



Proposta de um modelo de processo para aplicação de métodos de interpolação

Gustavo E. Kida¹, Raul T. Santos¹

¹Instituto de Computação – Universidade Federal do Mato Grosso (UFMT)
Cuiabá – MT – Brasil

gustavokida@live.com, raul@ic.ufmt.br

Abstract. *In the interpolation of spatial data there are several methods and each one of them have some advantages in specific scenarios. Furthermore, some methods needs treatments to perform a reliable interpolation, such as removing outliers. Hence, this work objective is to present a process model of the application of interpolation methods, serving as a good practice guide. The results demonstrate the efficacy in the maps quality, obtaining the best results when the process model is followed.*

Resumo. *Na interpolação de dados espaciais existem vários métodos e cada um deles possui algumas vantagens em cenários específicos. Além disso, alguns métodos precisam de tratamentos para realizar uma interpolação confiável, como a remoção de outliers e anisotropia. Dessa forma, o objetivo deste trabalho é propor um modelo de processo para aplicação de métodos de interpolação, servindo como um guia de boas práticas. Os resultados demonstram a efetividade dos tratamentos na qualidade dos mapas, obtendo o melhor resultado quando seguido o modelo de processo.*

1. Introdução

A Agricultura de Precisão busca atingir a melhor produtividade de forma sustentável, utilizando a tecnologia como apoio. Entre suas várias etapas é realizada a interpolação de amostras coletadas de uma área, a fim de reduzir o custo da coleta e produzir com isso um mapa de zonas de manejo [Bottega et al. 2017]. Devido a grande quantidade de métodos de interpolação é necessário realizar uma análise dos dados para decidir o método e tratar os dados de acordo com os requisitos do método selecionado [Webster and Oliver 2007; Li and Heap 2008; Adhikary et al. 2017; Li et al. 2015; Pebesma et al. 2011]. O objetivo desse artigo é propor um modelo de processo para aplicação adequada de métodos de interpolação, visando contribuir para obtenção de melhores mapas de

zonas de manejo, e conseqüentemente melhorando a produtividade do campo, reduzindo o uso de insumos e evitando o desperdício.

2. Metodologia

O modelo proposto foi baseado em uma revisão da literatura de livros e artigos com palavras chaves como “geostatistics”, “spatial interpolation”, “spatial interpolation methods”, “kriging” e “anisotropic inverse distance weighted”, buscando aqueles que detalhavam o processamento dos dados para melhorar os resultados de interpolações espaciais [Webster and Oliver 2007; Li and Heap 2008; Oliver and Webster 2014; Pebesma et al. 2011; Merwade et al. 2006]. Para a especificação do modelo de processo foi utilizado o *Business Process Model and Notation* (BPMN), oferecendo uma notação simples e de fácil entendimento [OMG 2011].

2.1. Construção do modelo

O modelo inicia-se pela remoção de outliers. Essa atividade é primordial para aumentar a precisão, pois os outliers podem gerar efeitos negativos substanciais nas análises estatísticas [Osborne and Overbay 2004]. Após essa etapa o modelo verifica se existem mais de 50 pontos de dados, considerado o mínimo para criar um variograma [Webster and Oliver 2007].

Quando houver menos de 50 pontos será feito a verificação da presença de anisotropia, característica que ocorre quando a distribuição espacial de um elemento varia mais intensamente para uma direção e menos em outra [Camargo et al. 2001]. Esse problema pode ser tratado realizando uma transformação das coordenadas, tornando-os isotrópicos [Boisvert 2010]. Após isso é utilizado o IDQ, método que interpola os dados atribuindo pesos para cada amostra com base na distância em que estas se encontram do ponto a ser interpolado [Alvarenga et al. 2010], e então é revertida a transformação da anisotropia caso tenha ocorrido e finalizado o processo.

Já para quando houver mais de 50 pontos, verifica-se a distribuição normal utilizando o teste de assimetria, considerando assimétricos os valores maiores que 1,0 [Webster and Oliver 2007]. Caso o conjunto de dados seja assimétrico deve ser feita uma transformação para distribuição normal [Oliver and Webster 2014]. A próxima etapa é construir um variograma e verificar se há tendência nele, ou seja, o variograma não deve crescer indefinitivamente. Caso exista tendência serão calculados os resíduos e verificado a presença de anisotropia nos resíduos, se houver anisotropia é realizada a remoção e em seguida a interpolação com o método Krigagem Universal (KU), que tem como característica conseguir tratar a tendência [Webster and Oliver 2007]. Após realizada a interpolação nos dados transformados é necessário reverter a transformação de assimetria [Webster and Oliver 2007].

Não existindo tendência, a anisotropia é analisada da mesma forma que nos casos anteriores, se não for encontrada então será gerado um variograma para analisar a presença do efeito pepita puro, que se caracteriza por um variograma em que o efeito pepita tem o mesmo valor do patamar [Webster and Oliver 2007; Oliver and Webster 2014]. Se não for detectado esse efeito será realizada a KO, a qual leva em conta a correlação espacial dos dados ao utilizar um variograma para determinar os pesos dos pontos vizinhos ao ponto que será estimado [Requia et al. 2019], então é revertida a transformação de assimetria e finalizado o processo.

Caso seja encontrada a anisotropia é preciso verificar se existem ao menos 250 pontos de dados, pois é a quantidade de amostras recomendadas para realizar o tratamento do variograma [Webster and Oliver 2007]. Se não houver 250 pontos, será desfeita a transformação de assimetria se foi realizada e o processo seguirá o caminho de testar a anisotropia e utilizar o método IDQ. Se existirem pontos suficientes será feita a remoção da anisotropia e então realizado a KO, finalizando o processo após reverter a transformação de assimetria se foi feita. Por fim, para ambos os casos em que o efeito pepita puro for detectado, será desfeita a transformação de assimetria e então o processo seguirá o caminho de testar a anisotropia e interpolar utilizando o IDQ.

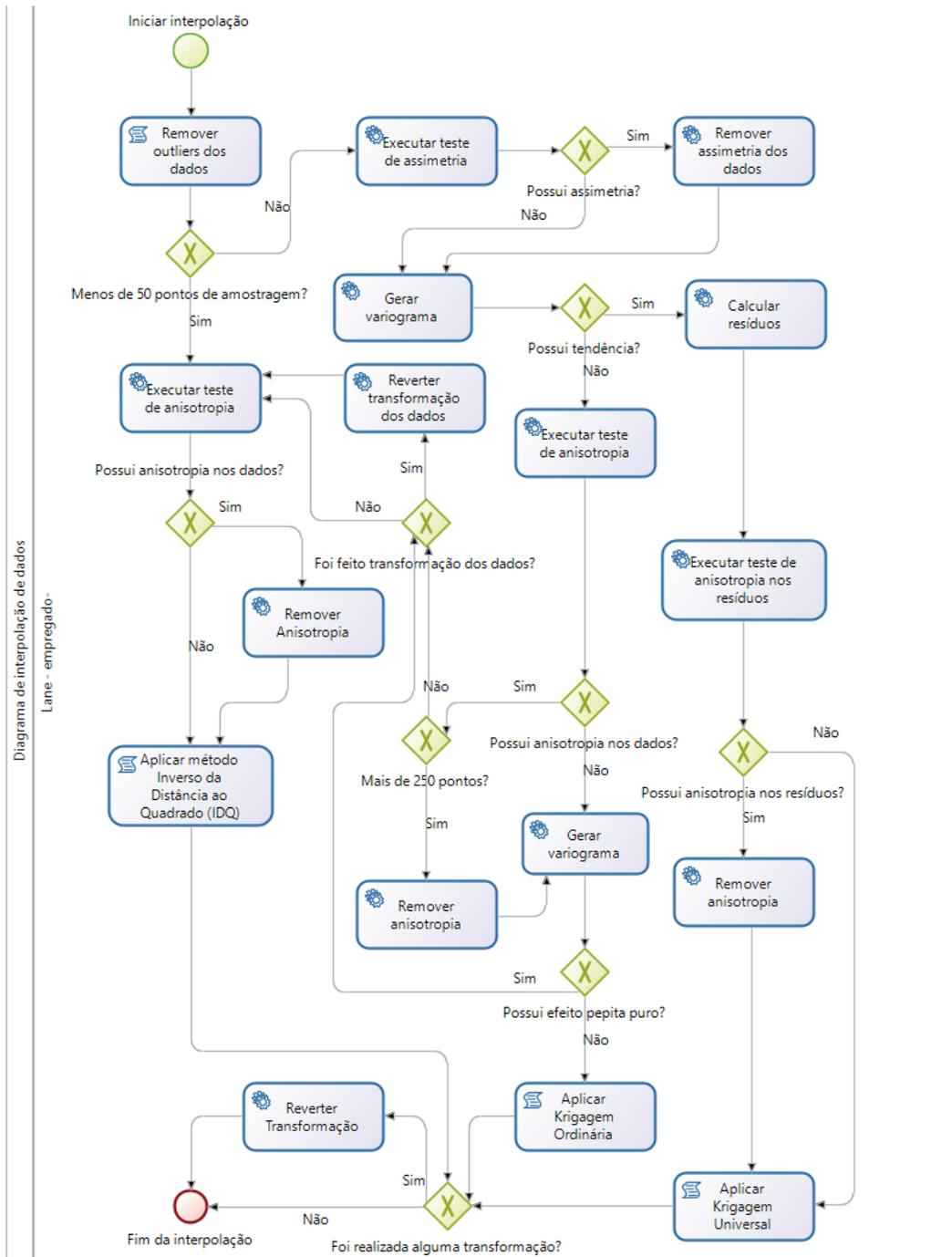


Figura 1. Modelo de processo para aplicação de métodos de interposição

O modelo está representado pela Figura 1, seguindo todas as etapas que foram explicadas.

2.2. Validação

O modelo foi validado utilizando validação cruzada pelo método *leave-one-out*, aplicando o *Root Mean Square Error* (RMSE) para verificar a acurácia, sendo que quanto mais próximo de 0, melhor a precisão. Os testes foram feitos seguindo o modelo proposto na Figura 1 para interpolar o conjunto de dados disponíveis para o projeto de pesquisa e comparados com os resultados dos testes que não seguem o modelo, utilizando o software R [Team 2013].

A validação utilizou um conjunto de dados de solo de um talhão de 133,96 ha, coletado próximo ao município de Tibagi no estado do Paraná, possuindo 62 pontos com distanciamento de 165 metros aproximadamente. Os atributos Acidez potencial (H+Al) e Fósforo (P) foram selecionados para a validação.

3. Estudo de caso

No atributo H+Al a remoção de outliers removeu 4 pontos, utilizando o método boxplot por ser mais estável [Bento and Santos 2018]. Já a assimetria foi de 1,904404 para o conjunto sem tratamentos e 0,8529702 para o conjunto com tratamentos, não necessitando transformar a assimetria. A anisotropia foi testada e não foi encontrada nos dados tratados.

Dessa forma, o caminho correto do modelo de processo é realizado pelo tratamento de outliers, verificação da assimetria, tendência no variograma, anisotropia e efeito pepita puro. Por fim os dados são interpolados com a KO.

Após a interpolação foi realizada a validação cruzada e foram obtidos valores de RMSE do atributo H+Al conforme a Tabela 1. A precisão melhorou 38,8% em relação ao IDQ sem tratamentos e 32,4% quando comparado com a KO sem tratamentos.

Tabela 1. Resultados RMSE da interpolação do atributo H+Al

Sem tratamento (KO)	11,10533
Sem tratamento (IDQ)	12,27034
Seguindo o modelo	7,516667 (KO)

A Figura 2 mostra os variogramas finais utilizados para a krigagem do atributo H+Al, e todos apresentaram um variograma aceitável, sem tendência ou efeito pepita [Oliver and Webster 2014].

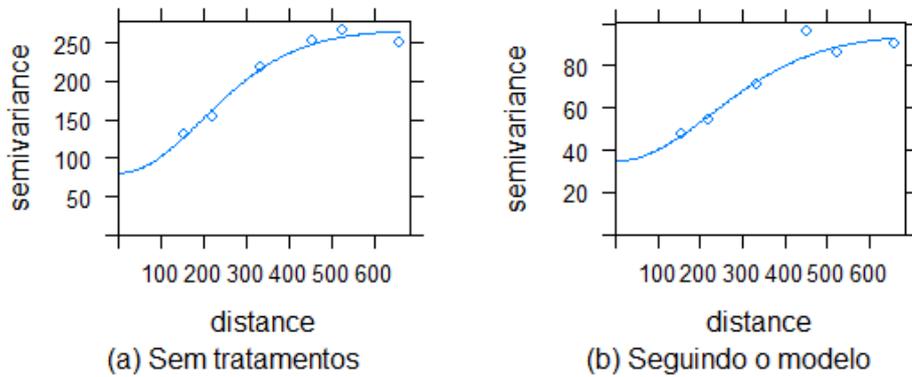


Figura 2. Variogramas do atributo H+AI

A Figura 3 apresenta os mapas interpolados do atributo H+AI, sendo que a Figura 3c foi considerada a melhor pela validação.

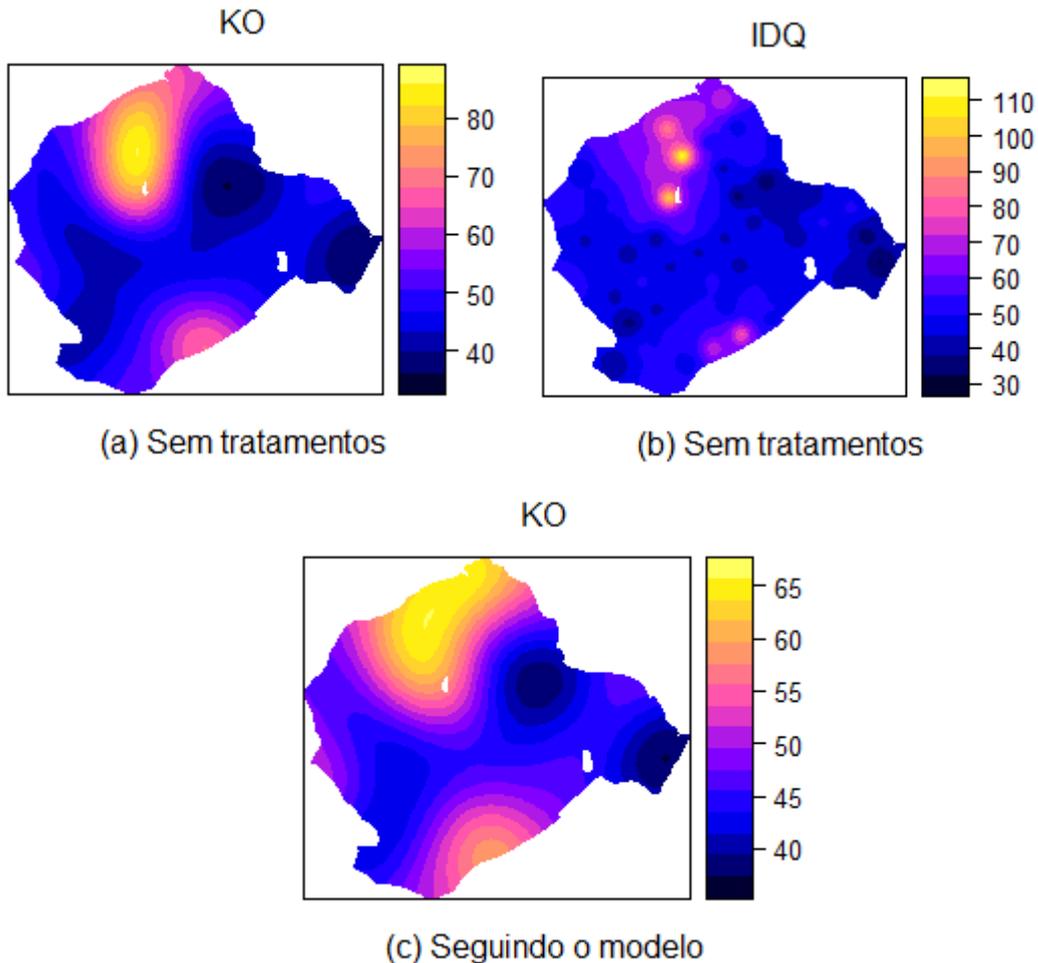


Figura 3. Mapas da interpolação do atributo H+AI

Para o atributo P, a remoção de outliers resultou em apenas 1 ponto removido, sobrando 61 pontos. A assimetria do atributo P foi de 0,764862 para o conjunto sem outliers, sendo assim, não houve a necessidade de uma transformação Box-Cox. O

software encontrou anisotropia nos dados sem outliers e retornou os valores para o ângulo de $-38,785155$ e $1,721492$ para razão. Assim, o modelo realizou a remoção de outliers, não encontrou assimetria e nem tendência, mas detectou a anisotropia. Como não há mais de 250 pontos a KO não foi aplicada, por isso foi feita a verificação de anisotropia no caminho para o IDQ, removeu a anisotropia detectada, realizou a interpolação com o método IDQ e reverteu as transformações.

Feito os testes, foram observados e comparados os valores RMSE obtidos pela validação cruzada apresentados na Tabela 2, demonstrando um aumento na acurácia seguindo o modelo de 12,1% em relação ao método KO e 9,5% se comparado ao IDQ, ambos os métodos sem tratamentos.

Tabela 2. Resultados RMSE da interpolação do atributo P

Sem tratamento (KO)	25,9916
Sem tratamento (IDQ)	25,23343
Seguindo o modelo	22,8612 (IDQ)

A Figura 4 mostra o variograma gerado para ser utilizado na KO do atributo P e apresentou um patamar estável.

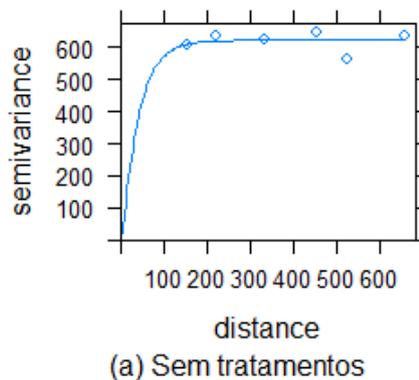


Figura 4. Variograma do atributo P

A Figura 5 mostra os mapas gerados pela interpolação de cada teste. Dentre os 3 testes, o melhor resultado é apresentado na Figura 5c que seguiu o modelo, pois obteve o menor valor de RMSE.

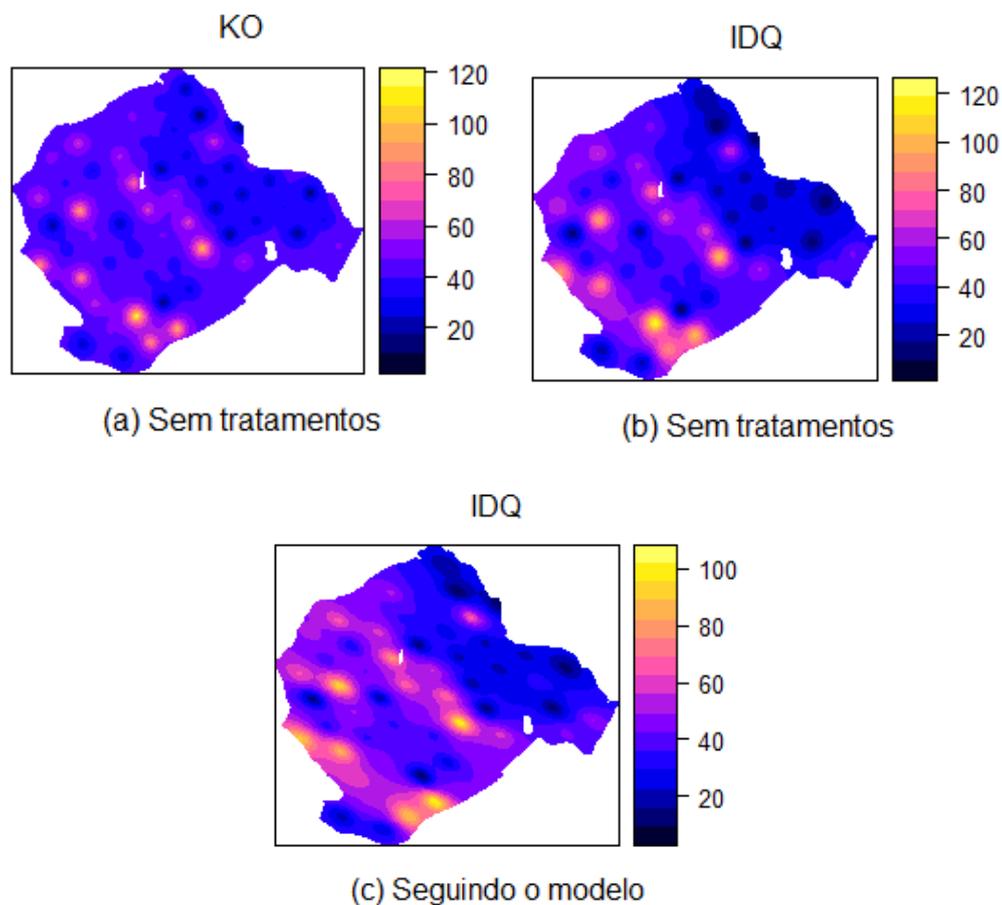


Figura 5. Mapas da interpolação do atributo P

4. Conclusões e trabalhos futuros

Os resultados obtidos seguindo o modelo foram melhores para ambos os atributos H+Al e P, demonstrando a aplicabilidade na área de ciências da terra, como na gestão ambiental que geralmente coletam pontos de uma área, mas necessitam de um mapa de superfície contínuo [Li and Heap 2008], na criação de mapas de zonas de manejo para reduzir a variância dos dados [Schenatto et al. 2016] e na redução de custos de amostragem de acordo com a estrutura espacial [Li and Heap 2008]. No atributo H+Al a melhora foi significativa utilizando apenas a remoção de outliers, já no atributo P houve um pequeno aumento na precisão com os tratamentos de anisotropia e outliers.

Os resultados também evidenciam que a remoção de outliers sempre melhora os resultados da interpolação, independente do atributo ou do método utilizado, além disso, no atributo H+Al a remoção de outliers removeu a assimetria como consequência, podendo assim reduzir a complexidade de todo o processo de interpolação e consequentemente contribuir para a redução de erros humanos.

Dessa forma, os resultados permitem concluir que o modelo consegue reduzir erros e escolher o melhor método de interpolação e seus tratamentos necessários, portanto é recomendado seguir o modelo de processo, servindo como um guia de boas práticas para a interpolação de mapas, permitindo reduzir os custos de amostragem de solo e interpolar mapas com a melhor representação possível da área.

Como trabalho futuro pretende-se validar as etapas que não foram utilizadas, buscando dados com tendência para testar o caminho do processo que utiliza o método KU. Além disso, será buscado melhorar o modelo como a inclusão de outros métodos de interpolação.

5. Agradecimentos

Os autores agradecem à FAPEMAT pelo apoio financeiro ao projeto e pela concessão de bolsa de iniciação científica através da cooperação técnica com a UFMT.

Referências

- Alvarenga, L. H. V., Pinto, A. L. R., Silva, S. T., Altoé, T. F., Morais, V. A., and MELLO, J. (2010). Comparação de procedimentos de amostragem na precisão de Inventário Florestal em fragmento de Floresta Estacional Semidecidual. In *XIX Congresso de Pós-graduação da UFPA*.
- Adhikary, S. K., Muttill, N., and Yilmaz, A. G. (2017). Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological processes*, 31(12), 2143-2161.
- Bento, G. M., and Santos, R. T. (2018) “Avaliação de Métodos de remoção de outliers e seus impactos na precisão dos métodos de interpolação”, Simpósio Mato-grossense de Mecanização Agrícola e Agricultura de Precisão.
- Boisvert, J. (2010). Geostatistics with locally varying anisotropy.
- Bottega, E. L., de Queiroz, D. M., de Assis de Carvalho Pinto, F., de Souza, C. M. A., and Valente, D. S. M. (2017). Precision agriculture applied to soybean: Part I- Delineation of management zones. *Australian Journal of Crop Science*, 11(5), 573-579.
- Camargo, E. C. G., Felgueiras, C. A., and Monteiro, A. M. V. (2001). A importância da modelagem da anisotropia na distribuição espacial de variáveis ambientais utilizando procedimentos geoestatísticos. *Simpósio Brasileiro de Sensoriamento Remoto*, 10, 395-402.
- Li, J., Duan, P., Sheng, Y., & Lv, H. (2015, December). Spatial interpolation approach based on IDW with anisotropic spatial structures. In *International Conference on Intelligent Earth Observing and Applications 2015* (Vol. 9808, p. 980810). International Society for Optics and Photonics.
- Li, J., and Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.
- Merwade, V. M., Maidment, D. R., & Goff, J. A. (2006). Anisotropic considerations while interpolating river channel bathymetry. *Journal of Hydrology*, 331(3-4), 731-741.
- Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, 113, 56-69.
- OMG (2011). Business Process Model and Notation (BPMN), Version 2.0 Object Management Group (Technical report, Object Management Group).

- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, 9(1), 6.
- Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G. B., Hristopulos, D., Pilz, J., ... & Skøien, J. O. (2011). INTAMAP: the design and implementation of an interoperable automated interpolation web service. *Computers & Geosciences*, 37(3), 343-352.
- Requia, W. J., Coull, B. A., and Koutrakis, P. (2019). Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM_{2.5} constituents over space. *Environmental research*, 175, 421-433.
- Schenatto, K., Souza, E. G., Bazzi, C. L., Bier, V. A., Betzek, N. M., & Gavioli, A. (2016). Data interpolation in the definition of management zones. *Acta Scientiarum. Technology*, 38(1), 31-40.
- Team, R. C. (2013). R: A language and environment for statistical computing.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.