

ALYNE NEVES SILVA

**DETECÇÃO DE OUTLIERS EM SÉRIES ESPAÇO-
TEMPORAIS: ANÁLISE DE PRECIPITAÇÃO EM MINAS
GERAIS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS-BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

S586d
2012

Silva, Alyne Neves, 1977-

Detecção de *outliers* em séries espaço-temporais: análise de precipitação em Minas Gerais / Alyne Neves Silva. – Viçosa, MG, 2012.

xii, 69f. : il. ; (algumas col.) ; 29cm.

Inclui anexos.

Orientador: Nerilson Terra Santos

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 51-54

1. Análise de séries temporais. 2. Precipitação (Meteorologia) - Previsão. 3. Mínimos quadrados.
I. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-Graduação em Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.55

ALYNE NEVES SILVA

**DETECÇÃO DE OUTLIERS EM SÉRIE ESPAÇO-
TEMPORAIS: ANÁLISE DE PRECIPITAÇÃO EM
MINAS GERAIS**

Dissertação apresentada à
Universidade Federal de Viçosa,
como parte das exigências do
Programa de Pós-Graduação em
Estatística Aplicada e Biometria,
para obtenção do título de *Magister
Scientiae*.

APROVADA: 24 DE JULHO DE 2012.

Prof. Silvio Bueno Pereira

Prof. Eric Batista Ferreira

Prof. Gérson Rodrigues dos Santos
(Presidente da Banca)

“O Bom Combate é aquele que é travado em nome de nossos sonhos. Quando eles explodem em nós com todo o seu vigor - na juventude - nós temos muita coragem, mas ainda não aprendemos a lutar. Depois de muito esforço, terminamos aprendendo a lutar, e então já não temos a mesma coragem para combater. Por causa disto, nos voltamos contra nós e combatemos a nós mesmos, e passamos a ser nosso pior inimigo. Dizemos que nossos sonhos eram infantis, difíceis de realizar, ou fruto de nosso desconhecimento das realidades da vida. Matamos nossos sonhos porque temos medo de combater o Bom Combate.”

Paulo Coelho (O Bom Combate)

À minha família.

AGRADECIMENTOS

A Deus, que se mostrou criador, que foi criativo. Seu fôlego de vida em mim me foi sustento e me deu coragem para questionar realidades e propor sempre um novo mundo de possibilidades.

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação me deu, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

Aos meus amigos, pelas alegrias, tristezas e dores compartilhadas. Com vocês, as pausas entre um parágrafo e outro de produção melhora tudo o que tenho produzido na vida.

Aos Professores Gérson Rodrigues dos Santos, Nerilson Terra Santos e Fernando Falco Pruski pela grande ajuda nessa empreitada. Posso dizer que a minha formação não teria sido a mesma sem a ajuda de vocês.

Aos meus companheiros de mestrado, ao Programa de Pós-Graduação em Estatística Aplicada e Biometria (PPESTBIO) da Universidade Federal de Viçosa e às pessoas com quem convivi nesse espaço de tempo em que vivi em Viçosa, MG. A experiência de uma produção compartilhada na comunhão com amigos nesses espaços foram a melhor experiência da minha formação acadêmica.

À Maola Faria e a Andréia Lopes por serem minhas grandes amigas.

A todos aqueles que de alguma forma estiveram e estão próximos de mim, fazendo esta vida valer cada vez mais a pena.

LISTA DE FIGURAS

Figura 1 - Exemplo de semivariograma típico.....	19
Figura 2 - Regiões climaticamente homogêneas do Estado de Minas Gerais..	25
Figura 3 - Mapa das posições das 43 estações pluviométricas utilizadas neste estudo.	27
Figura 4 - Processo de análise de séries temporais.	29
Figura 5 - Processo de análise geoestatística.	30
Figura 6 – Detecção de <i>outlier</i>	32
Figura 7 – Série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde para o período de janeiro de 2000 a novembro de 2005.	34
Figura 8 – Funções de autocorrelação da série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde. (a) Amostral; (b) Parcial.	35
Figura 9 – Periodograma.	36
Figura 10 – Funções de autocorrelação da série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde sem periodicidade. (a) Amostral; (b) Parcial.	38
Figura 11 – Análise exploratória dos valores da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas. (a) Valores divididos com peso igual; (b) valores divididos com peso proporcional; (c) Valores divididos pelos quartis e (d) Valores divididos pelos decis.	40
Figura 12– Análise de tendência da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas.	41
Figura 13 -Semivariograma experimental da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas para o mês de janeiro de 2000.	42
Figura 14 -Modelos de semivariograma empírico: (a) Exponencial, (b) Gaussiano e (c) Esférico.....	42
Figura 15 -Modelo de semivariograma estimado pelos métodos. (a) Esférico - Mínimos Quadrados Ordinários; (b) Esférico - Máxima Verossimilhança; (a) Exponencial - Mínimos Quadrados Ordinários; (b) Exponencial - Máxima Verossimilhança.....	43
Figura 16 - Semivariograma experimental dos resíduos do modelo de semivariograma esférico ajustado pelo método de mínimos quadrados. (a) Distância = 3; (b)Distância = 5 e (c) Distância = 7.	45

LISTA DE QUADROS

Quadro 1 - Municípios selecionados com base nas regiões climaticamente homogêneas do Estado de Minas Gerais de acordo com Aspiazu et. al. (1990, apud CARAM, 2007).....	26
Quadro 2 - Sumarização dos resultados da autovalidação.....	44

LISTA DE TABELAS

Tabela 1 - Frequências do periodograma	36
Tabela 2 – Modelo de regressão estimado para descrever a precipitação média mensal por número de dias chuvosos	37
Tabela 3 – Critério de Informação de Akaike dos modelos ajustados	39
Tabela 4 – Coeficientes estimados para o modelo <i>ARMA</i> (1,1).....	39
Tabela 5 -Teste de detecção de <i>outlier</i> para as 43 estações pluviométricas observadas, apresentando seus respectivos intervalos de confiança temporal de 95% e resíduos da autovalidação.....	47
Tabela 6 -Total de <i>outliers</i> detectados anualmente por estação pluviométrica	48
Tabela 7–Total mensal de <i>outliers</i> detectados anualmente.....	49

LISTA DE SIGLAS

ADF	Dickey-Fuller Aumentado
AIC	Critério de Informação de Akaike
ANA	Agência Nacional de Águas
AR	Autorregressivo
ARIMA	Autorregressivo Média Móvel Integrado
ARMA	Autorregressivo Média Móvel
BIC	Critério de Informação Bayesiano
BLUP	<i>Best Linear Unbiased Predictor</i>
DF	Dickey-Fuller
DHF	Dickey-Hasza-Fuller
FAC	Função de Autocorrelação Amostral
FACP	Função de Autocorrelação Parcial
IC	Intervalo de Confiança
ML	<i>Maximum Likelihood</i>
MA	Média Móvel
OLS	<i>Ordinary Least Squares</i>
PP	Phillips-Perron
SARIMA	Autorregressivo Média Móvel Integrado Sazonal
SSD	Simulação Sequencial Direta
ZCAS	Zona de Convergência do Atlântico Sul

ÍNDICE

AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT	vi
LISTA DE FIGURAS	vii
LISTA DE QUADROS	viii
LISTA DE TABELAS	ix
LISTA DE SIGLAS	x
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	4
2.1 PRECIPITAÇÃO PLUVIAL	4
2.2 SÉRIES TEMPORAIS	5
2.2.1 Notações e Conceitos	6
2.2.2 Tendência	6
2.2.3 Periodicidade	9
2.2.4 Modelo ARIMA	10
2.2.5 Identificação, Estimação e Adequação de um Modelo de Séries Temporais	11
2.3 GEOESTATÍSTICA.....	15
2.3.1 Notações e Conceitos	17
2.3.2 Estacionariedade	18
2.3.3 Modelos de Semivariograma.....	19
2.3.4 Seleção de Modelos.....	23
3 MATERIAL E MÉTODOS	25
3.1 REGIÃO DE ESTUDO	25
3.2 BASE DE DADOS	25
3.2 LINGUAGEM COMPUTACIONAL	28
3.3 DETECÇÃO DE <i>OUTLIERS</i>	28
3.3.1 Análise de Séries Temporais	28
3.3.2 Análise Geoestatística.....	30
3.3.3 Detecção de <i>Outliers</i>	31
4 RESULTADOS E DISCUSSÃO	34
4.1 RESULTADOS DE SÉRIES TEMPORAIS	34
4.2 RESULTADOS DE GEOESTATÍSTICA	40

4.3 DETECÇÃO DE <i>OUTLIER</i>	46
5 CONCLUSÃO	50
REFERÊNCIAS BIBLIOGRÁFICAS	51
ANEXO A – LOCALIZAÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS	55
ANEXO B – SÉRIES DE PRECIPITAÇÃO MÉDIA POR DIA	
CHUVOSO	56
ANEXO C - ROTINA DA ANÁLISE DE SÉRIES TEMPORAIS.....	66
ANEXO D - ROTINA DA ANÁLISE DE GEOESTATÍSTICA.....	68

RESUMO

SILVA, Alyne Neves, M.Sc., Universidade Federal de Viçosa, julho de 2012. **Detecção de *Outliers* em Séries Espaço-Temporais: Análise de Precipitação em Minas Gerais.** Orientador: Nerilson Terra Santos. Coorientadores: Gérson Rodrigues dos Santos e Fernando Falco Pruski.

Séries temporais são algumas vezes influenciadas por interrupções de eventos, tais como greves, eclosão de guerras, entre outras. Estas interrupções originam observações atípicas ou *outliers* que influenciam diretamente na homogeneidade da série, ocasionando interpretações e inferências errôneas da variável sob estudo, sendo muito comum em dados climatológicos. Assim, com o interesse de detectar *outliers* em séries temporais de precipitação, o presente trabalho teve por objetivo estabelecer um método de detecção *outliers*. Para tal, realizou-se a junção da modelagem ARIMA e de uma das metodologias clássicas de geoestatística, a autovalidação. O critério proposto compara os resíduos da análise de séries temporais com intervalos de confiança dos resíduos da autovalidação. Foram analisadas séries temporais da precipitação média mensal por dias chuvosos de 43 estações pluviométricas localizadas no estado de Minas Gerais, entre os anos de 2000 a 2005. Os procedimentos de análise vão da descrição da periodicidade por meio do periodograma até a obtenção da autovalidação, à partir da estimação dos modelos de semivariograma pelos métodos de mínimos quadrados ordinários e máxima verossimilhança. Pelos resultados, para o período sob estudo, foram detectado 165 *outliers*, espalhados entre as 43 estações pluviométricas. A estação Fazenda Campo Grande, localizada no município de Passa Tempo, foi a estação em que se registrou o maior número de *outliers*, 45 no total. Conforme os resultados obtidos considerou-se o método proposto muito eficiente na detecção de *outliers* e, conseqüentemente, na análise da homogeneidade das observações.

ABSTRACT

SILVA, Alyne Neves, M.Sc., Universidade Federal de Viçosa, julho de 2012. **Outliers Detection in Space-Time Series: Analysis of rainfall in Minas Gerais.** Adviser: Nerilson Terra Santos. Co-advisers: Gérson Rodrigues dos Santos and Fernando Falco Pruski.

Time series are sometimes influenced by disruptions of events, such as strikes, the outbreak of war, among others. These interrupts originate atypical observations or outliers that directly influence the homogeneity of the series, leading to erroneous inferences and interpretations of the variable under study, being very common in climatological data. So, in the interest of detecting outliers in time series of precipitation, this study aimed to establish a method of detecting outliers. For this, there was the junction of ARIMA models and methodologies of the classical geostatistics, the self-validation. The proposed criterion compares waste of time series analysis with confidence intervals of the residue of self-validation. We analyzed time series of average monthly rainfall for rainy days of 43 rainfall stations in the state of Minas Gerais, between the years 2000 to 2005. The analysis procedures ranging from the description of the periodicity through the periodogram to obtain validation, from the estimation of the semivariogram models by ordinary least squares methods and maximum likelihood. The results for the period under study, 165 were detected outliers, spread between the 43 rainfall stations. The station Campo Grande Ranch, located in the municipality of Passa Tempo, was the season in which they recorded the highest number of outliers, 45 in total. As the results, we considered the proposed method very efficient in detecting outliers, and therefore the analysis of the homogeneity of observations.

1 INTRODUÇÃO

Eventos da natureza produzem informações que necessitam de uma análise sofisticada para se entender a característica de um fenômeno, a fim de se realizar inferências com alguma margem de erro, possibilitando a tomada de decisões e/ou ações. Elementos climáticos como precipitação, temperatura, umidade relativa do ar, entre outros, exercem influência direta nas atividades cotidianas da população e no processo produtivo de bens (produtos e serviços).

Diversos estudos que caracterizam o comportamento de elementos climáticos consideram em suas metodologias o tempo e uma região de interesse. Quando um conjunto de observações, seja climático ou não, é analisado sequencialmente ao longo do tempo, denomina-se este como uma série temporal (MORETTIN e TOLÓI, 2002).

Em se tratando de Séries Temporais podem ocorrer intervenções no decorrer do registro temporal, tais como greves, eclosão de guerras, mudanças políticas ou crises econômicas, calor inesperado, ondas de frio, entre outros (WEI, 2006). As ocorrências destas intervenções causam a presença de observações inconsistentes das demais observações da série, denominadas *outliers*, que podem influenciar diretamente a homogeneidade da série, isto é, alterando a variabilidade entre as observações ao longo do tempo, conduzindo a uma interpretação errônea e impossibilitando conjecturas da variável sob estudo (WEI, 2006). Por definição, *outliers* são observações que se desviam acentuadamente dos outros membros da amostra em que ocorrem, ocasionando inconsistências. Quando a causa e o tempo da intervenção são conhecidos pode-se empregar a técnica denominada *análise de intervenção* (BOX e TIAO, 1975), para avaliar o efeito destes *outliers* na série temporal em estudo. No procedimento, supõe-se que a intervenção afeta o processo, alterando a função média ou a tendência da série temporal, de forma a se caracterizar uma quebra na estrutura da série. Contudo, nem sempre se conhece o tempo em que as intervenções ocorrem. Neste caso, é importante se obter procedimentos que irão detectar e remover os efeitos dos valores aberrantes que causam estragos na análise de dados, tornando a inferência resultante duvidosa. A detecção de *outliers* em séries temporais foi iniciada por

Fox (1972). Atualmente, na estatística, paramétrica ou não-paramétrica, existem diversos métodos para detectar e remover os efeitos dos *outliers* em séries temporais. No entanto, para séries de tempo observadas em diferentes pontos, as metodologias comumente empregadas não consideram a relação espacial que pode ocorrer entre as observações de cada ponto amostrado.

Na climatologia, a não-homogeneidade devido a presença de *outliers* pode ocorrer por diversos motivos, tais como mudança do local de observação, erros instrumentais, erros dos observadores, fechamento de estações de monitoramento, entre outros. Ou seja, confiabilidade dos dados sob análise pode ser indeterminada, mediante as inconsistências que ocorrem no momento em que se estabelece algum padrão para um fator climático. Em se tratando de um fator climático de grande importância, tal como a precipitação pluvial, a interpretação errada de seu comportamento é preocupante, pelo mesmo interferir em quase todos os processos de produção. Por exemplo, na agricultura, em que suas variações geográfica, temporal e sazonal interferem no plantio de um cultivar, que além de dar origem a alimentos, origina fibras, matéria-prima, dentre outros elementos utilizados na produção de bens e serviços.

A variação territorial é de grande influência em estudos relacionados na climatologia. Tobler (1979 apud CÂMARA e MEDEIROS, 1998) afirma que “no mundo, todas as coisas se parecem; mas coisas mais próximas são mais parecidas que aquelas mais distantes”. Assim, ao se realizar uma análise para detectar *outliers* em séries climatológicas é interessante levar em consideração observações no entorno do local pontualmente avaliado.

Uma ferramenta que vem sendo empregada na análise e na interpretação de elementos climáticos é a *Geoestatística*. A geoestatística considera a autocorrelação espacial que ocorre entre as observações regionalizadas para prever valores de atributos ou variáveis (CARAM, 2007), isto é, pode-se prever matematicamente, por meio das locações vizinhas, pontos amostrados ou não na área sob estudo. Tal como na estatística clássica, a geoestatística estabelece um algoritmo adequado para cada tipo de dado ou informação (MORAL, 2010).

Dentre os métodos para a detecção de *outliers* em séries temporais atualmente empregados têm-se os testes de Pettitt (PETTITT, 1979), da

Amplitude de Buishand (BUISSHAND, 1981 e 1982) e o de Homogeneidade Normal Padrão (ALEXANDERSSON, 1986). Estes não consideram em seus procedimentos a localização de observações vizinhas como uma componente que pode vir corroborar ou não a existência de *outliers*. Em contrapartida, Costa, Negreiros e Soares (2008) propõem a SSD como uma abordagem para a detecção de *outliers* em séries temporais considerando as observações espaciais e temporais da vizinhança de forma a detectar e classificar os *outliers*.

Neste contexto, este trabalho tem como objetivo geral a proposta de um método de detecção de *outliers* para avaliar a homogeneidade em séries espaço-temporais de precipitação observadas em estações de monitoramento pluviométrico localizadas no Estado de Minas Gerais. Para tal, os objetivos específicos são:

- Identificar, estimar e adequar modelos de séries temporais;
- Comparar modelos de semivariograma;
- Comparar métodos de estimação de modelos de semivariograma.

2 REFERENCIAL TEÓRICO

Neste capítulo apresenta-se o enquadramento teórico da temática necessária para a compreensão do objetivo deste estudo. Inicialmente apresenta-se a definição do elemento climático de interesse, a precipitação pluvial. Em seguida os conceitos sobre séries temporais, englobando o modelo a ser utilizado, seu procedimento de estimação e critérios de adequação são descritos. Subsequentemente, descreve-se a terminologia de geoestatística, suas notações, seus conceitos, dentre outros itens, para em seguida descrever os métodos de geoestatística.

2.1 PRECIPITAÇÃO PLUVIAL

A precipitação pluvial descreve qualquer tipo de fenômeno relacionado à queda de água da atmosfera, seja na forma líquida (valores expressos em milímetros - *mm*) ou sólida, sobre a atmosfera terrestre (VAREJÃO-SILVA, 2006).

O ciclo hidrológico terrestre é alimentado pelo fenômeno da precipitação, de forma a caracterizar os processos de escoamento superficial direto, infiltração, evaporação, transpiração, recarga de aquíferos, vazão básica de rios e outros. É considerado um processo aleatório, ao qual não permite previsão determinística a longo prazo (CARAM, 2007). Suas principais características são seu total, sua duração e suas distribuições temporal e espacial, onde o total não apresenta um significado que não seja em relação à sua duração (TUCCI, 2001).

No Brasil, as precipitações totais anuais em pontos localizados variam de 300mm no Nordeste árido, até 8.000mm na Região Amazônica (TUCCI, 2001). De acordo com Minuzzi et al. (2007), a região Sudeste do Brasil possui uma característica climática diversificada, devido a sua topografia, sua posição geográfica e, principalmente, os aspectos dinâmicos da atmosfera, que incluem os sistemas meteorológicos de micro, meso e grande escalas, que atuam direta ou indiretamente no regime pluvial, como a ZCAS e as Frentes Frias, principais responsáveis pela precipitação pluvial e o Anticiclone Subtropical do Atlântico

Sul e o Vórtice Ciclônico de Ar superior que, dependendo das suas posições, ocasionam grandes períodos de estiagens.

O tratamento de dados referentes à precipitação é de suma importância, visto que seu ciclo interfere, de forma geral, na qualidade de vida. Carvalho (2007) estabelece que o tratamento dos dados para grande parte dos problemas hidrológicos é realizado estatisticamente.

2.2 SÉRIES TEMPORAIS

Segundo Morettin e Tolói (2006) uma série temporal é um conjunto de observações realizadas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dado é a dependência serial entre as observações, ao longo do tempo observado, que é modelada matematicamente objetivando, principalmente, a obtenção de previsões de valores futuros para a série sob estudo. Sucintamente, a análise de séries temporais é utilizada para descrever, explicar, prever e controlar variáveis temporais de interesse. Esta é composta, basicamente, pela:

1. Análise da estrutura da série por meio de gráficos das funções de autocorrelação e autocorrelação parcial, onde se analisa indícios de tendência, sazonalidade e estacionariedade na série;
2. Proposição de modelos que descrevem o comportamento observado na estrutura da série;
3. Estimativa do modelo proposto;
4. Verificação da adequação dos pressupostos dos modelos por meio dos resíduos;
5. Previsão de observações futuras.

A seguir descrevem-se alguns conceitos importantes referentes à análise de séries temporais. Maiores informações podem ser vistas em Box e Jenkins (1970), Hamilton (1994), Brockwell e Davis (2002), Wei (2006), entre outros.

2.2.1 Notações e Conceitos

Seja $Z(\omega, t)$ uma variável aleatória pertencente ao espaço amostral ω e indexada no tempo t , formada pelo grupo finito de variáveis aleatórias $\{Z_{t_1}, \dots, Z_{t_m}\}$ do processo estocástico $\{Z(\omega, t): t = 0, \pm 1, \pm 2, \dots\}$ com função distribuição m -dimensional dada por:

$$F_{Z_{t_1}, \dots, Z_{t_m}}(z_1, \dots, z_m) = P\{\omega: Z_{t_1} \leq z_1, \dots, Z_{t_m} \leq z_m\}, \quad (1)$$

para $z_i, i = 1, \dots, m$, algum número real (WEI, 2006).

Define-se $E[Z_t]^m = \int \mathbf{z}^m dF(z_t)$ como o momento finito de m -ésima ordem da variável aleatória Z_t , considerando a existência da integral, tal que $dF(z_t)$ seja a derivada da probabilidade alocada para z_t pela distribuição F . Desta, segue que:

- $\mu_t = E[Z_t]$ o momento de primeira ordem da variável Z_t , denominado média ou esperança;
- $\sigma_t^2 = Var[Z_t] = E[Z_t - \mu_t]^2$ é o momento de segunda ordem em relação a esperança da variável Z_t , denominado variância;
- $\gamma(t_1, t_2) = Cov[t_1, t_2] = E\{[Z_1 - \mu_1][Z_2 - \mu_2]\}$ é a covariância entre Z_{t_1} e Z_{t_2} e $\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}}$ é a correlação entre Z_{t_1} e Z_{t_2} .

Uma série temporal pode conter componentes que influenciam sua estrutura de dependência temporal. Estes são denominados *tendência*, *sazonalidade* e *aleatoriedade*.

2.2.2 Tendência

Uma série ou um processo é dito ser tendencioso se, ao longo do tempo, apresentar um comportamento de crescimento (ou decréscimo) com vários possíveis padrões, tais como crescimento linear, crescimento exponencial, entre outros (EHLERS, 2009). Neste caso, o processo é dito ser não-estacionário.

De acordo com Box e Jenkins (1970), um processo estacionário se baseia na suposição de que o mesmo apresenta um equilíbrio estatístico. Um

processo é dito ser *estritamente estacionário* se suas propriedades não são afetadas por uma mudança no tempo, isto é, à medida que o tempo transcorre, tanto a média quanto a variância do processo são invariantes no tempo. Wei (2006) define um processo estacionário de primeira ordem se sua função distribuição unidimensional é invariante no tempo, ou seja,

$$F_{Z_{t_1}}(z_1) = F_{Z_{t_1+k}}(z_1),$$

para algum inteiro t_1 , k e $t_1 + k$. Da mesma forma, a série é dita ser estacionária de segunda ordem se $F_{Z_{t_1}, Z_{t_2}}(z_1, z_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(z_1, z_2)$ para algum inteiro t_1 , t_2 , k , $t_1 + k$ e $t_2 + k$, e, assim sucessivamente até a estacionariedade de ordem m , $m = 1, 2, \dots$. Diz-se então que um processo com função distribuição invariante para qualquer m é denominado *estritamente, fortemente* ou *completamente estacionário*.

Em um processo estritamente estacionário seus dois primeiros momentos são finitos, ou seja, $E(|Z_t|) = \mu < \infty$ e $E(Z_t^2) = \sigma^2 < \infty$ para todo t . Além disso, a covariância e a correlação de Z_t e Z_{t+k} são independentes da diferença entre os tempos (defasagem), k .

Com uma suposição não tão restrita pode-se estabelecer a *estacionariedade fraca de ordem f* , aonde os momentos até alguma ordem f dependem apenas de k . Destas, segue que processos estritamente estacionários com os dois primeiros momentos finitos também podem ser considerados processos estacionários. Em decorrência, por pressuposto assume-se que a distribuição de probabilidade normal é a distribuição “candidata” natural a descrever um processo estocástico, visto que sua distribuição de probabilidade conjunta também segue a normalidade e por seus dois primeiros momentos serem caracterizados como estritamente e fracamente estacionários (WEI, 2006).

Em relação aos momentos do processo sob estudo pode ocorrer que este apresente:

- média e variância invariantes com o tempo → estacionário homogêneo;
- média constante e variância heterocedástica → estacionário não-homogêneo;
- média variante e variância homocedástica → não-estacionário homogêneo;

- média e variância que variam com o tempo → não-estacionário não-homogêneo.

A tendência representa a função média de um processo não-estacionário, em esta pode ser classificada como determinística ou estocástica. No primeiro caso, um modelo de regressão usual pode ser aplicado considerando a série temporal como variável predita e uma variável tempo como variável preditora, de forma a descrever a tendência; no segundo, também referenciada como não-estacionariedade homogênea, os modelos *ARIMA* (em inglês *Autorregressive Integrated Moving Average*), propostos por Box e Jenkins (1970), são os comumente empregados. Basicamente, ser não-estacionário homogêneo implica o processo possui pelo menos uma *raiz unitária*¹. Neste caso, o procedimento utilizado é um filtro, denominado diferenciação (d), que remove a componente de tendência da série (EHLERS, 2009). O método consiste em diferenciar a série até que esta se torne estacionária. Geralmente, uma diferença é o suficiente para estacionarizar uma série, ou seja, $d = 1$. A série diferenciada é denominada integrada de ordem d , ou $I(d)$.

A estacionariedade pode ser verificada por meio das funções de autocorrelação. Mas podem-se empregar testes, tais como o DF e o ADF, propostos por Dickey e Fuller (1979 e 1981, respectivamente) ou o PP, proposto por Phillips e Perron (1988) que detectam a presença de raiz unitária em uma série, na ausência de sazonalidade. No caso da série apresentar indícios de sazonalidade, utiliza-se o teste de DHF, proposto por Dickey et al. (1984), que verifica a presença de raiz unitária sazonal.

Agora, no caso do processo não ser homogêneo, deve-se estabilizar sua variância antes de se realizar qualquer procedimento. Para tal é comum aplicar-se uma transformação aos dados. A transformação comumente empregada é a transformação poder, introduzida por Box e Cox (1964, apud WEI, 2006), também referida como transformação Box-Cox. Este método identifica diversos valores para o parâmetro de transformação. As transformações mais comuns são as transformações logarítmica, inversa e

¹ O termo raiz unitária refere-se a raiz do polinômio do operador de defasagem associado a componente autorregressiva, isto é, pelo menos uma raiz do polinômio autorregressivo é igual a 1.

quadrática. Uma observação importante é que caso seja necessário estabilizar a média e a variância do processo, primeiramente transforma-se a série para depois diferenciá-la.

2.2.3 Periodicidade

Periodicidade corresponde a ideia de repetição que continua indefinidamente, ou seja, define-se periodicidade como uma repetição que ocorre regular ou irregularmente em períodos de tempo. Quando a repetição em uma série temporal ocorre em períodos regulares de tempo diz-se que esta é sazonal. A sazonalidade é uma componente da série histórica difícil de ser estimada, pois é necessário compatibilizar a questão conceitual do fenômeno em estudo com a questão estatística (LATORRE e CARDOSO, 2001). Ambas, periodicidade ou sazonalidade, quando existentes na estrutura de uma série dificultam a obtenção de um modelo matemático que descreva a mesma. Uma forma de inserir esta componente ao modelo matemático é por meio dos modelos de regressão que incorporam funções senoidais à variável tempo. Para tal, a detecção da periodicidade na série de valores, além do ritmo em que esta aparece pode ser realizada através da análise espectral da mesma.

Por meio da análise espectral é possível identificar um padrão periódico, mesmo dentro de uma variabilidade aleatória. O procedimento utiliza um conjunto de funções que contêm seno e cosseno e tenta ajustá-las à variância observada em uma série de observações no tempo, levando em conta a amplitude das "ondas", o período em que elas se repetem e a fase em que se iniciam, isto é, decompõe-se a série de dados em componentes senoidais com coeficientes aleatórios não-correlacionados. Juntamente com essa decomposição – em senóides – existe a correspondente decomposição da função de autocovariância.

A decomposição espectral de um processo estacionário é análoga à representação de Fourier de funções determinísticas, que decompõe a série em um somatório de ondas de senos e cossenos de diferentes amplitudes e comprimentos de onda.

Neste caso, para se retirar o efeito da sazonalidade de uma série, por meio da análise espectral, é necessário estabelecer os períodos em que os picos ocorrem. Isso obtido a partir do *periodograma*. O *periodograma* é uma ferramenta de extrema importância, por meio dele se detectam os períodos significativos de uma série temporal composta por ondas de senos e cossenos, com frequências fixas mascaradas na série ou no resíduo do modelo de série temporal estimado (BOX e JENKINS, 1970).

2.2.4 Modelo ARIMA

Dentre os diversos modelos existentes destacam-se a classe de modelos *ARIMA*, propostos por Box e Jenkins (1970), ou *ARIMA*(p, d, q) em que as letras p e q referem-se, respectivamente, ao número de parâmetros *AR* e *MA* existentes no modelo e d representa quantas diferenciações foram necessárias para estacionarizar a série. Caso a série seja de natureza estacionária, isto é, $d = 0$, o *ARIMA* reduz-se ao modelo *ARMA*(p, q). A forma geral do modelo *ARIMA* é dada pela expressão

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)\varepsilon_t, \quad (2)$$

onde:

- Z_t é a série temporal ou processo estocástico sob estudo, tal que $\{Z_t\}$, $t = 1, \dots, n$ é o tempo e n é o número de observações;
- $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ e $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ representam os parâmetros autorregressivos e médias móveis, respectivamente, que não compartilham parâmetros em comum;
- $(1 - B)^d Z_t$ representa a série diferenciada, isto é, quando $d \geq 1$. Quando $d = 0$ o processo original é estacionário;
- ε_t é um processo ruído branco com média 0 e variância constante.

2.2.5 Identificação, Estimação e Adequação de um Modelo de Séries Temporais

A identificação, a estimação e os diagnósticos de adequação do modelo *ARIMA* são realizados da seguinte forma:

2.2.5.1 Identificação

As FAC e FACP são utilizadas para verificar indícios de tendência, sazonalidade e estacionariedade da série em estudo por meio da estrutura de dependência da mesma. Na FAC o coeficiente de correlação r_k , dado pela expressão

$$r_k = \frac{\sum_{t=1}^{n-k} \tilde{Z}_t \tilde{Z}_{t+k}}{\sum_{t=1}^n (\tilde{Z}_t)^2}, \quad \text{onde } \tilde{Z}_t = Z_t - \bar{Z} \text{ e } \bar{Z} \text{ é a média de } Z_t \quad (3)$$

mede o quanto os pares ordenados de observações (Z_{t+k}, Z_t) estão estatisticamente correlacionados.

A função de autocorrelação representa graficamente r_k , variando entre -1 e +1, indicando em que medida cada valor em um dado instante de tempo t se relaciona com os valores em $t + 1, t + 2, \dots, t + k$.

Já a FACP, que também um gráfico, representa a relação entre conjuntos de pares ordenados (Z_{t+k}, Z_t) levando-se em consideração os efeitos causados pela intervenção das observações intermediárias.

Ressalta-se que para propor ordens para os parâmetros do modelo, há a necessidade de que a série sob estudo seja estacionária, ou seja, sem a presença de heterogeneidade, periodicidade ou tendência.

No caso de variância explosiva² propõe-se uma transformação Box-Cox adequada, tal que exista um λ que maximize a função $T(Z_t)$ descrita a seguir.

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda}. \quad (4)$$

Para a periodicidade, o período de ocorrência for regular, insere-se na Equação (2) as componentes sazonais ocasionando a mudança do modelo *ARIMA*(p, d, q) para o *SARIMA*(p, d, q) \times (P, D, Q)_S, em que P e Q referem-se,

² Variância explosiva equivale a heterocedasticidade.

respectivamente, aos parâmetros autorregressivos e médias móveis sazonais, D a diferenciação sazonal e s o período de repetição. Em se tratando de periodicidade, trata-se a série temporal de interesse por meio por meio da análise espectral.

Primeiramente, define-se Z_t , uma série temporal de n observações, dada pela seguinte representação de Fourier (WEI, 2006):

$$Z_t = \sum_{k=0}^{\lfloor n/2 \rfloor} (a_k \cos \omega_k t + b_k \sin \omega_k t), \quad (5)$$

em que $\omega_k = 2\pi k/n$, $k = 0, 1, \dots, \lfloor n/2 \rfloor$ são frequências de Fourier e

$$a_k = \begin{cases} \frac{1}{n} \sum_{t=1}^n Z_t \cos \omega_k t, & k = 0 \text{ e } k = \frac{n}{2}, \text{ sen é par,} \\ \frac{2}{n} \sum_{t=1}^n Z_t \cos \omega_k t, & k = 1, 2, \dots, \lfloor \frac{n-1}{2} \rfloor, \end{cases}$$

e

$$b_k = \frac{2}{n} \sum_{t=1}^n Z_t \sin \omega_k t, \quad k = 1, 2, \dots, \lfloor \frac{n-1}{2} \rfloor,$$

são os coeficientes de Fourier, equivalentes as estimativas de mínimos quadrados dos coeficientes do seguinte modelo de regressão ajustado:

$$Z_t = \sum_{k=0}^{\lfloor n/2 \rfloor} (a_k \cos \omega_k t + b_k \sin \omega_k t) + \varepsilon_t, \quad (6)$$

em que ε_t é o erro aleatório.

A quantidade $I(\omega_k)$, dada por

$$I(\omega_k) = \begin{cases} na_0^2, & k = 0; \\ \frac{n}{2} (a_k^2 + b_k^2), & k = 1, \dots, \lfloor (n-1)/2 \rfloor; \\ na_{n/2}^2, & k = \frac{n}{2} \text{ quando } n \text{ é par,} \end{cases} \quad (7)$$

é chamada *periodograma*. Outra forma de se apresentar o periodograma é

$$I(\omega_k) = \frac{1}{2\pi} [d_{\cos}^2(\omega_k) + d_{\sin}^2(\omega_k)], \quad (8)$$

em que d_{\cos} e d_{\sin} são a transformada cosseno e seno de Fourier, respectivamente.

Esta quantidade representa a contribuição da frequência ω_k à soma de quadrados associados aos coeficientes seno e cosseno. Quando $I(\omega_k)$ são plotados contra ω_k em um gráfico formam-se periodogramas, isto é, estimadores do espectro.

Após a obtenção dos períodos, avalia-se se o período encontrado é significativo. A estatística de teste, proposta por Fisher (1929 apud WEI, 2006) baseia-se na comparação entre os valores de pico e de variabilidade total da série, a fim de se detectar sua significância. Basicamente, as hipóteses a serem testadas são

$$H_0: a_k = b_k = 0 \quad \text{versus} \quad H_1: a_k \neq 0 \quad \text{ou} \quad b_k \neq 0,$$

em que a hipótese nula, H_0 , é “não existe periodicidade”. A estatística de teste

$$T = \frac{\max[I(\omega_k)]}{\sum_{k=1}^{[n/2]} I(\omega_k)} \quad (9)$$

utiliza a ordenada máxima do periodograma, $\max[I(\omega_k)]$. Fisher (1929 apud WEI, 2006) descreve que, para algum $g \geq 0$ tem-se

$$P(T > g) \simeq N(1 - g)^{N-1} = \alpha,$$

onde $N = (n - 1)/2$ se n é ímpar e $N = (n/2) - 1$ se n é par. Se o valor T calculado de uma série for maior do que g , então rejeita-se a hipótese nula e conclui-se que Z_t contém uma componente periódica. O procedimento pode ser repetido para as demais ordenadas do periodograma, visto que a série pode conter mais de uma componente periódica. Após eliminar a heterogeneidade e a periodicidade, a próxima etapa é avaliar a tendência.

Como descrito anteriormente, empregam-se testes de raiz unitária para se avaliar a estacionariedade da série. O teste ADF assume o seguinte aspecto:

$$\Delta Z_t = \alpha + \beta t + \gamma Z_{t-1} + \sum_{j=1}^{p-1} \rho_{j+1} \Delta Z_{t-j} + \varepsilon_t, \quad (10)$$

em que Δ é o operador diferença $\gamma = \rho - 1$. Basicamente, o teste ADF estima uma autorregressão que incorpora defasagens em relação a variável que está sendo analisada. A hipótese nula do teste é de que existe pelo menos uma raiz unitária, ou seja, a variável não é estacionária e $\gamma = 0$.

Sendo constatada a não-estacionariedade da série propõe-se uma ordem para a diferenciação com base no decaimento das defasagens das funções de autocorrelação da série estacionária. A seguir é apresentada a primeira diferença da série Z_t ,

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (11)$$

2.2.5.2 Estimação

Com a série estacionária propõem-se o modelo provisório a ser ajustado, considerando as ordens dos parâmetros autorregressivos e médias móveis do modelo, de forma a se realizar suas respectivas estimações. Dentre os métodos de estimação existentes têm-se o método dos momentos, o método de mínimos quadrados e o método de máxima verossimilhança. Sendo o último o mais comum. Detalhes sobre as rotinas de estimação não serão abordados aqui, podendo ser vistas em Box e Jenkins (1970).

Paralelamente a estimação realizam-se testes de sobreajustamento. Estes consistem na inclusão de parâmetros extras ao modelo proposto e verificar sua significância estatística, sempre considerando a parcimônia³. Segundo Box e Draper (1987) “todos os modelos estão errados, mas alguns são úteis”. Isto é, os modelos estão errados, pois não são a realidade apresentada pelos dados, mas sim uma simplificação. Contudo, os modelos que simplificam a realidade são úteis para se interagir com ela. Para tal, como métodos de escolha entre diferentes modelos, costumam-se empregar os seguintes critérios:

- **AIC.** A ordem k do modelo deve minimizar o seguinte critério:

$$AIC(k) = n \ln \hat{\sigma}_k^2 + 2k, \quad (12)$$

em que $\hat{\sigma}_{k,1}^2$ é o estimador de máxima verossimilhança de σ_ε^2 ;

- **BIC.** A ordem k do modelo deve minimizar o seguinte critério:

$$BIC(k) = n \ln \hat{\sigma}_k^2 + \frac{k}{n} \ln n. \quad (13)$$

O modelo que obtiver o menor valor para qualquer um dos critérios ou para ambos é o modelo que apresenta variância estimada mínima. Ambos os critérios levam em consideração a quantidade de parâmetros estimados para o modelo em suas estatísticas. A utilização de um critério permite a comparação entre a estimação de diversos modelos, de forma a verificar qual dentre eles pode representar a série estudada.

³ O princípio da parcimônia nos diz que onde os melhores modelos são obtidos utilizando-se estruturas aceitáveis e simples, contendo em sua formulação um menor número de parâmetros.

2.2.5.3 Diagnósticos de Adequação

Após identificar a ordem e estimar eficientemente os parâmetros de um modelo é necessário verificar sua adequação. Esta checagem é realizada através dos resíduos ($\hat{\epsilon}_t$). Os resíduos, que são estimadores do erro aleatório, são obtidos pela diferença entre os valores reais da série Z_t e os valores estimados pelo modelo escolhido \hat{Z}_t , $\hat{\epsilon}_t = Z_t - \hat{Z}_t$.

Se o modelo escolhido for o adequado, espera-se que seus resíduos apresentem comportamento aleatório em torno da média zero, com variância aproximadamente constante, e serem não-correlacionados. É de suma importância que não haja violação dos pressupostos do modelo, normalidade, homogeneidade e independência, pois isso indica que o referido modelo não seja adequado.

Para se verificar se o modelo é adequado podem-se realizar testes sobre os resíduos. A suposição da normalidade pode ser avaliada por meio da aplicação do teste de Shapiro-Wilk (HOLLANDER e WOLFE, 1999) aos resíduos. Neste, a hipótese nula é de que a amostra sob estudo advém de uma população normalmente distribuída. A homogeneidade é avaliada inicialmente na série original, pois se a mesma não for homogênea necessitará de uma transformação. Aplicando uma transformação aos dados originais, implica-se que os resíduos obtidos pelo modelo estimado serão, conseqüentemente, homocedásticos se a transformação proposta for a ideal.

Já análise da independência é realizada pelo teste de Ljung-Box, conforme sugestão de Box e Jenkins (1970), em que se testa se um grupo de autocorrelações é significativamente diferente de zero.

2.3 GEOESTATÍSTICA

Modelos matemáticos têm por objetivo fornecer realizações de uma variável de interesse, ou seja, predizer. Na Geoestatística, as predições espaciais clássicas são realizadas por meio de métodos de interpolação. A interpolação é uma técnica utilizada para a estimativa do valor de um atributo em locais não amostrados, a partir de pontos amostrados na mesma área ou

região. A interpolação espacial converte dados de observações pontuais em campos contínuos, produzindo padrões espaciais que podem ser comparados com outras entidades espaciais contínuas. O raciocínio que está na base da interpolação é que, em média, os valores do atributo tendem a ser similares em locais mais próximos do que em locais mais afastados. Esse conceito também fundamenta a base das relações espaciais entre fenômenos geográficos, utilizando a correlação espacial como meio de diferença dos atributos estimados (CÂMARA e MEDEIROS, 1998). Basicamente, a Geoestatística pode ser caracterizada pelos seguintes pressupostos:

- Estacionariedade espacial do processo estocástico;
- Modelo estrutural da dependência espacial;
- Distribuição espacial da variável, com obtenção dos mapas que representam os valores da variável sob estudo.

Em termos práticos, um modelo de predição espacial pode ser utilizado como ferramenta de predição e/ou estimação, como no caso de séries pluviométricas de diferentes locais e períodos de ocorrência (ARAÚJO et al., 2001, e MELLO et al., 2008a). Contudo, existem diferentes modelos de dependência espacial. A diferença entre eles são os diferentes métodos de interpolação dos pontos espacialmente localizados.

A lógica da interpolação, fundamentada pela Primeira Lei da Geografia de Tobler (1979 apud CÂMARA e MEDEIROS, 1998), está no pressuposto de que, em média, os valores do atributo sob estudo tendem a ser similares em locais mais próximos do que em locais mais distantes. Os métodos de interpolação comumente empregados podem ser divididos em duas categorias: determinísticos e probabilísticos. No método determinístico destaca-se o inverso da distância e entre o probabilístico a krigagem. A krigagem leva em consideração a continuidade existente entre os valores observados nas locações próximas a locação de referência, o que faz com que o mesmo seja considerado um BLUP, ou seja, os pesos obtidos por seu estimador são considerados não-viesados e possuem erros de predição mínimos (LE e ZIDEK, 2006). Contudo, parte do seu processo de interpolação engloba a escolha do modelo teórico de semivariograma.

O semivariograma é um gráfico que analisa o grau de dependência espacial entre amostras de um campo experimental (HENGL, 2009). Le e Zidek

(2006) descrevem que a estrutura de covariância ou semicovariância de um campo aleatório⁴ espacial indica a força das relações entre variáveis que representam seus níveis em diferentes locais de domínio.

Como forma de se entender os procedimentos aqui propostos, segue-se uma apresentação das notações e dos conceitos básicos de estatística espacial para a caracterização das estruturas de covariância de processos espaciais, incluindo a definição dos momentos, da estacionariedade espacial e do variograma, foco principal do método proposto.

2.3.1 Notações e Conceitos

Conforme Hengl (2009), seja $Z(s_j)$ a variável aleatória resposta de interesse, formada pelo grupo de observações $z(s_1), z(s_2), \dots, z(s_n)$ tal que para qualquer $\{z(s_j), i = 1, \dots, n\}$, $s_j = (x_j, y_j)$ seja a coordenada geográfica x_j e y_j de cada uma das n estações de monitoramento pluviométrico observadas. Define-se \mathbb{A} , o domínio geográfico de interesse, como a área que contém as n estações de monitoramento nas quais foram realizadas medidas ao longo do tempo $t, t = 1, \dots, T$. Assume-se que os valores $z(s_j)$ sejam representativos, não-preferenciais e consistentes, de forma que o valor da variável resposta em alguma nova locação s_0 possa ser predito utilizando um modelo de predição espacial.

Diz-se que $Z(s_j)$ tem função de distribuição acumulada dimensional finita (f.d.a.) se para qualquer grupo de locações no domínio e para qualquer inteiro positivo n ,

$$F_{s_1, \dots, s_n}(x_1, \dots, x_n) \equiv P\{Z(s_1) \leq x_1, \dots, Z(s_n) \leq x_n\}, \quad (14)$$

em que P denota a probabilidade (LE e ZIDEK, 2006). Define-se que, para o momento finito de k -ésima ordem da variável aleatória $Z(s_j)$ em qualquer locação s_j , isto é, $E[Z(s_j)]^k \equiv \int x^k dF_s(x_j)$:

⁴Campo aleatório é uma generalização de um processo estocástico, tal que o parâmetro base não precisa mais ser um simples número real ou inteiro representando o tempo. Mas pode assumir valores que são vetores multidimensionais, ou alguns pontos múltiplos (KHOSHNEVISAN, 2002).

- $\mu(\mathbf{s}_j) \equiv E[Z(\mathbf{s}_j)]$ é o momento de primeira ordem da variável $Z(\mathbf{s}_j)$ em qualquer locação \mathbf{s}_j ;
- $Var[Z(\mathbf{s}_j)] \equiv E[Z(\mathbf{s}_j) - \mu(\mathbf{s}_j)]^2$ é o momento de segunda ordem sobre a esperança da variável Z em qualquer locação \mathbf{s}_j ;
- $Cov[\mathbf{s}_j, \mathbf{s}_l] \equiv E\{[Z(\mathbf{s}_j) - \mu(\mathbf{s}_j)][Z(\mathbf{s}_l) - \mu(\mathbf{s}_l)]\}$ é a covariância entre duas locações \mathbf{s}_j e \mathbf{s}_l , $j \neq l$. Se $j = l$, tem-se que $Cov[\mathbf{s}_j, \mathbf{s}_j] \equiv Var[Z(\mathbf{s}_j)]$;
- O variograma entre duas locações, \mathbf{s}_j e \mathbf{s}_l ($j \neq l$), no domínio sob estudo é definido como a metade do valor médio do quadrado da diferença entre pares de valores, tal como apresentado na Equação (15),

$$\gamma(\mathbf{s}_j, \mathbf{s}_l) = \frac{1}{2} E \left[\left(Z(\mathbf{s}_j) - Z(\mathbf{s}_l) \right)^2 \right]. \quad (15)$$

Da mesma forma que em processos temporais, a estacionariedade espacial é observada por meio dos momentos. Neste caso, a variável regionalizada se comporta de forma similar ou contínua na região de domínio, isto é, existe estacionariedade se os momentos estatísticos da variável aleatória $Z(\mathbf{s}_j)$ forem os mesmos em qualquer locação. No entanto, Le e Zidek (2006) afirmam que esta propriedade é muito forte, tal que alguns processos não irão apresentá-la totalmente. Assim, versões de estacionariedade foram definidas para permitir uma caracterização mais fina da natureza estocástica das locações (VIEIRA, 2000).

2.3.2 Estacionariedade

Para qualquer vetor de distância \mathbf{h} , o processo $Z(\mathbf{s}_j)$ é dito ser estritamente estacionário se a distribuição de probabilidade finito-dimensional de $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ e $\{Z(\mathbf{s}_1 + \mathbf{h}), \dots, Z(\mathbf{s}_n + \mathbf{h})\}$ são idênticas para qualquer n , ou seja, as distribuições são invariantes sob qualquer translação determinada pelo vetor \mathbf{h} .

A variável aleatória $Z(\mathbf{s}_j)$ é dita ser estacionária de segunda ordem se as seguintes condições forem satisfeitas (VIEIRA, 2000):

- a) A esperança $E[Z(\mathbf{s}_j)]$ existe e não depende de \mathbf{s}_j ;

b) Para qualquer par de variáveis aleatórias, $Z(s_j)$ e $Z(s_j + \mathbf{h})$, a covariância ou covariograma, $Cov(\mathbf{h})$, existe e for função de \mathbf{h} :

$$Cov(\mathbf{h}) = E[Z(s_j), Z(s_j + \mathbf{h})] - E[Z(s_j)]^2, \quad (16)$$

para qualquer j dentro do domínio.

Para um vetor de distâncias \mathbf{h} pode-se estabelecer o variograma pela expressão:

$$\gamma(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} [Z(s_j) - Z(s_j + \mathbf{h})]^2. \quad (17)$$

em que $n(\mathbf{h})$ é o número de pares de valores medidos $Z(s_j)$, $Z(s_j + \mathbf{h})$, separados por um vetor \mathbf{h} (VIEIRA, 2000). Os valores obtidos por $\gamma(\mathbf{h})$ são cruzados com seus respectivos valores de \mathbf{h} , originando uma função do vetor \mathbf{h} que depende da magnitude e da direção de \mathbf{h} , denominado semivariância,

$$2\gamma(\mathbf{h}) = E[Z(s_j) - Z(s_j + \mathbf{h})]^2. \quad (18)$$

2.3.3 Modelos de Semivariograma

O semivariograma é um gráfico das semivariâncias em função da distância a um ponto. A idéia por trás da semivariância e do semivariograma é de que a partir de certa distância a mesma não aumentará mais com a distância e se estabilizará num valor igual à variância média. Este é denominado silo ou patamar ($C_0 + C$). Já a distância entre o valor inicial do semivariograma e o início do patamar é denominado amplitude ou alcance (a). Agora, ao se extrapolar a curva do semivariograma para a distância zero, chega-se a um valor não nulo de semivariância, que recebe o nome de efeito pepita (C_0). Na Figura 1 apresenta-se um semivariograma típico.

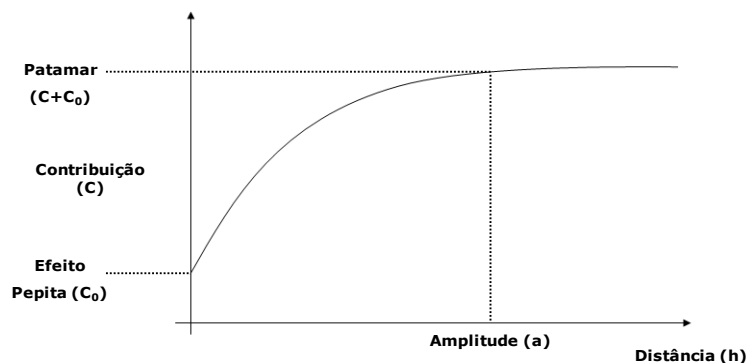


Figura 1 - Exemplo de semivariograma típico.

Em geral, as semivariâncias tendem a aumentar o seu valor à medida que o valor de h aumenta e estabilizar para certo valor h , visto que medidas mais próximas umas das outras tendem a apresentar um comportamento mais semelhante do que as medidas que estão mais afastadas. Assim, como a diferença entre as medidas tende a ser superior, o valor da semivariância tende a aumentar. Analogamente, à medida que a distância entre as medidas diminui, o valor da semivariância tende a diminuir, visto que a diferença entre medidas tende a ser inferior.

De acordo com Le e Zidek (2006) o variograma, $\gamma(\mathbf{h})$, é uma função do vetor \mathbf{h} especificada pelo comprimento e pela direção. Caso o semivariograma não apresente um comportamento único para as locações, em todas as direções, este fenômeno é denominado *anisotropia*. Contrariamente, caso a associação entre as locações seja a mesma em todas as direções, dependendo somente do comprimento de \mathbf{h} , denomina-se este fenômeno como *isotrópico*. É comum assumir em estudos geoestatísticos que a associação entre as locações sejam isotrópicas. Uma observação interessante é que pelo gráfico gerado para todas as direções por meio do se é possível estabelecer a isotropia ou não. Mais detalhes sobre anisotropia e isotropia podem ser encontrados detalhadamente em Isaaks e Srivastava (1989), Cressie (1993), Goovaerts (1997), Webster e Oliver (2001), Hengl (2009), entre outros.

Os modelos de semivariograma isotrópicos habitualmente utilizados em estudos geoestatísticos são:

1. **Efeito pepita puro.** Constante em todas as distâncias, ou seja, a correlação espacial é constante para qualquer distância no campo aleatório. O modelo é dado por

$$\gamma(\mathbf{h}) \equiv C_0 \geq 0, \text{ para todo } \mathbf{h} > 0.$$

2. **Linear.** O semivariograma aumenta linearmente à medida que o efeito pepita aumenta com a distância h . Seja $b = \frac{C}{a}$ o coeficiente angular, o modelo linear é dado por

$$\gamma(\mathbf{h}) \equiv C_0 + bh, \text{ para } \mathbf{h} > 0, C_0 \geq 0 \text{ e } b \geq 0.$$

3. **Exponencial.** Incrementa exponencialmente quando a distância h aumenta, sendo o patamar $C_0 + b$. Além disso, a covariância decresce a

medida que h aumenta. Para $h > 0$, $C_0 \geq 0$, $b \geq 0$ e $a \geq 0$ o modelo é dado por:

$$\gamma(h) = C_0 + b \left[1 - \exp\left(-\frac{h}{a}\right) \right], \text{ para } 0 < h \leq a.$$

- 4. Gaussiano.** O semivariograma incrementa exponencialmente quando h aumenta. Contudo, a amplitude máxima é de $\sqrt{3}a$ com patamar de $C_0 + b$. Para $h > 0$, $C_0 \geq 0$, $b \geq 0$ e $a \geq 0$ o modelo é dado por:

$$\gamma(h) = C_0 + b \left[1 - \exp\left(-\frac{h^2}{a}\right) \right], \text{ para } 0 < h \leq a.$$

- 5. Esférico.** O semivariograma aumenta continuamente a partir do efeito pepita C_0 para o patamar de $C_0 + b$, quando de $h \geq a$. Para $h > 0$, $C_0 \geq 0$, $b \geq 0$ e $a \geq 0$ o modelo é dado por:

$$\gamma(h) = \begin{cases} C_0 + b \left[\frac{3h}{2a} + \frac{h^3}{2a^3} \right], & 0 < h \leq a. \\ C_0 + b, & h > a \end{cases}$$

Previamente a determinar o modelo de semivariograma, constrói-se um semivariograma experimental através da Equação (16). Para tal, estima-se a semivariância de cada ponto em relação aos demais e inserem-se os valores obtidos no gráfico da semivariância pela distância. Por meio deste gráfico estima-se o modelo de variograma que melhor se aproxima da curva obtida.

2.3.3.1 Estimação do Modelo de Semivariograma

Após a construção do variograma experimental, é necessário estimar o modelo de semivariograma. Os métodos de ajuste mais empregados baseiam-se no semivariograma experimental ou no ajuste de um modelo direto aos dados, aonde os métodos OLS e o ML fazem parte, respectivamente.

2.3.3.1.1 Método OLS

O método de mínimos quadrados ordinários (sigla em inglês OLS), de acordo com Mello et al. (2005), consiste em obter os valores dos parâmetros de um modelo que minimizem a soma do quadrado da diferença entre os valores

observados e os estimados. No OLS, as estimativas dos parâmetros do semi-variograma são obtidas pela minimização da seguinte expressão:

$$Q(\theta) = \sum_{i=1}^k [g(h_i) - \gamma(h_i; \theta)]^2, \quad (19)$$

em que θ refere-se ao vetor dos parâmetros estimados, $g(h_i)$, cada semivariograma experimental, $\gamma(h_i; \theta)$ o modelo de semivariograma ajustado e k o número de “lags” do semivariograma experimental. A estimação dos parâmetros por OLS segue o procedimento abaixo, minimizando a Equação (19):

1. Obter as derivadas parciais de $Q(\theta)$ em relação ao modelo de semivariograma estimado;
2. Igualar a zero as equações de derivadas parciais;
3. Resolver o sistema de equações resultantes.

2.3.3.1.2 Método ML

O método de máxima verossimilhança (sigla em inglês ML) é uma técnica de estimação utilizada com frequência, por apresentar propriedades assintóticas interessantes como eficiência e consistência. Proposto por Hartley e Rao (1967), o método é iterativo e consiste em maximizar a função de densidade de probabilidades, em relação aos efeitos fixos e aos componentes de variâncias dos efeitos aleatórios do modelo, de forma a fornecer estimativas “mais verossímeis” dos parâmetros.

Seja (Z_1, \dots, Z_n) um conjunto de observações geradas a partir de um modelo linear gaussiano, definido pela soma de três componentes: determinística, estocástica e aleatória (DIGGLE e RIBEIRO JR., 2000). Isto é,

$$Z_i = \mu(s_i) + S(s_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (20)$$

em que $\mu(s_i)$ é a componente completamente determinística, determinada por um modelo linear da forma:

$$\mu(s_i) = \sum_{k=1}^p f_k(s_i)\beta_k, \quad (21)$$

tal que (f_1, \dots, f_p) seja um conjunto de p funções que descrevem a variável Z , deterministicamente, a partir das coordenadas espaciais (s_i, \dots, s_n) . $S(s_i)$ é a componente aleatória, entendida como um processo estocástico gaussiano

autocorrelacionado com média zero e variância σ^2 constante e, ϵ_i é a parte totalmente aleatória do modelo, sendo $\epsilon_1, \dots, \epsilon_n$ independentes, identicamente distribuídos, com média zero e variância τ^2 , denominada de *nugget* ou efeito pepita. Ressalta-se que $S(s_i)$ é a componente na Equação (20) que se refere à continuidade espacial aleatória da variável Z . A estimação dos parâmetros por ML segue o procedimento abaixo, maximizando a Equação (20):

1. Obter a função de verossimilhança $L(\beta; \theta)$, em que $\theta = \{C, C_0, a\}$;
2. Obter o logaritmo natural de $L(\beta; \theta)$, denominado $l(\beta; \theta)$;
3. Calcular as derivadas parciais de $l(\beta; \theta)$ em relação aos parâmetros do modelo de semivariograma estimado;
4. Igualar a zero as equações de derivadas parciais;
5. Resolver o sistema de equações resultantes.

2.3.4 Seleção de Modelos

Existem diversos métodos estatísticos que são empregados na escolha entre modelos lineares. Dentre as formas comumente utilizadas estão os critérios de informação. Estes consideram em seu procedimento a quantidade de parâmetros de cada modelo estimado, ou seja, são parcimoniosos. O AIC, descrito na Equação (12) é o critério mais utilizado, em que o modelo estimado que apresentar o menor valor para o critério é o modelo selecionado.

Outra forma de se selecionar um modelo é por meio da autovalidação ou validação cruzada. A autovalidação consiste em estimar os pontos observados, com base no modelo de semivariograma ajustado, possibilitando comparar o valor estimado ao real (MELLO et al., 2008a) e, conseqüentemente, o erro produzido pelo modelo. Pela autovalidação, conforme Ribeiro Junior (1995), além de se avaliar o modelo escolhido para descrever o semivariograma, também se avaliam todas as etapas do processo de modelagem, como estacionariedade, os estimadores e o modelo de semivariograma escolhido.

A autovalidação segue a seguinte rotina:

1. Retirar do conjunto de dados um elemento qualquer.
2. Calcular o valor estimado do ponto retirado empregando os dados restantes.

3. Calcular o erro de estimação, dado por:

$$\hat{\epsilon} = Z(s_i) - \hat{Z}(s_i) \quad (22)$$

em que $Z(s_i)$ é o valor observado e $\hat{Z}(s_i)$ é o respectivo valor .

4. Repetir os passos anteriores para todos os outros valores.

5. Análise dos erros de estimação, onde a mesma requer que as condições de tendência e de variância mínima sejam satisfeitas.

Caso haja indícios da existência de dependência espacial no resíduos, todo o procedimento desde a obtenção do semivariograma deverá ser novamente realizado.

3 MATERIAL E MÉTODOS

3.1 REGIÃO DE ESTUDO

A região de estudo compreendeu alguns dos municípios do Estado de Minas Gerais que possuem estações de monitoramento pluviométrico. A escolha desses municípios baseou-se em estudo apresentado por Aspiazu et al. (1990, apud CARAM, 2007), que separou por meio da análise de componentes principais Minas Gerais em 10 regiões climaticamente homogêneas, conforme observado na Figura 2. Dessas, 10 regiões foram utilizadas 5 (I.1, I.2, III.1.1, III.1.2 e III.2).

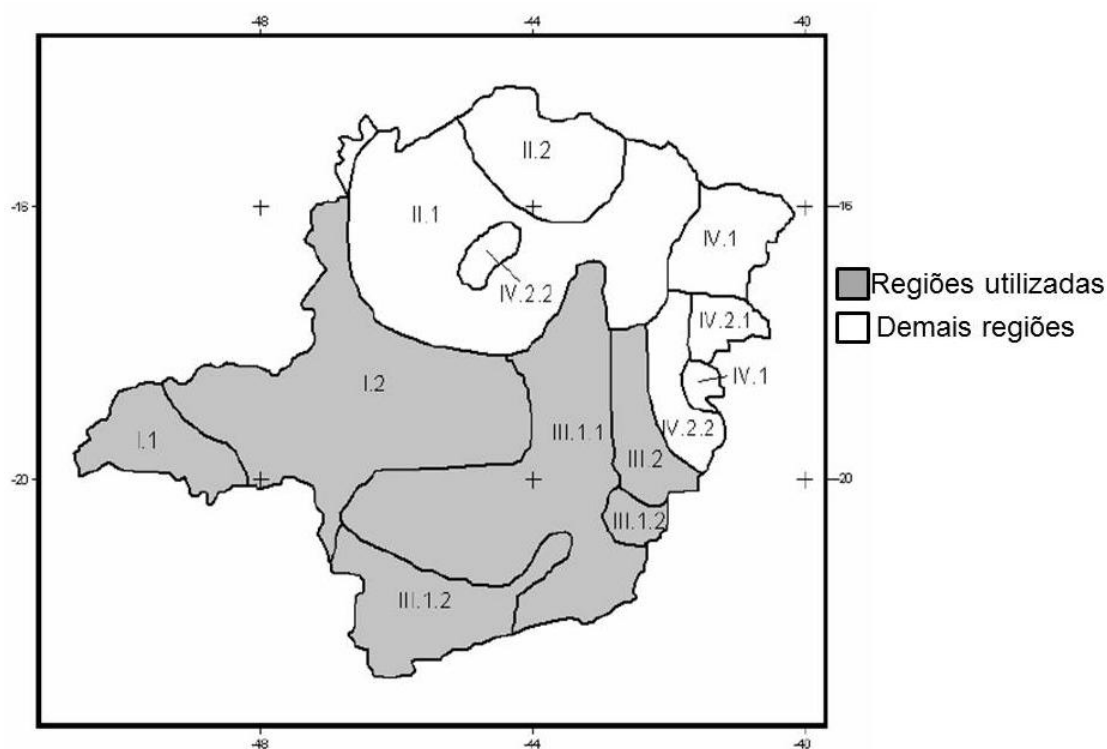


Figura 2 - Regiões climaticamente homogêneas do Estado de Minas Gerais
Fonte: Aspiazu et. al. (1990, apud CARAM, 2007)

3.2 BASE DE DADOS

A base de dados foi constituída por séries temporais referentes à precipitação média mensal por dias chuvosos de 43 estações pluviométricas do Estado de Minas Gerais, para o período de janeiro de 2000 a novembro de 2005 (ANEXO B). O referido período de estudo foi escolhido devido a não

existência de observações faltantes nesse intervalo de tempo para as 43 estações. De acordo com Sedyama (2005, apud CARAM, 2007), não é possível decidir se uma série é “absolutamente homogênea”, dispondo de observações de um único local e utilizando séries com dados faltantes. As informações referentes às estações pluviométricas foram obtidas na base de dados da ANA, disponibilizadas *online* em <http://hidroweb.ana.gov.br>. A Figura 3 apresenta a dispersão das estações pluviométricas nos municípios de Minas Gerais utilizados no estudo, conforme Quadro 1. O Anexo A destaca-se pela apresentação das coordenadas geográficas das estações pluviométricas apresentadas no Quadro 1.

Quadro 1 - Municípios selecionados com base nas regiões climaticamente homogêneas do Estado de Minas Gerais de acordo com Aspiazu et. al. (1990, apud CARAM, 2007)

Região	Município
I.1	Campina Verde
I.2	Patos de Minas, Paraopeba, Monte Carmelo, Onça de Pitangui, Estrela do Sul, Ibiá, Abadia dos Dourados, Patrocínio e Monte Alegre de Minas
III 1.1	Bocaiuva, Divinópolis, Guanhães, Entre Rios de Minas, Diamantina, Passa Tempo, Rio Piracicaba, Bambuí, Conceição do Mato Dentro, Iguatama, Jaboticatubas, Lamounier, Sabará, Astolfo Dutra, Caeté, Patrocínio do Muriaé, Taquaraçu de Minas, Rio Pomba, Morro do Pilar, Juiz de Fora, São Gonçalo do Rio Abaixo, Rio Preto, Acaiaca, Santa Rita de Jacutinga, Piranga, Carmo do Cajuru e Porto Firme
III 1.2	Caiana e Rio Casca
III 2	Coroaci, Bom Jesus do Galho, São Pedro do Suaçuí e Raul Soares.

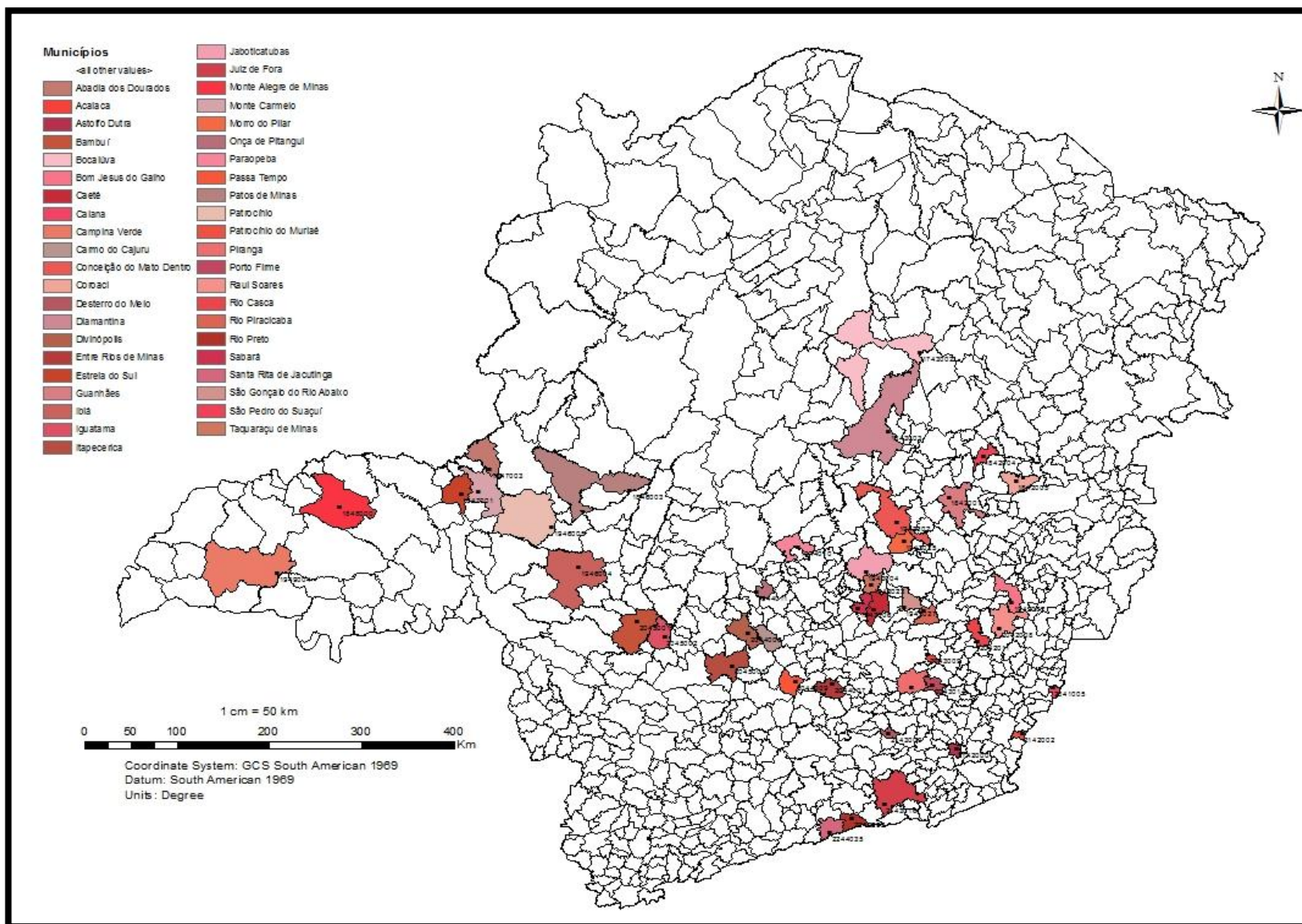


Figura 3 - Mapa das posições das 43 estações pluviométricas utilizadas neste estudo.

3.2 LINGUAGEM COMPUTACIONAL

Todas as análises foram realizadas na linguagem computacional livre *R* (*R Development Core Team*, 2012), considerando um nível de significância de 5%. Os pacotes utilizados foram: *tseries*, *TSA*, *geoR* e *scatterplot3d*. Os Anexos C e D apresentam um exemplo das rotinas realizadas no *R*.

3.3 DETECÇÃO DE *OUTLIERS*

A avaliação da homogeneidade por meio da detecção de *outliers* deu-se em 3 etapas:

1. Análise de Séries Temporais. Ajuste de um modelo de séries temporais para cada uma das 43 séries de precipitação média mensal por dias chuvosos e, conseqüentemente, obtenção dos seus respectivos resíduos;
2. Análise Geoestatística. Ajuste de um modelo de semivariograma da precipitação média mensal por dias chuvosos de cada estação pluviométrica em mês de cada ano sob estudo e, conseqüentemente, seus resíduos obtidos por meio da autovalidação;
3. Detecção de *Outlier*. Comparação entre os resíduos da análise de séries temporais e os resíduos da análise geoestatística

3.3.1 Análise de Séries Temporais

Esta etapa consiste em estimar um modelo *ARIMA* para cada uma das 43 séries temporais de precipitação média por dias chuvosos, conforme a expressão dada pela Equação (2). A Figura 4 ilustra o esquema do processo da análise de séries temporais, da interpretação dos dados até a adequação do modelo.

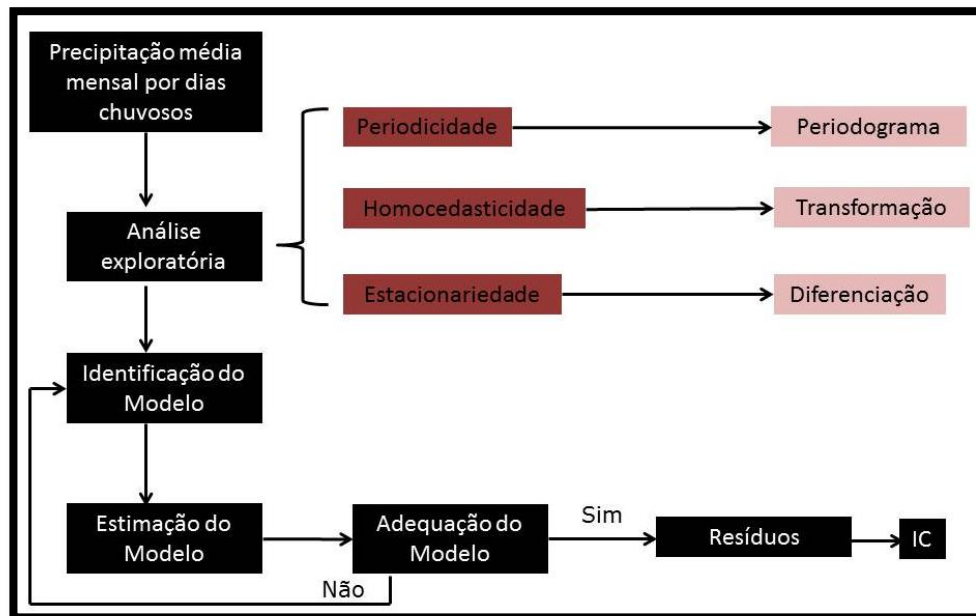


Figura 4 - Processo de análise de séries temporais.

Primeiramente, realizou-se uma análise explanatória das séries de precipitação média por dias chuvosos para avaliar as condições de estacionariedade, homocedasticidade e periodicidade das mesmas. Em que:

- A periodicidade foi observada por meio do periodograma, onde períodos significativos foram inseridos em um modelo de regressão linear em que seus coeficientes eram da forma de pares de senos e cossenos. Logo, os resíduos obtidos dessa regressão linear não apresentam nenhum efeito periódico;
- Em seguida, gráficos dos resíduos da regressão observados no tempo foram utilizados para avaliar a homocedasticidade;
- A estacionariedade foi verificada por meio das funções FAC e FACP e pelo teste ADF. Havendo presença de raiz unitária realizou a diferenciação.

Assim, eliminando componentes de tendência, periodicidade e homocedasticidade pode-se identificar o modelo, isto é, propor a ordem para os parâmetros autorregressivos e médias móveis por meio das funções FAC e FACP. A estimação do modelo proposto deu-se pelo método de máxima verossimilhança. Para verificar se o modelo proposto é realmente o indicado, estimaram-se outros modelos alterando as ordens dos parâmetros e empregou-se o AIC como critério de escolha entre os mesmos. Escolhido o modelo, realizou-se sua adequação por meio dos resíduos do modelo ajustado.

A normalidade foi verificada pelo teste de Shapiro-Wilk e a independência pelo teste de Ljung-Box.

Portanto, considerando que os resíduos dos modelos de séries temporais ajustados não violam os pressupostos de adequacidade, estes podem ser utilizados no restante deste trabalho.

3.3.2 Análise Geoestatística

Os procedimentos geoestatísticos foram realizados supondo isotropia na área de estudo, por simplicidade de justificação dos resultados obtidos. O esquema a seguir ilustra todo procedimento desta etapa.

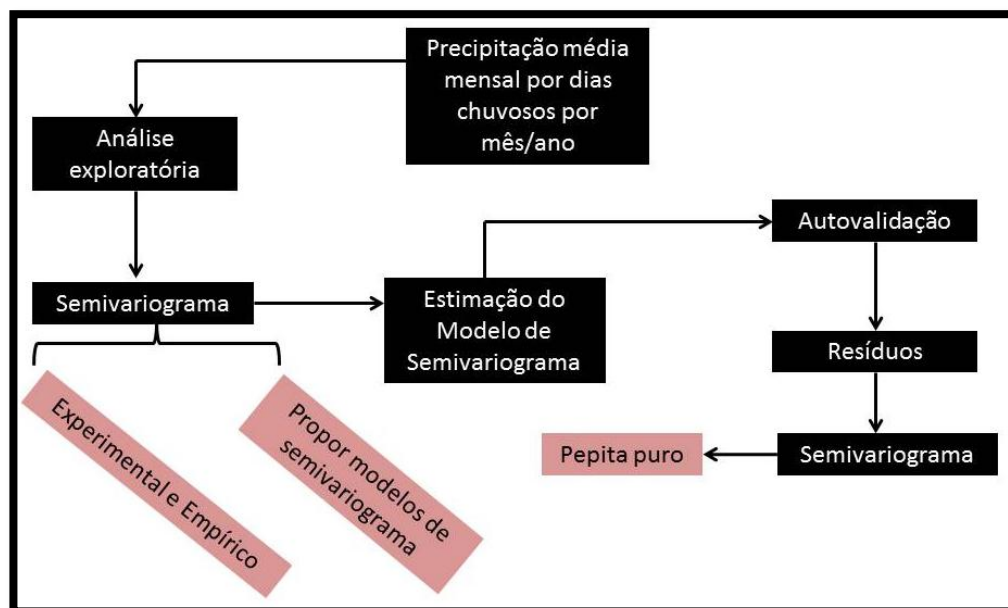


Figura 5 - Processo de análise geoestatística.

Inicialmente, realizou-se em cada mês por ano uma análise explanatória para verificar presença de possíveis tendências que poderiam atrapalhar a interpretação de dependência espacial entre a precipitação média mensal por dias chuvosos das estações pluviométricas. Em seguida, obteve-se para cada mês de cada ano um semivariograma empírico e observava-se sua estrutura, de forma a se propor um modelo de semivariograma a ser estimado. Com base no comportamento apresentado, propunha-se um ou mais modelos de semivariograma para descrever a dependência espacial entre as observações.

Neste trabalho optou-se em se trabalhar com os modelos de semivariograma exponencial, esférico e gaussiano.

No total realizou-se o ajuste de 72 modelos de semivariograma, uma para cada mês de cada ano (12 meses x 6 anos=72 meses).

A estimação dos modelos de semivariograma deu-se pelos métodos OLS e LM, de forma a caracterizar qual dos dois métodos de estimação é o mais eficiente. Os parâmetros obtidos pelo modelo de semivariograma empírico foram considerando nessa estimação.

A escolha entre os métodos de estimação deu-se pela autovalidação aplicada a todas as observações de precipitação média mensal por dias chuvosos. No caso, empregou-se a análise residual dos modelos de semivariograma estimados, onde se analisou a variância do erro de estimação e o AIC para escolher um dos métodos. Escolhido o modelo, obteve-se novamente o semivariograma empírico para os resíduos da autovalidação do modelo de semivariograma estimado, de forma a avaliar se ainda existem indícios de dependência espacial entre as localizações. Se ocorrer efeito pepita puro, o modelo estimado é o ideal; caso contrário, realizou-se todo o procedimento novamente.

3.3.3 Detecção de *Outliers*

A detecção de *outliers* em séries de precipitação média mensal por dias chuvosos deu-se por meio da estimação intervalar. A estimação intervalar, ou intervalo de confiança, é um intervalo estimado de um parâmetro de interesse em uma população. Em vez de se estimar o parâmetro por um único valor, este é dado por um intervalo de estimativas prováveis. O quanto estas estimativas são prováveis será determinado pelo coeficiente de confiança $(1-\alpha)$, para $\alpha \in (0,1)$. O nível α é chamado de nível de significância, e refere-se à probabilidade de se cometer o erro tipo I, rejeitar uma hipótese verdadeira.

Os estimadores são considerados variáveis aleatórias, e as estimativas obtidas quase certamente são distintas do valor do parâmetro, ou seja, quase certamente se comete um erro de estimação. Por esta razão torna-se necessário à construção de intervalos de confiança com probabilidade $(1-\alpha)$.

Neste trabalho, a detecção de *outliers* foi realizada considerando um IC dos resíduos obtidos pela análise de cada série temporal. Visto que os resíduos obtidos pela estimação de modelos de séries temporais não apresentam, por definição, nenhuma estrutura de dependência ou qualquer outro fator que desestabilize sua estrutura, estes são considerados ruídos branco, ou seja, normalmente distribuídos com média zero e variância constante. O que favorece seu emprego na construção do IC. Já os resíduos da autovalidação, consideram em sua estrutura o comportamento espacial, ou seja, se uma estação é discrepante em relação às estações vizinhas. Mas, se o comportamento discrepante destacar-se somente para poucas estações, isto pode ser indício da presença de valores discrepantes.

Neste contexto, o método proposto consiste em comparar o resíduo de cada estação obtido pela autovalidação com o IC dos resíduos da série temporal da mesma estação pluviométrica observada, conforme apresentado no esquema da Figura 6.

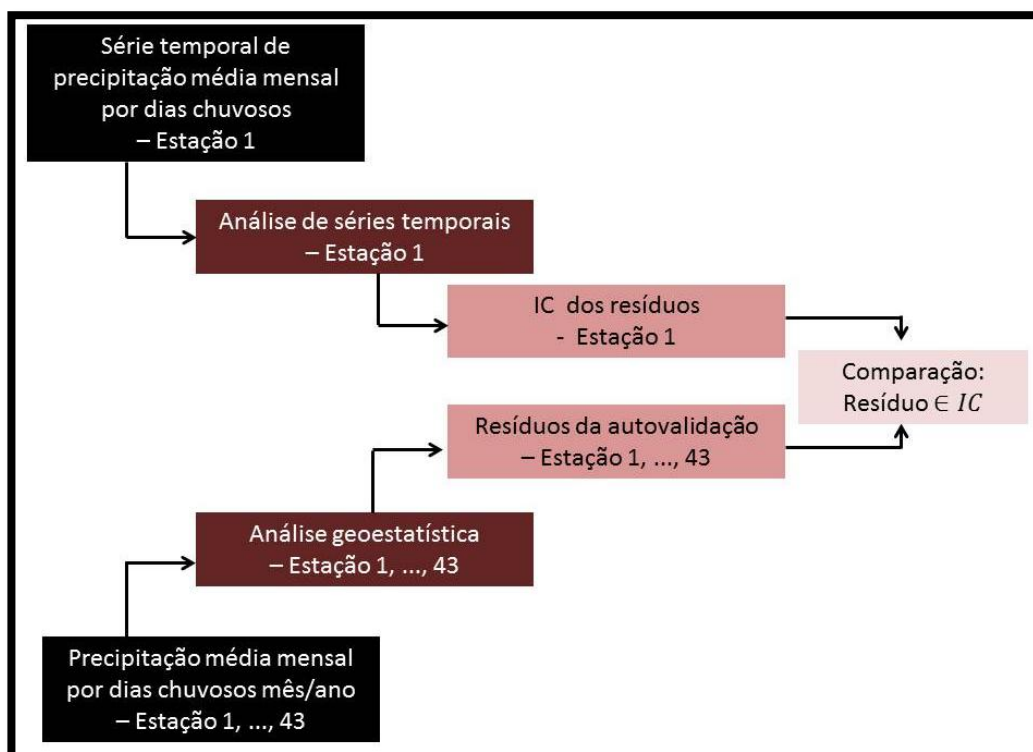


Figura 6 – Detecção de *outlier*.

O IC dos resíduos da série temporal de cada estação $(1-\alpha)\%$ foi obtido por

$$IC_{(1-\alpha)\%} = \left[\bar{\hat{\varepsilon}}_i - z_{\alpha/2} \frac{\sigma_{\hat{\varepsilon}}}{\sqrt{n}} \leq \hat{\varepsilon}_{s_i} \leq \bar{\hat{\varepsilon}}_i + z_{\alpha/2} \frac{\sigma_{\hat{\varepsilon}}}{\sqrt{n}} \right], \quad (23)$$

em que:

- $\bar{\hat{\varepsilon}}_i$ é a média residual da i -ésima estação pluviométrica, para $i = 1, \dots, 43$;
- $\sigma_{\hat{\varepsilon}}$ é o desvio padrão residual da i -ésima estação pluviométrica;
- $z_{\alpha/2}$ é o valor crítico para α na tabela da distribuição normal padronizada;
- $\hat{\varepsilon}_{s_i}$ é o resíduo da validação cruzada da estação pluviométrica i no ponto s_i para o mês m do ano a , em que $m = 1, \dots, 12$ e $a = 2000, \dots, 2005$.

Com base na Equação (23) o critério adotado para detecção de *outliers* constituiu em:

1. Avaliar se o resíduo do modelo espacial se encontra dentro do intervalo de confiança temporal obtido para cada mês em cada estação;
2. Se este estiver fora do intervalo será considerado um *outlier*.

Logo, havendo a presença de *outliers* pode-se considerar que a referente observação seja inconsistente.

4 RESULTADOS E DISCUSSÃO

Neste capítulo destacam-se a apresentação dos resultados obtidos pela aplicação da metodologia proposta. Primeiramente, apresentam-se os procedimentos da análise de séries temporais da precipitação média mensal por dias chuvosos para a estação pluviométrica Campina Verde (código 1949004) localizada no município de mesmo nome, conforme Figura 3. Em seguida, aplicam-se os procedimentos de Geoestatística para as 43 estações pluviométricas observadas no mês de janeiro de 2000. E, por último, os resultados da detecção de *outlier* são apresentados de forma geral.

4.1 RESULTADOS DE SÉRIES TEMPORAIS

A estação pluviométrica Campina Verde, localizada na longitude -49,28 e latitude -19,54, apresentou média de 9,69mm de precipitação média mensal por dias chuvosos, com mínima de 0 e máxima de 26,11mm. A Figura 7 apresenta o comportamento da série para o período do estudo, em que em vários meses são registrados valores altos para a precipitação média.

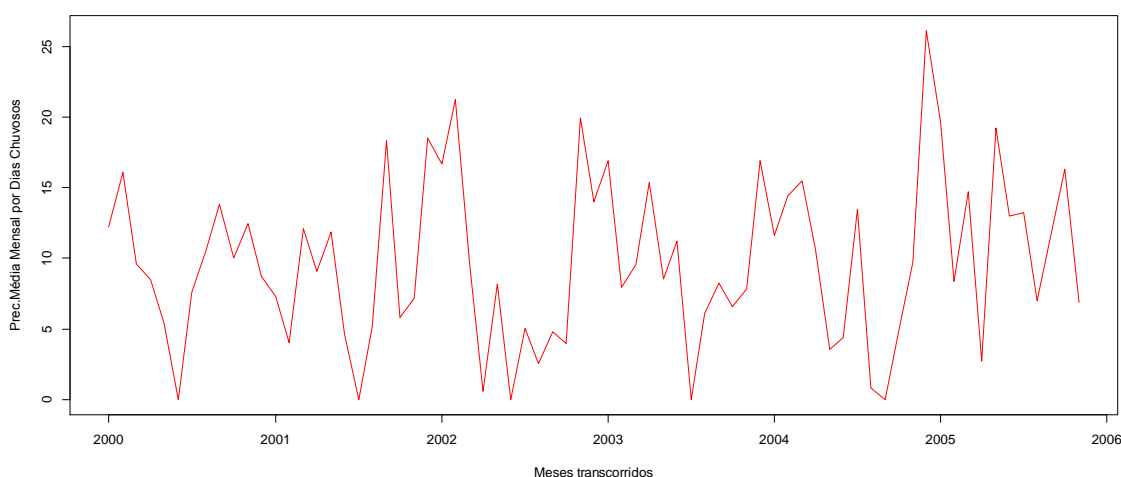


Figura 7 – Série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde para o período de janeiro de 2000 a novembro de 2005.

Como em quase todas as séries climatológicas, é possível observar pela Figura 7 que a série de precipitação apresenta periodicidades em sua estrutura

de dependência temporal. O mesmo pode ser observado nos gráficos de autocorrelação, apresentados na Figura 8.

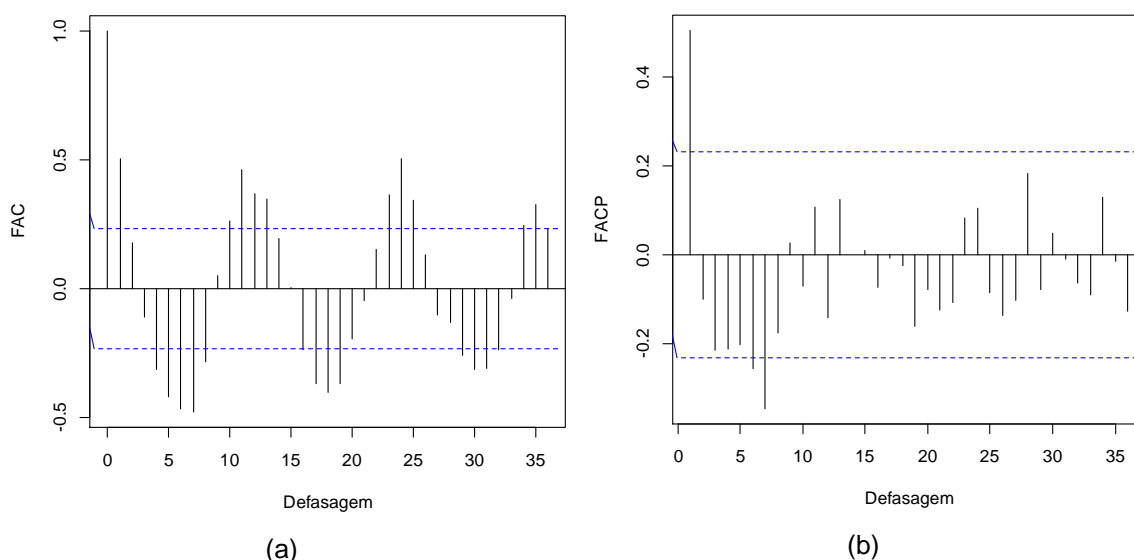


Figura 8 – Funções de autocorrelação da série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde. (a) Amostral; (b) Parcial.

Conforme Box e Jenkins (1970), uma série é considerada periódica se ao longo da FAC os valores apresentados nas defasagens tendem a se repetir em períodos regulares ou não. Pela Figura 8(a), observa-se a periodicidade irregular na FAC da série de Campina Verde, que, neste caso, a cada 11 defasagens ocorrem repetições do comportamento. Além disso, também pelas funções de autocorrelação é constatada a condição de estacionariedade da série. Pois, apesar da FAC apresentar picos de defasagens fora da área de significância, esta apresenta um decaimento aparentemente rápido entre as defasagens, indicando possivelmente a presença de componentes autorregressivas no modelo ajustado, conforme Brockwell e Davis (2002).

Assim, conforme os procedimentos já descritos eliminou-se a periodicidade da série por meio de uma regressão linear associada aos períodos significativos obtidos pelo periodograma. Pela Figura 9, verificam-se no periodograma os picos em que ocorreram as repetições.

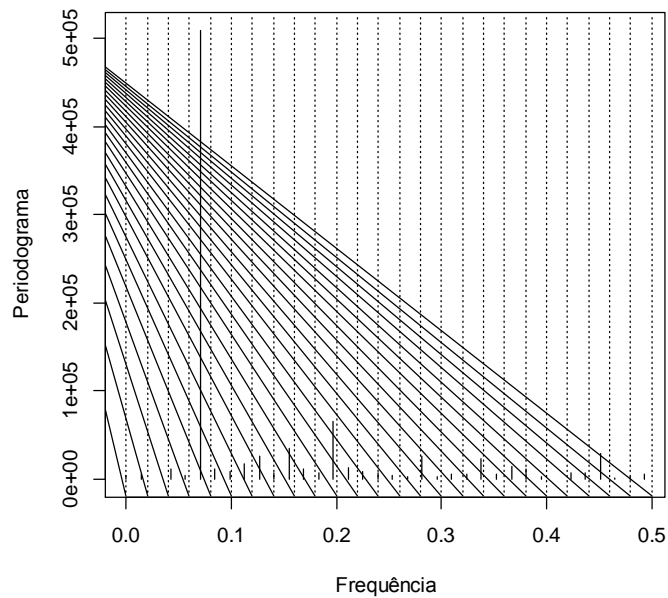


Figura 9 – Periodograma.

Conforme a Figura 9, pode-se considerar a existência que pelo menos 2 períodos possam ser significativos. Pela Tabela 1 observam-se 12 frequências e os períodos em que estas ocorrem, em que na posição 1 esta o equivalente ao maior pico visto na Figura 9. Como foi possível obter períodos significativos, tem-se que há periodicidade na série sob estudo, ou seja, existe periodicidade na série. Consequentemente, é possível descreve-la na forma de componentes senoidais em uma regressão, tal como Alpuim (1995) que utilizou o periodograma para obter os períodos significativos e, assim, descrever a temperatura da água do mar ao sul de Algarve, Portugal.

Tabela 1 - Frequências do periodograma

Posição	Frequência (f)	Período (p)
1	0,084507	14.113,81*
2	0,169014	1.487,09*
3	0,211268	973,1254*
4	0,352113	752,1064*
5	0,464789	677,3372*
6	0,197183	534,9083
7	0,112676	515,6694
8	0,140845	513,9512
9	0,323944	499,4033
10	0,098592	394,7112
11	0,225352	392,0304
12	0,295775	375,1233

*períodos significativos.

Sequencialmente, inseriram-se os respectivos períodos significativos na forma de pares de senos e cossenos em um modelo de regressão linear para descrever a precipitação média mensal por dias chuvosos, de acordo com a Equação (6). A tabela a seguir apresenta os coeficientes de regressão estimados e a análise do ajuste.

Tabela 2 – Modelo de regressão estimado para descrever a precipitação média mensal por número de dias chuvosos

Preditora	Parâmetro estimado	Erro Padrão	Estatística t	Valor-p
$Cos(2\pi * 0,0845t)$	84,9148	10,9113	7,782	0,0000*
$Sen(2\pi * 0,0845t)$	83,086	10,9113	7,615	0,0000*
$Cos(2\pi * 0,1690t)$	-0,7534	10,9113	-0,069	0,9452
$Sen(2\pi * 0,1690t)$	38,5554	10,9113	3,534	0,0008*
$Cos(2\pi * 0,2112t)$	-30,1275	10,9113	-2,761	0,0076*
$Sen(2\pi * 0,2112t)$	-8,0907	10,9113	-0,741	0,4612
$Cos(2\pi * 0,3521t)$	8,2214	10,9113	0,753	0,4541
$Sen(2\pi * 0,3521t)$	26,1632	10,9113	2,398	0,0196*
$Cos(2\pi * 0,4647t)$	12,6886	10,9113	1,163	0,2494
$Sen(2\pi * 0,4647t)$	22,7231	10,9113	2,083	0,0415*

Erro padrão residual: 65,01 em 61 GL
R² múltiplo: 0,7126, R² ajustado: 0,6654
Estatística F: 15,12 em 10 e 61 GL, valor-p: 4.096e-13

* Significativo ao nível de 5%.

Com base nesta estimativa, verificou-se que mais de 70% do ajuste de regressão por componentes senoidais foi explicado pelo modelo matemático estimado. Assim, estabeleceu-se uma nova série, sem a presença de periodicidade, o que pode ser visto na Figura 10 que apresenta as funções de autocorrelação da série temporal dos resíduos do modelo descrito na Tabela 2.

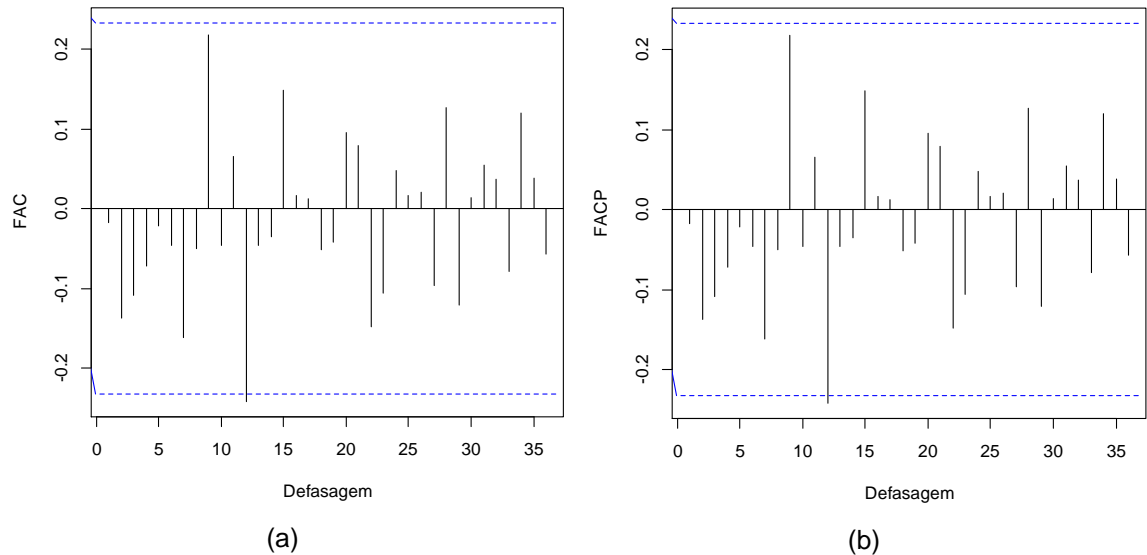


Figura 10 – Funções de autocorrelação da série temporal de precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde sem periodicidade. (a) Amostral; (b) Parcial.

Ainda analisando a Figura 10, tem-se que, conforme descrito anteriormente, a série de precipitação mensal média por dias chuvosos da estação Campina Verde é estacionária. Todas as suas defasagens encontram-se dentro do IC de significância estatística, o que é comprovado pelo teste de Dickey-Fuller (valor- $p < 0,05$). Logo, tem-se uma série sem periodicidade e estacionária, em que ainda existe dependência temporal entre as observações (Ljung-Box = 0,04373). Conseqüentemente, devem-se estabelecer as componentes autorregressivas e médias móveis do modelo *ARMA*.

Iniciando a identificação do modelo de séries temporais, os coeficientes das funções de autocorrelação na Figura 10 encontram-se dentro do intervalo de confiança da mesma, sugerindo que não há mais estrutura de dependência temporal, mas aplicando pelo teste de Ljung-Box verificou-se a presença de dependência temporal. Assim, propôs-se o modelo inicial *AR*(1) para descrever a nova série. Mas, como nenhum modelo é único optou-se por comparar o *AR*(1) estimado com outros modelos e escolher aquele que apresentar menor AIC (Tabela 3). O procedimento é o mesmo empregado por Sarnaglia e Reisen (2008), que utilizaram o AIC e o BIC para comparar modelos periódicos estimados em dados simulados, na presença e ausência de *outliers* do tipo aditivo.

Tabela 3 – Critério de Informação de Akaike dos modelos ajustados

Ordem		Modelo	AIC
<i>p</i>	<i>q</i>		
1	0	ARMA (1,0)	809,7976
2	0	ARMA (2,0)	810,5322
7	0	ARMA (7,0)	821,5860
0	1	ARMA (0,1)	809,7898
0	2	ARMA (0,2)	810,3982
0	7	ARMA (0,7)	821,4036
1	1	ARMA (1,1)	809,4925

Assim, alterando as respectivas ordens dos parâmetros autorregressivos e médias móveis estimaram-se 7 modelos. Conforme a tabela anterior, o modelo ajustado *ARIMA* (7,0,0) ou *ARMA* (7,0,) ou *AR* (7) foi que apresentou maior valor para o AIC, seguido do *MA* (7). O modelo inicialmente proposto, *AR* (1) apresentou AIC de 809,79, mas o *ARMA* (1,1), mesmo com mais parâmetros apresentou menor AIC, sendo este o escolhido. A Tabela 4 destaca a estimação do modelo *ARMA* (1,1).

Tabela 4 – Coeficientes estimados para o modelo *ARMA* (1, 1)

Coeficiente	Estimativa	Erro Padrão	Estatística t	Valor-p
ϕ_1	0,7725	0,2687	2,875	0,00404*
θ_1	-0,8870	0,1994	-4,447	8,7e-06*
Intercepto	0,1002	10,154	0,099	0,92143

$\hat{\sigma}^2 = 4812$;
 Soma de Quadrados Condicionais = 332.139,1
 AIC = 809,49

* Significativo ao nível de 5%.

Realizando a análise de adequação do modelo *ARMA* (1,1) estimado verificou-se que seus respectivos resíduos são normais (valor-p=0,4423) e independentes (valor-p=0,5963), ou seja, o modelo final adequado para descrever a precipitação média mensal por dias chuvosos da estação pluviométrica de Campina Verde é expresso da forma

$$\hat{Z}_t = 84,91\text{Cos}(2\pi * 0,0845t) + 83,08\text{Sen}(2\pi * 0,0845t) + 38,55\text{Sen}(2\pi * 0,1690t) - 30,12\text{Cos}(2\pi * 0,2112t) + 26,16\text{Sen}(2\pi * 0,3521t) + 22,72\text{Sen}(2\pi * 0,4647t) + 0,77Z_{t-1} + \hat{\varepsilon}_t + 0,88\hat{\varepsilon}_{t-1}.$$

Ressalta-se que todos os procedimentos, desde a constatação de periodicidade e estacionariedade até a validação dos pressupostos do modelo

resultou na obtenção de séries de resíduos para cada uma das estações pluviométricas sob estudo, ocasionando 43 modelos de séries temporais. Destes 43 modelos, todos possuíam componentes periódicas e somente 2 não eram originalmente estacionárias e necessitaram de diferenciação.

4.2 RESULTADOS DE GEOESTATÍSTICA

Inicialmente, apresenta-se o comportamento da precipitação média mensal por número de dias chuvosos. Ressalta-se novamente que a área sob estudo foi considerada isotrópica, sendo esta condição não verificada.

A Figura 11 apresenta a distribuição da precipitação média mensal por número de dias chuvosos, para as 43 estações pluviométricas no mês de janeiro de 2000, em categorias.

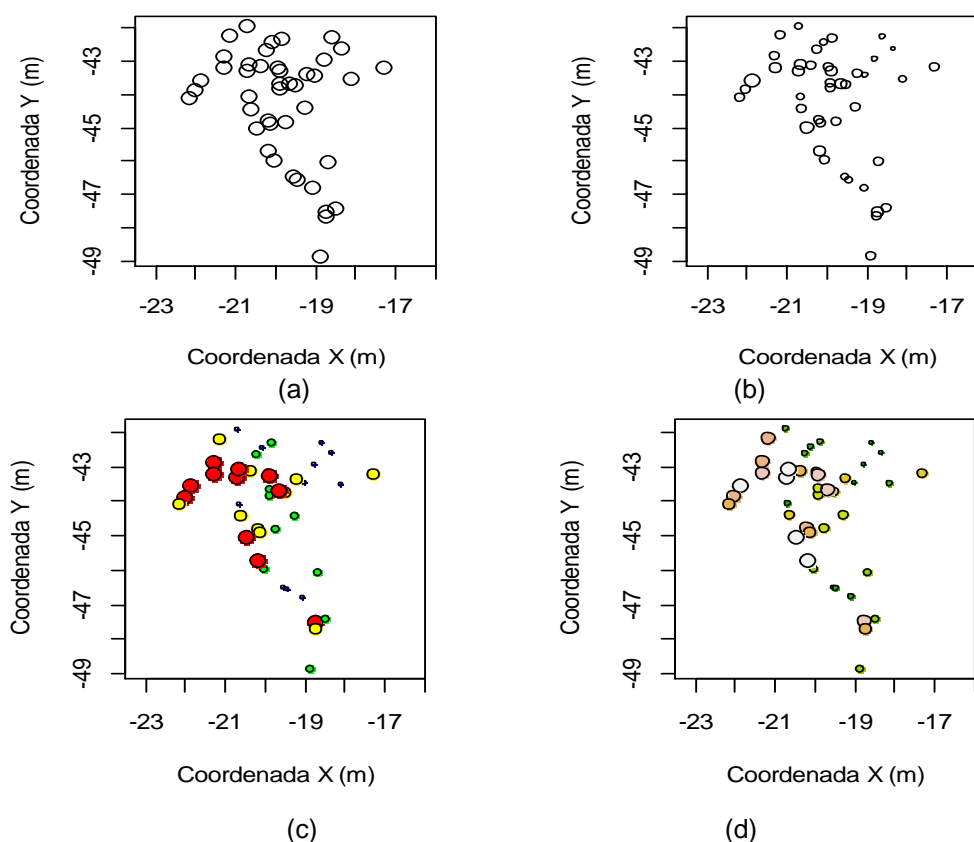


Figura 11 – Análise exploratória dos valores da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas. (a) Valores divididos com peso igual; (b) valores divididos com peso proporcional; (c) Valores divididos pelos quartis e (d) Valores divididos pelos decis.

Pela Figura 11 verificou-se que há variabilidade entre os valores apresentados pelas estações, visto que para cada tipo de separação

apresentada a forma em que os valores da precipitação são separadas em várias categorias. Mais precisamente, quando se separa os valores por meio dos quartis e dos decis fica mais visível a possível discrepância entre os valores apresentados para as locações observadas. Agora, visualizando os valores das precipitações nas estações considerando a tendência, vê-se pela Figura 12 a existência da mesma, de uma forma bem moderada o que não inviabiliza a continuidade da análise.

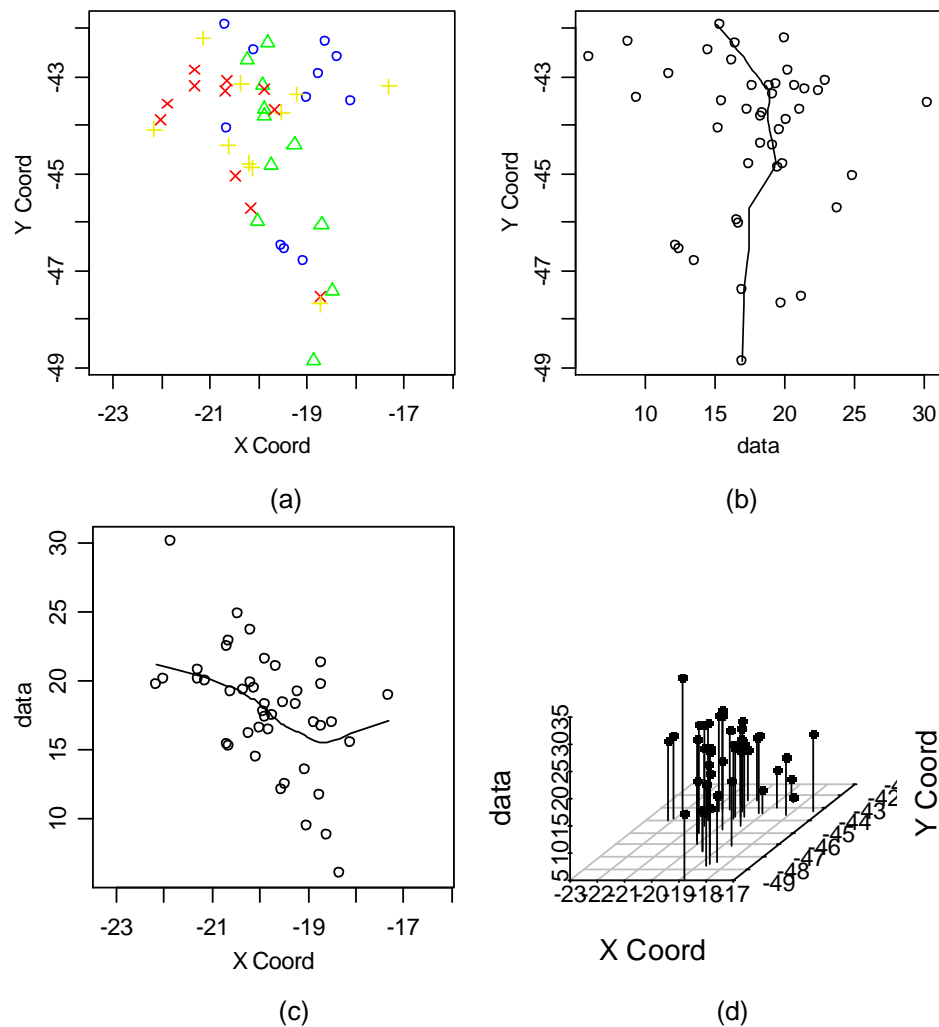


Figura 12– Análise de tendência da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas.

Caracterizando a construção do modelo de semivariograma apresenta-se na Figura 13 o semivariograma experimental, para distância máxima de defasagem 7. Um modelo, conforme a estrutura apresentada pelo semivariograma experimental, para descrever este semivariograma seria um modelo exponencial. Mas, analisando a Figura 14, em que são comparados os modelos de semivariograma exponencial, gaussiano e esférico, verificou-se

que o modelo esférico também é uma opção. Assim, ambos os modelos de semivariograma, exponencial e esférico, foram obtidos.

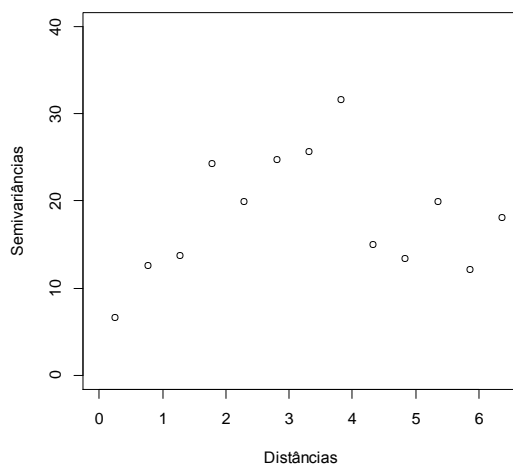


Figura 13 - Semivariograma experimental da precipitação média mensal por número de dias chuvosos para as 43 estações pluviométricas para o mês de janeiro de 2000.

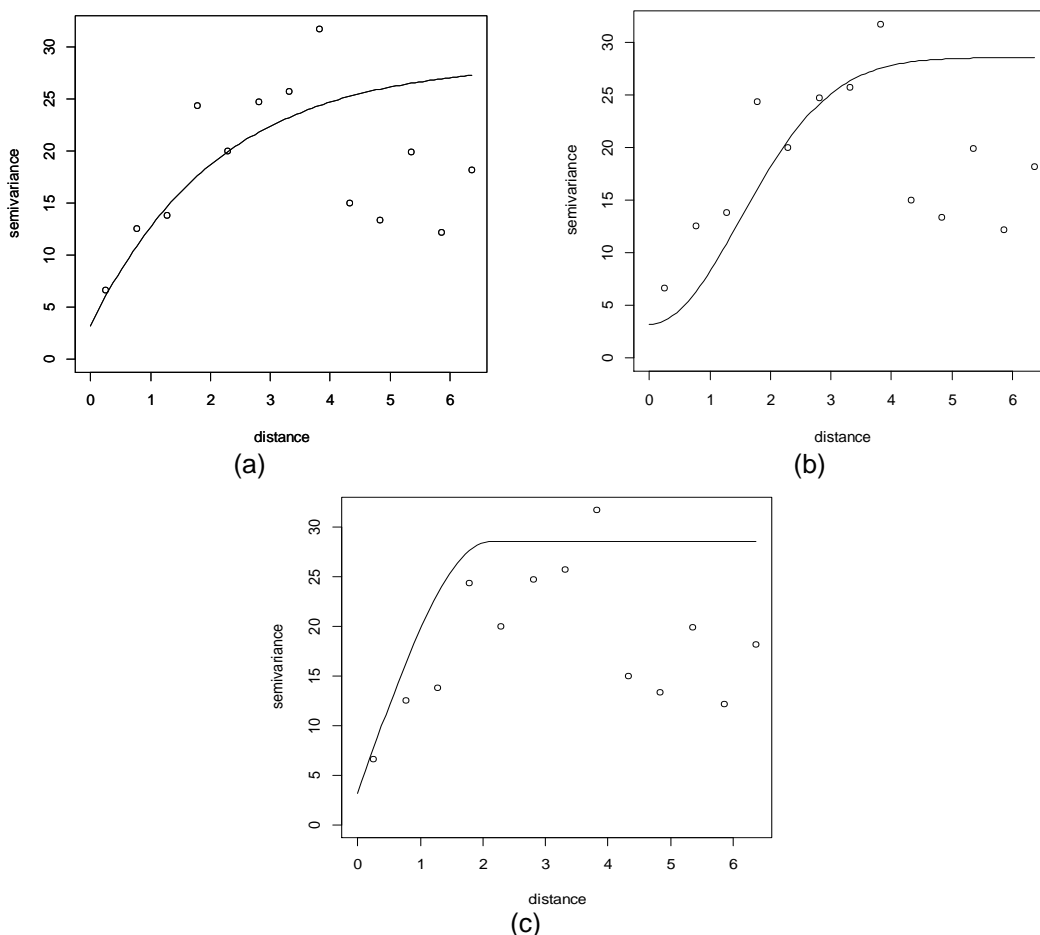


Figura 14 - Modelos de semivariograma empírico: (a) Exponencial, (b) Gaussiano e (c) Esférico.

O próximo passo foi estimar cada modelo de semivariograma por meio dos métodos de mínimos quadrados e de máxima verossimilhança,

considerando os seguintes valores para os parâmetros iniciais: contribuição = 25, alcance=2 e pepita =5.

Com base nos parâmetros do semivariograma empírico obtiveram-se os modelos de semivariograma estimados por OLS e ML, aonde o modelo estimado por OLS apresentou mais valor para o patamar (ver Figura 15).

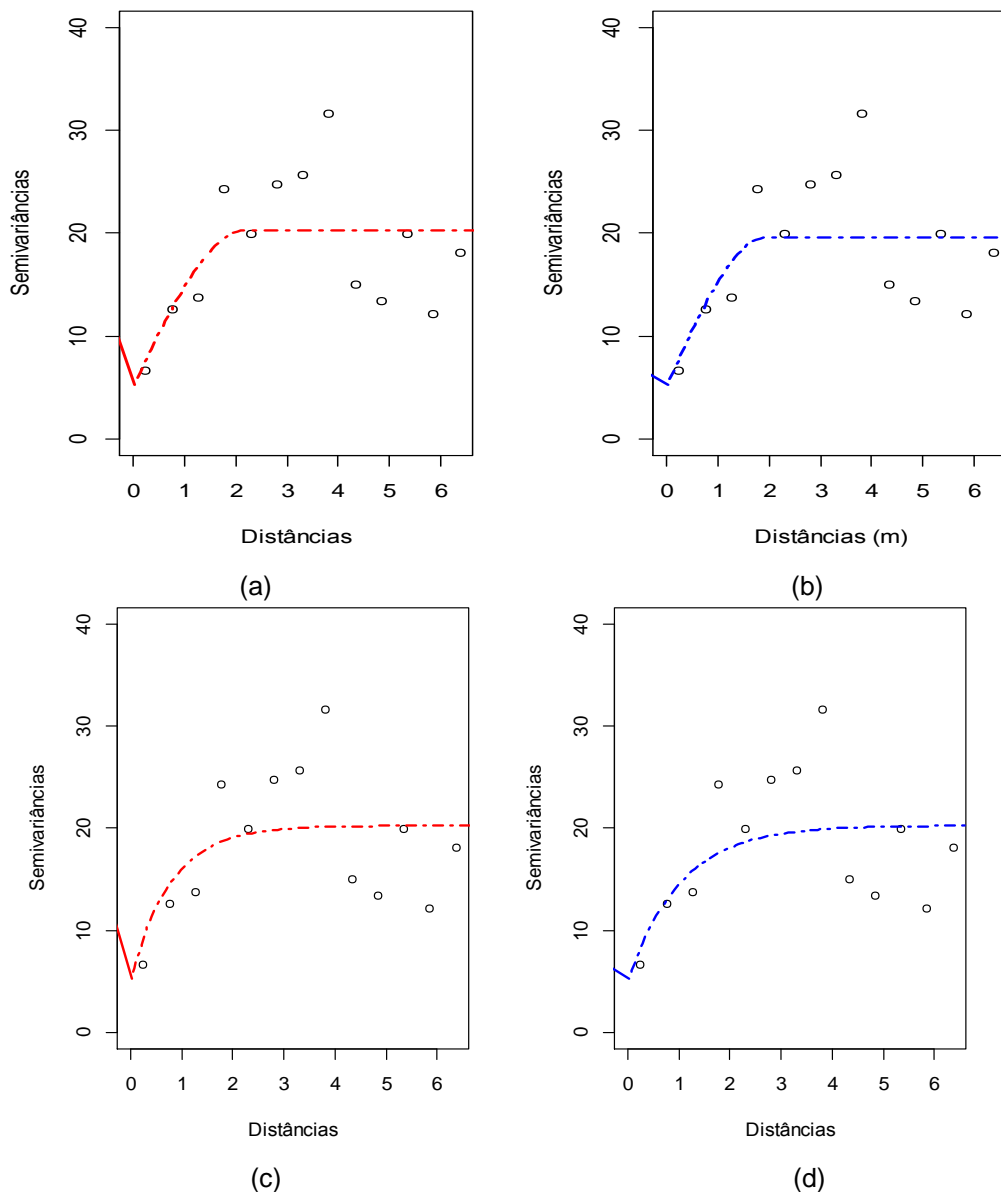


Figura 15 -Modelo de semivariograma estimado pelos métodos. (a) Esférico - Mínimos Quadrados Ordinários; (b) Esférico - Máxima Verossimilhança; (c) Exponencial - Mínimos Quadrados Ordinários; (d) Exponencial - Máxima Verossimilhança.

A escolha entre qual método de estimação e modelo de semivariograma deu-se por meio da autovalidação. Neste caso, ao se comparar ambos modelos de semivariograma pela autovalidação verificou-se que para qualquer método de estimação o modelo esférico apresentou menor erro padrão residual, conforme Quadro 2.

Quadro 2 - Sumarização dos resultados da autovalidação

Modelo Esférico									
Autovalidação - OLS					Autovalidação - LM				
Resíduos: Min 1Q Med 3Q Max -7,577 -1,818 -0,837 1,847 10,152					Resíduos: Min 1Q Med 3Q Max -7,7713 -2,0363 -0,6928 1,7547 10,2042				
Coefficientes: Estimativa Erro t Valor-t Intercepto -0,9941 3,2447 -0,306 0,761 β 1,0559 0,1794 5,887 6,27e ⁻⁰⁷ *					Coefficientes: Estimativa Erro t Valor-t Intercepto -1,4927 3,2764 -0,456 0,651 β 1,0838 0,1812 5,982 4,6e ⁻⁰⁷ *				
Erro padrão residual: 3,285 em 41 GL R ² Múltiplo: 0,4581 R ² Ajustado: 0,4448 Estatística F: 34,65 em 1e 41 GL Valor-p: 6,269e ⁻⁰⁷					Erro padrão residual: 3,26 em 41 GL R ² Múltiplo: 0,466 R ² Ajustado: 0,453 Estatística F: 35,78 em 1e 41 GL Valor-p: 4,596e ⁻⁰⁷				
Modelo Exponencial									
Autovalidação - OLS					Autovalidação - LM				
Resíduos: Min 1Q Med 3Q Max -7,6338 -2,0369 -0,7457 1,6734 10,4475					Resíduos: Min 1Q Med 3Q Max -7,5421 -2,0028 -0,7172 1,7930 10,3163				
Coefficientes: Estimativa Erro t Valor-t Intercepto -2,717 3,684 -0,738 0,465 β 1,152 0,204 5,645 1,38e ⁻⁰⁶ *					Coefficientes: Estimativa Erro t Valor-t Intercepto -1,6816 3,5151 -0,478 0,635 β 1,0942 0,1946 5,624 1,48e ⁻⁰⁶ *				
Erro padrão residual: 3,347 em 41 GL R ² Múltiplo: 0,4373 R ² Ajustado: 0,4236 Estatística F: 31,87 em 1e 41 GL Valor-p: 1,38e ⁻⁰⁶					Erro padrão residual: 3,35 em 41 GL R ² Múltiplo: 0,4355 R ² Ajustado: 0,4217 Estatística F: 31,63 em 1e 41 GL Valor-p: 1,479e ⁻⁰⁶				

* Significativo ao nível de 5%.

Escolhido o modelo de semivariograma esférico, avaliou-se seu método de estimação. Verificou-se que qualquer um dos dois métodos poderia ser utilizado, visto que suas estatísticas de adequacidade apresentarem valores muito próximos entre si. Mas, levando-se em consideração o parâmetro de regressão estimado, a estimação por mínimos quadrados foi a escolhida, por obter valor do parâmetro mais próximo de 1, ver Quadro 2.

Sequencialmente, após a escolha do modelo de semivariograma e de seu método de estimação realizou-se a avaliação da adequacidade do mesmo, em que se verificou a análise de dependência espacial dos resíduos da autovalidação. A Figura 16 apresenta o modelo de semivariograma experimental dos resíduos do modelo esférico estimado pelo método de mínimos quadrados.

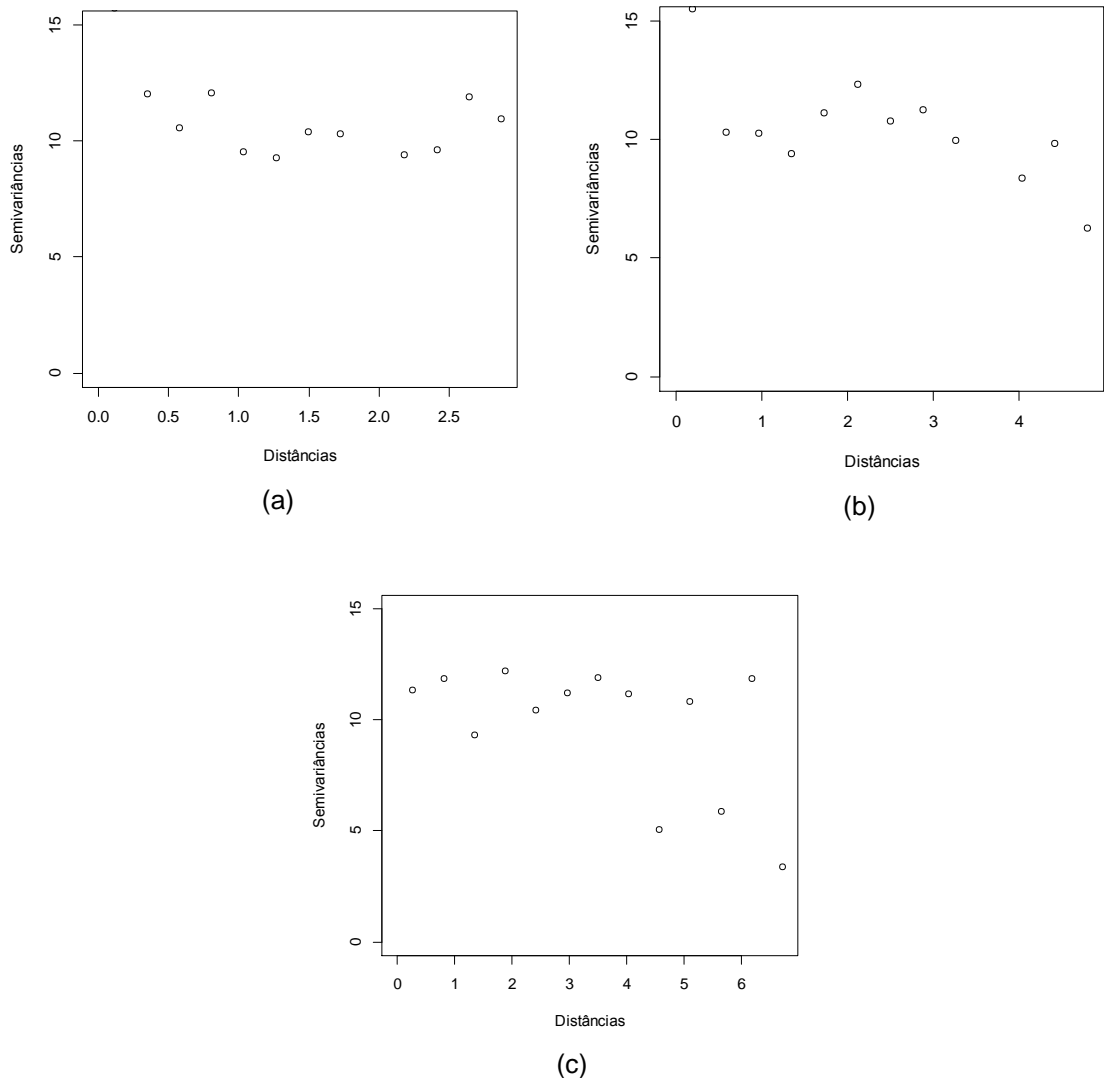


Figura 16 - Semivariograma experimental dos resíduos do modelo de semivariograma esférico ajustado pelo método de mínimos quadrados. (a) Distância = 3; (b) Distância = 5 e (c) Distância = 7.

Como é possível observar pela Figura 16, para qualquer valor de distância o semivariograma apresenta característica de um modelo de efeito pepita puro. Logo, considerando todos os resultados descritos, o modelo de semivariograma da precipitação média mensal por dias chuvosos para o mês de janeiro de 2000 das 43 estações pluviométricas foi descrito pelo modelo de semivariograma esférico ajustado pelo método de mínimos quadrados.

Analogamente, todos os procedimentos descritos aqui foram aplicados as demais estações pluviométricas, obtendo para cada mês de cada ano um modelo de semivariograma e um método de estimação para o mesmo.

4.3 DETECÇÃO DE *OUTLIER*

Para detectar os *outliers* realizou-se a construção de intervalos de confiança para a média residual de cada ajuste de série temporal realizado para as 43 estações pluviométricas. Em seguida, comparou-se se cada resíduo obtido da autovalidação do modelo esférico estimado pelo método de mínimos quadrados estava contido em seu respectivo intervalo de confiança. A Tabela 5 apresenta o teste para as 43 estações no mês de janeiro de 2000. Conforme o resultado, nesse mês houve apenas 1 estação com valores discrepantes das demais, a estação existente em Conceição do Mato Dentro.

Realizando as análises para as demais estações verificou-se que ocorreram no total 165 possíveis *outliers* para o período de janeiro de 2000 a novembro de 2005. As estações que mais apresentaram *outliers* foram as de Fazenda Campo Grande, Rio Piracicaba e Conceição do Mato Dentro, cada uma com 45, 11, e 10 meses em que foram detectados *outliers* (Tabela 4). Algumas estações não apresentaram grandes quantidades de valores discrepantes, mas quando detectado sua presença em um determinado ano, esta continua aparecendo nos demais anos, como no caso dos municípios de Horto Florestal, Sabará e Patrocínio do Muriaé.

Em relação ao anos, 2001 foi ano em que mais se detectou *outliers*, 33 no total. Em 2002 esse passou para 29, se estabilizando em 2003 e 2004 e voltando a crescer em 2005. Analisando os meses (Tabela 5), tem-se que em Janeiro, Fevereiro e Agosto detectou-se mais *outliers* do que nos outros meses, não estando o mês de Novembro não tão distante.

Tabela 5 -Teste de detecção de *outlier* para as 43 estações pluviométricas observadas, apresentando seus respectivos intervalos de confiança temporal de 95% e resíduos da autovalidação.

Código	Estação	Limite Inferior	Limite Superior	Resíduo	Situação
1949004	Campina Verde	-9.4197	9.439162	-1.69459	NORMAL
1847000	Major Porto	-9.37626	9.373992	1.878879	NORMAL
1847000	Monte Carmelo	-7.39733	7.558131	3.49329	NORMAL
1847001	Estrela do Sul	-5.70108	5.77568	0.670125	NORMAL
1847003	Abadia de Dourados	-10.397	10.37779	-2.52176	NORMAL
1848000	Monte Alegre de Minas	-7.71405	7.714723	-1.66787	NORMAL
1944010	Horto Florestal	-6.84049	6.916757	0.842045	NORMAL
1944011	Jaguaruna-Jusante	-7.42692	7.559532	-1.9429	NORMAL
1946004	Ibiá	-7.88812	7.890511	-0.96045	NORMAL
1946005	Salitre	-6.93279	7.019592	-1.68495	NORMAL
1743002	Vila Terra Branca-Jusante	-18.798	18.80619	3.353496	NORMAL
1842007	Guanhães	-8.764	8.712321	0.643343	NORMAL
1843003	Mendanha-Montante	-7.15691	7.991547	1.3564	NORMAL
1943001	Rio Piracicaba	-5.44249	5.533186	-1.7342	NORMAL
1943002	Conceição do Mato Dentro	-7.20609	7.188631	-7.57727	OUTLIER
1943004	Jaboticatubas	-8.40027	9.013436	-0.18699	NORMAL
1943006	Sabará	-7.64356	7.68459	-0.44034	NORMAL
1943010	Caeté	-9.12117	9.338143	-2.04968	NORMAL
1943023	Taquaraçu	-8.51329	8.501308	3.035362	NORMAL
1943025	Morro do Pilar	-7.15791	7.160965	5.432557	NORMAL
1943027	Usina Peti	-5.8825	6.090868	3.634917	NORMAL
2043009	Acaiaca-Jusante	-5.46911	5.416993	-0.83699	NORMAL
2043010	Piranga	-6.90087	6.949522	1.814242	NORMAL
2044003	Carmo do Cajuru	-8.21996	8.206356	-0.25674	NORMAL
2043014	Porto Firme	-5.20343	5.94548	2.553192	NORMAL
2044006	Divinópolis	-5.53853	5.298135	-1.06704	NORMAL
2044007	Entre Rios de Minas	-9.6462	9.57297	-5.22513	NORMAL
2044009	Fazenda Campo Grande	-3.73825	3.77462	0.202387	NORMAL
2045001	Bambuí	-8.98434	9.047581	-2.83653	NORMAL
2045002	Iguatama	-10.6204	10.61468	5.132621	NORMAL
2045005	Lamounier	-7.32116	7.313636	5.236341	NORMAL
2142000	Astolfo Dutra	-14.7693	14.5078	-1.52836	NORMAL
2142002	Patrocínio do Muriaé	-18.0777	18.17674	2.285785	NORMAL
2143000	Desterro do Melo	-8.41867	8.398461	-1.79965	NORMAL
2143016	Torreões	-10.9638	10.9739	10.15189	NORMAL
2243202	Fazenda São Gabriel	-12.2397	12.22983	-3.04967	NORMAL
2244035	Santa Rita do Jacutinga	-16.2598	15.43484	-1.03627	NORMAL
2041005	Caiana	-11.4327	11.46709	-1.97715	NORMAL
2042011	Rio Casca	-11.1167	11.12457	-1.42143	NORMAL
1842004	São Pedro do Suaçuí	-9.91266	9.880729	-5.75039	NORMAL
1842005	Coroaci	-8.4465	8.506022	-1.83728	NORMAL
1942002	Bom Jesus do Galho	-9.8873	9.899611	1.673653	NORMAL
2042008	Raul Soares-Montante	-8.33962	8.330458	-2.30692	NORMAL

Tabela 6 -Total de outliers detectados anualmente por estação pluviométrica

Código	Estação	Ano						Total
		2000	2001	2002	2003	2004	2005	
1949004	Campina Verde	0	0	0	0	0	0	0
1847000	Major Porto	3	0	0	3	2	0	8
1847000	Monte Carmelo	0	0	0	0	0	0	0
1847001	Estrela do Sul	0	0	0	0	0	0	0
1847003	Abadia de Dourados	0	0	0	0	0	0	0
1848000	Monte Alegre de Minas	0	0	0	0	0	0	0
1944010	Horto Florestal	1	1	0	1	1	1	5
1944011	Jaguaruna-Jusante	0	0	0	0	0	0	0
1946004	Ibiá	0	1	0	0	0	1	2
1946005	Salitre	0	1	0	0	0	1	2
1743002	Vila Tera Branca-Jusante	0	0	0	0	0	0	0
1842007	Guanhães	0	1	0	0	0	1	2
1843003	Mendanha-Montante	0	0	0	0	0	0	0
1943001	Rio Piracicaba	2	2	0	2	3	2	11
1943002	Conceição do Mato Dentro	2	1	2	2	2	1	10
1943004	Jaboticatubas	0	2	0	0	0	1	3
1943006	Sabará	1	1	0	1	1	1	5
1943010	Caeté	0	1	1	0	0	1	3
1943023	Taquaraçu	0	2	0	0	0	2	4
1943025	Morro do Pilar	0	0	0	0	0	0	0
1943027	Usina Peti	0	0	2	0	0	0	2
2043009	Acaiaca-Jusante	0	1	1	0	0	1	3
2043010	Piranga	0	1	0	0	0	1	2
2044003	Carmo do Cajuru	0	0	1	0	0	0	1
2043014	Porto Firme	0	2	1	0	0	2	5
2044006	Divinópolis	0	0	1	0	0	0	1
2044007	Entre Rios de Minas	0	0	0	0	0	0	0
2044009	Fazenda Campo Grande	8	7	9	7	8	6	45
2045001	Bambuí	0	0	0	0	0	0	0
2045002	Iguatama	0	0	1	0	0	0	1
2045005	Lamounier	0	0	1	0	0	0	1
2142000	Astolfo Dutra	0	0	1	0	0	0	1
2142002	Patrocínio do Muriaé	2	2	0	2	1	1	8
2143000	Desterro do Melo	0	1	1	0	0	1	3
2143016	Torreões	0	1	1	0	0	1	3
2243202	Fazenda São Gabriel	0	1	0	0	0	1	2
2244035	Santa Rita do Jacutinga	0	0	1	0	0	0	1
2041005	Caiana	1	2	1	1	1	2	8
2042011	Rio Casca	0	1	2	0	0	1	4
1842004	São Pedro do Suaçuí	0	0	0	0	0	0	0
1842005	Coroaci	0	0	0	0	0	0	0
1942002	Bom Jesus do Galho	0	0	0	0	0	0	0
2042008	Raul Soares-Montante	5	1	2	5	5	1	19
	Total	25	33	29	24	24	30	165

Tabela 7–Total mensal de *outliers* detectados anualmente

Ano	Mês												Total
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	
2000	1	5	1	1	1	0	4	0	1	3	4	4	25
2001	8	6	1	1	2	0	1	9	0	1	2	2	33
2002	4	1	6	5	1	0	2	2	1	1	4	1	29
2003	2	5	1	1	1	0	4	1	2	3	3	3	24
2004	1	5	1	1	1	0	4	1	1	4	4	4	24
2005	5	6	1	1	2	0	1	7	0	1	2		30
Total	21	28	11	10	8	0	16	20	5	13	19	14	165

5 CONCLUSÃO

Este trabalho teve por objetivo associar as análises de Séries Temporais e de Geoestatística para obter um método de detecção de *outliers*. Para tal, realizou-se a detecção de *outliers* em séries de precipitação observadas em estações pluviométricas existentes no estado de Minas Gerais. Como consequência do objetivo deste é possível estabelecer se um conjunto de observações é consistente ou não para serem empregadas em estudos.

Os resultados obtidos fornecem indícios de que o método proposto é adequado para a detecção de *outliers* em séries de precipitação. Para o período sob estudo constatou-se que a estação pluviométrica existente no município de Conceição do Mato Dentro apresentou maior quantidade de *outliers*. Ressalta-se que o método estabelece somente a existência ou não do *outlier*, sendo que estabelecer sua origem há a necessidade de estudos de campo de forma a verificar a estação, seus equipamentos, entre outros para se estabelecer a causa de discrepância entre as observações.

Dessa forma, o método de detecção de *outliers* apresentado pode ser considerado uma forma de validar a forma de coleta de observações referentes à precipitação. Como as técnicas aqui apresentadas são técnicas comumente empregadas em diversos tipos de estudos, o método em questão pode ser utilizado da mesma forma.

REFERÊNCIAS BIBLIOGRÁFICAS

AGÊNCIA NACIONAL DE ÁGUAS (ANA), download em <http://hidroweb.ana.gov.br>.

ALEXANDERSSON, H. **A homogeneity test applied to precipitation data.** *Journal of Climatology*, **6**: 661–675, 1986.

ALPUIM, T. **Detecção de periodicidades em séries temporais: o periodograma.** Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Working paper nº 19, 1995.

ARAÚJO, W. F.; ANDRADE, A. S. J.; MEDEIROS, R. D.; SAMPAIO, R.S. Precipitação pluviométrica mensal provável em Boa Vista, estado de Roraima, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v5, n.3, p.563-567, 2001.

BOX, G. E. P.; DRAPER, N. R. **Empirical Model-Building and Response Surfaces**, Wiley, 1987.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis, Forecasting and Control.** Holden-Day, San Francisco, California, 1970.

BOX, G.E.P.; G.C. TIAO. Intervention analysis with applications to economic and environmental problems. **J. Am. Statist. Assoc.**, 70: 70-79, 1995.

BROCKWELL, P.J.; DAVIS, R.A. **Introduction to time series and forecasting.** 2nd Edition, Springer, 2002.

BUISHAND, T.A. The analysis of homogeneity of long-term rainfall records in the Netherlands. **KNMI Scientific Report** WR 81-7, DeBilt, The Netherlands, 1981.

BUISHAND, T.A. Some methods for testing the homogeneity of rainfall records. **Journal of Hydrology**, 58: 11–27, 1982.

CÂMARA, G. ; MEDEIROS, J. S. **.Geoprocessamento em projetos ambientais.** 2a.. ed. São José dos Campos: INPE, 1998. v. único. 195p.

CARAM, R.O. **Reconstrução de Séries e Análise Geoestatística da Precipitação no Estado de Minas Gerais.** Dissertação (Mestrado em Meteorologia Agrícola), Universidade Federal de Viçosa, Viçosa, 2007.

CARVALHO, A.C.A. **Caracterização da rede Pluviométrica do Estado de Sergipe e Aplicação de Redes Neurais para Preenchimento de Falhas.** Dissertação (Mestrado em Desenvolvimento e Meio Ambiente), Universidade Federal de Sergipe, São Cristóvão, 2007.

COSTA, A. C., NEGREIROS, J.; SOARES, A. Identification of inhomogeneities in precipitation time series using stochastic simulation. In: SOARES, A., PEREIRA, M. J. and DIMITRAKOPOULOS, R. (Eds.), **geoENV VI - Geostatistics for Environmental Applications**, Springer, 275-282, 2008.

CRESSIE, N.A.C. **Statistics for Spatial Data**, revised edition. John Wiley & Sons, New York, 416 p, 1993.

DICKEY, D.; FULLER, W. Distribution of the estimators for autoregressive time series with a unit root, **Journal of the American Statistical Association**, 74(366), 1979.

DICKEY, D.A.; FULLER, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica**, 49:1057-1072, 1981.

DICKEY, D.A.; HASZA, H.P.; FULLER, W.A. Testing for unit roots in seasonal times series. **Journal of the American Statistical Association**. Vol. 79, n. 386, pág.355-367, 1984.

DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. Model-based geostatistics. In: **Simpósio de Probabilidade Estatística**, 14, 2000, Caxambú. Resumos... Caxambú: Associação Brasileira de Estatística, 2000. 129p

EHLERS, R. **Análise de Séries Temporais**. Notas de aula, 2009.

FOX, A. J. Outliers in time series, **Journal of the Royal Statistical Society**, Ser. B, 34 (1972), 350-363.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation** (Applied Geostatistics). Oxford University Press, New York, 496 p, 1997.

HAMILTON, J. D. **Time Series Analysis**. Princeton University Press, 1994.

HARTLEY, H.O., RAO, J.N.K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, 54(1/2): 93-108, 1967.

HENGL, T. **A Practical Guide to Geostatistical Mapping**, download em <http://www.lulu.com>, 2009.

HOLLANDER, M.; WOLFE, D.A. **Nonparametric Statistical Methods**. New York: John Wiley & Sons. Second edition, 1999.

ISAAKS, E.H.; SRIVASTAVA, R.M. **An Introduction to Applied Geostatistics**. Oxford University Press, New York, 542 p, 1989.

KHOSHNEVISAN, D. **Multiparameter Processes: An Introduction to Random Fields**, Springer, 2002.

LATORRE, M.R.D.O.; CARDOSO, M.R.A. Time series analysis in epidemiology: an introduction to methodological aspects. **Revista Brasileira de Epidemiologia**, v.4, n.3, p.145-152, 2001.

- LE, N. D.; ZIDEK, J. V. **Statistical Analysis of Environmental Space–Time Processes**. Springer, New York, 2006.
- MELLO, C. R. de; SILVA, A. M. Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.1, p.68-74, 2009.
- MELLO, C. R.; VIOLA, M. R.; MELLO, J. M.; SILVA, A. M. Continuidade espacial de chuvas intensas no estado de Minas Gerais. **Ciência e Agrotecnologia**, v.32, n.2, p.532-539, 2008a.
- MELLO, C. R.; VIOLA, M. R.; NORTON, L. D.; SILVA, A. M.; WEIMAR, F. A. Development and application of a simple hydrologic model simulation for a Brazilian headwater basin. **Catena**, v.75, p.235-247, 2008b.
- MELLO, J. M. de. **Geoestatística aplicada ao inventário florestal**. Piracicaba: ESALQ/USP, 2004. 110p. Tese Doutorado.
- MELLO, J. M. de; BATISTIA, J. L. F.; RIBEIRO JUNIOR, P. J.; OLIVEIRA, M. S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. **Scientia Florestalis**, v.1, n.69, p.25-37, 2005.
- MINUZZI, R.B.; SEDIYAMA, G. C.; COSTA, J. M. N.; VIANELLO, R. L. Influência da La Niña na estação chuvosa da Região Sudeste, **Revista Brasileira de Meteorologia**, São José dos Campos, v.22, n. 3 p.345-353, 2007
- MORETTIN, P.A.; TOLOI, Clélia M. C. **Análise de Séries Temporais**. 2ª Edição, 2002.
- PETTITT, A.N. A non-parametric approach to the change-point problem. **Appl. Statis.**, 28 (2): 126-135, 1979.
- PHILLIPS, P. C. B.; PERRON, P. Testing for a Unit Root in Time Series Regression. **Biometrika** 75, 335-346, 1988.
- R Development Core Team. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>, 2012.
- RIBEIRO JUNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995.99f. Dissertação de Mestrado, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, São Paulo.
- SARNAGLIA, A.J.Q; REISEN, V.A. Estimção Robusta em Processos Periódicos Auto-Regressivos na Presença de Outliers Aditivos, **In: 18º Simpósio Nacional de Probabilidade e Estatística**, Estância de São Pedro, São Paulo, 2008.

TUCCI, C. E. M. **Hidrologia: ciência e aplicações**. 2.ed.; 2.reimp. – Porto Alegre: Ed. Universidade/

UFRGS: ABRH: 2001.

VAREJÃO-SILVA, M. A. **Meteorologia e Climatologia**. Recife: 2006. (Versão Digital 2).

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R.F.; ALVAREZ V., V.H. & SCHAEFER, G.R., eds. **Tópicos em ciência do solo**. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, 2000. v.1. p.1-54.

WEI, W. **Time Series Analysis: Univariate and Multivariate Methods**. Pearson Books, second edition, 2006.

ANEXO A – LOCALIZAÇÃO DAS ESTAÇÕES PLUVIOMÉTRICAS

Código	Nome da Estação	Longitude	Latitude	Código	Nome da Estação	Longitude	Latitude
1949004	Campina Verde	-49,48	-19,54	2043010	Piranga	-43,3	-20,69
1846003	Major Porto	-46,04	-18,71	2044003	Carmo do Cajuru	-44,79	-20,19
1847000	Monte Carmelo	-47,52	-18,72	2043014	Porto Firme	-43,09	-20,67
1847001	Estrela do Sul	-47,69	-18,74	2044006	Divinópolis	-44,89	-20,14
1847003	Abadia de Dourados	-47,41	-18,49	2044007	Entre Rios de Minas	-44,07	-20,66
1848000	Monte Alegre de Minas	-48,87	-18,87	2044009	Fazenda Campo Grande	-44,43	-20,63
1944010	Horto Florestal	-44,4	-19,27	2045001	Bambuí	-45,97	-20,02
1944011	Jaguaruna-Jusante	-44,81	-19,73	2045002	Iguatama	-45,7	-20,18
1946004	Ibiá	-46,54	-19,48	2045005	Lamounier	-45,04	-20,47
1946005	Salitre	-46,8	-19,07	2142000	Astolfo Dutra	-42,86	-21,31
1743002	Vila Tera Branca-Jusante	-43,21	-17,31	2142002	Patrocínio do Muriaé	-42,22	-21,15
1842007	Guanhães	-42,93	-18,77	2143000	Desterro do Melo	-43,2	-21,31
1843003	Mendonha-Montante	-43,52	-18,11	2143016	Torreões	-43,56	-21,87
1943001	Rio Piracicaba	-43,18	-19,92	2243202	Fazenda São Gabriel	-43,88	-22,01
1943002	Conceição do Mato Dentro	-43,44	-19,02	2244035	Santa Rita do Jacutinga	-44,09	-22,15
1943004	Jaboticatubas	-43,74	-19,52	2041005	Caiana	-41,92	-20,69
1943006	Sabará	-43,82	-19,89	2042011	Rio Casca	-42,65	-20,23
1943010	Caeté	-43,67	-19,9	1842004	São Pedro do Suaçuí	-42,6	-18,36
1943023	Taquaraçu	-43,69	-19,66	1842005	Coroaci	-42,28	-18,61
1943025	Morro do Pilar	-43,37	-19,22	1942002	Bom Jesus do Galho	-42,32	-19,83
1943027	Usina Peti	-43,27	-19,88	2042008	Raul Soares-Montante	-42,44	-20,1
2043009	Acaiaca-Jusante	-43,14	-20,36				

ANEXO B – SÉRIES DE PRECIPITAÇÃO MÉDIA POR DIA CHUVOSO

Data	Código da Estação Pluviométrica													
	1949004	1846003	1847000	1847001	1847003	1848000	1944010	1944011	1946004	1946005	1743002	1842007	1843003	1943001
01/01/2000	256,20	285,40	490,40	416,20	272,40	322,60	312,50	333,40	299,70	258,50	303,90	246,40	279,10	373,00
01/02/2000	321,80	347,30	275,00	329,60	222,50	298,50	284,60	161,40	328,10	262,00	85,10	128,20	112,50	93,40
01/03/2000	173,40	247,40	323,30	252,50	337,10	192,10	196,90	210,80	227,60	206,90	207,40	363,70	166,50	252,00
01/04/2000	51,00	23,80	67,10	48,50	32,10	47,00	23,90	24,50	76,20	47,80	12,70	9,40	12,30	34,70
01/05/2000	5,40	0,00	0,00	0,00	0,00	0,20	2,70	0,00	5,50	0,00	0,60	4,40	1,20	1,20
01/06/2000	0,00	0,00	0,00	2,20	0,00	0,00	0,00	0,00	0,80	6,60	0,00	0,90	7,10	3,40
01/07/2000	15,20	0,00	16,70	13,30	10,00	10,90	0,00	7,70	5,80	12,30	2,10	6,30	3,70	6,60
01/08/2000	42,10	18,70	20,80	14,40	28,60	10,30	32,90	23,80	14,20	13,20	11,80	31,60	29,30	35,50
01/09/2000	124,40	85,60	122,20	86,30	61,30	56,40	51,70	50,60	100,40	123,30	47,10	81,80	45,20	84,70
01/10/2000	20,00	71,20	34,50	54,00	38,00	15,30	9,00	45,50	96,20	100,40	79,00	30,00	80,60	43,50
01/11/2000	187,00	254,60	249,80	251,90	257,30	180,10	232,60	308,50	170,70	232,00	279,60	317,90	275,30	413,70
01/12/2000	173,80	209,30	237,60	273,50	221,10	253,70	191,30	257,20	216,50	180,70	182,30	241,90	307,80	265,50
01/01/2001	102,20	121,10	204,20	195,80	89,90	145,00	183,00	122,90	225,40	74,00	61,70	37,10	129,70	112,30
01/02/2001	44,70	43,30	112,20	80,30	102,20	71,30	62,90	87,10	130,90	142,70	10,70	57,30	40,20	12,60
01/03/2001	217,50	171,80	90,70	125,00	254,70	275,60	66,30	64,70	148,40	235,70	32,70	113,90	79,00	124,10
01/04/2001	36,20	0,00	32,40	21,30	17,30	64,10	0,80	5,20	21,00	17,50	0,00	21,60	19,10	1,90
01/05/2001	83,00	25,90	29,60	34,50	40,10	65,40	19,70	35,40	46,40	58,00	36,50	67,80	49,40	37,90
01/06/2001	13,80	0,00	0,00	2,40	0,00	0,30	0,00	0,00	0,10	0,00	4,50	20,30	7,80	2,30
01/07/2001	0,00	0,00	0,10	0,00	0,00	0,00	0,00	4,50	19,30	1,60	0,30	1,00	0,00	2,10
01/08/2001	35,90	64,20	28,10	31,10	26,30	34,40	33,10	9,50	56,30	68,90	1,80	78,00	20,00	9,00
01/09/2001	91,70	17,00	86,80	14,20	34,30	74,80	64,30	52,10	55,20	87,50	0,80	33,60	26,00	180,60
01/10/2001	52,40	92,10	85,40	116,00	111,60	189,00	71,90	129,40	87,70	118,00	52,90	70,90	97,10	85,80
01/11/2001	122,10	172,10	177,00	269,80	255,60	168,80	144,40	102,40	244,50	262,90	268,40	271,30	148,60	324,70
01/12/2001	240,70	349,70	454,90	476,80	394,60	411,10	231,60	276,40	384,40	301,60	231,90	283,00	357,40	355,10
01/01/2002	233,00	192,30	274,30	241,50	246,60	240,40	191,00	201,90	161,90	281,90	158,00	236,60	330,50	344,40

01/02/2002	425,20	274,50	357,00	365,00	360,10	260,90	303,10	309,70	215,20	259,90	155,90	300,90	177,50	200,70
01/03/2002	105,60	83,00	134,80	186,10	117,10	133,00	53,60	32,00	90,20	112,30	117,80	35,80	38,80	25,00
01/04/2002	0,60	2,50	12,70	40,80	14,30	3,70	25,10	21,80	27,80	19,20	44,40	17,80	43,00	46,90
01/05/2002	32,70	22,80	47,70	36,80	14,50	62,00	25,50	15,30	34,70	61,60	5,80	12,00	4,50	24,40
01/06/2002	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	2,00	6,00	0,30	3,70
01/07/2002	10,10	29,50	4,00	2,60	0,60	3,70	16,70	7,60	11,50	1,80	0,10	1,10	10,50	6,80
01/08/2002	12,60	1,20	0,00	0,00	0,00	0,80	0,00	0,00	0,30	0,40	3,50	10,70	1,00	1,00
01/09/2002	48,20	77,00	31,00	25,50	42,10	32,10	34,00	74,90	42,70	57,60	57,60	121,70	97,00	102,60
01/10/2002	19,90	63,00	100,40	37,80	100,70	73,60	43,30	115,40	31,40	79,70	0,60	52,00	10,20	60,80
01/11/2002	219,50	81,10	126,00	95,50	81,40	126,00	125,80	189,90	50,70	145,80	269,10	220,40	195,60	324,80
01/12/2002	224,10	199,10	185,00	335,70	235,40	228,80	301,00	228,10	254,60	213,00	351,60	408,50	252,20	329,10
01/01/2003	354,50	364,80	480,70	651,90	425,70	478,70	344,70	296,60	374,50	395,10	270,80	306,60	300,80	477,00
01/02/2003	102,90	42,60	97,90	108,20	92,40	137,50	53,90	58,30	89,70	121,30	29,10	66,20	43,10	56,50
01/03/2003	171,60	242,10	333,00	252,60	149,20	219,90	186,40	154,80	146,60	216,20	121,60	98,40	159,00	165,50
01/04/2003	184,10	38,10	93,60	90,10	57,10	195,70	17,60	20,30	26,10	37,80	13,50	38,80	7,60	81,00
01/05/2003	25,60	57,40	11,90	21,50	10,90	23,50	28,70	44,80	27,50	42,90	3,80	7,60	10,00	12,00
01/06/2003	11,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	3,60
01/07/2003	0,00	0,00	0,00	0,00	0,00	0,00	0,00	3,10	13,80	3,50	0,40	0,90	0,00	5,80
01/08/2003	18,30	55,30	0,80	4,30	0,50	5,60	13,00	9,00	15,30	11,50	29,70	10,80	25,70	26,10
01/09/2003	32,90	18,50	116,30	140,60	36,20	38,90	21,60	16,70	46,70	68,10	2,30	17,50	5,30	37,10
01/10/2003	59,00	23,00	51,70	65,80	98,60	80,80	22,00	43,80	65,60	45,00	212,70	85,50	68,90	67,10
01/11/2003	93,60	167,60	144,20	151,20	99,40	180,60	136,40	89,60	223,90	144,00	82,80	167,30	68,70	181,30
01/12/2003	219,90	215,10	208,60	336,40	153,40	183,80	185,60	197,40	383,00	243,50	224,80	300,20	194,50	269,60
01/01/2004	243,70	270,30	425,30	362,10	339,70	227,40	286,50	280,50	501,60	420,30	199,00	245,30	280,10	441,90
01/02/2004	259,50	264,10	364,10	296,10	338,70	278,20	296,40	273,80	267,40	348,80	277,80	314,00	283,70	319,10
01/03/2004	155,00	250,00	218,60	204,30	141,20	83,90	218,80	117,20	291,30	216,80	150,40	145,60	131,30	156,30
01/04/2004	127,70	151,80	61,50	70,20	40,20	146,10	44,40	55,20	220,10	129,50	187,00	128,10	139,60	188,80
01/05/2004	28,50	18,00	8,80	33,10	15,60	30,40	4,20	7,10	38,80	21,70	0,30	6,20	4,50	34,80
01/06/2004	17,50	0,00	0,30	9,90	0,00	45,90	0,00	4,30	15,50	1,30	23,40	11,80	3,30	7,40
01/07/2004	53,80	19,90	25,80	37,80	21,70	27,40	39,40	37,70	40,40	38,90	4,60	40,90	13,60	37,90
01/08/2004	0,80	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,30	2,80

01/09/2004	0,00	0,00	0,70	0,70	1,00	3,10	0,00	0,10	2,70	3,60	0,00	0,00	0,00	1,00
01/10/2004	57,70	55,80	62,90	31,70	25,50	104,70	92,30	47,90	87,00	160,50	69,00	24,30	29,60	79,70
01/11/2004	146,00	102,50	99,40	110,20	112,50	164,90	111,90	77,20	152,50	151,20	129,30	112,40	140,60	92,60
01/12/2004	496,00	252,40	321,20	382,10	339,40	286,20	358,60	347,20	363,10	258,30	298,60	450,10	257,70	351,70
01/01/2005	412,70	167,40	324,40	280,60	332,90	460,90	235,30	284,30	371,50	295,20	173,60	268,60	185,00	264,70
01/02/2005	50,10	94,30	52,90	74,20	58,70	16,60	118,70	65,90	130,90	59,40	171,70	154,80	217,00	195,00
01/03/2005	161,70	175,20	155,80	210,50	313,60	358,90	266,90	311,90	168,40	261,10	227,20	267,50	212,80	341,90
01/04/2005	21,60	32,80	26,30	73,80	74,60	46,00	47,40	62,50	62,70	22,40	12,50	14,10	28,30	22,60
01/05/2005	96,20	14,90	35,80	24,80	19,60	73,10	30,50	41,90	88,00	59,30	18,40	35,20	9,60	41,90
01/06/2005	26,00	7,30	13,20	14,20	8,20	13,00	1,60	9,20	5,80	7,60	1,20	11,60	1,60	27,10
01/07/2005	13,20	8,00	0,00	0,00	0,00	0,00	6,60	0,00	1,70	1,30	15,00	2,30	1,10	3,00
01/08/2005	14,00	15,80	12,80	9,60	16,50	10,80	2,50	0,70	1,10	15,80	0,00	2,50	2,50	2,40
01/09/2005	94,40	80,80	56,80	32,00	51,20	41,20	28,80	99,70	76,70	53,40	18,20	20,70	48,80	128,40
01/10/2005	130,50	26,50	74,20	49,60	70,50	72,00	25,90	61,70	85,00	33,80	63,90	14,90	60,20	42,80
01/11/2005	103,10	312,10	325,70	294,00	326,20	439,70	226,40	229,40	338,50	327,90	307,60	336,50	399,40	358,10

Data	Código da Estação Pluviométrica													
	1943002	1943004	1943006	1943010	1943023	1943025	1943027	2043009	2043010	2044003	2043014	2044006	2044007	2044009
01/01/2000	189,60	313,10	366,30	329,90	338,00	365,90	431,30	330,70	473,30	337,90	367,70	351,30	305,50	384,70
01/02/2000	296,40	74,30	363,40	209,10	162,60	226,60	182,80	125,70	150,40	140,60	162,00	144,60	141,30	143,20
01/03/2000	259,70	183,20	157,80	128,30	186,30	327,60	189,60	185,00	192,00	202,40	158,00	213,80	164,00	151,50
01/04/2000	39,20	1,00	37,90	13,00	21,20	67,30	56,50	27,10	19,00	25,30	11,20	39,70	19,50	11,50
01/05/2000	2,80	8,30	10,10	8,90	1,30	10,60	2,00	5,00	21,50	5,80	0,00	8,30	0,30	0,10
01/06/2000	2,50	0,00	0,00	0,00	0,00	0,10	0,00	0,00	18,20	15,00	14,20	0,00	6,20	0,60
01/07/2000	15,10	10,40	63,80	2,70	7,50	20,60	7,60	13,20	9,90	9,00	10,70	23,40	2,00	10,30
01/08/2000	19,30	33,70	33,60	32,10	31,60	30,50	24,40	31,70	36,20	26,80	22,20	17,40	25,10	14,50
01/09/2000	33,50	64,60	68,20	81,80	52,40	72,50	77,00	109,10	97,00	48,10	105,60	66,50	112,10	68,60
01/10/2000	17,50	18,00	79,10	49,20	20,50	56,90	46,30	49,30	103,80	86,10	78,80	33,20	30,50	51,30
01/11/2000	340,20	192,90	289,70	243,00	273,50	398,20	367,20	275,50	265,90	202,40	214,60	236,40	179,50	236,60
01/12/2000	232,00	203,00	297,30	262,00	299,10	322,60	319,60	246,80	214,50	136,90	211,30	233,30	336,60	243,30
01/01/2001	212,40	132,90	193,10	142,50	169,20	182,70	147,20	132,90	122,10	213,70	302,00	184,30	199,80	263,20
01/02/2001	25,30	48,70	25,50	48,20	109,50	47,60	22,90	11,90	86,80	58,40	29,70	135,10	65,10	28,70
01/03/2001	189,90	82,60	117,50	142,90	104,50	112,10	154,90	123,40	171,90	177,20	107,90	129,30	103,50	134,00
01/04/2001	56,20	1,50	40,00	4,00	19,00	55,50	16,90	39,00	9,20	31,10	64,60	23,90	13,30	31,90
01/05/2001	23,50	71,60	26,90	39,90	18,50	30,50	25,10	30,80	59,90	34,90	51,10	46,90	72,40	70,40
01/06/2001	1,20	2,50	0,00	1,00	0,00	5,50	3,90	0,30	0,20	1,10	0,00	0,00	0,00	3,30
01/07/2001	0,80	1,30	5,60	14,40	0,00	0,50	0,80	0,80	0,80	3,60	3,20	4,10	7,80	7,20
01/08/2001	42,90	31,00	29,50	27,10	2,20	34,70	15,00	33,30	12,40	11,60	23,40	32,80	23,50	29,30
01/09/2001	19,90	51,20	72,60	67,60	30,70	56,10	45,00	66,70	47,60	27,10	72,70	35,00	75,50	44,60
01/10/2001	92,10	119,90	115,90	140,20	97,90	114,50	118,20	116,90	113,20	14,10	156,80	66,80	125,10	107,40
01/11/2001	215,20	194,80	278,90	328,90	258,30	341,20	308,90	266,80	266,50	275,80	368,00	130,10	199,00	170,10
01/12/2001	360,60	377,30	409,50	458,50	403,10	439,70	399,30	281,50	312,50	372,80	268,30	406,40	227,10	453,90
01/01/2002	281,40	154,60	398,40	328,90	313,20	407,50	463,30	459,80	347,40	225,70	295,90	338,50	263,70	170,20
01/02/2002	290,60	258,70	320,40	235,70	206,80	368,80	241,80	244,20	341,60	151,30	236,10	253,60	260,60	305,70
01/03/2002	35,00	30,60	96,50	64,10	60,20	48,80	46,80	111,40	77,90	69,20	75,30	122,40	106,00	67,60
01/04/2002	48,70	0,00	44,70	28,30	14,60	82,50	20,60	21,50	29,00	8,40	30,80	34,60	8,20	15,30
01/05/2002	10,70	23,80	28,30	34,70	27,00	24,70	14,30	15,60	18,80	4,30	34,50	13,60	10,10	36,60

01/06/2002	1,10	0,00	0,00	0,00	0,00	2,10	0,00	0,00	0,00	0,50	0,00	0,00	0,00	2,80
01/07/2002	8,90	27,50	6,30	3,80	11,40	4,40	29,40	0,20	9,00	3,10	5,00	6,50	2,70	8,90
01/08/2002	2,00	8,70	1,40	0,50	0,00	7,30	1,40	7,30	7,90	0,30	12,40	17,00	12,40	4,10
01/09/2002	182,50	46,50	87,40	76,20	91,80	98,80	109,30	124,50	124,80	76,70	93,50	87,60	85,60	56,50
01/10/2002	75,60	28,80	32,90	64,90	9,00	45,20	37,30	18,40	84,10	74,70	64,30	43,30	45,50	80,00
01/11/2002	300,20	268,10	185,40	201,30	243,20	286,70	233,90	240,50	244,80	204,10	311,20	152,60	359,80	302,20
01/12/2002	275,10	444,40	460,70	373,40	391,10	386,10	337,20	276,10	247,30	291,10	247,10	318,60	363,80	245,10
01/01/2003	450,90	662,90	741,00	817,30	540,90	619,50	653,10	393,20	471,60	450,50	394,20	409,10	339,00	477,50
01/02/2003	18,90	27,00	124,50	51,00	100,60	12,60	54,10	32,40	80,30	75,50	51,30	114,90	76,80	53,20
01/03/2003	75,60	51,50	157,60	166,80	195,10	209,60	107,90	161,20	124,70	148,80	111,00	170,60	165,00	235,20
01/04/2003	22,70	4,40	16,60	8,60	7,00	8,20	23,50	42,40	37,20	44,50	27,30	22,10	32,60	23,10
01/05/2003	13,20	17,10	18,00	20,30	18,50	48,70	17,20	12,00	26,20	33,50	17,60	23,70	25,90	33,30
01/06/2003	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	5,60	0,00	0,30	0,00	0,10	2,90
01/07/2003	0,00	0,00	0,70	0,70	0,20	2,70	4,30	3,90	11,40	5,30	6,80	3,50	38,70	5,00
01/08/2003	35,10	10,80	4,70	6,70	5,00	15,70	24,00	57,70	64,70	12,90	67,60	18,40	11,90	11,20
01/09/2003	13,90	27,40	18,40	14,30	17,80	40,70	32,00	51,00	45,90	31,00	52,10	25,10	25,50	15,20
01/10/2003	60,50	8,20	33,70	41,40	38,50	61,70	67,50	41,80	37,70	30,20	31,00	41,60	55,40	43,50
01/11/2003	163,60	175,60	219,90	239,40	234,50	182,20	261,20	164,20	156,10	302,30	183,60	200,30	180,70	222,30
01/12/2003	348,50	325,90	250,70	196,50	208,10	316,50	274,00	275,30	369,30	226,80	290,70	245,70	275,40	300,50
01/01/2004	319,10	357,50	505,10	532,60	353,80	431,70	503,90	413,00	310,50	313,00	420,00	339,30	288,90	248,70
01/02/2004	283,00	256,10	306,10	412,30	292,20	230,00	318,10	256,90	358,40	423,50	429,60	393,40	368,40	312,30
01/03/2004	150,90	117,90	232,60	116,90	106,90	158,00	296,40	86,20	149,10	60,00	194,30	79,50	101,00	61,00
01/04/2004	314,40	100,00	147,40	116,00	62,40	460,80	174,90	343,30	116,30	73,20	63,40	73,10	71,00	58,00
01/05/2004	6,70	33,30	30,50	26,10	30,30	8,60	43,80	47,10	31,30	32,60	36,00	38,40	44,40	53,80
01/06/2004	5,00	2,30	21,80	7,20	6,20	9,80	4,80	22,80	49,20	21,10	28,20	17,30	34,60	36,40
01/07/2004	61,40	50,80	55,50	42,70	29,10	50,60	34,50	33,80	36,80	27,40	37,30	24,20	48,90	46,80
01/08/2004	0,50	0,00	0,00	0,00	0,00	0,00	0,00	0,20	0,30	0,60	0,00	0,00	0,00	2,30
01/09/2004	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,60	0,00	0,00	0,00	0,00	0,20	23,80
01/10/2004	37,10	8,10	58,10	42,30	24,10	89,10	56,10	109,50	81,20	68,30	65,30	86,80	69,90	74,00
01/11/2004	107,20	115,40	108,50	123,20	137,90	173,90	176,70	121,60	172,80	99,00	180,20	89,40	178,20	122,90
01/12/2004	409,40	367,40	400,50	410,10	363,00	432,50	343,10	293,30	310,00	377,80	359,40	323,00	283,30	456,40

01/01/2005	353,30	185,60	223,50	242,20	151,00	312,10	193,60	242,00	222,50	365,30	251,50	291,70	478,20	376,90
01/02/2005	282,80	261,10	192,00	183,70	144,20	207,80	171,40	192,10	195,30	84,20	234,70	130,50	287,20	247,60
01/03/2005	533,60	290,90	345,10	336,30	367,60	379,00	385,40	252,40	423,30	230,90	269,00	205,10	177,60	280,10
01/04/2005	86,30	38,90	31,80	51,80	27,90	95,30	24,30	52,40	138,90	23,60	121,80	28,50	67,80	68,30
01/05/2005	17,40	34,30	35,20	46,40	19,00	19,90	39,90	46,20	49,20	35,60	58,30	88,70	84,70	85,10
01/06/2005	9,20	3,60	15,70	13,60	14,20	17,80	17,10	21,60	19,50	16,30	34,60	24,50	16,30	20,10
01/07/2005	3,10	1,00	3,30	0,00	3,10	4,70	5,00	11,40	22,40	0,00	13,80	0,20	10,80	11,60
01/08/2005	23,90	4,20	9,30	4,10	5,60	3,80	5,50	16,50	0,20	7,80	17,10	3,30	8,10	3,50
01/09/2005	40,20	105,80	132,00	137,30	102,90	86,00	139,60	99,40	109,00	66,80	82,50	47,80	131,20	113,40
01/10/2005	34,00	22,90	22,70	3,80	7,60	10,00	24,30	80,10	60,60	107,20	62,60	55,40	83,20	81,20
01/11/2005	300,80	279,50	272,50	261,10	248,50	433,00	333,50	224,10	136,90	104,30	193,90	176,00	271,90	265,40

Data	Código da Estação Pluviométrica													
	2045001	2045002	2045005	2142000	2142002	2143000	2143016	2243202	2244035	2041005	2042011	1842004	1842005	1942002
01/01/2000	300,10	356,90	373,60	283,40	260,80	354,10	332,80	363,00	315,90	169,30	260,00	102,90	150,20	280,10
01/02/2000	158,00	210,10	227,90	179,60	143,60	196,90	39,90	166,20	197,40	136,60	149,10	149,30	201,70	108,70
01/03/2000	261,10	291,20	193,50	248,00	166,90	215,50	153,50	266,30	172,50	227,50	235,80	145,50	151,00	213,10
01/04/2000	17,90	65,50	36,70	10,20	62,20	43,60	6,00	16,30	55,40	69,00	24,10	32,60	26,00	70,20
01/05/2000	7,80	0,00	0,00	5,00	34,20	0,70	11,20	8,10	0,00	25,40	8,60	16,10	17,10	6,50
01/06/2000	0,00	0,00	0,00	0,00	0,00	0,00	0,30	0,10	0,00	18,30	0,00	3,10	16,20	0,00
01/07/2000	1,30	4,60	3,60	18,60	7,30	9,50	19,20	11,40	68,00	13,30	1,10	8,90	17,50	1,60
01/08/2000	3,30	4,70	11,00	35,60	32,20	30,70	18,90	46,30	24,50	42,00	29,90	26,60	36,20	42,50
01/09/2000	85,20	60,60	72,40	91,20	77,40	110,20	175,10	127,80	137,80	82,00	64,70	29,50	51,60	27,20
01/10/2000	46,80	28,20	38,50	54,20	32,40	41,70	35,00	103,90	88,70	113,10	88,60	52,60	27,50	38,20
01/11/2000	229,90	224,50	187,10	236,20	434,80	162,60	226,10	194,20	208,70	125,80	295,00	357,20	317,30	194,80
01/12/2000	234,10	189,20	251,00	217,00	493,80	206,20	270,60	367,50	329,70	327,40	321,00	268,10	280,00	232,50
01/01/2001	112,00	125,90	208,50	99,80	259,80	221,20	217,70	209,30	223,40	252,00	221,40	36,10	31,60	163,10
01/02/2001	61,10	46,50	75,40	76,80	248,60	52,60	145,60	212,10	145,20	172,90	21,60	34,10	9,70	31,90
01/03/2001	182,10	148,70	150,90	98,60	103,90	142,40	234,80	307,70	305,40	51,60	65,40	47,10	154,90	63,80
01/04/2001	6,90	4,20	2,00	42,00	37,40	17,60	0,00	40,40	6,60	89,50	39,40	3,50	4,10	65,70
01/05/2001	69,10	53,40	62,00	60,40	50,70	98,20	45,10	49,90	39,80	109,80	28,10	26,30	29,80	22,40
01/06/2001	0,00	0,00	0,00	0,00	2,90	0,00	0,50	0,00	0,00	20,20	5,00	23,00	8,10	1,60
01/07/2001	6,60	23,50	11,60	0,00	27,70	0,90	7,20	17,60	8,70	6,20	2,80	2,60	9,80	2,00
01/08/2001	40,20	52,70	58,40	3,00	3,20	8,00	6,20	106,50	66,90	10,90	23,40	29,20	22,40	11,90
01/09/2001	36,70	41,20	72,40	67,00	184,00	56,30	51,50	62,70	31,30	84,90	85,00	25,00	74,60	43,00
01/10/2001	149,50	111,90	75,70	97,60	97,30	86,20	70,40	163,40	121,00	111,70	110,20	105,90	127,80	74,30
01/11/2001	138,50	199,90	225,30	378,00	539,00	392,40	220,50	247,10	171,20	346,50	225,80	346,00	383,90	295,80
01/12/2001	448,50	379,90	361,80	340,00	770,40	302,90	387,80	264,50	261,60	203,20	222,00	216,60	311,90	311,60
01/01/2002	150,00	179,30	152,00	164,80	394,50	295,10	453,40	413,70	306,90	190,30	461,50	310,70	270,90	350,40
01/02/2002	257,00	257,80	234,10	255,00	604,70	330,50	292,80	150,30	221,70	213,10	278,50	60,80	261,00	267,40
01/03/2002	67,50	39,20	149,60	99,40	267,00	79,80	67,90	215,60	235,20	56,10	73,00	52,80	64,00	77,00
01/04/2002	12,10	44,00	1,00	3,40	0,00	22,60	0,60	0,00	0,00	18,50	4,10	56,10	62,20	34,10
01/05/2002	54,10	71,90	79,60	29,40	101,00	45,60	73,50	98,90	21,30	44,50	51,90	49,70	61,30	23,90

01/06/2002	0,00	0,00	0,00	0,00	4,20	0,00	0,00	0,00	0,00	7,80	2,10	1,00	8,40	1,00
01/07/2002	26,50	17,00	16,50	2,40	24,80	14,20	2,00	17,70	39,60	31,60	0,50	12,40	14,40	3,80
01/08/2002	0,90	5,30	14,00	11,20	17,00	13,60	14,10	42,90	46,20	0,40	1,50	3,90	23,20	2,60
01/09/2002	65,00	98,00	27,50	69,90	176,90	49,70	94,90	97,60	72,10	128,10	109,90	96,70	101,40	112,30
01/10/2002	30,50	43,30	55,30	60,00	163,10	81,50	49,20	79,40	91,30	102,70	37,10	7,30	13,70	96,60
01/11/2002	193,50	318,00	260,10	209,80	574,60	208,10	94,00	247,10	194,50	200,70	233,40	172,60	108,30	431,50
01/12/2002	493,00	291,70	314,20	341,40	573,90	308,00	256,40	380,60	427,50	306,70	271,50	332,70	377,50	259,60
01/01/2003	291,30	444,40	610,00	420,30	435,70	467,80	486,20	571,50	567,90	522,20	369,60	268,00	278,40	526,50
01/02/2003	99,90	65,40	39,60	32,80	21,00	44,10	51,00	61,30	52,70	14,70	7,10	110,10	44,20	42,10
01/03/2003	132,20	163,40	232,00	210,30	141,50	143,60	363,50	276,30	309,60	86,60	82,00	65,10	124,80	81,00
01/04/2003	17,60	48,40	27,10	61,40	110,80	35,90	51,80	26,20	17,60	24,00	31,70	28,40	26,60	52,50
01/05/2003	42,40	33,20	6,20	25,80	81,00	40,80	20,50	26,30	27,10	7,80	4,00	20,80	14,30	3,00
01/06/2003	0,00	0,00	0,00	0,00	0,00	2,10	10,50	8,80	56,60	0,00	0,00	1,20	0,00	0,00
01/07/2003	13,90	0,00	0,10	17,20	12,70	13,30	11,80	44,20	21,10	6,10	0,20	3,10	5,60	0,30
01/08/2003	30,70	16,60	9,70	63,40	179,20	38,80	102,70	82,20	54,40	78,50	27,60	10,30	11,80	12,10
01/09/2003	123,20	33,80	15,70	50,80	146,90	44,40	29,00	79,60	16,60	41,70	21,40	22,70	18,20	22,80
01/10/2003	38,00	30,90	58,00	80,90	87,30	63,00	227,90	173,10	208,00	77,70	58,80	85,90	121,60	62,90
01/11/2003	115,60	165,70	154,60	244,30	107,50	183,50	254,80	369,70	266,90	278,90	261,50	84,50	99,20	147,20
01/12/2003	261,80	348,70	275,20	205,30	594,40	239,60	276,20	342,40	202,20	241,00	264,70	168,70	301,30	279,20
01/01/2004	248,80	245,10	273,30	287,80	586,90	198,20	297,40	379,80	292,90	237,50	340,20	188,70	196,60	246,20
01/02/2004	584,40	277,90	269,90	317,90	199,80	286,10	436,20	502,20	414,70	278,70	199,10	238,60	267,50	292,90
01/03/2004	206,80	169,10	191,10	239,60	181,40	262,00	178,10	301,40	199,90	238,40	228,30	264,80	262,80	128,60
01/04/2004	31,40	70,00	86,90	161,60	134,60	159,80	203,70	245,70	217,80	103,60	124,90	77,20	194,60	164,60
01/05/2004	6,40	55,00	37,90	63,40	45,90	65,40	50,20	50,80	50,20	21,90	27,80	4,80	33,80	13,70
01/06/2004	46,40	24,70	36,90	26,20	90,90	35,30	15,60	38,50	50,20	12,90	36,50	4,10	14,10	44,20
01/07/2004	45,20	46,30	47,60	25,70	176,20	25,10	62,70	110,50	61,90	85,20	45,00	52,30	38,60	37,70
01/08/2004	0,00	0,00	0,00	1,30	15,50	0,00	0,00	6,50	0,00	25,30	2,80	0,00	1,40	1,50
01/09/2004	4,80	2,70	11,50	0,00	0,50	0,00	5,70	42,90	27,70	0,00	0,00	0,70	0,00	0,00
01/10/2004	119,90	81,80	65,00	64,80	161,30	127,90	193,20	288,10	157,00	103,80	42,70	43,60	78,10	104,30
01/11/2004	211,90	186,00	157,80	152,60	103,70	283,70	294,30	329,80	208,20	241,10	71,30	68,00	211,50	90,60
01/12/2004	458,70	364,90	516,90	577,50	498,50	465,50	442,10	399,90	316,10	409,10	423,50	280,30	307,00	476,50

01/01/2005	302,10	411,00	472,80	285,30	206,80	226,40	241,60	302,60	328,90	229,80	298,80	235,20	333,40	223,90
01/02/2005	77,50	25,40	147,90	159,10	344,20	114,00	168,50	248,60	253,30	301,40	233,40	371,80	457,90	304,20
01/03/2005	147,90	206,20	281,80	220,10	292,70	245,10	205,60	263,00	260,70	463,20	368,60	235,40	346,70	247,30
01/04/2005	31,40	33,50	69,30	46,50	90,30	69,20	45,80	250,60	211,70	95,90	41,10	25,40	14,60	43,90
01/05/2005	118,00	83,60	75,20	46,50	43,40	41,10	30,20	110,40	87,60	60,40	59,60	18,20	48,50	32,70
01/06/2005	14,20	36,50	25,50	35,60	43,80	18,20	3,60	39,80	27,70	51,20	22,60	19,50	48,40	29,00
01/07/2005	9,20	4,30	6,80	17,40	45,20	19,80	40,90	36,70	57,00	25,80	2,90	3,50	27,00	0,30
01/08/2005	0,00	1,00	0,00	14,90	0,30	36,40	20,00	0,00	2,20	10,70	0,00	11,90	19,30	0,00
01/09/2005	55,80	45,60	44,10	39,10	88,80	73,80	65,80	108,20	77,00	102,10	152,70	27,60	41,50	111,70
01/10/2005	161,20	155,70	170,80	56,70	58,70	62,50	59,90	37,60	32,60	130,70	41,10	51,70	51,60	74,80
01/11/2005	156,80	388,20	149,40	207,40	289,10	297,70	187,30	306,30	196,20	269,40	174,50	342,60	313,50	357,00

Data	Código da Estação Pluviométrica				
	2042008				
01/01/2000	203,1	01/05/2002	11,8	01/09/2004	0
01/02/2000	135,3	01/06/2002	0,2	01/10/2004	95,8
01/03/2000	257,5	01/07/2002	1	01/11/2004	162,7
01/04/2000	16,1	01/08/2002	2,3	01/12/2004	470,2
01/05/2000	32,4	01/09/2002	93,5	01/01/2005	269,7
01/06/2000	0	01/10/2002	35,7	01/02/2005	186
01/07/2000	0,4	01/11/2002	259	01/03/2005	401,7
01/08/2000	31,7	01/12/2002	277,4	01/04/2005	40,9
01/09/2000	44,5	01/01/2003	517,8	01/05/2005	52,5
01/10/2000	101,7	01/02/2003	11,1	01/06/2005	42,1
01/11/2000	201,7	01/03/2003	67	01/07/2005	18,2
01/12/2000	362,6	01/04/2003	72,2	01/08/2005	47
01/01/2001	311,3	01/05/2003	4,7	01/09/2005	131,6
01/02/2001	43,9	01/06/2003	0	01/10/2005	22,5
01/03/2001	117,5	01/07/2003	0	01/11/2005	269,4
01/04/2001	20,4	01/08/2003	20,2		
01/05/2001	32,2	01/09/2003	40,3		
01/06/2001	7,9	01/10/2003	60,8		
01/07/2001	3,7	01/11/2003	207,3		
01/08/2001	10,7	01/12/2003	130,6		
01/09/2001	56,3	01/01/2004	352,6		
01/10/2001	118,7	01/02/2004	351,6		
01/11/2001	287,8	01/03/2004	290,6		
01/12/2001	309	01/04/2004	183,6		
01/01/2002	328,6	01/05/2004	13,2		
01/02/2002	284,3	01/06/2004	22		
01/03/2002	48,6	01/07/2004	24		
01/04/2002	21,4	01/08/2004	2,2		

ANEXO C - ROTINA DA ANÁLISE DE SÉRIES TEMPORAIS

```
# Leitura da pasta contendo dados e comandos de séries temporais
getwd()
setwd("F:/Dissertacao/Dados e comandos/Comandos_Series_Temporais")
getwd()

# Carregando os pacotes de series temporais
library(tseries)
library(TSA)

#Leitura das séries
dados=read.table("dadoscompletos2005.txt",h=T) #leitura das series de precipitação por dias
chuvosos
A1=dados$A1#selecionando a estação de Campina Verde
A1=ts(A1,frequency=12,start=c(2000,1),end=c(2005,11))#Estabelecendo o período de tempo
da série

#Gerando o gráfico da série
plot.ts(A1,main="",xlab="Meses    transcorridos",ylab="Prec.Média    Mensal    por    Dias
Chuvosos",col="red")

#Funções de autocorrelação
par(mfrow=c(1,2));m=acf(A1,lag.max=36,plot=F);m$lag=m$lag*12
plot(m,main="",xlab="Defasagem",ylab="FAC")
m=pacf(A1,lag.max=36,plot=F);m$lag=m$lag*12;plot(m,main="",xlab="Defasagem",ylab="FAC
P")

#Periodograma
n<-length(A1)
f = 0:(n/2)/n
per=periodogram(A1)
plot(f, per$spec, type="h", xlab="Frequência", ylab="Periodograma")
abline(v=seq(0,.5,.02), lty="dotted")

#ajuste de regressão
x<-A1-mean(A1) # Retirar a média dos dados
# Cálculo do periodograma
n<-length(x);t<-1:n
p<-abs(2*fft(x)/n)^2
p<-p[1:(n/2+1)]
f <- 0:(n/2)/n

per.1<-cbind(f,p);
max.1<-as.matrix(per.1)[order(p,decreasing=T)[1:12],] # cálculo de frequências significativas.
#Neste caso, só estão sendo observadas as 12 primeiras.
max.1

#Componentes de senos e cossenos. Deve-se calcular para cada frequência significativa e
testar quais componentes são diferentes de zero.
cos1<-cos(2*pi*max.1[1]*t); sin1<-sin(2*pi*max.1[1]*t);
cos2<-cos(2*pi*max.1[2]*t); sin2<-sin(2*pi*max.1[2]*t);
cos3<-cos(2*pi*max.1[3]*t); sin3<-sin(2*pi*max.1[3]*t);
cos4<-cos(2*pi*max.1[4]*t); sin4<-sin(2*pi*max.1[4]*t);
cos5<-cos(2*pi*max.1[5]*t); sin5<-sin(2*pi*max.1[5]*t);
cos6<-cos(2*pi*max.1[6]*t); sin6<-sin(2*pi*max.1[6]*t);
cos7<-cos(2*pi*max.1[7]*t); sin7<-sin(2*pi*max.1[7]*t);
cos8<-cos(2*pi*max.1[8]*t); sin8<-sin(2*pi*max.1[8]*t);
```

```

cos9<-cos(2*pi*max.1[9]*t); sin9<-sin(2*pi*max.1[9]*t);
cos10<-cos(2*pi*max.1[10]*t); sin10<-sin(2*pi*max.1[10]*t)

# Ajuste da tendência através de um modelo de regressão. Neste caso só inclui as
componentes com as 10 primeiras frequências.
# Observação: retirar a constante no modelo ajustado.
fit.trend<-lm(x~cos1+sin1+cos2+sin2+cos3+sin3+cos4+sin4+cos5+sin5-1);
summary(fit.trend)

#série temporal ajustada pelo modelo de regressão estimado
x.ajustado=fit.trend$coeff[1]*cos1+fit.trend$coeff[2]*sin1+
fit.trend$coeff[3]*cos2+fit.trend$coeff[4]*sin2 + ...

#Nova série temporal sem periodicidade
erro<-x-x.ajustado

#Funções de autocorrelação
par(mfrow=c(1,2))
m=acf(erro,lag.max=36,plot=F)
m$lag=m$lag*12
plot(m,main="",xlab="Defasagem",ylab="FAC")
m=pacf(erro,lag.max=36,plot=F)
m$lag=m$lag*12
plot(m,main="",xlab="Defasagem",ylab="FACP")

# Teste de raiz unitária - teste de Dickey-Fuller (ADF)
adf.test(A1)

#Ajuste dos modelos de séries temporais
modelo1=arma(erro,order=c(1,0));modelo2=arma(erro,order=c(2,0))
modelo3=arma(erro,order=c(3,0));modelo4=arma(erro,order=c(4,0))
modelo5=arma(erro,order=c(5,0));modelo6=arma(erro,order=c(6,0))
modelo7=arma(erro,order=c(7,0));modelo8=arma(erro,order=c(0,1))
modelo9=arma(erro,order=c(0,2));modelo10=arma(erro,order=c(0,3))
modelo11=arma(erro,order=c(0,4));modelo12=arma(erro,order=c(0,5))
modelo13=arma(erro,order=c(0,6));modelo14=arma(erro,order=c(0,7))
modelo15=arma(erro,order=c(1,1));modelo16=arma(erro,order=c(1,2))
modelo17=arma(erro,order=c(2,1));modelo18=arma(erro,order=c(2,2))

#Comparando AIC dos modelos de séries temporais
modelo1$aic;modelo2$aic;modelo3$aic;modelo4$aic;modelo5$aic;modelo6$aic;modelo7$aic;m
odelo8$aic
modelo9$aic;modelo10$aic;modelo11$aic;modelo12$aic;modelo13$aic;modelo14$aic;modelo1
5$aic;
modelo16$aic;modelo17$aic;modelo18$aic

#Modelo Final
modelo=arima(erro,order=c(1,0,1))
residuos=modelo$residuals

#Análise residual
qqnorm(residuos);qqline(residuos, col="red")
Box.test(residuos)
shapiro.test(residuos)

#Função de autocorrelação dos resíduos
par(mfrow=c(2,2))
m=acf(residuos,main="ACF erro",lag.max=36,plot=F);m$lag=m$lag*12;
plot(m,main="",xlab="Defasagem",ylab="FAC")
m=pacf(residuos,main="PACF erro",lag.max=36,plot=F) ;m$lag=m$lag*12;
plot(m,main="",xlab="Defasagem",ylab="FACP")

```

ANEXO D - ROTINA DA ANÁLISE DE GEOESTATÍSTICA

```
#Leitura da pasta de trabalho
getwd()
setwd("C:/Users/Acer/Desktop/Dissertacao")
getwd()

#Carregando pacotes necessários
library(geoR)
require(scatterplot3d)

# Leitura dos dados
dados1 <- read.geodata("jan2000.txt", header=T)
dados1
summary(dados1)

# Análise explanatória

attach(dados1)
par(mfrow=c(2,2))
points(dados1,xlab="Coordenada X (m)",ylab="Coordenada Y (m)", pt.divide="equal")
points(dados1,xlab="Coordenada X (m)",ylab="Coordenada Y (m)",
pt.divide="data.proportional")
points(dados1,xlab="Coordenada X (m)",ylab="Coordenada Y (m)", pt.divide="quartiles")
points(dados1,xlab="Coordenada X (m)",ylab="Coordenada Y (m)", pt.divide="deciles")

plot(dados1) # sem linha de tendência
plot(dados1, low=T) # com linha de tendência
plot(dados1, scatter=T, low=T) # visão 3D das realizações

# Semivariograma empírico
vario.emp.1 <- variog(dados1,direction="omnidirectional", max.dist=7)
plot(vario.emp.1,ylim=c(0,40),xlab="Distâncias",ylab="Semivariâncias")

# Ajuste do modelo teórico
plot(vario.emp.1,main="Variograma Empírico",ylim=c(0,40))
eyefit(vario.emp.1)

# Estimção: OLS e ML
# Obs.: Dentro do argumento "ini" deve-se colocar primeiro a "contribuição" e o "alcance
teórico".
ols.1 <- variofit(vario.emp.1, ini=c(25,2), fix.nugget=T, nug=5, wei="equal",cov.model="sph")
ml.1 <- likfit(dados1, ini=c(25,2), fix.nugget=T, nug=5, lik.method="ML",cov.model="sph")

par(mfrow=c(1,2))
plot(vario.emp.1, ylim=c(0,40), xlab="Distâncias", ylab=expression(Semivariâncias),
main="Variograma Empírico (MQO)")
lines(ols.1, max.dist=7, col="red", lty=4, lwd=2)
plot(vario.emp.1, ylim=c(0,40), xlab="Distâncias (m)", ylab=expression(Semivariâncias),
main="Variograma Ajustado (MV)")
lines(ml.1, max.dist=7, col="blue", lty=4, lwd=2)

plot(vario.emp.1, ylim=c(0,40), xlab="Distâncias", ylab=expression(Semivariâncias),
main="Variograma Empírico")
lines(ols.1, max.dist=7, col="red", lty=4, lwd=2)
lines(ml.1, max.dist=7, col="blue", lty=4, lwd=2)

summary(ols.1)
summary(ml.1)
```

```

# Autovalidação
autoval.1 <- xvalid(dados1, model=ols.1)
autoval.2 <- xvalid(dados1, model=ml.1)

summary(autoval.1)
summary(autoval.2)

names(autoval.1); names(autoval.2)
(reg.val.1 = lm(autoval.1$data~autoval.1$predicted)); (reg.val.2 =
lm(autoval.2$data~autoval.2$predicted))

#Gráfico dos ajustes
par(mfrow=c(1,2))
plot(autoval.1$predicted, autoval.1$data, xlab="Preditos", ylab="Observados",
main="Regressão de Autovalidação (MQO)")
abline(reg.val.1)
abline(coef=c(0,1), col="gray60", lty=3)
plot(autoval.2$predicted, autoval.2$data, xlab="Preditos", ylab="Observados",
main="Regressão de Autovalidação (MV)")
abline(reg.val.2)
abline(coef=c(0,1), col="gray60", lty=3)

#Obtendo os resíduos
res=reg.val.1$residuals
res

#Verificando os resíduos
res <- read.geodata("resJan2000.txt", header=T)

# Semivariograma empírico dos resíduos
vario.emp.2 <- variog(res,direction="omnidirectional", max.dist=7)
plot(vario.emp.2,ylim=c(0,15),xlab="Distâncias",ylab="Semivariâncias")

```