

Estudo comparativo dos interpoladores espaciais IDW e SPLINE para dados pluviométricos no Norte de Minas Gerais

Paulo Victor Alves de Oliveira – UFMG (paulovictor3491@gmail.com)

Marcos Antônio Timbó Elmiro – UFMG (timboelmiro@gmail.com)

Diego Rodrigues Macedo – UFMG (diegorm@ufmg.br)

Marcelo Antonio Nero – UFMG (marcelo.nero@gmail.com)

Resumo: A interpolação é um método matemático essencial na espacialização de dados ambientais. Trata-se de uma técnica amplamente difundida para estimar valores desconhecidos de fenômenos contínuos, entre pontos amostrados. Em dados de precipitação a interpolação é uma alternativa descomplicada e acessível de espacializar estes dados em grandes áreas. A presente pesquisa tem como objetivo principal avaliar dois métodos de interpolação determinísticos: Inverso da Distância Ponderada (IDW) e Spline, aplicados na análise de dados pluviométricos. A área de estudo compreende o Norte de Minas Gerais, uma região seca com temperaturas elevadas e pouca disponibilidade hídrica. Foram utilizados dados de 99 estações pluviométricas para o ano de 2022, sendo 68 delas dentro da mesorregião norte e as outras 31 fora, em um buffer de 50 km. O processamento envolveu o uso dos métodos IDW e SPLINE, com validação cruzada (Cross Validation) e métricas de acurácia o ME (Erro Médio), MAE (Erro Absoluto Médio), MSE (Erro Quadrático Médio) e RMSE (Raiz do Erro Quadrático Médio). A análise descritiva dos dados revelou alta variabilidade nas precipitações. As métricas de Cross Validation indicaram uma leve superioridade do Spline sobre o IDW. Conclui-se que o Spline é mais eficaz na interpolação de dados pluviométricos na região, e recomenda-se explorar métodos geoestatísticos, como a Krigagem, em estudos futuros para uma abordagem mais completa.

Palavras-chave: Interpolação; Modelos determinísticos; Cross Validation; Precipitação.

Comparative study of spatial interpolators IDW and SPLINE for pluviometric data in Northern Minas Gerais

Abstract: Interpolation is an essential mathematical method in the spatialization of environmental data. It is a widely used technique to estimate unknown values of continuous phenomena, between sampled points. In precipitation data, interpolation is a straightforward and accessible alternative for spatializing these data over large areas. The main objective of this research is to evaluate two deterministic interpolation methods: Inverse Distance Weighted (IDW) and Spline, applied to the analysis of pluviometric data. The study area encompasses the North of Minas Gerais, a dry region with high temperatures and limited water availability. Data from 99 rain gauge stations were used for the year 2022, 68 of which are within the northern mesoregion and the other 31 outside it, within a 50 km buffer. The processing involved the use of IDW and Spline methods, with cross-validation and accuracy metrics including Mean Error (ME), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The descriptive analysis of the data revealed significant variability in precipitation. Cross-validation metrics indicated a slight superiority of Spline over IDW. It is concluded that Spline is more effective in interpolating precipitation data in the region, and it is recommended to explore geostatistical methods, such as Kriging, in future studies for a more comprehensive approach.

Keywords: Interpolation; Deterministic models; Cross Validation; Precipitation.

1 - INTRODUÇÃO E OBJETIVOS

A interpolação pode ser compreendida como aplicação de métodos matemáticos e estatísticos para estimar valores desconhecidos, dentro de um conjunto de dados conhecidos ou amostrados. Em um universo limitado de dados, como um conjunto de pontos, a interpolação pode ser empregada para calcular ou estimar valores intermediários entre estes pontos, constituindo uma técnica particularmente valiosa na espacialização de uma informação de natureza contínua, pois permite a estimativa do atributo em qualquer local dentro dos limites dos dados coletados (AZPURUA; DOS RAMOS, 2010).

Existem na literatura diversos estudos que fazem uso dessa técnica, nos quais abrangem uma ampla variedade de métodos aplicáveis a dados ambientais expandindo o interesse da técnica em diferentes áreas de estudo podendo reduzir custos, simplificar processos e otimizar o tempo, tais como os apresentados nas pesquisas de Pereira e Macedo (2017), Fernandes e Macedo (2018), Monteiro et al. (2024), Fernandes et al. (2024), Purnomo et al. (2024), Wang et al. (2024). Esses métodos podem ser categorizados em dois grupos principais: o modelo geostatístico e os modelos determinísticos (BARRENA-GONZÁLEZ; LAVADO CONTADOR; PULIDO FERNÁNDEZ, 2022).

A interpolação de dados ambientais desempenha um papel crucial em contextos onde há escassez de tecnologias que não são capazes de mapear toda a superfície terrestre; a utilização desse método é fundamental para resolver o problema de dados georreferenciados ausentes; mesmo que essas técnicas enfrentem algumas limitações, tais como a estimativa de erros, elas ainda são muito úteis (AMINI et al. 2019).

Em dados de precipitação a aplicação da técnica se torna muito útil, pois cria uma superfície contínua de precipitação, proporcionando uma análise mais precisa e detalhada que pode melhorar a acurácia de modelos hidro meteorológicos, tendo em vista que a precipitação pode variar significativamente em pequenas escalas espaciais (FOEHN et al. 2018; SHARIFI, SAGHAFIAN, STEINACKER, 2019).

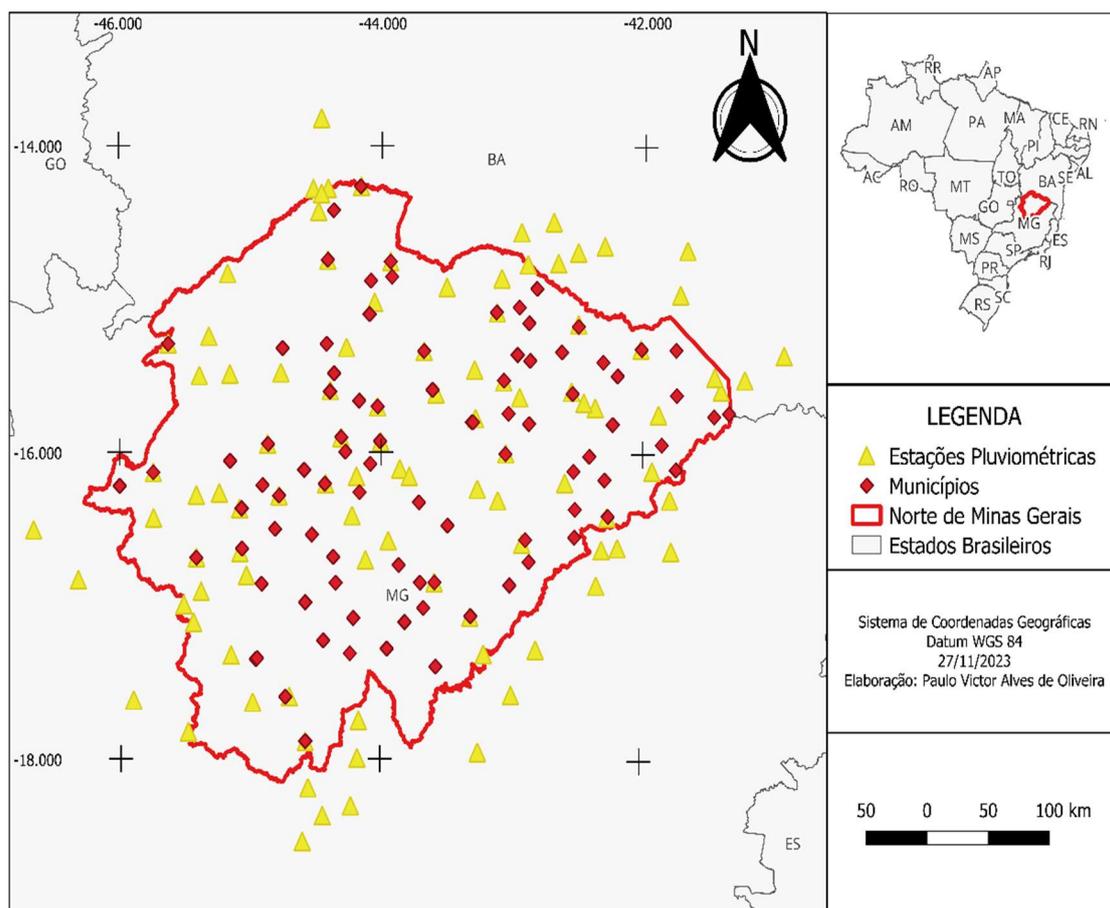
Neste contexto, o presente trabalho tem como objetivo principal avaliar a eficácia de dois métodos de interpolação determinísticos, o Inverse Distance Weighting (IDW) e o SPLINE, na análise de dados pluviométricos da região Norte de Minas Gerais coletados durante o ano de 2022. A escolha do método mais adequado será fundamentada na comparação de métricas estatísticas, visando identificar qual deles proporciona resultados mais precisos e robustos para a espacialização dessas informações.

2 – METODOLOGIA

2.1 - ÁREA DE ESTUDO

O Norte de Minas Gerais é uma das doze mesorregiões que compartmentam o território mineiro. Como o nome já indica é a região mais setentrional do estado, composta por 89 municípios dispostos em uma área de 127.816,15km², com uma população aproximada de 1.610.413 habitantes em 2010 (AMMESF, 2023; IBGE, 2018). O clima norte mineiro é caracterizado por ser seco, possuir temperaturas elevadas e pouca disponibilidade hídrica. Estes fatos somados a problemas econômicos e sociais como municípios com economias pouco diversas, populações pequenas e baixa renda, tornam a região um alvo da vulnerabilidade ocasionada pelas mudanças climáticas (FREITAS; CALHEIROS; DOS REIS, 2019). Na Figura 1 é possível ver relação entre os municípios que constituem a região e as estações pluviométricas com dados úteis em 2022.

Figura 1. Mapa de municípios e estações pluviométricas.



Fonte: Elaborado pelos autores.

2.2-PROCESSAMENTO

Para as análises, foram utilizados os dados de 99 estações pluviométricas presentes no entorno e no interior do perímetro do Norte do estado de Minas Gerais, das quais 68 estão dentro dos limites da mesorregião norte e as demais 31 estão situadas dentro de uma área de influência de 50 km fora região Norte do estado. Os dados das estações pluviométricas funcionais foram obtidos pelo banco de dados da ANA – Agência Nacional de Águas, através de um plugin do software QGIS chamado ANA Data Aquisition. A base da ANA oferece os dados de precipitação diária e o período selecionado para o estudo foi o ano de 2022.

Ao todo 548 estações pluviométricas foram obtidas, cujo 360 estações se localizam no interior da Mesorregião Norte e as demais 188 estavam na região da área de influência; as 449 estações não utilizadas foram descartadas no pré-processamento por não apresentarem valores válidos de precipitação, seja por falhas de medição, desuso dos equipamentos, dentre outros problemas que podem ter ocorrido nestas estações.

O pré-processamento foi realizado retirando as estações sem registro de valores, logo em seguida as estações consistentes foram processadas no Microsoft Excel (Microsoft, 2019), onde os dados foram organizados, falhas de entradas corrigidas e a precipitação anual pôde ser calculada somando os valores diários.

Com os dados de precipitação espacializados em ambiente do software ArcMap10.8 (ESRI, 2019);o processamento prosseguiu utilizando a ferramenta *Subset Features* da extensão *Geoestatistical Wizard* (ESRI, 2019).Essa ferramenta permite a separação da base de dados em dois grupos, sendo um deles utilizado para a produção dos modelos de interpolação e o outro para a comparação e validação dos resultados; para a separação da base de dados, 70 % das estações foram utilizadas para o treinamento dos modelos e os 30 % restantes para validação.

Os modelos de interpolações também foram realizados utilizando a extensão *Geoestatistical Wizard* (ESRI, 2019), os dois métodos são: o IDW - Inverso do Quadrado da Distância e o SPLINE. Ao fim das análises espaciais os resultados foram novamente processados no Microsoft Excel (Microsoft, 2019) para o cálculo das métricas de acurácia restantes.

2.1. – IDW

Os interpoladores do tipo IDW – Inverse Distance Weighting, utilizam um método que atribui pesos inversamente proporcionais à distância dos pontos conhecidos. A premissa é que os pontos interpolados são mais influenciados pelos pontos mais próximos e menos afetados pelos pontos mais distantes; isto significa que os pontos mais próximos têm um peso maior na estimativa do valor desconhecido do que os pontos mais distantes; o IDW é um método local, exato e determinístico (WANG et al. 2014; BORGES et al. 2016; HADI & TOMBUL,2018). A formula geral do IDW é dada pela Equação 1:

$$\hat{Z}(s_0) = \sum_{i=0}^N \lambda_i Z(s_i) \quad (1)$$

Onde:

$\hat{Z}(s_0)$ = Valor estimado na localização s_0 ;

N = Número de pontos localizados ao redor do ponto a ser estimado usado no cálculo;

λ_i = Peso correspondente a cada ponto conhecido.

O peso é calculado utilizando a Equação 2:

$$\lambda_i = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (2)$$

Onde:

P = Potência: controla a influência da distância entre os pontos de valor estimado;

N = Número de pontos usados na estimação

d_i = Distância entre o ponto a ser estimado e o ponto conhecido

Os parâmetros utilizados na interpolação do IDW foram potência, $p = 2$, com um tipo de vizinhança padrão e um número mínimo e máximo de vizinhos de 10 e 15 respectivamente.

2.2 – SPLINE

Os interpoladores do tipo SPLINE estimam valores utilizando uma sequência de polinômios definidos em intervalos adjacentes, apresentando derivadas contínuas nos extremos desses intervalos, garantindo a continuidade das derivadas nos pontos extremos desses intervalos, semelhante à citação original fornecida. (WOLD, 1974).

O SPLINE se baseia em duas premissas básicas; a primeira é que, a função interpoladora obrigatoriamente deve passar pelos pontos amostrados, a segunda é que o modelo busque sempre ser o mais suave possível. Em síntese o spline busca por meio de uma abordagem matemática gerar uma curva suave que se adapte aos dados, assegurando que a função resultante passe por todos os pontos específicos e mantenha uma transição suave em termos de derivadas (AMINI et al. 2019).

Os modelos do tipo RBF (Radial Basis Functions) tendem a ser mais eficazes em superfícies suaves, quando o tamanho dos dados de amostra disponível é maior. O desempenho é menos otimizado para superfícies com variações apreciáveis que abrangem curtos intervalos. Em discordância com outros métodos, como o IDW, os modelos RBF não forçam as estimativas a manterem a faixa dos dados amostrados (IKECHUKWU et al. 2017).

No processamento SPLINE foi utilizado o *Radial Basis Function* (RBF), Kernel de Spline Completamente Regularizado, com um parâmetro de Kernel = 0,0006863 e um número mínimo e máximo de vizinhos de 10 e 15, assim como no IDW.

2.3- VALIDAÇÃO E ACÚRACIA

Além da Validação Cruzada (Cross Validation) fornecida pelo ArcMap ao fim do processamento, foram utilizados outros quatro índices de avaliação já muito difundidos como: o Mean Error – ME (Erro Médio), Mean Absolute Error – MAE (Erro Absoluto Médio), Mean Squared Error – MSE (Erro Quadrático Médio) e Root Mean Squared Error – RMSE (Raiz do Erro Quadrático Médio); Das equações 3 a 6 é possível verificar as expressões dos índices, onde n é o número de observações, p são os valores estimados e o são os valores observados:

ME (3)

$$ME = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)$$

MAE (4)

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i|$$

MSE (5)

$$MSE = \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2$$

RMSE (6)

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \right]^{1/2}$$

O Erro Médio (ME) é frequentemente utilizado para avaliar o viés nas estimativas, mas sua compensação entre estimativas positivas e negativas pode levar a um ME menor do que o erro real, indicando cautela em sua interpretação como indicador de precisão (IKECHUKWU et al. 2017; LI & HEAP, 2008; ISAAKS et al. 1989).

Erro Quadrático Médio (RMSE) fornece uma medida do tamanho do erro, sendo sensível a valores atípicos (*outliers*) devido ao peso atribuído a erros grandes; tanto o RMSE quanto o Erro Quadrático Médio (MSE) compartilham essas limitações, já o Erro Médio Absoluto (MAE) é menos sensível a valores extremos, indicando a extensão do possível erro na estimativa (ULIANA, 2020; LI & HEAP, 2008; WILLMOTT, 1982).

Apesar de MAE e RMSE serem medidas semelhantes, fornecendo estimativas do erro médio, elas não oferecem informações sobre o tamanho relativo da diferença média e a natureza das diferenças; outra característica sobre MAE é a diferença absoluta média entre os dados observados e estimados, ou seja, não sofre alterações em seus resultados provenientes de outlier, nos valores analisados (STUCCHI, 2024; NALDER & WEIN, 1998; WILLMOTT, 1982).

A validação cruzada é empregada em conjunto com essas medidas para avaliar o desempenho dos métodos de interpolação, nela as informações referentes a um ponto são temporariamente removidas, sendo então estimadas a partir dos demais pontos de dados, e a diferença entre os valores reais e estimados é calculada (AMINI et al. 2019; BROWNE,

2000). Baseando-se nesta premissa a análise de acurácia foi feita para os valores de ME e RMSE.

3 – RESULTADOS E DISCUSSÃO

Os resultados da análise descritiva de dados estatísticos podem ser visualizados na Tabela 1; a análise apresenta um conjunto com alta amplitude variando entre 62,9 a 2440 mm nas estações, os dados apresentam uma média de 773,80 mm com desvio padrão de 345,87; isso sugere que os valores individuais no conjunto de dados estão mais distantes da média, indicando uma maior instabilidade ou variabilidade nos dados.

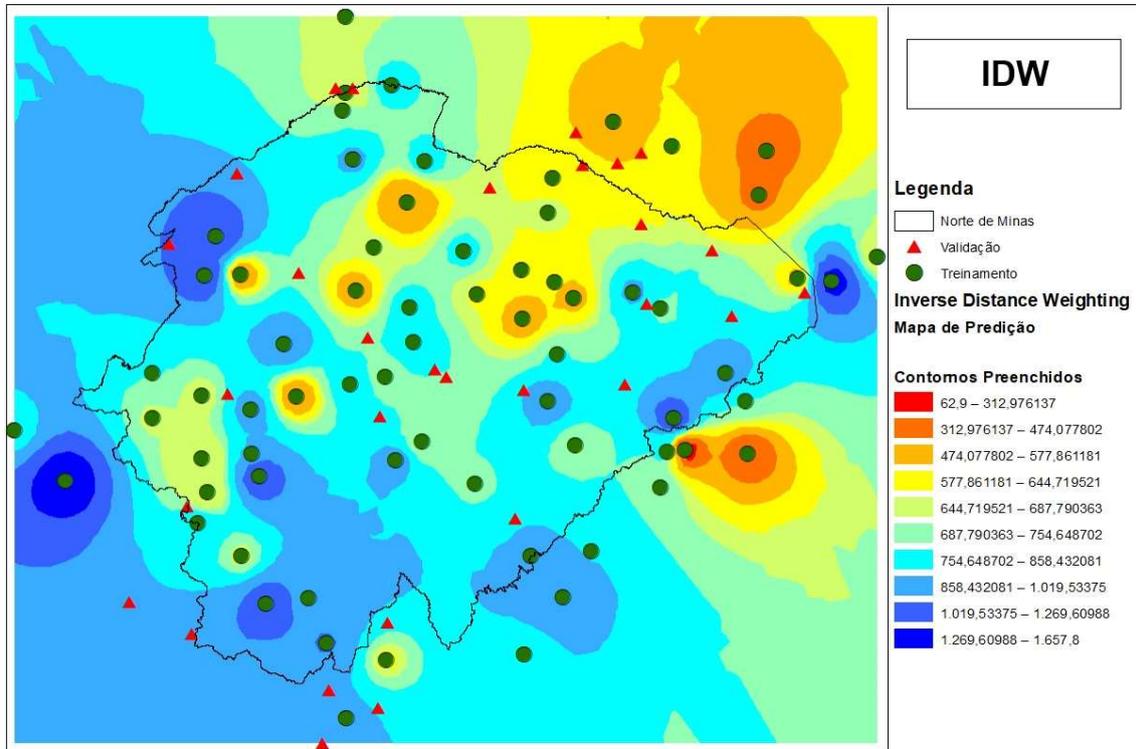
Tabela 1. Estatística Descritiva

Métrica	Valor
Média	773,8070707
Erro padrão	34,76174335
Mediana	754,1
Desvio padrão	345,8749793
Variância da amostra	119629,5013
Curtose	4,737961667
Assimetria	1,287901611
Intervalo	2377,1
Mínimo	62,9
Máximo	2440
Contagem	99

Fonte: Elaborado pelos autores.

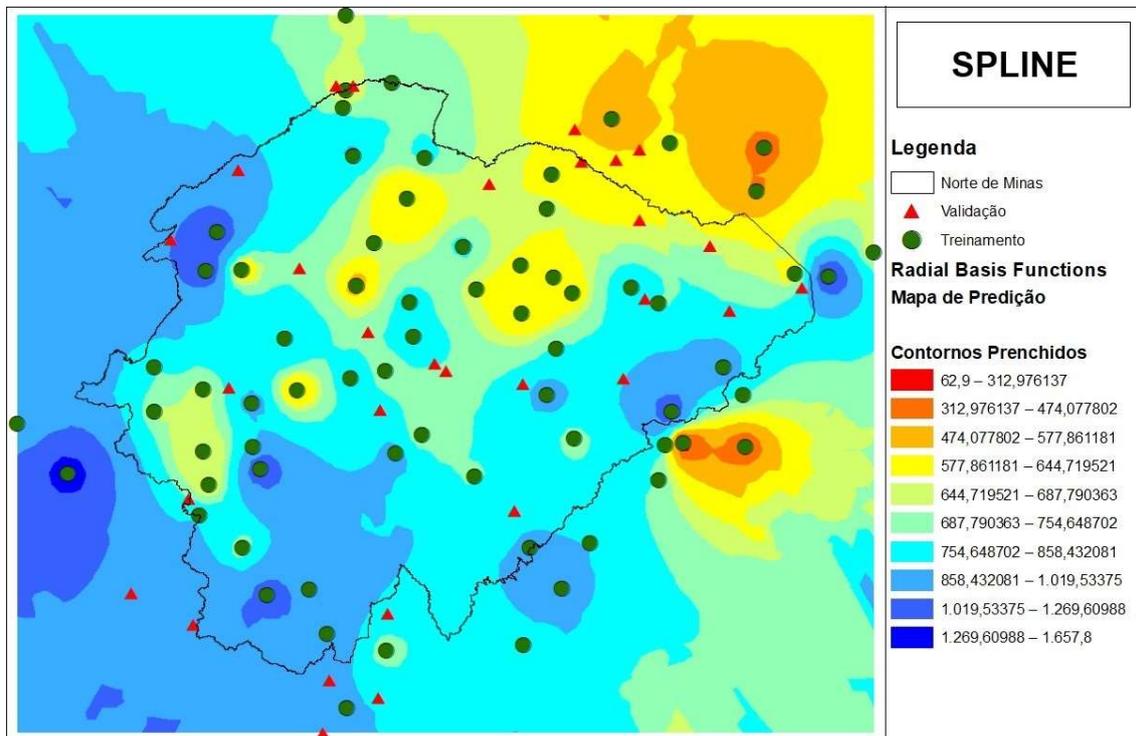
As Figuras 2 e 3 apresentam os resultados de espacialização dos interpoladores utilizados, nota-se que ambos apresentam resultados similares, mas a espacialização com SPLINE apresenta uma suavização de superfície um pouco mais homogênea do que a espacialização IDW.

Figura 2. Resultado de espacialização com o Interpolador IDW.



Fonte: Elaborado pelos autores.

Figura 3. Resultado de espacialização com o Interpolador SPLINE.



Fonte: Elaborado pelos autores.

Os resultados mostram que os valores de RMS foram altos em ambos os interpoladores o que indica um desempenho ruim na previsão ou ajuste dos dados. Os valores de Cross Validation (Validação Cruzada) do SPLINE se saem ligeiramente melhores, apresentando valores de Root Mean Squared e média mais baixos do que os obtidos no IDW, um desempenho similar pode ser visualizado nas métricas de acurácia mesmo a diferença sendo pouca em relação de números absolutos, eles também são elevados; os dados de interpolação do SPLINE apresentam um menor ME e RMSE (Tabela 2). De acordo com Shope et al. (2015) e Hadi & Tombul (2018), quando os valores de RMSE são maiores, indicam que o método utilizado operou de forma pior.

Tabela 2. Resultado das Métricas de Acurácia e Validação

Métrica	IDW	SPLINE
Média*	-7,996	-2,940
RMS*	338,108	327,508
ME	40,375	28,866
MAE	288,148	285,750
MSE	189074,184	188203,205
RMSE	434,827	433,824

Fonte: Elaborado pelos autores.

*Valores de Cross Validation fornecidos pela análise do ArcMap.

4 - CONSIDERAÇÕES FINAIS

As tecnologias matemáticas e estatísticas capazes de suprir essas lacunas existentes nas medições de campo são importantes, tornando-se indispensáveis em um cenário de poucos dados existentes devido à infraestrutura deficiente que se dispõe atualmente; a região em questão se encontra no limite do semiárido brasileiro, então seu monitoramento se faz necessário para verificar possíveis mudanças nos padrões fluviométricos locais.

Em suma, com base nos textos da bibliografia; a análise dos resultados das métricas de acurácia deste trabalho que corroboram com a literatura corrente conclui que o interpolador SPLINE produziu resultados melhores, apresentando um erro menor quando comparado ao interpolador IDW para este conjunto de dados de precipitação.

Ainda com base nos resultados vale ressaltar que a utilização do método geostatístico de interpolação; aplicando os outros diferentes tipos de Krigagem podem apresentar uma nova abordagem para os dados, o que pode ser feito em trabalhos futuros.

Adicionalmente com este estudo é possível obter um cenário das atuais estações em funcionamento, isto pode auxiliar o poder público no fomento de projetos que visem a implementação de novas estações em locais mais estratégicos ou na manutenção das estações já existentes.

5 – REFERÊNCIAS BIBLIOGRÁFICAS

AMINI, M. A. et al. Analysis of deterministic and geostatistical interpolation techniques for mapping meteorological variables at large watershed scales. **Acta Geophysica**, v. 67, n. 1, p. 191-203, 2019.

AMMES. Associação dos Municípios da Bacia do Médios São Francisco. Disponível em <<http://ammesf.org.br/municipios/>>; acesso em: 28 nov. 2023.

AZPURUA, M. A.; DOS RAMOS, K. A comparison of spatial interpolation methods for estimation of average electromagnetic field magnitude. **Progress In Electromagnetics Research M**, v. 14, p. 135-145, 2010.

BARRENA-GONZÁLEZ, J.; LAVADO CONTADOR, J. F.; PULIDO FERNÁNDEZ, M. Mapping Soil Properties at a Regional Scale: Assessing Deterministic vs. Geostatistical Interpolation Methods at Different Soil Depths. *Sustainability*, v. 14, n. 16, p. 10049, 2022.

BORGES, P. A. et al. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. **Theoretical and applied climatology**, v. 123, p. 335-348, 2016.

BROWNE, M. W. Cross-validation methods. **Journal of mathematical psychology**, v. 44, n. 1, p. 108-132, 2000.

ESRI. ArcGIS for Windows Version 10.8. License type ArcInfo. [S.I]: ESRI – Environmental Systems Research Institute, 2019.

FERNANDES, A. C. G. et al. Estimation of average annual precipitation considering spatio-temporal variability through the use of multiple linear regression in the State of Paraíba. **Environmental & Social Management Journal/Revista de Gestão Social e Ambiental**, v. 18, n. 4, 2024.

FERNANDES, L. C.; MACEDO, D. R. Interpolação de dados meteorológicos utilizando covariáveis para a Região Metropolitana de Belo Horizonte. **Cadernos do Leste**, [S. l.], v. 17, n. 17, 2018. DOI: 10.29327/249218.17.17-1. Disponível em: <https://periodicos.ufmg.br/index.php/caderleste/article/view/13126>. Acesso em: 6 ago. 2024.

FREITAS, L. O.; CALHEIROS, T.; DOS REIS, R. J. Vulnerabilidade da mesorregião Norte de Minas Gerais face às mudanças climáticas\The vulnerability of the Northern Minas Gerais mesoregion in the face of climate change. **Caderno de Geografia**, v. 29, n. 56, p. 134-134, 2019.

FOEHN, A. et al. Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. **Journal of Hydrology**, v. 563, p. 1092-1110, 2018.

HADI, S. J.; TOMBUL, M. Comparison of spatial interpolation methods of precipitation and temperature using multiple integration periods. **Journal of the Indian Society of Remote Sensing**, v. 46, p. 1187-1199, 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cidades: Estimativa da População. 2018.** Brasília: IBGE 2010. Disponível em: <<http://www.cidades.ibge.gov.br/xtras/home.php?lang=>>; acesso em: 28 nov. 2023.

IKECHUKWU, M. N. et al. Accuracy assessment and comparative analysis of IDW, spline and kriging in spatial interpolation of landform (topography): an experimental study. **Journal of Geographic Information System**, v. 9, n. 3, p. 354-371, 2017.

ISAACS, E. H. et al. Applied geostatistics. 1989.

LI, J.; HEAP, A. D. A review of spatial interpolation methods for environmental scientists. 2008.

MICROSOFT OFFICE 2019 Project for Windows. [S. l.]: Microsoft Corporation, 2019.

MONTEIRO, F. et al. Análise espacial das chuvas no Estado do Rio Grande do Norte, Brasil. **Revista Geográfica de América Central**, n. 72, p. 59-78, 2024.

NALDER, I. A.; WEIN, R. W. Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. **Agricultural and forest meteorology**, v. 92, n. 4, p. 211-225, 1998.

PEREIRA, A. L. S.; MACEDO, D. R. Análise de métodos de interpolação para o mapeamento da precipitação na bacia do Reservatório de Nova Ponte/MG. **Cadernos do Leste**, [S. l.], v. 17, n. 17, 2017. DOI: 10.29327/249218.17.17-3. Disponível em: <https://periodicos.ufmg.br/index.php/caderleste/article/view/13128>. Acesso em: 6 ago. 2024.

PURNOMO, H. et al. Comparison of Global and Local Deterministic Interpolation Methods to Predict Spatial Distribution of Ni, Fe and MgO in Nickel Laterite Deposits. **Jurnal Geografi Gea**, v. 24, n. 1, p. 75-90. Disponível em: <https://ejournal.upi.edu/index.php/gea/article/download/61886/27104>. Acesso em: 6 ago. 2024.

SHARIFI, E.; SAGHAFIAN, B.; STEINACKER, R. Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. **Journal of Geophysical Research: Atmospheres**, v. 124, n. 2, p. 789-805, 2019.

SHOPE, C. L. et al. Modeling spatiotemporal precipitation: Effects of density, interpolation, and land use distribution. **Advances in Meteorology**, v. 2015, 2015.

STUCCHI, G. Validação de medidas de evapotranspiração coletadas por meio de um lisímetro de lençol freático constante. 2024.

ULIANA, E. M. et al. MODELOS HIDROLÓGICOS SAC-SMA E IPH II: CALIBRAÇÃO E AVALIAÇÃO DO DESEMPENHO NA ESTIMATIVA DE VAZÕES NA BACIA DO RIO PIRACICABA (MG). **IRRIGA**, v. 25, n. 2, p. 202-222, 2020.

WANG, S. et al. Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada. **International Journal of Climatology**, v. 34, n. 14, p. 3745-3751, 2014.

WANG, Y. et al. Research Progress on Spatiotemporal Interpolation Methods for Meteorological Elements. **Water**, v. 16, n. 6, p. 818, 2024.

WILLMOTT, C. J. Some comments on the evaluation of model performance. **Bulletin of the American Meteorological Society**, v. 63, n. 11, p. 1309-1313, 1982.

WOLD, S. Spline functions in data analysis. **Technometrics**, v. 16, n. 1, p. 1-11, 1974.