BHSF é um dos passos para se entender como as condições climáticas podem interferir na distribuição das chuvas, destacando as áreas com maior risco climático a serem inseridas em estratégias sociais de resiliência climática (CHEN et al., 2008).

Os dados de precipitação mensal na região nordeste apresentam, em sua maioria, uma distribuição com alto grau de assimetria positiva, valores não negativos e uma grande frequência de observações que são exatamente iguais a zero. Diante do exposto, propõe-se nesta tese modelar a dinâmica e dependência espaço-temporal da precipitação por meio de uma metodologia geoestatística (CRESSIE; WIKLE, 2015; IACO; PALMA; POSA, 2015). Primeiramente, é ajustado o componente de tendência, responsável pela variação em larga escala do fenômeno, em seguida é realizada uma análise espaço-temporal sobre os resíduos que são autocorrelacionados no espaço-tempo.

O foco desta pesquisa é considerar, no ajuste do componente de tendência, diferentes modelos probabilísticos, uma vez que a estimativa precisa da distribuição espaço-temporal da variável resposta requer uma distribuição de probabilidade adequada. Para isso são utilizados os modelos aditivos generalizados para posição, escala e forma (do inglês, *Generalized Additive Models for Location, Scale and Shape - GAMLSS*), pois estes permitem, no ajuste da regressão, modelar a superdispersão e excesso de zeros nos dados (STASINOPOULOS et al., 2017).

Assim, esta tese propõe uma nova abordagem (GAMLSS) para modelar o componente de tendência, presente na geoestatística espaço-temporal, o qual é responsável pela variação em larga escala do fenômeno em estudo. Esta abordagem é inovadora, pois até o presente momento, não encontrou-se na literatura trabalhos que utilizaram deste tipo de modelagem para o ajuste da tendência na geoestatística espaço-temporal.

A primeira parte da tese consiste em uma introdução, referencial teórico e considerações finais de toda a tese. A segunda parte consiste em três artigos. No primeiro artigo tem-se uma análise geoestatística espaço-temporal, em que para modelar a tendência é utilizada a regressão linear normal, tendo como covariáveis as coordenadas geográficas (longitude e latitude) e um índice temporal, o qual é descrito com detalhes na seção métodos deste artigo. A metodologia empregada neste artigo está bem detalhada na literatura da estatística espacial.

Como mencionado anteriormente, os dados de precipitação mensal do NEB, em geral, apresentam como aspectos marcantes a alta dispersão, excesso de zeros e uma distribuição assimétrica. Estas características podem acarretar dificuldades no modelagem do componente de tendência. Para contornar isto, surge a necessidade de utilizar novas metodologias no ajuste da tendência. Assim, no segundo artigo, para o ajuste deste componente é considerada uma abordagem inovadora, os modelos GAMLSS, em que é possível estar inserindo diferentes distribuições de probabilidade no ajuste da regressão. O objetivo deste artigo é de verificar se

com a regressão GAMLSS, resulta em modelos com melhores desempenhos para a krigagem com regressão espaço-temporal. Após removida a tendência, à dependência espaço-temporal dos resíduos é sugerido o modelo de covariância “product-sum” generalizado, o qual não leva em consideração o efeito de anisotropia. Para medir o desempenho da krigagem é realizada a validação cruzada e extraída diversas estatísticas para medir a qualidade do ajuste final.

No terceiro artigo é considerada uma área de estudo mais ampla (a BHSF). Os modelos GAMLSS são aqui comparados com a regressão linear normal. Os resíduos produzidos por este ajuste são modelados pela função de covariância “sum-metric”, a qual contempla o efeito da anisotropia geométrica. Para determinar o melhor ajuste, na krigagem com regressão, é proposta na validação cruzada utilizar um método inovador proposto por Bastos (2017).

A krigagem espaço-temporal com regressão permitirá reconstruir o padrão espaço-temporal da precipitação em toda a área de estudo. Assim, a metodologia proposta nesta tese permite a obtenção de uma estimativa da precipitação em toda a área de estudo com uma resolução espacial e temporal e pode ser facilmente implementada e aplicada em novos conjuntos de dados para outras regiões.

2 REFERENCIAL TEÓRICO

2.1 Modelo estocástico espaço-temporal

Ao modelar um fenômeno no espaço-tempo, obtém-se vantagens significativas ao considerar a existência de uma dinâmica no espaço e no tempo, sendo possível compreender a variabilidade espaço-temporal do processo. Uma vez considerada apenas a distribuição espacial em um dado momento, tem-se um processo meramente espacial, ou modelar um dado local ao longo do tempo, tem-se apenas um processo meramente temporal (MARTÍNEZ; MELO; MELO, 2017; HU et al., 2017; IACO; PALMA; POSA, 2015).

Uma localização espaço-temporal, em geral, é observada como um ponto em $\mathbb{R}^d \times \mathbb{R}$, com \mathbb{R}^d sendo o espaço euclidiano d -dimensional e \mathbb{R} a dimensão do tempo. Sabe-se que, do ponto de vista matemático $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$, entretanto, na prática não faz sentido considerar os aspectos espaciais e temporais da mesma maneira, uma vez que existem diferenças significativas entre os dois eixos de coordenadas. Em geral, no tempo os dados são observados apenas em uma direção (passado), com o objetivo de extrapolar seu valor para o futuro. No caso espacial, em geral, os dados são observados em duas direções e tem por objetivo interpolar para um local não amostrado. Adicionalmente, nem as escalas e nem as unidades das distâncias no tempo e no espaço são geralmente comparáveis (MONTERO; FERNANDEZ-AVILES; MATEU, 2015). A seguir é apresentada a caracterização de um processo estocástico espaço-temporal.

Seja um processo espaço-temporal $\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq \mathbb{R}^d \times \mathbb{R}\}$ em que, $\mathbf{s} \in \mathbb{R}^d$, em geral $d \leq 3$, representando as dimensões no espaço e $t \in \mathbb{R}$, representando apenas uma dimensão, o tempo (CRESSIE; WIKLE, 2015).

Seja a variável Z observada na coordenada espaço-tempo (s_i, t_j) , $i = 1, \dots, n$ e $j = 1, \dots, m$. Comumente, um dos principais objetivos na análise de dados espaço-temporais é obter uma predição para uma localização espacial s_0 em um dado momento no tempo t_0 , em que s_0 ou t_0 não foi observado, de acordo com o conjunto de dados observados no espaço-tempo (JOVEIN; HOSSEINI, 2017). Um outro aspecto, não menos importante, é sobre o conhecimento da dependência espaço-temporal de Z . Sabe-se que o conhecimento do grau de dependência espaço-temporal já serve como uma base para formular interpretações acerca da variação espacial, temporal e espaço-temporal, as quais se tornam essenciais para a otimização de projetos de monitoramento espaço-temporal (HEUVELINK et al., 2012; KILIBARDA et al., 2015).

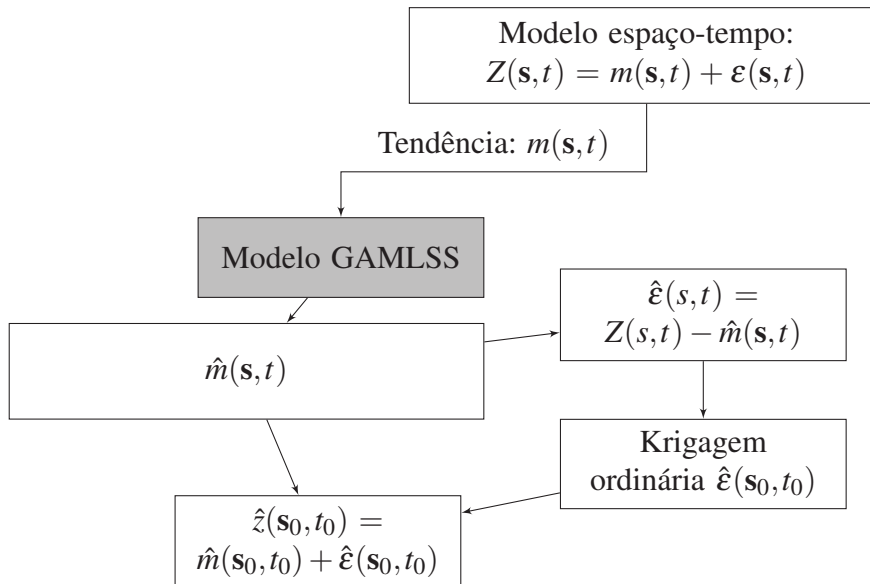
A variação espaço-temporal de Z pode ser estabelecida pela decomposição:

$$Z(\mathbf{s};t) = m(\mathbf{s};t) + \varepsilon(\mathbf{s};t), \quad (2.1)$$

em que o componente de tendência $m(\mathbf{s};t)$, que é não-constante no espaço e no tempo, representa a variação em larga escala, enquanto a componente residual $\varepsilon(\mathbf{s};t)$ descreve as flutuações aleatórias em pequena escala (HEUVELINK et al., 2012).

Nesta tese a modelagem sobre $Z(\mathbf{s};t)$ é realizada em dois estágios (Figura 2.1). No primeiro estágio é realizada a modelagem para o componente de tendência $m(\mathbf{s};t)$. No segundo estágio, os resíduos oriundos do primeiro estágio serão modelados utilizando uma abordagem geoestatística espaço-temporal, seguida de uma krigagem ordinária desses resíduos. Para uma predição final da variável Z será realizada a combinação do ajuste da tendência com o resíduo, a qual denomina-se krigagem com regressão. Na seção a seguir são apresentados detalhes teóricos da metodologia utilizada no ajuste do componente de tendência, dando destaque para os modelos GAMLSS.

Figura 2.1 – Fluxograma da krigagem com regressão com modelo GAMLSS



2.2 Componente de tendência

O componente de tendência $m(\mathbf{s},t)$, dado na Equação 2.1, é comumente modelado por regressão linear múltipla, via modelo Normal. Diversos trabalhos utilizaram este tipo de método para modelar o componente de tendência na geoestatística espaço-temporal. Dentre estes, destacam-se os trabalhos de Gräler et al. (2012), que analisaram um indicador da

qualidade do ar PM_{10} , medido diariamente, em função da altitude das estações e de um índice temporal para ajuste da periodicidade anual presente nos dados, $\cos(dia/365 \times 2\pi)$; Hengl et al. (2011), que em seu trabalho sobre previsão espaço-temporal das temperaturas diárias na Croácia, modelaram a tendência utilizando imagens MODIS para extrair informações da latitude, longitude, radiação solar, entre outras, e também criaram o índice temporal para modelar a flutuação sazonal da temperatura diária, $\cos(dia/365 \times 2\pi)$; Hu et al. (2015) ao estudarem a prevalência da esquistossomose no espaço-tempo notaram que esta variável estava significativamente relacionada com a distância de cada aldeia até o rio e com variáveis ambientais; Iaco, Palma e Posa (2015) utilizaram o fator região como tendência para a taxa de fertilidade das mulheres na França; Hu et al. (2017) ao estudarem a distribuição espaço-temporal mensal da precipitação, notaram que esta apresentava relação linear significativa com índice de vegetação obtida por imagens MODIS e também com o índice temporal, $\cos(mês/12 \times 2\pi)$. Entretanto, em todos estes trabalhos supracitados os autores não levaram em consideração de que a regressão linear Normal pode não ser a mais indicada para modelar a componente de tendência.

Assim, estudos recentes têm abordado, na modelagem do componente de tendência, estruturas de modelos não normais, utilizando os Modelos Lineares Generalizados (do inglês, *Generalized Linear Models* - GLM) com variável resposta Gama (MENEZES et al., 2015; MONTEIRO; MENEZES; SILVA, 2017). Em ambos os trabalhos, os autores modelaram a concentração de NO_2 em Portugal, a qual tinha um comportamento assimétrico com excesso de zeros. Assim, para contornar esta característica, eles propuseram a adição de uma constante nessa variável.

Contudo, uma vez assumido o modelo Gama, o domínio da variável Z tem que ser estritamente positivo, e este não inclui o valor zero. Dessa forma, surge a necessidade de utilizar uma modelagem mais completa, tais como os modelos GAMLSS (STASINOPOULOS et al., 2017).

2.2.1 Modelos GAMLSS

Na literatura surgiram algumas técnicas de regressão, as quais aceitavam na modelagem que fossem utilizadas diferentes distribuições de probabilidades para ajustar-se a variável resposta. Dentre essas técnicas, destacam-se os modelos GLM e os Modelos Aditivos Generalizados (do inglês, *Generalized Additive Models* - GAM). Entretanto, ambos os

modelos assumem que a distribuição a ser utilizada deve pertencer a família exponencial (RIGBY; STASINOPOULOS, 2005).

Uma outra importante classe são os modelos lineares mistos com efeitos aleatórios que podem ser utilizados em dados com amostragem espacial, hierárquica e longitudinal. Estes modelos assumem normalidade para a distribuição condicional da variável resposta dado os efeitos aleatórios e portanto, não podem modelar a assimetria e a curtose. Também surgiram outras técnicas que combinavam estas duas classes: modelo linear generalizado misto (do inglês, *Generalized Linear Mixed Model* - GLMM) e os Modelos Aditivos Generalizados Mistos (do inglês, *Generalized Additive Mixed Models* - GAMM). Embora mais flexíveis do que as classes GLM e GAM, também pressupõem uma distribuição condicional da família exponencial para a variável resposta e, à exceção da média, raramente permitem a modelagem de outros parâmetros da distribuição (RIGBY; STASINOPOULOS, 2005).

No trabalho de Rigby e Stasinopoulos (2005) apresentou-se uma classe geral denominada GAMLSS, em que não é requerida a suposição de que a variável pertença a família exponencial e é apresentada uma classe ampla de distribuições que podem ser utilizadas. Estes modelos permitem modelar todos os parâmetros de uma dada distribuição de probabilidade. Os parâmetros da distribuição condicional a variável resposta são modelados com funções paramétricas e/ou não paramétricas (suavização) e, inclusive, pode-se inserir no modelo termos de efeitos aleatórios.

Na regressão GAMLSS assume-se que as observações $\mathbf{z}^T = (z_1, z_2, \dots, z_n)$, $i = 1, 2, \dots, n$, são independentes, com função de probabilidade (ou densidade) $f(z_i|\boldsymbol{\theta}^i)$, condicionada a $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i})^T = (\mu_i, \sigma_i, \nu_i, \tau_i)^T$, um vetor com quatro parâmetros de uma dada distribuição. Os parâmetros μ, σ, ν, τ são classificados como os parâmetros de localização, escala, assimetria e curtose, respectivamente, sendo que os dois últimos representam a forma da distribuição (STASINOPOULOS et al., 2017). Rigby e Stasinopoulos (2005)

definiram os modelos GAMLSS da seguinte forma:

$$g_w(\boldsymbol{\theta}_w) = \boldsymbol{\eta}_w = \mathbf{X}_w \boldsymbol{\beta}_w + \sum_{j=1}^{J_w} \mathbf{Y}_{jw} \boldsymbol{\phi}_{jw}, \Rightarrow \left\{ \begin{array}{l} g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Y}_{j1} \boldsymbol{\phi}_{j1} \\ g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Y}_{j2} \boldsymbol{\phi}_{j2} \\ g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Y}_{j3} \boldsymbol{\phi}_{j3} \\ g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Y}_{j4} \boldsymbol{\phi}_{j4} \end{array} \right. \quad (2.2)$$

em que:

- i) $g_w(\cdot)$ é uma função de ligação monótona com $w = 1, 2, 3, 4$., e $\boldsymbol{\theta}_w$ está relacionado com as variáveis explicativas e os efeitos aleatórios;
- ii) $\boldsymbol{\theta}_w$ e $\boldsymbol{\eta}_w$ são vetores de tamanho n ;
- iii) $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T, \boldsymbol{\beta}_4^T)^T$, sendo $\boldsymbol{\beta}_w^T = (\beta_{1w}, \beta_{2w}, \dots, \beta_{j'_w})$ um vetor de parâmetros de tamanho J'_w ;
- iv) \mathbf{X}_w e \mathbf{Y}_{jw} são as matrizes do delineamento de efeitos fixos de tamanho $n \times J'_w$ e $n \times q_{ij}$, respectivamente;
- v) $\boldsymbol{\phi} = (\boldsymbol{\phi}_{11}^T, \dots, \boldsymbol{\phi}_{1J_1}^T, \dots, \boldsymbol{\phi}_{41}^T, \dots, \boldsymbol{\phi}_{4J_4}^T)^T$, em que $\boldsymbol{\phi}_{jw}$ é uma variável aleatória de dimensão q_{jw} , assumindo-se que $\boldsymbol{\phi}_{jw} \sim N_{q_{jw}}(\mathbf{0}, \mathbf{G}_{jw}^-)$, em que \mathbf{G}_{jw}^- é a inversa generalizada de uma matriz simétrica $\mathbf{G}_{jw} = \mathbf{G}_{jw}(\boldsymbol{\lambda}_{jw})$, a qual pode depender de um vetor de hiperparâmetros, $\boldsymbol{\lambda}_{jw}$. Se \mathbf{G}_{jw} é singular, então $\boldsymbol{\lambda}_{jw}$ tem uma função a priori imprópria, a qual é proporcional a $\exp\left(-\frac{1}{2} \boldsymbol{\lambda}_{jw}^T \mathbf{G}_{jw} \boldsymbol{\lambda}_{jw}\right)$.

No modelo (2.2) é possível modelar cada parâmetro da distribuição como uma função linear de variáveis explicativas e/ou como funções lineares de variáveis com efeitos aleatórios (STASINOPOULOS; RIGBY, 2007). No GAMLSS é possível a obtenção de vários submodelos. Para um modelo linear puramente paramétrico, tem-se que $\sum_{j=1}^{J_w} \mathbf{Y}_{jw} \boldsymbol{\phi}_{jw} = \mathbf{0}$, ou seja, não existem no modelo termos aditivos associados aos parâmetros da distribuição. Nesta tese será utilizada a formulação do GAMLSS que engloba funções de suavização.

Seja $\mathbf{Y}_{jw} = \mathbf{I}_n$, em que \mathbf{I}_n é a matriz identidade de ordem $n \times n$ e $\boldsymbol{\phi}_{jw} = \mathbf{v}_{jw} = v_{jw}(x_{jw})$. O modelo (2.2) pode ser reescrito na forma:

$$g_w(\boldsymbol{\theta}_w) = \boldsymbol{\eta}_w = \mathbf{X}_w \boldsymbol{\beta}_w + \sum_{j=1}^{J_w} v_{jw}(x_{jw}), \quad (2.3)$$

em que v_{jw} é uma função desconhecida da variável explicativa X_{jw} e $v_{jw}(x_{jw})$ é um vetor que avalia a função v_{jw} em x_{jw} . Na literatura, esse modelo é denominado de GAMLSS aditivo semi-paramétrico e pode conter termos paramétricos, não paramétricos e de efeitos aleatórios.

2.2.2 Estimação do modelo

No modelo 2.2 assume-se que $\boldsymbol{\phi}_{jw}$ tem distribuição a priori normal e independente, $\boldsymbol{\phi}_{jw} \sim N_{q_{jw}}(\mathbf{0}, \mathbf{G}_{jw}^-)$, sendo $\mathbf{G}_{jw} = \mathbf{G}_{jw}(\boldsymbol{\lambda}_{jw})$, em que \mathbf{G}_{jw}^- é a inversa generalizada de uma matriz simétrica \mathbf{G}_{jw} que depende de um vetor de hiperparâmetros $\boldsymbol{\lambda}_{jw}$. A suposição de independência entre diferentes vetores de efeitos aleatórios ($\boldsymbol{\lambda}_{jw}$) é de extrema importância dentro da estrutura do GAMLSS (RIGBY; STASINOPOULOS, 2005). No trabalho destes autores eles demonstraram que, utilizando de resultados bayesianos, o método da estimação máxima a posteriori para os vetores de parâmetros $\boldsymbol{\phi}_{jw}$, $\boldsymbol{\beta}_w$ e $\boldsymbol{\lambda}_{jw}$ é análogo ao método da máxima verossimilhança penalizada (do inglês, *Penalized Likelihood Estimation* - PLE).

Nesta tese o método PLE será utilizado no ajuste dos modelos para determinar os estimadores dos parâmetros $\boldsymbol{\phi}_{jw}$, $\boldsymbol{\beta}_w$ e $\boldsymbol{\lambda}_{jw}$ o qual é obtido pela seguinte expressão:

$$l_p = l - \frac{1}{2} \sum_{w=1}^p \sum_{j=1}^{J_k} \boldsymbol{\phi}_{jw}^T \mathbf{G}_{jw} \boldsymbol{\phi}_{jw}, \quad (2.4)$$

em que l_p é o logaritmo da função de verossimilhança penalizada e $l = \sum_{i=1}^n \ln\{f(z_i|\boldsymbol{\theta}^i)\} = \sum_{i=1}^n \ln\{f(z_i|\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i, \mathbf{v}_i, \boldsymbol{\sigma}_i)\}$ é o logaritmo da função de verossimilhança dos dados condicionados a $\boldsymbol{\theta}^i$, $\forall i = 1, 2, \dots, n$. A maximização da PLE, expressa na Equação 2.4, em relação aos parâmetros $\boldsymbol{\beta}_w$ e $\boldsymbol{\phi}_{jw}$, é realizada em dois passos que são descritos a seguir:

Passo 1: A primeira e segunda derivada da expressão 2.4 é obtida por meio do algoritmo Newton-Raphson. A r -ésima interação deste algoritmo é dada por:

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1p} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2p} \\ \vdots & \cdots & \cdots & \vdots \\ \mathbf{A}_{p1} & \mathbf{A}_{p2} & \cdots & \mathbf{A}_{pp} \end{pmatrix}^r \begin{pmatrix} \boldsymbol{\alpha}_1^{(r+1)} - \boldsymbol{\alpha}_1^{(r)} \\ \boldsymbol{\alpha}_2^{(r+1)} - \boldsymbol{\alpha}_2^{(r)} \\ \vdots \\ \boldsymbol{\alpha}_p^{(r+1)} - \boldsymbol{\alpha}_p^{(r)} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_p \end{pmatrix}^r,$$

em que $\boldsymbol{\alpha}_w^T = (\boldsymbol{\beta}_w^T, \boldsymbol{\phi}_{1w}^T, \boldsymbol{\phi}_{2w}^T, \dots, \boldsymbol{\phi}_{jkw}^T)$, $\mathbf{a}_w = \partial l_p / \partial \boldsymbol{\alpha}_w$ e $\mathbf{A}_{ws} = \partial^2 l_p / \partial \boldsymbol{\alpha}_w \partial \boldsymbol{\alpha}_s^T$, sendo $s = (1, 2, 3, 4)$. A matriz \mathbf{A}_{ws} é dada por:

$$\begin{pmatrix} \mathbf{X}_w^T \mathbf{W}_{ws} \mathbf{X}_s & \mathbf{X}_w^T \mathbf{W}_{ws} \mathbf{Y}_{1s} & \cdots & \mathbf{X}_w^T \mathbf{W}_{ws} \mathbf{Y}_{J_s s} \\ \mathbf{Z}_{1w}^T \mathbf{W}_{ws} \mathbf{X}_s & \mathbf{Z}_{1w}^T \mathbf{W}_{ws} \mathbf{Y}_{1s} + \mathbf{G}_{1w} \text{(se } s = k) & \cdots & \mathbf{Y}_{1w}^T \mathbf{W}_{ws} \mathbf{Y}_{J_s s} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{Z}_{J_w w}^T \mathbf{W}_{ws} \mathbf{X}_s & \mathbf{Z}_{J_w w}^T \mathbf{W}_{ws} \mathbf{Y}_{1s} & \cdots & \mathbf{Z}_{J_w w}^T \mathbf{W}_{ws} \mathbf{Y}_{J_s s} + \mathbf{G}_{J_w w} \text{(se } s = k) \end{pmatrix},$$

em que a matriz $\mathbf{W}_{ws} = -\partial^2 l / \partial \boldsymbol{\eta}_w \partial \boldsymbol{\eta}_s^T$.

O vetor $\mathbf{a}_w^{(r)}$ é escrito da seguinte forma:

$$\mathbf{a}_w^{(r)} = \begin{pmatrix} \mathbf{X}_w^T \mathbf{u}_w^{(r)} \\ \mathbf{Y}_{1w}^T \mathbf{u}_w^{(r)} - \mathbf{G}_{1w} \boldsymbol{\phi}_{1w}^{(r)} \\ \vdots \\ \mathbf{Y}_{J_w w}^T \mathbf{u}_w^{(r)} - \mathbf{G}_{J_w w} \boldsymbol{\phi}_{J_w w}^{(r)} \end{pmatrix},$$

em que o vetor $\mathbf{u}_w = \partial l / \partial \boldsymbol{\eta}_w$.

Passo 2: A atualização do vetor $\boldsymbol{\phi}_{jw}$ é obtida da seguinte maneira:

$$\mathbf{G}_{jw} \left(\boldsymbol{\phi}_{jw}^{(r+1)} - \boldsymbol{\phi}_{jw}^{(r)} \right) + \mathbf{Y}_{jw}^T \sum_{s=1}^4 \mathbf{W}_{ws}^{(r)} \left(\boldsymbol{\eta}_s^{(r+1)} - \boldsymbol{\eta}_s^{(r)} \right) = \mathbf{Y}_{jw}^T \mathbf{u}_w^{(r)} - \mathbf{G}_{jw}^T \boldsymbol{\phi}_{jw}^{(r)}$$

A atualização $\boldsymbol{\phi}_{jw}$ pode ser reescrita como:

$$\mathbf{Y}_{jw} \boldsymbol{\phi}_{jw}^{(r+1)} = \mathbf{S}_{jw}^{(r)} \boldsymbol{\varepsilon}_{jw}^{(r)},$$

em que \mathbf{S}_{jw} é a matriz de encolhimento (alisamento) dada por:

$$\mathbf{S}_{jw} = \mathbf{Y}_{jw} (\mathbf{Y}_{jw}^T \mathbf{W}_{ww} \mathbf{Y}_{jw} + \mathbf{G}_{jw})^{-1} \mathbf{Y}_{jw}^T \mathbf{W}_{ww}. \quad (2.5)$$

A matriz \mathbf{W}_{ww} é a matriz de pesos iterativos. Diferentes formas de \mathbf{Y}_{jw} e \mathbf{G}_{jw} correspondem a diferentes tipos de termos aditivos para $\boldsymbol{\eta}_w$. Para os termos de suavização *spline* cúbico tem-se que $\boldsymbol{\phi}_{jw} = \mathbf{v}_{jw}$, $\mathbf{Z}_{jw} = \mathbf{I}_n$ e $\mathbf{G}_{jw} = \lambda_{jw} \mathbf{K}_{jw}$, sendo \mathbf{K}_{jw} uma matriz estruturada.

Para mais detalhes do algoritmo de estimação no GAMLSS consultar o Apêndice C do trabalho de Rigby e Stasinopoulos (2005).

2.2.3 Termos Aditivos

Os componentes aditivos no modelo dado na expressão (2.2) podem modelar diversos termos, como de suavização e de efeitos aleatórios, bem como termos úteis em séries temporais, tais como passeios aleatórios (RIGBY; STASINOPOULOS, 2005). Nos artigos 2 e 3 que compõem esta tese foram utilizadas, no ajuste da regressão GAMLSS, as funções *splines*. A seguir são detalhados os termos com funções de suavização *spline* cúbica (cs) e *spline* penalizada (p-*spline*).

2.2.3.1 Splines cúbicos

Para os termos de *splines* cúbicos, assume-se que no modelo dado na expressão (2.3) as funções $v(t)$ são contínuas, duas vezes diferenciáveis e que a maximização da função de verossimilhança está sujeita aos termos de penalização da forma: $\lambda \int_{-\infty}^{+\infty} v''(t)^2 dt$. As funções de maximização $v(t)$ são todas *splines* cúbicas e são expressas como combinação linear de suas funções bases $B_i(t)$ (RIGBY; STASINOPOULOS, 2005). Dessa forma, $v(t) = \sum_{i=1}^n \delta_i B_i(t)$.

Considere ainda que $\mathbf{v} = v(\mathbf{x})$ o vetor de avaliações da função $v(t)$ nos valores \mathbf{x} da variável explicativa \mathbf{X} . Seja \mathbf{N} uma matriz não-singular de dimensão $n \times n$ e que em cada uma de suas n colunas contém os valores das funções avaliadas, $B_i(t)$, em \mathbf{x} , com $i = 1, 2, \dots, n$. Dessa forma, \mathbf{v} pode ser então expresso por meio do vetor de coeficientes denotado por $\boldsymbol{\delta}$, cujo resultado é obtido pela combinação linear $\mathbf{v} = \mathbf{N}\boldsymbol{\delta}$.

Seja $\mathbf{\Omega}$ uma matriz de ordem $n \times n$ dos produtos internos das segundas derivadas das funções base *splines* cúbicas nos r -ésimos e s -ésimos registros, expressa da forma:

$$\mathbf{\Omega}_{rs} = \int B_r''(t)B_s''(t)dt.$$

A penalidade é dada pela forma quadrática:

$$Q(\mathbf{v}) = \lambda \int_{-\infty}^{+\infty} v''(t)^2 dt = \lambda \boldsymbol{\delta}^T \mathbf{\Omega} \boldsymbol{\delta} = \lambda \mathbf{v}^T \mathbf{N}^{-T} \mathbf{\Omega} \mathbf{N}^{-1} \mathbf{v} = \lambda \mathbf{v}^T \mathbf{K} \mathbf{v}, \quad (2.6)$$

em que $\mathbf{K} = \mathbf{N}^{-T} \mathbf{\Omega} \mathbf{N}^{-1}$ é uma matriz de penalidade conhecida que depende apenas dos valores do vetor explicativo \mathbf{x} . Para ajuste da regressão GAMLSS com efeitos aleatórios para o Modelo (2.2) é necessário que $\boldsymbol{\phi} = \mathbf{v}$, $\mathbf{Y} = \mathbf{I}_n$ e $\mathbf{G} = \lambda \mathbf{K}$, de tal forma que $\mathbf{v} \sim N_n(\mathbf{0}, \lambda^{-1} \mathbf{K}^-)$, sendo a matriz \mathbf{K}^- uma inversa generalizada da matriz \mathbf{K} .

2.2.3.2 *Splines* de penalização

Suavizadores em que a quantidade de funções bases é menor que a quantidade de observações, mas seus coeficientes de regressão são penalizados, são denominados *splines* penalizados ou *P-splines*. Para garantir a flexibilidade nas suavizações das curvas tem-se a imposição de penalidades nos parâmetros $\boldsymbol{\phi}$ das funções bases *B-splines* com o intuito de assegurar a suavidade do ajuste (STASINOPOULOS et al., 2017). Resumidamente, assume-se que $\mathbf{D}_r \boldsymbol{\phi} \sim N_{n-r}(0, \lambda^{-1} \mathbf{I})$ em que \mathbf{D}_r é uma matriz de ordem $(q-r) \times q$ que fornece r -ésimas diferenças do vetor q -dimensional $\boldsymbol{\phi}$ (RIGBY; STASINOPOULOS, 2005).

2.3 Geoestatística espaço-temporal

Após especificado o componente de tendência e obtidas as estimativas dos parâmetros, pode-se então subtrair esta de Z , restando então o resíduo estocástico espaço-tempo ε . Uma das vantagens de inserir componentes na tendência está no fato de que estas retiram dos resíduos comportamentos que levam a grande variação do erro à krigagem, uma vez que esta contabiliza as incertezas nos coeficientes de tendência (HEUVELINK; GRIFFITH, 2010).

Definição 1 O processo $\varepsilon(\mathbf{s};t)$ é dito ser estritamente estacionário se sua distribuição de probabilidade é invariante a translação. Assim, para quaisquer dois vetores, \mathbf{h}_s e h_t , tem-se

que

$$\varepsilon(\mathbf{s}_1; t_1), \varepsilon(\mathbf{s}_2; t_2), \dots, \varepsilon(\mathbf{s}_n; t_n)$$

e

$$\varepsilon(\mathbf{s}_1 + \mathbf{h}_s; t_1 + h_t), \varepsilon(\mathbf{s}_2 + \mathbf{h}_s; t_2 + h_t), \dots, \varepsilon(\mathbf{s}_n + \mathbf{h}_s; t_n + h_t)$$

possuem a mesma função de distribuição multivariada.

Definição 2 O processo $\varepsilon(\mathbf{s}; t)$ é estacionário de segunda ordem se:

$$E[\varepsilon(\mathbf{s}; t)] = 0$$

e

$$C[\varepsilon(\mathbf{s}_i; t_i), \varepsilon(\mathbf{s}_j; t_j)] = C(\mathbf{h}_s, h_t).$$

Dessa forma, a função de covariância dependerá apenas de \mathbf{h}_s e h_t .

Definição 3 O processo $\varepsilon(\mathbf{s}; t)$ é intrinsecamente estacionário se o processo de incremento $\varepsilon(\mathbf{s} + \mathbf{h}_s, t + h_t) - \varepsilon(\mathbf{s}, t)$ é estacionário de segunda ordem no espaço-tempo para qualquer (\mathbf{h}_s, h_t) fixo.

Definição 4 Se o processo espaço-temporal $\varepsilon(\mathbf{s}; t)$ tem uma função de covariância estacionária para o espaço e para o tempo, então tem-se uma função de covariância estacionária. Assim, a função de covariância é expressa da forma:

$$C[\varepsilon(\mathbf{s}_i; t_i), \varepsilon(\mathbf{s}_j; t_j)] = C(\mathbf{h}_s, h_t),$$

em que $\mathbf{h}_s = \mathbf{s}_i - \mathbf{s}_j$ e $h_t = t_i - t_j$ correspondem as distâncias no espaço e no tempo, respectivamente.

2.3.1 Variograma espaço-temporal

Por definição o processo $\varepsilon(\cdot; \cdot)$ é dito ser intrinsecamente estacionário, se tem esperança constante e seu variograma é estacionário. Se $\varepsilon(\cdot; \cdot)$ for estacionário de segunda ordem com função de covariâncias estacionárias, $C(\mathbf{h}_s, h_t)$, então, como no ajuste puramente espacial, $\varepsilon(\cdot, \cdot)$ é intrinsecamente estacionário com variograma dado por:

$$\gamma(\mathbf{h}_s, h_t) = C(\mathbf{0}, 0) - C(\mathbf{h}_s, h_t), \quad \mathbf{h}_s \in \mathbb{R}^d, \quad h_t \in \mathbb{R}. \quad (2.7)$$

Entretanto, a estimativa do variograma espaço-temporal engloba características que devem ser levadas em consideração, tais como, escalas distintas no espaço e no tempo e, princípios de causalidades diferentes (HENGL et al., 2011). Assim, o variograma espaço-temporal é definido como:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \frac{1}{2} E [\varepsilon(\mathbf{s}, t) - \varepsilon(\mathbf{s} + \mathbf{h}_s, t + h_t)]^2, \quad (2.8)$$

em que $\mathbf{h}_s = \mathbf{s} - \mathbf{s}'$ e $h_t = t - t'$, para quaisquer (\mathbf{s}, t) e (\mathbf{s}', t') no domínio espaço-temporal.

As covariâncias são usualmente descritas usando um variograma espaço-temporal $\gamma_{st}(\cdot, \cdot)$, que mede a diferença média entre dados separados no domínio espaço-temporal usando o vetor distâncias (MONTERO; FERNANDEZ-AVILES; MATEU, 2015). A estimativa desse variograma obtido pelo método dos momentos é dada por:

$$\hat{\gamma}_{st}(\mathbf{r}_s, r_t) = \frac{1}{2|L(\mathbf{r}_s, r_t)|} \sum_{L(\mathbf{r}_s, r_t)} [\varepsilon(\mathbf{s} + \mathbf{h}_s, t + h_t) - \varepsilon(\mathbf{s}, t)]^2, \quad (2.9)$$

em que $|L(\mathbf{r}_s, r_t)|$ é a cardinalidade para o conjunto $L(\mathbf{r}_s, r_t)$. Nota-se na Equação 2.9, que o variograma empírico depende apenas das distâncias espaciais e temporais e isso pode não resultar em um variograma válido. Dessa forma, para obter estimativas de γ_{st} em qualquer defasagem arbitrária, o variograma empírico deve ser suavizado. Assim, para solucionar essa limitação encontrada no variograma empírico, deve-se usar um modelo teórico, $\gamma_{st}(\mathbf{h}_s, h_t, \boldsymbol{\omega})$, que ajuste, o melhor possível, a nuvem de pontos do variograma amostral apresentado na Equação 2.9. O vetor $\boldsymbol{\omega}$ contém todos os parâmetros desconhecidos a serem estimados.

2.4 Modelos de covariâncias

A noção intuitiva de que os valores tendem a ser fracamente correlacionados quando estão “distantes” pode ser capturada usando funções de covariâncias que tendem a zero à medida que suas coordenadas espaço-temporais se afastam. Nos últimos anos têm surgindo algumas classes de modelos de covariâncias espaço-tempo (IACO; PALMA; POSA, 2015; KILIBARDA et al., 2015).

Uma função de covariâncias é uma função não negativa definida. Portanto, investigar funções não- negativas definidas no espaço e no tempo tem sido um esforço importante nos

estudos espaço-temporais (IACO; MYERS; POSA, 2001). A seguir são apresentadas algumas definições estabelecidas no trabalho desses autores.

Definição 5 A função $\{f(u, v) : u, v \in D\}$ definida em $D \times D$ é dita ser não-negativa-definida, se para qualquer número complexo $\{a_i : i = 1, \dots, m\}$, qualquer $\{u_i : i = 1, \dots, m\}$ em D e qualquer inteiro m , tem-se:

$$\sum_{i=1}^m \sum_{j=1}^m a_i \bar{a}_j f(u_i, u_j) \geq 0, \quad (2.10)$$

em que \bar{a} denota o conjugado complexo de a .

A estacionariedade forte de $\varepsilon(\cdot, \cdot)$ corresponde à equivalência das duas medidas de probabilidade que definem os processos aleatórios $\varepsilon(\cdot, \cdot)$ e $\varepsilon(\cdot + \mathbf{h}_s, \cdot + h_t)$, respectivamente, $\forall \mathbf{h}_s \in \mathbb{R}^d$ e $\forall h_t \in \mathbb{R}$. Após reescrever o resultado da definição 5 em termos de $C(\cdot; \cdot)$, observa-se que a propriedade definida para uma função de covariâncias espaço-temporal estacionária $C(\cdot; \cdot)$ sobre $\mathbb{R}^d \times \mathbb{R}$ torna-se:

$$\sum_{i=1}^m \sum_{j=1}^m a_i \bar{a}_j C(\mathbf{h}_{s_i} - \mathbf{h}_{s_j}; h_{t_i} - h_{t_j}) \geq 0. \quad (2.11)$$

Note que, $C(\mathbf{h}_s; 0)$, com $\mathbf{h}_s \in \mathbb{R}^d$, corresponde a uma função de covariâncias espacial estacionária e que $C(0; h_t)$, $h_t \in \mathbb{R}$ corresponde a uma função de covariâncias temporal estacionária. A função de correlação espaço-temporal estacionária associada a $C(\cdot; \cdot)$ é:

$$\rho(\mathbf{h}_s; h_t) \equiv \frac{C(\mathbf{h}_s, h_t)}{C(\mathbf{0}, 0)}, \quad \mathbf{h}_s \in \mathbb{R}^d, h_t \in \mathbb{R}. \quad (2.12)$$

É possível que a estacionariedade da função de covariâncias seja considerada separadamente para o espaço e o tempo.

No caso espacial, para modelar a estrutura de dependência espacial, a discussão se concentra na escolha da função de variograma. Entretanto, na modelagem espaço-temporal a ferramenta mais comum é a obtenção da função de covariância (MONTERO; FERNANDEZ-AVILES; MATEU, 2015). Por esta razão, serão apresentados a seguir modelos de covariâncias válidos e que estão sendo utilizados em fenômenos indexados no espaço-tempo. Na abordagem espaço-temporal existem basicamente duas classes gerais de modelos de covariância: os separáveis e os não separáveis.

Os modelos separáveis pressupõem que os processos espaciais e temporais são não correlacionados e que também os variogramas são compostos de modelos puramente espacial e puramente temporal. Como exemplo, a função de covariância espaço-temporal pode ser expressa como o produto entre os componentes puramente espacial e puramente temporal, sendo a maior desvantagem desse tipo de modelo, a não incorporação do componente de interação espaço-temporal (HUANG et al., 2007; MONTERO; FERNANDEZ-AVILES; MATEU, 2015). Devido à sua facilidade de definição e dos cálculos matriciais na análise estatística espacial serem muito mais simples, os modelos de covariâncias separáveis dominaram a literatura até o início do século XXI. Entretanto, eles têm propriedades muito particulares que raramente são vistas em estudos empíricos de dependência espaço-temporal. Por outro lado, os modelos não separáveis assumem que o fenômeno espaço-temporal esteja correlacionado, ou seja, considera em sua estrutura interação espaço-tempo (IACO; MYERS; POSA, 2001; SHERMAN, 2010; CRESSIE; WIKLE, 2015). A seguir são apresentados alguns modelos separáveis e não separáveis.

2.4.1 Modelo soma

A função de covariância no modelo soma, introduzido por Rouhani e Hall (1989), é do tipo separável e representa a covariância espaço-temporal como a soma dos modelos puramente espacial e puramente temporal. O modelo é escrito na forma:

$$C[(\mathbf{s}_i, t_i)(\mathbf{s}_j, t_j)] = C_s(\mathbf{s}_i, \mathbf{s}_j) + C_t(t_i, t_j), \quad (\mathbf{s}_i, \mathbf{s}_j) \in D \subseteq \mathbb{R}^d, \forall (t_i, t_j) \in T \subseteq \mathbb{R},$$

sendo C_s uma função de covariância espacial definida no \mathbb{R}^d e C_t uma função de covariância temporal definida em \mathbb{R} . Considerando que estas duas funções de covariâncias são estacionárias, o modelo soma fica expresso por:

$$C(\mathbf{h}_s, h_t) = C_s(\mathbf{h}_s) + C_t(h_t), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}.$$

O variograma para esta classe de modelo separável, admitindo estacionariedade, é dado por:

$$\gamma(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s) + \gamma_t(h_t), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}.$$

O modelo soma, assim que introduzido na literatura, foi muito utilizado em diversos trabalhos por apresentar algumas facilidades, em especial, no aspecto da implementação computacional. Entretanto, surgiram diversas críticas a respeito desse modelo, sendo que uma delas era que a soma de dois modelos de covariâncias, um espacial e outro temporal, não resultava em uma função de covariância positiva definida e sim em uma função de covariância positiva semi-definida, fazendo com que a matriz de covariâncias nas equações de krigagem possa não ser inversível para certos conjuntos de localizações (DIMITRAKOPOULOS; LUO, 1994).

2.4.2 Modelo métrico

A função de covariância no modelo métrico, introduzida por Dimitrakopoulos e Luo (1994), é expressa da forma:

$$C(\mathbf{h}_s, h_t) = C(|\mathbf{h}_s| + c|h_t|), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}, c > 0, \quad (2.13)$$

em que $|\mathbf{h}_s| + c|h_t|$ é a distância sobre $\mathbb{R}^d \times \mathbb{R}$ e c é uma constante positiva. Considerando como distância $|c_1^2 \mathbf{h}_s|^2 + c_2^2 |h_t|^2$ ao invés de $|\mathbf{h}_s| + c|h_t|$ que em termos topológicos é equivalente ao acima (MONTERO; FERNANDEZ-AVILES; MATEU, 2015), o modelo métrico é reescrito da forma:

$$C(\mathbf{h}_s, h_t) = C(c_1^2 |\mathbf{h}_s|^2 + c_2^2 |h_t|^2), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}. \quad (2.14)$$

A hipótese sobre esse modelo é que a função de covariância espaço-temporal possui uma estrutura uniforme em todo o domínio (por exemplo, exponencial, esférico e gaussiano), resultando que o modelo de variograma para o domínio espacial é o mesmo para o domínio temporal e que possuem o mesmo patamar, mesmo apresentando alcances diferentes (MONTERO; FERNANDEZ-AVILES; MATEU, 2015).

Tem-se que a constante c na Equação (2.13) e as constantes c_1 e c_2 na Equação (2.14) definindo as métricas espaço-temporais representam os fatores anisotrópicos (anisotropia geométrica ou elíptica) entre as distâncias espaciais e as defasagens temporais, os quais determinam os diferentes intervalos espaciais e temporais (DIMITRAKOPOULOS; LUO,

1994). Para a Equação (2.14) o variograma é expresso da forma:

$$\gamma(\mathbf{h}_s, h_t) = \gamma(c_1^2 \|\mathbf{h}_s\|^2 + c_2^2 |h_t|^2), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}.$$

2.4.3 Modelo “product-sum”

O modelo de covariância produto, introduzido por Cesare, Myers e Posa (2001), é escrito da forma: $C(\mathbf{h}_s, h_t) = C_s(\mathbf{h}_s)C_t(h_t)$.

Esses autores realizaram um trabalho que enfatizaram algumas funções de covariâncias, em especial, as classes de modelos “product-sum”, relatando que os coeficientes dessa classe incorporam os parâmetros dos patamares com relação ao espaço e tempo. Adicionalmente, estes autores comparam os modelos dessa classe com outras estruturas de covariâncias. De acordo com estes autores o modelo de covariância “product-sum” é dado por:

$$C(\mathbf{h}_s, h_t) = k_1 C_s(\mathbf{h}_s)C_t(h_t) + k_2 C_s(\mathbf{h}_s) + k_3 C_t(h_t), \quad (2.15)$$

em que C_s e C_t são funções de covariância espacial e temporal, respectivamente. Com as restrições $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$.

Quando $\varepsilon(\mathbf{s}, t)$ é estacionário de segunda ordem, tem-se que:

$$\text{i) } \gamma_s(\mathbf{h}_s) = C_s(0) - C_s(\mathbf{h}_s) \Rightarrow C_s(\mathbf{h}_s) = C_s(0) - \gamma_s(\mathbf{h}_s).$$

$$\text{ii) } \gamma_t(h_t) = C_t(0) - C_t(h_t) \Rightarrow C_t(h_t) = C_t(0) - \gamma_t(h_t).$$

Substituindo na Equação 2.15:

$$\begin{aligned} C(\mathbf{h}_s, h_t) &= k_1 C_s(\mathbf{h}_s)C_t(h_t) + k_2 C_s(\mathbf{h}_s) + k_3 C_t(h_t) \\ &= -\gamma_s(\mathbf{h}_s) [k_1 C_t(0) + k_2] - \gamma_t(h_t) [k_1 C_t(0) + k_3] + \\ &\quad + k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) + C(\mathbf{0}, 0). \end{aligned} \quad (2.16)$$

Na Equação 2.16, aplicando $\gamma(\mathbf{h}_s, u) = C(\mathbf{0}, 0) - C(\mathbf{h}_s, h_t)$, tem-se o variograma espaço-temporal:

$$\gamma(\mathbf{h}_s, h_t) = (k_2 + k_1 C_t(0)) \gamma_s(\mathbf{h}_s) + (k_3 + k_1 C_s(0)) \gamma_t(h_t) - k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t). \quad (2.17)$$

Na Equação 2.17, $C_s(\mathbf{0})$ e $C_t(0)$ são os efeitos da contribuição espacial e temporal, respectivamente. É suficiente assumir a estacionariedade de segunda ordem para garantir que estes variogramas tenham um patamar (MONTERO; FERNANDEZ-AVILES; MATEU, 2015).

O modelo de “product-sum” dá origem às seguintes relações:

$$\gamma(\mathbf{h}_s, 0) = \gamma_s(\mathbf{h}_s)(k_2 + k_1 C_t(0)) = k_s \gamma_s(\mathbf{h}_s) \quad (2.18)$$

e

$$\gamma(0, h_t) = \gamma_t(h_t)(k_3 + k_1 C_s(\mathbf{0})) = k_t \gamma_t(h_t). \quad (2.19)$$

As Equações 2.18 e 2.19 são os variogramas marginais espacial e temporal, respectivamente. Por definição, tem-se que, $\gamma_{st}(0, 0) = \gamma_s(0) = \gamma_t(0) = 0$. Pela Equação 2.17 tem-se que:

$$\begin{aligned} \gamma_{st}(\mathbf{h}_s, 0) &= [k_2 + k_1 C_t(0)] \gamma_s(\mathbf{h}_s) \\ &= k_s \gamma_s(\mathbf{h}_s) \end{aligned} \quad (2.20)$$

$$\begin{aligned} \gamma_{st}(0, h_t) &= [k_3 + k_1 C_s(\mathbf{0})] \gamma_t(h_t), \\ &= k_t \gamma_t(h_t), \end{aligned} \quad (2.21)$$

em que k_s e k_t podem ser vistos como coeficientes de proporcionalidade entre os variogramas $\gamma_s(h, 0)$ e $\gamma_t(\mathbf{0}, h_t)$ e os variogramas marginais $\gamma_s(h_s)$ e $\gamma_t(h_t)$.

Esses dois relacionamentos podem ser simplificados impondo três restrições:

$$k_2 + k_1 C_t(0) = 1 \quad (2.22)$$

$$k_3 + k_1 C_s(\mathbf{0}) = 1 \quad (2.23)$$

$$k_1 + k_2 + k_3 = 1 \quad (2.24)$$

o que facilita a estimativa e o modelo $\gamma(\mathbf{h}_s, 0)$ e $\gamma(0, h_t)$ utilizando $\gamma_s(\mathbf{h}_s)$ e $\gamma_t(h_t)$, isto é, determinando k_1 , k_2 e k_3 .

2.4.4 Modelo “product-sum” generalizado

De maneira semelhante, Iaco, Myers e Posa (2001) estabeleceram uma generalização para o modelo de covariância “product-sum”. De acordo com estes autores as restrições impostas ao modelo “product-sum” (Equações 2.22, 2.23 e 2.24) podem ser simplificadas em um único parâmetro. Assim, eles mostram que os três coeficientes do modelo podem ser escritos em termos dos patamares e dos parâmetros k_s e k_t , os quais podem ser combinados, formando um único parâmetro k . A expressão do variograma espaço-temporal, $\gamma(\mathbf{h}_s, h_t)$, pode ser então simplificada usando as expressões 2.18 e 2.19:

$$\begin{aligned}
 \gamma(\mathbf{h}_s, h_t) &= (k_2 + k_1 C_t(0)) \gamma_s(\mathbf{h}_s) + (k_3 + k_1 C_s(0)) \gamma_t(h_t) - k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) \\
 &= k_s \gamma_s(\mathbf{h}_s) + k_t \gamma_t(h_t) - k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) \\
 &= \gamma_s(\mathbf{h}_s, 0) + \gamma_t(0, h_t) - k_1 \frac{\gamma_s(\mathbf{h}_s)}{k_s} \frac{\gamma_t(h_t)}{k_t} \\
 &= \gamma_s(\mathbf{h}_s, 0) + \gamma_t(0, h_t) - k \gamma_s(\mathbf{h}_s) \gamma_t(h_t).
 \end{aligned} \tag{2.25}$$

Na Equação 2.25, $k = \frac{k_1}{k_s k_t}$, sendo que k pode ser expresso como:

$$k = \frac{\frac{k_t C_t(0) + k_s C_s(0) - C(0,0)}{C_s(0) C_t(0)}}{k_s k_t} = \frac{k_t C_t(0) + k_s C_s(0) - C(0,0)}{[k_s C_s(0)] [k_t C_t(0)]}. \tag{2.26}$$

O comportamento assintótico dos variogramas $\gamma(h, u)$, $\gamma(h, 0)$ e $\gamma(0, u)$ sugerem que eles não possuem os mesmos patamares. Esse resultado pode ser comprovado pelo Teorema 1.

Teorema 1 (IACO; MYERS; POSA, 2001) *Seja Z um campo aleatório espaço-temporal estacionário de segunda ordem. Suponha que a covariância espaço-temporal C_{st} tem a forma dada pela Equação 2.15 e supondo que C_{st} é contínuo no espaço-tempo. Com base nas equações (2.20) e (2.21) tem-se os seguintes resultados:*

$$\begin{aligned}
 \lim_{(\mathbf{h}_s, h_t) \rightarrow \infty} \gamma_{st}(\mathbf{h}_s, h_t) &= C_{st}(0,0), \\
 \lim_{\mathbf{h}_s \rightarrow \infty} \gamma_{st}(\mathbf{h}_s, 0) &= k_s C_s(0), \\
 \lim_{h_t \rightarrow \infty} \gamma_{st}(0, h_t) &= k_t C_t(0).
 \end{aligned}$$

Utilizando as Equações (2.17), (2.20) e (2.21) a expressão para $\gamma_{st}(\mathbf{h}_s, h_t)$ pode ser simplificada da forma:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_{st}(\mathbf{h}_s, 0) + \gamma_{st}(0, h_t) - k\gamma_{st}(\mathbf{h}_s, 0)\gamma_{st}(0, h_t), \quad (2.27)$$

em que $k = \frac{k_1}{k_s k_t}$. O Teorema 2 estabelece as condições necessárias e suficientes para os limites de k .

Teorema 2 (IACO; MYERS; POSA, 2001) *Seja Z um campo aleatório espaço-temporal estacionário de segunda ordem. Suponha que a covariância espaço-temporal $C(\mathbf{h}_s, h_t) = k_1 C_s(\mathbf{h}_s)C_t(h_t) + k_2 C_s(\mathbf{h}_s) + k_3 C_t(h_t)$ e supondo que $C(\mathbf{h}_s, h_t)$ é contínuo no domínio espaço-temporal. Supondo que o parâmetro k assume o valor $k = \frac{k_1}{k_s k_t}$ no variograma espaço-temporal $\gamma(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s, 0) + \gamma_t(0, h_t) - k\gamma_s(\mathbf{h}_s, 0)\gamma_t(0, h_t)$. Então $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$ se, e somente se, k satisfaz a condição:*

$$0 < k \leq \frac{1}{\max\{sill[\gamma(h, 0)]; sill[\gamma(0, u)]\}}. \quad (2.28)$$

O modelo de covariância produto-soma generalizado são mais flexíveis que o modelo produto quando o intuito é modelar e estimar estruturas de dependência espaço-temporal (MONTERO; FERNANDEZ-AVILES; MATEU, 2015).

2.4.5 Modelo “sum-metric”

Outra classe importante de modelos de covariância não separáveis é a do modelo “sum-metric”, o qual consiste na combinação dos modelos soma e métrico. Este por sua vez apresentava-se como candidato para solucionar as restrições dos dois modelos. A função de covariância espaço-temporal é dada por:

$$C(\mathbf{h}_s, h_t) = C_s(\mathbf{h}_s) + C_t(h_t) + C_{st}(h_{st}), \quad (\mathbf{h}_s, h_t) \in \mathbb{R}^d \times \mathbb{R}, \quad (2.29)$$

em que a defasagem h_{st} é estimada inserindo um parâmetro (α) relacionado à razão de uma anisotropia geométrica, dada por:

$$h_{st} = \sqrt{||\mathbf{h}_s||^2 + (\alpha|h_t|)^2}.$$

Para a Equação (2.29), o variograma é expresso da forma:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s) + \gamma_t(h_t) + \gamma_j(h_{st}). \quad (2.30)$$

Na Equação (2.30) γ_s , γ_t e γ_j são os variogramas espacial, temporal e conjunto. A vantagem desse modelo é que ele tem componentes espaciais, temporais e da interação que podem ser interpretados (SNEPVANGERS; HEUVELINK; HUISMAN, 2003). O parâmetro α denota a razão de anisotropia geométrica que converte unidade de distância de tempo, h_t , em distância espacial, \mathbf{h}_s , assim este faz-se necessário uma vez que uma unidade de distância no espaço não é a mesma que uma unidade de distância no tempo (HEUVELINK; GRIFFITH, 2010). Por exemplo, se $\hat{\alpha}$ é igual a 20 km por dia, então dois pontos que estão separados a 100 km de distância no espaço e 0 dias de distância no tempo tem a mesma correlação de dois pontos que estão separados a cinco dias de distância no tempo e zero quilômetros no espaço.

2.4.6 Modelos de Cressie-Huang

Cressie e Huang (1999) afirmam que os modelos separáveis apresentam uma grande desvantagem na modelagem por não considerar a interação espaço-temporal. Como solução propuseram uma classe de modelos paramétricos de covariâncias não separáveis, a qual é baseada no teorema de Bochner.

Teorema 3 (Bochner) *Seja C uma função em $\mathfrak{R}^d \times \mathfrak{R}$ contínua e simétrica. Assim, C é uma função de covariância espaço-temporal se, e somente se, é expressa da seguinte forma:*

$$C(\mathbf{h}_s, h_t) = \int \int e^{i\mathbf{h}_s' \boldsymbol{\psi} + h_t \varphi} dF(\boldsymbol{\psi}, \varphi), \quad (\mathbf{h}_s, h_t) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (2.31)$$

em que F é uma função finita, não-negativa e simétrica em $\mathfrak{R}^d \times \mathfrak{R}$ e, comumente, chamada de medida espectral. Assim, classe de funções estacionárias de covariâncias espaço-temporal definidas em $\mathfrak{R}^d \times \mathfrak{R}$ é semelhante classe das transformadas de Fourier. Se C é integrável então F é absolutamente contínua com a densidade de Lebesgue dada por:

$$f(\boldsymbol{\psi}, \varphi) = (2\pi)^{-(d+1)} \int \int e^{-i\mathbf{h}_s' \boldsymbol{\psi} + h_t \varphi} C(\mathbf{h}_s, h_t) d\mathbf{h}_s h_t, \quad (\boldsymbol{\psi}, \varphi) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (2.32)$$

em que $f(\boldsymbol{\psi}, \varphi)$ é denominada de densidade espectral. Utilizando os resultados da Equação 2.31, a função de covariância espaço-temporal estabelecida por Cressie e Huang (1999) é escrita

da forma:

$$C(\mathbf{h}_s, h_t) = \int \int e^{i\mathbf{h}_s' \boldsymbol{\psi} + h_t \varphi} g(\boldsymbol{\psi}; \varphi) d\boldsymbol{\omega} d\varphi, \quad (2.33)$$

em que $C(\mathbf{h}_s, h_t)$ é uma função de covariância espaço-temporal estacionária e $g(\boldsymbol{\psi}; \varphi) \geq 0$ uma densidade espectral da função de distribuição espectral de C . Admitindo que $C(\mathbf{h}_s, h_t)$ é integrável, então a função de covariância pode ser escrita como:

$$C(\mathbf{h}_s, h_t) = \int e^{i\mathbf{h}_s' \boldsymbol{\psi}} \rho(\boldsymbol{\psi}; h_t) l(\boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (2.34)$$

em que $\rho(\boldsymbol{\psi}; h_t)$ é uma função de autocorrelação contínua para cada $\boldsymbol{\psi} \in \mathfrak{R}^n$. Para a expressão 2.34 tem-se as seguintes condições:

- $\rho(\boldsymbol{\psi}; h_t)$ é uma função de autocorrelação contínua, $\int \rho(\boldsymbol{\psi}; h_t) l(\boldsymbol{\psi}) dh_t < \infty$ e $l(\boldsymbol{\psi}) > 0$
- $\int l(\boldsymbol{\psi}) d\boldsymbol{\psi} < \infty$.

Assim, levando em consideração estas duas condições, é possível obter modelos válidos de covariâncias espaço-temporais por meio das funções, $\rho(\boldsymbol{\psi}; h_t)$ e $l(\boldsymbol{\psi})$. Uma extensão a esta classe de modelos é proposta por Gneiting (2002). Estes autores mencionam que os modelos de Cressie-Huang apresentam uma limitação no que diz respeito a integral de Fourier, com base nesta limitação eles estabeleceram uma classe mais geral de modelos de covariância espaço-temporal.

2.5 Mínimos quadrados generalizados

Como o objetivo é ajustar um dos modelos de variograma paramétrico, o método de mínimos quadrados é a abordagem mais comum. Porém, existem outras técnicas paramétricas como: máxima verossimilhança, equações de estimação generalizada, verossimilhança composta, entre outras (SCHABENBERGER; GOTWAY, 2005).

No método de mínimos quadrados não se têm suposições a respeito da distribuição de probabilidade dos valores gerados do variograma empírico. O método considera o modelo estatístico da forma $\hat{\varepsilon}'(\mathbf{h}_s, h_t) = \gamma(\mathbf{h}_s, h_t) - \hat{\gamma}(\mathbf{h}_s, h_t)$. Assume que $\varepsilon'(\mathbf{h}_s, h_t)$ tem média nula e matriz de covariâncias $\mathbf{R} = \mathbf{R}(\boldsymbol{\omega})$, a qual dependente de $\boldsymbol{\omega}$. Uma das formas de obter a diagonal

de $\mathbf{R}(\boldsymbol{\omega})$ é dada por:

$$\frac{\text{Var}[\hat{\gamma}(\mathbf{h}_s, h_t)]}{2} \approx \frac{\gamma(\mathbf{h}_s, h_j, \boldsymbol{\omega})^2}{|N(\mathbf{h}_s, h_t)|}, \quad (2.35)$$

em que $N(\mathbf{h}_s, h_t)$ é o número de pares a cada distância espaço-temporal, de tal forma que os pesos maiores são atribuídos aos pontos do variograma empírico, calculados usando um maior número de pares. O vetor $\boldsymbol{\omega}$ contém todos parâmetros que estão presentes no variograma teórico relacionados aos efeitos do patamar, pepita e alcance.

O método dos mínimos quadrados ponderados (do inglês, *Weighted Least Squares* - WLS) substitui a $\mathbf{R}(\boldsymbol{\omega})$ pela matriz diagonal $\mathbf{W}(\boldsymbol{\omega})$. Os elementos de $\mathbf{W}(\boldsymbol{\omega})$ são obtidos pela Equação 2.35. Assim, utilizando os pesos dados na expressão 2.35, o método WLS fica dado por:

$$\begin{aligned} & (\hat{\gamma}(\mathbf{h}_s, h_t) - \gamma(\mathbf{h}_s, h_t, \boldsymbol{\omega}))' \mathbf{W}(\boldsymbol{\omega})^{-1} (\hat{\gamma}(\mathbf{h}_s, h_t) - \gamma(\mathbf{h}_s, h_t, \boldsymbol{\omega})) \\ &= \sum_{i=1}^l \sum_{j=1}^k \frac{|N(\mathbf{h}_{s_i}, h_{t_j})|}{\gamma(\mathbf{h}_{s_i}, h_{t_j}, \boldsymbol{\omega})^2} \{ \hat{\gamma}(\mathbf{h}_{s_i}, h_{t_j}) - \gamma(\mathbf{h}_{s_i}, h_{t_j}, \boldsymbol{\omega}) \}^2. \end{aligned} \quad (2.36)$$

2.6 Krigagem espaço-temporal

Uma vez construído o variograma empírico a partir dos dados e ajustado a este um modelo de variograma teórico, o próximo passo consiste na krigagem espaço-temporal dos resíduos. As únicas ferramentas de que se fazem necessárias são o preditor espaço-temporal ideal, as equações de krigagem espaço-temporais, que fornecem os pesos dos dados espaço-temporais da regionalização observada em tal preditor, e a variância de predição (MONTERO; FERNANDEZ-AVILES; MATEU, 2015). Desta forma, o modelo de variograma é crucial na krigagem espaço-temporal para calcular o melhor preditor linear não-viesado, sendo necessário obter as equações de krigagem espaço-temporal, as quais fornecem os pesos das observações em tal preditor e a variação de predição que indica a precisão das predições (KILIBARDA et al., 2015).

Nesta seção é apresentado o preditor de krigagem espaço-temporal e as equações de krigagem espaço-temporais acompanhadas da variância de predição. No entanto, estes resultados são uma generalização do caso puramente espacial. Seja $\{\varepsilon(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq \mathbb{R}^2 \times \mathbb{R}\}$. Considere que $\varepsilon(\mathbf{s}, t)$ são observados nos n locais espaço-temporais $\varepsilon(\mathbf{s}_1, t_1), \varepsilon(\mathbf{s}_2, t_2), \dots, \varepsilon(\mathbf{s}_n, t_n)$.

A krigagem ordinária espaço-temporal (MONTERO; FERNANDEZ-AVILES; MATEU, 2015) consiste em prever o valor $\varepsilon(\mathbf{s}_0, t_0)$ em um ponto espaço-temporal não amostrado (\mathbf{s}_0, t_0) , a partir do componente estocástico $\varepsilon(\mathbf{s}, t)$. Para isso, é utilizado o preditor linear, $\varepsilon^*(\mathbf{s}_0, t_0) = \sum_{i=1}^n \lambda_i \varepsilon(\mathbf{s}_i, t_i)$. λ_i são os pesos da krigagem obtidos impondo as condições de que a esperança do erro de predição seja zero e que tenha variância mínima, ou seja, ser o melhor preditor linear não-viesado (do inglês, *Best Linear Unbiased Prediction* - BLUP).

2.6.1 Equações da krigagem espaço-temporal

Semelhante ao caso espacial, as equações de krigagem espaço-temporal dependem da hipótese de estacionariedade atribuída ao processo estocástico (HENGL; HEUVELINK; ROSSITER, 2007). Seja $\varepsilon(\mathbf{s}, t)$ um processo estocástico espaço-temporal estacionário de segunda ordem, com média constante e função de covariância $C(\mathbf{h}_s, h_t)$. Assim, as equações de krigagem espaço-temporal são expressas da forma:

$$\sum_{j=1}^n \lambda_j C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) = C(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0), \forall i = 1, 2, \dots, n, \quad (2.37)$$

em que λ_j são os pesos da krigagem (mais adiante falaremos um pouco mais destes pesos). A variação do erro de predição é expressa da seguinte forma:

$$V[\varepsilon^*(\mathbf{s}_0, t_0) - \varepsilon(\mathbf{s}_0, t_0)] = C(\mathbf{0}, 0) - \sum_{i=1}^n \lambda_i C(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0). \quad (2.38)$$

Se $\varepsilon(\mathbf{s}, t)$ é estacionário de segunda ordem com média constante, embora desconhecido, e função de covariância conhecida $C(\mathbf{h}_s, h_t)$, ou com uma média constante e desconhecida, mas com variância não limitada, a krigagem espaço-temporal não pode ser executada porque a média do processo estocástico é desconhecida. Assim, deve-se impor uma condição de não-viés (MONTERO; FERNANDEZ-AVILES; MATEU, 2015). Nessas situações, utilizam-se as equações de krigagem espaço-temporais ordinárias para obter os pesos ótimos do preditor de krigagem. Na obtenção dos pesos podem ser utilizadas as equações de krigagem espaço-temporais ordinárias que serão expressas em termos do variograma:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_{st}(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) + \alpha = \gamma_{st}(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0), \forall i = 1, 2, \dots, n. \\ \sum_{i=1}^n \lambda_i = 1, \end{cases} \quad (2.39)$$

em que α é o multiplicador Lagrange. A variância de predição correspondente é:

$$V(\boldsymbol{\varepsilon}^*(s_0, t_0) - \boldsymbol{\varepsilon}(s_0, t_0)) = \sum_{i=1}^n \lambda_i C(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0) + \alpha. \quad (2.40)$$

O preditor final, denotado por $\hat{z}(\mathbf{s}_0, t_0)$, para a variável Z , na localização (\mathbf{s}_0, t_0) , é calculado por meio da krigagem com regressão, a qual combina métodos de regressão com krigagem. O método da krigagem espaço-temporal por regressão é descrito com detalhes em Hengl et al. (2011), Heuvelink et al. (2012) e Kilibarda et al. (2014). A vantagem da krigagem com regressão é que esta permite utilizar diferentes métodos de regressão, possibilitando a interpretação separada dos componentes preditos (HENGL; HEUVELINK; ROSSITER, 2007). Assim, o preditor final, $\hat{z}(\mathbf{s}_0, t_0)$, é expresso por:

$$\hat{z}(\mathbf{s}_0, t_0) = \hat{m}(\mathbf{s}_0, t_0) + \hat{\boldsymbol{\varepsilon}}(\mathbf{s}_0, t_0), \quad (2.41)$$

em que $\hat{m}(\mathbf{s}_0, t_0)$ é o valor estimado para a localização (\mathbf{s}_0, t_0) obtido por meio do ajuste de um modelo de regressão e $\hat{\boldsymbol{\varepsilon}}(\mathbf{s}_0, t_0)$ é o valor predito por meio da krigagem ordinária.

As Equações (2.39) e (2.40) são semelhantes à krigagem espacial uma vez que a formulação matemática é a mesma (HEUVELINK et al., 2012; KILIBARDA et al., 2015). Entretanto, uma das vantagens que a krigagem espaço-temporal tem sobre a krigagem puramente espacial é que esta, além de fazer interpolação para locais não amostrados, realiza predições para valores futuros (IACO; PALMA; POSA, 2015; JOVEIN; HOSSEINI, 2017).

2.6.2 Análise de desempenho

A análise do desempenho da interpolação na krigagem espaço-temporal será realizada utilizando o método de validação “leave-one-out” (HENGL et al., 2011; HEUVELINK et al., 2012; KILIBARDA et al., 2014; HU et al., 2015; IACO; PALMA; POSA, 2015; MARTÍNEZ; MELO; MELO, 2017). Esse método consiste em retirar um ponto observado no espaço-tempo, denotado por $z(\mathbf{s}_i, t)$ e, em seguida, predizer o seu valor $\hat{z}(\mathbf{s}_i, t)$. Este processo é então repetido,

até que todos os pontos tenham sido escolhidos. Note que, na validação espaço-temporal, ao retirar um ponto espacial, estão sendo retiradas todas as observações temporais medidas naquele ponto.

As estatísticas que serão calculadas para medir a qualidade do ajuste serão a Raiz Quadrada do Erro Quadrático Médio (do inglês, *Root Mean Square Error* - RMSE), o Erro Médio Absoluto (do inglês, *Mean Absolute Error* - MAE), o Coeficiente de Determinação (R^2), Coeficiente de Correlação de Pearson (CCP) e o Coeficiente de Correlação de Concordância (CCC). No artigo 3 é utilizada uma metodologia inovadora proposta por Bastos (2017) que combina as estatísticas RMSE, MAE, R^2 e CCC por meio de uma técnica de classificação.

3 CONSIDERAÇÕES FINAIS

Nesta pesquisa, para analisar observações amostradas no espaço-tempo, em especial dados de chuvas, foi empregada uma metodologia geoestatística espaço-temporal. A variação em larga escala, que compreende o componente de tendência do fenômeno em estudo, foi modelada utilizando covariáveis espaciais e temporais, a fim de retirar a tendência espaço-temporal nos resíduos da regressão. No ajuste do componente de tendência foi proposta a modelagem com regressão linear gaussiana e com modelos GAMLSS com diferentes distribuições para a variável resposta. Os resultados da validação cruzada indicaram melhores resultados quando considerado, no ajuste da tendência, os modelos GAMLSS. A metodologia apresentada neste artigo possibilitou a criação de cenários futuros para a área em estudo, interpolando em locais não observados e prevendo estimativas em anos futuros.

REFERÊNCIAS

- ALMEIDA, H. A. de; MEDEIROS, E. A. Variabilidade no regime pluvial em duas mesorregiões da Paraíba e sua relação com o fenômeno EL Niño Oscilação Sul. **Journal of Environmental Analysis and Progress**, Journal of Environmental Analysis and Progress - JEAP, v. 2, n. 3, p. 177–185, 2017.
- ANA. **Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas (ANA)**. Brasília, 2017.
- ARAUJO, M. V. O.; CELESTE, A. B. Rescaled range analysis of streamflow records in the São Francisco River Basin, Brazil. **Theoretical and Applied Climatology**, Springer Nature, 2018. ISSN 0177-798X.
- BASTOS, R. L. **Métodos para seleção de modelos de semivariograma em campos aleatórios gaussianos**. Tese (Doutorado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 2017. Disponível em: <<http://repositorio.ufla.br/handle/1/15568>>.
- CESARE, L.; MYERS, D.; POSA, D. Estimating and modeling space–time correlation structures. **Statistics & Probability Letters**, Elsevier BV, v. 51, n. 1, p. 9–14, 2001.
- CHEN, M. et al. Assessing objective techniques for gauge-based analyses of global daily precipitation. **Journal of Geophysical Research**, American Geophysical Union (AGU), v. 113, n. D4, 2008.
- COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: The lms method and penalized likelihood. **Statistics in Medicine**, Wiley, v. 11, n. 10, p. 1305–1319, 1992.
- CRESSIE, N.; HUANG, H.-C. Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. **Journal of the American Statistical Association**, Informa UK Limited, v. 94, n. 448, p. 1330–1339, 1999.
- CRESSIE, N.; WIKLE, C. K. **Statistics for spatio-temporal data**. Hoboken, Nova Jersey, EUA: John Wiley & Sons, 2015.
- DIMITRAKOPOULOS, R.; LUO, X. Spatiotemporal Modelling: Covariances and Ordinary Kriging Systems. In: **Geostatistics for the Next Century**. [S.l.]: Springer Netherlands, 1994. p. 88–93.
- GNEITING, T. Nonseparable, Stationary Covariance Functions for Space–Time Data. **Journal of the American Statistical Association**, Informa UK Limited, v. 97, n. 458, p. 590–600, 2002.
- GRÄLER, B. et al. Spatio-temporal analysis and interpolation of PM10 measurements in Europe for 2009. **ETC/ACM Technical Paper**, v. 8, p. 1–29, 2012.
- HENGL, T.; HEUVELINK, G. B.; ROSSITER, D. G. About regression-kriging: From equations to case studies. **Computers & Geosciences**, Elsevier BV, v. 33, n. 10, p. 1301–1315, 2007.
- HENGL, T. et al. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. **Theoretical and Applied Climatology**, Springer Nature, v. 107, n. 1-2, p. 265–277, 2011.

- HEUVELINK, G. et al. Sampling Design Optimization for Space-Time Kriging. In: _____. **Spatio-Temporal Design**. Hoboken, Nova Jersey, EUA: Wiley-Blackwell, 2012. cap. 9, p. 207–230. ISBN 9781118441862.
- HEUVELINK, G. B. M.; GRIFFITH, D. A. Space-Time Geostatistics for Geography: A Case Study of Radiation Monitoring Across Parts of Germany. **Geographical Analysis**, Wiley-Blackwell, v. 42, n. 2, p. 161–179, 2010.
- HU, D. et al. Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. **Cluster Computing**, Springer Nature, v. 20, n. 1, p. 347–357, 2017.
- HU, Y. et al. Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. **PLOS Neglected Tropical Diseases**, Public Library of Science (PLoS), v. 9, n. 2, p. e0003470, 2015.
- HUANG, H.-C. et al. Model comparison and selection for stationary space–time models. **Computational Statistics & Data Analysis**, Elsevier BV, v. 51, n. 9, p. 4577–4596, 2007.
- IACO, S.; MYERS, D.; POSA, D. Space–time analysis using a general product–sum model. **Statistics & Probability Letters**, Elsevier BV, v. 52, n. 1, p. 21–28, 2001.
- IACO, S. D.; PALMA, M.; POSA, D. Spatio-temporal geostatistical modeling for French fertility predictions. **Spatial Statistics**, Elsevier BV, v. 14, p. 546–562, 2015.
- JOVEIN, E. B.; HOSSEINI, S. M. Predicting saltwater intrusion into aquifers in vicinity of deserts using spatio-temporal kriging. **Environmental Monitoring and Assessment**, Springer Nature, v. 189, n. 2, p. 81, 2017.
- KILIBARDA, M. et al. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. **Journal of Geophysical Research: Atmospheres**, American Geophysical Union (AGU), v. 119, n. 5, p. 2294–2313, 2014.
- KILIBARDA, M. et al. Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. **Spatial Statistics**, Elsevier BV, v. 14, p. 22–38, 2015.
- LOPES, I. et al. Variation of aridity index and climatic trend to desertification for the semi-arid region of the Brazilian Northeast. **Revista Brasileira de Geografia Física**, Revista Brasileira de Geografia Física, v. 10, n. 4, p. 1014–1026, 2017.
- MARTÍNEZ, W. A.; MELO, C. E.; MELO, O. O. Median Polish Kriging for space–time analysis of precipitation. **Spatial Statistics**, Elsevier BV, v. 19, p. 1–20, feb 2017.
- MENEZES, R. et al. Spatial–temporal modellization of the NO_2 concentration data through geostatistical tools. **Statistical Methods & Applications**, Springer Nature, v. 25, n. 1, p. 107–124, 2015.
- MONTEIRO, A.; MENEZES, R.; SILVA, M. E. Modelling spatio-temporal data with multiple seasonalities: The NO_2 Portuguese case. **Spatial Statistics**, Elsevier BV, v. 22, p. 371–387, 2017.
- MONTERO, J. M.; FERNANDEZ-AVILES, G.; MATEU, J. **Spatial and spatio-temporal geostatistical modeling and kriging**. Hoboken, Nova Jersey, EUA: John Wiley & Sons, 2015.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley-Blackwell, v. 54, n. 3, p. 507–554, 2005.

ROUHANI, S.; HALL, T. J. Space-Time Kriging of Groundwater Data. In: **Geostatistics**. [S.l.]: Springer Netherlands, 1989. p. 639–650.

SCHABENBERGER, O.; GOTWAY, C. A. **Statistical Methods for Spatial Data Analysis**. Boca Raton, Flórida, EUA: CRC press, 2005.

SHERMAN, M. **Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties**. Hoboken, Nova Jersey, EUA: John Wiley & Sons, Ltd, 2010.

SILVEIRA, C. et al. Climate change in the São Francisco river basin: analysis of precipitation and temperature. **Revista Brasileira de Recursos Hídricos**, Associação Brasileira de Recursos Hídricos - ABRH, v. 21, n. 2, p. 416–428, 2016.

SNEPVANGERS, J.; HEUVELINK, G.; HUISMAN, J. Soil water content interpolation using spatio-temporal kriging with external drift. **Geoderma**, Elsevier BV, v. 112, n. 3-4, p. 253–271, 2003.

STASINOPOULOS, D. M.; RIGBY, R. A. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. **Journal of Statistical Software**, Foundation for Open Access Statistic, v. 23, n. 7, p. 1–46, 2007.

STASINOPOULOS, M. D. et al. **Flexible Regression and Smoothing: Using GAMLSS in R**. Boca Raton, Flórida, EUA: CRC Press, 2017.

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Rainfall spatiotemporal variability analysis covering recent decades in Paraíba, Brazil

**Artigo redigido conforme as normas da
revista *International Journal of
Climatology*
(versão submetida)**

Rainfall spatiotemporal variability analysis covering recent decades in Paraíba, Brazil

Elias Silva de Medeiros¹, Renato Ribeiro de Lima¹, Ricardo Alves de Olinda², Carlos Antônio Costa dos Santos³

¹Department of Statistics, Federal University of Lavras, Lavras, Minas Gerais, Brazil

²Department of Statistics, Paraíba State University, Campina Grande, Paraíba, Brazil

³Center for Technology and Natural Resources, Federal University of Campina Grande, Campina Grande, Paraíba, Brazil

Correspondence to: Elias Silva de Medeiros, Department of Statistic, Federal University of Lavras, Lavras, Minas Gerais, 37200-000, Brazil (eliasestatistica@gmail.com).

Abstract

Knowledge of rainfall estimates is very important in several areas of hydrology, and the rain distribution can show high variability over both space and time. Thus, several studies not only in hydrology but also in other areas have analysed spatiotemporal datasets from two perspectives by fixing space and studying its behaviour over time or by fixing time and studying space. In the present study, the rainfall distribution in the state of Paraíba, which is located in the north-eastern region of Brazil, was analysed jointly, because this area presents high variability in the North-South and East-West directions throughout the year. The mean precipitation values observed in several irregularly spaced pluviometric stations during the years 1994-2014 showed great variation between their mesoregions and during all the months of the year. To model this behaviour, we propose that rainfall should be modelled over space and time. For this purpose, a spatiotemporal geostatistical model is proposed that is capable of not only performing interpolations for places that have not been observed but also for future times. The results indicate a high degree of correlation for the spatiotemporal interaction. Finally, rainfall estimates from the Paraíba region indicate the dramatic spatial and temporal variability that directly affects the water resources of the entire region, providing a detailed visual analysis of sectors undergoing anywhere from a scarcity to an excess of rainfall. These results should motivate further research, including in the north-eastern region of Brazil, with the aim predicting which places could experience droughts or floods.

Keywords: Droughts, Spatiotemporal kriging, Precipitation

1 Introduction

The analysis of spatiotemporal datasets has been performed from two perspectives: to keep the space as a variable and study its behaviour over time; or to keep the time and studying space. Banerjee *et al.* (2014) mention that spatiotemporal modelling has being approached in theoretical studies as well as in applied studies, which has gained visibility in recent years, because phenomena indexing over both space and time is poorly understood. Thus, the concerning on the spatiotemporal dynamics has led to development spatiotemporal models in

many areas of knowledge such as epidemiology (Hu *et al.*, 2015), demography, environmental science (Monteiro *et al.*, 2017) and hydrology (Hu *et al.*, 2017; Martínez *et al.*, 2017).

Once spatiotemporal observations are considered, their interpolation presents benefits in relation to purely spatial predictions. One of these benefits is that interpolation can be performed in georeferencing positions over space-time (Hu *et al.*, 2015; Iaco *et al.*, 2015). In this study, spatiotemporal geostatistics methodology will be applied to the monthly total rainfall data from the state of Paraíba, north-eastern Brazil.

The rainfall distribution in the state of Paraíba presents high variability in the North-South and East-West directions entire the year (Gomes *et al.*, 2015). Therefore, the rainfall pattern in this region indicates variations in both space and time. Thus, due to this behaviour, there is a need to model this phenomenon jointly. The novelty of this study is the modelling of the spatiotemporal of monthly rainfall in the state of Paraíba using spatiotemporal geostatistical methodology. It is noteworthy that no research has applied this methodological approach has not been taking into consideration for rainfall studies in Brazil.

The theoretical foundations of spatiotemporal geostatistics are presented in Cressie and Wikle (2015). Some studies have approached the computational implementation of the spatiotemporal geostatistics model. Gräler *et al.* (2016) and Gräler *et al.* (2012) discuss this type of issue, noting that the methodology has been little disseminated in areas other than statistics. However, in recent years, packages for spatiotemporal geostatistical analysis have been implemented in R software, RandomFields (Schlather *et al.*, 2015), spTimer (Bakar and Sahu, 2015), INLA (Cameletti, 2015) and spacetime (Pebesma, 2012). The gstat package (Gräler *et al.*, 2016) uses the classes of the spacetime package from the model fitting of spatiotemporal covariance to spatiotemporal interpolation.

Some studies have applied spatiotemporal geostatistics to rainfall data around the world. Raja *et al.* (2017) performed the spatiotemporal kriging of rainfall in Turkey, observing that this country is located in one of the regions with higher risk of possible effects from climatic change, similar to the state of Paraíba, in Brazil. Thus, they report that to understand the temporal and spatial characteristics of rainfall is crucial for risk management as well as for water resource planning and management, which essentially depend on rainfall. Hu *et al.* (2017) used spatiotemporal regression by kriging in the Xinjiang Uygur Autonomous region, north-western China, to model the monthly rainfall from 2004 to 2013. The authors identified strong spatiotemporal correlation in the rainfall distribution in this region, thus justifying the use of the

methodology proposed in the study. Martínez *et al.* (2017) performed a spatiotemporal analysis of rainfall data in Colombia, identifying regions that presented irregular rainfall distributions.

Thus, the present study aims to analyse the spatiotemporal distribution of the rainfall in the state of Paraíba, Brazil. In this regard, the following issues were approached: investigate the variations over both space and time to understand the rainfall mechanism in this region; identify the degree of spatiotemporal correlation in rainfall; and identify priority regions the state for a thorough investigation of those undergoing everything from scarcity to excess rainfall.

2 Material

The dataset used in this study were obtained from the Agency of Water Management of Paraíba State (AESA), which is responsible for rainfall information in the state. The state of Paraíba is located in the north-eastern Brazil, between parallels 6° and 8° south latitude, meridian 34° and 39° west longitude, and accounting an area of approximately $56,585 \text{ km}^2$. This state is in the tropical region, being divided into the following four mesoregions: Zona da Mata, Agreste, Borborema and Sertão.

The dataset, refer to the series of monthly rainfall from 269 rainfall stations over the period from 1994-2014.

The rainfall data matrix for the 1994-2014 period is shown in Figure 1. There are 269 pluviometric stations, of which 215 (80% of the stations) have missing data (white tones). Thus, we propose that for the trend adjustment, followed by the residual variogram, the 54 stations with records for all the months of the study period should be considered, because these locations adequately represent the variability of the data. The other stations that were not included in the adjustment of the variogram were used to perform space-time kriging.

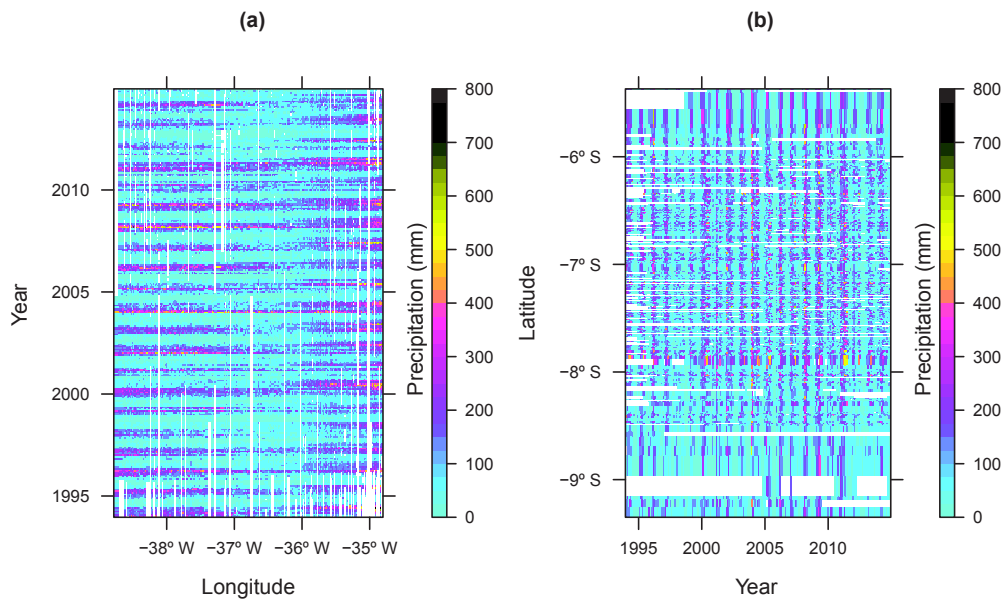


Figure 1: Monthly rainfall of the 269 stations in the period of 1994-2014.

The spatial distribution of the 269 rainfall stations is shown in Figure 2, with identification of the mesoregions in Paraíba state. The points within the map represent the 269 rain stations. The major points are the 54 rainfall stations that were used to model the trend and the variogram.

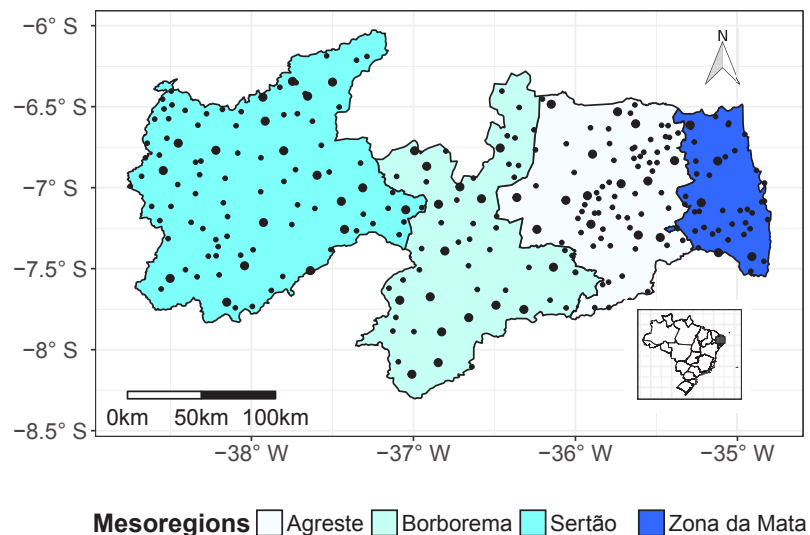


Figure 2: Spatial distribution of the 269 rainfall stations sampled (black dots) in the State of Paraíba divided into four mesoregions. The major points are the 54 rainfall stations used to model the trend and the variogram.

3 Methods

3.1 Spatiotemporal model

In this study, the spatiotemporal process $\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq R^d \times R\}$ was assumed, where $\mathbf{s} \in R^d$, being $d = 2$, representing the two dimensions in space (latitude and longitude), and $t \in R$ representing only one dimension, the time (Cressie and Wikle, 2015). Whether the variable Z was observed was reflected in spatiotemporal coordinates (\mathbf{s}_i, t_j) , $i = 1, \dots, n$ and $j = 1, \dots, m$.

Commonly, one of the primary purposes of spatiotemporal data analysis is to predict a spatial location \mathbf{s}_0 at a given point in time t_0 , in which (\mathbf{s}_0, t_0) it not observed (Bahrami Jovein and Hosseini, 2017). Another aspect, which is no less important, is about the knowledge of the spatiotemporal dependence of Z . Spatiotemporal dependence degree serves as a basis for formulating interpretations about spatial, temporal and spatiotemporal variations, which become essential for the optimization of spatiotemporal monitoring projects (Heuvelink *et al.*, 2012; Kilibarda *et al.*, 2015). Thus the spatiotemporal variation of Z can be shared into a trend m and stochastic residual ε as follows:

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

It is assumed that Z has first and second-order moments. The trend component of is equal to $E[Z(\mathbf{s}, t)]$ is the expectation variable Z . This component can be modeled by a regression model fitted by the ordinary least squares (OLS) method. The residual ε includes the following components: spatial, temporal and spatiotemporal (Iaco *et al.*, 2015). For modelling, it was stated that these three components are second-order stationary, mutually independent and spatially isotropic.

The observed rainfall values are considered as an accomplishment of the spatiotemporal random field defined in Eq. (1), where the deterministic component $m(\mathbf{s}, t)$, which is non-constant over space and time, represents large-scale variation, while the residual $\varepsilon(\mathbf{s}, t)$ describes random fluctuations at a small scale (Heuvelink *et al.*, 2012).

3.2 Components of the trend

The trend component $m(\mathbf{s}, t)$ can be estimated by multiple linear regression. Several studies have used this method to model this trend component in spatiotemporal geostatistics,

including the study of Gräler *et al.* (2012) that analysed the air quality indicator PM_{10} , which was measured daily, according to the altitude of the seasons and a temporal index to adjust the annual periodicity present in the data by $\cos(2\pi \times day/365)$. In their research on the spatiotemporal prediction of daily temperatures in Croatia, Hengl *et al.* (2012) modelled the trend using MODIS images to extract information on the latitude, longitude, and solar radiation, among others, and they created the temporal index for modelling seasonal fluctuations in the daily temperature using $\cos(2\pi \times day/365)$. Hu *et al.* (2015) studied the prevalence of Schistosomiasis, which was significantly related to the distance of each village from the river along with environmental variables. De Iaco *et al.* (2015) used the region factor as a trend for the fertility rate of women in France; Hu *et al.* (2017) studied the monthly spatiotemporal distribution of rainfall and observed that it has presented a significant linear relationship with the vegetation index obtained by MODIS images and with the temporal index, using $\cos(2\pi \times month/12)$.

In this study, the geographic coordinates (latitude and longitude) and a temporal index to contour the effect of the annual seasonality on the precipitation were considered as covariates. The adjusted trend model is given by the following:

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

where $x_1 = \cos x = \cos(\text{month}/12 \cdot 2\pi)$ and $x_2 = \text{sen} x = \sin(\text{month}/12 \cdot 2\pi)$. The $\cos x$ and $\text{sen} x$ functions were used as covariates in the multiple regression model to explain the seasonal variation. The variables x_3 , x_4 and x_5 represent the latitude, longitude and the quadratic effect of the longitude, respectively. Although the covariates exhibit spatial, temporal and spatio-temporal variations, the regression model cannot account for all the variations. The residuals of this model may have space-time dependency, which indicates that a space-time variogram can be estimated and then used to interpolate the residues by means ordinary kriging.

3.3 Spatiotemporal variogram

After specifying the trend component and obtaining the parameter estimates, this result could then be subtracted from Z , leaving the spatiotemporal stochastic residue ε . One of the advantages of inserting components into the trend is that they remove factors from the residuals

behaviour that lead to a large variation in the kriging error, because this approach counts the uncertainties in the trend coefficients (Heuvelink and Griffith, 2010).

The residuals produced by this regression will then be used for the construction of a sample variogram. For a more detailed model, the monthly rainfall (Z) was modeled as a sum of the trend and the spatiotemporal residue $\varepsilon(\mathbf{s}, t)$ as defined by the following:

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

After analysing and removing the spatial and temporal trends in the data, the spatiotemporal interest is shown in terms of the empirical evaluation of the covariance function (Martínez *et al.*, 2017). The spatiotemporal covariance is usually described using spatiotemporal variogram γ , which measures the mean difference between the separated data in the spatiotemporal domain using the distance vector (\mathbf{h}_s, h_t) . Nevertheless, the estimation of the spatiotemporal variogram encompasses features that must be taken into consideration, such as the distinct scales in space and time and the principles of different causalities (Hengl *et al.*, 2012). The spatiotemporal variogram is defined as follows:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \frac{1}{2} E [\varepsilon(\mathbf{s}, t) - \varepsilon(\mathbf{s} + \mathbf{h}_s, t + h_t)]^2. \quad (4)$$

In Equation (4), $\mathbf{h}_s = \mathbf{s} - \mathbf{s}'$ and $h_t = t - t'$, for any (\mathbf{s}, t) and (\mathbf{s}', t') in the spatiotemporal domain.

An estimation for the variogram $\gamma_{st}(\mathbf{h}_s, h_t)$ is established as:

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

According to the sample variogram presented in Eq. (5), $|L(\mathbf{r}_s, r_t)|$ is the cardinality for the set $L(\mathbf{r}_s, r_t)$, that is

$$L(\mathbf{r}_s, r_t) = \{(\mathbf{s} + \mathbf{h}_s, t + h_t) \in A, (\mathbf{s}, t) \in A : \mathbf{h}_s \in Tol(\mathbf{r}_s) \text{ e } h_t \in Tol(r_t)\},$$

where $Tol(\mathbf{r}_s)$ and $Tol(r_t)$ are the vectors for the spatial and temporal lag of \mathbf{r}_s and r_t , respectively. It is worth noting that there is no definition of spatiotemporal, because the pairs

of points separated by (\mathbf{h}_s, h_t) are detected by computing the spatial and temporal distances separately (Iaco *et al.*, 2015).

3.4 Generalized product-sum model

The next step of the analysis is to determine a theoretical model to be adjusted to the spatiotemporal sample variogram based on the covariance functions.

It is known that spatial and temporal domains have different scales. An example is the data analysed in this study, in which the spatial scale is in kilometres, whereas the time is given in months. This type of characteristic causes the variogram appraisal and the interpolation to have different methods (Hengl *et al.*, 2012). Statistical assumptions, which are generally defined in the literature, will address the estimation of the spatiotemporal variogram, and moreover, they will ensure that the defined non-negative covariance function (Cressie and Wikle, 2015; Huang *et al.*, 2007; Iaco *et al.*, 2001). In the spatiotemporal approach, there are basically two types of general models, separable and non-separable. Separable models assume that the spatial and temporal processes are not correlated and the variograms are composed of purely both spatial and temporal factors (Huang *et al.*, 2007). However, the non-separable models assume that the spatiotemporal phenomena are correlated and the variogram structures include the marginal models of space and time as well as the interaction of these structures (Cressie and Wikle, 2015; Iaco *et al.*, 2001; Sherman, 2010). In this study, the non-separable models, which were of a generalized product sum type, were used to model the spatiotemporal rainfall structures in the state of Paraíba. The formulation of this model is described by Gräler *et al.* (2016).

Cesare *et al.* (2001) performed a study that shows some covariance functions, especially the product-sum model classes, reporting that the coefficients of this class should incorporate the spatial and temporal and the interaction between then parameters. Furthermore, these authors compared the models of this class with other covariance structures. Similarly, Iaco *et al.* (2001) established a generalization for the product-sum covariance model ($C_{st}(\mathbf{h}_s, h_t)$) and reported that the product-sum model incorporates peculiar characteristics involving the values of the sills. Therefore, this model is written as follows:

$$C_{st}(\mathbf{h}_s, h_t) = kC_s(\mathbf{h}_s)C_t(h_t) + C_s(\mathbf{h}_s) + C_t(h_t). \quad (6)$$

In Equation (6), C_s and C_t are spatial and temporal covariance functions, respectively.

The spatiotemporal variogram ($\gamma_{st}(\mathbf{h}_s, h_t)$) can thus be expressed as follows:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_{st}(\mathbf{h}_s, 0) + \gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h}_s, 0)\gamma_{st}(\mathbf{0}, h_t). \quad (7)$$

In Equation (7), $\gamma_{st}(\mathbf{h}_s, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$ are spatial and temporal marginal variograms, respectively. The great advantage of this type of model is that marginal fitting is based on the sample and only one parameter of the global sill is incorporated in the spatiotemporal interaction (Iaco *et al.*, 2015). The parameter k is positive and has the following identity that involves the global sill ($sill_{st}$), together with the spatial and temporal sills, $sill_s$ and $sill_t$, respectively:

$$sill_{st} = k \times sill_s \times sill_t + sill_s + sill_t. \quad (8)$$

In Equation (8), k can be estimated by using the sill information. This is an alternative that Gräler *et al.* (2016) described for the formulation of the model product sum proposed in Eq. (7). In this study, the authors proposed the following restructuring for Eq. (7), which is used in this study:

$$\gamma_{st}(\mathbf{h}_s, h_t) = (k \cdot sill_t + 1)\gamma_{st}(\mathbf{h}_s, \mathbf{0}) + (k \cdot sill_s + 1)\gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h}_s, \mathbf{0})\gamma_{st}(\mathbf{0}, h_t). \quad (9)$$

Based on the sample spatiotemporal variogram, $\gamma_{st}(\mathbf{r}_s, r_t)$, the marginal variograms in this study were outlined. Through a visual analysis, it possible to define some theoretical models, which will be part of the Eq. (9), such as Gaussian, Exponential and Spherical models. In this way, it is possible to direct the structures for $\gamma_{st}(\mathbf{h}_s, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$.

3.5 Spatiotemporal kriging

After establishing the trend function and the covariance for the residue, the spatiotemporal interpolation can be calculated. At the beginning of the 21st century, the first theoretical works on spatiotemporal geostatistics were published (Cressie and Wikle, 2015). The spatiotemporal regression method is described in detail in Hengl *et al.* (2012), Heuvelink *et al.* (2012), and Kilibarda *et al.* (2015). According to Kilibarda *et al.* (2015), the variogram model is crucial in spatiotemporal kriging to calculate the best non-biased linear predictor, which is given by the following:

$$\hat{\epsilon}(\mathbf{s}_0, t_0) = \mathbf{c}_0^T \mathbf{C}_n^{-1} \bar{\epsilon}. \quad (10)$$

In Equation (10), $\varepsilon(\mathbf{s}_0, t_0)$ is the linear predictor, as obtained by kriging, for the location (\mathbf{s}_0, t_0) , \mathbf{C}_n is the covariance matrix of order $n \times n$ of the residues for the n points observed in space-time as derived from the spatiotemporal variogram, \mathbf{c}_0 is a covariance vector of residues for the observed and predicted points and $\bar{\boldsymbol{\varepsilon}}$ is a vector of residues in the n observed points.

The final predictor, $z(\mathbf{s}_0, t_0)$, for the rainfall variable (Z), at the location (\mathbf{s}_0, t_0) , which was given by Kilibarda *et al.* (2014), is defined as follows:

$$\hat{z}(\mathbf{s}_0, t_0) = \hat{m}(\mathbf{s}_0, t_0) + \hat{\varepsilon}(\mathbf{s}_0, t_0). \quad (11)$$

In Equation (11), $m(\mathbf{s}_0, t_0)$ is the estimated amount for the location (\mathbf{s}_0, t_0) obtained by the Eq. (2) and $\varepsilon(\mathbf{s}_0, t_0)$ as defined in the Eq. (10).

Kilibarda *et al.* (2015) defined the error variance of the prediction in spatiotemporal kriging as follows:

$$\begin{aligned} \sigma^2(\mathbf{s}_0, t_0) &= \text{Var} [Z(\mathbf{s}_0, t_0) - \hat{Z}(\mathbf{s}_0, t_0)] \\ &= C(\mathbf{0}, 0) - \mathbf{c}_0^T \mathbf{C}_n^{-1} \mathbf{c}_0 + \\ &+ (\mathbf{m}_0 - \mathbf{M}^T \mathbf{C}_n^{-1} \mathbf{c}_0)^T (\mathbf{M}^T \mathbf{C}_n^{-1} \mathbf{M})^{-1} (\mathbf{m}_0 - \mathbf{M}^T \mathbf{C}_n^{-1} \mathbf{c}_0). \end{aligned} \quad (12)$$

The Equation (12) is similar to the spatial kriging once the mathematical formulation is the same (Heuvelink *et al.*, 2012; Kilibarda *et al.*, 2015). We have that $C(\mathbf{0}, 0)$ is the variance for Z , \mathbf{M} is the matrix of the order design $n \times p$ of variable predictors in the observed points, and \mathbf{m}_0 is the vector of predictors in the prediction point.

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

- 1^o step:** the values estimated by Eq. (2) are extracted, followed by the extraction of the residues for all the observations;
- 2^o step:** after obtaining these residues, the values of the sample spatiotemporal variogram are calculated as defined in Eq. (5), followed by the theoretical model fitting defined in Eq. (9); and
- 3^o step:** the residues are then interpolated using Eq. (10) and then adding the trend component, resulting in the final predictor defined in Eq. (11).

3.6 Selection and Performance of the Model

To determine the best theoretical variogram model that fit the sample variogram, the RMSE (Root Mean Square Error) was calculated, which was obtained by taking the difference in the point clouds between the sampling variogram and the theoretical variogram.

After selecting the model that best fit the cloud of points from the sample variogram, space-time kriging was performed to measure the performance of the fitted model. Thus, interpolation was performed by accounting for the leave-one-out method (Hengl *et al.*, 2012; Kilibarda *et al.*, 2014; Martínez *et al.*, 2017) as applied to all 269 pluviometric stations. This method consists of removing an observed point from the rainfall in space-time ($z(\mathbf{s}_0, t)$) and then predicting its value ($\hat{z}(\mathbf{s}_0, t)$). This process is then repeated for all the remaining points and residues of this process, and then the difference between the observed and predicted values are taken at each location ($z(\mathbf{s}_0, t) - \hat{z}(\mathbf{s}_0, t)$). After the leave-one-out method was applied to measure the kriging performance, a scatter plot of the observed values versus the values predicted by space-time kriging was constructed as well as the coefficient of determination. Additionally, the application of the methodology proposed in this paper allows missing data to be entered.

One of the advantages that spatial-temporal kriging has relative to purely spatial kriging is that it does not only interpolate to non-sampled locations, but it also predicts future values in time (Bahrami Jovein and Hosseini, 2017; Iaco *et al.*, 2015). Spatio-temporal kriging were implemented in the R package *gstat* (Graler et al. 2016).

4 Results

4.1 Exploratory analysis

Maps of the monthly average rainfall value, in millimetres (mm), for the 1994-2014 period are shown in Figure 3.

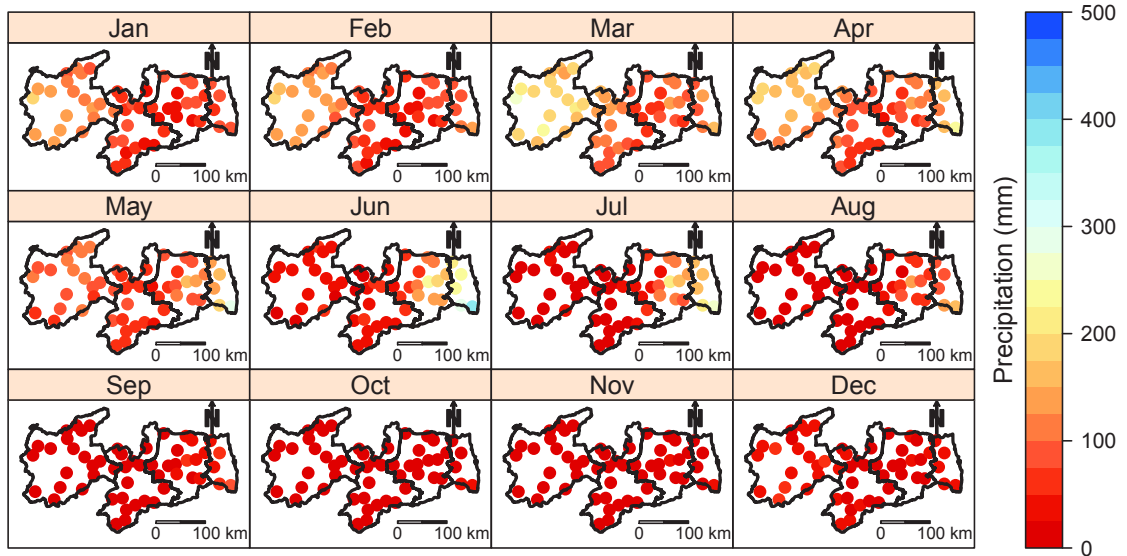


Figure 3: Maps of the monthly average values of rainfall in each station for the period of 1994-2014.

Based on the average value, there is a difference in the rainfall distribution over months (Fig. 3). It is also observed that, within each month, there is a distinct behaviour for the average monthly rainfall in each mesoregion. Similar results are observed in the studies by Menezes *et al.* (2010) and Macedo *et al.* (2010). In this figure, it is evident that each mesoregion has a distinct rainy period and distinguishing mechanisms in the rainfall distribution. For example, in Sertão state, the rainfall occurs from January to April, and, in the Zona da Mata, the rainfall is well distributed throughout the year, on average. Based on this finding, the great spatial and temporal variability in the rainfall in the state is well-known.

4.2 Regression analysis for the trend component

To model the trend in the rainfall data, the regression model was used as indicated in Eq. (2). The parameters estimation, standard error, statistical t, p value and the coefficient of determination (R^2) for the monthly total rainfall for the 1994-2014 period are presented in Table 1.

Table 1: Estimation of the parameters of the model described in the Eq. (2) of the rainfall in the region analyzed for the period of 1994-2014.

Variable	Estimation	Standard error	t-value	p-value	R^2
Intercept	323.193	124.843	2.589	<0.01	0.29
cosx	-19.981	0.854	-23.392	<0.01	
senx	52.554	0.854	61.525	<0.01	
latitude	0.061	0.013	4.884	<0.01	
longitude	-2.308	0.079	-29.330	<0.01	
longitude ²	0.002	0.001	30.082	<0.01	

Considering the results presented in Table 1, at the 0.01 level of significance, the covariates cosx, senx, latitude, and longitude and the quadratic effect of the longitude are associated with precipitation. The coefficient of determination estimated by the model shows that 29% of the variation in the total monthly precipitation is explained in this trend model. This low coefficient of determination leads to strong indications that the trend component was not able to explain the spatio-temporal variability in the precipitation. It is also known that it is very difficult to provide a precise estimate for climatic variables, especially precipitation, because the spatial and temporal distribution presents high variability, causing the trend model to present a low R^2 (Raja *et al.*, 2017). It is noteworthy that the methodology proposed in this paper to estimate precipitation accounts not only for the trend but also for the dependence of spatio-temporal data. Next, we will model the residues produced by this trend component.

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

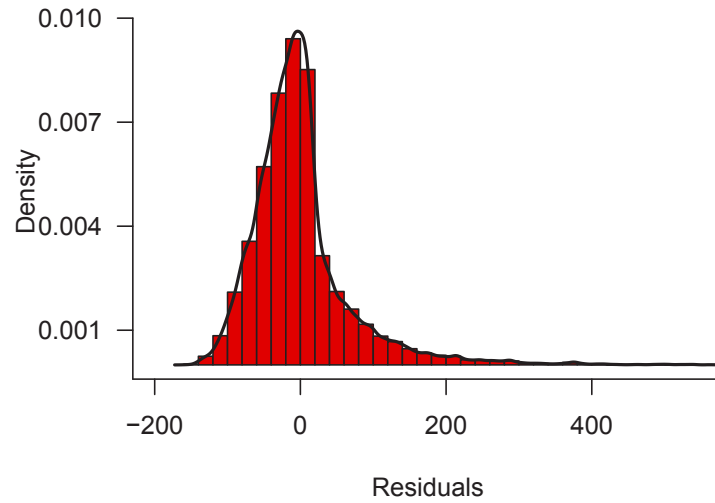


Figure 4: Histogram of the residues of the fitted model.

Next, the residuals from this regression were used to analyse the spatio-temporal autocorrelation by means of the variogram construction described in Eq. (5).

4.3 Spatiotemporal variogram of the residuals

After removing the spatiotemporal rainfall trend, the variogram is calculated using the residuals. The empirical spatiotemporal variogram model (Fig. 5a) and the fitted product sum model (Fig. 5b) are shown in Figure 5. As shown, the residues show a clear correlation for both space and time, and the total variation in these residues is explained by spatial and temporal components. Note that the spatial structure becomes weaker as the time differences increase, and the temporal structure becomes weaker as the spatial differences increase.

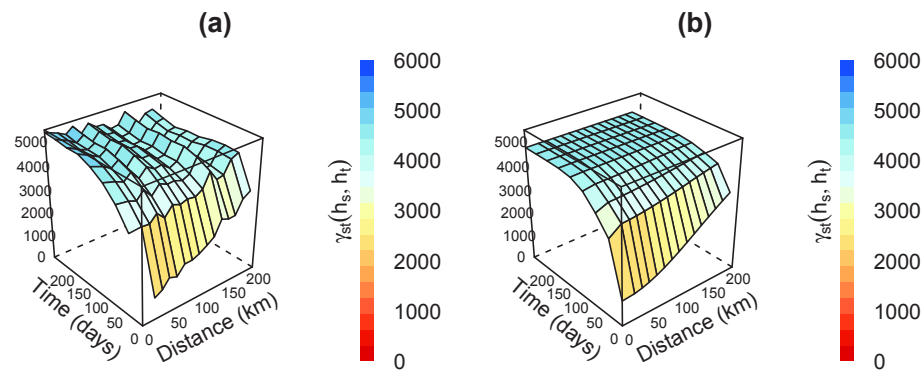


Figure 5: Sample spatiotemporal variogram (a) and the fitted product sum model (b), obtained from the residuals of the multiple linear regression.

The sample variogram can be visualized in terms of its marginal variograms, with one for space and another for time. The increasing trend in the spatial and temporal dimension in the sample variogram indicates the presence of a strong spatiotemporal correlation. Hu *et al.* (2015) stated that the characteristics present in Figure 5a further reinforce the need to apply spatiotemporal kriging to the residues. In this sample spatiotemporal variogram, different structures were fitted for the generalized product sum model class. Figure 5b shows one of these results when using the Gaussian model for the space and the Gaussian model for the time.

An exploratory analysis was performed in the empirical spatiotemporal variogram, to build the marginal variograms (Fig. 6).

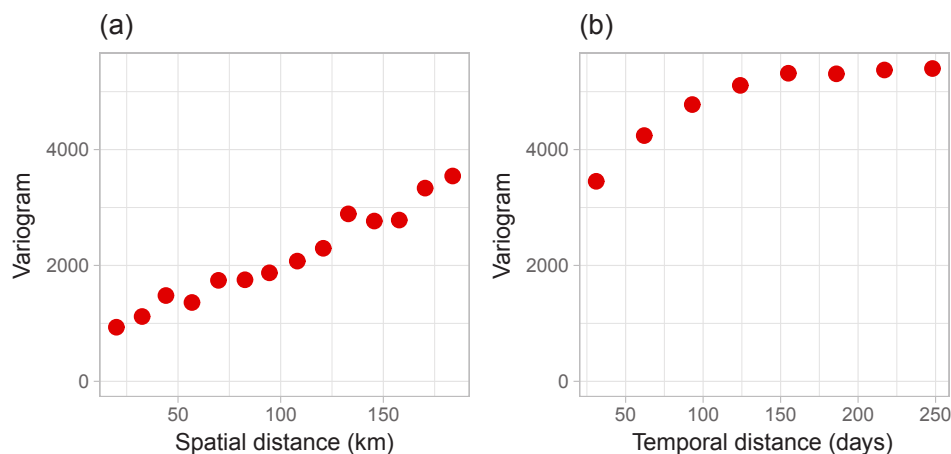


Figure 6: Marginal variograms of the residues: purely spatial (a) and purely temporal (b).

For the analysed data, the spatial variability exhibited by the spatial marginal variogram (Fig. 6a) is different from the temporal variability shown by the temporal marginal variogram (Fig. 6b), because the two variograms present different sill values. De Iaco *et al.* (2015) observed similar characteristics compared to those presented in Figures 6a and 6b and stated that, in this case, the spatiotemporal correlation of the residues can be adequately fitted by the generalized product-sum variogram model, as defined in Eq. (9). In this case, as mentioned above, the structures of the spatial and temporal correlation are noticeable. However, some studies related to the rainfall distribution do not present this significant correlation. Muthusamy *et al.* (2017) showed the development of the spatial variogram and did not show the spatial correlation of the rainfall data for the Bradford, England region.

In the adjustment of the generalized product-sum model, different variogram models were evaluated. For space and time, the following models were used: Gaussian (Gau), Exponential (Exp) and Spherical (Sph).

4.4 Model validation

The estimation of the parameters for the generalized product sum model is presented in Table 2, when considering the different structures for the spatial and temporal components. The RMSE was used to compare the models. Studies such as those presented in Hu *et al.* (2015), Iaco *et al.* (2015), Kilibarda *et al.* (2015) and Bahrami Jovein and Hosseini (2017) used only a single spatiotemporal correlation structure. In this study, a more comprehensive analysis was proposed by using different structures in the model formation.

Table 2: Parameters estimates of the product-sum variogram model fitted for the residues of the monthly rainfall regression.

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

For the generalized product-sum variogram model, the best structure (for a lower RMSE) was found when considering the Gaussian model for both the spatial and temporal components. According to the selected model, the parameters related to the range showed the spatial and temporal variation in the rainfall distribution of Paraíba. The spatial range of 180 km suggested that the spatial correlation became insignificant after this sill. On the other hand, the time interval of 71 days (approximately two months) indicated that the temporal correlation became insignificant after this interval.

After the best model was selected by variogram adjustment, the cross-validation leave-one-out was performed, and the results are presented below. In Figure 7, the spatiotemporal distribution of the mean RMSE values based on the 1994-2014 period is presented for each month, in which the high values indicate high variability between the estimated and observed amounts in the rainfall distribution.

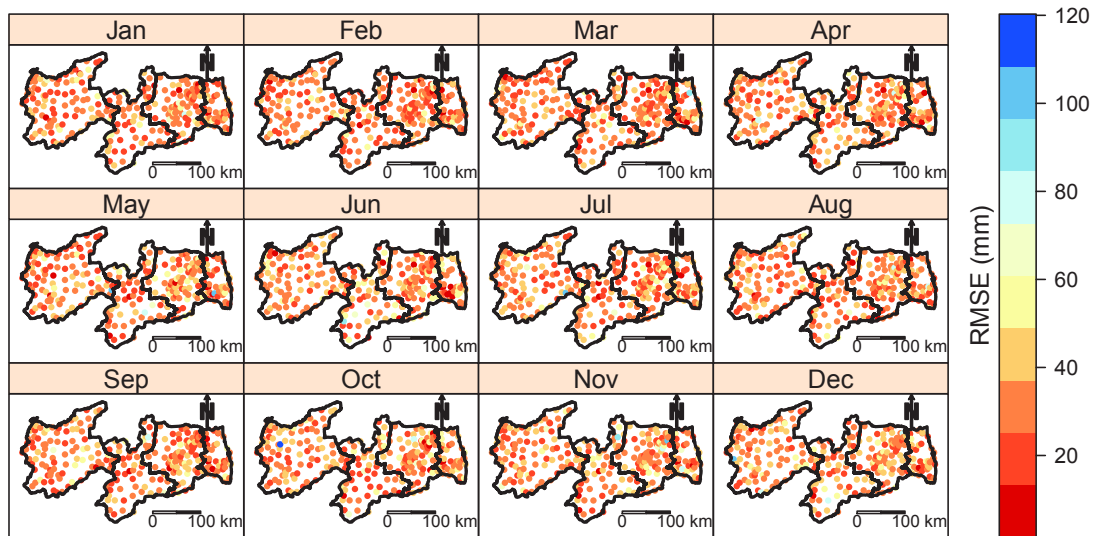


Figure 7: Spatial distribution of the mean value of RMSE of each station in the period of 1994-2014.

The highest values of the RMSE are shown between the months of January and May. In Sertão, from December to May, the greatest differences were noted between the predicted and observed rainfall values due to the occurrence of the rainy season in the region. These results are similar to those found by Menezes *et al.* (2010) and Gomes *et al.* (2015). This rainy season presents high spatial variability due to the presence of different weather systems, such as the Intertropical Convergence Zone (ITCZ) and the Upper Level Cyclonic Vortex (VCAN), as described by Macedo *et al.* (2010).

In most of the study region, the months from June to November presented the smallest RMSE, indicating that the predicted values were close to the observed rainfall values. These results are consistent with Gomes *et al.* (2015), who reported that the lowest rainfall rates in Paraíba occurred during these periods, causing the rainfall in most of the region to be evenly distributed. It is also possible to observe the long monthly statistical variation.

To evaluate the quality of fit for the model that was selected earlier, graphics have been created using all the values predicted by space-time kriging according to their actual values and the histogram of the difference between these values (Fig. 8), while considering all 269 pluviometric stations.

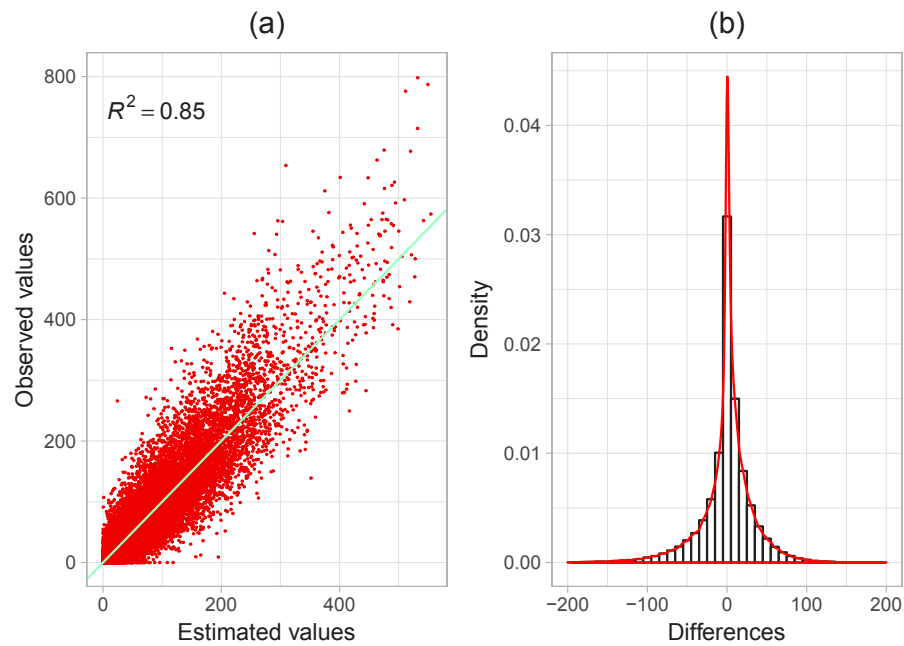


Figure 8: Comparison between the observed and estimated values (a) and the histogram of the difference between these values (b).

Figure 8 shows a strong linear relationship of the $y=x$ type between the estimated and observed values, which was supported by the coefficient of determination of 0.85. This finding indicates that 85% of the total precipitation variability can be explained by spatio-temporal kriging. Additionally, a visual analysis of the difference between these values, as presented in Fig. 8b, shows that the differences are distributed symmetrically with a frequency near zero. In other words, the scatter plot of the observed versus the estimated values during the 1994-2014 period as well as the histogram of the differences between these values (Fig. 8b) show that the predictions are not significantly biased.

4.5 Spatiotemporal kriging

The maps of the estimated rainfall for the month of January in the years 2015-2018 are shown in Figure 9.

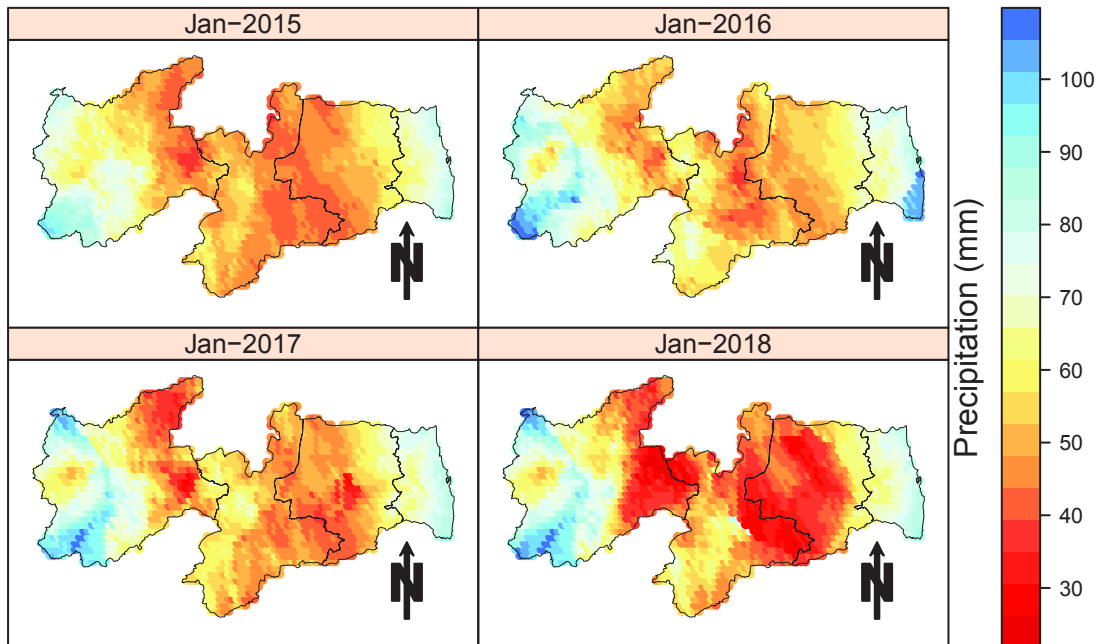


Figure 9: Prediction maps of precipitation in January 2015-2018.

The application of the methodology proposed in this study presents a great advantage, because enables the use of spatiotemporal kriging techniques to make predictions at any point in space-time within the chosen region. Thus, it allows for the evaluation of how the rainfall distribution patterns change in relation to space and time, besides presenting values in places that are not monitored by weather stations. This application is illustrated in Fig. 9, which represents the amount of spatiotemporal rainfall foreseen in Paraíba for the month of January in the years 2015-2018. January was chosen because it is the month that presents the highest rainfall values in the whole region, and it shows high variability in space and time. Rainfall presents a strong spatial and temporal pattern here, evidencing the success achieved when applying the geostatistical spatiotemporal methodology.

Regardless of the selected period, the highest rainfall level is distributed in the Zona da Mata mesoregion (Fig. 9). The rainfall in the north-south direction presents low variability. However, there is high variability in the east-west direction for this prediction. The greatest variation in the rainfall is diagnostic in Sertão of Paraíba.

Figure (10) shows the maps of the standard error for rainfall predictions in the month of January, in the years 2015-2020.

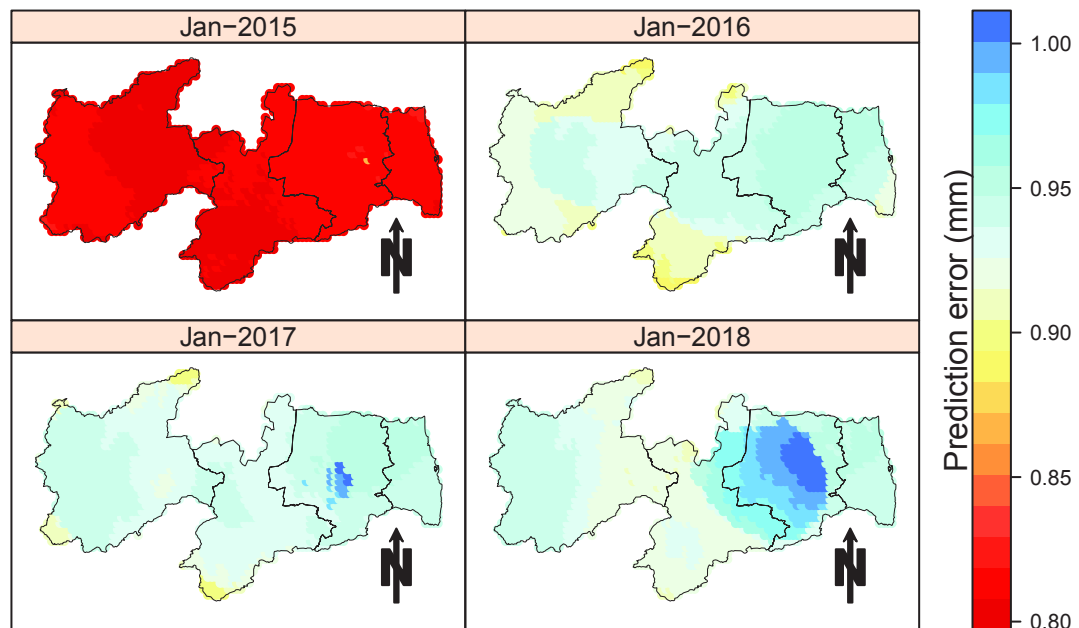


Figure 10: Maps of the standard error of precipitation prediction in January 2015-2018.

Along with the spatiotemporal kriging related to the predictions for January 2015-2018, the map of the standard error is given, which indicates the uncertainty (Fig. 10). The error presents variations in the east-west direction, and within each mesoregion, there is a variation in this estimation. It is also noted that the Borborema mesoregion presented the highest indexes for the standard error of prediction. This behaviour occurs because the rainfall in this area is more intense.

In Figures S1 and S2, which are available in the supplementary material, the prediction and the error values are shown for all the months of the 2015-2020 period.

5 Discussion

In this study, the spatiotemporal modelling of rainfall in the state of Paraíba was performed using well-known geostatistics techniques, such as kriging. For the state of Paraíba, some studies have already been performed involving the spatial (Santos *et al.*, 2012) or temporal (Dantas *et al.*, 2016) study of the rainfall distribution in the state, not taking into account the joint spatiotemporal variability of this phenomenon. Thus, this research presents the joint spatiotemporal study of the rainfall distribution in this region.

Here, the trend component represented 29% of the rainfall temporal variability. Studies by Gräler *et al.* (2012), Hengl *et al.* (2012) and Hu *et al.* (2017) indicated that the modelling of this component influences the fitting of the spatiotemporal variogram of the residues. A possible improvement in this study would be the consideration of environmental factors, such as the temperature, in the most adequate description of the trend component, probably resulting in an expected increase in the phenomenon variability percentage being modelled. Thus, the model for the trend somewhat explained the variability of the data, because it is less than half the variability of the precipitation space-time in the study area. Several studies have shown that including climatic variables in precipitation predictions generates more accurate models, reducing error metrics and increasing R^2 values (Apaydin *et al.*, 2011; Raja *et al.*, 2017). However, for the study area, the 269 pluviometric stations that were used here only have precipitation levels for the study region. In Brazil, there are other sources of data. For Paraíba state, there are only 8 meteorological stations, become impossible to use few stations in this work, because their data could negatively affect the predictions in the model. For future study, we can compare the results obtained in this article with the application of data obtained by simulation. Another important topic addressed in this research is that the variability in the data can be explained, along with the trend component, by fitting the covariance functions for the residuals obtained by trend adjustment, because they have space-time dependence (Kilibarda *et al.*, 2015; Martínez *et al.*, 2017).

The distribution of the rainfall in the state of Paraíba shows a clear correlation for both space and time, thus justifying the use of spatiotemporal kriging. When constructing the spatiotemporal variogram of the residues, several researchers found that the spatial dependence became weaker as the time increased, and the temporal dependence became insignificant for large spatial distances (Kilibarda *et al.*, 2015; Menezes *et al.*, 2010; Monteiro *et al.*, 2017; Tonini *et al.*, 2016). In the aforementioned studies, the phenomena that were modelled did not present a joint aspect in the spatiotemporal structure in the construction of this variogram. After the removal of the trend component, the rainfall data from Paraíba showed a clear spatial correlation structure for different temporal distances and a temporal correlation structure for different spatial distances. Regarding the theoretical aspect, some studies fitted models to the marginal sample variograms obtained from the empirical spatiotemporal variogram and then inserted these individual fittings into the product sum model (Bahrami Jovein and Hosseini, 2017; Iaco *et al.*, 2015). Nevertheless, this study proposes a fitting of the spatiotemporal

theoretical model, because the spatiotemporal interaction acts directly on the parameters estimations; when theoretical models are fitted to the marginal variograms, it is evident that there is no interference in the estimation from this interaction. However, this type of procedure results in problems in the optimization and determination of the parameter estimations. Studies using this type of fitting can be found in Hu *et al.* (2015) and Gräler *et al.* (2016).

This research represents a breakthrough for climatological studies because it employs more accurate techniques compared to traditional statistical techniques. The application of the statistical methods in this paper constitutes an important contribution and advance in understanding precipitation variability and the water resources of a region. The methodology considers that the spatio-temporal variability is decomposed by the trend component and the stochastic residue. However, the methodology presents some limitations, one of which is the difficulty in capturing the variability in locations that are separated by great distances between the locations sampled in the dataset. Additionally, for work involving this type of modelling, for the trend component to present better metrics, it will be necessary to include auxiliary variables, such as humidity and temperature.

The interpolation of climate data is widely used for research and management purposes in several areas of knowledge, including social science (Iaco *et al.*, 2015), hydrology (Martínez *et al.*, 2017) and forest science (Tonini *et al.*, 2016). The importance of modelling, both spatial and temporal, of the rainfall distribution in the state of Paraíba is extremely important for the management of priority areas that may present rainfall scarcity caused by periods of drought or excess rainfall, resulting in flooding, which implies environmental and socioeconomic damages (Tavares Diniz, 2013). The application of spatiotemporal kriging to obtaining rainfall records from unobserved sites should provide favourable conditions for the planning and management of water resources, because it will be possible to analyse regions that have a high impact on the state economy. Biondi (2013) performed spatiotemporal rainfall kriging and presented results that directly address the planning needs for the distribution of waters and natural resources once they have diagnosed the occurrence of drought in the river basin, because it is possible to analyse which geographic areas are more likely to be affected by severe droughts.

6 Conclusions

The rainfall mechanisms throughout the analysed region presented variability over both space and time. Thus, it was necessary to analyse the rainfall from the point of view of the

spatiotemporal interaction, uncovering information that includes the distribution of rainfall over space and time. The great advantage of the proposed modelling is that it is able to show interpolations for locations that were not observed at any time, including in the future. Finally, the rainfall estimations in the Paraíba region clarify the dramatic spatial and temporal variability of water resources in all the mesoregions of the state, thereby providing a detailed visual analysis of sectors undergoing rainfall scarcity and excess. These results should motivate further research, including along the north-eastern region of Brazil, with the aim of making predictions for diagnosing which sites might experience droughts or floods.

In this paper, a data analysis of rainfall is proposed through spatiotemporal geostatistics, and it was applied to irregularly spaced weather stations by considering rainfall distributions over both space and time. Notably, the rainfall variability may be related to spatial and weather covariates, such as the temperature, humidity, and altitude, among others, suggesting the addition of covariates for future studies to model the effect of the rainfall trend. Moreover, the monthly rainfall in the analysed region presented a series of records with null values, which affects, to a certain extent, the application of a normal regression model. Exploring these issues in detail will require a non-normal model that allows for a better explanation regarding the variability of rainfall in the region. This model will then be the next step of the investigation, and it will be reported in the near future.

8 Acknowledgements

This work was conducted during a scholarship supported by the Coordination of Improvement of Higher Education Personnel (CAPES) at Federal University of Lavras.

References

- Apaydin H, Anli AS, Ozturk F. 2011. Evaluation of topographical and geographical effects on some climatic parameters in the Central Anatolia Region of Turkey. *International Journal of Climatology* **31**(9): 1264–1279. DOI: 10.1002/joc.2154.
- Bahrami Jovein E, Hosseini SM. 2017. Predicting saltwater intrusion into aquifers in vicinity of deserts using spatio-temporal kriging. *Environmental Monitoring and Assessment*. *Environmental Monitoring and Assessment* **189**(2). DOI: 10.1007/s10661-017-5795-8.

Bakar KS, Sahu SK. 2015. spTimer: Spatio-Temporal Bayesian Modeling Using R. *Journal of Statistical Software* **63**(15): 32. DOI: 10.18637/jss.v063.i15.

Banerjee S, Bradley CP, Alan GE. 2014. *Hierarchical Modeling and Analysis for Apatial Data, Second Edition*. Chapman and Hall/CRC: Boca Raton.

Biondi F. 2013. Space-time kriging extension of precipitation variability at 12 km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrology and Earth System Sciences Discussions* **10**(4): 4301–4335. DOI: 10.5194/hessd-10-4301-2013.

Cameletti MBM. 2015. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. JohnWiley & Sons, Ltd Registered. DOI: 10.1017/CBO9781107415324.004.

Cesare L De, Myers DE, Posa D. 2001. Estimating and modeling space-time correlation structures. *Statistics and Probability Letters* **51**(1): 9–14. DOI: 10.1016/S0167-7152(00)00131-0.

Cressie N, Wikle CK. 2015. *Statistics for spatio-temporal data*. John Wiley and Sons: Hoboken.

Dantas LG, Santos CAC dos, Olinda RA. 2016. *Revista Brasileira de Geografia Física* **9**: 997–1006. DOI: 10.5935/1984-2295.20140015.

Gomes OM, Dos Santos CAC, De Souza FDAS, De Paiva W, Olinda RA. 2015. Análise comparativa da precipitação no estado da paraíba utilizando modelos de regressão polinomial. *Revista Brasileira de Meteorologia* **30**(1): 47–58. DOI: 10.1590/0102-778620120454.

Gräler B, Gerharz L, Pebesma E. 2012. Spatio-temporal analysis and interpolation of PM10 measurements in Europe. *ETC/ACM Technical Paper 2011/10* (November 2011): 37.

Gräler B, Pebesma E, Heuvelink G. 2016. Spatio-Temporal Interpolation using gstat. *Wp* **8**: 1–20. DOI: 10.1007/978-3-319-17885-1.

Hengl T, Heuvelink GBM, Tadić MP, Pebesma EJ. 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology* **107**(1–2): 265–277. DOI: 10.1007/s00704-011-0464-2.

Heuvelink GBM, Griffith DA. 2010. Space-time geostatistics for geography: A case study of radiation monitoring across parts of Germany. *Geographical Analysis* **42**(2): 161–179. DOI: 10.1111/j.1538-4632.2010.00788.x.

Heuvelink GBM, Griffith DA, Hengl T, Melles SJ. 2012. Sampling Design Optimization for Space-Time Kriging. *Spatio-temporal Design: Advances in Efficient Data Acquisition* (OCTOBER): 207–230. DOI: 10.1002/9781118441862.ch9.

Hu D, Shu H, Hu H, Xu J. 2017. Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Cluster Computing* **20**(1): 347–357. DOI: 10.1007/s10586-016-0708-0.

Hu Y, Li R, Bergquist R, Lynn H, Gao F, Wang Q, Zhang S, Sun L, Zhang Z, Jiang Q. 2015. Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. *PLoS Neglected Tropical Diseases* **9**(2). DOI: 10.1371/journal.pntd.0003470.

Huang HC, Martinez F, Mateu J, Montes F. 2007. Model comparison and selection for stationary space-time models. *Computational Statistics and Data Analysis* **51**(9): 4577–4596. DOI: 10.1016/j.csda.2006.07.038.

Iaco S, Myers DE, Posa D. 2001. Space-time analysis using a general product-sum model. *Statistics and Probability Letters* **52**(1): 21–28. DOI: 10.1016/S0167-7152(00)00200-5.

Iaco S, Palma M, Posa D. 2015. Spatio-temporal geostatistical modeling for French fertility predictions. *Spatial Statistics* **14**: 546–562. DOI: 10.1016/j.spasta.2015.10.002.

Kilibarda M, Hengl T, Heuvelink GBM, Gräler B, Pebesmatadić EP, Bajat B. 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research* **119**(5): 2294–2313. DOI: 10.1002/2013JD020803.

Kilibarda M, Tadić MP, Hengl T, Luković J, Bajat B. 2015. Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. *Spatial Statistics*. Elsevier Ltd **14**: 22–38. DOI: 10.1016/j.spasta.2015.04.005.

Macedo MJH, Guedes RVS, Sousa FAS, Dantas FRC. 2010. Análise da índice padronizado de precipitação para o estado da Paraíba, Brasil. *Revista Ambiente e Agua* **5**(1): 204–214. DOI: 10.4136/1980-993X.

Martínez WA, Melo CE, Melo OO. 2017. Median Polish Kriging for space–time analysis of precipitation. *Spatial Statistics* **19**: 1–20. DOI: 10.1016/j.spasta.2016.10.003.

Menezes HE a., Brito JIB De, Lima R a. FD a. 2010. Veranico e a produção agrícola no Estado da Paraíba, Brasil. *Revista Brasileira de Engenharia Agrícola e Ambiental* **14**(2): 181–186. DOI: 10.1590/S1415-43662010000200009.

Monteiro A, Menezes R, Silva ME. 2017. Modelling spatio-temporal data with multiple seasonalities: The NO₂ Portuguese case. *Spatial Statistics*. DOI: 10.1016/j.spasta.2017.04.005.

Muthusamy M, Schellart A, Tait S, Heuvelink GBM. 2017. Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models. *Hydrology and Earth System Sciences* **21**(2): 1077–1091. DOI: 10.5194/hess-21-1077-2017.

Pebesma E. 2012. Spacetime: Spatio-Temporal Data in R. *Journal of Statistical Software* **51**(7): 1–30. DOI: 10.18637/jss.v051.i07.

Raja NB, Aydin O, Türkoğlu N, Çiçek I. 2017. Space-time kriging of precipitation variability in Turkey for the period 1976–2010. *Theoretical and Applied Climatology* **129**(1–2): 293–304. DOI: 10.1007/s00704-016-1788-8.

Santos CAC dos, Gomes OM, Souza F de AS de, Paiva W De. 2012. Análise Geoestatística da Precipitação Pluvial do Estado da Paraíba. *Revista Brasileira de Geografia Física* **4**(4): 692–702. DOI: 10.26848/rbgf.v4i4.232712.

Schlatter M, Malinowski A, Menck PJ. 2015. Analysis, Simulation and Prediction of Multivariate Random Fields with Package RandomFields. *Journal of Statistical Software* **63**: 1–25. DOI: <http://dx.doi.org/10.18637/jss.v063.i08>.

Sherman M. 2010. *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. DOI: 10.1002/9780470974391.

Tavares Diniz JM. 2013. Variabilidade da precipitação e do número de dias com chuvas de duas cidades distintas da Paraíba. *Holos* **3**(0): 171. DOI: 10.15628/holos.2013.1291.

Tonini F, Dillon WW, Money ES, Meentemeyer RK. 2016. Spatio-temporal reconstruction of missing forest microclimate measurements. *Agricultural and Forest Meteorology*. Elsevier B.V. **218–219**: 1–10. DOI: 10.1016/j.agrformet.2015.11.004.

Supplementary material

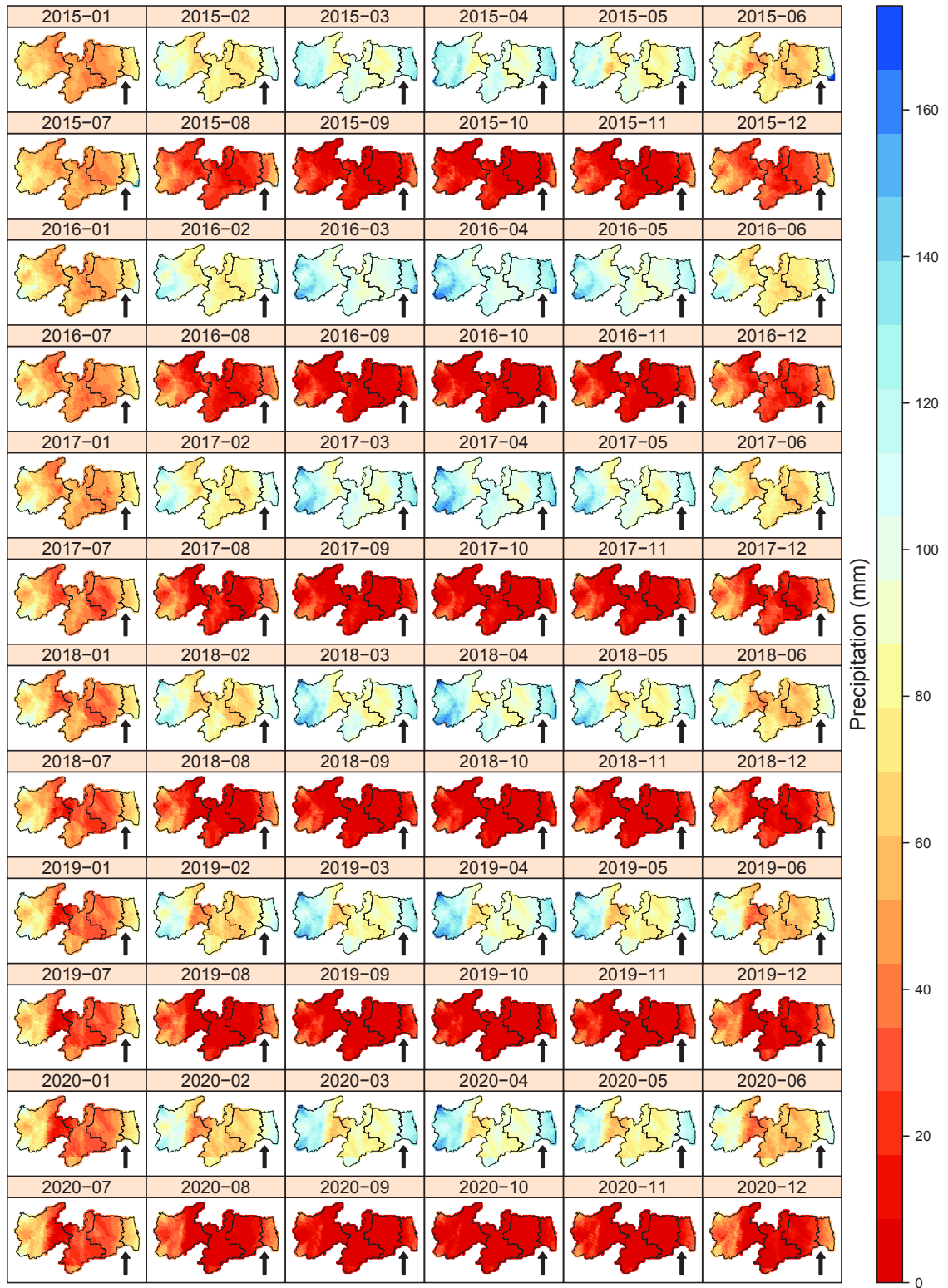


Figure S1: Prediction maps of precipitation for the period from 2015 to 2020.

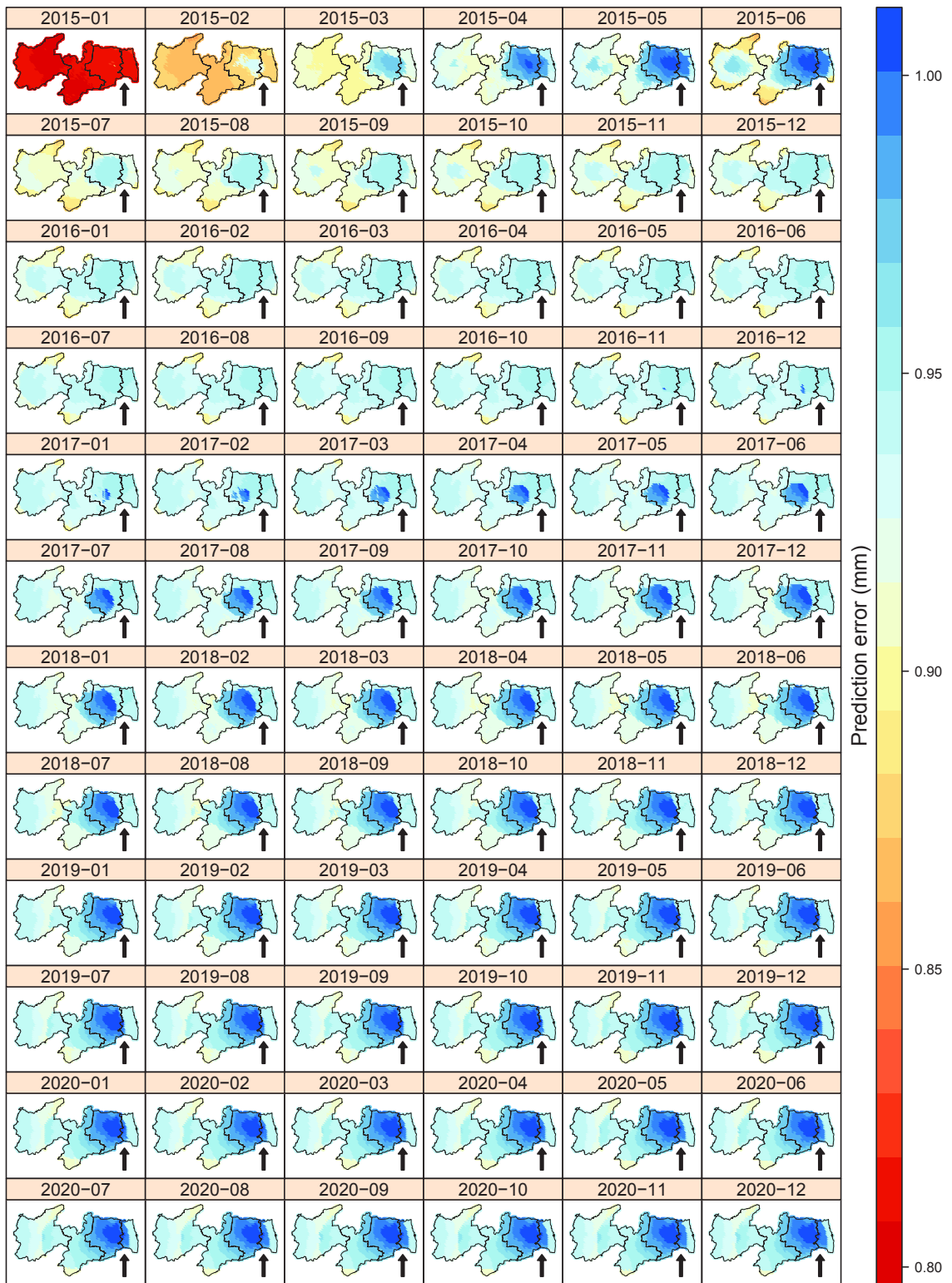


Figure S2: Maps of the standard error of precipitation prediction in the period from 2015 to 2020.

SEGUNDA PARTE - ARTIGOS**ARTIGO 2**

Space-time kriging of precipitation: modeling the large-scale variation with GAMLSS

**Artigo redigido conforme as normas da
revista *Theoretical and Applied
Climatology*
(versão submetida)**

Space-time kriging of precipitation: modeling the large-scale variation with GAMLSS model

Elias Silva de Medeiros¹, Renato Ribeiro de Lima¹, Ricardo Alves de Olinda², Carlos Antônio Costa dos Santos³

¹Department of Statistic, Federal University of Lavras, Lavras, Minas Gerais, Brazil

²Department of Statistic, Paraíba State University, Campina Grande, Paraíba, Brazil

³Center for Technology and Natural Resources, Federal University of Campina Grande, Campina Grande, Paraíba, Brazil

Correspondence to: Elias Silva de Medeiros, Department of Statistic, Federal University of Lavras, Lavras, Minas Gerais, 37200-000, Brazil. (eliasestatistica@gmail.com).

Abstract

Knowing the dynamics of the space-time distribution of precipitation is of paramount importance for water resources management. To model the large-scale variation of precipitation, several models have been based on the normal distribution, thus without to consider into account the excess of zeros observations, which is common in a daily time recording or even in monthly. This work proposes a new approach for modeling the trend component using a zero inflated gamma distribution. The residuals of this regression usually have space-time dependence and were modeled by a space-time covariance function. The results demonstrate that our approach produced reliable and accurate estimates for precipitation, surpassing the methods which have been used in the literature. The modeling provided estimates of precipitation in non-sampled locations and in unobserved times, thus serving as a tool to help government agencies related to water management to anticipate society needs and prevent water crises.

Keywords: ZAGA model, spatio-temporal dependence, trend component.

1 - Introduction

With interest in analyzing a phenomenon together in space and time, works have emerged in several areas of knowledge, especially in the climate sciences (Kilibarda et al. 2015; Martínez et al. 2017; Stauffer et al. 2017). It is well known that the spatial-temporal behavior of rainfall is of great importance for the water resources of a region and that it has a direct influence on the human activities, such as agriculture and commerce (Marengo et al. 2017b).

The northeast region of Brazil (NEB) is characterized by irregular distribution of rainfall in space and time, causing high spatial and temporal variability of precipitation. In this region it is common the occurrence of high rainfall indices in a location and no record of rains in its surroundings (Borges et al. 2014). The state of Paraíba, located in this region, presents the same characteristics, with periods of drought in rainy seasons with behavior different from the

precipitation between the mesoregions that constitute the region (Almeida and Medeiros 2017). Consequently, this region presents a problem with regard to the spatio-temporal variability of rainfall.

In the literature, several researches have presented some solutions about the precipitation study in this region. For this, the papers use spatial statistics techniques (Santos et al. 2012; Francisco et al. 2016). However, the above-mentioned works did not take spatial and temporal dependence into account simultaneously. In view of the above, it is proposed in this research to model the dynamics and space-time dependence of precipitation by means of a methodology of spatial statistics (Cressie and Wikle 2015; Iaco et al. 2015). First, the trend component will be fitted, responsible for the large-scale variation of the phenomenon, and then a space-time geostatistics analysis is performed on the residuals that are autocorrelated in space-time. Several works used multiple linear regression to model the deterministic component, considering the Gaussian model (Gräler et al. 2012; Hengl et al. 2012; Hu et al. 2017). These works did not assume that the variable response to be studied presented a distribution that could not be correctly modeled by a Gaussian model.

However, recent studies have been approached in this modeling, structures of non-Gaussian models, considering Generalized Linear Models (GLM) with variable response Gamma. Once the Gamma model is assumed, the variable domain has to be strictly positive and it does not include zero. Among these works Menezes et al. (2016) and Monteiro et al. (2017) analyzed the concentration of NO_2 , which had asymmetric behavior with excesses of zero. However, when using GLM with Gamma distribution, it is not possible to model the frequency of zeroes in the variable under study. To prevent this limitation, there are models that capture the frequency of zeroes in continuous variables. Stauffer et al. (2017) analyzed the spatial-temporal distribution of daily precipitation at 117 stations located in Tirol, Austria, over a 42-year period. These authors drew attention to the high frequency of zero observations in the data, suggesting the use of the normal censored model of zero, however, they modeled only the trend component, not taking into account the space-time dependence of precipitation in this region.

In this research, the analyzed variable will be the total monthly rainfall measured in the State of Paraíba, Brazil, which presents an asymmetric behavior and has a high occurrence of zeros, due to the climatic characteristic in the region under study. Thus, to model this behavior, a new approach in the regression fitted is proposed through generalized additive models for

location, scale and shape – GAMLSS (Stasinopoulos et al. 2017). These models have several advantages, such as: all distribution parameters can be modeled as covariate functions; accommodates cases of overdispersion and excess of zeroes in the data; allows to adjust of any distribution to the response variable. In this work, the Zero Adjusted Gamma (ZAGA) distribution was used. The ZAGA distribution is an extension of the Gamma distribution and has three parameters, one of which models the probabilities for zero occurrences. This is an innovative approach since we did not find works in the literature, which used the GAMLSS models to model the trend in space-time geostatistics.

For a preliminary data analysis of the data, graphs will be constructed to verify the behavior of the precipitation in general and for each pluviometric station. It will also be analyzed if the precipitation distribution is associated with spatial (longitude and latitude) and temporal (temporal index) factors. Significant covariables will compose the trend component adjusted by GAMLSS. The spatial and temporal variability of this regression residuals will be modeled using geostatistical techniques. Then, space-time kriging will allow the reconstruction of the space-time pattern of precipitation in the study area. This methodology will provide information for an adequate evaluation of the phenomenon in non-sampled places (interpolation) and in times not observed (prediction).

2 - Material

The region of study comprises the Paraíba State, Brazil (Figure 01), located in the NEB, which is located between the parallels 6° and 8° of south latitude and between the meridians of 34° and 39° west longitude, with an area of approximately 56,500 km^2 . Paraíba belongs to the Tropical Region and its area is divided into four mesoregions: Zona da Mata, Agreste, Borborema and Sertão.

In this work, the total monthly precipitation data measured in 54 pluviometric stations will be considered during the period from 1994 to 2014. Table 1 shows the identification number, name of pluviometric station, latitude, longitude, altitude above of sea level and quantiles of precipitation (mm) at 2.5% , 50% and 97.5%.

Table 1: Information of the pluviometric stations with identification number (ID), name of the mesoregion, name of rainfall station, latitude (Lat), longitude (Long), Altitude (Alt) and precipitation quantiles 2.5%, 50%, 97.5 %.

ID	Mesoregion	Station	Lat	Long	Alt (m)	Quantiles		
						2.5%	50.0%	97.5%
1	Zona da Mata	Alhandra	-7.43	-34.91	56.28	6.49	119.85	481.71
2	Zona da Mata	Jacaraú	-6.61	-35.29	181.19	0.88	55.90	311.01
3	Zona da Mata	Mamanguape	-6.84	-35.12	16.62	2.60	66.70	403.02
4	Zona da Mata	Pedras de Fogo	-7.40	-35.12	175.36	4.71	79.25	399.52
5	Zona da Mata	Sapé	-7.09	-35.22	116.84	1.86	64.35	324.14
6	Agreste	Alagoinha	-6.96	-35.55	168.64	0.00	70.50	306.56
7	Agreste	Araçagi	-6.83	-35.39	104.62	0.00	56.95	290.73
8	Agreste	Araruna	-6.53	-35.74	575.41	0.00	53.40	235.32
9	Agreste	Areia	-6.98	-35.72	572.28	4.34	95.55	331.73
10	Agreste	Areial	-7.05	-35.93	693.09	0.00	38.55	182.72
11	Agreste	Boa Vista	-7.26	-36.24	486.00	0.00	20.30	127.07
12	Agreste	Campina Grande	-7.23	-35.90	544.37	0.93	48.00	244.28
13	Agreste	Casserengue	-6.79	-35.89	396.63	0.00	18.00	137.88
14	Agreste	Cuité	-6.49	-36.15	668.96	0.00	33.35	190.79
15	Agreste	Dona Inês	-6.61	-35.63	421.72	0.00	51.25	242.04
16	Agreste	Ingá	-7.29	-35.61	155.56	0.00	44.35	191.81
17	Agreste	Mogeiro	-7.31	-35.48	108.28	0.00	40.95	194.68
18	Agreste	Pocinhos	-7.08	-36.06	650.32	0.00	21.45	140.05
19	Agreste	Soledade	-7.06	-36.36	523.63	0.00	17.75	139.61
20	Borborema	Barra de S. M.	-7.75	-36.32	488.63	0.00	15.30	142.21
21	Borborema	Boqueirão	-7.49	-36.14	355.08	0.00	21.65	148.11
22	Borborema	Camalaú	-7.89	-36.83	519.16	0.00	11.15	185.53
23	Borborema	Caraúbas	-7.73	-36.49	442.24	0.00	10.60	161.65
24	Borborema	Congo	-7.80	-36.66	491.99	0.00	12.25	164.82
25	Borborema	Juazeirinho	-7.07	-36.58	553.96	0.00	20.10	175.97
26	Borborema	Junco do Seridó	-7.00	-36.71	589.56	0.00	22.40	210.18
27	Borborema	Pedra Lavrada	-6.76	-36.46	521.82	0.00	13.55	180.72
28	Borborema	Prata	-7.70	-37.08	584.00	0.00	18.85	220.19
29	Borborema	Salgadinho	-7.10	-36.85	430.45	0.00	13.60	240.85
30	Borborema	Santa Luzia	-6.87	-36.92	311.24	0.00	9.80	246.99
31	Borborema	São João do Tigre	-8.08	-36.85	572.84	0.00	13.80	161.03
32	Borborema	São J. dos Cord.	-7.39	-36.81	530.49	0.00	14.30	313.51
33	Borborema	São S. do Umb.	-8.15	-37.01	595.79	0.00	20.10	200.60
34	Borborema	Sumé	-7.67	-36.90	519.88	0.00	15.70	264.53
35	Borborema	Várzea	-6.77	-36.99	267.64	0.00	10.65	270.29
36	Sertão	Água Branca	-7.51	-37.64	732.80	0.00	41.40	274.52
37	Sertão	Bom Sucesso	-6.44	-37.93	289.12	0.00	27.20	296.80
38	Sertão	Brejo do Cruz	-6.35	-37.50	200.67	0.00	27.70	320.60
39	Sertão	Cajazeiras	-6.89	-38.54	299.44	0.00	38.80	409.20
40	Sertão	Catolé do Rocha	-6.34	-37.75	298.89	0.00	34.55	301.66
41	Sertão	Conceição	-7.56	-38.50	388.28	0.00	27.90	272.48
42	Sertão	Condado	-6.92	-37.59	260.82	0.00	23.75	336.19

43	Sertão	Lagoa	-6.59	-37.91	275.56	0.00	31.30	314.85
44	Sertão	Mãe D'Água	-7.26	-37.43	411.10	0.00	15.65	277.71
45	Sertão	Manaíra	-7.71	-38.15	767.40	0.00	28.95	284.65
46	Sertão	Nova Olinda	-7.48	-38.04	321.00	0.00	27.55	326.82
47	Sertão	Passagem	-7.14	-37.05	305.04	0.00	14.65	273.24
48	Sertão	Patos	-7.00	-37.31	256.69	0.00	26.25	303.92
49	Sertão	Piancó	-7.21	-37.93	261.16	0.00	24.90	321.85
50	Sertão	Pombal	-6.77	-37.80	191.56	0.00	22.85	353.35
51	Sertão	Riacho dos C.	-6.44	-37.65	206.93	0.00	24.45	269.48
52	Sertão	Santa Teresinha	-7.08	-37.45	307.16	0.00	22.55	368.42
53	Sertão	São J. R. Peixe	-6.73	-38.45	248.20	0.00	31.90	325.64
54	Sertão	Sousa	-6.77	-38.22	235.44	0.00	30.30	327.17

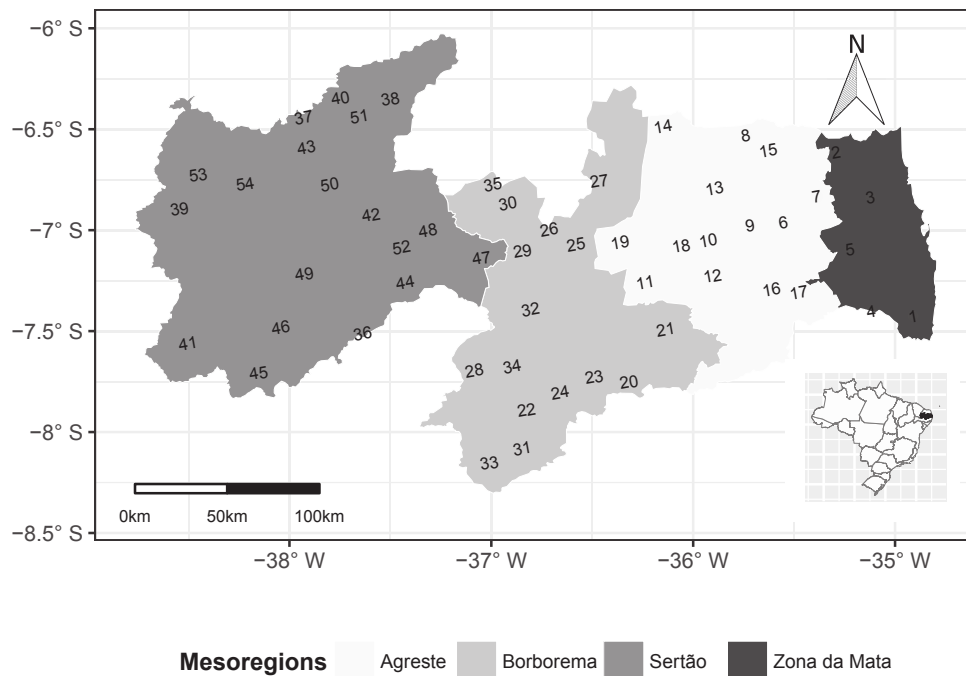


Figure 1 - Location of the 54 pluviometric stations in the studied area distributed in the four mesoregions of Paraíba, Brazil. The information for each respective number is presented in Table 1.

3 - Methods

In space-time geostatistics, observations are modeled as a stochastic process by means of a random function $\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq R^d \times R\}$. In this work it is assumed that Z has the first and second moments. Thus, $Z(\mathbf{s}, t)$ is decomposed by the sum of the trend components and the stochastic residual:

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

In Equation (1), the trend component $m(\mathbf{s}, t) = E[Z(\mathbf{s}, t)]$ is not constant in space and time, representing the large scale of variation. The residual component $\varepsilon(\mathbf{s}, t)$ describes random fluctuations on a small scale and encompasses the three components: spatial, temporal and interaction.

3.1 - Trend component

The trend component is responsible for the large-scale variation of the stochastic process. In spatio-temporal datasets this component is not a deterministic function, so it is necessary to specify stochastic patterns that represent the observed variability. Thus, in the specification of the trend component were included the geographic coordinates themselves as covariables, as well as time. The purpose of this paper is to fit this component using GAMLSS models. In the GAMLSS models it is assumed that the observations z_i are independent, with $i = 1, 2, \dots, n$ and conditioned to $\boldsymbol{\theta}_k = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\tau})^T$, with $k = 1, 2, 3, 4$. So, $Z|\boldsymbol{\theta}_k \sim G(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\tau})$, where G represents Z distribution. The parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{v}, \boldsymbol{\tau}$ can be classified as the parameters of location, scale, shape and kurtosis, respectively, where the latter two represents the distribution shape (Stasinopoulos et al. 2017). The GAMLSS models are expressed in the form,

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Y}_{jk} \boldsymbol{\phi}_{jk}, \quad (2)$$

where $g_k(\cdot)$ is a monotonic link function, and $\boldsymbol{\theta}_k$ is related to explanatory variables and random effects; $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of size n ; $\boldsymbol{\beta}_k$ is a parameter vector of size J'_k ; \mathbf{X}_k and \mathbf{Y}_{jk} are the fixed (covariate) fixed matrices of size $n \times J'_k$ and $n \times q_{ij}$, respectively; $\boldsymbol{\gamma}_{jk}$ is a random variable of dimension q_{ij} . In a purely parametric linear model we have that $\sum_{j=1}^{J_k} \mathbf{Y}_{jk} \boldsymbol{\phi}_{jk} = 0$, that is, there are no additive terms associated with the distribution parameters in the model. In the space-time case, we have $\boldsymbol{\eta}(\mathbf{s}, t) = g(m(\mathbf{s}, t))$.

In the trend component modeling we considered $\mathbf{Y}_{jk} = \mathbf{I}_n$, where \mathbf{I}_n is the identity matrix of order $n \times n$ and $\boldsymbol{\phi}_{jk} = \mathbf{v}_{jk} = v_{jk}(x_{jk})$. Thus, Equation (2) is rewritten as:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} v_{jk}(x_{jk}). \quad (3)$$

The model given in (3) was fitted to the data of this research, where \mathbf{v}_{jk} is an unknown function of the explanatory variable x_{jk} and $\mathbf{v}_{jk}(x_{jk})$ is a vector that evaluates the function v_{jk}

in x_{jk} . In the literature, this model is called the semiparametric additive GAMLSS model and may contain parametric, non-parametric and random effects terms (Stasinopoulos et al. 2017).

In this work, the geographic coordinates (latitude and longitude) and the temporal index (time) were used as covariables to model the trend. Time index was constructed as follows: An index equal to 1 denoted Jan/1994, the index equal to 2 denoted Feb/1994, and so on. In all covariables, cubic splines (cs) were inserted, including the covariate time that was designed to model the seasonal effect of precipitation. For the trend component were fitted several models, and through the Generalised Akaike information criterion (GAIC), the selected model for the trend is given by:

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \ln(\boldsymbol{\mu}) = \beta_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} \\ g_2(\boldsymbol{\sigma}) &= \beta_{20} \\ g_3(\mathbf{v}) &= \ln\left(\frac{\mathbf{v}}{1-\mathbf{v}}\right) = \beta_{30}x_{13}, \end{aligned} \quad (4)$$

where x_{11} , x_{12} and x_{13} represent the covariables of longitude, latitude and time, respectively. The parameters β_{10} , β_{11} , β_{12} and β_{13} are related to the location vector of the GAMLSS model with log link function. β_{20} is the parameter related to the intercept for the scale vector with identical link function. β_{30} is the parameter associated to covariable time and that is responsible for modeling the occurrence or no zeroes in the response variable with logit link function.

To obtain the parameter estimates for this regression model, the penalized maximum likelihood method was used. This differs from conventional methods because it assigns the value zero to the coefficients of non-significant variables. However, this method does not take into account the assumption that errors are uncorrelated. Thus, it is necessary to model the residuals of this model that takes into account this correlation structure. For this, it is proposed to fit a valid non-separable space-time variogram model.

3.2 - Spatio-temporal variogram

When inserting structures in the trend component, one has the advantage that it contemplates the large-scale variation of the phenomenon being studied, causing the residuals to present small scale variations, consequently reducing the prediction error in the kriging. After obtaining the estimates of the trend component parameters, the stochastic residual is obtained by difference, $\varepsilon(\mathbf{s}, t) = Z(\mathbf{s}, t) - m(\mathbf{s}, t)$. Therefore, the next step of the analysis is the

empirical estimation of the covariance function or spatio-temporal variogram to model the residuals (Martínez et al. 2017).

Spatio-temporal covariates are usually described using a space-time variogram (γ_{st}), which measures the mean difference between separated data in the space-time domain using the distance vector. The variogram estimate, obtained by the moments method, is given by:

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

where $|L(\mathbf{r}_s, r_t)|$ is the cardinality for the set $L(\mathbf{r}_s, r_t)$. It is noted in this equation that the empirical variogram depends only on spatial and temporal distances, and this may not result in a valid variogram. Thus, to obtain estimates of γ_{st} in any arbitrary lag, the empirical variogram must be smoothed. As a solution to this limitation found in the empirical variogram, one should use a theoretical model, $\gamma_{st}(\mathbf{h}_s, h_t, \boldsymbol{\omega})$, which fits, as best as possible, the space-time dependence structure of the residuals (Montero et al. 2015). The vector $\boldsymbol{\omega}$ contains all the unknown parameters to be estimated.

The common approach is to fit a theoretical model to the empirical variogram. The least squares estimation method is the most trivial approach. However, there are other methods such as: maximum likelihood, generalized estimation equations, composite likelihood, among others (Schabenberger and Gotway 2004). In this work, the least squares method was used.

In the least method squares there are no assumptions about the probability distribution of the generated values of the empirical variogram. The method considers the statistical model of the form $\hat{\gamma}(\mathbf{h}_s, h_t) = \gamma(\mathbf{h}_s, h_t) + \varepsilon'(\mathbf{h}_s, h_t)$. Assume that $\varepsilon'(\mathbf{h}_s, h_t)$ has null mean and covariance matrix $\mathbf{R} = \mathbf{R}(\boldsymbol{\omega})$, which depends on $\boldsymbol{\omega}$. In this work it is suggested that the diagonal of $\mathbf{R}(\boldsymbol{\omega})$ be approximated to:

$$\frac{\text{Var}[\gamma(\mathbf{h}_{s_i}, h_{t_j})]}{2} \approx \frac{\gamma(\mathbf{h}_{s_i}, h_{t_j}, \boldsymbol{\omega})^2}{|N(\mathbf{h}_{s_i}, h_{t_j})|}, \quad (6)$$

where $N(\mathbf{h}_{s_i}, h_{t_j})$ is the number of pairs at each space-time distance. Thus, the weights assigned in the weighting matrix will be proportional to the number of pairs in a given space-time distance.

In this work, in the adjustment of the theoretical space-time variogram model, the weighted least squares (WLS) method, which replaces $\mathbf{R}(\boldsymbol{\omega})$ with the diagonal matrix $\mathbf{W}(\boldsymbol{\omega})$. The elements of $\mathbf{W}(\boldsymbol{\omega})$ are obtained from Equation (6). Thus, using the weights given in (6),

the WLS method is given by:

$$\begin{aligned} & (\boldsymbol{\gamma}(\mathbf{h}_s, h_t) - \boldsymbol{\gamma}(\mathbf{h}_s, h_t, \boldsymbol{\omega}))' \mathbf{W}(\boldsymbol{\omega})^{-1} (\boldsymbol{\gamma}(\mathbf{h}_s, h_t) - \boldsymbol{\gamma}(\mathbf{h}_s, h_t, \boldsymbol{\omega})) \\ &= \sum_{i=1}^I \sum_{j=1}^J \frac{|N(\mathbf{h}_{si}, h_{tj})|}{\gamma(\mathbf{h}_{si}, h_{tj}, \boldsymbol{\omega})^2} \{ \gamma(\mathbf{h}_{si}, h_{tj}) - \gamma(\mathbf{h}_{si}, h_{tj}, \boldsymbol{\omega}) \}^2. \end{aligned}$$

Generally, statistical assumptions are considered, which are defined in the literature, which comprise a space-time covariance function, even, ensure that this function is defined positive (Iaco et al. 2001; Huang et al. 2007; Cressie and Wikle 2015). Nowadays, the construction of valid space-time covariance functions has been the subject of study by several researchers (Iaco et al. 2002; Kilibarda et al. 2014; Montero et al. 2015). In addition, although this type of construction is a difficult task, the complexity becomes greater when it is desired to determine valid models of space-time covariance. In the space-time approach there are basically two types of general model classes: the separable and the non-separable.

The separable models assume that spatial and temporal processes are uncorrelated and that variograms are composed of purely spatial and purely temporal models. As an example, the space-time covariance function can be expressed as the product between purely spatial and purely temporal components. The major drawback of this type of model is the non-incorporation of the spatial-temporal interaction component (Huang et al. 2007). However, the non-separable models assume that the space-time phenomenon is correlated in space-time (Sherman 2010). In this research we used the generalized sum product model, which belongs to the non-separable covariance model class.

The stationary covariance model product-sum (Cesare et al. 2001; Iaco et al. 2001) is given by: $C(\mathbf{h}_s, h_t) = k_1 C_s(h_s) C_t(h_t) + k_2 C_s(h_s) + k_3 C_t(h_t)$. Assuming that $\varepsilon(s, t)$ is second order stationary, we have that:

- $\gamma_s(\mathbf{h}_s) = C_s(\mathbf{0}) - C_s(\mathbf{h}_s) \Rightarrow C_s(\mathbf{h}_s) = C_s(\mathbf{0}) - \gamma_s(\mathbf{h}_s)$
- $\gamma_t(h_t) = C_t(0) - C_t(h_t) \Rightarrow C_t(h_t) = C_t(0) - \gamma_t(h_t),$

in which, based on the assumption of second order stationarity, we have that the product-sum covariance function is defined as:

$$C(\mathbf{h}_s, h_t) = -\gamma_s(\mathbf{h}_s) [k_1 C_t(0) + k_2] - \gamma_t(h_t) [k_1 C_t(0) + k_3] + k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) + C(\mathbf{0}, 0), \quad (7)$$

where $k_1 > 0$, $k_2 \geq 0$ e $k_3 \geq 0$ are constants that ensure that the covariance function is positive definite. $C_s(\mathbf{0})$ e $C_t(0)$ are the spatial and temporal sills, respectively. Applying the relation, $\gamma(\mathbf{h}_s, h_t) = C(\mathbf{0}, 0) - C(\mathbf{h}_s, h_t)$, we have the product-sum variogram. This model gives origin to the following relations:

$$\begin{aligned}\gamma(\mathbf{h}_s, 0) &= \gamma_s(\mathbf{h}_s) (k_2 + k_1 C_t(0)) = k_s \gamma_s(\mathbf{h}_s) \text{ and} \\ \gamma(\mathbf{0}, h_t) &= \gamma_t(h_t) (k_3 + k_1 C_s(\mathbf{0})) = k_t \gamma_t(h_t),\end{aligned}\quad (8)$$

where k_s and k_t are the proportionality coefficients between the space-time variograms, $\gamma_s(\mathbf{h}_s, 0)$ and $\gamma_t(\mathbf{0}, h_t)$, and the spatial and temporal variograms, $\gamma_s(\mathbf{h}_s)$ and $\gamma_t(h_t)$, respectively. Therefore, modeling $\gamma(\mathbf{h}_s, 0)$ is equivalent to modeling $\gamma_s(\mathbf{h}_s)$. Likewise, estimating and modeling $\gamma(\mathbf{0}, h_t)$ is equivalent to estimating and modeling $\gamma_t(h_t)$. The two relations established in Equation (8) can be simplified by imposing three constraints: $k_2 + k_1 C_t(0) = 1$, $k_3 + k_1 C_s(\mathbf{0}) = 1$ and $k_1 + k_2 + k_3 = 1$.

These constraints facilitate the estimation of the parameters of the models $\gamma(\mathbf{h}_s, 0)$ and $\gamma(\mathbf{0}, h_t)$ using $\gamma_s(\mathbf{h}_s)$ and $\gamma_t(h_t)$, for this it is necessary to determine k_1 , k_2 and k_3 . These three coefficients of the model can be written in terms of the sills and parameters k_s and k_t (Iaco et al. 2001). In this particular case, this would lead to the modeling of $\gamma_s(\mathbf{h}_s)$ and $\gamma_t(h_t)$, using the models of $\gamma(\mathbf{h}_s, 0)$ and $\gamma(\mathbf{0}, h_t)$, respectively. In addition, these two parameters are combined into a single parameter k , resulting in the generalized product-sum model.

By combining the parameters k_s and k_t to form a single parameter k , the space-time variogram expression, $\gamma(\mathbf{h}_s, h_t)$, can be simplified, that is,

$$\begin{aligned}\gamma(\mathbf{h}_s, h_t) &= (k_2 + k_1 C_t(0)) \gamma_s(\mathbf{h}_s) + (k_3 + k_1 C_s(\mathbf{0})) \gamma_t(h_t) - k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) \\ &= k_s \gamma_s(\mathbf{h}_s) + k_t \gamma_t(h_t) - k_1 \gamma_s(\mathbf{h}_s) \gamma_t(h_t) \\ &= \gamma_s(\mathbf{h}_s, 0) + \gamma_t(\mathbf{0}, h_t) - k_1 \frac{\gamma_s(\mathbf{h}_s)}{k_s} \frac{\gamma_t(h_t)}{k_t} \\ &= \gamma_s(\mathbf{h}_s, 0) + \gamma_t(\mathbf{0}, h_t) - k \gamma_s(\mathbf{h}_s) \gamma_t(h_t),\end{aligned}\quad (9)$$

where $k = \frac{k_1}{k_s k_t}$. Thus, the parameter k can be expressed as:

$$k = \frac{\frac{k_t C_t(0) + k_s C_s(\mathbf{0}) - C(\mathbf{0}, 0)}{C_s(\mathbf{0}) C_t(0)}}{k_s k_t} = \frac{k_t C_t(0) + k_s C_s(\mathbf{0}) - C(\mathbf{0}, 0)}{[k_s C_s(\mathbf{0})] [k_t C_t(0)]}. \quad (10)$$

In this study we used the covariance generalized product sum model. The variogram function of this model is expressed by:

$$\gamma_{st}(\mathbf{h}_s, h_t) = (kC_t(0) + 1)\gamma_{st}(\mathbf{h}_s, 0) + (kC_s(\mathbf{0}) + 1)\gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h}_s, 0)\gamma_{st}(\mathbf{0}, h_t), \quad (11)$$

where $\gamma_{st}(\mathbf{h}_s, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$ are the marginal spatial and temporal variograms, respectively. The great advantage of this type of model is that sample-based marginal adjustments are used and only one global sill parameter is incorporated for space-time interaction (Iaco et al. 2015). The parameter k is positive and has an identity involving the global sill, $C_{st}(\mathbf{0}, 0)$ together with the spatial and temporal sill, $C_s(\mathbf{0})$ and $C_t(0)$, respectively, given by:

$$k = \frac{C_{st}(\mathbf{0}, 0) - C_s(\mathbf{0}) - C_t(0)}{C_s(\mathbf{0})C_t(0)}. \quad (12)$$

3.3 - Geostatistical prediction

Once the trend component has been determined and estimated and a valid space-time variogram model fitted, the next step consists of the space-time kriging of the residuals using the ordinary kriging method. The variogram model is crucial in space-time kriging to calculate the best invariant linear predictor. Thus, it is necessary to obtain the space-time kriging equations, which provide the weights of the observations in such a predictor and the prediction variation that indicates the prediction accuracy (Kilibarda et al. 2015).

That is $\{\varepsilon(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq R^2 \times R\}$. Consider that $\varepsilon(\mathbf{s}, t)$ are observed in the space-time places $\varepsilon(\mathbf{s}_1, t_1), \varepsilon(\mathbf{s}_2, t_2), \dots, \varepsilon(\mathbf{s}_n, t_n)$. The ordinary space-time kriging (Montero et al. 2015) consists of predicting the value $\varepsilon(\mathbf{s}_0, t_0)$ at the non-sampled space-time point (\mathbf{s}_0, t_0) , from the stochastic component $\varepsilon(\mathbf{s}, t_0)$. For this, the linear predictor, $\varepsilon^*(\mathbf{s}_0, t_0) = \sum_{i=1}^n \lambda_i \varepsilon(\mathbf{s}_i, t_i)$ is used, where λ_i are the kriging weights obtained by imposing the conditions that the prediction error expectation is zero and that it has minimum variance, that is, it is the Best Linear Unbiased Predictor (BLUP). In this research, to obtain the weights, we used the ordinary space-time kriging equations that are calculated in terms of the variogram:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_{st}(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) + \alpha = \gamma_{st}(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0), \forall i = 1, 2, \dots, n. \\ \sum_{i=1}^n \lambda_i = 1. \end{cases} \quad (13)$$

In the system of equations (13), α is the Lagrange multiplier required to minimize the prediction variance. The prediction variance is given by:

$$V(Z^*(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0)) = \sum_{i=1}^n \lambda_i C(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0) + \alpha, \quad (14)$$

in which the final predictor ($z(\mathbf{s}_0, t_0)$) for the precipitation variable (Z), at the location (\mathbf{s}_0, t_0), is defined as:

$$z(\mathbf{s}_0, t_0) = m(\mathbf{s}_0, t_0) + \varepsilon(\mathbf{s}_0, t_0),$$

where $m(\mathbf{s}_0, t_0)$ is the estimated value for the location (\mathbf{s}_0, t_0) obtained by fitting GAMLSS models.

In this study, the interpolation performance in space-time kriging was evaluated using the *leave-one-out* method. This method consists in removing an observed point from the precipitation in space-time ($p(\mathbf{s}_0, t)$) and then predicting its value ($\hat{p}(\mathbf{s}_0, t)$). This process is then repeated for all remaining points and the residual (E) of this procedure is then obtained by the difference between the observed and predicted values at each location ($E = p(\mathbf{s}_0, t) - \hat{p}(\mathbf{s}_0, t)$). After that, the RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) and R^2 (Coefficient of Determination) were extracted as selection criteria.

All statistical analyzes were implemented in the R packages *gamlss* (Stasinopoulos et al. 2017), *gstat* (Graler et al. 2016) and *sp* (Bivand et al. 2013).

4 - Results

4.1 - Descriptive analysis

The monthly precipitation in each of the 54 locations, in the period 1994-2014, presented a median value of 29.05 (mm) and an interquartile range (3.60 - 90.20) mm. In the whole historical series, approximately 20% of the observations presented 0.00 mm of total monthly precipitation. To evaluate the precipitation distribution of the region under study, a histogram was initially constructed for the dataset (Figure 2).

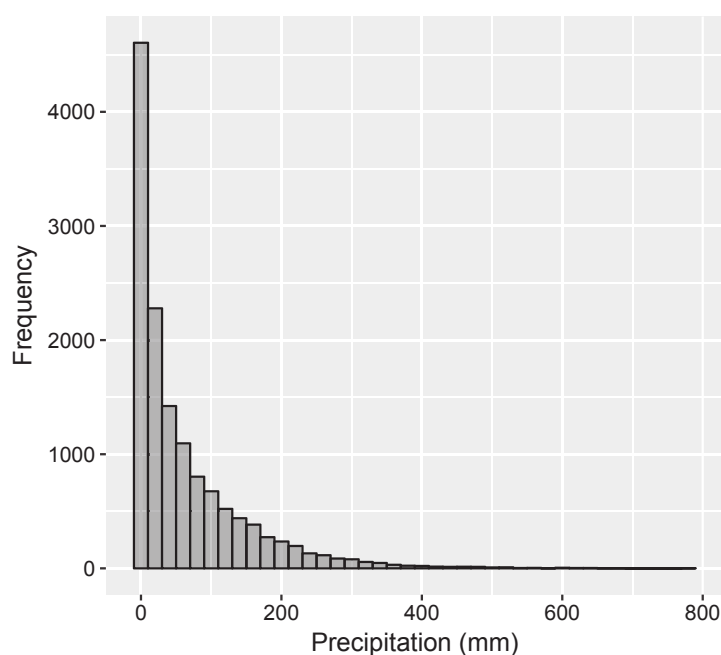


Figure 2 - Histogram of the precipitation data of the 54 pluviometric stations sampled in the period from 1994 to 2014.

In Figure 2, it is possible to perceive the excess of zeroes values in the dataset, and consequently, the presence of an asymmetric behavior, leading to indications that the Gaussian model is not the most suitable for modeling. It is known that 80% of the entire state of Paraíba is located in the Brazilian semi-arid region, which presents low rainfall and irregular rains. Consequently, the scarcity of rainfall directly affects water resources, resulting in serious problems in the region's economy, which focuses on family farming and livestock farming (Macedo et al. 2010). One of the main agricultural techniques in this region is rainfed agriculture, a technique that is based on production in places with a low rainfall amount, which is totally dependent on rainfall and has been directly affected by low water supply and distribution, impacting the productive potential throughout the region (Litre et al. 2017). Thus, with agriculture being hampered by pluviometric irregularity, extensive livestock farming appears as the main economic source and means of subsistence in periods of prolonged drought. However, agriculture and livestock farming are experiencing strong negative impacts throughout the region due to prolonged periods of drought, for example: low income in the regional and national economies, loss of human lives and livestock due to starving and rural exodus (Marengo et al. 2017b). In this way, agriculture and livestock are totally dependent on the spatio-temporal distribution of precipitation.

Irregular rainfall, resulting in periods of prolonged drought, is associated with meteorological phenomena. The years of strong El Niño intensity have caused intense droughts (Marengo et al. 2017a). Rainfall periods are concentrated in a few months, occurring mainly in winter, which are associated with the Intertropical Convergence Zone (ITCZ) operation and the passage of cold fronts of the South Atlantic Ocean (Rodrigues and McPhaden 2014).

In order to investigate the precipitation behavior in the 54 pluviometric stations, box plots were constructed for each of these locations (Figure 3).

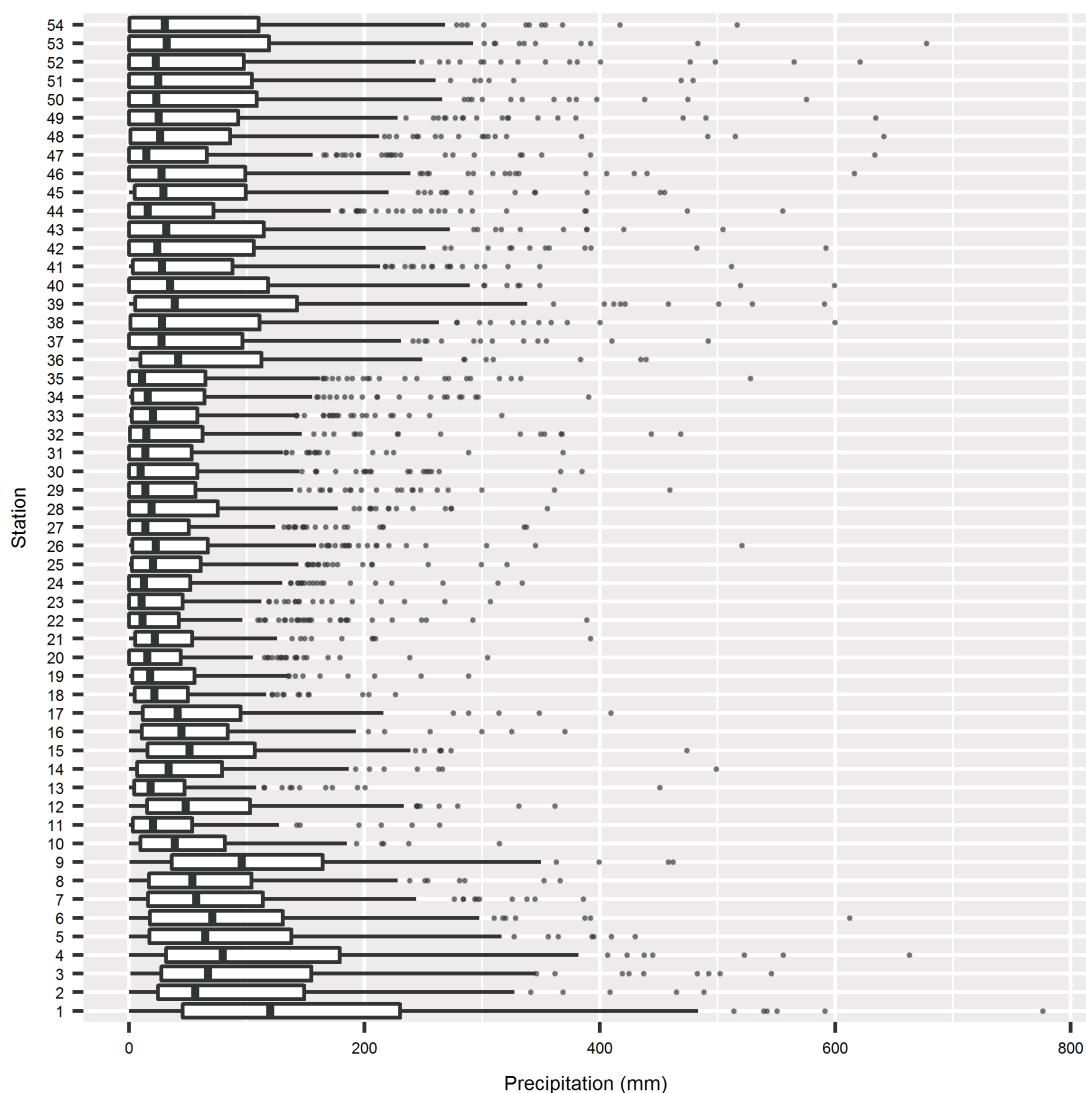


Figure 3 - Box plot of the pluviometric stations used in this study.

Due to the asymmetric distribution and the large variability of precipitation, the State of Paraíba suffers from irregular distribution of rainfall throughout the region. This distribution

presents great variability among the mesoregions analyzed in this work. A positive asymmetry is observed in all the locations, as well as the presence of atypical values.

4.2 - Trend Analysis

The distribution of precipitation over geographical coordinates (latitude and longitude) and in relation to the temporal index, with the respective adjustments by means of cubic splines, is presented in Figure 4.

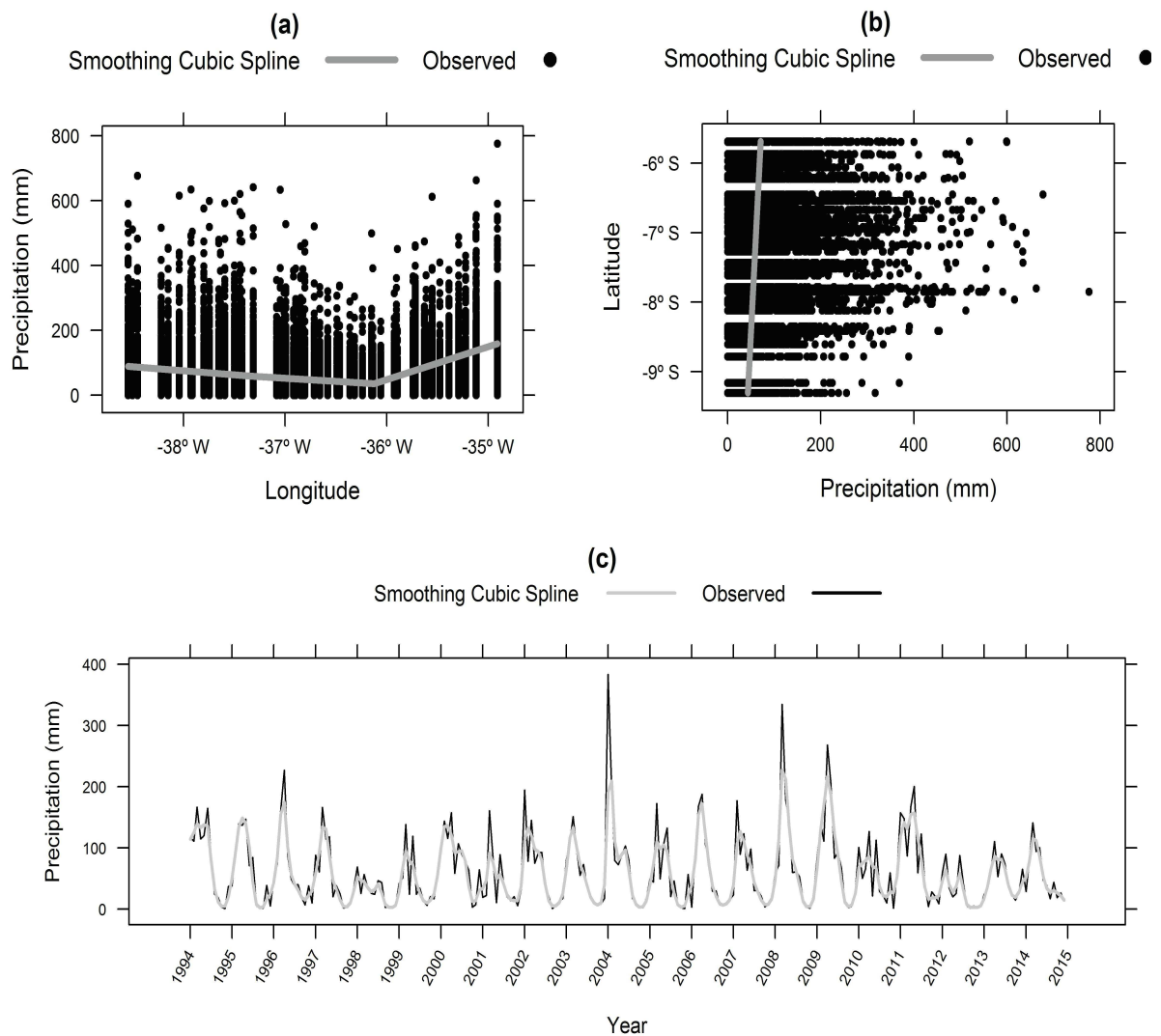


Figure 4 - Distribution of precipitation data (mm) (black dots) in the longitude direction (a) and in the latitude (b) direction and respective trend lines (gray line) using cubic smoothing splines function over the study region; time series of precipitation data (mm) (black line) and adjusted temporal trend (gray line) (c).

An exploratory analysis through the adjustment of the cubic smoothness spline function shows that, in relation to the longitude, the precipitation occurs more frequently in the extremes of the studied region that correspond to the mesoregions of the Sertão and Zona da Mata, and that occurs few rainfall in regions located near the meridian 36° west longitude, which corresponds to Borborema mesoregion (Figure 4a).

In relation to latitude, there is no behavior that stands out in relation to precipitation, at most a first degree relation (Figure 4b). Thus, the spatial distribution of precipitation in the study region presents a spatial tendency that varies with the latitudinal and longitudinal gradients. For the analysis of the time trend, the graph of the time series of precipitation was calculated, based on the average calculated on the basis of the 54 pluviometric stations (Figure 4c). As it can be observed, this series shows an evident seasonal behavior of 12 months. Thus, for this series, a GAMLSS model was fitted using cubic smoothing splines function.

To model the trend of precipitation data, the GAMLSS model was used, via ZAGA distribution, as indicated in the system of Equations (4). Table 2 presents the coefficient estimates, standard error, t value, p-value and the coefficient of determination (R^2) in the model adjustment.

Table 2: Estimates, standard errors (SE), p-values and the coefficient of determination of the GAMLSS regression model for the data set used in this study.

Component	Parameter	Estimative	SE	t value	p-value	R^2
μ	β_{10}	-16.760	0.517	-32.399	$< 10^{-4}$	0.480
	β_{11}	9.2×10^{-4}	7.2×10^{-5}	12.717	$< 10^{-4}$	
	β_{12}	0.002	5.6×10^{-5}	39.081	$< 10^{-4}$	
	β_{13}	-6.7×10^{-4}	1.1×10^{-4}	-5.955	$< 10^{-4}$	
σ	β_{20}	0.855	0.005	163.700	$< 10^{-4}$	
ν	β_{30}	-0.010	0.002	-59.350	$< 10^{-4}$	

The results presented in Table 2 indicate, at the significance level of 0.01, that the covariates longitude, latitude and the temporal index are related to the precipitation. The coefficient of determination shows that 48% of the large-scale variation of precipitation is explained by the trend component. This information corroborates the results found through the descriptive analysis.

After estimating the trend component, which was described in Equation (1), we can then analyze the stochastic residual, $\varepsilon(\mathbf{s}, t) = Z(\mathbf{s}, t) - m(\mathbf{s}, t)$. After, the space-time sampling variogram of these residual is calculated, taking as reference the fit of the generalized product-sum theoretical model.

4.3 - Geostatistical analysis

The residuals, obtained from the regression via GAMLSS models, present an evident pattern of spatio-temporal correlation (Figure 5a). There is a strong spatial correlation between nearby sites, and the spatial structure has become weaker as time differences increase. Similarly, the same occurs with the temporal structure (Figure 5b). Thus, the spatial-temporal kriging of the residuals is required.

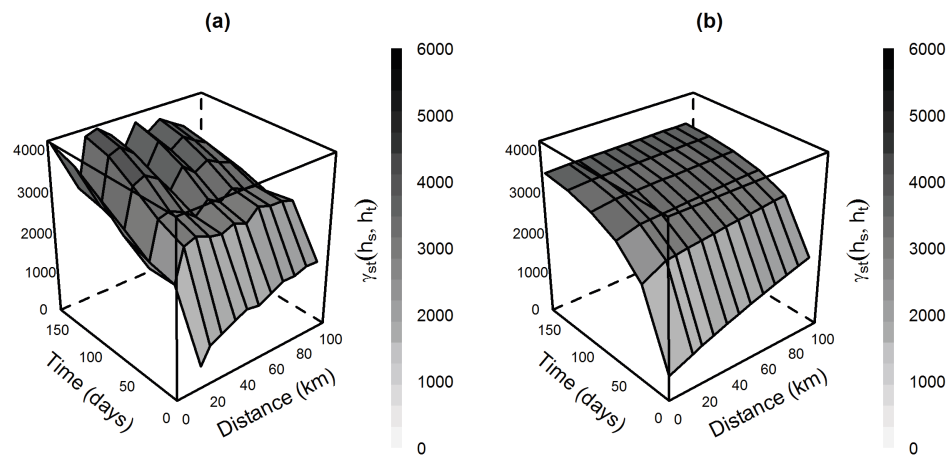


Figure 5 - Empirical space-time variogram (a) and adjusted variogram (b).

The behavior of the empirical space-time variogram estimates is presented in terms of the marginal effects (Figure 6).

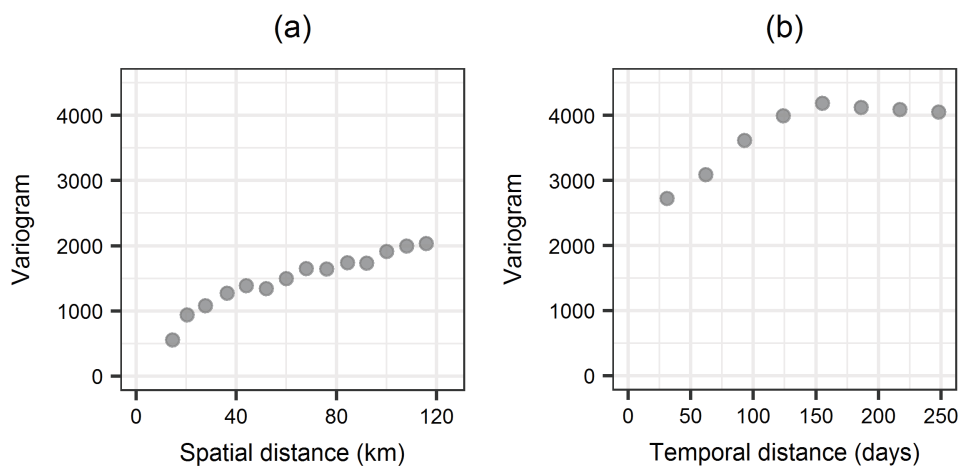


Figure 6 - Marginal experimental variograms of the residuals: purely spatial (a) and purely temporal (b).

It is clearly seen that the spatial variability exhibited by the marginal spatial variogram (Figure 6a) differs from the temporal variability shown in the marginal temporal variogram (Figure 6b), since a difference, with respect to the sill, is observed between the two marginal variograms. For this reason, the spatio-temporal variability of the residuals can be modeled through the generalized product-sum model. One of the advantages of constructing marginal variograms is the fact that it is possible to identify some theoretical models of variogram (for example, Gaussian, Exponential and Spherical) that will compose the structure in the generalized product-sum model (Equation 11).

As described previously, for the large-scale variation, the Normal, Gamma and ZAGA models were fitted. For the residuals obtained by each model, different structures of the generalized product-sum model were fitted, followed by the spatial-temporal kriging of the residuals. The results of the cross-validation are presented in Table 3. The results of the validation were calculated after the values adjusted by the trend component were added to the values interpolated in space-space kriging.

Table 3 - Results of space-time kriging for different trend models with their respective values of the “leave-one-out” cross validation.

MODEL	NORMAL			GAMMA			ZAGA		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
GAU + EXP	36.089	21.881	0.813	36.652	22.065	0.807	36.116	21.920	0.812
GAU + SPH	36.881	22.414	0.805	36.708	22.090	0.807	37.596	22.895	0.797
EXP + EXP	34.710	21.011	0.827	35.381	21.253	0.821	34.598	20.970	0.828
EXP + SPH	34.991	21.181	0.824	35.388	21.262	0.821	35.342	21.430	0.820

The results presented in Table 3 indicate that the Exponential model was the best fit for the spatial ($\gamma_{st}(\mathbf{h}_s, 0)$) and temporal ($\gamma_{st}(\mathbf{0}, h_t)$) structures in all regression models, since it presented higher R^2 and smaller RMSE and MAE. Note that the ZAGA model was the one that presented the best results with the RMSE of 34.598 mm, the MAE of 20.970 mm and the R^2 equal to 82.8% . These results demonstrate the effectiveness of adjusting to the precipitation data an adequate model that contemplates the characteristics presented by the phenomenon.

The components of the generalized product-sum model (Equation 11) and their estimates are shown in Table 4. The components were chosen based on the results in Table 3.

Table 4 - Parameter estimates of the generalized product-sum variogram model fit at residuals. The final model is obtained by replacing these estimates in Equation (11).

Component	Variogram model	Nugget	Sill	Range	k
Spatial ($\gamma_{st}(\mathbf{h}_s, 0)$)	Exponential	0.757	4.350	171 km	15.689
Temporal ($\gamma_{st}(\mathbf{0}, h_t)$)	Exponential	17.479	52.058	47 days	

The parameter estimates (Table 4) indicate that all components contribute on the spatial-temporal pattern of the ZAGA regression residuals. The parameter estimation related to the spatial extent was high, indicating that the residuals are correlated in distances of up to 171 km. The time scale estimate indicated that sites with a delay of up to 47 days had residual autocorrelation.

The major advantage of the methodology proposed in this work is that with the space-time kriging technique, predictions can be made for unobserved locations and times (Kilibarda et al. 2014; Iaco et al. 2015). Thus, it allows to evaluate the behavior in the distribution of the precipitation along the space and the time, being able to calculate estimates non monitored places. As an example of this methodology applicability, the spatial-temporal kriging of precipitation was carried out in the period 2015-2020 (Figure 7).

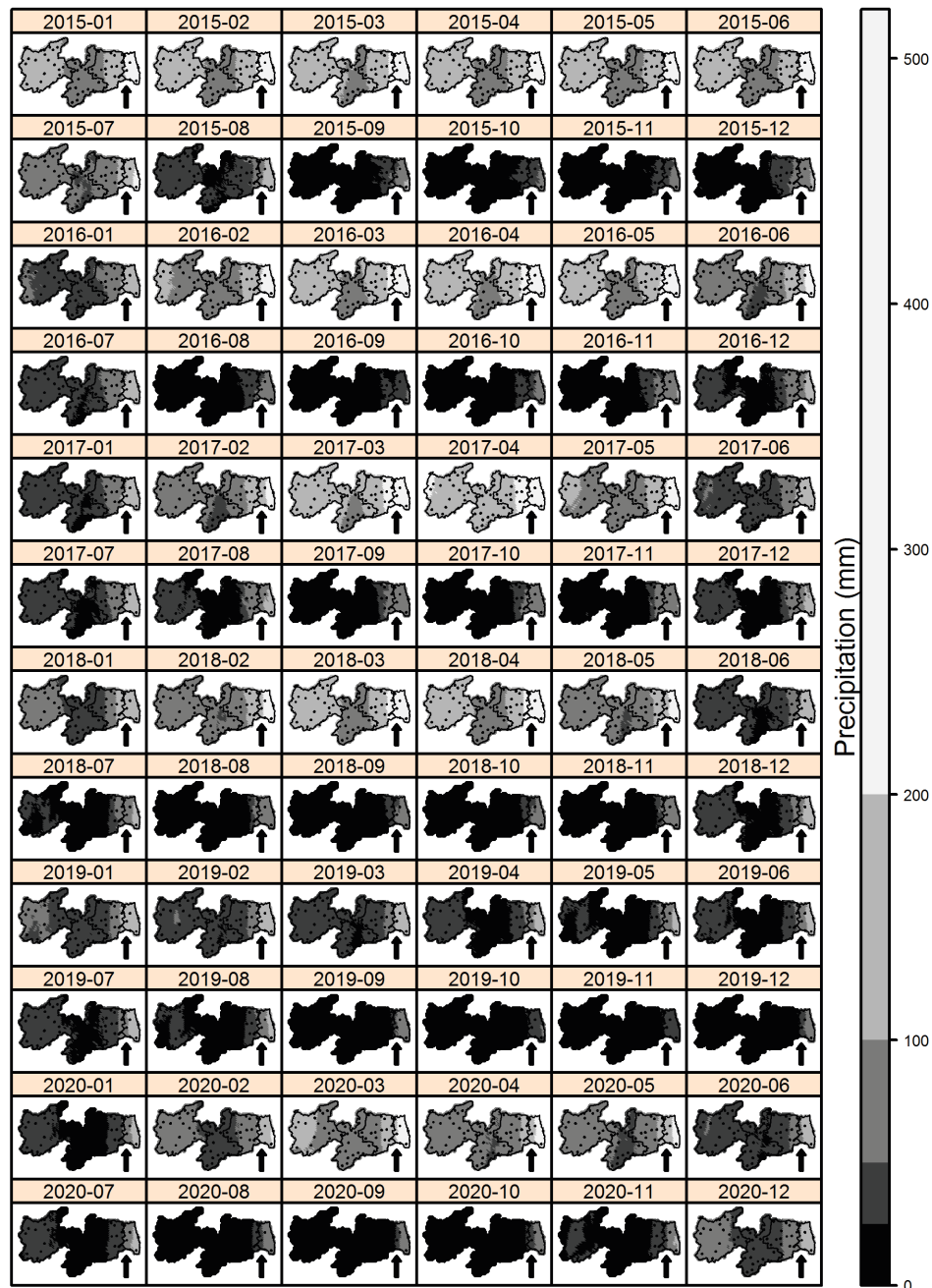


Figure 7- Spatial-temporal prediction of total monthly rainfall (mm) in the study region for the years 2015-2020. The crosses (+) represent the positions of each pluviometric station.

In Figure 7 it is possible to evaluate the precipitation behavior in non-sampled sites, as well as in times not observed in the sample, helping public policies of water projects and management systems to make decisions related to the distribution of water to the present and to the future. It is noted that the spatial distribution of rainfall presents variability within each mesoregion and also among the four mesoregions. Rainfall in the region occurs more intensely

during the first half of the year, but during the same period there is an irregularity in the distribution of rainfall between the mesoregions. In Sertão, the high spatial variability can be justified by meteorological systems in the area, such as the ITCZ and the VCAN. It is noted that in this region, the rainfall distribution during the months of February to May occurs irregularly. In years when El Niño and La Niña occur with strong intensity, there is strong evidence that rainfall occurs, respectively, above or below the expected value, resulting in a high temporal variability of precipitation (Marengo et al. 2017a). In Borborema, the lowest rainfall rates are found throughout the period. The rainy season in this area is concentrated between the months of February and May, presenting high variation in space and time. The low volume of rainfall and the irregularity of its distribution have contributed, for example, to the occurrence of the desertification phenomenon (Borges et al. 2014; Souza et al. 2015).

5 - Discussion

The state of Paraíba has irregularities in the distribution of precipitation, with high spatial and temporal variability, resulting in intense rains with short duration and prolonged periods of drought. Low rainfall over a long period of time has resulted in dry periods, not only in the state of Paraíba but largely the NEB. The impacts caused by the droughts have not only affected the region's economy, but also contributed to the farmers indebtedness, who are seeking loans from banks to cover the damage caused by long periods of drought. In addition, drought has affected the water distribution for human consumption, as well as for the electricity supply, the occurrence of diseases, migration, among others, resulting in an unprecedented water crisis (Gutiérrez et al. 2014). Thus, a space-time distribution modeling of rainfall becomes of paramount importance for the water resources management in order to reduce the impacts caused by drought in this region.

Rainfall in the State of Paraíba stands out due to the great variability between the mesoregions: Zona da Mata and Agreste have high probability values for rainfall, but the Borborema mesoregion presents low precipitation probability. The chances of precipitation occur again in Sertão (Francisco et al. 2016). In this work the authors used the gamma distribution, which presents asymmetric behavior, to model precipitation data. However, it is known that this probability distribution does not include values equal to zero, not having a parameter that can be used to model the excess of zero that is common in data of monthly total precipitation throughout the NEB. As an advantage, the methodology that was proposed in this

paper allowed to model the occurrence of null values, resulting in better estimates when compared with models that do not contemplate the excess of zeroes in the data set.

Several works in the literature that deal with spatial-temporal modeling and which have null values in the database have used the regression model with Normal distribution (Guo et al. 2015; Hu et al. 2017) or Gamma regression model (Menezes et al. 2016; Monteiro et al. 2017). However, the Normal regression does not solve the data asymmetry problem, and the Gama regression does not contemplate in its domain the occurrence of zeroes. In this work, we propose a new way of modeling the trend component through GAMLSS models, especially using the ZAGA distribution, allowing one of its three parameters to model the excess of zeroes that occurs with great frequency in data of monthly total precipitation.

The trend component alone does not model the spatio-temporal dependence that is present in the regression residuals. Thus, there is a need for spatiotemporal modeling in these residuals. In the literature, several works did not take into account this spatial and/or temporal residuals dependence (Santos et al. 2012; Gomes et al. 2015). Thus, this work proposed to model the spatiotemporal dependence structure of this stochastic component, by means of a covariance function. However, this approach has as characteristic the excess of parameters present in the space-time theoretical variogram, resulting in great computational difficulties to determine the parameter estimates.

As a great advantage of the proposed methodology is that it is able to predict in non-sampled locations and in times not observed, thus helping projects related to water resources in the region. An aggravating factor in the irregular rainfall distribution in the study region is found in the Agreste and Borborema mesoregions, with a low precipitation volume. The great part of the state of Paraíba is part of the Brazilian semi-arid, which in turn has the climate characterized by low humidity and low rainfall volume. As a consequence of this worsening, in recent years studies have revealed the desertification process in regions that are located in the semi-arid Northeast (Rodrigues and McPhaden 2014). Thus, this study can identify regions that, in the near future, will suffer from the water supply for human and animal consumption. These results can act as a supportive tool, so that government agencies related to water management can anticipate the society needs and avoid water crises, which is intensifying more and more not only in the State of Paraíba, but in the NEB as one all.

6 - Conclusion

In this work, total monthly precipitation in the state of Paraíba was analyzed using spatiotemporal geostatistical tools. In this paper, rainfall data were analyzed in 54 pluviometric stations from 1994 to 2014. Large-scale variation was modeled using spatial and temporal covariates in order to remove the spatiotemporal trend in the regression residuals. In this regression it was proposed the adjustment of GAMLSS models with different distributions for the response variable (Normal, Gamma and ZAGA). The results of the cross-validation indicated that the ZAGA distribution is the one that best fit the data with the exponential model for the spatial component and also for the time component of the generalized product-sum variogram model. In addition, the results indicated the presence of the space-time correlation in the phenomenon, implying the adequate use of space-time kriging proposed in this work. Finally, space-time kriging in Paraíba state for the period from 2015 to 2020 evidenced the great irregularity in the distribution of rainfall throughout the region, providing a detailed visual analysis of sectors suffering from scarcity or excessive rainfall, supporting government policies that address water resources.

7 - Acknowledgments

This work was conducted during a scholarship supported by the Coordination of Improvement of Higher Education Personnel (CAPES) at Federal University of Lavras.

References

- Almeida HA de, Medeiros EA (2017) Variabilidade no regime pluvial em duas mesorregiões da Paraíba e sua relação com o fenômeno EL Niño Oscilação Sul. *J Environ Anal Prog* 2:177. doi: 10.24221/jeap.2.3.2017.1340.177-185
- Bivand RS, Pebesma E, Gómez-Rubio V (2013) Hello World: Introducing Spatial Data. In: *Applied Spatial Data Analysis with R, Second*. Springer New York, New York, NY, pp 1–16
- Borges BC, Baptista GM de M, Meneses PR (2014) Identification of Hydromorphic Areas by Means of Spectral Analysis of Remote Sensing Data, as Support for the Preparation of Forensic Reports Issued to Evaluation of Rural Properties. *Rev Bras Geogr Física* 7:1062–1077. doi: 10.5935/1984-2295.20140015

- Cesare LD, Myers D., Posa D (2001) Estimating and modeling space–time correlation structures. *Stat Probab Lett* 51:9–14. doi: 10.1016/S0167-7152(00)00131-0
- Cressie N, Wikle CK (2015) *Statistics for spatio-temporal data*. John Wiley and Sons, Hoboken
- Francisco PRM, Mello V da S, Bandeira MM, et al (2016) Discriminação de cenários pluviométricos do estado da Paraíba utilizando distribuição Gama Incompleta e Teste Kolmogorov-Smirnov. *Rev Bras Geogr Física* 09:41–61
- Gomes OM, Santos CAC dos, Souza F de AS de, et al (2015) ANÁLISE COMPARATIVA DA PRECIPITAÇÃO NO ESTADO DA PARAÍBA UTILIZANDO MODELOS DE REGRESSÃO POLINOMIAL. *Rev Bras Meteorol* 30:47–58. doi: 10.1590/0102-778620120454
- Gräler B, Gerharz L, Pebesma E (2012) Spatio-temporal analysis and interpolation of PM10 measurements in Europe. *ETC/ACM Tech Pap* 2011/10 37
- Graler B, Pebesma E, Heuvelink G (2016) Spatio-Temporal Interpolation using gstat. *Wp* 8:1–20. doi: 10.1007/978-3-319-17885-1
- Guo L, Lei L, Zeng Z-C, et al (2015) Evaluation of Spatio-Temporal Variogram Models for Mapping Xco₂ Using Satellite Observations: A Case Study in China. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:376–385. doi: 10.1109/JSTARS.2014.2363019
- Gutiérrez APA, Engle NL, De Nys E, et al (2014) Drought preparedness in Brazil. *Weather Clim Extrem* 3:95–106. doi: 10.1016/j.wace.2013.12.001
- Hengl T, Heuvelink GBM, Perčec Tadić M, Pebesma EJ (2012) Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor Appl Climatol* 107:265–277. doi: 10.1007/s00704-011-0464-2
- Hu D, Shu H, Hu H, Xu J (2017) Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Cluster Comput* 20:347–357. doi: 10.1007/s10586-016-0708-0
- Huang H-C, Martinez F, Mateu J, Montes F (2007) Model comparison and selection for stationary space–time models. *Comput Stat Data Anal* 51:4577–4596. doi: 10.1016/j.csda.2006.07.038

Iaco SD, Myers DE, Posa D (2001) Space–time analysis using a general product–sum model. *Stat Probab Lett* 52:21–28. doi: 10.1016/S0167-7152(00)00200-5

Iaco SD, Myers DE, Posa D (2002) Nonseparable Space-Time Covariance Models: Some Parametric Families. *Math Geol* 34:23–42. doi: 10.1023/A:1014075310344

Iaco SD, Palma M, Posa D (2015) Spatio-temporal geostatistical modeling for French fertility predictions. *Spat Stat* 14:546–562. doi: 10.1016/j.spasta.2015.10.002

Kilibarda M, Hengl T, Heuvelink GBM, et al (2014) Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *J Geophys Res Atmos* 119:2294–2313. doi: 10.1002/2013JD020803

Kilibarda M, Tadić MP, Hengl T, et al (2015) Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. *Spat Stat* 14:22–38. doi: 10.1016/j.spasta.2015.04.005

Litre G, Curi M, Mesquita PS, et al (2017) O desafio da comunicação da pesquisa sobre riscos climáticos na agricultura familiar: a experiência de uso de cartilha educativa no Semiárido nordestino. *Desenvolv e Meio Ambient* 40:207–228. doi: 10.5380/dma.v40i0.49069

Macedo MJ, Guedes R, Sousa F de A, Dantas F (2010) Analysis of the standardized precipitation index for the Paraíba state, Brazil. *Ambient e Agua - An Interdiscip J Appl Sci* 5:204–214. doi: 10.4136/ambi-agua.130

Marengo JA, Alves LM, Alvala RC., et al (2017a) Climatic characteristics of the 2010-2016 drought in the semiarid Northeast Brazil region. *An Acad Bras Cienc.* doi: 10.1590/0001-3765201720170206

Marengo JA, Torres RR, Alves LM (2017b) Drought in Northeast Brazil—past, present, and future. *Theor Appl Climatol* 129:1189–1200. doi: 10.1007/s00704-016-1840-8

Martínez WA, Melo CE, Melo OO (2017) Median Polish Kriging for space–time analysis of precipitation. *Spat Stat* 19:1–20. doi: 10.1016/j.spasta.2016.10.003

Menezes R, Piairol H, García-Soidán P, Sousa I (2016) Spatial–temporal modellization of the $\{NO_2\}$ concentration data through geostatistical tools. *Stat Methods Appt* 25:107–124. doi: 10.1007/s10260-015-0346-3

Monteiro A, Menezes R, Silva ME (2017) Modelling spatio-temporal data with multiple seasonalities: The NO₂ Portuguese case. *Spat Stat* 22:371–387. doi: 10.1016/j.spasta.2017.04.005

Montero J-M, Fernández-Avilés G, Mateu J (2015) *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. John Wiley & Sons, Ltd, Chichester, UK

Rodrigues RR, McPhaden MJ (2014) Why did the 2011-2012 La Niña cause a severe drought in the Brazilian Northeast? *Geophys Res Lett* 41:1012–1018. doi: 10.1002/2013GL058703

Santos CAC dos, Gomes OM, Souza F de AS de, Paiva W De (2012) Análise Geoestatística da Precipitação Pluvial do Estado da Paraíba (Geostatistical Analysis of Precipitation in Paraíba State). *Rev Bras Geogr Física* 4:692. doi: 10.26848/rbgf.v4i4.232712

Schabenberger O, Gotway CA (2004) *Statistical Methods for Spatial Data Analysis*. CRC Press, Boca Raton

Sherman M (2010) *Spatial Statistics and Spatio-Temporal Data*. John Wiley & Sons, Ltd, Chichester, UK

Souza BI, Menezes R, Artigas RC (2015) Efeitos da desertificação na composição de espécies do bioma Caatinga, Paraíba/Brasil. *Investig Geográficas* 45–59. doi: 10.14350/rig.44092

Stasinopoulos M, Rigby RA, Heller GZ, et al (2017) *Flexible regression and smoothing using GAMLSS in R*. CRC Press, Boca Raton

Stauffer R, Mayr GJ, Messner JW, et al (2017) Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. *Int J Climatol* 37:3264–3275. doi: 10.1002/joc.4913

SEGUNDA PARTE - ARTIGOS**ARTIGO 3**

Krigagem com regressão espaço-temporal com modelos GAMLSS

**Artigo será traduzido e submetido à
uma revista na área de estatística
(versão preliminar)**

Krigagem com regressão espaço-temporal com modelos GAMLSS

Resumo

A variabilidade espaço-temporal de um fenômeno por ser decomposta pelo componente de tendência e pelo resíduo estocástico. O foco deste artigo é considerar diferentes modelos probabilísticos no ajuste do componente de tendência, uma vez que a estimativa precisa da distribuição espaço-temporal da variável resposta requer uma distribuição de probabilidade adequada. Para isso, foram utilizados os modelos aditivos generalizados para posição, escala e forma (GAMLSS), pois estes permitem, no ajuste da regressão, modelar a superdispersão e excesso de zeros nos dados. Para aplicação da metodologia apresentada neste artigo, foi utilizado um conjunto de dados de precipitação da Bacia do Rio São Francisco (BHSF), avaliados no espaço e no tempo. A BHSF ocupa 8% de todo o território brasileiro, sendo a maior parte localizada na Região Nordeste do Brasil. Os dados de precipitação apresentam, em sua maioria, uma distribuição com alto grau de assimetria positiva e uma grande frequência de observações que são exatamente iguais a zero. Assim, para o componente de tendência é proposto o ajuste de modelos GAMLSS, comparando diferentes famílias de distribuições de probabilidades que contemplam, em sua maioria, as características destes dados. Os resíduos produzidos por esta regressão foram modelados por estruturas de covariâncias que levam em consideração a dependência espacial, temporal e espaço-temporal. A krigagem com regressão espaço-temporal que combinou a krigagem sobre os resíduos com a regressão GAMLSS apresentou bons resultados ao analisar os dados de precipitação da BHSF. A metodologia apresentada neste artigo possibilitou a criação de cenários futuros para a área em estudo, interpolando locais não observados e prevendo estimativas em anos futuros. Sabe-se que a projeção destes cenários para a BHSF é de extrema importância para o planejamento e desenvolvimento de atividades relacionadas aos recursos hídricos, identificando as mudanças climáticas e conscientizando o governo e a população sobre os impactos e a vulnerabilidade destas mudanças climáticas.

Palavras-chave: Regressão GAMLSS. Distribuições assimétricas. Climatologia

1 - Introdução

Nos últimos anos, o aumento da frequência e intensidade das secas e sua relação com o aquecimento global vem sendo foco de várias pesquisas associadas ao risco climático, provocando uma crise hídrica sem precedentes. Diversos estudos têm sido realizados para estudar a variabilidade temporal e espacial da precipitação (MARTÍNEZ et al., 2017; RAJA et al., 2016; STAUFFER et al., 2016). Observar informações de precipitação, por meio de instrumentos de medições em locais específicos, pode ser considerada uma tarefa fácil. Entretanto, tem-se a necessidade de obter estimativas para locais não amostrados (interpolação) e em tempos futuros (previsões). Existem diversas metodologias que auxiliam na obtenção dessas estimativas, tais como a análise de regressão (STAUFFER et al., 2016) e os

modelos geoestatísticos que consideram a interação espaço-temporal (KILIBARDA et al., 2014; MARTÍNEZ et al., 2017).

Uma vez identificado que o fenômeno em estudo apresenta uma variabilidade espacial e temporal, os modelos que consideram a relação espaço-tempo têm apresentado melhores resultados por considerar em sua estrutura os componentes espaciais, temporais e espaço-temporais (IACO et al., 2001; KILIBARDA et al., 2015; MONTEIRO et al., 2017; SNEPVANGERS et al., 2003). Os fenômenos indexados no espaço-tempo podem ser decompostos em dois componentes, o de tendência e o do resíduo estocástico. Os resíduos podem ser modelados por funções de covariâncias. O componente de tendência é, na maioria dos trabalhos, ajustados por meio de uma regressão linear normal (GRÄLER et al., 2012; HU et al., 2017; KILIBARDA et al., 2014).

Estudos recentes têm levado em consideração que a variável resposta pode apresentar um comportamento de não-normalidade (MENEZES et al., 2016; MONTEIRO et al., 2017), propondo a utilização de um modelo linear generalizado com família Gama. O foco deste estudo é considerar diferentes modelos probabilísticos, uma vez que uma boa estimativa da distribuição espaço-temporal da variável resposta requer uma distribuição de probabilidade adequada. Para isso, serão utilizados os modelos aditivos generalizados para posição, escala e forma (GAMLSS), pois estes permitem, no ajuste da regressão, modelar a superdispersão e excesso de zeros nos dados (STASINOPOULOS et al., 2017). Para aplicação da metodologia apresentada neste artigo, será utilizado o conjunto de dados de precipitação, da Bacia do Rio São Francisco (BHSF), avaliados no espaço e no tempo.

Os dados de precipitação total mensal na BHSF apresentam, em sua maioria, uma distribuição com alto grau de assimetria positiva e uma grande frequência de observações iguais a zero. Assim, para o componente de tendência é proposto o ajuste de modelos GAMLSS, comparando diferentes famílias de distribuições de probabilidades que contemplam, em sua maioria, as características de dados de precipitação. Os resíduos produzidos por esta regressão são modelados por estruturas que levam em consideração a dependência espacial, temporal e espaço-temporal. A krigagem com regressão será utilizada para obter estimativas mensais em qualquer ano, inclusive no futuro, para qualquer local dentro da área de estudo.

Desde 2012, a Bacia do Rio São Francisco (BHSF) enfrenta condições hidrológicas adversas, com vazões e precipitações abaixo da média, reduzindo o volume armazenado em

seus reservatórios. Entre 2014 e 2016 foram registradas as menores vazões naturais médias anuais no reservatório de Sobradinho, desde o início das observações pelo Operador Nacional de Sistemas Elétrico (ONS), em 1931 (ANA, 2017). O Rio São Francisco é de suma importância para o semiárido do Nordeste Brasileiro (NEB) para o abastecimento de água visando ao consumo humano e irrigação, na geração de energia pelas usinas hidrelétricas de Paulo Afonso e Sobradinho, para o turismo e para a pesca (ARAUJO; CELESTE, 2018; SILVEIRA et al., 2016).

Desenvolver cenários do comportamento da precipitação na BHSF é um dos passos para se entender como as condições climáticas podem interferir na distribuição das chuvas, destacando as áreas com maior risco climático e que devem ser inseridas em estratégias sociais de resiliência climática (CHEN et al., 2008). Assim, o objetivo deste artigo é estimar a precipitação, com alto grau de precisão, em toda a área de estudo, tanto no espaço quanto no tempo e que pode ser facilmente implementada e aplicada em novos conjuntos de dados para esta ou outras regiões.

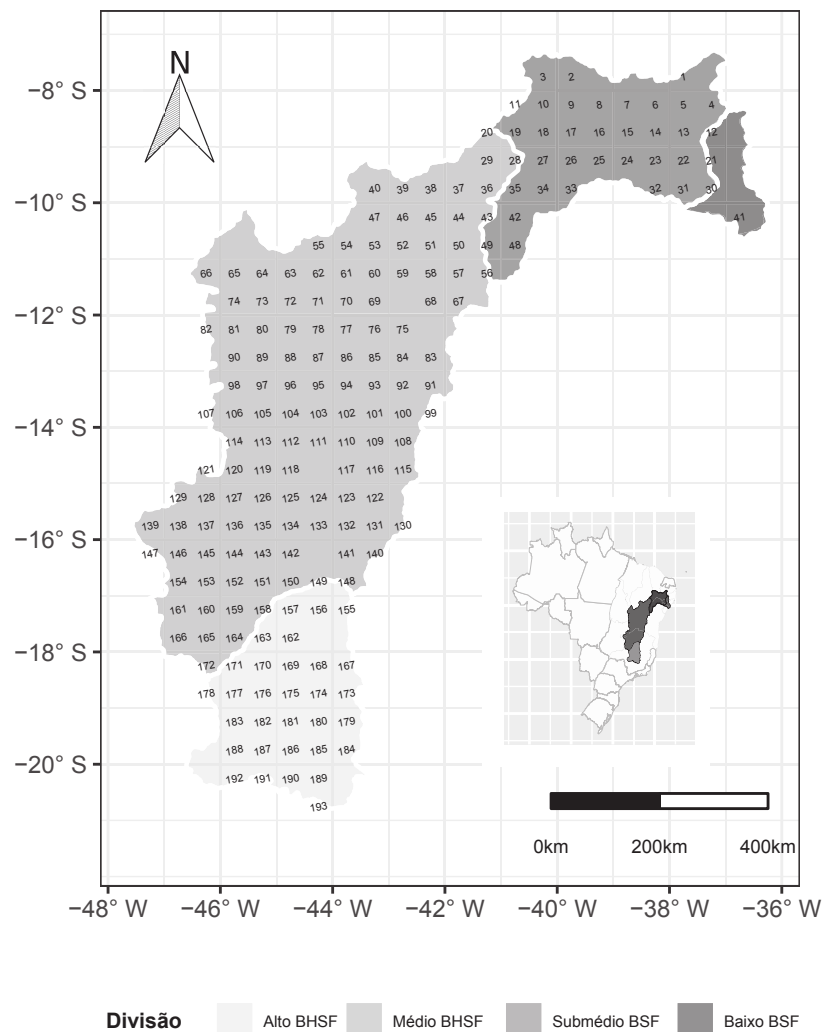
2 - Material e Métodos

2.1 - Área de estudo e o conjunto de dados

Neste trabalho foram utilizados dados de precipitação total mensal em 193 pontos da BHSF (Figura 1), localizada no Brasil, entre as coordenadas $7,0^{\circ}$ – $21,0^{\circ}$ S e $35,0^{\circ}$ – $47,7^{\circ}$ W, e no período de 1979 a 2014. O conjunto de dados foi disponibilizado pelo Centro de Previsão Climática (CPC) em pontos de grade $0,5^{\circ} \times 0,5^{\circ}$ calculados por meio de observações de calibre (CHEN et al., 2008).

O principal rio da BHSF é o Rio São Francisco, que nasce na Serra da Canastra, localizada no Estado de Minas Gerais, e cobre uma área de drenagem de 639.219 km^2 em uma extensão de 2.863 km, correspondendo a 8% de todo o território brasileiro. Durante o seu percurso natural até o oceano, percorre os estados de Minas Gerais, Bahia, Pernambuco, Sergipe e Alagoas, antes de desaguar no Atlântico. Para fins de planejamento, a BHSF é dividida em quatro regiões fisiográficas: Alto, Médio, Submédio e Baixo São Francisco (MMA, 2006). Na Figura 1 tem-se a representação da BHSF, com a identificação dessas quatro regiões e dos 193 pontos amostrados.

Figura 1 - Mapa da Bacia do Rio São Francisco e a distribuição dos postos amostrados que foram identificados por números de 1 a 193.



Segundo a classificação de Köppen-Geiger (PEEL et al., 2007), a bacia apresenta cinco regiões climáticas diferentes: temperado, clima tropical de savana (Aw), clima desértico (BWh), Clima semiárido (BSh), clima temperado com inverno seco e verão quente (Cwa) e clima com inverno seco e verão ameno (CWb). Devido a essa característica, diferentes ecossistemas são observados ao longo da região como Floresta Atlântica (Cabeceira da BHSF), Cerrado (Alto e Médio São Francisco), Caatinga (Médio e Submédio São Francisco), Caatinga, Floresta Atlântica, mangue e vegetação costeira (Baixo São Francisco) (BEZERRA et al., 2018).

Devido a sua extensão, vários sistemas atmosféricos atuam na distribuição das chuvas sobre a BHSF: Zona de Convergência Intertropical (ZCIT) e Vórtices Ciclônicos de Altos Níveis (VCAN) no Baixo São Francisco principalmente nos primeiros meses do ano; a Zona

de Convergência do Atlântico Sul (ZCAS) e sistemas frontais (frentes frias) no Alto, Médio e Submédio São Francisco; durante o verão austral Distúrbios Ondulatórios de Leste (DOL), Linhas de Instabilidade (LI's) e brisas no decorrer do ano (CAVALCANTI et al., 2016). A região do Submédio São Francisco é a mais vulnerável a ocorrência de secas severas nos períodos de El Niño forte (BEZERRA et al., 2018).

2.2 - Modelo espaço-temporal

As observações da precipitação foram modeladas de acordo com processo estocástico espaço-temporal, considerando a função aleatória $\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in D \subseteq R^d \times R_+\}$, sendo $d = 2$ correspondendo à dimensão do domínio espacial em que foram utilizadas as coordenadas geográficas (latitude e longitude) de cada ponto amostrado. O processo $Z(\mathbf{s}, t)$ é decomposto pela soma do componente de tendência $m(\mathbf{s}, t)$ e do resíduo estocástico $\varepsilon(\mathbf{s}, t)$:

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

Na Equação (1), $m(\mathbf{s}, t) = E[Z(\mathbf{s}, t)]$ representa o valor esperado do processo $Z(\mathbf{s}, t)$, representando a variação em larga escala do processo.

2.3 - Componente de tendência

Na maioria dos fenômenos observados, o componente $m(\mathbf{s}, t)$ não pode ser modelado por uma função determinística, em virtude da aleatoriedade envolvida no processo. Assim, surge a necessidade de modelar este componente por funções estocásticas que têm o intuito de modelar os padrões dos dados observados no espaço-tempo (MONTEIRO et al., 2017). Na modelagem desse componente são incluídas covariáveis que foram coletadas em conjunto com a variável resposta. Neste artigo utilizou-se as próprias coordenadas geográficas de cada ponto amostrado bem como inserido efeitos anuais por meio de um índice temporal, uma vez que a precipitação apresentou uma sazonalidade de 12 meses. Para modelar essa regressão foi considerado um modelo GAMLSS, uma vez que permite uma maior flexibilidade no ajuste da regressão. A título de comparação foram realizados ajustes utilizando o modelo linear Normal (LM), e os modelos GAMLSS com resposta Normal (NO), Gama Inflacionada de Zeros (ZAGA) e Gama Generalizada (GG).

2.3.1 - Modelos GAMLSS

Seja $\mathbf{z}^T = (z_1, z_2, \dots, z_n)$ um vetor para a variável resposta com n observações. Assume-se que as observações z_i ($i = 1, 2, \dots, n$) são independentes condicionado a $\boldsymbol{\theta}_k$ ($k = 1, 2, \dots, p$), em que p representa a quantidade de parâmetros. Uma função $g_k(\cdot)$ é definida como uma função de ligação monótona que relaciona $\boldsymbol{\theta}_k$ com as variáveis explicativas e os efeitos aleatórios por meio de um modelo aditivo expresso por:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Y}_{jk} \boldsymbol{\phi}_{jk}, \quad (2)$$

em que $\boldsymbol{\theta}_k$ e $\boldsymbol{\eta}_k$ são vetores de tamanho n , $\boldsymbol{\beta}_k$ é um vetor de parâmetros de tamanho J'_k ; \mathbf{X}_k e \mathbf{Y}_{jk} são as matrizes do delineamento (covariáveis) fixas de tamanho $n \times J'_k$ e $n \times q_{ij}$, respectivamente, e $\boldsymbol{\phi}_{jk}$ é uma variável aleatória q_{ij} -dimensional. O Modelo (2) é chamado de GAMLSS (RIGBY et al., 2005).

Se $\mathbf{Y}_{jk} = \mathbf{I}_n$, em que \mathbf{I}_n é uma matriz identidade $n \times n$ e $\boldsymbol{\phi}_{jk} = h_{jk} = h_{jk}(\mathbf{x}_{jk})$ para todas as combinações de j e k , o Modelo (2) é da forma:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \quad (3)$$

em que \mathbf{x}_{jk} são vetores de tamanho n , h_{jk} é uma função desconhecida das covariáveis X_{jk} e $h_{jk} = h_{jk}(\mathbf{x}_{jk})$ é o vetor que avalia a função h_{jk} em \mathbf{x}_{jk} . O Modelo (3) é chamado de modelos GAMLSS semiparamétrico, uma vez que contém termos paramétricos e semiparamétricos (RIGBY et al., 2005). Neste artigo quando se ajustaram os modelos GAMLSS com as famílias ZAGA e a Normal foram utilizadas as funções *P-Splines* (EILERS et al., 2015) com sua versão mais atual que utiliza a decomposição de valores singulares (STASINOPOULOS et al., 2017).

No ajuste da regressão GAMLSS foram utilizadas as covariáveis latitude (x_{11}), longitude (x_{12}), índice temporal (x_{13}) e as variáveis sazonais, $(\sum_{q=1}^{12} x_{14q})$, representando cada um dos doze meses do ano. O índice temporal foi construído da seguinte forma: o número 1 jan/1979, número 2 para fev/1979 e assim por diante. Diante disso, os modelos que foram

ajustados, a título de comparação, para o componente de tendência são expressos da forma:

$$\begin{aligned}
 &g_1(\boldsymbol{\mu}) = \ln(\boldsymbol{\mu}) = \beta_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} \\
 \text{Modelo I:} &g_2(\boldsymbol{\sigma}) = \ln(\boldsymbol{\sigma}) = \beta_{20} \\
 &g_3(\mathbf{v}) = \ln(\mathbf{v}) = \beta_{30} \\
 &g_1(\boldsymbol{\mu}) = \ln(\boldsymbol{\mu}) = \beta_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} + \sum_{q=1}^{12} x_{14q} \\
 \text{Modelo II:} &g_2(\boldsymbol{\sigma}) = \ln(\boldsymbol{\sigma}) = \beta_{20} + \sum_{q=1}^{12} x_{14q} \\
 &g_3(\mathbf{v}) = \ln(\mathbf{v}) = \beta_{30}
 \end{aligned} \tag{4}$$

Na Expressão (4), $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ e \mathbf{v} representam, nos modelos GAMLSS, os parâmetros de uma distribuição de probabilidade relacionados a posição, escala e forma, respectivamente. Note que, ao utilizar a distribuição Normal no ajuste da Regressão GAMLSS, não é necessário modelar o parâmetro \mathbf{v} , pois esta distribuição contém apenas os parâmetros de posição e escala.

2.3.2 - Método de estimação

Para as covariáveis latitude, longitude e índice temporal foram utilizadas as funções *P-Splines* com dois graus de liberdade para latitude e longitude e 3 graus de liberdade para o índice temporal. Estas funções são extremamente úteis para modelar as relações entre os preditores lineares e a variável resposta, em especial, quando estas relações são de polinômios de alto grau ou até mesmo não-lineares.

No Modelo (2), assume-se que $\boldsymbol{\phi}_{jk}$ tem distribuições normais a priori independentes $\boldsymbol{\phi}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^-)$, sendo \mathbf{G}_{jk}^- é a inversa generalizada de uma matriz simétrica de dimensão $q_{jk} \times q_{jk}$, $\mathbf{G}_{jk}^- = \mathbf{G}_{jk}^-(\boldsymbol{\lambda}_{jk})$, a qual pode depender de um vetor de hiperparâmetros $\boldsymbol{\lambda}_{jk}$. A suposição de independência entre diferentes vetores de efeitos aleatórios $\boldsymbol{\lambda}_{jk}$ é essencial dentro da estrutura do GAMLSS (RIGBY et al., 2005). O método da máxima verossimilhança penalizada (MPL) foi utilizado no ajuste dos modelos para determinar os estimadores para os parâmetros $\boldsymbol{\lambda}_{jk}$, $\boldsymbol{\beta}_k$ e $\boldsymbol{\phi}_{jk}$. O MPL é dado por $l_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \boldsymbol{\phi}_{jk}^T \mathbf{G}_{jk}(\boldsymbol{\lambda}_{jk}) \boldsymbol{\phi}_{jk}$, em que $l = \sum_{i=1}^n \ln[f(z_i, \boldsymbol{\theta}^i)]$ é a função log-verossimilhança das observações dado $\boldsymbol{\theta}^i$.

Para maximizar a função de verossimilhança penalizada l_p foi utilizado o algoritmo CG (COLE; GREEN, 1992), sendo este baseado no algoritmo de Newton-Raphson (RIGBY et al., 2005). No trabalho destes autores, demonstra-se que a maximização de l_p resulta em uma

matriz de suavização \mathbf{S}_{jk} , sendo esta aplicada aos resíduos parciais ε_{jk} para atualizar a estimativa do preditor aditivo $\mathbf{Y}_{jk}\boldsymbol{\phi}_{jk}$ por meio de um algoritmo de *backfitting*, $\mathbf{S}_{jk} = \mathbf{Y}_{jk} \left(\mathbf{Y}_{jk}^T \mathbf{W}_{kk} \mathbf{Y}_{jk} + \mathbf{G}_{jk} \right)^{-1} \mathbf{Y}_{jk}^T \mathbf{W}_{kk}$, em que $j = 1, 2, \dots, J_k$, $k = 1, 2, \dots, p$ e \mathbf{W}_{kk} é uma matriz diagonal de pesos iterativos. Para funções de suavização *splines* cúbicas tem-se que $\boldsymbol{\lambda}_{jk} = h_{jk}$, $\mathbf{Y}_{jk} = \mathbf{I}_n$ e $\mathbf{G}_{jk} = \boldsymbol{\lambda}_{jk} \mathbf{K}_{jk}$, em que \mathbf{K}_{jk} é uma matriz estruturada.

Neste artigo o ajuste para o Modelo (1) é realizado separadamente, incluindo uma abordagem de dois passos. Primeiro estima-se o componente de tendência e em seguida modela-se o resíduo espaço-temporal, o qual apresenta dependência espaço-temporal. Na próxima seção são apresentados detalhes da modelagem desses resíduos.

2.4 - Resíduo estocástico

Ao ajustar uma regressão para o componente de tendência, utilizou-se o método da máxima verossimilhança. Ao utilizar esse método é pressuposto que os erros são não correlacionados. Entretanto, sabe-se que os resíduos produzidos por este ajuste apresentam uma autocorrelação espaço-temporal. Assim, com o intuito de modelar esses resíduos foi proposto nesta pesquisa o ajuste de um modelo de variograma espaço-temporal válido, o qual leva em consideração as estruturas de dependências espacial, temporal e espaço-temporal dos resíduos. O modelo de variograma teórico foi ajustado ao modelo de variograma empírico:

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

em que $|L(\mathbf{r}_s, r_t)|$ é a cardinalidade para o conjunto $L(\mathbf{r}_s, r_t)$, \mathbf{r}_s e r_t são as distâncias espacial e temporal, respectivamente.

Na literatura são apresentados diversos métodos que incorporam uma estrutura de dependência espaço-tempo para o resíduo estocástico espaço-temporal (CRESSIE; HUANG, 1999; IACO et al., 2001; GNEITING, 2002; SNEPVANGERS et al., 2003). A principal diferença entre os modelos de covariâncias espaço-tempo consiste na formulação de suas estruturas, pois alguns admitem que o espaço e o tempo atuem separadamente, já outros consideram a presença da não separabilidade entre o espaço e o tempo, como IACO (2010). Na classe de modelos separáveis, as funções de covariâncias espaço-tempo podem ser escritas como a soma ou o produto dos componentes puramente espaciais e temporais. Embora esta classe de modelos apresente facilidade no aspecto computacional, ele não admite a presença da

interação espaço-tempo, sendo esta a principal desvantagem desses modelos (MONTERO et al., 2015). Assim nosso estudo propõe uma abordagem com a classe de modelos não separáveis, utilizando o modelo de covariância espaço-temporal do tipo “*sum-metric*” (KILIBARDA et al., 2014; SNEPVANGERS et al., 2003). Este modelo assume que o resíduo $\varepsilon(s, t)$ pode ser decomposto em três componentes estacionários de segunda ordem e mutuamente independentes (HEUVELINK; GRIFFITH, 2010; SNEPVANGERS et al., 2003):

$$\varepsilon(\mathbf{s}, t) = \varepsilon_s(\mathbf{s}) + \varepsilon_t(t) + \varepsilon_{st}(\mathbf{s}, t), \quad (6)$$

em que $\varepsilon_s(\mathbf{s})$ é o processo puramente espacial, $\varepsilon_t(t)$ é o processo puramente temporal e $\varepsilon_{st}(\mathbf{s}, t)$ é o processo espaço-temporal que incorpora a anisotropia geométrica. O variograma para $\varepsilon(\mathbf{s}, t)$ é escrito da forma:

$$\gamma(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s) + \gamma_t(h_t) + \gamma_{st}\left(\sqrt{\mathbf{h}_s^2 + (\alpha h_t)^2}\right), \quad (7)$$

em que $\gamma(\mathbf{h}_s, h_t)$ é o variograma associado às distâncias espaciais (\mathbf{h}_s) e temporais (h_t), γ_s e γ_t são os componentes puramente espacial e temporal e γ_{st} é o componente da interação espaço-tempo. O parâmetro α denota a anisotropia geométrica, assumindo que as distâncias espacial e temporal podem ser reduzidas a uma única distância espaço-tempo. Supondo que para o Modelo (7) seja ajustado o modelo exponencial para os três componentes, tem-se neste caso que o modelo “*sum-metric*” terá 10 parâmetros a serem estimados: três pepitas, três patamares e três alcances para os variogramas espacial, temporal e espaço-temporal, respectivamente, e o parâmetro da anisotropia geométrica. Para ajustar o Modelo (7) aos dados do variograma empírico foi utilizado o método de mínimos quadrados ponderados. Os modelos teóricos de variogramas utilizados no ajuste foram os Esférico, Gaussiano, Exponencial e Periódico.

2.5 - Krigagem com regressão

A Krigagem Espaço-Temporal (KST) não se diferencia matematicamente da krigagem puramente espacial (KILIBARDA et al., 2015). Vale ressaltar que a escolha do modelo de variograma para cada componente do Modelo (7) é de fundamental importância na KST para obter o melhor preditor não-viesado, dado por:

$$\hat{\varepsilon}(\mathbf{s}_0, t_0) = \mathbf{c}_0^T \mathbf{C}_n^{-1} \bar{\boldsymbol{\varepsilon}}, \quad (8)$$

em que $\hat{\varepsilon}(\mathbf{s}_0, t_0)$ é o valor obtido pelo método da krigagem para o ponto espaço-tempo (\mathbf{s}_0, t_0) , \mathbf{c}_0^T é o vetor de tamanho n das covariâncias entre os resíduos para os pontos estimados e observados, \mathbf{C}_n é a matriz de covariância dos resíduos de dimensão $n \times n$ e $\bar{\boldsymbol{\varepsilon}}$ é o vetor dos resíduos obtidos pelo ajuste do Modelo (1). O preditor final da variável Z para a localização (\mathbf{s}_0, t_0) é dado por:

$$z(\mathbf{s}_0, t_0) = \hat{m}(\mathbf{s}_0, t_0) + \hat{\varepsilon}(\mathbf{s}_0, t_0), \quad (9)$$

em que $\hat{m}(\mathbf{s}_0, t_0)$ é o valor estimado para o componente de tendência para a localização (\mathbf{s}_0, t_0) .

Note que a krigagem com regressão consiste no ajuste, separadamente, do componente de tendência e dos resíduos. Ao ajustar um modelo de regressão, obtemos os valores preditos para a variável resposta, seguindo-se a obtenção dos resíduos para todos os pontos amostrados no espaço-tempo. Para estes resíduos é obtido o variograma empírico, seguido do ajuste de um modelo teórico. Os resíduos são então interpolados e adiciona-se à tendência estimada. Nos últimos anos, para a krigagem de regressão, têm surgido avanços teóricos e várias aplicações em fenômenos reais (HENGL et al., 2011; HEUVELINK et al., 2012; KILIBARDA et al., 2014).

2.6 - Validação cruzada

Para determinar a melhor estrutura para o Modelo (7) utilizou-se o método da validação cruzada “*leave-one-out*” (IACO et al., 2015; KILIBARDA et al., 2014). O método consiste em retirar uma observação espacial e realizar a predição para esta localidade. O processo é repetido até que todas as localidades tenham sido deixadas de fora em uma única vez. No final teremos o conjunto de dados com os valores observados e os valores preditos. Após isto foram calculadas a Raiz Quadrada do Erro Quadrático Médio (RMSE), o Erro Médio Absoluto (MAE), o Coeficiente de Correlação de Concordância (CCC) e o Coeficiente de Determinação (R^2). Para determinar o melhor modelo utilizou-se o método de ranqueamento proposto por Bastos (2017). Este método consistiu-se da seguinte forma:

- 1º Passo:** Selecionou-se 4 métodos de seleção: RMSE, MAE, CCC e R^2 ($E_g, g = 1, 2, 3, 4$);
- 2º Passo:** Ajustou-se 30 modelos, que era resultado das combinações de diferentes modelos teóricos de variograma ($q = 1, 2, \dots, 30$);
- 3º Passo:** Em seguida, calculou-se a seguinte estatística: $R_{qg} = |(E_g - E_{qg}) / E_g|$, em que R_{qg} , com $q = 1, 2, \dots, 30$ e $g = 1, 2, 3, 4$, é o valor da estatística para o q -ésimo modelo no

g -ésimo método de seleção, E_g é a estimativa do melhor modelo para o g -ésimo método de seleção, E_{qg} é o valor do q -ésimo modelo para o g -ésimo método de seleção.

4º Passo: Calculou-se as médias e os desvios padrões do passo 3 com relação a cada modelo:

Modelo	Método				Média	Desvio Padrão
	RMSE	MAE	CCC	R^2		
1	R_{11}	R_{12}	R_{13}	R_{14}	\bar{R}_1	S_1
2	R_{21}	R_{22}	R_{23}	R_{24}	\bar{R}_2	S_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
30	R_{301}	R_{302}	R_{303}	R_{304}	\bar{R}_{30}	S_{30}

5º Passo: Obteve-se a seguinte estatística:

$$\text{Rank}_q = \sqrt{\bar{R}_q^2 + S_q^2}$$

6º Passo: O melhor modelo foi dado pelo $\min(\text{Rank}_1, \text{Rank}_2, \dots, \text{Rank}_{30})$.

2.7 - Programa estatístico

Todas as análises dessa pesquisa foram feitas no *programa R* (R Core Team, 2018). Para o ajuste da tendência foi utilizado à biblioteca *gamlss* (STASINOPOULOS et al., 2017). Para o ajuste do componente residual utilizou-se a biblioteca *gstat* (GRÄLER; PEBESMA; HEUVELINK, 2016).

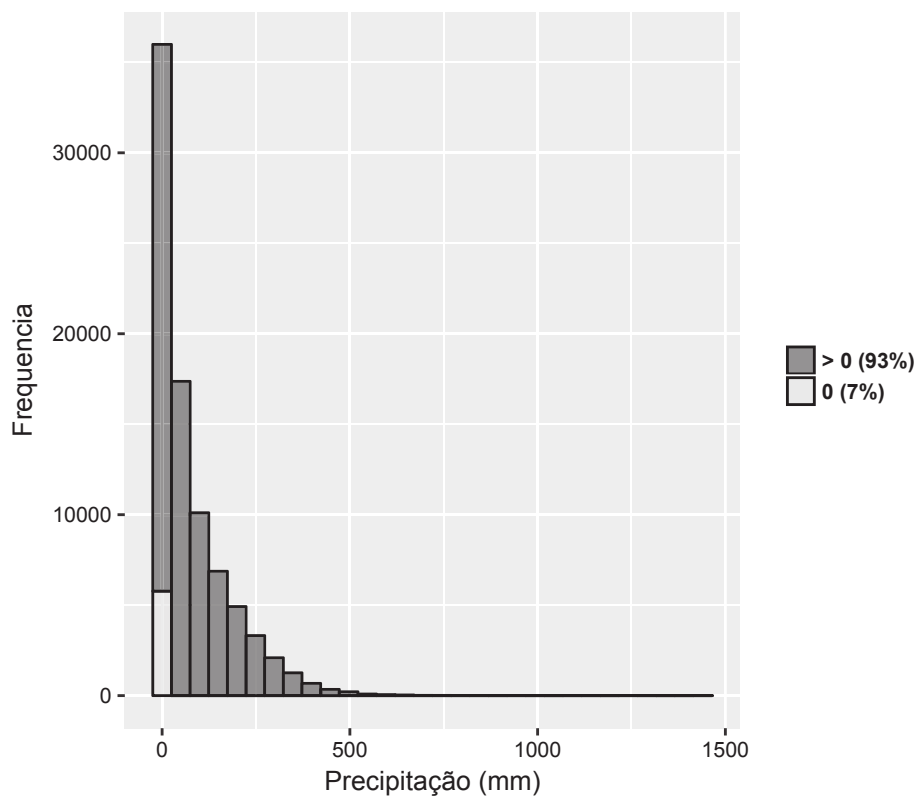
3 - Resultados

3.1 - Análise exploratória

Na Figura 2 tem-se o histograma da variável precipitação total mensal observada em todos os 193 pontos amostrados durante 35 anos. Nota-se, para esta variável, uma assimetria positiva, sendo que 7% das observações, são iguais a zero. O valor mediano da precipitação foi de 37,470 mm/mês com desvio padrão de 100 mm/mês. Adicionalmente, 25% das observações estiveram abaixo de 5 mm/mês de precipitação. Essas características climáticas da BHSF estão associadas à atuação e ao comportamento dos diversos sistemas atmosféricos e padrões das teleconexões que influenciam a média anual da precipitação variando entre 350 mm (Submédio

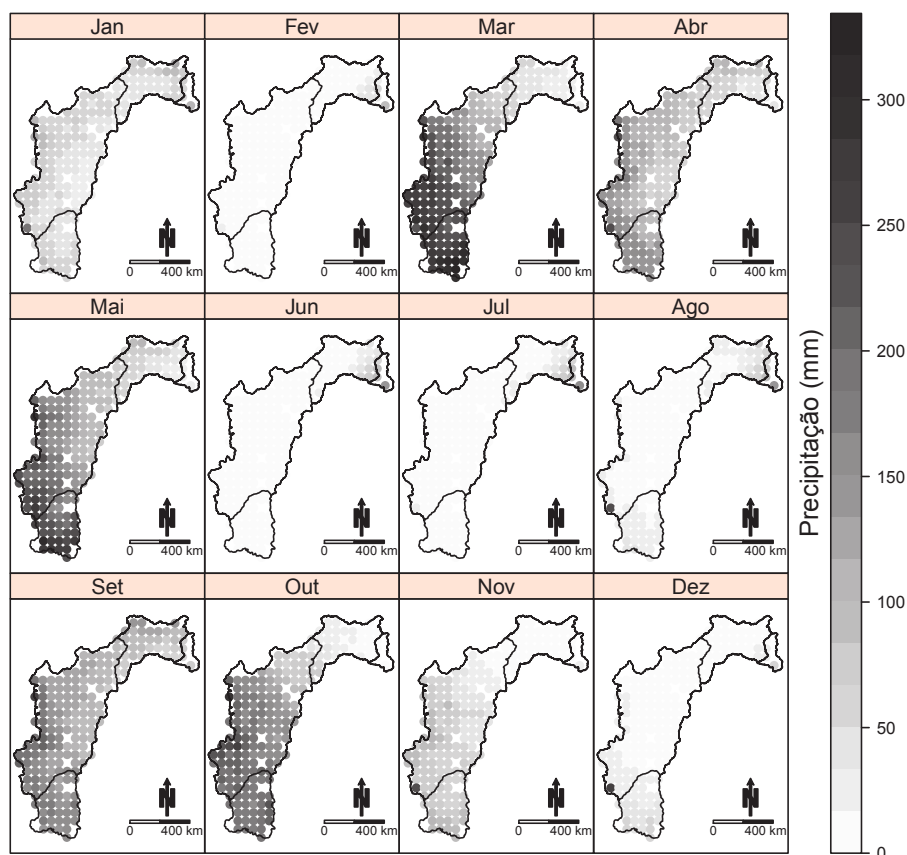
São Francisco) a 1500 mm (Alto São Francisco), deste modo caracterizando a alta variabilidade climática espacial da região (BEZERRA et al., 2018), como pode ser visto na Figura 3.

Figura 2 - Histograma da precipitação. As diferentes cores na figura representam as porcentagens de valores iguais ou maiores do que zero no conjunto de dados.



A Figura 3 destaca os mapas dos valores medianos, obtidos a partir da série histórica de 1979-2014, da precipitação para todos os 12 meses. Nota-se que, no geral, os maiores índices pluviométricos estão presentes na região do Alto São Francisco, seguida pelo Médio São Francisco, diferentemente das regiões do Baixo e Submédio São Francisco, que apresentam índices pluviométricos inferiores.

Figura 3 – Mapa do valor mediano da precipitação na BHSF, no período de 1979-2014.



Além da variabilidade espacial, há também uma alta variabilidade temporal na BHSF, justificando dessa forma a aplicação de um modelo que contemple a variabilidade espacial, temporal e, além disso, a interação espaço-tempo, uma vez que a distribuição das chuvas em determinadas áreas da BHSF apresenta uma variabilidade mensal.

3.2 - Geoestatística espaço-temporal

Após identificada a presença de variabilidade espacial e temporal, o próximo passo consistiu no ajuste do componente de tendência. Para o ajuste deste componente foram propostos dois modelos conforme descritos na Expressão (4). Para a modelagem dos resíduos oriundos desse ajuste, foi utilizada a função de covariância “*sum-metric*”. Por final, foi realizada a krigagem com regressão espaço-temporal, que combinou a krigagem sobre os resíduos com o ajuste do componente de tendência. A validação cruzada foi utilizada para determinar a escolha da melhor estrutura, que combinou diferentes modelos de regressão para a tendência com diferentes modelos teóricos de variograma para o resíduo estocástico.

Todos os resultados da validação cruzada são apresentados na Tabela 1. Cada linha dessa tabela contém os resultados relacionados ao ajuste do modelo estocástico espaço-temporal, utilizando a função de covariância “*sum-metric*”, para duas estruturas no componente de tendência (Modelo I e Modelo II). Ao todo, na krigagem com regressão, foram ajustadas 30 estruturas espaço-temporal para cada um dos dois modelos de tendência, resultando no ajuste de 60 modelos espaço-temporais.

Para o Modelo I, quando comparadas todas as estatísticas aplicadas na validação cruzada, em especial a estatística *Rank* o melhor ajuste é dado quando se utilizou o GAMLSS com a família ZAGA para o componente de tendência, e, para o resíduo estocástico, a função de covariância “*sum-metric*” com os variogramas Esférico, Periódico e Gaussiano nos seus componentes espacial, temporal e espaço-temporal, respectivamente. Neste modelo, a RMSE foi de 25,070 mm/mês, a MAE de 10,929 mm/mês, o CCC de 0,937 e o R^2 de 0,967.

Para o Modelo II, no ajuste do componente de tendência, mostra-se que o GAMLLS com família ZAGA apresenta os melhores resultados e, no resíduo estocástico, a função de covariância “*sum-metric*” com os variogramas Gaussiano, Periódico e Exponencial nos seus componentes espacial, temporal e espaço-temporal, respectivamente. Neste modelo a RMSE foi de 24,344 mm/mês, a MAE foi de 9,790 mm/mês, o CCC de 0,940 e o R^2 de 0,969.

Portanto, a família ZAGA se destaca, o que se deve principalmente às características da série, por esta apresentar dados mensais de precipitação iguais a zero em algum momento na escala temporal. Em ambos os modelos de tendência, todas as estatísticas utilizadas na validação cruzada indicam o melhor ajuste de um único modelo. Quando comparada as estatísticas entre os melhores ajustes para cada um dos modelos de tendência, nota-se que o Modelo II apresenta os melhores resultados, sendo este o escolhido para realizar as previsões e previsões, em locais e tempos não observados.

Tabela 1- Medidas calculadas na validação cruzada considerando os modelos teóricos de variogramas Esférico (Sph), Gaussiano (Gau), Exponencial (Exp) e Periódico (Per) para o espaço (S), tempo (T) e espaço-tempo (S x T) e diferentes distribuições de probabilidade (FDP), bem como considerando os Modelos I e II descritos na Equação (4).

FDP	Variograma “ <i>sum-metric</i> ”			Modelo 1					Modelo 2				
	S	T	S x T	RMSE	MAE	R^2	CCC	Rank	RMSE	MAE	R^2	CCC	Rank
ZAGA	Sph	Per	Gau	25,070	10,929	0,937	0,967	0,000	25,437	10,951	0,935	0,966	0,069
ZAGA	Sph	Per	Exp	27,088	12,685	0,926	0,961	0,097	24,838	10,356	0,938	0,968	0,033
ZAGA	Gau	Per	Exp	25,716	11,391	0,933	0,965	0,027	24,344	9,790	0,940	0,969	0,000
ZAGA	Gau	Per	Sph	27,436	13,084	0,925	0,960	0,118	27,596	13,002	0,924	0,960	0,192
ZAGA	Exp	Per	Exp	29,249	14,563	0,914	0,954	0,201	24,897	10,462	0,938	0,968	0,039
ZAGA	Exp	Per	Sph	27,249	13,238	0,924	0,960	0,124	29,089	14,287	0,915	0,955	0,271
GG	Sph	Per	Gau	30,321	15,220	0,908	0,951	0,240	25,687	11,176	0,934	0,966	0,083
GG	Sph	Per	Exp	25,706	11,316	0,934	0,965	0,023	24,916	10,402	0,938	0,968	0,036
GG	Gau	Per	Exp	27,993	13,410	0,921	0,958	0,138	24,495	9,916	0,940	0,969	0,008
GG	Gau	Per	Sph	27,554	13,135	0,924	0,960	0,122	27,550	12,894	0,924	0,960	0,186
GG	Exp	Per	Exp	28,168	13,556	0,920	0,958	0,146	24,940	10,474	0,937	0,968	0,040
GG	Exp	Per	Sph	27,585	13,251	0,924	0,959	0,127	28,904	14,093	0,916	0,956	0,259
LM	Sph	Per	Gau	26,829	12,266	0,928	0,962	0,076	25,582	11,099	0,934	0,966	0,078
LM	Sph	Per	Exp	25,523	11,084	0,934	0,966	0,012	24,845	10,360	0,938	0,968	0,034
LM	Gau	Per	Exp	25,626	11,191	0,934	0,966	0,018	26,586	12,188	0,929	0,963	0,142
LM	Gau	Per	Sph	27,633	13,160	0,924	0,959	0,123	41,370	24,159	0,911	0,902	0,881
LM	Exp	Per	Exp	26,659	12,183	0,929	0,963	0,071	26,010	11,607	0,932	0,964	0,108
LM	Exp	Per	Sph	27,602	13,201	0,924	0,959	0,125	29,087	14,199	0,915	0,955	0,266
NO	Sph	Per	Gau	27,487	12,894	0,924	0,960	0,110	25,425	10,884	0,935	0,966	0,065
NO	Sph	Per	Exp	27,979	13,431	0,921	0,958	0,139	24,676	10,148	0,939	0,968	0,021
NO	Gau	Per	Exp	26,349	11,979	0,930	0,964	0,059	24,425	9,858	0,940	0,969	0,004
NO	Gau	Per	Sph	27,433	13,055	0,925	0,960	0,117	27,289	12,845	0,925	0,960	0,182
NO	Exp	Per	Exp	27,435	12,958	0,924	0,960	0,113	24,920	10,462	0,938	0,968	0,040
NO	Exp	Per	Sph	27,491	13,200	0,924	0,960	0,124	28,637	13,919	0,918	0,956	0,248
GA	Sph	Per	Gau	27,526	13,015	0,924	0,960	0,116	26,005	11,443	0,932	0,965	0,099
GA	Sph	Per	Exp	26,327	11,966	0,930	0,964	0,058	25,002	10,417	0,937	0,968	0,038
GA	Gau	Per	Exp	31,990	16,775	0,897	0,945	0,325	24,693	10,061	0,939	0,968	0,017
GA	Gau	Per	Sph	27,433	13,057	0,925	0,960	0,117	28,457	13,519	0,919	0,957	0,226
GA	Exp	Per	Exp	28,942	14,258	0,916	0,955	0,184	25,691	11,221	0,934	0,965	0,085
GA	Exp	Per	Sph	27,488	13,202	0,924	0,960	0,124	29,735	14,686	0,911	0,953	0,296

Com base nos resultados apresentados na Tabela 1, foi selecionado o ajuste utilizando GAMLSS, com família ZAGA, para a componente de tendência. Para o resíduo estocástico, o melhor ajuste foi obtido com os modelos de variogramas Gaussiano, Periódico e Exponencial para os componentes espacial, temporal e espaço-temporal, respectivamente.

O variograma ajustado é dado por:

$$\gamma(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s) + \gamma_t(h_t) + \gamma_{st}\left(\sqrt{\mathbf{h}_s^2 + (\alpha h_t)^2}\right),$$

em que:

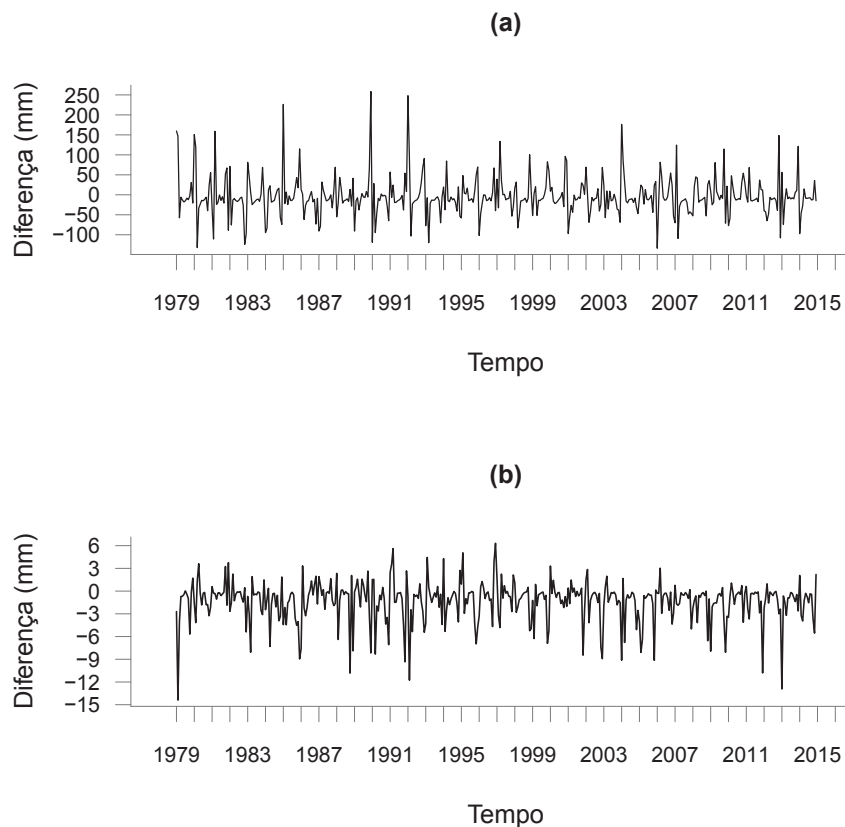
- $\gamma_s(\mathbf{h}_s) = 0,000 + 378,053 \left[1 - e^{-\left(\frac{\mathbf{h}_s}{76}\right)^2}\right]$
- $\gamma_t(h_t) = 2328,774 + 159,813 \left[1 - \cos\left(\frac{2\pi h_t}{371}\right)\right]$
- $\gamma_{st}\left(\sqrt{\mathbf{h}_s^2 + \alpha h_t^2}\right) = 0,000 + 1948,242 \left[1 - e^{-\left(\frac{\mathbf{h}_s^2 + (13h_t)^2}{196}\right)}\right]$

De acordo com estes resultados, a maioria da variação total dos resíduos é explicada pela componente espaço-temporal, o que indica que os padrões espacial e temporal estão presentes em sua interação, sugerindo uma correlação espacial de até 76 km e uma correlação temporal de até 1 ano, sendo este último justificado pela presença de um comportamento sazonal da precipitação. A anisotropia geométrica mostrou que as estações com uma defasagem temporal de 1 mês exibem uma correlação similar com estações separadas a 390 km.

3.3 - Desempenho do modelo

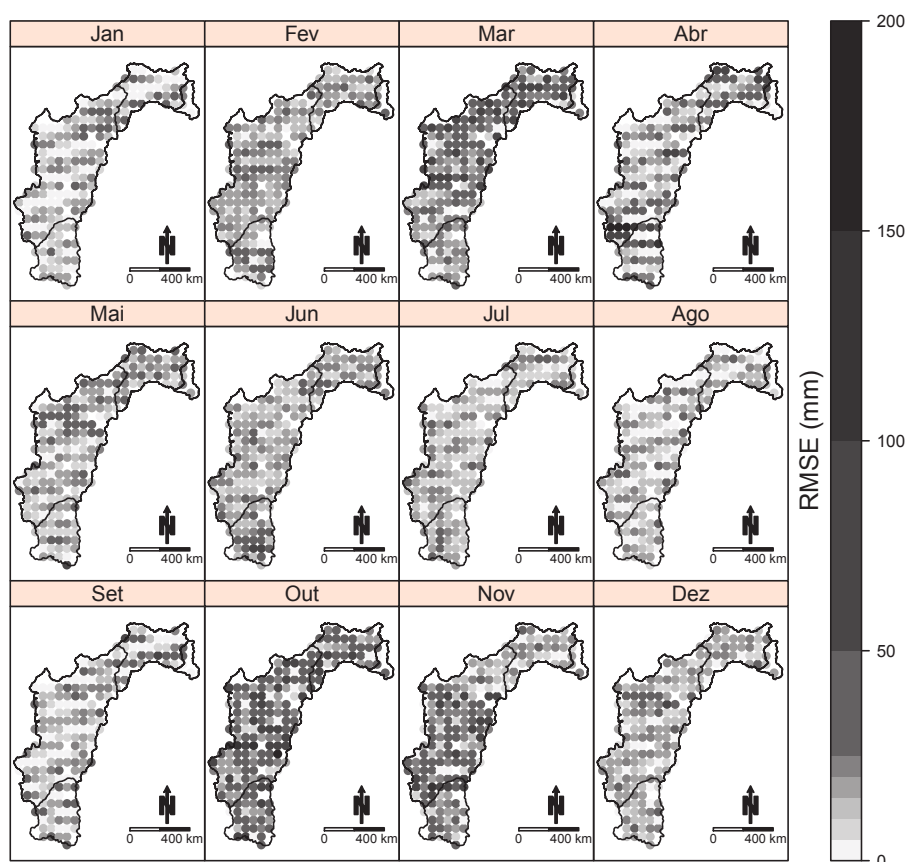
Na Figura 4 pode-se observar as diferenças entre os valores observados da precipitação e os valores da precipitação estimados por meio do componente de tendência (Figura 4a), e as diferenças entre os valores observados e os valores da precipitação preditos pela krigagem com regressão (Figura 4b). Nota-se que, não há um bom ajuste quando utilizado apenas o componente de tendência para prever a precipitação. Conforme o esperado, quando se utiliza a krigagem com regressão, nota-se que os valores preditos estão próximos dos observados, fornecendo a evidência de que a metodologia proposta apresentou um bom desempenho.

Figura 4 – Diferenças entre os valores observados da precipitação e os valores preditos pelo componente de tendência (a) e pela krigagem com regressão (b), durante o período de 1979-2014.



O desempenho do modelo pode ser verificado por meio da estatística RMSE obtida na validação cruzada (Figura 5). Nesta figura pode ser verificado que o modelo estima de forma razoável os dados observados de precipitação em quase todos os meses do ano. As exceções são algumas áreas da região do Médio e Submédio São Francisco nos meses chuvosos de março, abril, outubro e novembro. A distribuição das chuvas nesses períodos são mais sensíveis ao comportamento da temperatura do Oceano Pacífico Equatorial, com influência significativa de fenômenos cíclicos, especialmente os associados ao ENOS (El Niño-Oscilação Sul). Assim, é importante associar a previsão (ou prognóstico) com um intervalo de confiança para as localidades, reduzindo assim a incerteza nestes períodos.

Figura 5 – Mapa do valor médio, com base no período de 1979-2014, da estatística RMSE obtida por meio da validação cruzada.

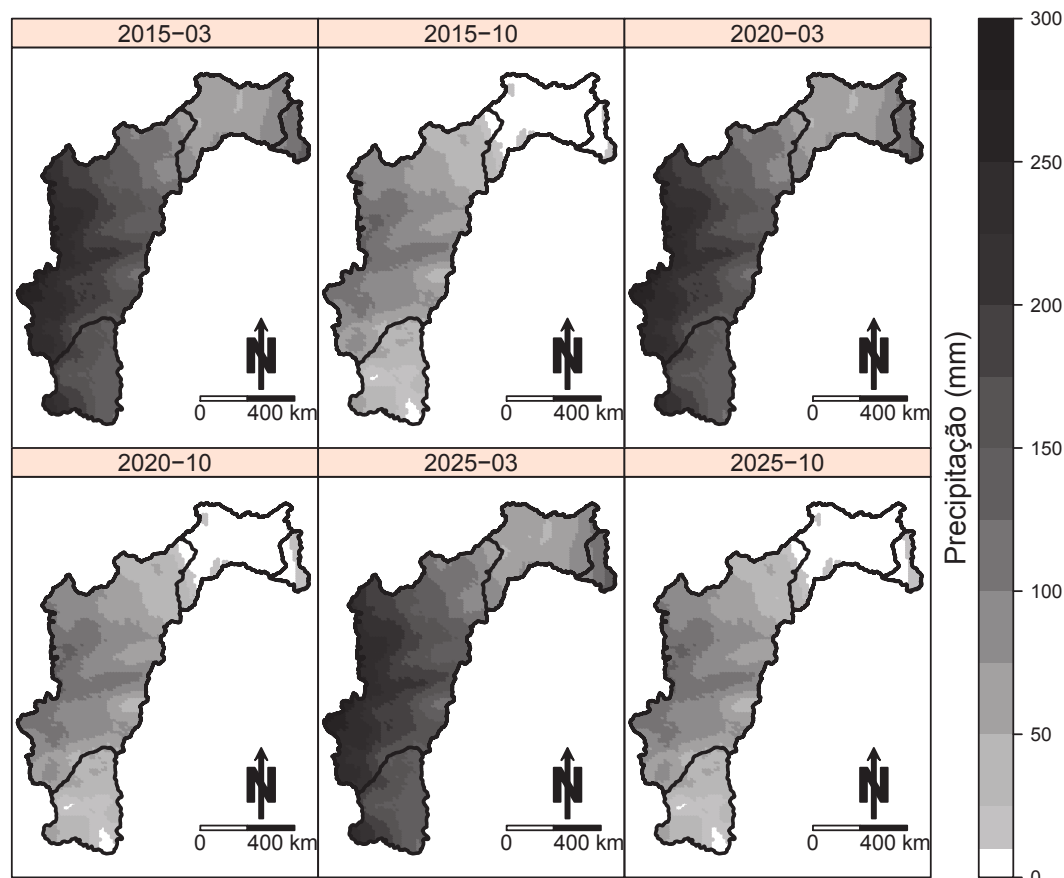


3.4 - Predição espaço-temporal

Como ilustração da krigagem com regressão, na Figura 6 é possível observar a predição dos valores de precipitação da BHSF nos meses de março e outubro para o ano de 2015, 2020 e 2025, em que o erro associado a predição e o prognóstico podem ser verificados na Figura 7.

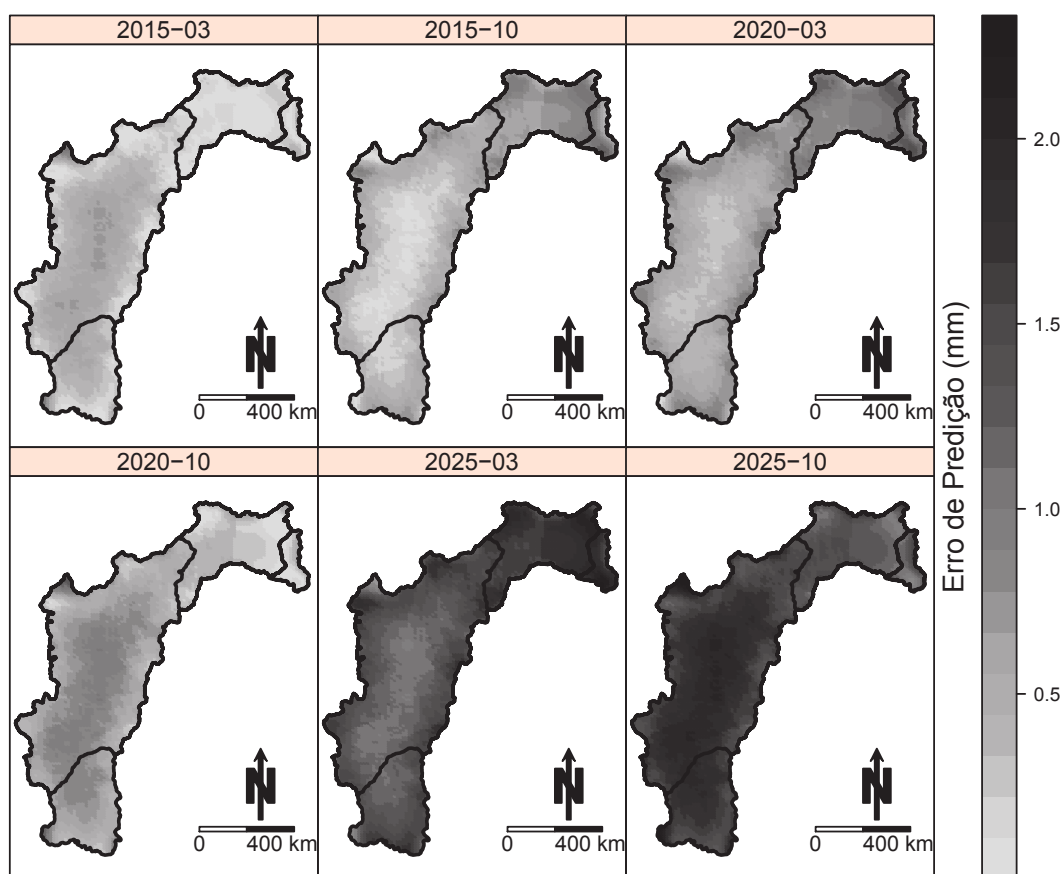
Esse tipo de abordagem é essencial, a fim de detectar possíveis alterações na distribuição das chuvas nessa região, possibilitando desta forma implementações de ações mitigatórias preventivas (ARAÚJO; CELESTE, 2018).

Figura 6 – Mapas dos valores previstos da precipitação para BHSF nos meses de março e outubro, nos anos de 2015, 2020 e 2025. A cor branca dentro do mapa representa o nível acumulado de precipitação abaixo de 10 mm/mês.



Nota-se que o volume total mensal da precipitação tende a decrescer em todas as quatro regiões da bacia (Figura 6). Resultados similares foram encontrados por Marengo et al. (2012), onde foi observado projeções decrescentes na precipitação, com redução de 15% no volume da precipitação anual até 2040 na BHSF, podendo chegar até 20% no período chuvoso (BEZERRA et al., 2018; SILVEIRA et al., 2016; SANTOS et al., 2018). Este resultado pode significar que a diminuição no volume da recarga do Rio aumentará, gerando riscos climáticos em um futuro bem próximo, ocasionando ainda mais problemas na economia do país por falta dos recursos hídricos principalmente para consumo humano, agricultura e produção de energia no NEB.

Figura 7 – Mapas dos erros de previsão da precipitação para BHSF nos meses de março e outubro, nos anos de 2015, 2020 e 2025.



4 - Discussões e conclusão

Neste artigo foi apresentada uma nova forma de modelar o componente de tendência presente no processo estocástico espaço-temporal. Os modelos de regressão GAMLSS (STASINOPOULOS et al., 2017) tornam-se uma poderosa ferramenta na modelagem dessa tendência, uma vez que permitem a presença de termos não paramétricos e a utilização de diferentes famílias de distribuições de probabilidade no ajuste da regressão. A modelagem desse componente é de suma importância, pois este é responsável pela variação em larga escala do fenômeno a ser estudado. Os resultados demonstraram que os modelos de regressão GAMLSS são adequados e preferíveis no ajuste da regressão. Estudos recentes têm sugerido o ajuste desses modelos em fenômenos que apresentam características que fogem de uma distribuição normal, como dados quantitativos discretos, censurados e com excessos de zero em variáveis quantitativas contínuas (SÁ et al., 2018; STAUFFER et al., 2016).

Para aplicação da metodologia proposta neste artigo foi utilizado o conjunto de dados de precipitação total mensal, avaliado no espaço-tempo na região da BHSF. Uma análise descritiva destes dados identificou que a precipitação apresentava uma distribuição assimétrica positiva, com excesso de valores zero e uma distribuição irregular do valor mediano mensal no espaço e no tempo, indicando a necessidade de uma modelagem espaço-temporal nestes dados. Devido a estes fatores, partiu-se da hipótese de que o modelo de regressão Normal poderia não ser o mais adequado no ajuste da regressão múltipla para o componente de tendência, sugerindo assim a modelagem da regressão GAMLSS com distribuições ZAGA, GG, GA e a distribuição Normal. O modelo de regressão Normal tem sido amplamente utilizado no ajuste do componente de tendência em fenômenos indexados no espaço-tempo (IACO et al., 2015; GRÄLER et al., 2012; HU et al., 2015; KILIBARDA et al., 2014). Entretanto, estudos recentes têm sugerido o ajuste de modelos lineares generalizados ao avaliar variáveis com distribuições assimétricas (MENEZES et al., 2016; MONTEIRO et al., 2017). Assim, a proposta desse artigo consiste em considerar os modelos GAMLSS, com diferentes distribuições de probabilidades, no ajuste do componente de tendência presente no processo estocástico espaço-temporal.

Para o conjunto de dados analisados neste artigo foi visto que, após a modelagem do componente de tendência, os resíduos produzidos apresentaram dependência espaço-temporal, os quais foram modelados pela função de covariância “sum-metric”. Semelhantemente, trabalhos que sugeriram modelar um processo estocástico espaço-temporal utilizaram esta mesma função (GRÄLER et al., 2012; KILIBARDA et al., 2014). A grande complicação de se utilizar este modelo consiste na quantidade de parâmetros (10 parâmetros) à serem estimados, resultando em dificuldades de convergência do algoritmo utilizado para obtenção das estimativas.

A krigagem com regressão espaço-temporal, que combinou a krigagem sobre os resíduos com a regressão GAMLSS, foi utilizada para analisar os dados de precipitação da BHSF. A metodologia apresentada neste artigo possibilitou a criação de mapas de krigagem para a área em estudo, interpolando locais não observados e predizendo estimativas da precipitação em anos futuros. Sabe-se que a construção destes mapa, em especial à BHSF, que fica localizada, em sua maioria, na região do semiárido nordestino o qual é caracterizado pelas altas temperaturas e períodos de estiagem, é de extrema importância para o planejamento e desenvolvimentos de atividades relacionadas com os recursos hídricos da região, identificando

o efeito de mudanças climáticas e conscientizando o governo e a população sobre os impactos e a vulnerabilidade destas mudanças.

REFERÊNCIAS

- ANA. **Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas (ANA)**. Brasília, 2017.
- ARAÚJO, M. V. O.; CELESTE, A. B. Rescaled range analysis of streamflow records in the São Francisco River Basin, Brazil. **Theoretical and Applied Climatology**, Springer Nature, 2018. ISSN 0177-798X.
- BASTOS, R. L. **Métodos para seleção de modelos de semivariograma em campos aleatórios gaussianos**. Tese (Doutorado em Estatística e Experimentação Agropecuária) — Universidade Federal de Lavras, Lavras, 2017. Disponível em: <<http://repositorio.ufla.br/handle/1/15568>>.
- BEZERRA, B. G. et al. Changes of precipitation extremes indices in São Francisco River Basin, Brazil from 1947 to 2012. **Theoretical and Applied Climatology**, Springer Nature, 2018.
- CAVALCANTI, I. F. A. **Tempo e clima no Brasil**. São Paulo, SP: Oficina de textos, 2016.
- CHEN, M. et al. Assessing objective techniques for gauge-based analyses of global daily precipitation. **Journal of Geophysical Research**, American Geophysical Union (AGU), v. 113, n. D4, 2008.
- CRESSIE, N.; HUANG, H.-C. Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. **Journal of the American Statistical Association**, Informa UK Limited, v. 94, n. 448, p. 1330–1339, 1999.
- EILERS, P. H.; MARX, B. D.; DURBÁN, M. Twenty years of P-splines. v. 39, n. 2, p. 149–186, 2015.
- GNEITING, T. Nonseparable, Stationary Covariance Functions for Space–Time Data. **Journal of the American Statistical Association**, Informa UK Limited, v. 97, n. 458, p. 590–600, 2002.
- GRÄLER, B.; PEBESMA, E.; HEUVELINK, G. Spatio-temporal interpolation using gstat. **R Journal**, v. 8, n. 1, p. 204–218, 2016.
- GRÄLER, B. et al. Spatio-temporal analysis and interpolation of PM10 measurements in Europe for 2009. **ETC/ACM Technical Paper**, v. 8, p. 1–29, 2012.

HENGL, T. et al. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. **Theoretical and Applied Climatology**, Springer Nature, v. 107, n. 1-2, p. 265–277, 2011.

HEUVELINK, G. et al. Sampling Design Optimization for Space-Time Kriging. In: _____. **Spatio-Temporal Design**. Hoboken, Nova Jersey, EUA: Wiley-Blackwell, 2012. cap. 9, p. 207–230. ISBN 9781118441862.

HEUVELINK, G. B. M.; GRIFFITH, D. A. Space-Time Geostatistics for Geography: A Case Study of Radiation Monitoring Across Parts of Germany. **Geographical Analysis**, Wiley-Blackwell, v. 42, n. 2, p. 161–179, 2010.

HU, D. et al. Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. **Cluster Computing**, Springer Nature, v. 20, n. 1, p. 347–357, 2017.

HU, Y. et al. Spatio-temporal Transmission and Environmental Determinants of Schistosomiasis Japonica in Anhui Province, China. **PLOS Neglected Tropical Diseases**, Public Library of Science (PLoS), v. 9, n. 2, p. e0003470, 2015.

IACO, S.; MYERS, D.; POSA, D. Space–time analysis using a general product–sum model. **Statistics & Probability Letters**, Elsevier BV, v. 52, n. 1, p. 21–28, 2001.

IACO, S. D. Space–time correlation analysis: a comparative study. **Journal of Applied Statistics**, Informa UK Limited, v. 37, n. 6, p. 1027–1041, 2010.

IACO, S. D.; PALMA, M.; POSA, D. Spatio-temporal geostatistical modeling for French fertility predictions. **Spatial Statistics**, Elsevier BV, v. 14, p. 546–562, 2015.

KILIBARDA, M. et al. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. **Journal of Geophysical Research: Atmospheres**, American Geophysical Union (AGU), v. 119, n. 5, p. 2294–2313, 2014.

KILIBARDA, M. et al. Global geographic and feature space coverage of temperature data in the context of spatio-temporal interpolation. **Spatial Statistics**, Elsevier BV, v. 14, p. 22–38, 2015.

MANETA, M. et al. Water demand and flows in the São Francisco River Basin (Brazil) with increased irrigation. **Agricultural Water Management**, Elsevier BV, v. 96, n. 8, p. 1191–1200, aug 2009.

MARENGO, J. A. et al. 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. **Climate Dynamics**, Springer Nature, v. 38, n. 9-10, p. 1829–1848, aug 2011.

MARTÍNEZ, W. A.; MELO, C. E.; MELO, O. O. Median Polish Kriging for space–time analysis of precipitation. **Spatial Statistics**, Elsevier BV, v. 19, p. 1–20, 2017.

MENEZES, R. et al. Spatial–temporal modellization of the NO_2 concentration data through geostatistical tools. **Statistical Methods & Applications**, Springer Nature, v. 25, n. 1, p. 107–124, 2016.

MMA. **São Francisco - Caderno da Região Hidrográfica**. Brasília, 2006. Disponível em: <<http://www.mma.gov.br/>>.

MONTEIRO, A.; MENEZES, R.; SILVA, M. E. Modelling spatio-temporal data with multiple seasonalities: The NO_2 Portuguese case. **Spatial Statistics**, Elsevier BV, v. 22, p. 371–387, 2017.

MONTERO, J. M.; FERNANDEZ-AVILES, G.; MATEU, J. **Spatial and spatio-temporal geostatistical modeling and kriging**. Hoboken, Nova Jersey, EUA: John Wiley & Sons, 2015.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, Copernicus GmbH, v. 11, n. 5, p. 1633–1644, 2007.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>.

RAJA, N. B. et al. Space-time kriging of precipitation variability in Turkey for the period 1976–2010. **Theoretical and Applied Climatology**, Springer Nature, v. 129, n. 1-2, p. 293–304, 2016.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley-Blackwell, v. 54, n. 3, p. 507–554, 2005.

SÁ, A. C. L.; TURKMAN, M. A. A.; PEREIRA, J. M. C. Exploring fire incidence in Portugal using generalized additive models for location, scale and shape (GAMLSS). **Modeling Earth Systems and Environment**, Springer Nature, v. 4, n. 1, p. 199–220, 2018.

SANTOS, C. A. G. et al. Integrated spatiotemporal trends using TRMM 3B42 data for the Upper São Francisco River basin, Brazil. **Environmental Monitoring and Assessment**, Springer Nature, v. 190, n. 3, feb 2018.

SILVEIRA, C. et al. Climate change in the São Francisco river basin: analysis of precipitation and temperature. **Revista Brasileira de Recursos Hídricos**, Associação Brasileira de Recursos Hídricos - ABRH, v. 21, n. 2, p. 416–428, 2016.

SNEPVANGERS, J.; HEUVELINK, G.; HUISMAN, J. Soil water content interpolation using spatio-temporal kriging with external drift. **Geoderma**, Elsevier BV, v. 112, n. 3-4, p. 253–271, 2003.

STASINOPOULOS, M. D. et al. **Flexible Regression and Smoothing: Using GAMLSS in R**. Boca Raton, Flórida, EUA: CRC Press, 2017.

STAUFFER, R. et al. Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. **International Journal of Climatology**, Wiley, v. 37, n. 7, p. 3264–3275, 2016.