



Ministério da
**Ciência, Tecnologia
e Inovação**



VALIDAÇÃO DE UMA KRIGEAGEM ORDINÁRIA POR MEIO DO PRODUTO TRMM 3B43 DE PRECIPITAÇÃO MENSAL – UM ESTUDO DE CASO DO BIOMA CERRADO

Alana Carla Toniol

Monografia para disciplina de Introdução
ao Geoprocessamento do programa de Pós
Graduação em Sensoriamento Remoto.

INPE
São José dos Campos
2015

RESUMO

O Bioma Cerrado é marcado pela alta sazonalidade entre a estação seca e chuvosa, esta ocorrendo entre os meses de outubro a março, de forma irregular. A precipitação é uma variável que pode ser espacializada em uma determinada região a partir de um conjunto de pontos amostrais coletados *in situ*, bem como pode ser avaliada sinopticamente por meio de produtos de imagens de satélite. Contudo, há uma carência de uma validação estatisticamente consistente de uma modelagem geoestatística para o Bioma Cerrado. Para suprir essa necessidade, este trabalho teve como objetivo geral validar uma Krigeagem ordinária por meio do produto de sensoriamento remoto orbital TRMM-3B43 de precipitação mensal. Como objetivos específicos, comparou-se duas krigeagens ordinárias (no software SPRING, e no ArcGIS, otimizada), e os erros associados à correlação existente entre os dados coletados *in situ* e os dados orbitais. Os resultados obtidos neste trabalho foram satisfatórios, os quais demonstraram uma alta correlação entre o produto TRMM e a Krigeagem otimizada, em que foi apresentado um p-valor $7,1e-114$, um coeficiente de determinação (R^2) de 0,64 e um NRMSE de 0,09. A análise de regressão linear entre os dados *in situ* e os dados orbitais também apresentou-se satisfatória, sendo o R^2 de 0,55, p-valor $1,75e-22$ e NRMSE de 0,07. Os altos valores de incerteza associados ao mapa de variância gerado pelo SPRING estiveram relacionados, principalmente, à esparsa distribuição dos pontos amostrais na região norte do Bioma. Contudo, apenas 20% dos pixels do Bioma encontraram-se no range de maior incerteza. Pelo mapa de diferença (anomalia) entre a imagem TRMM e a Krigeagem otimizada, constatou-se que aproximadamente 75% dos pixels amostrados apresentaram valores de diferença em torno de -50 e 50 mm/mês entre os percentis 12,5 e 87,5%. Pode-se concluir que o produto TRMM mostrou-se um bom preditor da precipitação no Bioma cerrado, possibilitando a validação dos dados de modelagem geoestatística para essa região. Além disso, os erros inerentes aos dados *in situ* e ao produto TRMM influenciaram na calibração e validação da Krigeagem ordinária, porém, todos mantiveram-se dentro do limiar de significância.

LISTA DE FIGURAS

Figura 1. Parâmetros de semivariograma	13
Figura 2. Distribuição Normal Padrão dos dados de precipitação no Cerrado.....	18
Figura 3. Coeficiente de assimetria, média e mediana caracterizando a distribuição simétrica dos dados.....	19
Figura 4. Ajuste do semivariograma empírico por meio de um modelo teórico esférico.	20
Figura 5. Obtenção do alcance e dos valores de Efeito Pepita e Akaike para o modelo teórico ajustado.....	20
Figura 6. Krigeagem gerada por meio do semivariograma empírico.....	21
Figura 7. Gráfico de dispersão representando a correlação entre a Krigeagem obtida por meio do ajuste do semivariograma empírico (Spring) e a Krigeagem obtida por meio do ajuste otimizado (Arcgis).....	22
Figura 8. Mapa de incerteza associado à modelagem estocástica. Os círculos pretos representam as áreas de maior variância na área de estudo.....	23
Figura 9. Mapa de diferença entre a imagem TRMM e a Krigeagem otimizada.....	24
Figura 10. Representação do histograma indicando que a distribuição da maior parte das amostras encontra-se no segundo e terceiro quartil.....	25
Figura 11. Análise de regressão entre a imagem TRMM e a Krigeagem otimizada.....	25
Figura 12. (a) Produto TRMM recortado para o Brasil. (b) Produto TRMM ao fundo e Krigeagem otimizada recortada para o Bioma Cerrado.	26
Figura 13. Gráfico de dispersão representando a correlação existente entre a imagem TRMM e os dados coletados <i>in situ</i>	27
Figura 14. Boxplot representando a dispersão das observações em cada variável.....	27

LISTA DE TABELAS

Tabela 1. Estimadores de erros (MSE) e (RMSE).....	17
Tabela 2. Percentual de pixels em cada range de incerteza.....	23

SUMÁRIO

1	INTRODUÇÃO.....	6
1.1	Hipótese	8
1.2	Objetivo Geral.....	8
1.3	Objetivos Específicos	8
2	METODOLOGIA.....	8
2.1	Área de Estudo.....	8
2.2	Dados de aquisição	9
2.2.1	Coletas <i>in situ</i>	9
2.2.2	Dados orbitais	10
2.3	Modelagem Geoestatística (Krigagem).....	10
2.4	Mapa de Variância	15
2.5	Álgebra de mapas.....	15
2.6	Análise estatística	16
2.6.1	Regressão Linear	16
2.6.2	Análise do Erro	17
3	RESULTADOS E DISCUSSÃO	18
3.1	Modelagem Geoestatística	18
3.2	Mapa de Variância	22
3.3	Álgebra de Mapas e Processamento Digital de Imagens	24
4	CONCLUSÕES	28

1 INTRODUÇÃO

Considerado o segundo maior bioma brasileiro em termos de extensão, o Cerrado se destaca pela grande diversidade florística. Ribeiro e Walter (2008) caracterizam a vegetação do Cerrado como um mosaico representado por formações florestais, savânicas e campestres (LIMA et al., 2010). O clima no Cerrado é marcado pela alta sazonalidade existente entre os períodos de seca e o período chuvoso exigindo, portanto, uma adaptação dos diversos hábitos da vegetação nos períodos de déficit hídrico na superfície.

Sarmiento (1984) e Oliveira (1998) destacam que o ciclo de vida das espécies arbóreo-arbustivas do cerrado é perene sendo que estas possuem mecanismos que facilitam sua sobrevivência e reprodução durante o período seco (LENZA; KLINK, 2006). Dentre as diversas adaptações pode-se citar o desenvolvimento de um sistema radicular profundo, possibilitando o alcance do lençol freático onde há maior disponibilidade de água. As espécies do estrato lenhoso, por sua vez, podem superar o estresse hídrico sazonal perdendo completamente sua folhagem, por certo tempo, durante o período de seca. (OLIVEIRA, 1998; SARMIENTO et al., 1985; MIRANDA, 1995; MORAIS et al., 1995; FRANCO et al., 2005).

Segundo Marcuzzo e Melo (2012), a distribuição da precipitação média mensal no Cerrado apresenta uma grande estacionalidade, a qual se concentra entre os meses de outubro a março, marcados pela estação chuvosa. Em geral, a precipitação média anual fica entre 1200 e 1800 mm contribuindo para que os lençóis freáticos do Cerrado alimentem as nascentes que originam seis das oito maiores bacias hidrográficas brasileiras (SANTOS et al., 2010).

O estudo da distribuição da precipitação no Bioma apresenta como finalidade a tentativa de regionalização de grandes padrões pluviométricos dos cerrados, bem como o estudo de probabilidade de ocorrência de veranicos (MOREIRA, 1985). Outro aspecto a ser observado é a perda da capacidade produtiva de áreas agrícolas em decorrência da erosão hídrica. Assad et al. (1992), ressalta ainda a importância do conhecimento da chuva máxima diária para aplicações imediatas no dimensionamento de barragens, nos estudos de irrigação e drenagem e no controle de enxurradas.

Uma vez que se tem o conhecimento da distribuição de chuvas numa região, torna-se necessário um planejamento a médio e longo prazo para que haja um melhor aproveitamento do recurso. Para tanto, Bertoni e Tucci (2001) ressaltam a importância de um posto de medição de chuvas, o qual visa obter uma série ininterrupta de precipitações ao longo dos anos.

A Agência Nacional de Águas (ANA) recomenda que a coleta tradicional de dados pluviométricos com a devida consistência de seus valores deve ser realizada a nível diário, com conseqüente reflexo nas séries mensais, anuais e históricas de cada estação. A condição espacial da precipitação sugere sempre a necessidade de analisar os dados de conjuntos de estações de medição pluviométricas próximas para permitir o preenchimento de lacunas nos registros ou a substituição de dados observados e considerados errôneos.

Uma das vantagens da coleta tradicional de dados pluviométricos é a acurácia na obtenção das informações, uma vez que os dados de chuva são coletados *in situ*. No entanto, o processo de mensuração da quantidade de chuva pode ser comprometido por falhas na coleta sistemática dos dados.

A coleta de dados pluviométricos pode ser realizada ainda através de meios alternativos, como por exemplo, fazendo-se o uso de produtos de imagens orbitais. Uma das vantagens de se utilizar esse método consiste no baixo custo de acesso às informações, além do seu caráter sinóptico, isto é, da abrangência do máximo de informações em um único produto. Dentre as limitações para a obtenção de dados por meio de sistemas orbitais pode-se citar os possíveis erros de instrumentação associados.

Como exemplo tem-se o satélite Tropical Rainfall Measuring Mission (TRMM), o qual oferece imagens para acompanhamento, previsão e análise de precipitação. Outra implicação do TRMM consiste na utilização das informações para a veracidade de dados *in situ*, o que poderia suprir a ausência de validações estatisticamente consistente para a modelagem geoestatística da precipitação acumulada mensal para o Bioma Cerrado.

1.1 Hipótese

Visto que os dados obtidos pelo satélite são validados em campo, tomamos o produto TRMM-3B43 como preditor da precipitação acumulada numa modelagem geoestatística para o Bioma Cerrado.

1.2 Objetivo Geral

Esse trabalho objetiva validar, por meio do sensoriamento remoto, medidas de precipitação no Cerrado, coletadas pelo Instituto Nacional de Meteorologia (INMET) e Agência Nacional de Águas (ANA).

1.3 Objetivos Específicos

Dentre os objetivos específicos estão:

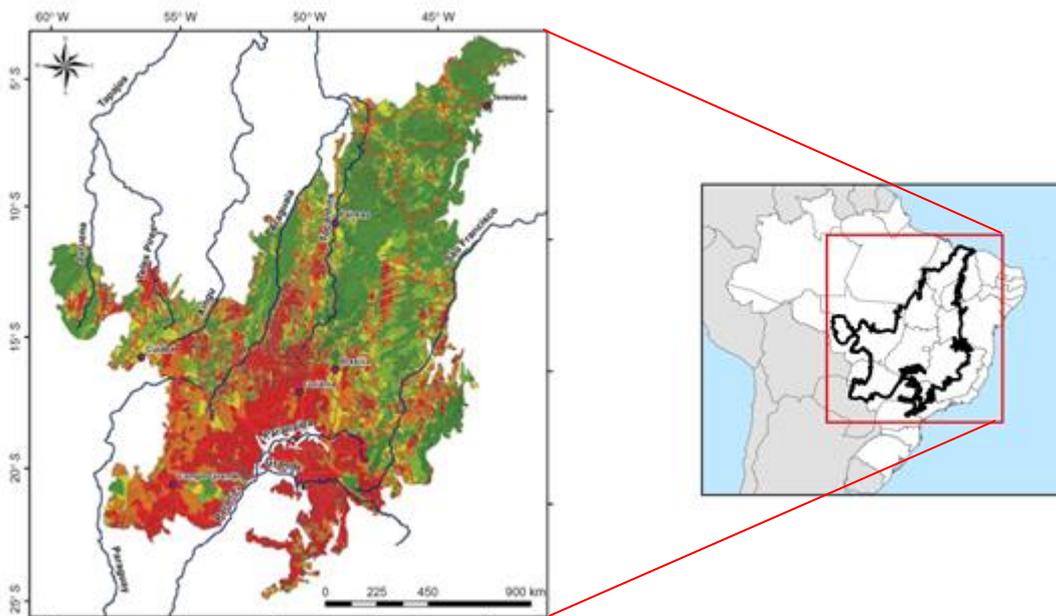
- Comparar duas metodologias geoestatísticas;
- Avaliar o erro existente na correlação entre dados *in situ* e dados orbitais.

2 METODOLOGIA

2.1 Área de Estudo

O Cerrado abrange uma área de quase 208 milhões de hectares, correspondendo a aproximadamente 22% do território nacional. Localizado entre as coordenadas 25° S e 45° W, o Bioma se insere nos estados de Goiás, Tocantins, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Bahia, Maranhão, Piauí, Rondônia, Paraná, São Paulo e Distrito Federal, além dos enclaves no Amapá, Roraima e Amazonas.

No espaço territorial do Cerrado encontram-se as nascentes das três maiores bacias hidrográficas da América do Sul (Amazônica/Tocantins, São Francisco e Prata), o que resulta em um elevado potencial aquífero e favorece a sua biodiversidade. O clima do Cerrado é estacional marcado por um período chuvoso (outubro a março) seguido de um período seco (abril a setembro). A precipitação média anual varia de 1200 a 1800 mm e as temperaturas são geralmente amenas ao longo do ano, entre 22°C e 27°C em média.



Fonte: Ferreira et al., (2011)

2.2 Dados de aquisição

2.2.1 Coletas *in situ*

A ANA e o INMET são os órgãos oficiais responsáveis pela divulgação da quantidade de chuvas medida em mm/mês. Os dados de precipitação são adquiridos por meio de equipamentos específicos, tais como pluviômetros, de maior simplicidade e menor custo; e pluviógrafos, os quais estão presentes em estações meteorológicas, exigindo um alto custo e operação mais complexa. Enquanto nos pluviômetros é lida a lâmina de água acumulada em mm/dia ou mm/chuva, os pluviógrafos medem a quantidade e a duração da precipitação registrando, portanto, a intensidade de chuva.

Uma vez que os dados pluviométricos mensais não seguem, em geral, uma distribuição de probabilidades do tipo normal, adota-se, apenas, a compatibilização da série mensal consolidada com sua correspondente anual. Esta apresenta uma menor variabilidade, constituindo um ótimo referencial para detecção de prováveis inconsistências.

A condição espacial da precipitação sugere sempre a necessidade de analisar os dados de conjuntos de estações de medição pluviométricas próximas para permitir o preenchimento de lacunas nos registros ou a substituição de dados observados e considerados errôneos (ANA, 2012). Portanto, a análise de consistência dos dados de

precipitação deve ter como objetivo a identificação e correção de erros, bem como o preenchimento de falhas das séries pluviométricas.

2.2.2 Dados orbitais

Desenvolvido pela National Aeronautics and Space Administration (NASA) e a Japan's National Space Development Agency (JAXA/NASDA), a plataforma TRMM tem o objetivo de avaliar diversas características da precipitação tanto em regiões tropicais, quanto subtropicais, possibilitando a aquisição de dados de maneira sistemática (VIANA et al., 2010).

O satélite contém cinco instrumentos básicos: o radar de precipitação, o qual fornece mapas tridimensionais da estrutura de uma tempestade; o imageador de micro-ondas, um sensor passivo que permite a quantificação do vapor de água, a água de nuvens e a intensidade da precipitação na atmosfera; o scanner de radiação visível e infravermelha, o qual fornece indicação da chuva e se conecta a outras observações; o sistema de nuvens e energia radiante da Terra, responsável por monitorar a energia trocada entre o Sol, a atmosfera terrestre, a superfície e nuvens, e o espaço; e finalmente, o sensor de imageamento de raios, o qual consiste num instrumento altamente sofisticado capaz de detectar e localizar relâmpagos na região tropical do globo.

Os dados de precipitação em mm/hora foram obtidos no site da NASA, por meio do produto TRMM 3B43, sendo que esse produto apresenta a resolução espacial de 0,25 de grau.

2.3 Modelagem Geoestatística (Krigagem)

Técnicas de Krigagem podem ser entendidas como uma predição linear ou uma forma de inferência bayesiana (KRIGE, 1952). Ela parte do princípio de que os pontos próximos no espaço tendem a ter valores mais parecidos do que os pontos mais afastados. Porém, a partir de determinada distância, certamente não se encontrarão valores próximos porque a correlação espacial pode deixar de existir.

Os passos para o emprego da geoestatística são:

- Análise exploratória dos dados;
- Análise estrutural (cálculo e modelagem do variograma);

- Realização de inferências (Krigagem)

Burrough (1987) demonstrou que a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes:

- Uma componente estrutural, associada a um valor médio constante ou a uma tendência constante;
- Uma componente aleatória, especialmente correlacionada;
- Um ruído aleatório ou erro residual.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z , em x , é dada por:

$$Z(x) = m(x) + e'(x) + e''$$

onde:

$m(x)$ é uma função determinística que descreve a componente estrutural de Z em x

$e'(x)$ é um termo estocástico, que varia localmente e depende espacialmente de $m(x)$

e'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

Diferentemente de outras técnicas de inferência, a Krigagem ordinária requer algumas hipóteses (DAVID, 1977). No caso mais simples, admite-se que a componente determinística $m(x)$, é constante (não há tendências na região). Então, $m(x)$, é igual ao valor esperado da variável aleatória Z na posição x e a diferença média entre os valores observados em x e $(x+h)$ separados por um vetor de distância h (módulo e direção), é nula. Logo, pode-se escrever:

$$E[Z(x) - Z(x + h)] = 0$$

Admite-se também neste contexto a estacionariedade da covariância, isto é, a covariância entre dois pares quaisquer $Z(x)$ e $Z(x + h)$, separados por um vetor distância h , existe, e depende somente de h . Então:

$$C(h) - Cov[Z(x), Z(x + h)] = E[Z(x) * Z(x + h)] - m^2, \forall x$$

A estacionariedade da covariância também implica na estacionariedade do variograma, definido por:

$$2\gamma(h) = E[Z(x) - Z(x + h)]^2$$

As considerações acima resumem a hipótese de estacionariedade de 2ª ordem. Porém, na geoestatística adota-se uma hipótese de estacionariedade menos restritiva denominada Intrínseca. Neste caso, admite-se que $E[Z(x)] = \mu$, $\forall x$, e que a variância das diferenças depende somente do vetor distância h , isto é:

$$\text{Var}[Z(x) - Z(x + h)] = E\{[Z(x) - Z(x + h)]^2\} = 2\gamma(h)$$

Onde $2\gamma(h)$ é o variograma, definido.

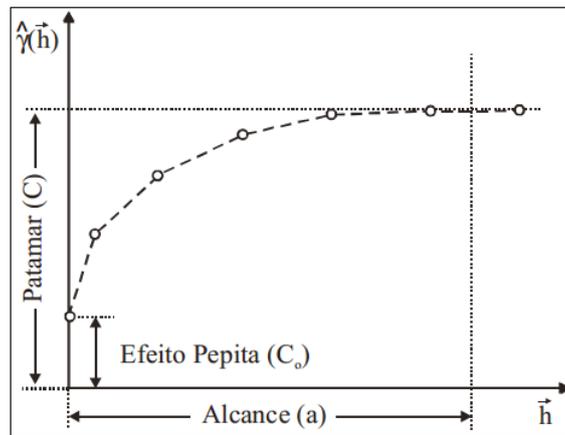
Se as condições especificadas pela hipótese intrínseca são contempladas, então o semivariograma pode ser estimado a partir do dado amostral:

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

Onde $N(h)$ é o número de pares de pontos amostrais separados pelo vetor distância (h).

O gráfico $\gamma(h)$ versus h , conforme apresentado na **Error! Reference source not found.**, é conhecido como o semivariograma experimental. O seu padrão representa o que, intuitivamente, se espera de dados reais, isto é, que as diferenças $\{Z(x_i) - Z(x_i + h)\}$ decresçam à medida que h diminui.

Figura 1. Parâmetros de semivariograma



É esperado que observações, mais próximas geograficamente, tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Desta maneira, é esperado que $\gamma(h)$ aumente com a distância h .

O procedimento de ajuste, do semivariograma experimental, não é direto e automático, mas sim, iterativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Resumidamente os modelos de ajuste estão divididos em dois tipos (ISAACS; SRIVASTAVA, 1989): modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Dentre esses modelos, está o modelo esférico, que foi escolhido para esse estudo. Modelos do segundo tipo, não atingem o patamar, e continuam aumentando enquanto a distância aumenta. Tais modelos são utilizados para modelar fenômenos que possuem capacidade infinita de dispersão.

Quanto aos métodos de inferência de Krigagem ordinária, seu estimador é conforme a equação a seguir (JOURNAL, 1988):

$$Z_{x_0}^* = \sum_{i=1}^n \lambda_i Z(x_i)$$

com $\sum_{i=1}^n \lambda_i = 1$

Journal (1988) mostrou que o sistema de Krigagem ordinária pode ser escrito em notação matricial como:

$$K * \lambda = k$$

onde:

K e k são matrizes de covariâncias (ou variogramas); e λ é o vetor dos pesos.

Determinados os pesos, aplica-se o estimador da equação (8). Além da inferência de valores, esta técnica fornece também uma medida de imprecisão denominada de variância de Krigeagem (JOURNEL, 1988)

O módulo de procedimentos geoestatísticos, integrado ao sistema SPRING, tem como objetivo a análise em duas dimensões de dados espacialmente distribuídos, no que diz respeito à inferência de superfícies geradas a partir de amostras georreferenciadas. Inicialmente sobre o dado de entrada, neste estudo proveniente do PI cadastral, é realizada uma análise exploratória através de estatísticas univariadas e bivariadas.

As estatísticas univariadas fornecem um meio de organizar e sintetizar um conjunto de valores, que se realiza principalmente através do histograma. As estatísticas bivariadas fornecem meios de descrever o relacionamento entre duas variáveis, isto é, entre dois conjuntos de dados ou de duas distribuições. Essa relação pode ser visualizada através do diagrama de dispersão.

Seguindo, uma análise de continuidade espacial baseada no semivariograma é realizada. A análise completa do semivariograma compreende os seguintes passos:

1. Levantamento do semivariograma experimental;
2. Ajuste a uma família de modelos de semivariograma;
3. Validação do modelo a ser utilizado nos procedimentos de Krigeagem.

Para finalizar, a inferência dos dados é realizada utilizando as técnicas de Krigeagem e o resultado é armazenado num PI do sistema SPRING.

Após a execução da Krigeagem e da aplicação do algoritmo sobre os dados a serem calibrados, o resultado das operações são incorporados à rotina de validação do SPRING.

2.4 Mapa de Variância

A interpolação consiste na predição do valor de um atributo, em uma localização espacial não amostrada, a partir de medidas feitas em outras localizações, que se situam dentro de uma vizinhança dada (BURROUGH; MCDONNELL, 1998)

Associado à distribuição normal dos dados está o mapa de variância ou incerteza. A incerteza é uma estimativa da magnitude do erro condicionada aos conhecimentos a priori, por exemplo, variabilidade espacial e amostras vizinhas, em relação ao atributo. A informação de incerteza, relacionada ao modelo de inferência adotado é útil, portanto, para avaliar a qualidade do modelo assumido para um determinado atributo (FELGUEIRAS et al., 2001). O mapa de incerteza revela, para cada valor de pixel, o erro pertinente em decorrência do processo da modelagem estocástica.

2.5 Álgebra de mapas

A álgebra de mapas pode ser vista como uma extensão da álgebra tradicional, sendo esta aplicada a dados geográficos. Denota o conjunto de operadores que manipulam campos geográficos tais como, imagens, mapas temáticos e modelos numéricos de terreno (BARBOSA, 1997)

Filho (2000) destaca sobre a utilização de funções e operadores utilizados pela álgebra de mapas, os quais permitem o processamento cartográfico. Este envolve a transformação, a combinação de camadas, o envolvimento destas com filtros de várias maneiras, tais como máscaras de vizinhança, conectividade, proximidades e rotas ótimas, e ainda a realização de análises estatísticas (BERRY, 1988).

Uma linguagem computacional de alto nível utilizada em uma modelagem cartográfica para representar uma sequência de processamento foi desenvolvida por Tomlin (1990). Esta linguagem é composta de operadores e funções espaciais classificadas em quatro tipos: local, focal, incremental e zonal.

A função local tem como característica a adoção de um novo valor para a localização de uma camada em função do valor originalmente associado àquele local na mesma camada ou em camadas diferentes. Dentre os operadores usados pelas funções locais, estão os de álgebra booleana, sendo os principais a interseção (AND) e a união (OR).

Funções focais e incrementais calculam um novo valor para cada localização em função de sua vizinhança. Essas funções correspondem ao conceito de filtragens espaciais em processamento de imagens.

A função zonal, por sua vez, inclui as operações que calculam um novo valor para uma localização em relação aos valores associados com uma zona contendo aquela localização. As zonas atuam como se fossem vizinhanças estendidas com formas e extensões singulares (FILHO, 2000).

Como procedimento algébrico neste trabalho foi realizada a diferença pixel a pixel de dois mapas a fim de verificar a semelhança existente entre esses.

2.6 Análise estatística

2.6.1 Regressão Linear

A análise de regressão consiste em verificar a relação funcional existente entre duas variáveis. Os valores da variável resposta (Y) em função da variação dos níveis do fator quantitativo (X) podem ser representados através de um gráfico de dispersão a fim de se estabelecer uma equação representativa do fenômeno em estudo. O objetivo da regressão é, portanto, obter um modelo matemático que melhor se ajuste aos valores observados de Y em função da variação dos níveis da variável X.

Uma informação auxiliar ao resultado da análise de variância da regressão consiste no coeficiente de determinação (R^2). Este varia no intervalo de 0 a 1 e tem por finalidade verificar se o modelo proposto é adequado ou não para descrever o fenômeno. O modelo de regressão linear simples é representado pela equação abaixo:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Em que,

Y_i : Variável dependente (variável resposta)

X_i : Variável independente

ε_i : Componente aleatório

β_0 : Intercepto populacional

β_1 : Inclinação populacional

Para que seja feita a análise de regressão linear é necessário que haja linearidade no fenômeno em estudo; que a variância dos termos de erro seja constante; que os erros apresentem normalidade e independência; e que não existam observações muito desproporcionais.

A análise de regressão neste trabalho teve como propósito avaliar a correlação existente entre a imagem TRMM e a Krigagem otimizada; a correlação entre duas modelagens geoestatísticas; e a correlação entre os dados coletados *in situ* e os dados orbitais.

2.6.2 Análise do Erro

A avaliação dos erros inerentes à inferência de dados torna-se tão importante quanto a própria análise estatística. Sendo assim, a existência de erros pode estar condicionada à seleção dos dados amostrais, bem como à maneira como esses se distribuem no espaço sob uma relação inter ou intraespecífica.

O erro associado à validação da modelagem geoestatística e à correlação entre os dados do satélite TRMM e os dados coletados *in situ* foram calculados pelos estimadores: Erro Médio Quadrático (MSE), e Raíz do Erro Médio Quadrático (RMSE).

Tabela 1. Estimadores de erros (MSE) e (RMSE)

Estimador	Fórmulas
MSE	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2$
RMSE	$RMSE = \sqrt{MSE}$

Onde, y_i e x_i são os valores de precipitação observados e estimados para i th amostras.

O MSE consiste numa medida frequentemente usada na verificação da acurácia de modelos numéricos. Este erro eleva as diferenças individuais ao quadrado sendo, portanto, sensível aos grandes erros. O MSE é sempre positivo e, quando apresenta valor igual a zero indica simulação perfeita.

A RMSE em adição ao MSE é usada, por sua vez, para expressar a acurácia dos resultados numéricos com a vantagem de apresentar valores de erro nas mesmas dimensões da variável analisada (HALLAK; FILHO, 2011).

Neste trabalho o Erro Quadrático Médio Normalizado (NRMSE) foi gerado para os dados de regressão entre a Krigagem otimizada e a imagem TRMM e para a análise entre os dados *in situ* e os dados orbitais.

3 RESULTADOS E DISCUSSÃO

3.1 Modelagem Geoestatística

A precipitação no Bioma Cerrado apresentou, pela análise estatística descritiva dos dados, um comportamento estacionário. Esse caráter representa a suposição de um comportamento homogêneo da estrutura de correlação espacial na região de estudo. Portanto, o processo é considerado estacionário se os efeitos de 1º e 2º ordem, isto é, média do processo no espaço e covariância são constantes em toda a região estudada. Sendo constantes, pode-se dizer que não há tendências na distribuição dos dados.

A curva de distribuição dos dados mostrou-se levemente assimétrica à direita com valor de coeficiente de assimetria de 1,057 (Figura 2). No entanto, os valores próximos de média (199,3) e mediana (192,5) demonstram uma distribuição gaussiana das amostras (Figura 3).

Figura 2. Distribuição Normal Padrão dos dados de precipitação no Cerrado.

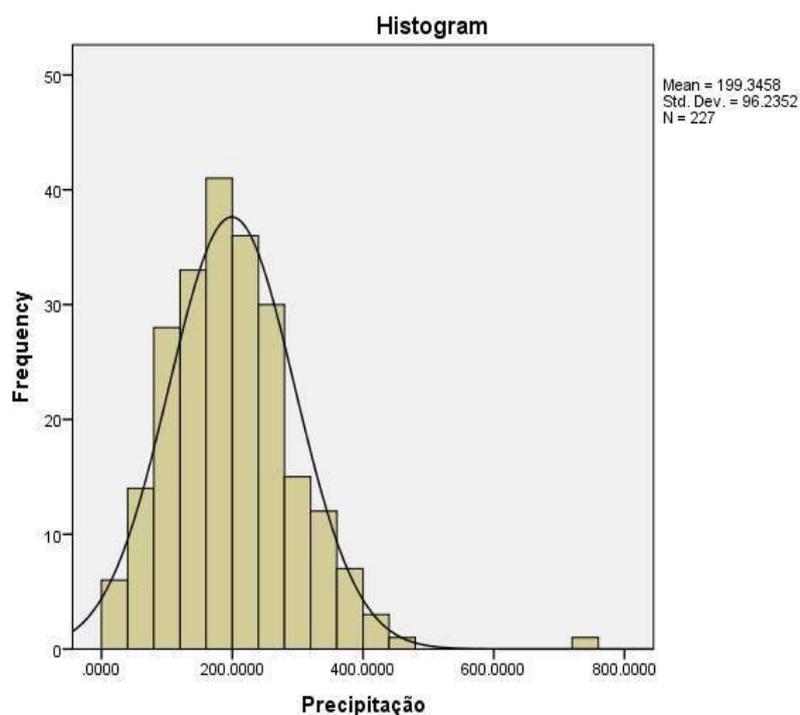


Figura 3. Coeficiente de assimetria, média e mediana caracterizando a distribuição simétrica dos dados.

```
      E S T A T I S T I C S: Precipitation
Number of Points ..... 219
Number of Valid Points ... 219
Mean Value ..... 199.30682532
Variance Value ..... 9379.53817140
Standart Desviation Value ..... 96.84801584
Coefficient of Variation ... 0.48592423
Skewness ... 1.05730379
Kurtose ..... 6.99093973
Minimum Value ..... -0.00002350
Lower Quartile ..... 128.99996948
Median Value ..... 192.59997559
Upper Quartile ..... 258.14996338
Maximum Value ..... 757.79998779
```

O semivariograma resulta do estudo da variabilidade espacial de um atributo, por meio de um conjunto amostral, levando-se em consideração a distância e a direção de ocorrência no espaço (FELGUEIRAS et al., 2001). Neste trabalho a modelagem do semivariograma empírico baseou-se na análise anisotrópica e a tolerância foi mantida, de acordo com a literatura (CAMARGO, et al., 2004), como metade do valor do incremento.

Por meio do ajuste do semivariograma empírico pelo modelo teórico esférico (Figura 4), pode-se dizer que a distância na qual os pares de amostras estão correlacionados (CAMARGO et al., 2004), isto é, o alcance, manteve-se dentro da área de estudo (171.342,471 metros).

Dentre os principais parâmetros obtidos com o ajuste do semivariograma empírico estão o Efeito Pepita e o Akaike. Quanto maior o primeiro, maior a probabilidade da variabilidade dos dados ser explicada pelo acaso. Assim, um efeito pepita baixo revela que a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras é baixa. O Akaike, por sua vez, infere sobre a aplicabilidade do modelo teórico para o ajuste da distribuição dos dados.

O ajuste do semivariograma empírico resultou em baixos valores de Efeito Pepita (28,43) e Akaike (-52,19) predizendo, portanto, uma boa representatividade do modelo. A

Figura 5 ilustra os valores obtidos para os devidos parâmetros. A Krigagem gerada por meio do ajuste do semivariograma empírico é observada na

Figura 6.

Figura 4. Ajuste do semivariograma empírico por meio de um modelo teórico esférico.

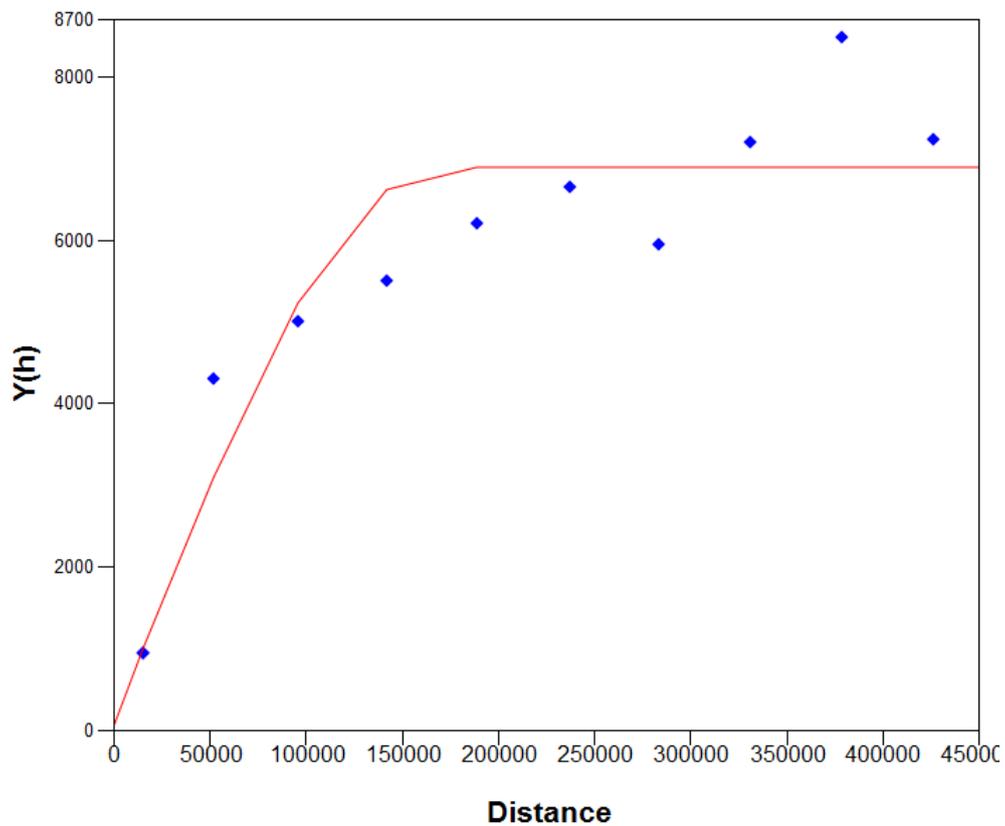


Figura 5. Obtenção do alcance e dos valores de Efeito Pepita e Akaike para o modelo teórico ajustado.

```

FITTING OF SEMIVARIOGRAM
*****
Data file summary:
Data file name: C:\Users\Diogo Amore\Desktop\Piccina\DATA\Geo_coursework\Precipitation\Precipitation/G
No. of variables: 3
No. of points: 12
No. of points used: 12

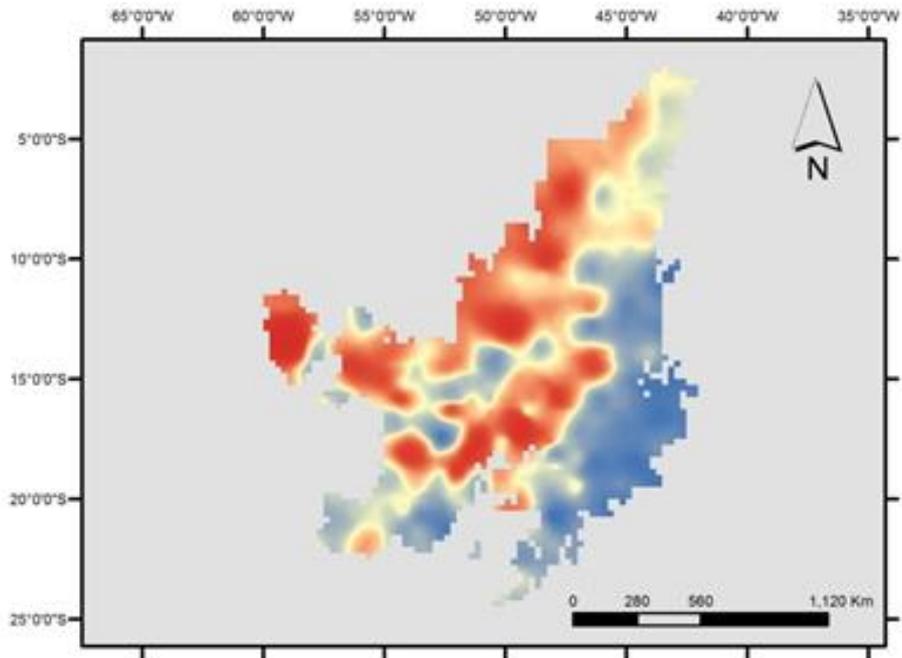
Initial Parameters:
Nugget effect (Co): 0.000
For transtive model: Spherical
Sill-nugget (C1): 7767.262
Range (a): 260096.142

Spherical semivariogram model

```

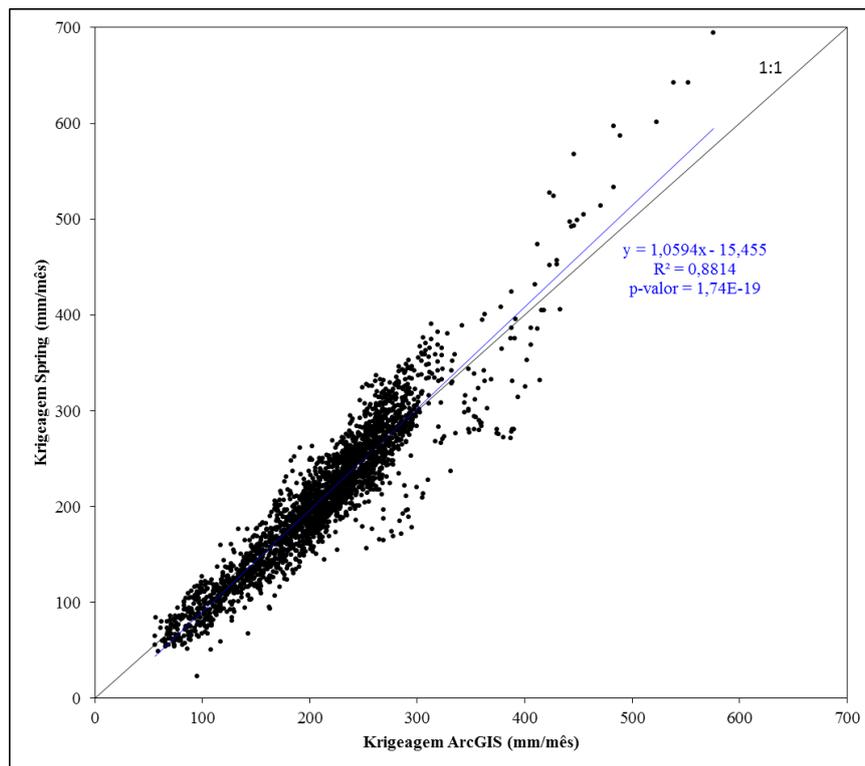
No.	Akaike	Nugget Effect	Contribution	Range
1	-51.226	235.517	6773.641	208605.361
2	-51.906	117.907	6824.981	185235.496
3	-52.191	28.437	6854.678	171342.471

Figura 6. Krigeagem gerada por meio do semivariograma empírico.



Ao contrário da avaliação de erros obtida por uma única validação cruzada pelo semivariograma empírico, o ajuste otimizado permitiu, por meio de inúmeras validações cruzadas, a obtenção de valor zero para o Efeito Pepita. Dessa forma, foi possível estabelecer uma correlação existente entre a modelagem geoestatística gerada através do semivariograma empírico e outra gerada pelo semivariograma otimizado. O gráfico de dispersão, representado pela Figura 7, aponta um coeficiente de determinação (R^2) igual a 0,88 e um p-valor baixo, o que demonstra uma alta correlação entre as Krigeagens.

Figura 7. Gráfico de dispersão representando a correlação entre a Krigeagem obtida por meio do ajuste do semivariograma empírico (Spring) e a Krigeagem obtida por meio do ajuste otimizado (Arcgis).



3.2 Mapa de Variância

Assim, percebe-se pela **Error! Reference source not found.** as incertezas inerentes a cada pixel conforme sua disposição na área de estudo. Pela Tabela 2, nota-se o percentual de pixels em cada range de incerteza.

Figura 8. Mapa de incerteza associado à modelagem estocástica. Os círculos pretos representam as áreas de maior variância na área de estudo.

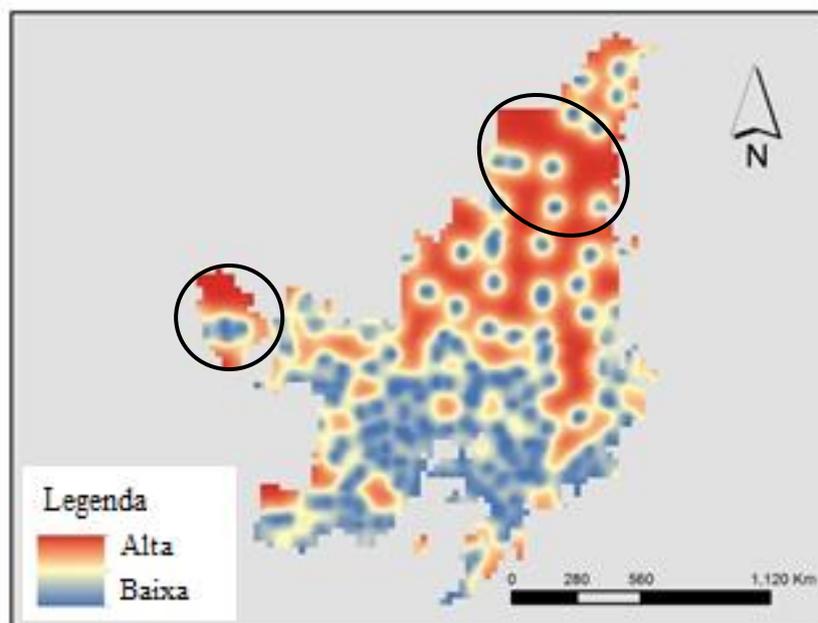


Tabela 2. Percentual de pixels em cada range de incerteza.

Range de variância (incerteza)	Número de pixels em cada range	Percentual de cada range sobre o total de pixels	Percentual cumulativo
$7000 < x < 7587$	147	5,35%	99,93%
$6000 < x < 7000$	391	14,23%	94,58%
$5000 < x < 6000$	428	15,58%	80,34%
$4000 < x < 5000$	480	17,47%	64,76%
$3000 < x < 4000$	497	18,09%	47,29%
$2000 < x < 3000$	535	19,48%	29,20%
$1000 < x < 2000$	261	9,50%	9,72%
$0 < x < 1000$	6	0,22%	0,22%

O mapa de incerteza sugere que as áreas de ocorrência de maior variância estão relacionadas principalmente as poucas ocorrências de amostras e à distribuição esparsa destas nas regiões circuladas no mapa da área de estudo.

Pela Tabela 2 tem-se que o maior percentual de pixels (70%) encontra-se no range de variância entre 2000 e 6000, o que indica uma equidade na distribuição das amostras dentro das classes de incerteza. Assim, pode-se dizer que aproximadamente 20% dos pixels encontram-se no range de maior variância, entre 6000 e 7587 e que, cerca de 10% dos pixels, estão presentes no range de menor incerteza.

3.3 Álgebra de Mapas e Processamento Digital de Imagens

Pela álgebra de mapas, a partir da diferença pixel a pixel entre a imagem TRMM e a Krigeagem otimizada (**Error! Reference source not found.**), foi possível concluir que aproximadamente 75% dos pixels amostrados encontram-se entre os valores de -50 e +50 de um total de -299,42 e +209,49, conforme representado pela distribuição das amostras nos quartis do histograma (Figura 10).

Dessa forma, pode-se dizer que os pixels que obtiveram valores de diferença de até -50 concentram-se no primeiro quartil (12,5) e aqueles que obtiveram valores de diferença acima de +50 encontram-se distribuídos no último quartil, equivalente a 87,5.

Figura 9. Mapa de diferença entre a imagem TRMM e a Krigeagem otimizada

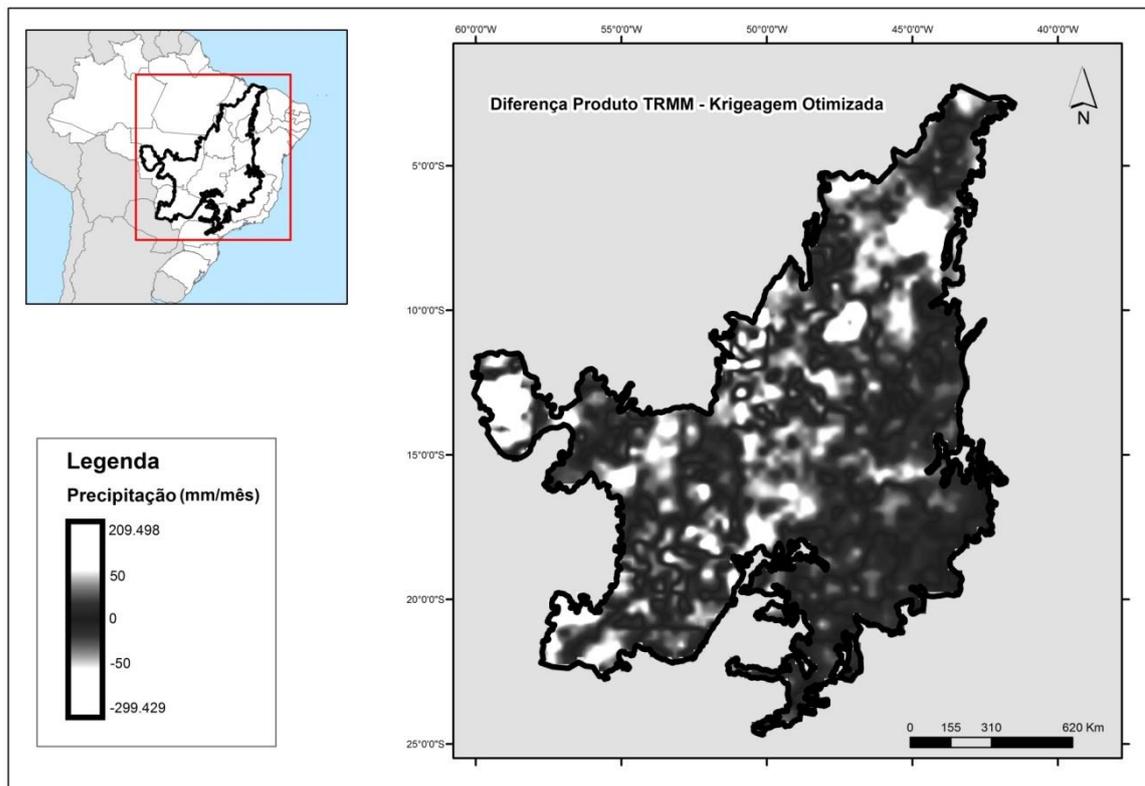
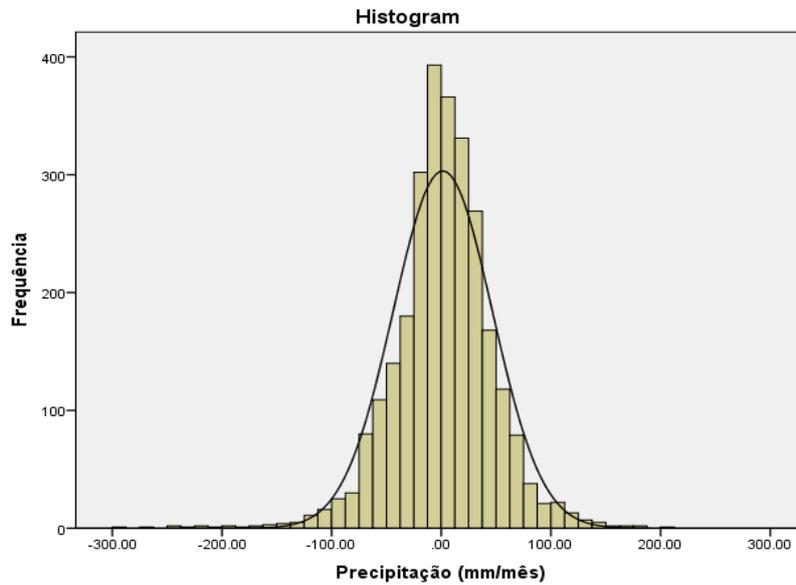


Figura 10. Representação do histograma indicando que a distribuição da maior parte das amostras encontra-se no segundo e terceiro quartil.



A análise de regressão entre a imagem TRMM e a Krigagem otimizada apresentou p-valor baixo e valores de R^2 igual a 0,64 e NRMSE de 0,09. Pela análise do gráfico de dispersão (Figura 11), significa dizer que existe uma alta correlação entre as duas variáveis, possibilitando a validação do modelo geoestatístico proposto. Pela análise visual da comparação entre a imagem TRMM e a Krigagem otimizada, nota-se uma semelhança na distribuição da precipitação para Bioma Cerrado (Figura 12).

Figura 11. Análise de regressão entre a imagem TRMM e a Krigagem otimizada.

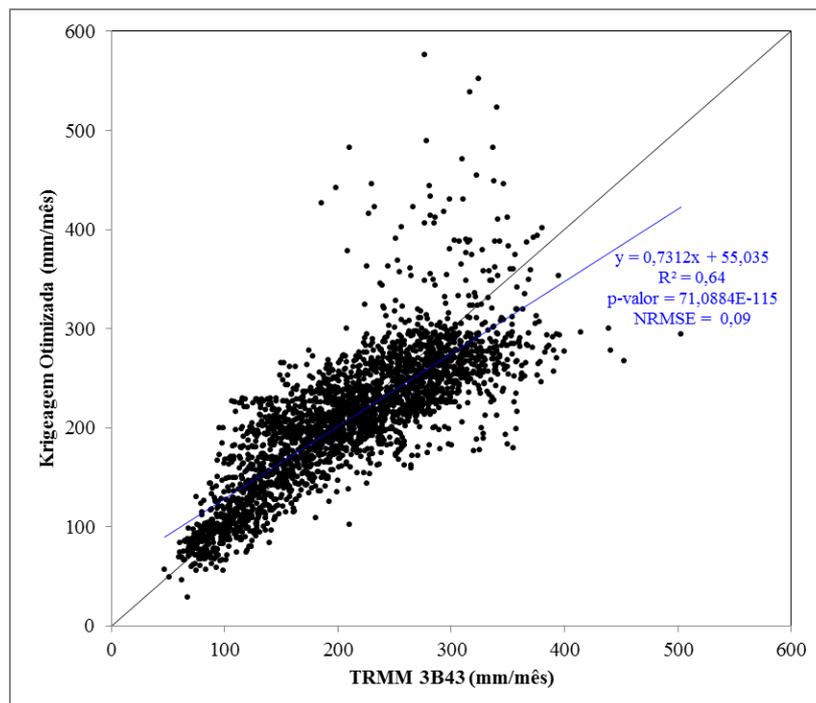
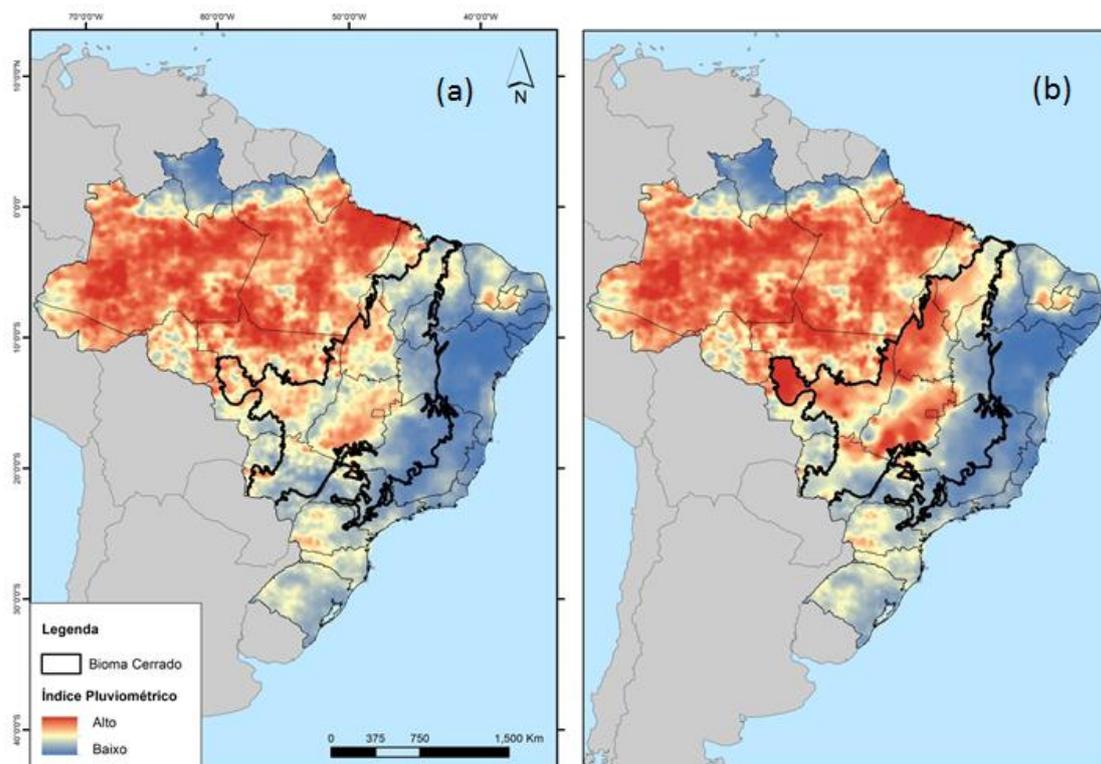


Figura 12. (a) Produto TRMM recortado para o Brasil. (b) Produto TRMM ao fundo e Krigeagem otimizada recortada para o Bioma Cerrado.



A análise de regressão entre os dados *in situ* e os dados obtidos pelo satélite TRMM (Figura 13), observa-se a obtenção de um p-valor baixo, coeficiente de determinação de 0,55 e um NRMSE de 0,07. Pela análise pode-se afirmar que ambos apresentam alta correlação entre si e um pequeno erro associado o que explica, de certa forma, parte do erro observado na validação da Krigeagem por meio da imagem TRMM.

Por fim, a representação da dispersão de observações é dada por meio do gráfico de Boxplot (Figura 14). Neste é possível comparar a distribuição de valores de cada variável, bem como avaliar a ocorrência de possíveis outliers associados à variável estudada.

Figura 13. Gráfico de dispersão representando a correlação existente entre a imagem TRMM e os dados coletados *in situ*.

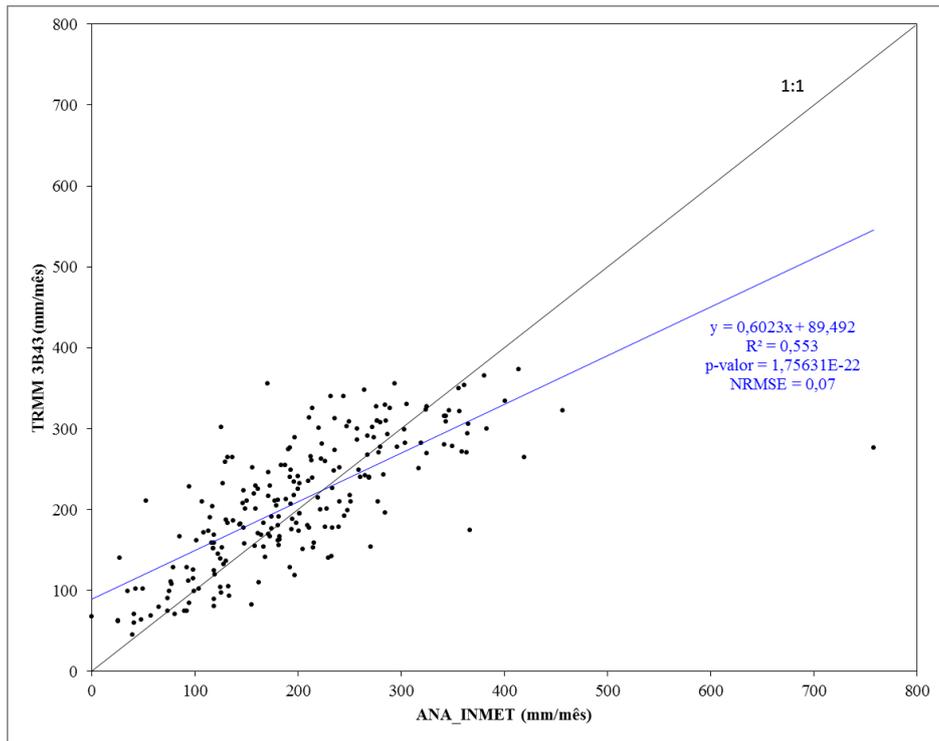
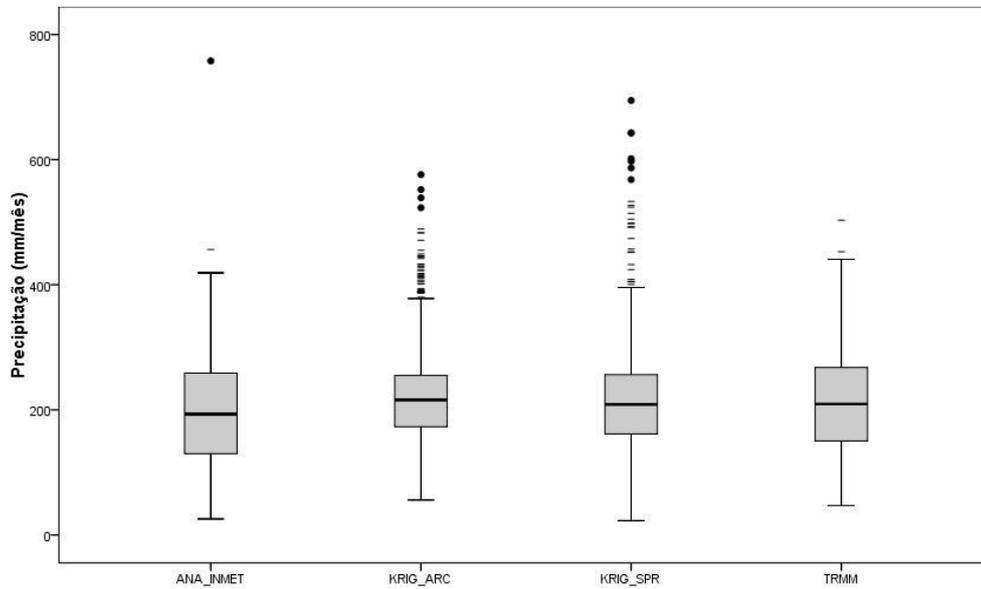


Figura 14. Boxplot representando a dispersão das observações em cada variável.



4 CONCLUSÕES

A análise estatística descritiva dos dados implicou na estacionariedade dos dados de precipitação no Bioma Cerrado admitindo, assim, uma distribuição gaussiana. Essa condição pressupõe um comportamento homogêneo da estrutura de correlação espacial na região de estudo, denotando constantes a média e a variância dos dados.

O alcance gerado pelo ajuste do semivariograma empírico manteve-se dentro da área de estudo e o valor do efeito pepita para este esteve abaixo do limite de um quarto do patamar, indicando resultados aceitáveis. No entanto, a Krigeagem otimizada apresentou melhor robustez na interpolação dos dados de precipitação devido, provavelmente, à distribuição normal dos dados, bem como às inúmeras validações cruzadas aplicadas sobre o modelo a fim de se obter valores ótimos para os parâmetros de interesse.

Pelo mapa de variância conclui-se que as áreas de ocorrência de maior variância estão relacionadas principalmente às poucas ocorrências de amostras e à distribuição esparsa destas na região norte do Bioma, principalmente. A maior porcentagem de pixels encontra-se distribuída de forma equivalente para os ranges de variância.

A diferença pixel a pixel entre a imagem TRMM e a Krigeagem otimizada obteve um resultado satisfatório, já que a maior parte das amostras apresentou valores de diferença em torno de zero.

As análises de regressão linear apresentaram valores de coeficientes de determinação altos, assim como p-valores baixos para as correlações propostas neste estudo.

Quanto aos erros inerentes à estimação dos valores, estes estão associados ao número de amostras e à distância entre os pares correlacionados. Pela análise dos resultados pode-se concluir que os erros inerentes aos dados meteorológicos *in situ* e ao produto TRMM influenciaram na calibração e validação da Krigeagem ordinária, porém, todos estiveram dentro do limiar de significância.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANA. Orientações para consistência de dados pluviométricos. , 2012.
- ASSAD, E. D.; MASUTOMO, R.; ASSAD, M. L. L. Estimativa das precipitações máximas prováveis com duração de 24 horas e de 30 minutos. Caso dos Cerrados Brasileiros. . Brasília, 1992.
- BARBOSA, C. C. F. **Álgebra de mapas e suas aplicações em sensoriamento remoto e geoprocessamento**. INPE, 1997.
- BERRY, J. K. . Assessing Spatial Impacts of Land Use Plans. **Environmental Management**, v. 27, p. 1–9, 1988.
- BERTONI, J. C.; TUCCI, C. E. M. Precipitação. **Hidrologia: ciência e aplicação**. 2^a ed. Porto Alegre, 2001.
- BURROUGH, P. A. Spatial aspects of ecological data. **Data analysis in community and landscape ecology**. p.213–251, 1987.
- BURROUGH, P. A.; MCDONNELL, R. A. Principles of Geographical Information Systems. . New York, 1998.
- CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. Análise Espacial de Superfícies. **Análise Espacial de Dados Geográficos**. p.3.1–3.37. São José dos Campos, 2004.
- DAVID, M. Geostatistical Ore Reserve Estimation. . The Netherlands, 1977.
- FELGUEIRAS, C. A.; FUKS, S. D.; MONTEIRO, A. M. V. Métricas de incertezas em modelagem de atributos espaciais. , 2001.
- FERREIRA, L. G.; ASNER, G. P.; KNAPP, D. E.; et al. Equivalent water thickness in savanna ecosystems: MODIS estimates based on ground and EO-1 Hyperion data. **International Journal of Remote Sensing**. doi: 10.1080/01431161.2010.523731, 2011.
- FILHO, B. S. S. Modelagem de dados espaciais. , 2000.
- FRANCO, A. C.; BUSTAMANTE, M.; CALDAS, L. S.; et al. Leaf functional traits of Neotropical savanna trees in relation to seasonal water deficit. **Trees**, v. 19, p. 326–335, 2005.
- HALLAK, R.; FILHO, A. J. P. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: sensibilidade a variações com os esquemas de advecção e assimilação de dados. **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 591–608. doi: 10.1590/S0102-77862011000400009, 2011.
- ISAAKS, E. H.; SRIVASTAVA, R. M. An introduction to applied geostatistics. . New York, 1989.

JOURNAL, A. G. New distance measures: The route toward truly non-Gaussian geostatistics. **Mathematical Geology**, v. 20, n. 4, p. 459–475. doi: 10.1007/BF00892989, 1988.

KRIGE, D. G. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**, p. 201–215. doi: 10.2307/3006914, 1952.

LENZA, E.; KLINK, C. A. Comportamento fenológico de espécies lenhosas em um cerrado sentido restrito de Brasília, DF. **Revista Brasileira de Botânica**, v. 29, n. 4, p. 627–638. doi: 10.1590/S0100-84042006000400013, 2006.

LIMA, T. A.; PINTO, J. R. R.; LENZA, E.; PINTO, A. D. S. Florística e estrutura da vegetação arbustivo-arbórea em uma área de cerrado rupestre no parque estadual da Serra de Caldas Novas, Goiás. **Biota Neotropica**, v. 10, n. 2, p. 159–166. doi: 10.1590/S1676-06032010000200020, 2010.

MARCUZZO, F. F. N.; MELO, D. C. . R. Sazonalidade e Distribuição Espaço-Temporal das Chuvas no Bioma do Cerrado do Estado do Mato Grosso do Sul. **Revista Brasileira de Recursos Hídricos**, v. 17, n. 1, p. 77–86, 2012.

MIRANDA, I. S. Fenologia do estrato arbóreo de uma comunidade de cerrado em Alter-do-Chão. **Revista Brasileira de Botânica**, v. 18, n. 2, p. 235–240, 1995.

MORAIS, H. C.; DINIZ, I. R.; BAUMGARTEN, L. Padrões de produção de folhas e sua utilização por larvas de Lepidoptera em um Cerrado de Brasília. **Revista Brasileira de Botânica**, v. 18, p. 163–170, 1995.

MOREIRA, M. **Metodologia para definir padrões pluviométricos. Caso: Cerrado brasileiro.** UnB, 1985.

OLIVEIRA, P. E. Fenologia e biologia reprodutiva das espécies de cerrado. **Cerrado: Ambiente e Flora**. p.169–192. Embrapa - CPAC, 1998.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do Bioma Cerrado. **In Cerrado: ecologia e flora (S.M. Sano, S.P. Almeida & J.F. Ribeiro, eds.). Embrapa Cerrados**. p.151–212, 2008.

SANTOS, M. A. DOS; BARBIERI, A. F.; CARVALHO, J. A. M. DE; MACHADO, C. J. O Cerrado Brasileiro: Notas Para Estudo. , p. 6–15, 2010.

SARMIENTO, G. The ecology of neotropical savannas. . Cambridge, 1984.

SARMIENTO, G.; GOLDSTEIN, G.; MEINZER, F. Adaptive strategies of woody species in neotropical savannas. , 1985.

TOMLIN, C. D. Geographic Information Systems and Cartographic Modelling. . New Jersey, 1990.

VIANA, D. R.; FERREIRA, N. J.; CONFORTE, J. C. Avaliação das estimativas de precipitação 3B42 e 3B43 do satélite TRMM na Região Sul do Brasil. **Congresso Brasileiro de Meteorologia**. Belem, PA, 2010.