

Validação de uma Krigeagem Ordinária por meio do produto TRMM 3B43 de precipitação mensal: um estudo de caso para o Bioma Cerrado

Alana Carla Toniol¹
Diogo de Jesus Amore¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
alana.toniol@inpe.br, amore@dsr.inpe.br

Abstract. Biome Cerrado is characterized by its high inter-seasonal variability. Precipitation is an important feature of this biome, and such parameter can be spatialized from point-based samples. In such context, this study aimed to validate the Ordinary Kriging (OK) interpolation technique via the use of TRMM 3B43 monthly precipitation product. It also aimed to compare two geostatistical models (optimized and non-optimized OK interpolation), as well as evaluate errors associated with the correlation between *in situ* and orbital data. Scatterplots and Normalized Root Mean Square Errors (NRMSE) were created for the models statistical analysis. Uncertainty maps were also produced in order to verify the visual quality of the models. It has been found that there is high correlation between the TRMM product and the optimized OK interpolation model ($R^2=0.64$, $NRMSE=0.07$, $p\text{-value}<0.01$). Considering data variance of non-optimized OK, only 20% of the pixels fell into the highest variance *range* (6000-7587). TRMM and *in situ* inherent errors influenced the calibration and validation of both geostatistical models, however, both were located within the statistical significant *range* of 1%. Overall, TRMM 3B43 product proved efficient in predicting Biome Cerrado precipitation which allowing the validation of geostatistical modelling for that area.

Palavras-chave: remote sensing, Kriging, rainfall, Cerrado, sensoriamento remoto, Krigeagem, precipitação, Cerrado.

1. Introdução

Considerado o segundo maior bioma brasileiro em termos de extensão, o Cerrado se destaca por sua grande diversidade florística representada pelos estratos herbáceo, arbustivo e arbóreo (Ribeiro e Walter, 1998). O Bioma é marcado ainda pela alta sazonalidade existente entre os períodos chuvoso e seco exigindo, portanto, uma adaptação dos diversos hábitos da vegetação durante o déficit hídrico na superfície.

A precipitação média anual no Cerrado varia entre 1200 mm e 1800 mm, o que contribui para a alimentação de seis das oito maiores bacias hidrográficas brasileiras (Santos et al., 2010). O estudo da distribuição da precipitação apresenta-se como uma tentativa de regionalização de grandes padrões pluviométricos do Cerrado. Outros aspectos associados à precipitação no Bioma incluem estudos quanto à perda da capacidade produtiva de áreas agrícolas, decorrente de erosões hídricas; ao dimensionamento de barragens; à implantação de sistemas de irrigação e drenagem; e ao controle de enxurradas (Assad et al., 1992).

O conhecimento da distribuição de chuvas numa região permite a elaboração de um planejamento de médio e longo prazo para que se obtenha um melhor aproveitamento do recurso hídrico. Assim, Bertoni e Tucci (2001) ressaltam a importância dos postos de medições de chuvas, os quais visam obter uma série ininterrupta de precipitações ao longo dos anos. Dentre as vantagens da coleta tradicional de dados pluviométricos está a acurácia na obtenção das informações, uma vez que os dados de chuva são coletados *in situ*. No entanto, o processo de mensuração pode ser comprometido por falhas na coleta sistemática dos dados.

Outro método utilizado para a obtenção de dados pluviométricos consiste, portanto, no 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. 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 uso do TRMM

consiste em utilizar os dados das imagens orbitais para a validação de dados *in situ*, o que poderia suprir a ausência de validações estatisticamente consistentes para a modelagem geostatística da precipitação acumulada mensal para o Bioma Cerrado.

Dentre as técnicas utilizadas na modelagem geostatística destaca-se a Krigagem, a qual pode ser entendida como uma predição linear ou uma forma de inferência bayesiana (Krige, 1952). A técnica 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, não mais existiriam valores próximos porque a correlação espacial deixaria de existir.

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), a partir do produto TRMM-3B43. Assume-se a hipótese de que o produto do satélite TRMM seja um preditor da precipitação acumulada numa modelagem geostatística para o Cerrado.

1. Metodologia de Trabalho

2.1 Área de estudo

A área de estudo consiste no Bioma Cerrado, com aproximadamente 2 milhões de Km², correspondente a 22% do território nacional (Figura 1). Localizado entre as coordenadas 25° S e 45° W, o Bioma engloba doze estados brasileiros e abriga nascentes das três maiores bacias hidrográficas da América do Sul (Amazônica/Tocantins, São Francisco e Prata). 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.

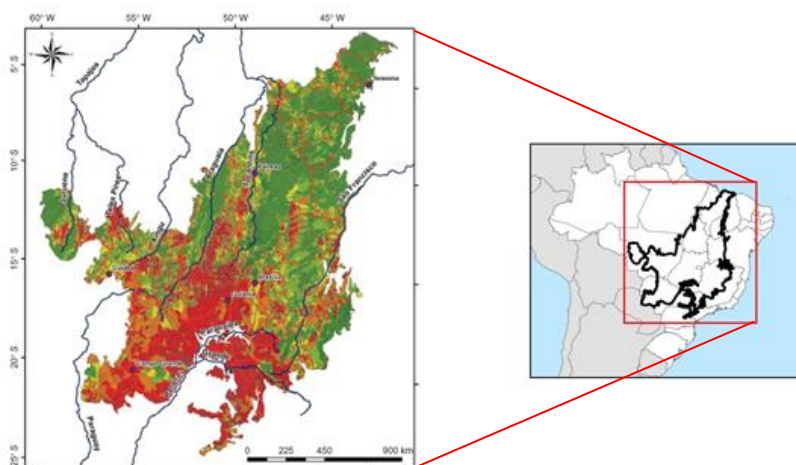


Figura 1. Área de estudo representada pelo Bioma Cerrado.

2.1 Aquisições de dados

Os dados de precipitação *in situ* foram adquiridos no site do INMET e ANA, os quais foram coletados por meio de pluviômetros, nos quais é lida a lâmina de água acumulada em mm/dia ou mm/chuva; e pluviógrafos, que medem a quantidade e a duração da precipitação, registrando a intensidade de chuva. Já os dados orbitais de precipitação, em mm/hora, foram obtidos no site da NASA, por meio do produto TRMM-3B43, o qual apresenta uma resolução espacial de 0,25 grau.

2.2 Modelagem geoestatística

A análise exploratória e estrutural dos dados, bem como a realização de inferências são elementos essenciais à aplicação geoestatística. Dentre as hipóteses requeridas para a inferência dos dados por meio da Krigagem, admite-se que a componente determinística (que descreve a componente estrutural) seja constante de forma que não haja tendências na região. Admite-se ainda a estacionariedade da covariância entre dois pares quaisquer $Z(x)$ e $Z(x+h)$, separados por um vetor distância h . Assim, pela Equação 1 temos:

$$C(h) - Cov[Z(x), Z(x+h)] = E[Z(x) * Z(x+h)] - m^2, \forall x \quad (1)$$

A estacionariedade da covariância também implica na estacionariedade do variograma porém, na geoestatística, adota-se uma hipótese de estacionariedade menos restritiva denominada intrínseca. Neste caso, admite-se que $E[Z(x)] = \forall x$, e que a variância das diferenças depende somente do vetor distância h , dado pela Equação 2:

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

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

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). Se as condições especificadas pela hipótese intrínseca são contempladas, então o semivariograma pode ser estimado a partir do dado amostral (Equação 3):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} * \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2 \quad (3)$$

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

Os estimadores dos métodos de inferência da Krigagem Ordinária (KO) são baseados na Equação 4, podendo o sistema de KO ser escrito em notação matricial conforme Equação 5 (Journel, 1988):

$$Z_{x_0}^* = \sum_{i=1}^n \lambda_i Z(x_i) \quad (4)$$

Em que: $\sum_{i=1}^n \lambda_i = 1$

$$K * \lambda = k \quad (5)$$

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

2.3 Análise dos dados

Para a análise dos dados foi elaborado um mapa de variância ou incerteza, associado à distribuição normal das amostras. O mapa de incerteza revelou, para cada valor de *pixel*, o erro pertinente em decorrência do processo de modelagem.

Para a análise estatística foi realizada uma regressão linear, estabelecendo-se uma equação representativa do fenômeno em estudo. A análise de regressão neste trabalho teve como propósito avaliar as correlações existentes entre duas modelagens geoestatísticas; entre a imagem TRMM e a KO otimizada (realizada no *software* Arcgis); e entre os dados coletados *in situ* e os dados orbitais.

A análise de erros, por sua vez, condicionou-se à seleção dos dados amostrais, bem como à maneira como esses se distribuíram 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 Raiz do Erro Médio Quadrático (RMSE). Neste trabalho o Erro Quadrático Médio Normalizado (NRMSE) foi gerado para os dados de regressão entre a KO otimizada e a imagem TRMM e para a análise entre os dados *in situ* e os dados orbitais.

2. 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.

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

Pelo ajuste do semivariograma empírico, realizado pelo modelo teórico esférico (Figura 2), observa-se que a distância na qual os pares de amostras estiveram correlacionados (Camargo et al., 2004) manteve-se dentro da área de estudo (171.342,471 metros).

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. Em seguida observa-se a KO empírica gerada no *software* SPRING por meio do ajuste do semivariograma empírico.

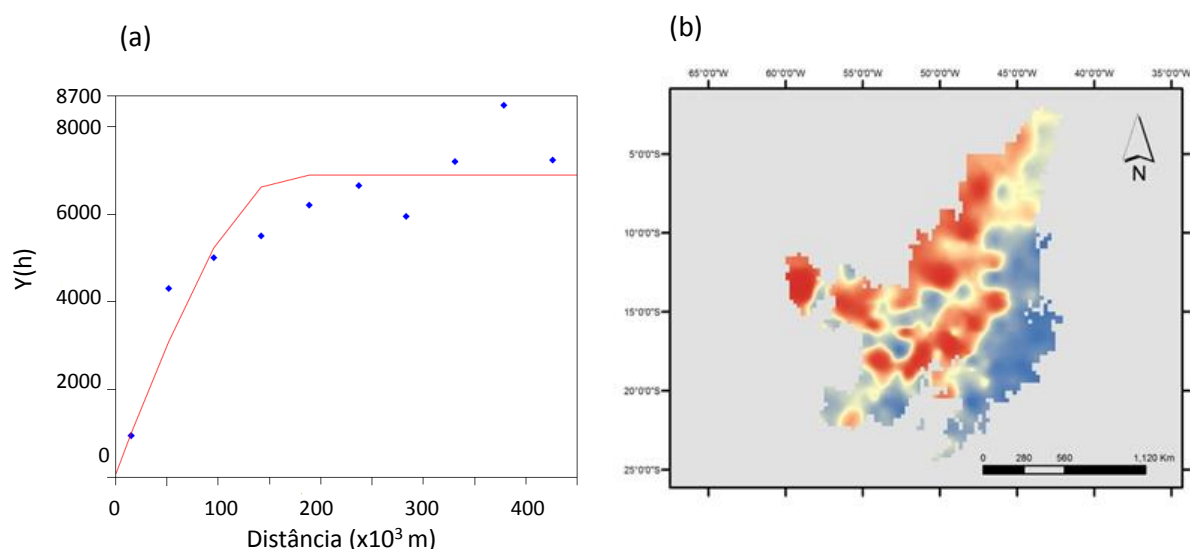


Figura 2. Ajuste do semivariograma empírico por meio de um modelo teórico esférico (a) e resultado da KO empírica (b).

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 3, apontou um coeficiente de determinação (R^2) igual a 0,88 e um p-valor baixo, o que demonstra uma alta correlação entre as Krigeagens Ordinárias.

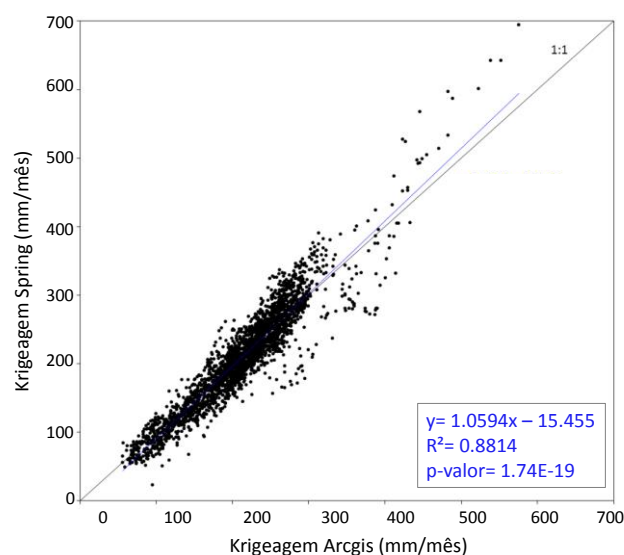


Figura 3. Gráfico de dispersão representando a correlação entre a KO obtida por meio do ajuste do semivariograma empírico (Spring) e a KO obtida por meio do ajuste otimizado (Arcgis).

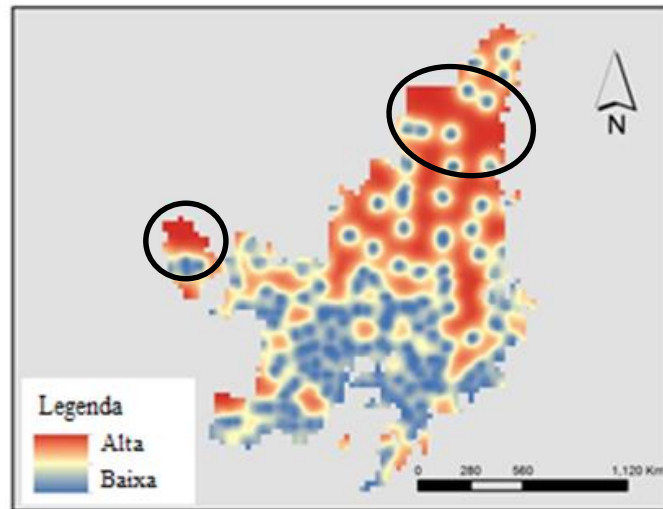
3.2 Análise estatística

Pelos dados apresentados na Tabela 1 tem-se que o maior percentual de *pixels* (70%) ocorreu no intervalo de variância entre 2000 e 6000, indicando uma equidade na distribuição das amostras dentro das classes de incerteza. Pode-se dizer ainda que aproximadamente 10% dos *pixels* estiveram dentro do intervalo de menor incerteza, enquanto que 20% desses ocuparam o intervalo de maior incerteza (6000-7587). Pelo mapa de incerteza da Figura 4, nota-se que as maiores variâncias estiveram associadas às áreas com poucas amostras de campo e às suas distribuições esparsas.

Tabela 1. Percentual de *pixels* associados aos intervalos 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%

Figura 4. Mapa de incerteza associado à KO não-otimizada. Os círculos pretos representam as áreas de maior variância na área de estudo.



A análise de regressão entre a imagem TRMM e a KO 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 5), pode-se dizer que existe uma alta correlação entre as duas variáveis, possibilitando a validação do modelo geoestatístico proposto.

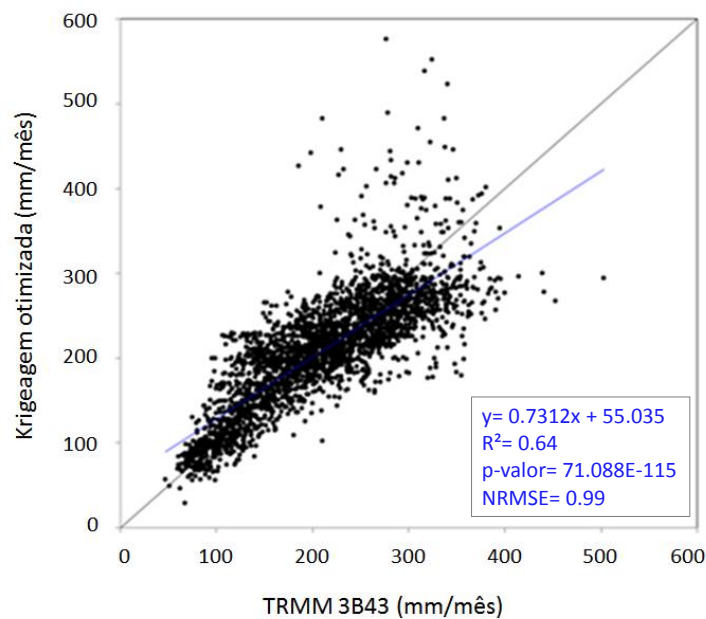


Figura 5. Análise de regressão entre a imagem TRMM e a KO otimizada.

Pela análise visual da comparação entre a imagem TRMM e a KO otimizada, nota-se uma semelhança na distribuição da precipitação para Bioma Cerrado (Figura 6).

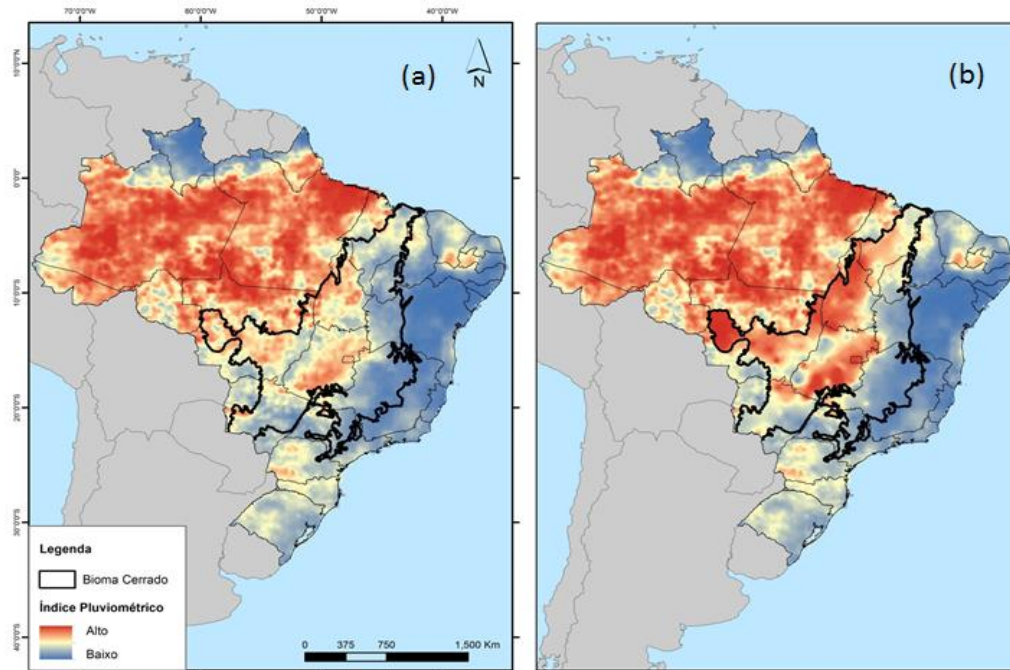


Figura 6. Produto TRMM recortado para o Brasil (a). Produto TRMM ao fundo e KO otimizada recortada para o Bioma Cerrado (b).

A Figura 7 mostra a análise de regressão entre os dados *in situ* e os dados obtidos pelo satélite TRMM e a representação da dispersão das observações, que é dada por meio do *Boxplot*. Por esses resultados observa-se um p-valor baixo, coeficiente de determinação de 0,55 e um NRMSE de 0,07. Ambos os dados apresentaram alta correlação entre si e um pequeno erro associado. Este justifica parte do erro observado na validação da KO por meio da imagem TRMM. Pelo *Boxplot* foi possível comparar a distribuição de valores de cada variável, bem como avaliar a ocorrência de possíveis *outliers* associados à variável estudada.

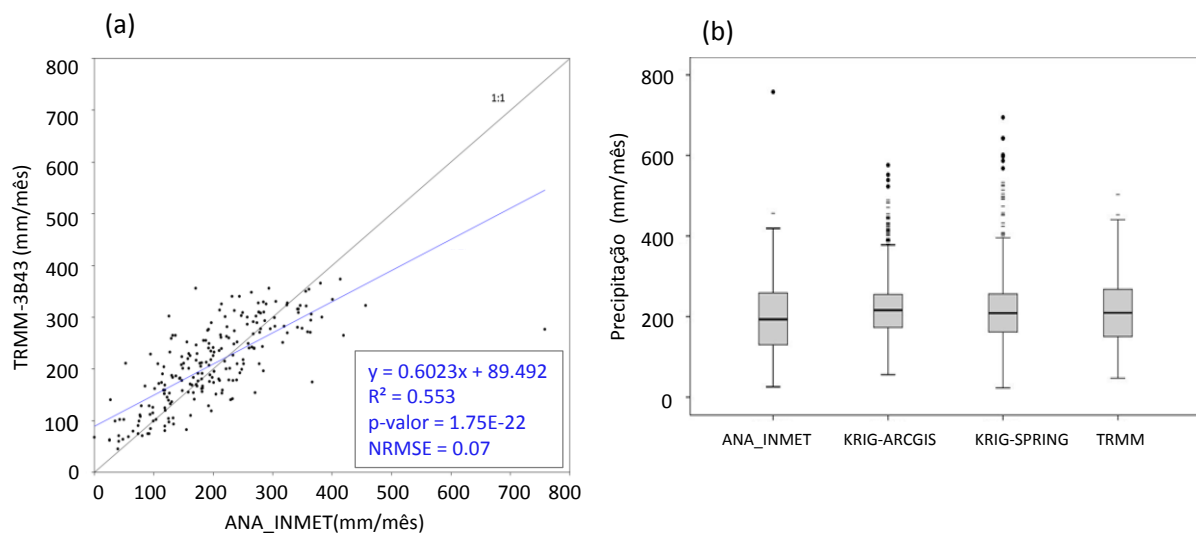


Figura 7. Gráfico de dispersão representando a correlação existente entre a imagem TRMM e os dados coletados *in situ* (a) e *Boxplot* (b) representando a dispersão das observações em cada variável.

3. Conclusões

A análise estatística descritiva 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 esteve abaixo do limite de um quarto do patamar, indicando resultados aceitáveis. No entanto, a KO 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.

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

As análises de regressão linear apresentaram valores de coeficientes de determinação altos, assim como p-valores menores que 0.01. Quanto aos erros inerentes à estimação dos valores, esses estiveram 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 KO, porém, todos estiveram dentro do limiar de significância.

Agradecimentos

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa para o desenvolvimento desse trabalho durante o programa de pós graduação no Instituto Nacional de Pesquisas Espaciais (INPE).

Referências Bibliográficas

- 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.
- BERTONI, J. C.; TUCCI, C. E. M. Precipitação. **Hidrologia: ciência e aplicação**. 2^a ed. Porto Alegre, 2001.
- 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.
- FELGUEIRAS, C. A.; FUKS, S. D.; MONTEIRO, A. M. V. Métricas de incertezas em modelagem de atributos espaciais. , 2001.
- JOURNEL, 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.
- RIBEIRO, J. F.; WALTER, B. M. T. Fitofisionomias do bioma cerrado. p.89–166, 1998.
- 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.