



UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA AGRÍCOLA

**AGDA LOUREIRO GONÇALVES OLIVEIRA**

**INFLUÊNCIA DO RIGOR GEOESTATÍSTICO NA  
QUALIDADE DO MAPEAMENTO EM AGRICULTURA DE  
PRECISÃO**

**INFLUENCE OF GEOSTATISTICAL RIGOR ON THE  
QUALITY OF MAPS USED IN PRECISION AGRICULTURE**

Campinas

2020

**AGDA LOUREIRO GONÇALVES OLIVEIRA**

**INFLUÊNCIA DO RIGOR GEOESTATÍSTICO NA  
QUALIDADE DO MAPEAMENTO EM AGRICULTURA DE  
PRECISÃO**

Dissertação apresentada à Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Engenharia Agrícola na Área de Máquinas Agrícolas.

Orientador: Prof. Dr. Lucas Rios do Amaral

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA AGDA LOUREIRO GONÇALVES OLIVEIRA, E ORIENTADA PELO PROF. DR. LUCAS RIOS DO AMARAL.

Campinas

2020

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Luciana Pietrosanto Milla - CRB 8/8129

OL4i Oliveira, Agda Loureiro Gonçalves, 1993-  
Influência do rigor geoestatístico na qualidade do mapeamento em agricultura de precisão / Agda Loureiro Gonçalves Oliveira. – Campinas, SP : [s.n.], 2020.

Orientador: Lucas Rios do Amaral.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola.

1. Geologia - Métodos estatísticos. 2. Agricultura de precisão. I. Amaral, Lucas Rios do, 1985-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Influence of geostatistical rigor on the quality of maps used in precision agriculture

**Palavras-chave em inglês:**

Geostatistics

Precision agriculture

**Área de concentração:** Máquinas Agrícolas

**Titulação:** Mestra em Engenharia Agrícola

**Banca examinadora:**

Lucas Rios do Amaral [Orientador]

Célia Regina Grego

Alessandro Samuel Rosa

**Data de defesa:** 28-02-2020

**Programa de Pós-Graduação:** Engenharia Agrícola

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-8761-9716>

- Currículo Lattes do autor: <http://lattes.cnpq.br/7112574061987202>

Este exemplar corresponde à redação final da **Dissertação de Mestrado** defendida por **Agda Loureiro Gonçalves Oliveira**, aprovada pela Comissão Julgadora em 28 fevereiro de 2020, na Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas.

**FEAGRI**

---

**Prof. Dr. Lucas Rios do Amaral – Presidente e Orientador**

---

**Dra. Célia Regina Grego – Membro Titular**

---

**Dr. Alessandro Samuel Rosa – Membro Titular**

**Faculdade de**  
**Engenharia Agrícola**  
**Unicamp**

**A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do discente.**

À Deus que guia todos os meus caminhos e me oferece forças para continuar buscando novas aventuras.

À minha família que me incentivou a seguir meus sonhos e lutar para alcançar meus objetivos. Obrigada por me ensinarem a não me acomodar com situações e a ter forças para correr atrás de tudo aquilo que sou capaz de mudar.

À vó Neném, Tia Van e minha mainha que não medem esforços para me apoiar. Espero que um dia consiga retribuir todo o carinho e atenção que me dedicam.

*Dedico.*

## AGRADECIMENTOS

À Universidade Estadual de Campinas e à Faculdade de Engenharia Agrícola (FEAGRI) pela oportunidade de cursar a pós graduação.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão de bolsa de mestrado com processo nº 2018/25473-2.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de mestrado.

Ao meu orientador Prof. Dr. Lucas Rios do Amaral por todos os ensinamentos, correções, direcionamentos e incentivos na construção dessa dissertação.

Aos membros da banca de defesa, Célia Regina Grego e Alessandro Samuel-Rosa, por todas as contribuições para esse trabalho.

Aos professores da FEAGRI por todo o aprendizado. Em especial à Prof<sup>a</sup>. Dr<sup>a</sup>. Gleyce K. Figueiredo e ao Prof. Dr. Luiz Henrique Antunes Rodrigues por todo o incentivo.

À minha querida avó, exemplo de mãe e mulher que me mostrou desde pequena o poder da força de vontade e do trabalho árduo para conquistar objetivos na vida.

Aos meus pais, Nena e Osvaldo, pelo apoio e por sempre acreditarem em meus sonhos, mesmo nos que pareciam impossíveis. Aos meus irmãos, Thiago e Maria Vitória, pelo companheirismo que sempre tivemos.

Aos meus Tios, Ivanilde, Cida, Rogério, Antônio (in memoriam) e Jailson (in memoriam), pelo exemplo e suporte que me ofereceram. E aos primos, meus primeiros amigos, Iago, Rilciara, Richard e Rildson por todas as experiências que compartilhamos desde quando ainda não sabíamos o que era responsabilidade.

Ao meu namorado Artur F. de Vito Junior, meu companheiro, que esteve do meu lado me apoiando, oferecendo ajuda e suporte para alcançar todos os meus objetivos. Obrigada por todo o carinho.

À Mara e Artur que me acolheram nesses anos longe de minha primeira casa. Obrigada por toda a paciência, pelas conversas e pelos conselhos. A vida longe de casa não é fácil e vocês fizeram os meus dias mais felizes, muito obrigada.

À toda a família Cavichioli: Vó Teresa, Tia Su, Fábio, Tio Quinho, Silvana, Renan, Mayara, Rogério, Jucão, Marjorie, Lucas e Gabriel. Muito obrigada pelas risadas nos almoços de domingo, pelas conversas, por trazer leveza à dificuldade do cotidiano.

Aos colegas e amigos do grupo de pesquisa – GITAP: Especialmente a Maiara Pusch, Julyane Fontenelli, Igor Valente, Joaquim Pedro, Rodrigo Freitas, Thiago Brasco e Yane

Freitas. Obrigada pelas trocas de ideias nas pausas para o café, as risadas nos momentos de descontração e por todo os conhecimentos compartilhados. Sem vocês, sem dúvida, a caminhada de descoberta de conhecimento não seria a mesma.

Aos colegas e amigos Ana Clara, Fernanda, Paulo e Marcos que me acompanharam nos anos de caminhada pela graduação.

À Valéria, Ritinha e Rebeca por toda a amizade, disponibilidade, atenção. Obrigada por todo o convívio nesses anos.

“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.”

Marie Curie

## RESUMO

Os mapas utilizados pela agricultura de precisão para aplicação localizada têm se baseado na amostragem em grade e interpolação dos dados. A interpolação por krigagem apresenta dificuldades no seu uso, visto que é um método complexo e que leva em consideração diversos parâmetros nos cálculos (i.e., anisotropia, tendência, efeito pepita, contribuição, alcance) que, se não considerados, a análise é prejudicada, reduzindo a confiabilidade dos mapas gerados por esse método. Nesse sentido, objetiva-se avaliar se considerar esses parâmetros na modelagem geoestatística permite ganho expressivo na qualidade dos mapas empregados na agricultura de precisão. Para isso foram testados o ganho na qualidade preditiva ao tratar anisotropia e tendência nos dados, levando em conta duas densidades amostrais (1 amostra/hectare e 1 amostra/ 4 hectares), além da modelagem variográfica pelo método dos momentos e método da máxima verossimilhança restrita. Esses parâmetros foram testados em campos virtuais simulados que foram produzidos pelo processo de Simulação Incondicional Gaussiana pelo pacote *gstat* no software R, e validados em dois campos experimentais que apresentam condições de anisotropia e tendência. Os resultados obtidos pelas análises foram avaliados a partir da comparação dos valores preditos pela interpolação e os valores observados nos campos simulados, além de comparados os valores estimados nos campos experimentais com conjunto de validação. Foi observado que a adição de efeitos direcionais na modelagem variográfica contribui para incremento da precisão da interpolação por modelos geoestatísticos. Inclusive, em alguns casos, o modelo determinístico (utilizado para comparação), superou o modelo geoestatístico em que não foi considerado nenhum efeito direcional. Além disso, ao reduzir o número de amostras, o efeito direcional na modelagem continuou contribuindo para aumento da precisão dos modelos geoestatísticos. Mas, ao alterar a forma de modelagem variográfica, a importância de cada efeito direcional também foi alterada, com anisotropia influenciando mais o método da máxima verossimilhança restrita, enquanto a tendência influenciou mais o método dos momentos. Portanto, considerar efeitos direcionais na modelagem semivariográfica proporciona maior qualidade/precisão dos mapas interpolados por geoestatística, o que eleva a maior confiabilidade dos mapas que são utilizados em Agricultura de Precisão.

Palavras-chave: efeitos direcionais; anisotropia; densidade amostral; krigagem; máxima verossimilhança; semivariograma.

## ABSTRACT

Maps used by precision agriculture to prescribe local application are based on sampling grid and interpolation. Interpolation by kriging presents some difficulties related to its usage since it is a complex method that considers several parameters for calculation (i.e. anisotropy, trend, nugget effect, partial sill, range) and, when they are not properly considered, it might jeopardize the quality of the predictions, reducing the reliability of maps created by this method. Accordingly, this study aims to evaluate if considering directional effects in geostatistical modelling leads to an expressive gain in the quality of the maps used by precision agriculture. Thus, we studied the gain in predictive quality by treating anisotropy and trend found in data, considering two sampling grids (1 sample/ ha and 1 sample/4ha), besides calculating the semivariogram by the method of moments and the method of restricted maximum likelihood. These parameters were tested in a simulated virtual field produced by the process of Gaussian Unconditional Simulation calculated by the *gstat* package on R software. In addition, these effects were also tested in two experimental fields that presented trend and anisotropy on the dataset. The results obtained by the analyses were evaluated through comparison of predicted values by interpolation and observed values obtained in the virtual fields, and on experimental fields the predicted values were compared with validation sets. Our results showed that adding directional effects on semivariogram modelling contributes to increase the accuracy on geostatistical interpolation. Indeed, we found better results of deterministic models than geostatistical models when none of the directional effects were regarded in modelling. However, the importance of directional effects also changed according to the method of semivariogram modelling. As the restricted maximum likelihood method seemed to have a stronger relationship with anisotropy effect to achieve higher precision, the method of moments achieved better results with trend. Thus, the process of accounting for directional effects in the semivariogram modelling can provide higher prediction accuracy on the maps interpolated by kriging. In this way, it is possible to obtain more reliable maps used for Precision Agricultural practices.

Keywords: directional effects; anisotropy; sampling size; kriging; maximum likelihood; semivariogram.

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>13</b>
1.2 OBJETIVOS .....	15
<b>2 REVISÃO DE LITERATURA.....</b>	<b>16</b>
2.1 AMOSTRAGEM.....	16
2.2 INTERPOLAÇÃO DOS DADOS .....	17
2.2.1 Estimativa determinística de dados espaciais .....	18
2.2.2 Estimativa estocástica de dados espaciais .....	18
2.2.3 Teoria das variáveis regionalizadas – da covariância à semivariância .....	19
2.2.4 O semivariograma .....	21
2.2.5 A anisotropia como efeito direcional no semivariograma e nas predições espaciais .....	23
2.2.6 A tendência como efeito direcional no semivariograma e nas predições espaciais .....	23
2.2.7 Estimativa do semivariograma pelo REML .....	24
<b>3 MATERIAL E MÉTODOS .....</b>	<b>26</b>
3.1 CAMPO VIRTUAL.....	27
3.1.1 Parametrização do Modelo.....	27
3.1.2 Amostragem .....	29
3.1.3 Modelagem variográfica .....	29
3.1.4 Interpolação.....	31
3.2 CAMPOS EXPERIMENTAIS .....	31
3.2.1. Processamento dos dados .....	33
3.3 FORMA DE ANÁLISE DOS RESULTADOS .....	34
3.3.1 Estatística para estimativa do modelo do semivariograma .....	34
3.3.2 Estatísticas utilizadas para validação externa .....	34
3.3.3 Testes de alcance para o método da máxima verossimilhança restrita .....	35
<b>4 RESULTADOS E DISCUSSÃO .....</b>	<b>36</b>
4.1 INFLUÊNCIA DOS EFEITOS DIRECIONAIS NA QUALIDADE DOS MAPAS INTERPOLADOS .....	36

4.2 INFLUÊNCIA DOS EFEITOS DIRECIONAIS NAS ESTIMATIVAS AO VARIAR A DENSIDADE DE PONTOS AMOSTRADOS. ....	38
4.3 IMPORTÂNCIA DOS EFEITOS DIRECIONAIS CONFORME O MÉTODO DE ESTIMATIVA DOS PARÂMETROS DO SEMIVARIOGRAMA .....	42
4.4 CONSIDERAÇÕES GERAIS .....	44
<b>5 CONCLUSÃO .....</b>	<b>48</b>
<b>REFERÊNCIAS .....</b>	<b>49</b>
<b>APÊNDICE 1 – Modelagem dos Campos Virtuais – Simulação Incondicional Gaussiana .....</b>	<b>52</b>
<b>Apêndice 2 – Características dos Campos Virtuais e Experimentais .....</b>	<b>54</b>
<b>Apêndice 3 – Superfícies interpoladas pelos cenários propostos. ....</b>	<b>60</b>
<b>Apêndice 4 – Estatísticas das interpolações nas 100 realizações dos Campos Virtuais ..</b>	<b>64</b>

## 1 INTRODUÇÃO

A agricultura de precisão (AP) considera a heterogeneidade dos talhões e, dentre diversas práticas, faz amostragem em grade e aplica os fertilizantes em taxas variadas através de mapas de prescrição (MOLIN et al., 2015). A construção de mapas utilizados pela AP para recomendação de fertilizantes conta com algumas etapas. A primeira é a coleta de dados, na maioria das vezes a partir de amostragens, em que é realizado planejamento do número de amostras de solo que serão coletadas e sua localização, sendo que isso é feito pois amostrar a área toda é economicamente inviável. Após esse processo são realizadas as análises laboratoriais e, a partir delas, é obtida a representação da área analisada a partir de interpolação dos dados. Essa interpolação pode ser realizada de maneira matemática (forma determinística) ou reconhecendo a dependência espacial dos atributos do solo a partir de análises geoestatísticas (método da krigagem). Ambas as formas levam a uma representação aproximada e, de certa forma, irreal do comportamento dos atributos do solo, principalmente quando a amostragem é muito esparsa, o que leva a mapas com uma série de erros acumulados (NANNI et al., 2011; CHERUBIN et al., 2014). Entretanto, já é bem estabelecido pela literatura que a maneira em que é realizado o planejamento amostral impacta na modelagem geoestatística, devido ao maior número de amostras geralmente levar a mapas mais confiáveis (OLIVER e WEBSTER, 2014).

A literatura aponta que a krigagem é o método de interpolação mais confiável, sendo considerado como interpolador geoestatístico não enviesado (WEBSTER e OLIVER, 1990). Porém, a análise geoestatística é, de certo modo, complexa. Além disso, toda a teoria desenvolvida – variáveis regionalizadas de Matheron (1971) – não foi direcionada à agricultura, onde ocorrem interações entre os fatores antrópicos e os naturais, o que levam a características diferentes ao que ocorre no campo de estudo em que o método foi gerado – a geologia e mineração. Além disso, para melhor confiabilidade dos mapas, a geoestatística tem se mostrado como um método interessante e pesquisadores (KRAVCHENKO e BULLOCK, 1999; BETZEK *et al.*, 2019) constataram que a qualidade de suas estimativas supera métodos matemáticos como o Inverso da Distância Ponderada (Inverse Distance Weight – IDW), por considerar, além do peso em função das distâncias entre as amostras (assim como ocorre no IDW), o comportamento espacial da variável em estudo. No entanto, em dados agrícolas tem sido notado que tal superioridade no desempenho nem sempre é observada (ASSIS SILVA et al., 2008; AMARAL; JUSTINA, 2019). Nesse sentido, comparação entre krigagem ordinária (OK) e IDW vem sendo discutida por cientistas que defendem que a escolha do interpolador deve ocorrer de acordo com análise criteriosa dos dados (KRAVCHENKO e BULLOCK, 1999)

e que não existe melhor interpolador para todos os casos (ROBINSON e METTERNICHT, 2006). Para automatizar essa escolha, BETZEK *et al.* (2019) apresentaram rotina que analisa todos os dados amostrados e fornece o melhor interpolador de acordo com o menor índice de seleção de interpolador (ISI), calculado a partir do erro médio e do erro médio padrão. Não obstante, nenhum desses trabalhos leva em consideração os impactos de efeitos direcionais na escolha do método de interpolação por geoestatística, principalmente a ocorrência de anisotropia nos dados, que é destacado como limitação no estudo de Betzek *et al.* (2019).

Na geoestatística, para descrever o comportamento de uma variável ao longo da área de estudo, após a realização de amostragem, é feito o cálculo (modelagem) do semivariograma para posterior realização da interpolação. Essa modelagem pode ocorrer por diversos métodos, sendo que o método dos momentos (MoM) descrito por Matheron calcula o semivariograma em passos (semivariograma experimental) e depois adapta um modelo matemático ao semivariograma experimental, dando origem ao semivariograma teórico (OLIVER e WEBSTER, 2014). Já o método da máxima verossimilhança restrita ou residual (REML) calcula o semivariograma teórico diretamente nos dados, isto é, a estimativa do semivariograma teórico ocorre no semivariograma de nuvem, sem simplificação (WEBSTER e OLIVER, 1992).

Além disso, no cálculo do semivariograma devem ser considerados os efeitos direcionais dos dados. Essa análise é realizada pois o comportamento dos dados no campo pode variar conforme direção, por exemplo o comportamento espacial dos elementos pode ser influenciado pela direção das operações agrícolas (i.e., linha de plantio). Esse comportamento diferencial pode ser constatado na modelagem semivariográfica e reconhecido no momento da interpolação na porção estocástica da variável regionalizada. A denominação para essa variação em relação à direção é anisotropia (KRŪMINIENE, 2006). E é analisada de forma a constatar se a continuidade ou dependência espacial se altera conforme a direção dos pares de pontos analisados para preparação do semivariograma experimental. Geralmente, esse efeito é avaliado satisfatoriamente a partir do cálculo e análise de covariância nos ângulos de 0°, 45°, 90° e 135° (YAMAMOTO e LANDIM, 2013). E, devido ao comportamento dos dados agrícolas, essa análise deve ser realizada a partir do mapa semivariográfico, o qual apresenta a covariância em todos os ângulos, para assim encontrar a direção preferencial. Quando essa análise não é realizada de maneira criteriosa, pode levar à diminuição da confiabilidade da aplicação da geoestatística (OLIVER e WEBSTER, 2014).

Outro efeito direcional é a tendência, que pode ser entendida como variação gradual na semivariância (constatada na modelagem semivariográfica) que ocorre na porção

determinística da variável regionalizada. Quando a média e variância no conjunto de dados não se alteram ao longo da área e a correlação entre qualquer par de pontos depende apenas do espaço entre eles (não ocorre variação gradual), é utilizada krigagem ordinária (YAMAMOTO e LANDIM, 2013). Assim, com a variação gradual, a tendência é considerada tanto no semivariograma quanto na interpolação, sendo que na última fase o método interpolador é alterado de krigagem ordinária para krigagem universal (KIS, 2016), a fim de modelar o efeito de tendência a partir das coordenadas. Ou seja, não identificar esse efeito de tendência nos dados pode levar ao método equivocado de interpolação geoestatística. Na prática, ao não considerar tendência nos dados, a interpolação não identificará a variabilidade em pequena escala, uma vez que a variabilidade em macro escala predominará na modelagem (MOLIN et al., 2015). Assim, é preciso remover este efeito macro dos dados (i.e., tendência) para então se obter mapas com feições mais bem delineadas.

Dessa forma, todos os detalhamentos citados acima são necessários para garantir boa qualidade na interpolação de dados por krigagem. Porém, tais análises geram alta complexidade para sua adoção, o que limita seu uso por usuários de agricultura de precisão. Além disso, devido ao fato das variáveis de interesse da AP poderem apresentar alta influência antrópica (AMARAL e JUSTINA, 2019), o que limita a qualidade das interpolações por conta de sua porção aleatória na distribuição espacial dos dados, existe a dúvida se a maior complexidade da krigagem realmente gera retorno na criação de mapas. Assim sendo, a meta principal desta pesquisa é verificar se o rigor na modelagem geoestatística, levando em consideração efeitos direcionais, permite ganho expressivo na qualidade dos mapas empregados na agricultura de precisão.

## 1.2 OBJETIVOS

Desse modo, o objetivo principal deste trabalho foi avaliar se a qualidade dos mapas obtidos por interpolação é influenciada significativamente pela consideração de efeitos direcionais na modelagem variográfica.

Assim, alguns objetivos específicos se apresentam:

- Avaliar se ao considerar a tendência e anisotropia nos dados se obtém melhor representação espacial da variável em estudo.
- Verificar se a influência dos efeitos direcionais é impactada ao variar densidade amostral;
- Analisar se ao utilizar métodos de modelagem variográfica distintos (MoM ou REML) há alteração na importância dos efeitos direcionais.

## 2 REVISÃO DE LITERATURA

A Agricultura de Precisão é uma forma de manejo que reconhece a variação no comportamento da lavoura (MOLIN *et al.*, 2015), fazendo aplicação de insumos de acordo com a necessidade de cada porção individualmente através de aplicação em taxa variada. Para estudar a característica variável é necessário coletar dados ao longo da lavoura e como obter informações de todos os pontos da lavoura é inviável (preço elevado de análises), é comum utilizar de amostragem de dados para representar as áreas. Não obstante, a amostragem é pontual e não traz informações da lavoura como um todo. Para contornar esse problema é possível realizar predição de pontos que não foram coletados a partir de técnicas de interpolação. Desta maneira é obtido o mapa de variação de atributos ao longo de toda a lavoura, produto que agrega informações gerenciais importantes, como por exemplo locais onde se deve aplicar maior ou menor quantidade de fertilizantes. Sem embargo, o procedimento de interpolação culmina em estimativas que podem conter erros, o que prejudica a qualidade de mapas fornecidos para aplicação de insumos agrícolas. Uma das causas desse erro de interpolação é a escolha do modelo de interpolação para estimativa em que podem ser utilizados modelos determinísticos ou geoestatísticos. Outra causa é a escolha da forma de modelar o comportamento espacial da variável em estudo. E, por fim, ao realizar tal modelagem geoestatística, reconhecer ou não efeitos direcionais que podem ser causados pela direção das operações agrícolas, por exemplo.

### 2.1 AMOSTRAGEM

As técnicas de amostragem que são utilizadas na AP vêm da estatística e como na área de sua origem, o objetivo é estimar uma população a partir de uma porção dela (amostra). Quando a amostragem é realizada de maneira adequada, deve demonstrar o comportamento da população, de sua variabilidade e distribuição no domínio estudado (WEBSER e OLIVER, 2007). Ou seja, o planejamento amostral é uma etapa muito importante para representar o comportamento da população estudada. Para isso, é possível realizar planejamento de forma sistemática, aleatória simples ou estratificada. Destas, a amostragem aleatória simples é realizada de maneira que as coordenadas dos pontos amostrais são escolhidas de maneira casual. A aleatória estratificada é realizada em estratos e as amostras são aleatoriamente escolhidas dentro de cada estrato. Já a amostragem sistemática é realizada a partir de uma malha regular e os pontos são posicionados nos nós da malha (YAMAMOTO e LANDIM, 2013). Esse última

forma de planejamento amostral é muito utilizada por usuários de AP e é conhecida como amostragem em grade (BRUS *et al.*, 2006).

Além da forma de amostrar, também é importante definir o número de amostras necessário para representar adequadamente o comportamento das áreas. Com número maior de amostras é possível obter melhor caracterização das áreas, com o contraponto de custos mais elevados. O estudo de variáveis do solo tem mostrado que são necessárias grades densas para representar adequadamente a variabilidade espacial de tais propriedades (CHERUBIN *et al.*, 2014). Nesse sentido, é necessário definir o número de amostras que ao mesmo tempo obtenha previsões satisfatórias e com custos razoáveis.

## 2.2 INTERPOLAÇÃO DOS DADOS

De acordo com Mendenhall (1968), a estimativa de valores não amostrados pode ocorrer a partir de modelos matemáticos determinísticos como o Inverso das Distâncias Ponderadas (IDW) ou estocásticos (krigagem). Assim, no modelo determinístico, os modelos estimam novos valores a partir de uma equação linear (Eq. 1). Um modelo comumente utilizado por usuários de AP é o IDW (inverso do quadrado da distância ponderada). Esse se baseia na premissa de que amostras mais próximas são mais parecidas do que amostras mais distantes; assim, o peso que se dá para amostras mais próximas é maior, reduzindo conforme a distância entre as amostras e ponto a ser estimado aumenta. Enquanto com o modelo estocástico é possível determinar a confiabilidade da estimativa de um novo valor (Webster e Oliver, 2007). Assim, os modelos estocásticos fornecem a informação de variância da previsão da krigagem que representa o quanto o modelo está incerto a respeito de cada previsão realizada, ou seja, é possível investigar se a estimativa que o modelo realizou é confiável.

Assim, a previsão dos valores não amostrados ocorre a partir da Eq. 1 com uso de peso associado ao valor observado e às distâncias entre as amostras. A representação formal da estimativa é dada por (WEBSTER e OLIVER, 2007):

$$Z(x_0) = \sum_{i=1}^N w_i z(x_i) \quad \text{Eq.1}$$

Em que  $Z(x_0)$  corresponde ao valor que será estimado;  $w_i$  corresponde aos pesos que serão adicionados na estimativa; e  $z(x_i)$  corresponde aos valores observados nos pontos amostrados.

### 2.2.1 Estimativa determinística de dados espaciais

A estimativa dos pontos não amostrados é realizada a partir de um modelo que atribui maiores pesos aos vizinhos mais próximos e menores pesos aos vizinhos mais distantes. Isaaks e Srivastava (1989) descreveram que uma maneira de fazer isso é fazer com que o peso de cada amostra seja proporcional ao inverso da distância de qualquer potência (Eq. 2) (i.e. quadrática, cúbica). Desta maneira, os pesos utilizados para estimativas são determinados:

$$w_i = \frac{1}{|x_i - x_0|^\beta} \text{ com } \beta > 0 \quad \text{Eq.2}$$

Em que  $w_i$  é o peso que será atribuído na determinação dos valores não amostrados, sendo adicionado à Eq.1;  $x_i$  são os locais amostrados e  $x_0$  os locais não amostrados; e  $\beta$  são os expoentes.

A estimativa dos novos valores é realizada de forma que os valores são estimados com base nos pontos circundantes. Assim, são contabilizados, além dos pontos amostrados, o peso  $w_i$  estimado (Eq. 2) e a estimativa do novo valor é dada de acordo com a Eq. 1. Quando  $x_i$  e  $x_0$  são iguais, o valor estimado é igual a  $Z(x_i)$ , mantendo os valores observados inalterados. Ademais, quando  $\beta$  é igual a zero, o peso diminui rapidamente enquanto a distância aumenta, obtendo-se estimativa local.

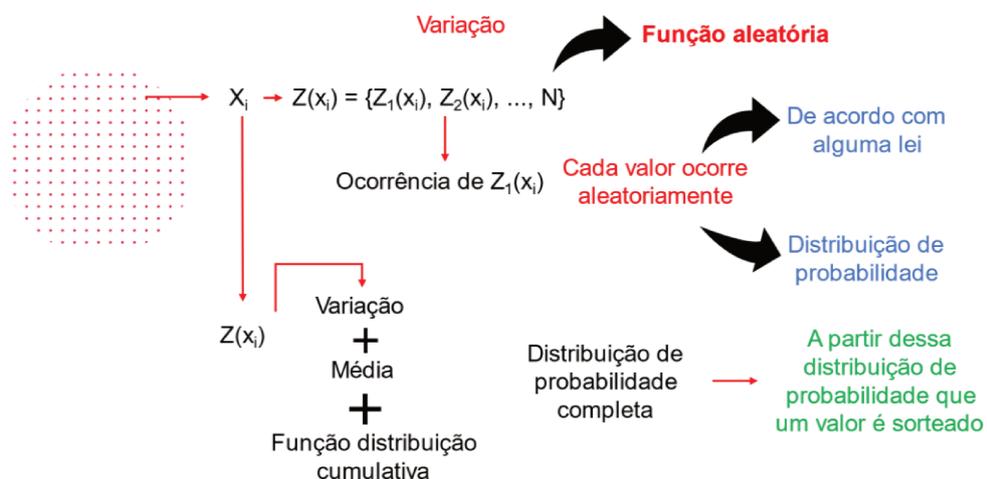
### 2.2.2 Estimativa estocástica de dados espaciais

A adoção da geoestatística ligada ao estudo dos solos é descrita por Webster e Oliver (2007). Os autores destacam que sua complexidade é elevada por acontecerem interações entre processos físicos e químicos que deixam a variação da área com aparência aleatória. Isso faz com que simples modelos matemáticos não sejam suficientes para definir a variabilidade das variáveis. Devido a esse comportamento aleatório, uma das formas para trabalhar com as variáveis espacializadas é no cenário estocástico (i.e. probabilístico). O cenário estocástico pode ser entendido na forma que a ocorrência de um valor da variável ocorre em função do acaso. Assim, a ocorrência de um valor obtido da amostragem, em dada localização da área, é uma realização do processo aleatório.

### 2.2.3 Teoria das variáveis regionalizadas – da covariância à semivariância

Sob esse pressuposto (Figura 1), cada ponto no espaço ( $x_i$ ) não possui apenas um valor, mas um conjunto de valores ( $Z(x_i)$ ), que é obtido a partir de uma distribuição probabilística e que são considerados como a função aleatória. Isto é, a probabilidade de um certo valor aparecer em uma amostragem de solo é decorrente de uma série de fatores que são desconhecidos e por isso são considerados como aleatórios. Assim, em um ponto  $x_i$ , a propriedade  $Z(x)$  é considerada como variável aleatória que possui média das realizações ( $\mu$ ), variância entre as realizações ( $\sigma^2$ ), momento de alta ordem e função de distribuição cumulativa (cdf), contendo distribuição probabilística completa (Figura 1). Nesse contexto, a distribuição dos dados é conhecida e, por isso, é possível estimar valores não amostrados a partir dos seus vizinhos, adicionando a componente de erro ( $\mathcal{E}$ ) na estimativa (WEBSTER e OLIVER, 2007). Dessa maneira, a variável regionalizada é caracterizada como conjunto de valores dentro de cada realização da função aleatória  $Z(x)$  (também chamada de processo aleatório ou estocástico).

**Figura 1** – O processo estocástico pode ser elucidado como ocorrência de  $Z_i(x)$  (variável aleatória) na local  $x_i$ , de acordo com uma distribuição probabilística de apresentar um certo valor (do conjunto de valores de  $Z(x)$ ) desde que cada valor de  $Z(x)$  ocorre ao acaso.



Fonte: WEBSTER e OLIVER (2007).

Apesar do termo probabilístico, estocástico, as variáveis regionalizadas têm característica de apresentarem relações de dependência entre pontos mais próximos, ou seja, os pontos  $x_i$ ,  $i = 1, 2, 3, \dots$ ,  $Z(x_i)$  não são independentes. A relação entre duas variáveis  $z_1$  e  $z_2$  para

observações pareadas, para  $n$  pares de observações,  $z_{i,1}$ ,  $i = 1, 2, \dots, n$ , pode ser determinada a partir da covariância:

$$\hat{C}(z_1 z_2) = \frac{1}{n} \sum_{i=1}^n \{z_{i,1} - \bar{z}_1\} \{z_{i,2} - \bar{z}_2\} \quad \text{Eq. 3}$$

Onde  $\bar{z}_1$  e  $\bar{z}_2$  são as médias de  $z_1$  e  $z_2$ , respectivamente. E  $\hat{C}(z_1, z_2)$  coestima a covariância da população.

De acordo com os mesmos autores, a covariância pode ser utilizada para pares de variáveis regionalizadas. Assim,  $z_1$  e  $z_2$  se tornam  $Z(x_1)$  e  $Z(x_2)$  que são valores da mesma variável  $Z$  em dois pontos no espaço  $x_1$  e  $x_2$ . A forma de descrever a covariância se torna:

$$C(x_1, x_2) = E[\{Zx_1 - \mu(x_1)\}\{Zx_2 - \mu(x_2)\}] \quad \text{Eq. 4}$$

Onde  $\mu(x_1)$  e  $\mu(x_2)$  são as médias de  $Z$  nos pontos  $x_1$  e  $x_2$ . O problema com esse uso é que normalmente se tem conhecimento apenas de uma realização de  $Z$  e, portanto, não é possível a obtenção da média ( $\mu$ ). Por isso é feita a suposição do processo estacionário que significa que cada localização  $x_i$ ,  $i = 1, 2, 3 \dots, n$ , é uma realização de  $Z(x_i)$ . Dessa forma, na estacionaridade, a média  $\mu$  é constante em  $x_i$ ,  $i = 1, 2, 3 \dots, n$ , e pode ser obtida a partir de diversas amostragens. Com a média constante e a covariância e variância dependendo apenas da separação dos pontos tem-se a estacionaridade de segunda ordem:

$$C(x_i, x_j) = E[\{Z(x_i) - \mu\}\{Z(x_j) - \mu\}] \quad \text{Eq. 5}$$

Se  $x_i$  e  $x_j$  não são coincidentes, então a covariância é dependente apenas da distância entre eles  $h = x_i - x_j$ . Então podemos substituir esses termos na Eq. 7:

$$\begin{aligned} \text{cov}[Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})] &= E[\{Z(\mathbf{x}) - \mu\} \{Z(\mathbf{x} + \mathbf{h}) - \mu\}], \\ &= E[\{Z(\mathbf{x})\} \{Z(\mathbf{x} + \mathbf{h}) - \mu^2\}], \\ &= C(\mathbf{h}) \end{aligned} \quad \text{Eq. 6}$$

Dessa maneira obtemos a função auto covariância, pois representa a relação de  $Z$  com ele mesmo. Essa função descreve a relação de dependência de  $Z$  ao mudar a localização de  $x_i$ ,  $i = 1, 2, 3, \dots, n$ . Nesse contexto, com a estacionaridade definida, é possível descrever o processo aleatório estacionário a partir de:

$$Z(x) = \mu + \varepsilon(x) \quad \text{Eq.7}$$

Sendo  $\mu$  correspondente à média do processo; e  $\mathcal{E}(x)$  corresponde à porção aleatória.

O problema com a estacionaridade de primeira ordem é que nem sempre a média  $\mu$  é constante em todos os lugares. Todavia, Matheron (1965) propôs que a média pode não ser constante em todos os locais, mas pode ser constante em pequenas distâncias. Assim, foi proposta a hipótese intrínseca de Matheron com a alteração de covariância por variância:

$$\begin{aligned}\text{var}[Z(x) - Z(x + h)] &= E[\{Z(x) - Z(x + h)\}^2] \\ &= 2\gamma(h)\end{aligned}\tag{Eq. 8}$$

Onde  $\gamma$  é a semivariância que corresponde à metade da variância. A semivariância é representada por (OLIVER e WEBSTER, 2014):

$$\begin{aligned}\gamma(h) &= \frac{1}{2} \text{var}[Z(x) - Z(x + h)] \\ &= \frac{1}{2} E[\{Z(x) - Z(x + h)\}^2]\end{aligned}\tag{Eq. 9}$$

#### 2.2.4 O semivariograma

A semivariância experimental dos dados é obtida a partir da Eq. 10, sendo que o semivariograma é calculado ao alterar a distância  $h$  e pode ser estimado por algumas formas. Alguns exemplos de formas de estimar os parâmetros do semivariograma teórico são o Método dos Momentos (MoM) (MATHERON, 1965) e Método da Máxima Verossimilhança Residual ou Restrita (REML) (PARDO-IGÚZQUIZA e DOWD, 1997). Para o MoM, a estimativa da semivariância ocorre a partir da Eq. 10, em que é possível obter a semivariância de todos os pares de pontos, separados por distância chamada *lag*.

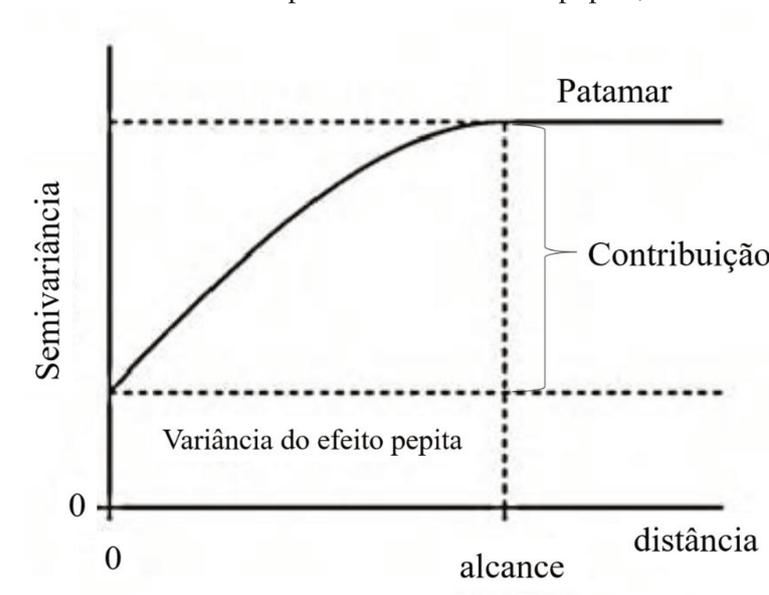
$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} \{z(x_i) - z(x_i + h)\}^2\tag{Eq. 10}$$

Em que  $z(x_i)$  e  $z(x_i + h)$  são os valores reais de  $Z$  nos locais  $(x_i)$  e  $(x_i + h)$  e  $m(h)$  é o número de pares separados por distância  $h$  (*lag*).

O semivariograma do MoM caracterizado por ser contínuo e monotônico (a semivariância aumenta de acordo com o aumento da distância do *lag*), de tal modo que quando o semivariograma passa a apresentar aumento dissimilar a um certo ponto, temos o alcance que é considerado o ponto máximo em que o processo é espacialmente dependente ou auto

correlacionado. O limite máximo, no eixo das ordenadas, em que a variância aumenta é chamado de **patamar** (variância *a priori*), sendo que a partir dele ocorre estabilização da semivariância em função dos *lags*. Na abscissa (distância), esse limite é chamado de **alcance**, que também é chamado de limite de correlação, sendo o ponto em que a auto correlação se torna zero (CRESSIE, 1990). E, no início do semivariograma do MoM pode ser observada descontinuidade, que é chamada de **efeito pepita**, em que os valores de semivariância parecem ocorrer aleatoriamente e independente dos outros valores (Figura 2).

**Figura 2**– Semivariograma Experimental gerado pela modelagem semivariográfica pelo Método dos Momentos destacando os parâmetros de efeito pepita, contribuição e alcance.



Portanto, para estimativa do semivariograma teórico é importante a determinação dos parâmetros efeito pepita, patamar, contribuição (patamar menos efeito pepita) e alcance através do chute inicial. Após esse processo é realizada a adaptação de um modelo matemático que melhor explique o comportamento dos dados presente no semivariograma experimental (LARK e WEBSTER, 2006). Para isso são testados diversos modelos, sendo que os mais comuns são modelo gaussiano, exponencial e esférico. O modelo pode ser escolhido através de validação cruzada, em que deve apresentar melhor resultado em relação as estatísticas estabelecidas (i.e., menor valor de erro e maior capacidade de explicar a variância dos dados).

Outra forma de estimar os parâmetros do semivariograma teórico é a partir do REML que não utiliza de *lags*, mas da variância de todos os pares de pontos distribuídos no espaço amostral. Por isso o processo de geração do semivariograma pelo REML é

computacionalmente mais intensivo do que pelo MoM (WEBSTER e OLIVER, 2007). O detalhamento da estimativa por REML é dada após a definição de tendência na próxima seção.

### 2.2.5 A anisotropia como efeito direcional no semivariograma e nas predições espaciais

Além da variação no espaço estimada pelo semivariograma omnidirecional tratado até aqui, a variação espacial pode obter comportamento diferenciado de acordo com a direção avaliada. Esse comportamento é dito como anisotrópico. Na natureza são encontrados alguns tipos de anisotropia, como zonal, geométrica e mista, sendo que a anisotropia do tipo geométrica é a mais encontrada (KRŪMINIENĖ, 2006) e a única que a correção está geralmente implementada nos softwares para análise geoestatística. Esse tipo de anisotropia ocorre, de acordo com a definição descrita por ISAACS e SRIVASTAVA (1989), quando as direções apresentam dois alcances e um único patamar, ocorrendo maior continuidade em uma direção do que em outra (YAMAMOTO e LANDIM, 2013). A direção de maior continuidade é a principal (maior alcance) e a de menor continuidade é a direção secundária. Na prática, ao não reconhecer o efeito anisotrópico nos dados, as estimativas por krigagem podem mascarar o efeito de linhas de contorno presentes na lavoura de cana-de-açúcar (SANCHES *et al.*, 2018), reduzindo a qualidade da estimativa da krigagem ordinária.

### 2.2.6 A tendência como efeito direcional no semivariograma e nas predições espaciais

Além de anisotropia, o semivariograma pode apresentar tendência. Em geoestatística a tendência significa variação gradual no espaço (OLIVER e WEBSTER, 2014). Sua ocorrência pode ser observada no semivariograma quando ocorre aumento da variância mesmo após o momento que deveria atingir a patamar, mascarando o alcance real da variável. A equação que define a variável regionalizada é alterada, pois a média não é mais constante, mas varia de acordo com as coordenadas. O modelo para variação espacial é então transformado para:

$$Z(x) = m(x) + \varepsilon(x) \quad \text{Eq. 11}$$

Em que  $\mu$  é substituído por  $m(x)$ , pois a média passa a variar. Ao reconhecer essa variação gradual, os resíduos passam a ser estimados como:

$$\varepsilon(x) = Z(x) - m(x), \quad \text{Eq. 12}$$

E a tendência  $m(x)$  pode ser expressa como:

$$m(x) = \sum_{k=0}^K \beta_k f_k(x), \quad \text{Eq. 13}$$

Em que  $\beta_k$ ,  $K = 1, 2, \dots, K$ , são coeficientes desconhecidos e que podem ser estimados e  $f_k(x)$  são funções conhecidas para  $x$ . Unindo as duas equações acima (Eq. 12 e Eq. 13), pode-se obter o processo de tendência com o modelo:

$$Z(x) = u(x) + \varepsilon(x) = \sum_{k=0}^K \beta_k f_k(x) + \varepsilon(x) \quad \text{Eq. 14}$$

Assim, a tendência pode ser modelada nas coordenadas geográficas a partir de um polinômio de primeira ordem:

$$Z(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon(x) \quad \text{Eq. 15}$$

Em que  $x_1$  e  $x_2$  são as coordenadas espaciais e para  $K+1 = 3$ , ou seja, com três termos no polinômio.

Se dentro da Eq. 19  $K = 0$ , então  $f_0 = 1$ ,  $m(x) = \beta_0 = \mu$ , e assim se tem processo estacionário representado pela equação (Eq. 7), com semivariograma usual. Assim, o processo pode ser considerado não enviesado e pode ser utilizada na krigagem ordinária. Já quando  $K > 0$ , isto é, quando ocorre tendência nos dados, o processo é mais complexo. Isso pois na presença de tendência, o semivariograma dos resíduos é enviesado pois os resíduos dependem das estimativas dos parâmetros de  $\beta$ . A solução encontrada por Matheron (1969) foi utilizar da krigagem universal para estimativa, que será detalhada na seção de interpolação.

### 2.2.7 Estimativa do semivariograma pelo REML

Uma outra maneira de modelar a tendência de forma separada é a partir do REML. Nessa forma de modelagem o conjunto das variáveis aleatórias é considerado com distribuição multivariada normal com matriz de covariância. Uma nova variável aleatória pode ser obtida a partir desse pressuposto da criação de nova variável, e os parâmetros de variância podem ser obtidos ao aplicar a máxima verossimilhança a essa nova variável, produzindo a verossimilhança residual (LARK e WEBSTER, 2006). A estimativa do semivariograma por REML é dada pelo modelo misto linear, sendo que os termos  $\beta_0 + \beta_1 x_1 + \beta_2 x_2$  da Eq. 15 são substituídos por  $W\beta$ .

$$Z(x) = W\beta + \varepsilon(x) \quad \text{Eq. 16}$$

Em que o vetor  $W$  contém os elementos  $k+1, x_l, \dots, x_k$  da função da tendência e o vetor  $\beta$  contém os coeficientes.

Essa mudança ocorre, pois, a estimativa do semivariograma pelo REML ocorre de forma que primeiramente os resíduos ( $\mathcal{E}$ ) e a tendência ( $\mu$ ) são separados. Então os parâmetros semivariográficos são obtidos ao maximizar a máxima verossimilhança dos resíduos. A partir dos parâmetros é obtida a matriz de covariância. Com os parâmetros e a matriz de covariância são estimados os valores dos coeficientes  $\beta$  a partir dos mínimos quadrados generalizados. Por fim é realizada a estimativa de  $Z(x)$ . (WEBSTER e OLIVER, 2007). Assim, o procedimento para estimativa do REML ocorre de forma em que ocorre a obtenção os parâmetros da variância (efeito pepita, patamar e alcance) a partir de da aplicação da máxima verossimilhança a essa nova variável; computa a matriz de covariância; estima os coeficientes  $\beta$  a partir dos Mínimos Quadrados Generalizados, que produz mínima variância e  $\mathcal{E}$  não enviesado; e por fim substitui a covariância e os coeficientes no modelo de predição. Assim é obtido o BLUP empírico (E-BLUP) (LARK e WEBSTER, 2006).

Após modelagem semivariográfica, a interpolação por modelos geoestatísticos pode ser realizada a partir da krigagem. De acordo com Cressie (1990), a krigagem é definida como melhor preditor espacial não enviesado (BLUP), no sentido de minimizar a variância das estimativas A estimativa espacial krigagem é dada a partir de

$$Z(x_0) = \sum_{i=1}^n w(s_0)^T Z(x_i) \quad \text{Eq. 17}$$

Em que  $Z(x_0)$  corresponde ao valor que será estimado;  $w(s_0)^T$  corresponde aos pesos da krigagem que serão adicionados na estimativa de modo a reduzir o erro de predição na condição de não enviesamento, quando os pesos são forçados a somar um (HENGL *et al.*, 2018); e  $z(x_i)$  corresponde aos pontos amostrados.

### 3 MATERIAL E MÉTODOS

Para responder se o rigor geoestatístico realmente direciona para mapas com maior qualidade é importante que sejam feitos testes em dados conhecidos. Por não se saber exatamente o comportamento dos nutrientes do solo em uma lavoura pois, invariavelmente, se obtém tal dado por meio de amostragem (comportamento real das variáveis), não é possível se calcular com precisão o desempenho das estimativas realizadas pela interpolação.

Nesse sentido, e com o avanço de ferramentas computacionais, a criação de campos virtuais (GEBBERS e HERBST, 2009) tem se tornado uma ferramenta interessante por possibilitar o conhecimento do comportamento das variáveis espaciais ao longo de uma determinada área, permitindo avaliar a influência de efeitos específicos nos dados e a verificação da qualidade final dos mapas a partir de avaliação dos erros. Ademais, possibilitam maior eficiência do que estudos em campos reais (GEBBERS e HERBST, 2009), já que permitem a obtenção dos erros de estimativa ao invés da incerteza, além de permitir o estudo da utilização de diversos parâmetros isoladamente (testar métodos de modelagem de semivariograma, por exemplo). Assim, sabendo-se como os parâmetros influenciam na estimativa, é possível verificar a relevância da consideração dos efeitos direcionais nos dados na qualidade final do mapa obtido. Assim sendo, trabalhei com a modelagem desses fatores em dados simulados, com posterior validação dos resultados em dados reais.

Para melhor compreensão dos modelos analisados nesse estudo, essa seção será dividida em duas partes: Campo Virtual e Campo Experimental. Optei por separá-las pois a primeira é estudada a partir de áreas virtuais em que a eficiência dos modelos ocorre apenas em função dos parâmetros fixados. Enquanto nos campos experimentais as relações são mais complexas com diversos processos interagindo simultaneamente para ocorrência da variabilidade da área.

A simulação incondicional é uma aplicação da técnica de Montes Carlo e corresponde à criação de valores a partir de uma função de covariância ou semivariograma (WEBSTER; OLIVER, 2007). No caso de Simulação incondicional gaussiana, as variáveis aleatórias são geradas de acordo com um modelo semivariográfico (contendo efeito pepita, contribuição e alcance) sob a distribuição global gaussiana, apresentando distribuição normal entre os dados. Para cada simulação desses valores, tem-se uma realização da função aleatória  $Z(x)$ .

### 3.1 CAMPO VIRTUAL

#### 3.1.1 Parametrização do Modelo

Escolhi utilizar Simulação Incondicional Gaussiana (SIG) (WEBSTER e OLIVER, 2007) devido ao objetivo de representar o comportamento espacial de uma variável qualquer, mas que não necessariamente representasse um atributo real sob distribuição gaussiana global. O modelo utilizado para a simulação requer a determinação de parâmetros do semivariograma (alcance, contribuição e efeito pepita), além dos parâmetros de tendência (coeficientes  $\beta$ ) e anisotropia. Os parâmetros semivariográficos foram escolhidos a partir da análise de uma área experimental de 116 hectares (área que será detalhada na seção de campos experimentais, correspondente à Área 1) na qual foi realizada amostragem em grade irregular. A variável do solo avaliada para essa adoção foi o potássio, o qual foi escolhido por apresentar as características que pretendo avaliar nesse estudo (anisotropia e tendência) e que não foi utilizado na análise dos campos experimentais. Os parâmetros do semivariograma encontrados foram: efeito pepita = 0, contribuição = 0,28, e alcance = 391,50. Os coeficientes da tendência foram obtidos a partir da adaptação de um modelo de Mínimos Quadrados Generalizados (GLS); dessa maneira, obtive os coeficientes  $\beta = 1,6746, 0,0016, 0,0004$ . Já a anisotropia foi definida a partir da análise do mapa semivariográfico e plotagem do semivariograma em todas as direções. O ângulo de ocorrência de maior continuidade foi  $30^\circ$  e a razão entre o alcance de maior continuidade e seu alcance perpendicular foi de 0,5. O modelo semivariográfico utilizado foi o esférico (Eq. 18), pois é o mais frequentemente encontrado (OLIVER e WEBSTER, 2014).

$$\begin{aligned} \gamma(h) &= c_0 + c \left\{ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right\} \text{ para } 0 < h \leq a, & \text{Eq. 18} \\ \gamma(h) &= c_0 + c \text{ para } h > a, \\ \gamma(h) &= 0 \end{aligned}$$

Em que:  $h$  é o tamanho do passo,  $c_0$  é o efeito pepita,  $c$  é a contribuição e  $a$  é o alcance.

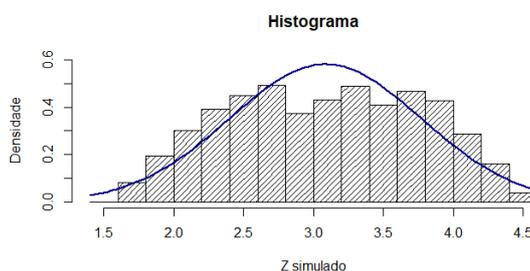
Nesse contexto, estou trabalhando sob o pressuposto de cenário estocástico e os valores “z” (variável simulada) são criados a partir desse pressuposto, com distribuição gaussiana (isto é, distribuição normal). Dessa maneira, cada modelagem, ou seja, cada simulação realizada pelo modelo descrito acima, corresponde a uma realização da variável. Caso ocorra outra simulação, temos outra realização e assim por diante até  $n$  simulações e  $n$  realizações, dentro de um cenário probabilístico/estocástico. Cada realização obtida dessa SIG é correspondente a um campo virtual (VF). Então, a partir daqui leia que um campo virtual é equivalente a uma realização da SIC, totalizando 100 realizações da SIC. As repetições dos

pontos no VF poderiam ser coletadas a partir de várias amostragens (realizadas aleatoriamente), ou de diversas simulações (i.e., 100 campos virtuais gerados para o mesmo modelo) com valores coletados sob o mesmo ponto. No estudo de Lark (2000) foram realizadas 100 amostragens, porém aqui optei por obter 100 realizações do VF, resultando em 100 campos virtuais.

Para implementação do código que gera os VFs utilizei o software R (código está descrito no Apêndice 1). Existem algumas funções no software R em que a simulação já está implementada, como por exemplo no pacote *RandomFields*, *geoR* e *gstat*. Dentre esses pacotes escolhi implementar a modelagem no *gstat* (Apêndice 1), pois este pacote apresenta documentação mais detalhada e é amplamente utilizado.

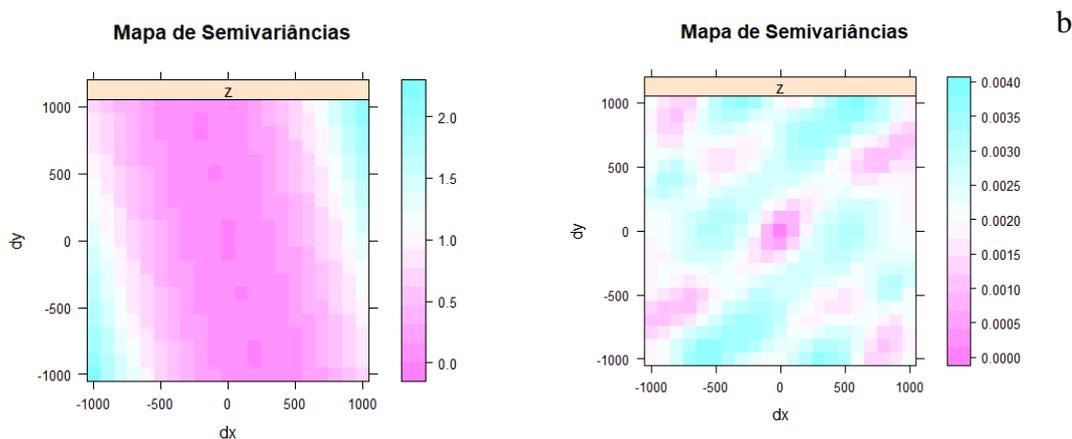
Nesse sentido, obtive 100 VFs gerados a partir do processo de Simulação Incondicional Gaussiana, com 200 hectares cada (1400m por 1450m) e com pixels de 10 metros de resolução. O tamanho da área foi determinado de modo a oferecer os requisitos mínimos (número de pontos amostrais) para trabalhar com Método dos Momentos e Método da Máxima Verossimilhança Restrita (REML), assim é possível obter o mínimo de 50 pontos para construção do semivariograma, mesmo ao amostrar 1 amostra/4ha. Escolhi utilizar essa resolução espacial (estimativas a cada 10 m) para reduzir custo computacional tanto do processo de geração dos VFs quanto do processamento dos dados após amostragem. A variável regionalizada nas suas 100 realizações, apresenta distribuição gaussiana global (Figura 3 a e b), ou seja, apresenta comportamento normal, o que é pré-requisito para utilização da interpolação por krigagem.

**Figura 3** – Comportamento normal apresentado pela e-type dos Campos Virtuais gerados.



Além disso, os campos virtuais gerados apresentam comportamento espacial contendo tendência e anisotropia (Figura 4 a e b).

**Figura 4** – Mapa semivariográfico da média e-type das realizações dos campos virtuais sendo a) comportamento com tendência e b) comportamento com anisotropia.



Com os campos virtuais criados, simulei um procedimento de amostragem comumente realizado em campo. Assim, foram coletadas amostras (i.e., amostragem sistemática) e, a partir delas, realizadas análises espaciais e interpolação para elaboração de mapas para aplicação de fertilizantes, por exemplo.

### 3.1.2 Amostragem

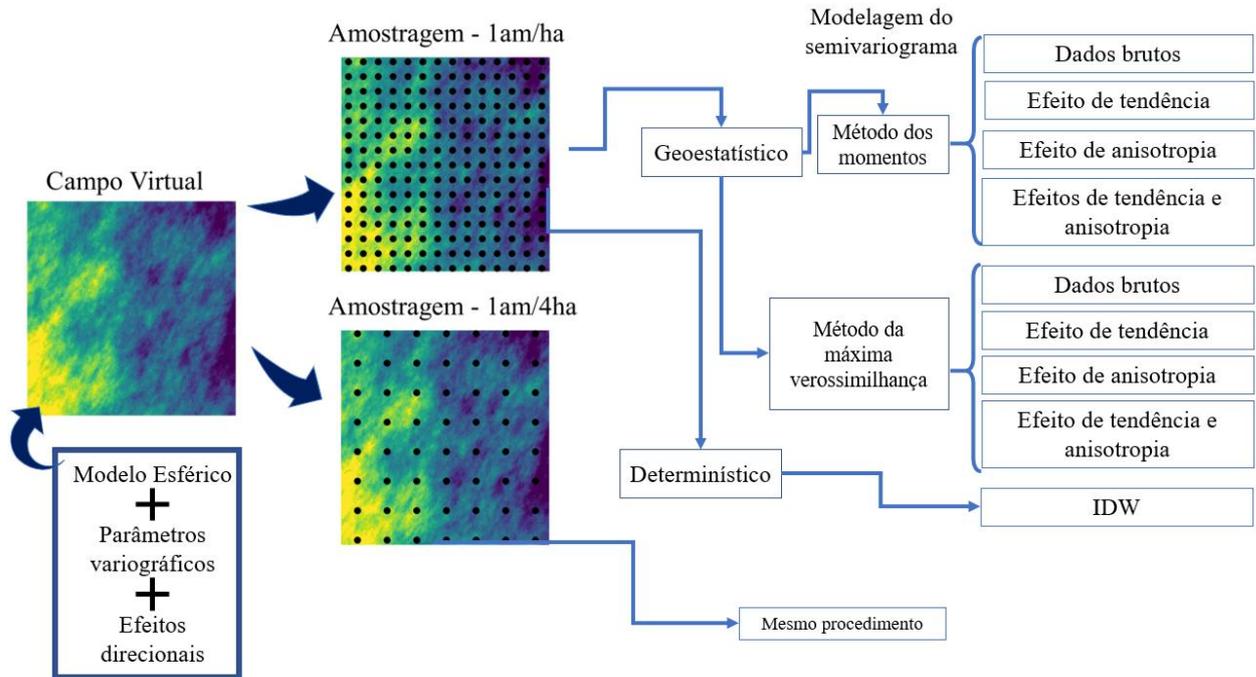
Realizei o planejamento amostral para duas densidades (uma amostra por hectare e uma amostra a cada quatro hectares). O planejamento foi para amostragem sistemática que é conhecida na AP por amostragem em grade regular. Dessa forma, foi criado um ponto a cada 100m (sobreposto ao nó das grades) para a amostragem mais densa, e um ponto a cada 200m para amostragem mais esparsa. O procedimento que adotei foi extrair os valores dos pixels dos VFs colocados com os pontos de amostragem gerados. Assim, os pontos amostrados dos VFs foram tabelados com coordenadas e valores do atributo simulado (“z”), à exemplo do que ocorreria caso o solo tivesse sido amostrado e enviado para análise de laboratório. As amostragens dos campos virtuais foram armazenadas em banco de dados separados para prosseguimento das análises.

### 3.1.3 Modelagem variográfica

A forma de modelagem geoestatística interfere no produto da interpolação por krigagem, que são os mapas de prescrição de insumos, por exemplo. Nesse sentido, a fim de estudar o impacto de cada parâmetro (forma de estimativa do semivariograma, contabilização de tendência e anisotropia) na qualidade dos mapas obtidos, diferentes cenários foram estudados: modelagem geoestatística por MoM e REML; dados brutos (não considerei os efeitos direcionais) e é chamado de condição A; tratamento de tendência dos dados, chamado de condição B; tratamento de anisotropia dos dados, condição C; tratamento da tendência e

anisotropia, condição D. Além disso, para cada amostragem realizada, também foi realizada interpolação por IDW para comparar com os mapas interpolados por krigagem. Esse processo resultou em um conjunto de 18 cenários (Figura 5), e 1800 áreas interpoladas.

**Figura 5**– Fluxograma experimental, mostrando as etapas da amostragem e análise geoestatística, tendo como produto 3600 mapas interpolados pelos diversos cenários estudados para cada campo virtual.



Convencionalmente o semivariograma é modelado em duas etapas. Na primeira, o semivariograma é estimado a partir de pares de pontos dispostos em classes de passos separados por distância  $h$  e calculados em todas as direções (semivariograma omnidirecional), dando origem ao semivariograma experimental (OLIVER e WEBSTER, 2007). Na segunda é feita a adaptação de modelo teórico (modelo matemático) que melhor representa o comportamento do semivariograma gerado na primeira fase, gerando o semivariograma teórico (LARK, 2000). Esse procedimento é denominado de MoM de Matheron (1965) e é calculado a partir da Eq. 12. Já o REML estima a variável regionalizada a partir de combinações espaciais dos dados originais, denominadas incremento generalizado, que são utilizadas para retirar a deriva do cálculo e trabalhar apenas com o resíduo, produzindo estimativas menos enviesadas do que o método anterior (o detalhamento do cálculo de REML pode ser encontrado em (PARDO-IGÚZQUIZA e DOWD, 1997)).

### 3.1.4 Interpolação

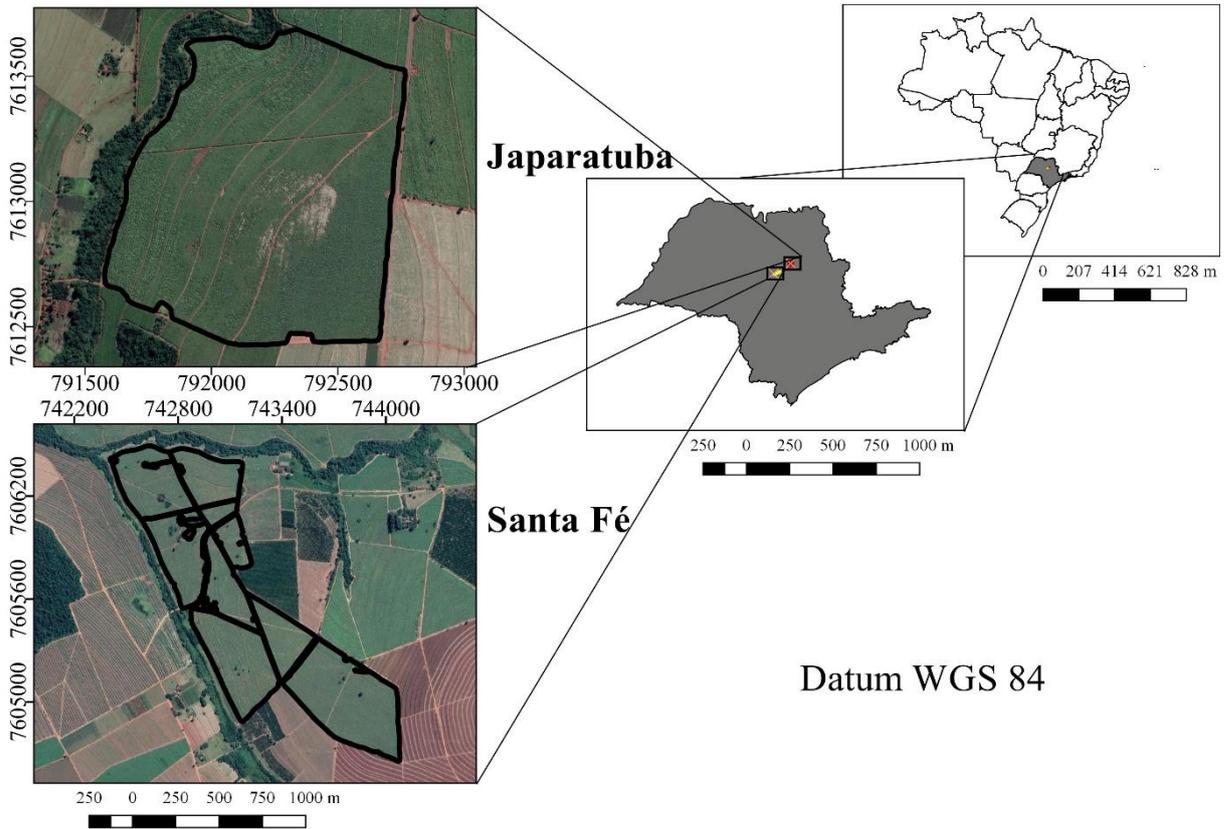
Após a modelagem variográfica foi realizada a interpolação dos dados, conforme o cenário estudado. A krigagem Ordinária foi utilizada quando a tendência não foi contabilizada nos modelos, já quando houve reconhecimento de tendência (contabilização em relação às coordenadas), utilizei krigagem universal para predição (CRESSIE, 1994). As interpolações foram realizadas de forma automatizada para todos os cenários e condições de todas as 100 realizações no software R.

Além disso, realizei a média (estatística do tipo e – e-type) de todos os 100 valores para cada ponto/coordenada/pixel do campo virtual e obtive uma superfície interpolada. Os resultados das métricas comparadas nas tabelas dos resultados e discussão são equivalentes aos dados provenientes do e-type dos campos virtuais. Por outro lado, na análise de variância o e-type não é utilizado e cada realização é considerada como repetição do dado.

## 3.2 CAMPOS EXPERIMENTAIS

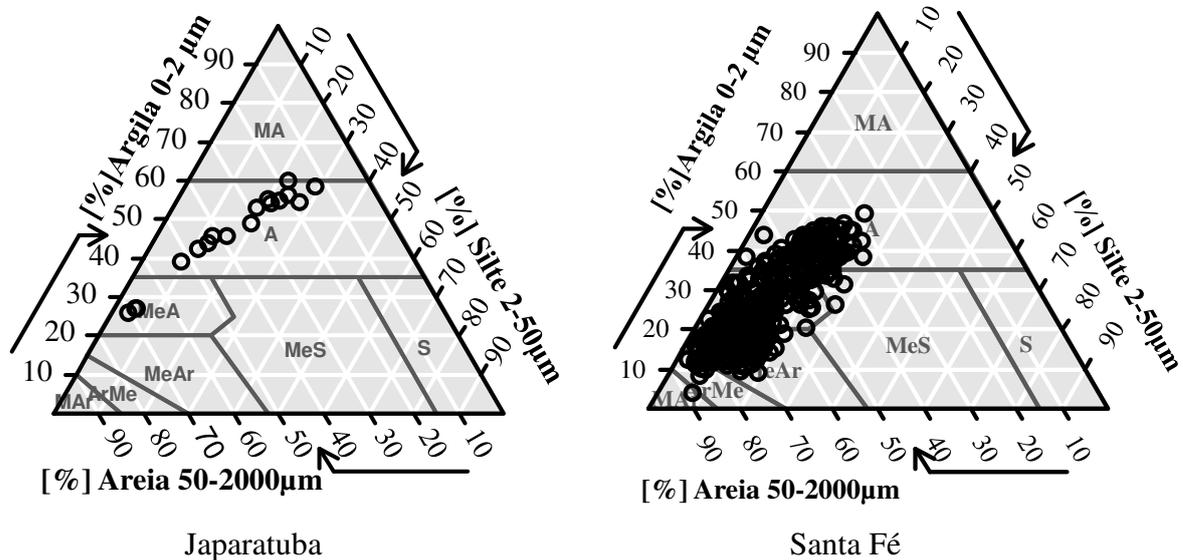
Optamos por analisar dois campos experimentais que apresentaram os efeitos direcionais estudados (tendência e anisotropia), sendo que a variável do solo que apresentou essas condições em ambos os campos foi a matéria orgânica (MO). Por isso, todos os dados apresentados para essas áreas são referentes à MO. As duas áreas são cultivadas com cana-de-açúcar (dados coletados para outras pesquisas). Uma das áreas está localizada no município de Motuca no estado de São Paulo, chamada Japaratuba e a outra área está localizada no município Nova Europa, no estado de São Paulo, chamada Santa Fé (Figura 6).

**Figura 6** – Localização das áreas experimentais e localização das amostras coletas.



A fazenda Japarutuba (Figura 8), possui 116 hectares e tem predominância de Latossolo Vermelho e solo de textura argilosa (Figura 9). O solo foi amostrado em novembro de 2011 no período de reforma do canavial, sendo o planejamento amostral realizado em grade regular, com amostras deslocadas para evitar carregadores ou bordas, e com densidade de cerca de uma amostra por hectare (118 amostras) e profundidade de coleta de 0 a 0,4m, sendo coletadas 6 subamostras por ponto amostrado. Enquanto a fazenda Santa Fé, possui 102 hectares com solo que varia desde textura arenosa média até argilosa (Figura 7). A amostragem de solo ocorreu em 2016 em grade irregular e foram coletadas três subamostras por ponto amostrado (100 pontos amostrados no total) na profundidade de 0-0,20m.

**Figura 7** – Triângulo o de classificação do solo de acordo com a classificação de Santos et al. (2013) dos campos experimentais.



### 3.2.1. Processamento dos dados

Os dados foram analisados a partir de estatística descritiva a fim de compreender o comportamento da variável, além de identificar a presença de possíveis outliers, pois a presença de valores anômalos no conjunto de dados influencia nos mapas interpolados (BORSSOI *et al.*, 2009). A análise de outliers adotada foi univariada, a partir de boxplot, observando intervalo interquartílico. Nas áreas analisadas não foram removidos outliers por não apresentarem valores discrepantes espacialmente em relação aos pontos circundantes.

Posteriormente os dados das amostras de solo foram separados espacialmente em dois conjuntos, calibração (80% dos dados) e validação (20% dos dados). Essa separação é feita por ser necessário avaliar a qualidade das interpolações para conjunto de dados diferente dos utilizados na modelagem do semivariograma (conjunto de calibração), obtendo assim os valores preditos pela interpolação em locais onde os valores reais são conhecidos.

Os pontos de calibração passaram pelo mesmo procedimento de modelagem geoestatística que foi utilizada nos dados simulados, ou seja, foram utilizados cenários onde os efeitos direcionais são ou não considerados, sendo então interpolados ao final por krigagem universal e IDW. A avaliação dos modelos geoestatísticos foi realizada a partir da interpolação dos pontos de calibração para as coordenadas dos pontos de validação, sendo comparada as diferenças entre os valores preditos e observados.

### 3.3 FORMA DE ANÁLISE DOS RESULTADOS

Essa seção será dividida em três partes, a primeira relativa à capacidade de predição do modelo (obtida através de validação cruzada) no conjunto de calibração. A segunda corresponde à aplicação do modelo no conjunto de validação. A última seção é relativa à análise de variância dos VFs, comparando a média das métricas de desempenho obtidas.

#### 3.3.1 Estatística para estimativa do modelo do semivariograma

Os modelos de predição/interpolação dos dados gerados a partir dos diferentes cenários que criei foram calibrados em um conjunto de dados que contém as amostras extraídas nos VFs e, no caso dos campos experimentais 80% das amostras. Nesse conjunto é realizada a modelagem semivariográfica a partir de diversos tipos de modelos (i.e., esférico, gaussiano e exponencial) e é realizada validação cruzada *Leave one out* para selecionar o modelo semivariográfico a ser utilizado na interpolação por krigagem.

Os critérios adotados para análise de desempenho do modelo foram: raiz quadrada do erro médio (RMSE), erro médio (ME), erro médio absoluto (MAE), quantidade de variância explicada (AVE) e coeficiente da regressão ( $R^2$ ). Essas estatísticas foram adotadas pois expressam informações diferentes a respeito do desempenho do modelo. RMSE corresponde ao desvio padrão da diferença entre valores preditos e observados (SCHMITT *et al.*, 2015), penalizando valores muito extremos estimados pelo modelo.

#### 3.3.2 Estatísticas utilizadas para validação externa

As estatísticas adotadas para obter o desempenho do modelo calibrado foram: RMSE, MAE, ME, AVE,  $R^2$  (assim como citado acima) além do índice de Willmott (apenas para os VFs), sendo que a razão da escolha desse índice, para os VFs, detalhada após o AVE. AVE denota a quantidade de variação que é explicada pelo modelo, quanto mais próximo a 1, melhor é o modelo (HENGL *et al.*, 2018). AVE é calculado como:

$$AVE = 1 - \frac{\sum_{i=1}^n \left( y(s_j) - \hat{y}(s_j) \right)^2}{\sum_{i=1}^n \left( y(s_j) - \bar{y}(s_j) \right)^2} \quad \text{Eq. 19}$$

Onde  $\hat{y}(s_j)$  é o valor estimado e  $y(s_j)$  é o valor observado em dada localização na validação cruzada, e  $\bar{y}(s_j)$  corresponde à média dos valores observados.

Nos VFs é possível comparar os mapas interpolados com os mapas verdadeiros (VFs). Para isso utilizei método de comparação de mapas, o índice de Willmott modificado. Esse índice (WILLMOTT, 2012) calcula a concordância entre mapas, sendo adimensional e possui valores que variam no intervalo de -1 a 1. Valores mais próximos ou menores do que zero indicam pouca concordância e quanto mais próximos de um, maior é a concordância entre os mapas. Ele é dado pelo valor de concordância entre os mapas ( $dr$ ) a partir das Eq. 20 e 21.

$$dr = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{c \sum_{i=1}^n |O_i - \bar{O}|}, \text{ quando } \sum_{i=1}^n |P_i - O_i| \leq c \sum_{i=1}^n |O_i - \bar{O}| \quad \text{Eq. 20}$$

$$dr = \frac{c \sum_{i=1}^n |O_i - \bar{O}|}{\sum_{i=1}^n |P_i - O_i|} - 1, \text{ quando } \sum_{i=1}^n |P_i - O_i| > c \sum_{i=1}^n |O_i - \bar{O}| \quad \text{Eq. 21}$$

Em que:  $P_i$  é valor predito,  $O_i$  é o valor observado,  $\bar{O}$  é a média dos valores observados nos pixels,  $n$  é o número de pixels comparados e  $dr$  é o valor relativo ao índice de Willmott.

A Eq. 20 é utilizada para quando o somatório da diferença entre o valor predito e o observado for menor ou igual ao somatório da diferença entre o valor observado e a média. E a Eq. 21 é utilizada para condições em que Eq. 20 não são atendidas.

### 3.3.3 Testes de alcance para o método da máxima verossimilhança restrita

A fim de verificar o impacto do alcance fixo no REML pelo pacote *gstat*, foram fixados três alcances de 200m, 400m e 955m para todos os cenários. Escolhemos fixar esses alcances ao avaliar o alcance de cada condição de modelagem proposta. Encontramos alcances de cerca de 955m nas condições sem reconhecimento de tendência e, entre 200 e 400m nas condições com reconhecimento de tendência na modelagem semivariográfica. Isto é, o alcance de 200m foi fixado para condição A, B, C e D e foi avaliado o erro (RMSE) associado à fixação desse alcance. O mesmo procedimento foi realizado para os demais dois alcances fixados.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 INFLUÊNCIA DOS EFEITOS DIRECIONAIS NA QUALIDADE DOS MAPAS INTERPOLADOS

Por meio dos resultados dos campos virtuais, foi possível concluir que a qualidade dos mapas obtidos através de interpolação de pontos amostrais é influenciada pela consideração de efeitos direcionais na modelagem semivariográfica. Os dados interpolados nos cenários mais criteriosos, em relação a consideração dos efeitos de tendência e anisotropia (Condições A, B, C e D), geraram superfícies com menores erros (MAE, ME, RMSE – Tabela 1 e Apêndice 4) e maior similaridade com o campo virtual (VF) medidos através do índice de Willmott (Tabela 1). Assim, ocorre menor desempenho dos modelos geoestatísticos nos cenários sem consideração dos efeitos direcionais. Esse menor desempenho ocorre, pois, ao desconsiderar efeitos importantes na determinação dos parâmetros do semivariograma, isto é, contribuição, alcance e efeito pepita, o modelo semivariograma gerado é prejudicado (KRAVCHENKO e BULLOCK, 1999), assim como pode ser observado no Apêndice 2. Dessa forma, ocorreu maior erro no cenário em que nenhum efeito direcional é reconhecido nos campos (Tabela 1 – MAE Condição A) do que quando os efeitos são reconhecidos (Tabela 1 – MAE Condição D). Ao adicionar o efeito de tendência e da anisotropia na modelagem obtivemos melhora no desempenho (Tabela 1 – Condição D); contudo, o efeito de anisotropia considerado separadamente obteve menor desempenho na representação da variação espacial (Tabela 1 – Condição C), com resultado superior à Condição A, onde nenhum efeito foi considerado, mas inferior às condições B (tendência) e D (ambos os efeitos). Quando apenas o efeito anisotrópico (Tabela 1 – Condição C) é considerado (isto é, efeito direcional ocasionado pelas linhas de plantio), o alcance do semivariograma é superestimado, pois a tendência é desconsiderada e o semivariograma não atinge patamar. Esse efeito faz com que a estimativa de valores do VF seja prejudicada por considerar que ocorre dependência em uma distância superior a real (alcance dos semivariogramas no Apêndice 2). Desse modo, o número de vizinhos considerados é maior, podendo adicionar ruídos por considerar vizinhos que não são dependentes entre si. De maneira similar, ao modelar o semivariograma apenas com o efeito da tendência (Tabela 1 – Condição B) os vizinhos considerados na estimativa por krigagem não são diferenciados de acordo com a direção preferencial (anisotropia), e é assumido que o comportamento espacial dos dados ocorre da mesma maneira em qualquer direção (isotropia), prejudicando o reconhecimento da direção preferencial da dependência espacial diferenciada causada pelas linhas de plantio (mapa semivariográfico - Figura 4<sup>a</sup> e b).

**Tabela 1** – Desempenho<sup>1</sup> das estimativas geoestatísticas a partir de modelagem de semivariograma por MoM considerando os cenários<sup>2</sup> de consideração de efeitos direcionais na amostragem densa (1 amostra/ha) nos campos virtuais (VF) e experimentais (Japarutuba e Santa Fé).

Condição <sup>2</sup>	MAE			ME			AVE		
	VF	JAP <sup>3</sup>	SF <sup>3</sup>	VF	JAP	SF	VF	JAP	SF
A	0,28	1,21	3,54	-0,01	-0,13	-1,11	0,83	0,77	0,62
B	0,23	1,23	3,55	-0,01	-0,11	-1,08	0,89	0,86	0,63
C	0,26	1,19	3,50	-0,01	-0,22	-1,06	0,86	0,86	0,63
D	0,22	1,20	3,49	-0,01	-0,20	-1,02	0,90	0,86	0,62
IDW	0,28	1,78	3,52	-0,01	-0,30	-1,03	0,83	0,86	0,62
Condição	RMSE			R2			Willmott		
	VF	JAP	SF	VF	JAP	SF	VF	JAP	SF
A	0,35	1,79	4,65	0,83	0,86	0,65	0,77	--	--
B	0,29	1,80	4,65	0,89	0,86	0,64	0,83	--	--
C	0,32	1,77	4,61	0,86	0,87	0,65	0,79	--	--
D	0,27	1,76	4,60	0,90	0,87	0,65	0,84	--	--
IDW	0,36	2,27	4,63	0,86	0,88	0,64	0,74	-	-

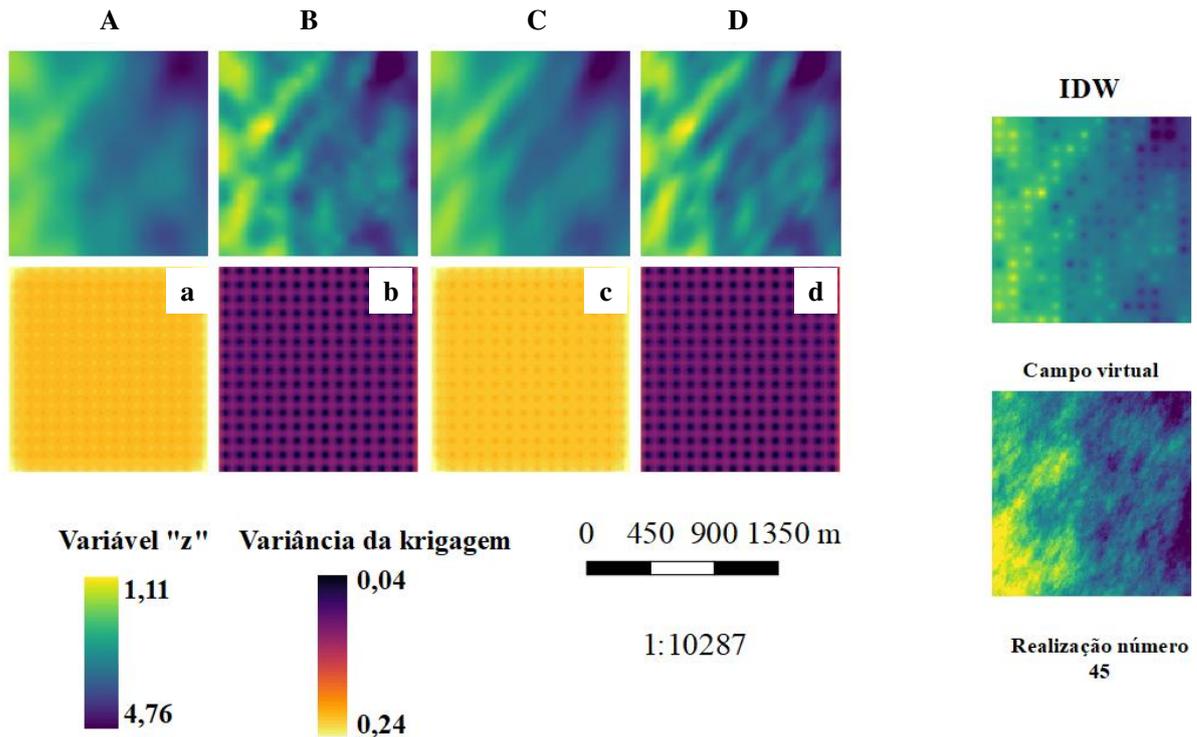
<sup>1</sup>MAE – erro médio absoluto; RMSE – raiz quadrada do erro médio; R<sup>2</sup> – coeficiente de determinação da regressão; AVE – quantidade de variância explicada; Willmott – índice de Willmott de concordância entre mapas.

<sup>2</sup>Condições: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

Os campos com efeito de tendência e ambos os efeitos direcionais (Condição B e D) apresentam melhores predições e menores valores de variância da krigagem (Figura 8 – b e d). Essa baixa variância explicita que os mapas gerados a partir do método dos momentos (MoM) com tendência e ambos os efeitos direcionais (Condição B e D) sejam mais assertivos e, portanto, mais confiáveis para aplicação de insumos no campo. Entretanto, vem sendo encontrado na literatura (AMARAL e JUSTINA, 2019) que a krigagem nem sempre é o melhor modelo de interpolação, visto que o IDW apresenta resultados similares (Tabela 1), obtendo vantagens quanto à menor complexidade de aplicação do modelo determinístico. Todavia, o modelo determinístico apresenta falhas na representação espacial (Figura 8) devido a mudanças abruptas de valores encontrados no campo ocasionadas pelo efeito das atividades agrícolas. Em dados ambientais (i.e. tipo de solo e variação textural), essa ocorrência não é natural, visto que os atributos tendem a variar suavemente de um pixel para o outro. Mas com intervenções

antrópicas, as propriedades do solo podem ter sua variabilidade alterada (CHEN *et al.*, 2020). Então, apesar de apresentar resultados similares (Tabela 1), a representação espacial do modelo é prejudicada (Tabela 1), fato evidenciado pela concordância de Willmott, que obteve IDW como pior modelo de predição dos dados não amostrados.

**Figure 8**– Representação espacial dos dados da amostragem densa (1 amostra/ha) interpolados pelos condições<sup>1</sup> propostas e relação de variância da krigagem com cada cenário. Ao lado, representação da interpolação da variável Z pelo IDW e o campo virtual tido como referência.



<sup>1</sup> Condições: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

#### 4.2 INFLUÊNCIA DOS EFEITOS DIRECIONAIS NAS ESTIMATIVAS AO VARIAR A DENSIDADE DE PONTOS AMOSTRADOS.

Além da amostragem densa (1 amostra/ha), foram realizadas amostragens de 0,25 amostras/ha no VF, e de 0,5 amostra/ha na Japarutuba e Santa Fé (amostragem esparsa). Ao reduzir a densidade amostral, a contabilização dos efeitos direcionais (Condição D) na modelagem semivariográfica continuou apresentando melhor desempenho (Tabela 2 e Apêndice 4). Inclusive, ao não considerar os efeitos direcionais (Condição A) na modelagem semivariográfica, houve redução do desempenho de forma que as predições por IDW superaram o desempenho da krigagem no VF. Assim, ao não considerar os efeitos, com número escasso de amostras, é preferível utilizar o modelo simples, determinístico.

**Tabela 2** – Desempenho<sup>1</sup> dos modelos geoestatísticos a partir de modelagem de semivariograma por MoM considerando os cenários<sup>2</sup> de consideração de efeitos direcionais na amostragem esparsa.

Condição <sup>2</sup>	MAE			ME			AVE		
	VF	JAP <sup>3</sup>	SF <sup>3</sup>	VF	JAP	SF	VF	JAP	SF
A	0,40	1,48	3,57	0,00	-0,20	-0,80	0,67	0,74	0,59
B	0,37	1,46	3,60	-0,01	-0,17	-0,77	0,70	0,83	0,61
C	0,38	1,48	3,57	0,00	-0,22	-0,76	0,69	0,83	0,61
D	0,36	1,46	3,57	-0,01	-0,18	-0,73	0,72	0,83	0,61
IDW	0,35	1,90	3,65	0,00	-0,37	-0,81	0,73	0,74	0,61
Condição	RMSE			R <sup>2</sup>			Willmott		
	VF	JAP	SF	VF	JAP	SF	VF	JAP	SF
A	0,50	1,97	4,72	0,69	0,84	0,63	0,61	--	--
B	0,47	1,98	4,70	0,71	0,83	0,63	0,69	--	--
C	0,48	1,96	4,69	0,70	0,84	0,63	0,64	--	--
D	0,45	1,99	4,68	0,72	0,83	0,63	0,64	--	--
IDW	0,44	2,44	4,80	0,76	0,82	0,61	0,66	--	--

<sup>1</sup>MAE – erro médio absoluto; RMSE – raiz quadrada do erro médio; R<sup>2</sup> – coeficiente de determinação da regressão; AVE – quantidade de variância explicada; Willmott – índice de Willmott de concordância entre mapas. MoM – Método dos Momentos; IDW – inverso do quadrado da distância.

<sup>2</sup>Condições: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

<sup>3</sup>JAP – Japarutuba; SF – Santa Fé

Contudo, essa relação de desempenho do modelo, efeitos direcionais e número de amostras não é observada nos campos experimentais. Nos campos experimentais a geoestatística é preferível, independentemente do número de amostras e mesmo com menor critério em relação aos efeitos direcionais (Tabela 2). De fato, Kerry e Oliver (2007) afirmaram que a modelagem é influenciada pelo número de amostras, sendo indicado utilizar no mínimo 100 amostras na modelagem geoestatística. Todavia, ao considerar efeitos direcionais, a qualidade da modelagem é melhorada (Tabela 2), obtendo bons resultados com krigagem mesmo com apenas 50 amostras. Essa melhoria quando algum dos efeitos direcionais é considerado é evidenciada ao verificar que ao adicionar tendência no VF e na Japarutuba ocorre variância mais próxima aos dados reais (Tabela 3 – Condição B). Por outro lado, na Santa Fé a consideração dos efeitos direcionais contribuiu para reduzir a variância dos dados preditos em relação aos dados amostrados, produzindo dados interpolados com maior efeito de suavização (Tabela 3 – Condição D). Nesse sentido, não é evidente um padrão em relação do número de amostras e a consideração de cada efeito separadamente.

**Tabela 3** - Variância dos dados da amostragem densa e esparsa nos campos virtuais (VF) e nos campos experimentais (Japarutuba e Santa Fé) e dos dados interpolados pelos diferentes cenários propostos nos VFs e nos Campos Experimentais.

		VF	Japarutuba	Santa Fé
		Amostragem densa		
Observado <sup>1</sup>		0,80	19,06	64,83
	A	0,53	18,16	43,06
	B	0,66	18,03	42,65
Predito <sup>2</sup>	C	0,55	17,78	42,91
	D	0,67	17,69	42,94
	IDW	0,42	9,03	37,54
		Amostragem esparsa		
Observado <sup>1</sup>		0,80	20,06	66,68
	A	0,29	17,02	43,81
	B	0,44	17,18	41,83
Predito <sup>2</sup>	C	0,32	16,61	42,01
	D	0,44	16,75	42,24
	IDW	0,35	9,23	39,26

<sup>1</sup>Dados amostrados nas densidades (densa e esparsa) nos VFs, Japarutuba e Santa Fé.

<sup>2</sup>Dados interpolados a partir de IDW e das diferentes condições propostas com geostatística: Condições: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

A qualidade do modelo de predição não pode ser atribuída somente ao modelo de interpolação escolhido, mas à característica do conjunto de dados. Isto é, no VF os efeitos direcionais foram isolados, o que permite encontrar modelos com melhor desempenho ao considerar esses efeitos. Já nos campos experimentais, esses efeitos não acontecem isolados, mas ocorre uma combinação de diversos outros fatores (i.e., material de formação do solo, microrganismos, clima, manejo do solo etc.). Além disso, a área Santa Fé possui efeito aleatório mais pronunciado do que a Japarutuba; uma forma de avaliar tal efeito é através da razão entre efeito pepita e patamar (Tabela 4 -  $c_0/(c_0 + c)$ ), sendo que quanto maior o valor, maior a suavização dos dados pelo interpolador (OLIVER e WEBSTER, 2014). A área Japarutuba apresenta a menor relação efeito pepita-patamar e menor suavização enquanto a Santa Fé apresenta a maior relação efeito pepita-patamar. Isso faz com que a Santa Fé apresente aleatoriedade dos dados e esta característica faz com que a porção de variância explicada no

semivariograma (contribuição) seja pequena em relação ao efeito pepita, o que resulta em predições espaciais com valores de variância inferiores aos reais (Tabela 3).

**Tabela 4** – Parâmetros do semivariograma estimados a partir do MoM para os campos virtuais e experimentais (Japarutuba e Santa Fé) para as duas densidades amostrais.

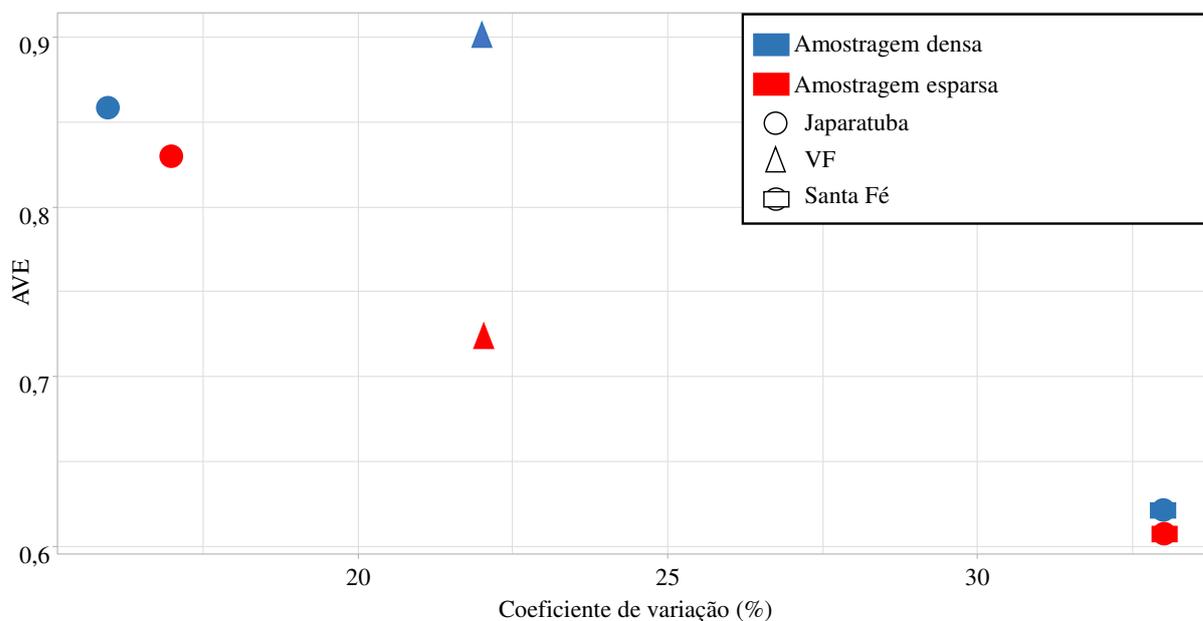
Amostragem densa									
	VF			Japarutuba			Santa Fé		
	$c_0^1$	$c^1$	$c_0/(c_0+c)$	$c_0^1$	$c^1$	$c_0/(c_0+c)$	$c_0^1$	$c^1$	$c_0/(c_0+c)$
A	0,15	5,58	0,03	1,06	1434,16	0,00	20,18	1236,60	0,02
B	0,04	0,24	0,14	0,00	5,72	0,00	15,18	18,30	0,45
C	0,14	6,89	0,02	0,00	4191,26	0,00	20,77	1572,22	0,01
D	0,04	0,24	0,14	0,00	5,69	0,00	15,14	18,26	0,45
Amostragem esparsa									
A	0,62	2,21	0,22	0,81	1882,99	0,00	21,96	1479,24	0,01
B	0,21	0,10	0,68	0,00	6,78	0,00	17,85	22,12	0,45
C	0,65	2,58	0,20	0,63	2213,31	0,00	22,57	1934,62	0,01
D	0,21	0,10	0,68	0,00	6,67	0,00	17,12	21,38	0,44

<sup>1</sup> $c_0$  – Efeito pepita;  $c$  – contribuição.  $c_0/(c_0+c)$  – razão do efeito pepita pelo patamar.

Além dos parâmetros de semivariância, alto coeficiente de variação (CV) influencia na redução do desempenho dos modelos geoestatísticos e determinísticos (LI e HEAP, 2011). Dessa maneira, a Santa Fé (Tabela 3) apresentou o modelo com pior desempenho, em relação as demais áreas devido a presença do CV mais elevado dentre as áreas avaliadas (Figura 9). Enquanto a área Japarutuba, mesmo com número de amostras equivalente (100 amostras) à metade do número de amostras extraídas do VF (200 amostras) apresenta elevado desempenho dos modelos devido ao menor CV (Figura 9). Essa redução de desempenho, onde o CV é elevado, ocorreu principalmente nos dados interpolados por krigagem ordinária (cenários sem consideração de efeito direcional (Condição A) ou consideração de apenas anisotropia (Condição C), devido a maior relação do CV com a redução do desempenho da krigagem ordinária do que com o estimador krigagem universal (cenários com ambos os efeitos e apenas efeito de tendência) (LI e HEAP, 2011). Além disso, o CV elevado pode estar relacionado com fatores de formação do solo (i.e., variação textural) que não foram incluídos nesse estudo, já que buscamos compreender a relação entre os efeitos direcionais e a qualidade da interpolação do modelo univariado (krigagem). Mas, em estudos paralelos a esse documento identificamos que parte da variação pode ser explicada por variáveis ambientais, sendo que essas variáveis podem auxiliar nas estimativas dos valores não amostrados a partir de modelos multivariados

como cokrigagem, krigagem com deriva externa/krigagem universal ou regressão da krigagem (GARCÍA-TOMILLO et al., 2017; SANCHES et al., 2018; JURADO-EXPÓSITO et al., 2019).

**Figura 9** – Relação entre o coeficiente de variação dos dados amostrados da amostragem densa e a quantidade de variância explicada (AVE) dos dados interpolados com modelagem por MoM no VF e nos Campos Experimentais (Japarutuba e Santa Fé).



#### 4.3 IMPORTÂNCIA DOS EFEITOS DIRECIONAIS CONFORME O MÉTODO DE ESTIMATIVA DOS PARÂMETROS DO SEMIVARIOGRAMA

Os efeitos direcionais influenciam a modelagem semivariográfica tanto por MoM quanto por Método da Máxima Verossimilhança Restrita (REML – Tabela 5 e Apêndice 4). Porém, ao analisar o comportamento dos modelos, encontramos que a importância dos efeitos direcionais é diferente para cada método de modelagem semivariográfica (Tabela 5 - ganho relativo das condições com efeitos direcionais – B, C e D – em comparação com a condição sem efeito direcional - A). Enquanto o MoM se mostrou muito mais dependente do efeito de adição de tendência (Condição B), o REML resultou em melhores estimativas ao considerar o efeito de anisotropia (Condição C) ou ao combinar ambos os efeitos direcionais (Condição D) nos dados. Os resultados de ME dos modelos do REML apresentaram valores mais próximos a zero (Tabela 5). Ademais, o ganho de desempenho do modelo REML ao adicionar os efeitos direcionais (Condição D) na modelagem é inferior ao ganho encontrado ao utilizar o MoM (Tabela 5) em áreas que apresentam maior CV (VF e Santa Fé) (Figura 10). Na área de baixo CV (Japarutuba – Figura 10) o ganho com os dois métodos é similar. Assim, quando se trabalha

com áreas com maior CV, o uso do REML deve ser priorizado por ser menos sensível aos efeitos direcionais, e menos influenciado caso os efeitos não sejam contabilizados por erro humano.

**Tabela 5** – Ganho de desempenho<sup>1</sup> dos métodos de estimativa de parâmetros do semivariograma ao considerar efeitos direcionais na modelagem em comparação com as condições que o efeito direcional não é considerado.

Condição	Modelo	VF		Japarutuba		Santa Fé	
		RMSE	GR <sup>2</sup>	RMSE	GR <sup>2</sup>	RMSE	GR <sup>2</sup>
A	MoM	0,354	GR <sup>2</sup>	1,789	GR <sup>2</sup>	4,651	GR <sup>2</sup>
	REML	0,284		1781		4,641	
B	MoM	0,287	-0,067	1,801	0,012	4,653	0,002
	REML	0,302	0,018	1,790	0,009	4,670	0,029
C	MoM	0,323	-0,031	1,768	-0,021	4,615	-0,036
	REML	0,275	-0,009	1,742	-0,039	4,618	-0,023
D	MoM	0,274	-0,080	1,764	-0,025	4,595	-0,056
	REML	0,266	-0,018	1,760	-0,021	4,628	-0,013

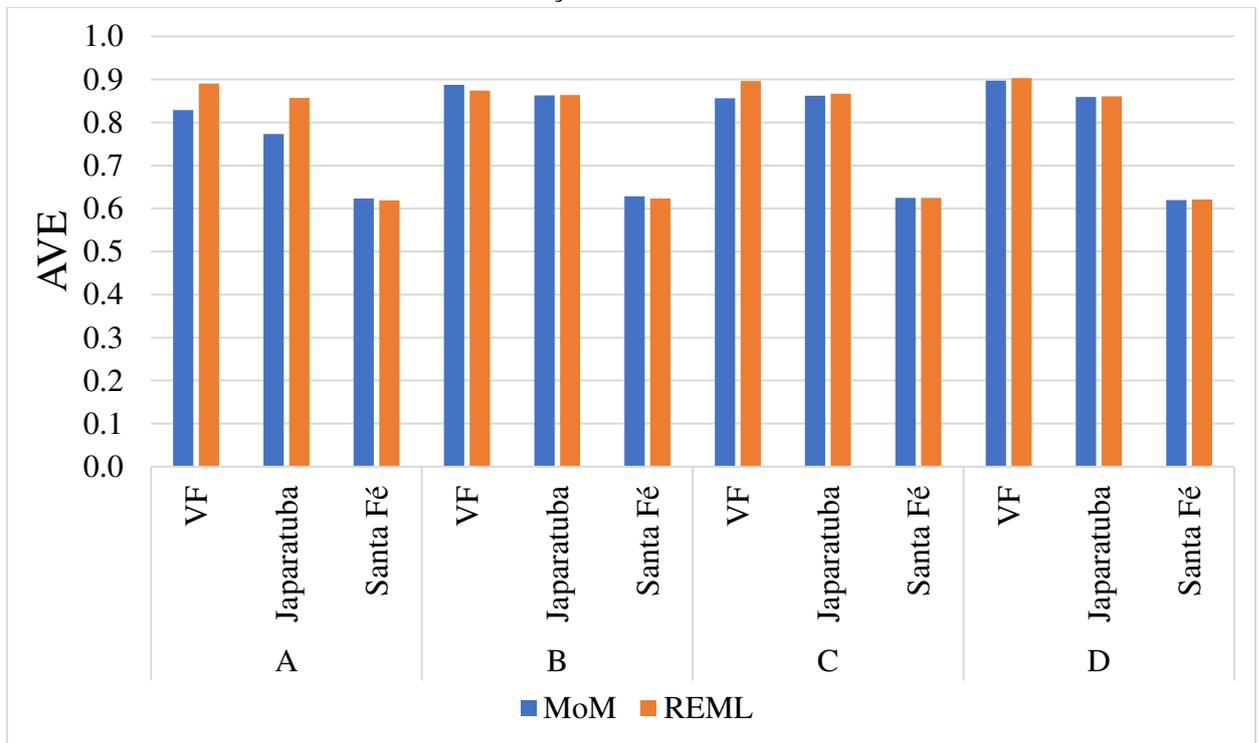
<sup>1</sup>RMSE – raiz quadrada do erro médio (g/kg)

<sup>2</sup>Redução do erro de predição (RMSE) em comparação à condição onde nenhum efeito direcional é contabilizado. (Condição A).

Ao utilizar o REML como forma de modelar o semivariograma ocorreu alteração na importância dos efeitos direcionais. Ao utilizar todos os pontos na modelagem, esse modelo não se beneficia tanto do reconhecimento da tendência quanto o MoM (Tabela 5), mas tem seu desempenho aumentado quando a anisotropia (Condição C) é reconhecida nos dados. A medida de quantidade de melhoria varia de acordo com a característica dos dados interpolados, ou seja, da área VF ou campos experimentais. Por exemplo, o VF apresenta AVE elevado (Figura 10) isto é, acima de 0,80 para todos os cenários avaliados na amostragem densa. Ao utilizar o REML a quantidade de variância explicada (Figura 10) é de no mínimo 0,874 (Condição A – Figura 10) e no máximo de 0,903 (Condição D – Figura 10), enquanto para o MoM, o efeito de melhora do desempenho é menos pronunciado com o valor mínimo de AVE é de 0,829 (Condição A) e o máximo é de 0,897 (Condição D). De fato, a literatura (PARDO-IGÚZQUIZA e DOWD, 1997; OLIVER e WEBSTER, 2007; OLIVER e WEBSTER, 2014) estabelece o REML como mais robusto do que o MoM, daí os resultados similares do REML ao utilizar uma modelagem mais criteriosa em relação ao reconhecimento de efeitos direcionais nas áreas (Tabela 5 e Figura 10). Nas áreas Japarutuba e Santa Fé o efeito é similar ao VF, mas com diferenças menores de ordem de valor, principalmente na Santa Fé (Figura 10), devido ao comportamento mais aleatório. Além disso, nessas áreas o efeito de tendência reduziu o

desempenho do modelo de REML (Condição B), assim quando o cenário de ambos os efeitos direcionais foi adotado, o desempenho do modelo foi pior do que quando utilizado o cenário de apenas anisotropia (Tabela 5), apesar da diferença ser pequena entre as condições. Nesse sentido, entendemos que a modelagem semivariográfica, com efeitos direcionais, está relacionada ao método de estimativa dos parâmetros do modelo semivariográficos (REML ou MoM), sendo mais importante reconhecer o efeito anisotrópico quando se utiliza o REML e o efeito de tendência quando se utiliza o MoM.

**Figura 10**– Desempenho<sup>1</sup> da interpolação utilizando a modelagem semivariográfica com o MoM e REML e os cenários<sup>2</sup> de consideração dos efeitos direcionais.



<sup>1</sup>AVE – quantidade de variância explicada

<sup>2</sup>Cenários: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

#### 4.4 CONSIDERAÇÕES GERAIS

Efeitos direcionais devem ser considerados na modelagem variográfica, independentemente do método utilizado na modelagem (MoM ou REML). A importância de observação da anisotropia já é destacada por Oliver e Webster (2014) que, além desse efeito, ressaltam que deve ocorrer a verificação de tendência na modelagem. Nesse estudo não foi diferente, a contabilização de efeitos direcionais é importante na modelagem, pois aumenta a precisão dos modelos geoestatísticos estudados ao reconhecer melhor algumas fontes de variação na área.

Outro fator essencial para utilização dos modelos geoestatísticos é a necessidade de avaliação da característica dos dados (i.e. estatística descritiva, normalidade), quesito que é requerimento para utilizar a krigagem (WEBSTER e LARK, 2019). Essa investigação dos dados é inclusive debatida por Webster, *et al.* (2016) que criticam a maneira com que muitos usuários de geoestatística fazem suas análises, sem a investigação criteriosa dos dados para aplicar as técnicas. Criticam ainda que muitas vezes os procedimentos são realizados na forma de caixa preta, ou seja, não é observado como o semivariograma é construído ou como a krigagem funciona, apenas aplica-se a técnica a partir das configurações padrão dos softwares para obtenção dos resultados. Além disso, é da análise da característica dos dados que é possível entender a superfície interpolada. Por exemplo, a Santa Fé apresentou pior desempenho das formas de interpolação testadas (geoestatísticos ou determinísticos) do que todas as outras áreas, e esse comportamento de menor desempenho é explicado tanto pelo semivariograma (Tabela 4 e Apêndice 2) quanto pela variância dos dados amostrados e interpolados e coeficientes de variação dos dados amostrados. Isto é, a qualidade da interpolação é dependente da característica dos dados. Desta maneira, cada etapa necessita de investigação, o que leva a um processo minucioso por somar complexibilidade e necessidade de seguir um procedimento passo a passo. Desde a análise dos dados, da modelagem das estruturas de covariância ou semivariograma (WEBSTER *et al.*, 2016) a priori à etapa de interpolação. De fato, a observação de cada parâmetro na modelagem semivariográfica é importante a fim de obter estimativas mais realistas e com maior assertividade.

Além disso, a utilização de modelos geoestatísticos para predição e elaboração de mapas possuem maior complexidade do que os modelos determinísticos, por adicionar a etapa de análise de semivariograma na estimativa dos valores não amostrados. Além do mais, mesmo com maior estabilidade de resultados, o uso do REML na modelagem semivariográfica pode sofrer limitações e ter seu desempenho influenciado pela experiência do usuário. Isso porque, ao utilizar o pacote *gstat*, o resultado da modelagem dos parâmetros (efeito pepita, contribuição e alcance) apresenta sensibilidade ao variar o chute inicial alcance (Tabela 6 – GR). Isso ocorre, pois, a função fixa o alcance determinado pelo usuário, ao contrário do que ocorre no pacote *geoR*, o qual calcula os três parâmetros. Assim, a qualidade das predições com *gstat* é reduzida, principalmente quando o efeito da tendência não é reconhecido nos dados e em amostragem esparsa (Figura 11), ou seja, nos cenários de dados brutos (Condição A) e dados com anisotropia (Condição C). Isto é, quando esses chutes são dados de forma equivocada e os efeitos direcionais estão presentes nos dados, o desempenho da modelagem por REML é reduzido (Tabela 6).

**Tabela 6** - Erro<sup>1</sup> de predição dos métodos de estimativa de parâmetros do semivariograma ao considerar efeitos direcionais na modelagem com pacote *gstat* e ao variar o chute inicial do alcance no REML.

Condição	Modelo	<sup>2</sup> 200 m		<sup>2</sup> 400 m		<sup>2</sup> 950 m	
		RMSE	GR <sup>3</sup>	RMSE	GR <sup>3</sup>	RMSE	GR <sup>3</sup>
Sem efeito	MoM	0,354	GR <sup>3</sup>	0,354	GR <sup>3</sup>	0,354	GR <sup>3</sup>
	REML	0,297		0,283		0,283	
Tendência	MoM	0,287	-0,067	0,287	-0,067	0,287	-0,067
	REML	0,279	-0,019	0,302	0,020	0,289	0,005
Anisotropia	MoM	0,323	-0,031	0,323	-0,031	0,323	-0,031
	REML	0,454	0,156	0,317	0,035	0,275	-0,008
Ambos os efeitos	MoM	0,274	-0,080	0,274	-0,080	0,274	-0,080
	REML	0,323	0,026	0,266	-0,017	0,273	-0,010

<sup>1</sup>RMSE – raiz quadrada do erro médio (g/kg)

<sup>2</sup>Chutes iniciais do alcance fixados para estimativa do semivariograma

<sup>3</sup>GR – Ganho relativo aos dados brutos

Ademais, a consideração de efeito anisotrópico é importante, mas ainda é comum encontrar na literatura a modelagem sem considerar tal efeito (OLIVER e WEBSTER, 2014). Por exemplo, BETZEK *et al.*, (2019) encontram bons resultados ao automatizarem a seleção dos parâmetros semivariográficos através do MoM, mesmo sem considerar presença de anisotropia nos dados. Nossos resultados mostram que para o método utilizado apenas o uso da tendência é suficiente para promover melhor desempenho (Tabela 5) do modelo geoestatístico ao se trabalhar com o MoM. Dessa maneira, os resultados de BETZEK *et al.*, (2019) não foram prejudicados ao não considerar efeito anisotrópico. Nesse sentido, como um dos métodos mais utilizados para estimar os parâmetros semivariográficos é o MoM, a interpolação traz resultados similares quando os efeitos são ou não considerados (Tabelas 1, 2 e 5) e o fato de cientistas e, principalmente usuários, ainda desprezarem o efeito anisotrópico na modelagem pode ser atribuído ao ganho pequeno de desempenho do modelo enquanto a complexidade do modelo é elevada, pois é necessária análise criteriosa do mapa variográfico para encontrar o ângulo e posteriormente o maior e menor alcance do semivariograma direcional.

Além disso, a maior complexidade da interpolação geoestatística é somada a interface do software estatístico utilizado nesse estudo (R + RStudio) que não está na forma mais amigável para usuários. E para melhor uso do software é interessante que as ferramentas estejam de forma reproduzível e que forneçam informações claras e concisas para evitar problemas de configurações equivocadas na análise, como por exemplo, ao utilizar chute inicial

equivocado utilizando o método REML. Por outro lado, alguns softwares que possuem interface mais amigável assim como o ArcGIS, não são de livre acesso e apresentam procedimento de interpolação na forma de caixa preta no sentido de que, mesmo sem muito conhecimento dos processos é possível gerar mapas interpolados.

## 5 CONCLUSÃO

A qualidade dos mapas obtidos por interpolação de dados é influenciada significativamente pela consideração de efeitos direcionais na modelagem de semivariogramas. Assim, os efeitos direcionais devem ser considerados quando a krigagem for utilizada, permitindo melhor representação da variabilidade presente nas áreas estudadas. Ademais, a consideração dos efeitos direcionais continua sendo importante mesmo ao reduzir o número de pontos na área. Todavia, a importância dos efeitos direcionais é diferente em função do método de estimativa dos parâmetros do semivariograma (MoM e REML). Assim, sugerimos que ao utilizar o MoM como método de estimativa dos parâmetros do semivariograma, geralmente se mostra mais interessante considerar apenas o efeito de tendência; enquanto considerar apenas o efeito da anisotropia nos dados geralmente resulta em melhor desempenho das estimativas por REML.

## REFERÊNCIAS

AMARAL, L. R. do; JUSTINA, D. D. Della. Spatial Dependence Degree and Sampling Neighborhood Influence on Interpolation Process for Fertilizer Prescription Maps. **Engenharia Agrícola**, v. 39, n. spe, p. 85–95, 2019.

ASSIS SILVA, S. et al. Avaliação de interpoladores estatísticos e determinísticos na estimativa de atributos do solo em agricultura de precisão. **Idesia**, v. 26, n. 2, p. 75–82, 2008.

BETZEK, N. M. et al. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, v. 157, n. November 2018, p. 49–62, 2019a. Disponível em: <<https://doi.org/10.1016/j.compag.2018.12.004>>.

BETZEK, N. M. et al. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, v. 157, p. 49–62, 1 fev. 2019b.

BORSSOI, J. A.; URIBE-OPAZO, M. A.; ROJAS, M. G. Diagnostic Techniques Applied in Geostatistics for Agricultural Data Analysis. **R. Brasil. Ci solo**, v. 33, n. 1, p. 1561–1570, 2009.

BRUS, D. J.; DE GRUIJTER, J. J.; VAN GROENIGEN, J. W. Chapter 14 Designing Spatial Coverage Samples Using the k-means Clustering Algorithm. **Developments in Soil Science**, v. 31, n. C, p. 183–192, 2006.

CHERUBIN, M. R. et al. Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. **Ciencia Rural**, v. 44, n. 3, p. 425–432, 2014.

CRESSIE, N. The Origins of Kriging 1. v. 22, n. 3, p. 239–252, 1990.

CRESSIE, N. A. C. **Statistics for Spatial Data, Revised Edition**. **Biometrics**, 1994. .

GARCÍA-TOMILLO, A. et al. Estimating soil organic matter using interpolation methods with a electromagnetic induction sensor and topographic parameters: a case study in a humid region. **Precision Agriculture**, v. 18, n. 5, p. 882–897, 2017.

GEBBERS, R.; HERBST, R.; AGRICULTURE, F. P. Sensitivity analysis of soil nutrient mapping. **EFITA conference**, v. 519, p. 513–519, 2009.

HENGL, T. et al. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. **PeerJ**, v. 2018, n. 8, 2018.

JURADO-EXPÓSITO, M. et al. Papaver rhoeas L. mapping with cokriging using UAV imagery. **Precision Agriculture**, n. 0123456789, 2019. Disponível em: <<https://doi.org/10.1007/s11119-019-09635-z>>.

KIS, I. M. Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable ( a case of linear spatial trend ) , case study of the Šandrovac Field. **The Mining-Geology-Petroleum Engineering Bulletin**, p. 41–58, 2016.

KRAVCHENKO, A.; BULLOCK, D. G. A comparative study of interpolation methods for mapping soil properties. **Agronomy Journal**, v. 91, n. 3, p. 393–400, 1999.

KRŪMINIENE, I. Analysis of anisotropic variogram models for prediction of the curonian lagoon data. **Mathematical Modelling and Analysis**, v. 11, n. 1, p. 73–86, 2006.

LARK, R. M. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. **European Journal of Soil Science**, v. 51, n. 4, p. 717–728, 2000.

LARK, R. M.; CULLIS, B. R.; WELHAM, S. J. On spatial prediction of soil properties in the presence of a spatial trend: The empirical best linear unbiased predictor (E-BLUP) with REML. **European Journal of Soil Science**, v. 57, n. 6, p. 787–799, dez. 2006.

LARK, R. M.; WEBSTER, R. Geostatistical mapping of geomorphic variables in the presence of trend. **Earth Surface Processes and Landforms**, v. 31, n. 7, p. 862–874, 2006.

LI, J.; HEAP, A. D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. **Ecological Informatics**, v. 6, n. 3–4, p. 228–241, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.ecoinf.2010.12.003>>.

MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. F. **Agricultura de precisão**. 1. ed. São Paulo: Oficina de Textos, 2015a.

MOLIN, J. P.; AMARAL, L. R.; COLAÇO, A. **Agricultura de Precisão Autores: José Paulo Molin; Lucas Rios do Amaral; André Freitas Colaço**. [s.l: s.n.]

NANNI, M. R. et al. Optimum size in grid soil sampling for variable rate application in site-specific management. **Scientia Agricola**, v. 68, n. 3, p. 386–392, 2011.

OLIVER, M. A.; WEBSTER, R. **A tutorial guide to geostatistics: Computing and modelling variograms and kriging** Catena Elsevier B.V., , 2014. . Disponível em: <<http://dx.doi.org/10.1016/j.catena.2013.09.006>>.

PARDO-IGÚZQUIZA, E.; DOWD, P. A. AMLE3D: A computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. **Computers and Geosciences**, v. 23, n. 7, p. 793–805, 1997.

ROBINSON, T. P.; METTERNICHT, G. Testing the performance of spatial interpolation techniques for mapping soil properties. **Computers and Electronics in Agriculture**, v. 50, n. 2, p. 97–108, 2006.

SANCHES, G. M. et al. Potential of apparent soil electrical conductivity to describe the soil pH and improve lime application in a clayey soil. **Soil and Tillage Research**, v. 175, n. September 2017, p. 217–225, 2018. Disponível em: <<http://dx.doi.org/10.1016/j.still.2017.09.010>>.

SCHMITT, P.; MANDEL, J.; GUEDJ, M. A Comparison of Six Methods for Missing Data Imputation. **Journal of Biometrics & Biostatistics**, v. 06, n. 01, p. 1–6, 2015.

TAYLOR, P.; OLIVER, M. A.; WEBSTER, R. International journal of geographical information systems Kriging : a method of interpolation for geographical information systems. **Geographical**, n. October 2011, p. 37–41, 2007.

WEBSTER, R. OLIVER, M. A. **Geostatistics for Environmental Scientists**. [s.l: s.n.]v. 1

WEBSTER, R.; LARK, R. M. Analysis of variance in soil research : Examining the assumptions. n. October 2018, p. 990–1000, 2019.

WEBSTER, R.; M.A. OLIVER; LARK, R. M. Editorial : statistics in the journal. n. March, p. 133–134, 2016.

WEBSTER, R.; OLIVER, M. A. **Statistical methods in soil and land resource survey**. Oxford: Oxford University Press, 1990.

WEBSTER, R.; OLIVER, M. A. Sample adequately to estimate variograms of soil properties. **Journal of Soil Science**, v. 43, n. 1, p. 177–192, 1992.

YAMAMOTO, JORGE K.; LANDIM, P. M. B. **Jorge Kazuo Yamamoto Paulo M. Barbosa Landim**. [s.l: s.n.]

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geostatística: conceitos e aplicações**. 1. ed. São Paulo: Oficina de Textos, 2013.

## APÊNDICE 1 – Modelagem dos Campos Virtuais – Simulação Incondicional Gaussiana

A simulação incondicional é uma aplicação da técnica de Montes Carlo e corresponde à criação de valores a partir de uma função de covariância ou semivariograma (WEBSTER; OLIVER, 2007). No caso de Simulação incondicional gaussiana, as variáveis aleatórias são geradas de acordo com um modelo semivariográfico (contendo efeito pepita, contribuição e alcance) sob a condição gaussiana, apresentando distribuição normal entre os dados. Para cada simulação desses valores, tem-se uma realização da função aleatória  $Z(x)$ .

Utilizei software R versão 3.4.4, combinado com RStudio, para geração dos campos. O pacote geostatístico que foi utilizado foi o `gstat`. Mas utilizei outros pacotes ao longo do código como o pacote *raster e sp*.

Gerei grade para geração dos valores estimados pelo modelo:

```
x = seq (1, 1400, by = 10)
y = seq (1, 1450, by = 10)
area = expand.grid(x = x, y = y)
n = nrow(area)
names(area) <- c('x','y')
```

Transformei essa grade em polígono para fixar resolução. O polígono em `.shp` foi gerado no software QGIS versão 2.18.26

```
boundary = shapefile("../cotorno.shp")
r = raster(boundary, res = 10)
rp = rasterize(boundary, r, 0)
grid = as(rp, "SpatialPixelsDataFrame")
xy1 = as.data.frame(grid)
xy = xy1[, c(2,3)]
```

Determinei o modelo para realizar a simulação

Os coeficientes  $\beta$  foram extraídos de uma área de 116 hectares com dependência espacial, tendência de primeira ordem e anisotropia. A direção preferencial, contribuição e alcance também foram baseados nessa área. A tendência é adicionada ao adicionar “ $\sim x+y$ ” na fórmula.

```
set.seed(93)
model = gstat(formula= z ~ x+y,
              locations = ~ x+y,
```

```
dummy=T,  
beta=c(1.6746398,0.0016447,0.0003655),  
model=vgm(nugget =0, psill= 0.2833884,  
          range= 391.5041, model='Sph',  
          anis = c(30, 0.5)),  
nmin = 4, nmax=500, data = xy)
```

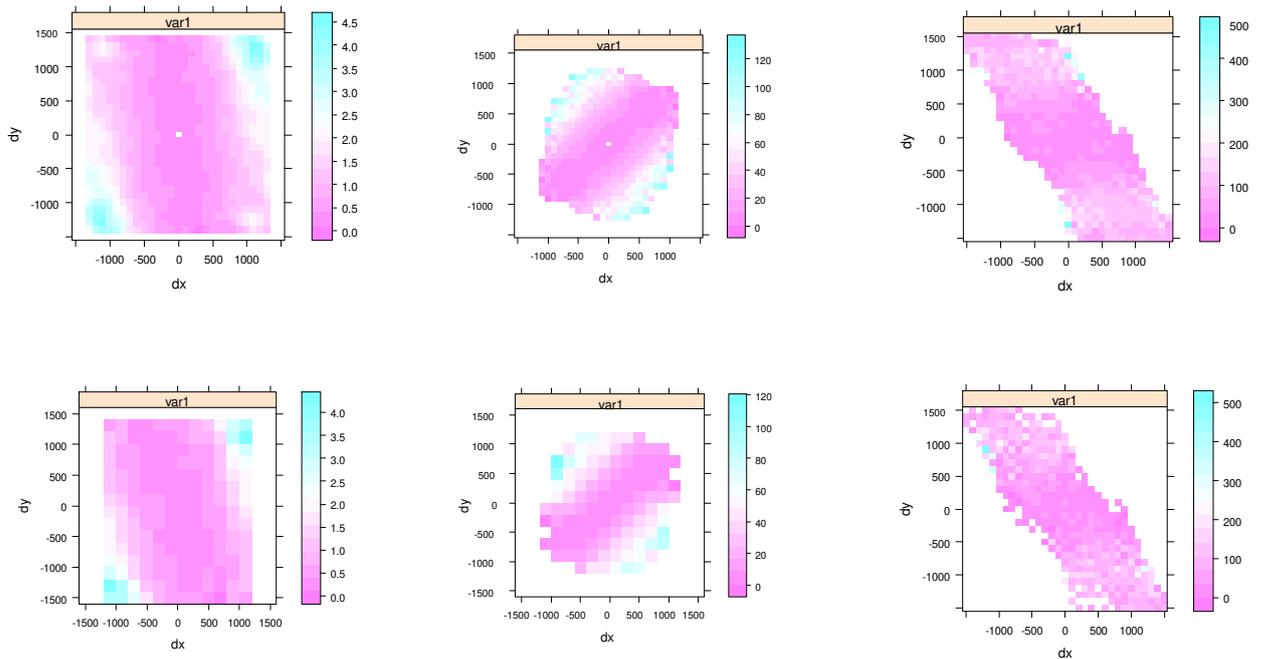
Com o modelo determinado, realizei a Simulação das 100 realizações da Simulação Incondicional Gaussiana.

```
set.seed(1993)  
vf = predict(model, newdata=xy, nsim= 100)
```

Esse processo levou aproximadamente 21 minutos em um computador com 8 GB de memória RAM, processador Inter® Core™ i-7 4700HQ CPU @ 2.40 GHz. O sistema instalado é Windows 10 com sistema operacional de 64 bits. Versão utilizada do R foi a 3.4.4 e do RStudio 1.1.463.

## Apêndice 2 – Características dos Campos Virtuais e Experimentais

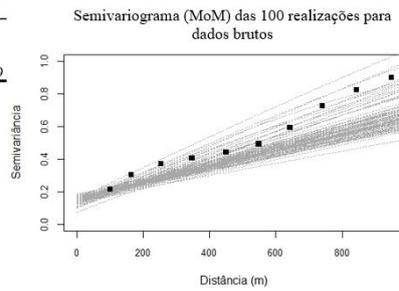
Os Campos Virtuais (VFs) gerados conforme código descrito no Apêndice 1 e os Campos Experimentais (Japarutuba e Santa Fé) foram amostrados nas densidades de uma amostra/ha e 0,25 amostras/ha. Ao analisar os dados provenientes dessas amostragens, para estimativa na presença de anisotropia, foi encontrada anisotropia em todos os dados a partir do mapa semivariográfico:



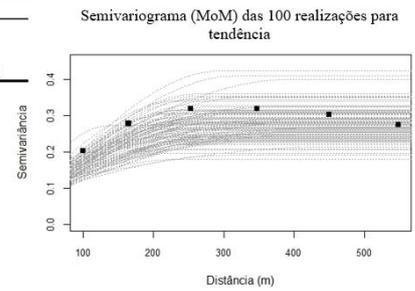
Além da anisotropia, foi verificada a presença de tendência em todos os dados, incluindo nas 100 realizações dos VFs, e foram gerados os semivariogramas nas seguintes condições: A) sem consideração de efeitos direcionais; B) consideração do efeito da tendência; C) consideração do efeito anisotrópico e D) consideração de ambos os efeitos direcionais.

Assim, temos os semivariogramas dos VFs para amostragem densa:

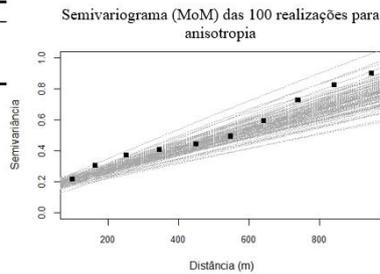
Dados brutos	
Efeito Pepita	0,1502
Contribuição	5,5830
Alcance	14594,3329



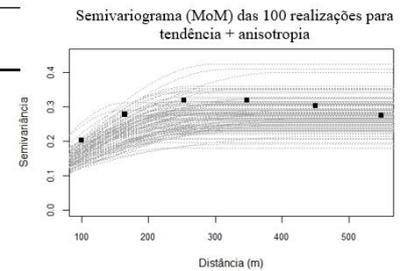
Dados Tendência	
Efeito Pepita	0,0392
Contribuição	0,2389
Alcance	285,4521



Dados Anisotropia	
Efeito Pepita	0,1445
Contribuição	6,8889
Alcance	15777,4128



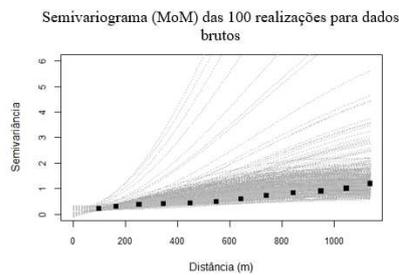
Dados Ambos os Efeitos	
Efeito Pepita	0,0389
Contribuição	0,2392
Alcance	284,1263



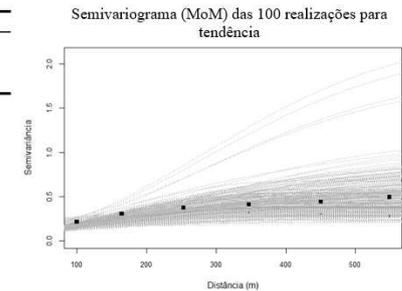
\*Parâmetros de semivariograma são referentes à média das 100 realizações do VF.

Ao realizar a modelagem dos VFs na amostragem esparsa:

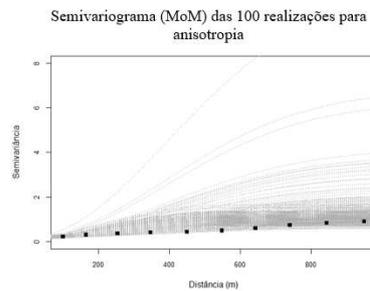
Dados brutos	
Efeito Pepita	0,6208
Contribuição	2,2074
Alcance	8945,8187



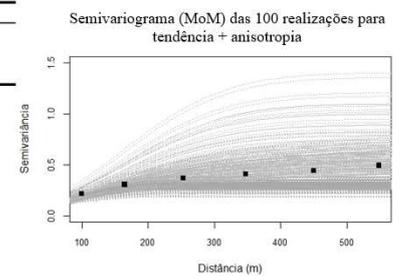
Dados Tendência	
Efeito Pepita	0,2101
Contribuição	0,1001
Alcance	568,4788



Dados Anisotropia	
Efeito Pepita	0,6516
Contribuição	2,5831
Alcance	9432,6371

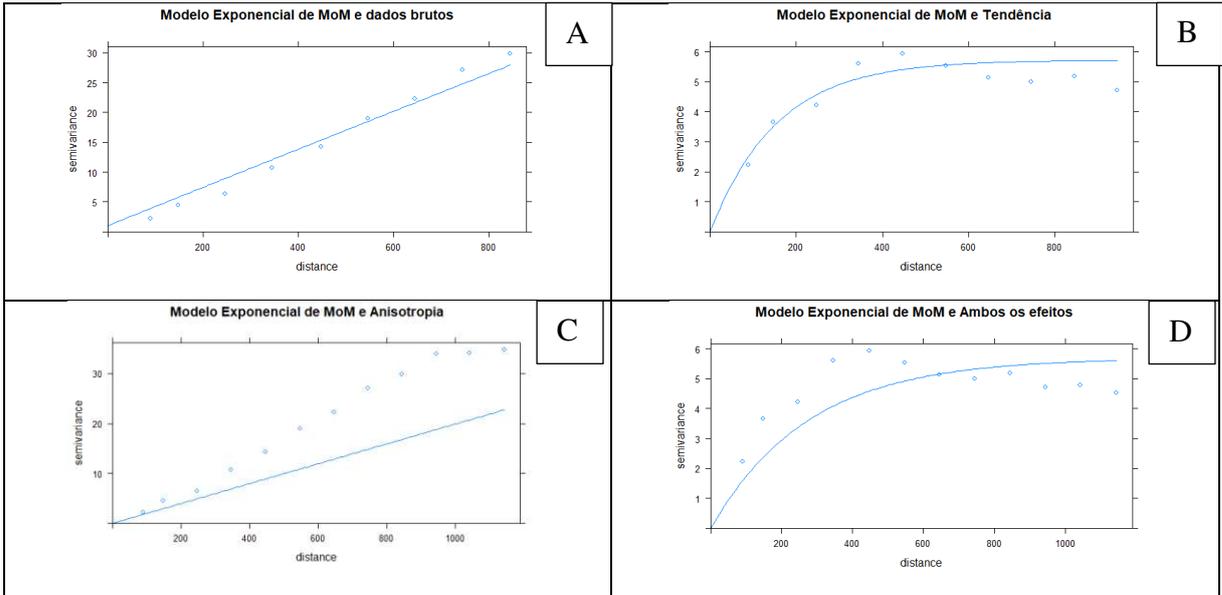


Dados Ambos os Efeitos	
Efeito Pepita	0,2082
Contribuição	0,0974
Alcance	572,8431



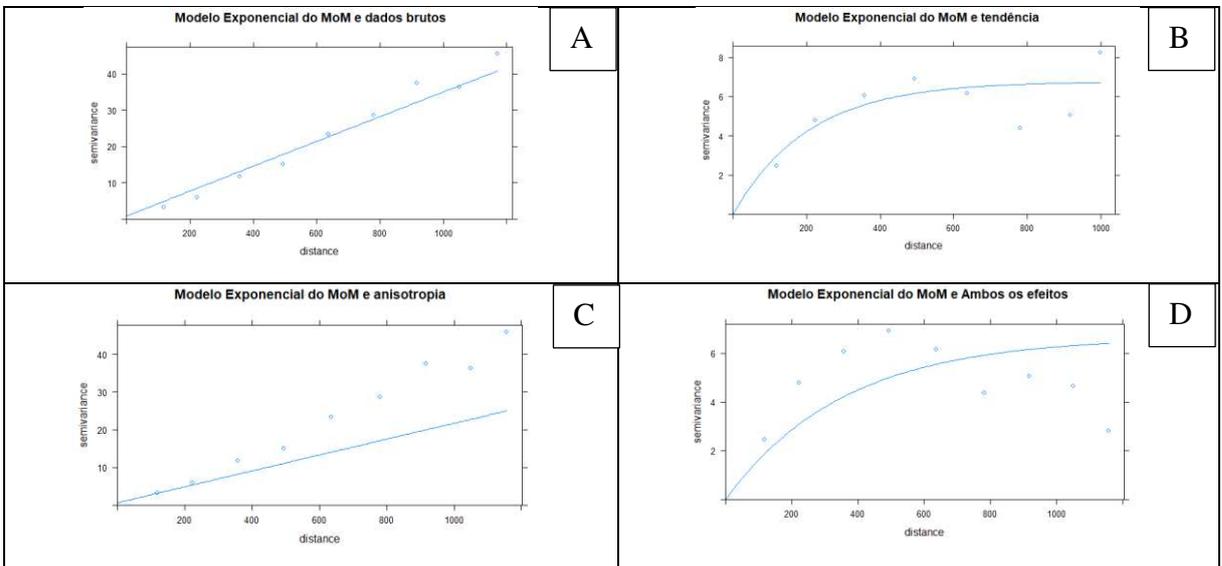
\*Parâmetros de semivariograma são referentes à média das 100 realizações do VF.

Os semivariogramas da consideração dos efeitos direcionais, nos diferentes cenários, na Japarutuba para amostragem densa foram:



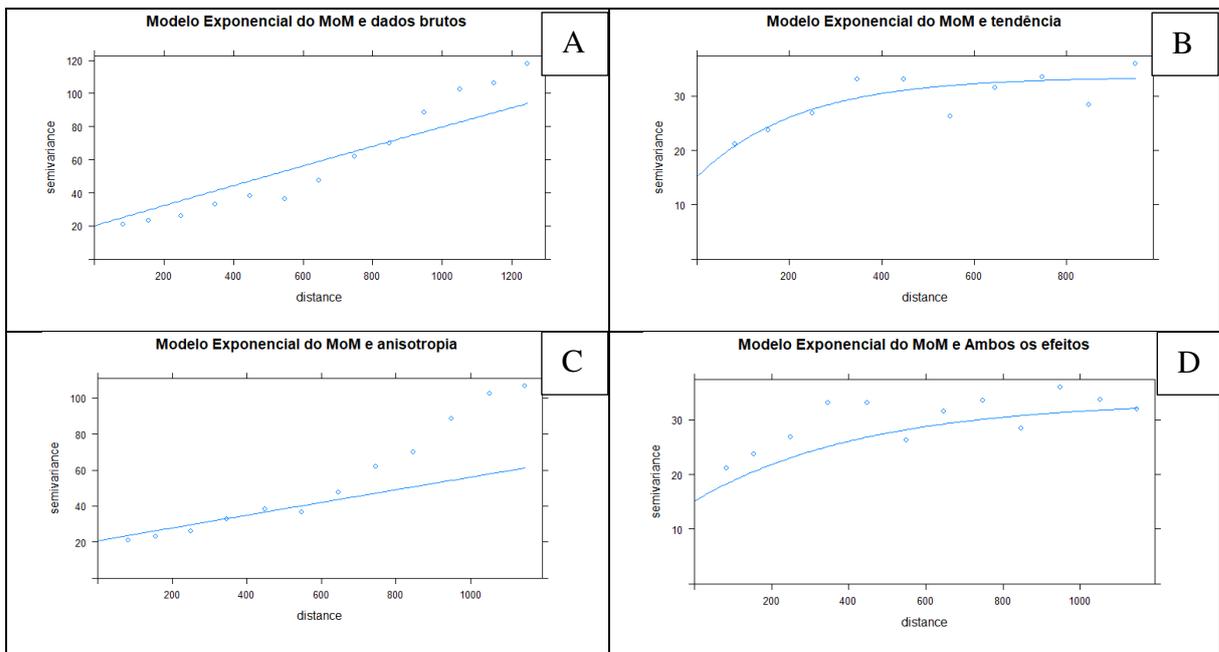
Amostragem densa				
Modelo	Condição	efeito pepita	contribuição	alcance
MoM	A	1,058	1434,157	44694,88
	B	0	5,722	154,599
	C	0,002	4191,259	210403,2
	D	0	5,692	273,53
REML	A	0,135	31,279	1000
	B	0,36	11,795	420
	C	0,142	26,266	1100
	D	0,228	9,307	400

Ao realizar a modelagem na amostragem esparsa:



Amostragem esparsa				
Modelo	Condição	efeito pepita	contribuição	alcance
MoM	A	0,806	1882,99	54576,64
	B	0	6,776	205,001
	C	0,632	2213,305	104582,3
	D	0	6,668	356,207
REML	A	2,218	51,25	1000
	B	2,658	24,388	420
	C	2,789	51,747	1200
	D	3,11	23,785	500

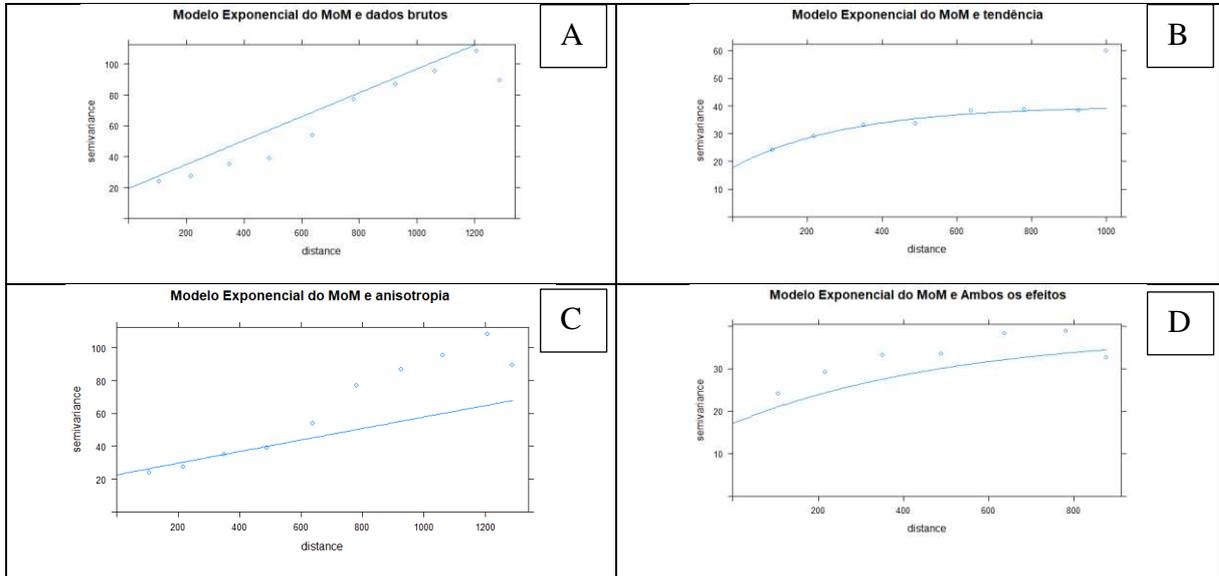
A Área Santa Fé, apresentou os semivariogramas da consideração dos efeitos direcionais para amostragem densa:



Amostragem densa				
Modelo	Condição	efeito pepita	contribuição	alcance
MoM	A	20,179	1236,599	20220,1
	B	15,177	18,297	221,739
	C	20,767	1572,223	44088,8
	D	15,141	18,264	439,382
REML	A	16,347	60,109	1200
	B	13,77	23,921	250

C	14,337	49,265	1100
D	10,571	24,12	250

Ao realizar a modelagem na amostragem esparsa:



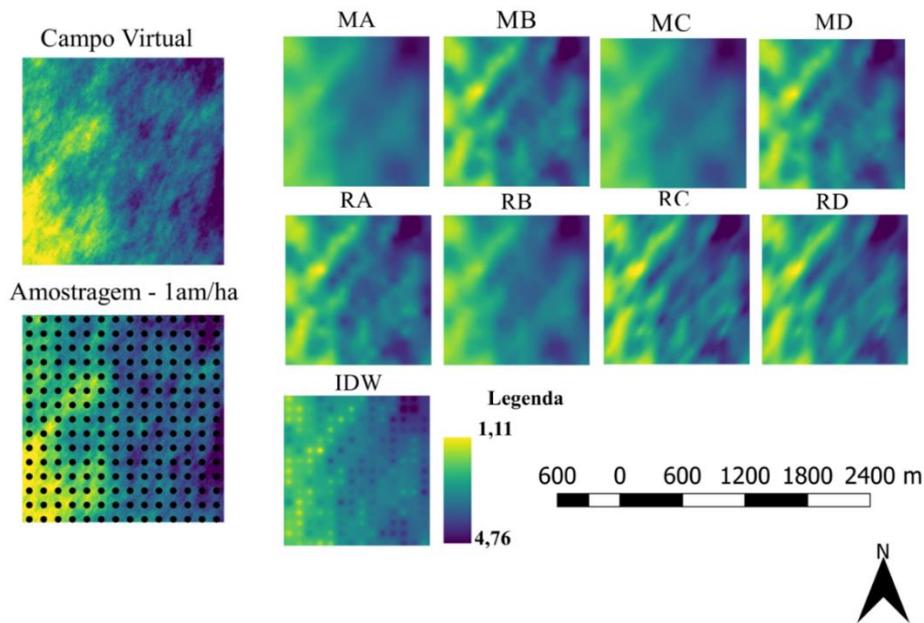
Amostragem esparsa				
Modelo	Condição	efeito pepita	contribuição	alcance
MoM	A	21,962	1479,239	24061,8
	B	17,853	22,116	307,904
	C	22,57	1934,619	54281,7
	D	17,117	21,376	523,627
REML	A	18,144	62,618	1200
	B	11,846	38,251	300
	C	15,61	52,561	1100
	D	10,351	48,741	550

Os resultados da de validação cruzada *leave-one-out* por REML e por MoM com modelo selecionados foram:

Campos virtuais					Área 1/ Japarutuba					Área 2/Santa Fé				
Amostragem 1					Amostragem 1					Amostragem 1				
Modelo	Condição	RMSE	R <sup>2</sup>	Modelo	Modelo	Condição	RMSE	R <sup>2</sup>	Modelo	Modelo	Condição	RMSE	R <sup>2</sup>	Modelo
MoM	Dados brutos	0,428	0,75	Esférico	MoM	Dados brutos	1,838	0,821	Exponencial	MoM	Dados brutos	4,785	0,643	Exponencial
	Efeito tendência	0,39	0,792	Esférico		Efeito tendência	1,796	0,829	Exponencial		Efeito tendência	4,833	0,636	Exponencial
	Efeito anisotropia	0,424	0,755	Esférico		Efeito anisotropia	1,871	0,815	Exponencial		Efeito anisotropia	4,687	0,658	Exponencial
	Ambos os efeitos	0,391	0,791	Esférico		Ambos os efeitos	1,826	0,823	Exponencial		Ambos os efeitos	4,7	0,656	Exponencial
REML	Dados brutos	0,393	0,789	Esférico	REML	Dados brutos	1,844	0,82	Exponencial	REML	Dados brutos	4,789	0,643	Exponencial
	Efeito tendência	0,405	0,776	Esférico		Efeito tendência	1,823	0,824	Exponencial		Efeito tendência	4,838	0,636	Exponencial
	Efeito anisotropia	0,362	0,82	Esférico		Efeito anisotropia	1,88	0,813	Exponencial		Efeito anisotropia	4,684	0,658	Exponencial
	Ambos os efeitos	0,351	0,831	Esférico		Ambos os efeitos	1,835	0,822	Exponencial		Ambos os efeitos	4,734	0,651	Exponencial
IDW	-	0,5	0,727	-	IDW	-	2,253	0,823	-	IDW	-	4,902	0,632	-
Amostragem 2					Amostragem 2					Amostragem 2				
Modelo	Condição	RMSE	R <sup>2</sup>	Modelo	Modelo	Condição	RMSE	R <sup>2</sup>	Modelo	Modelo	Condição	RMSE	R <sup>2</sup>	Modelo
MoM	Dados brutos	0,552	0,595	Esférico	MoM	Dados brutos	1,897	0,818	Exponencial	MoM	Dados brutos	5,166	0,592	Exponencial
	Efeito tendência	0,527	0,618	Esférico		Efeito tendência	1,883	0,82	Exponencial		Efeito tendência	5,227	0,582	Exponencial
	Efeito anisotropia	0,55	0,595	Esférico		Efeito anisotropia	1,928	0,812	Exponencial		Efeito anisotropia	5,092	0,603	Exponencial
	Ambos os efeitos	0,527	0,618	Esférico		Ambos os efeitos	1,914	0,815	Exponencial		Ambos os efeitos	5,113	0,6	Exponencial
REML	Dados brutos	0,567	0,557	Esférico	REML	Dados brutos	1,961	0,808	Exponencial	REML	Dados brutos	5,167	0,592	Exponencial
	Efeito tendência	0,539	0,602	Esférico		Efeito tendência	1,963	0,81	Exponencial		Efeito tendência	5,144	0,595	Exponencial
	Efeito anisotropia	0,581	0,536	Esférico		Efeito anisotropia	1,902	0,818	Exponencial		Efeito anisotropia	5,046	0,611	Exponencial
	Ambos os efeitos	0,539	0,602	Esférico		Ambos os efeitos	1,847	0,829	Exponencial		Ambos os efeitos	4,937	0,627	Exponencial
IDW	-	0,603	0,58	-	IDW	-	2,271	0,819	-	IDW	-	5,367	0,569	-

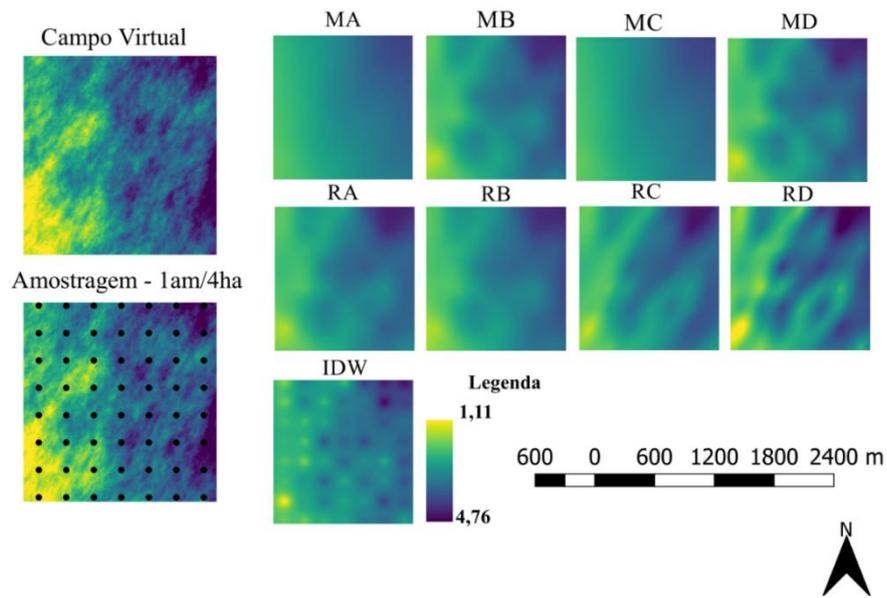
### Apêndice 3 – Superfícies interpoladas pelos cenários propostos.

A modelagem geoestatística (com semivariograma de acordo com Apêndice 2) e determinística (IDW) foi aplicada aos dados amostrados no VF (campo virtual) e Campo Experimental (Japarutuba e Santa Fé). As áreas interpoladas pelos diferentes modelos apresentaram comportamento similiar na amostragem densa. Já densidade esparsa apresenta superfícies mais homogêneas.



\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

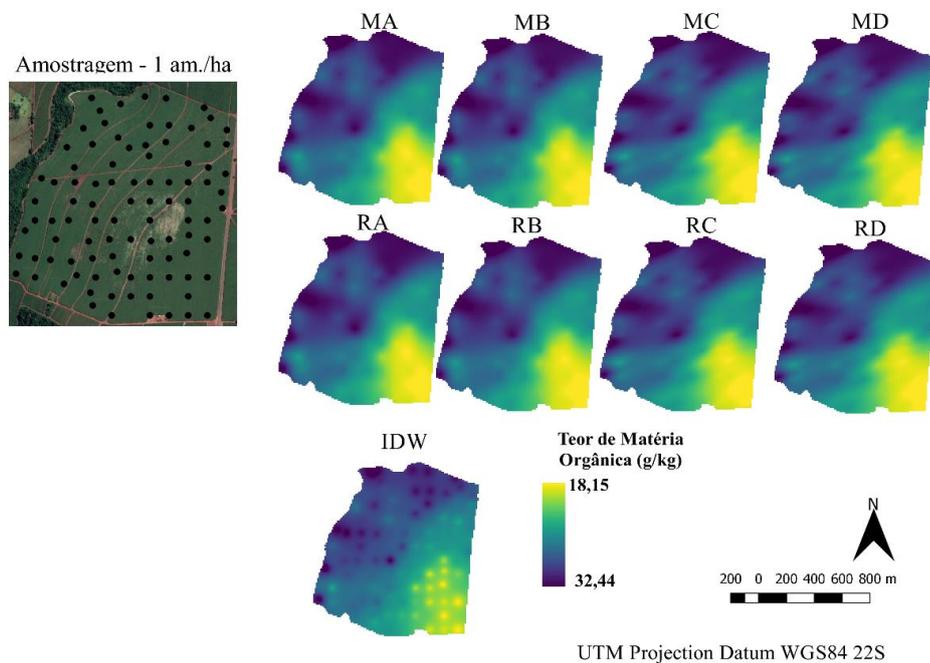
\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;



\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

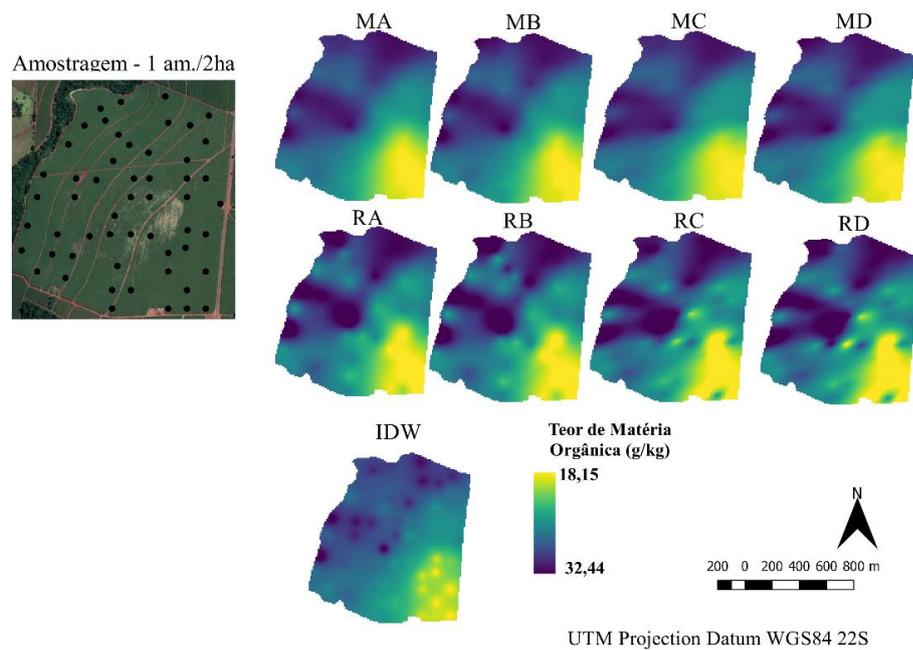
\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;

### Área Japaratuba:



\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

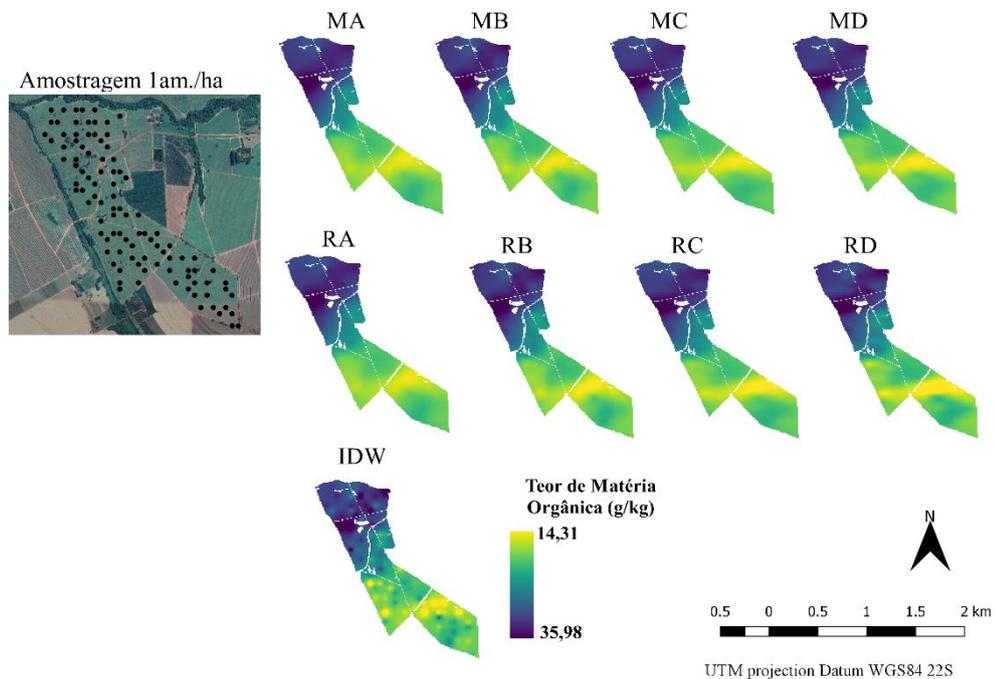
\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;



\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

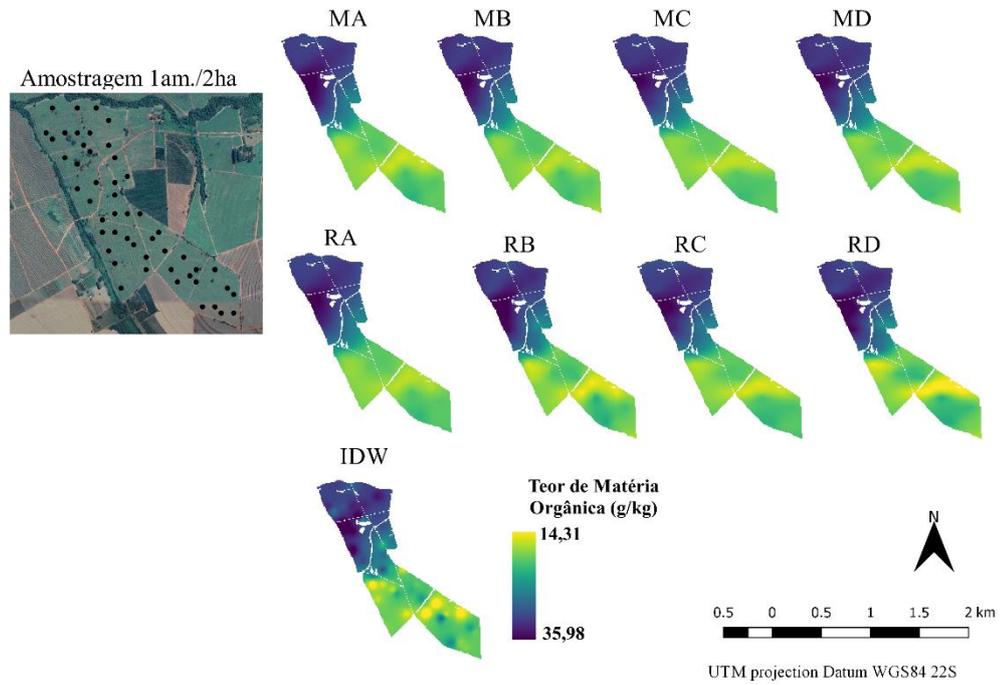
\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;

#### Área Santa Fé:



\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;

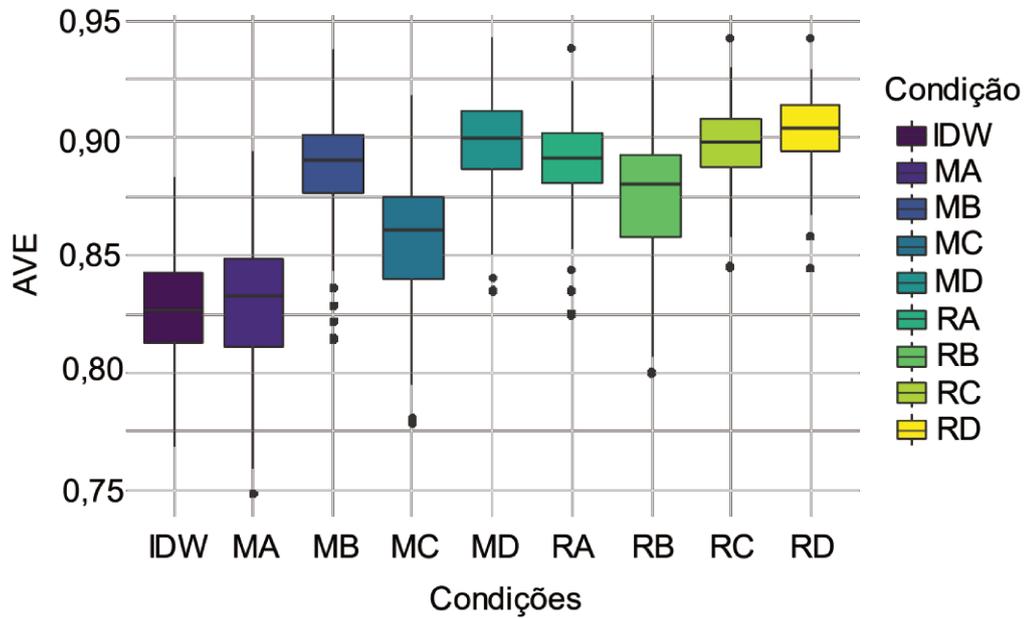


\*As condições são representadas por: A – sem reconhecimento de efeito direcional; B – reconhecimento de apenas efeito de tendência; C – reconhecimento de apenas efeito de anisotropia; D – reconhecimento de ambos os efeitos direcionais.

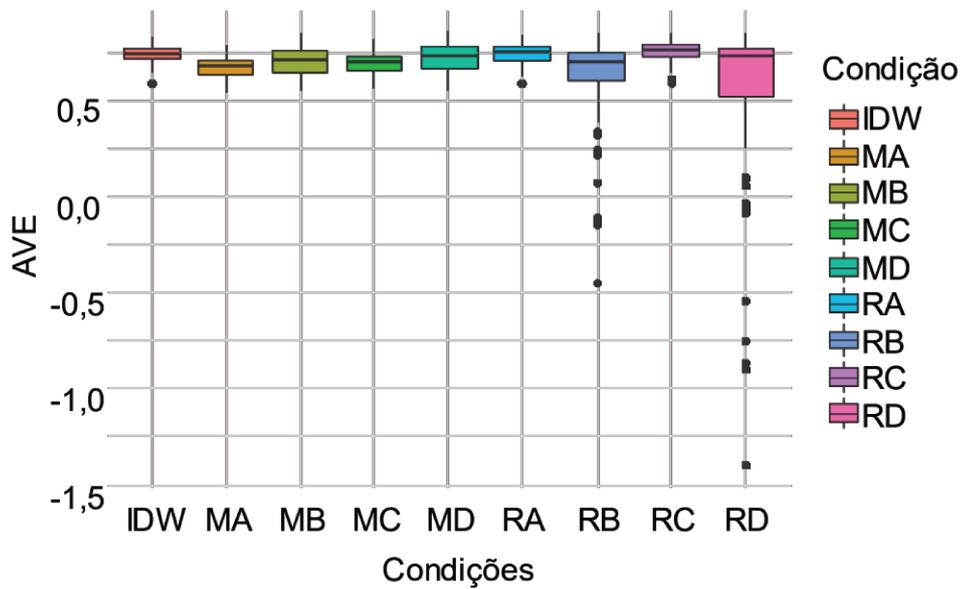
\*\*Os métodos de estimativas são representados pela inicial R (REML) e M (MoM) junto com as condições em cada superfície interpolada;

#### Apêndice 4 – Estatísticas das interpolações nas 100 realizações dos Campos Virtuais

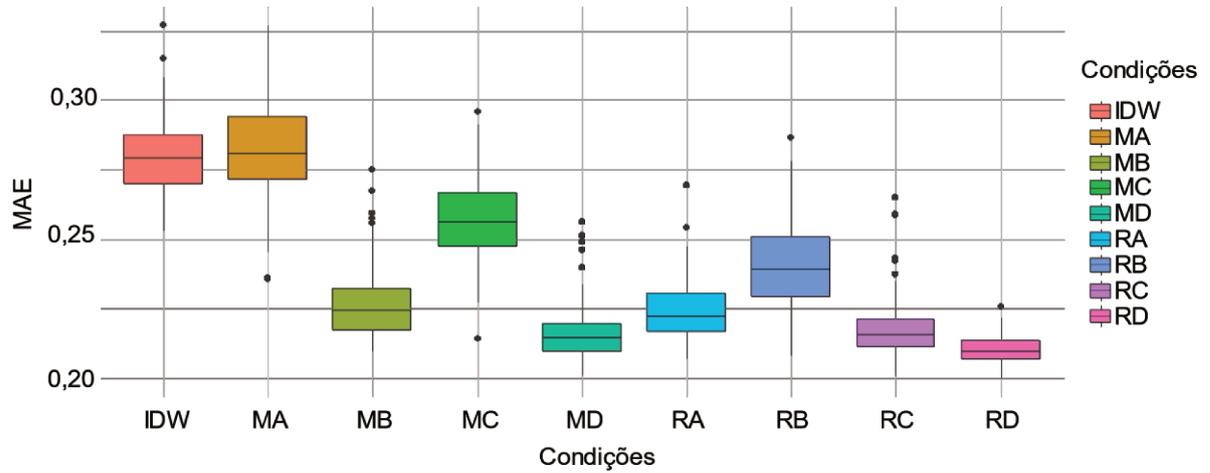
Os resultados de AVE das 9 condições dos 100 VFs na amostragem densa:



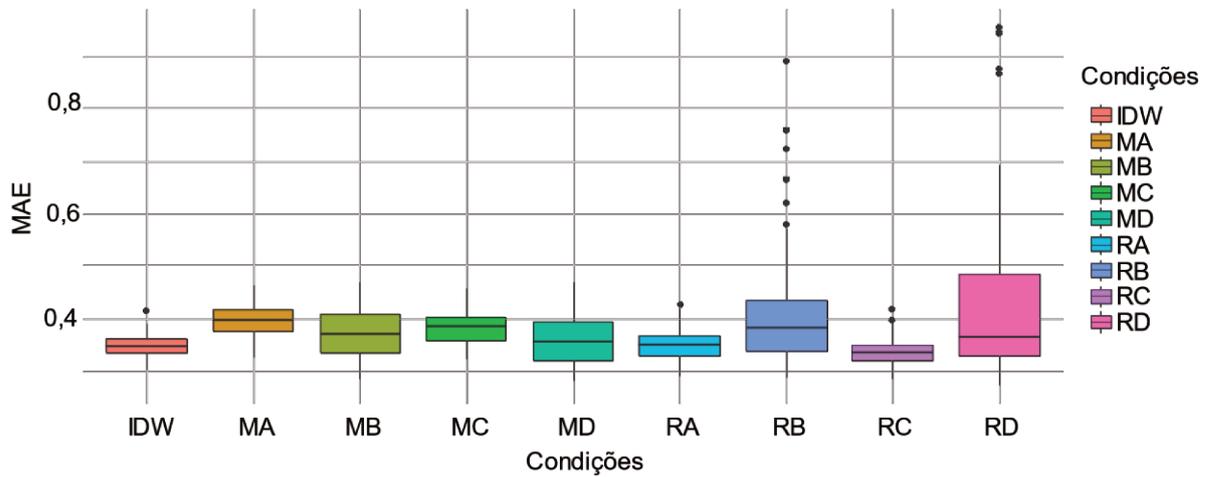
Os resultados de AVE das 9 condições dos 100 VFs na amostragem esparsa:



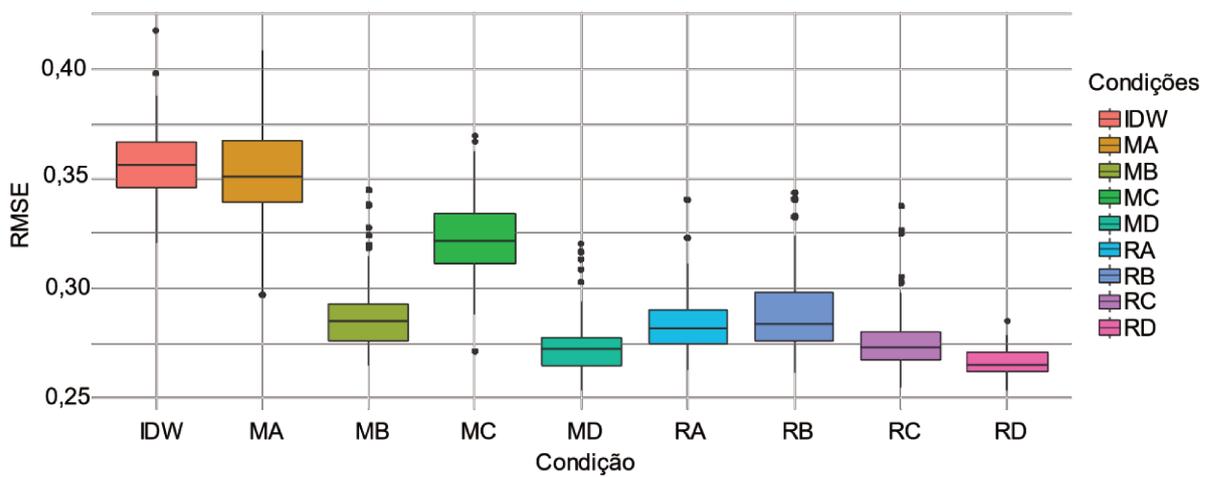
Os resultados de MAE das 9 condições dos 100 VFs na amostragem densa:



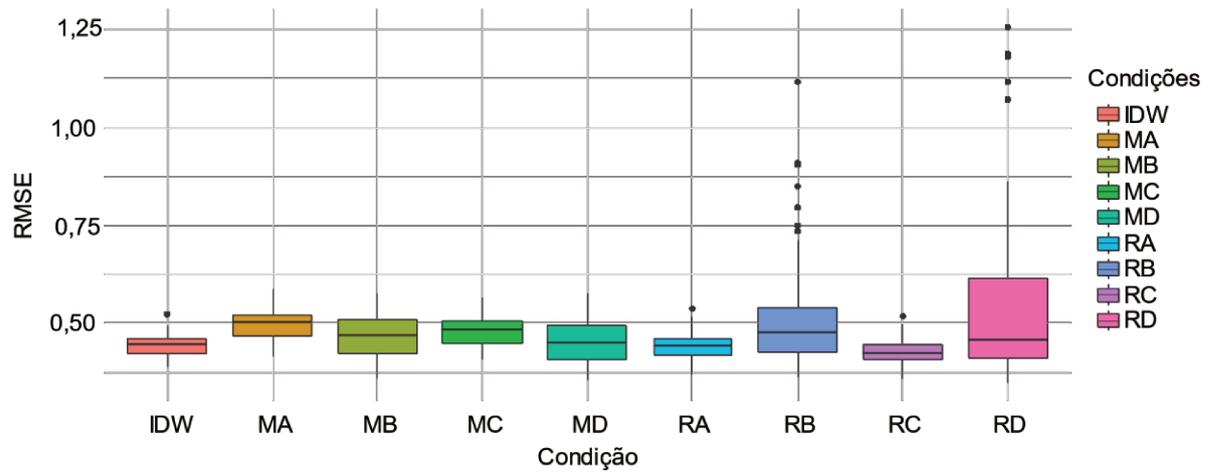
Os resultados de MAE das 9 condições dos 100 VFs na amostragem esparsa:



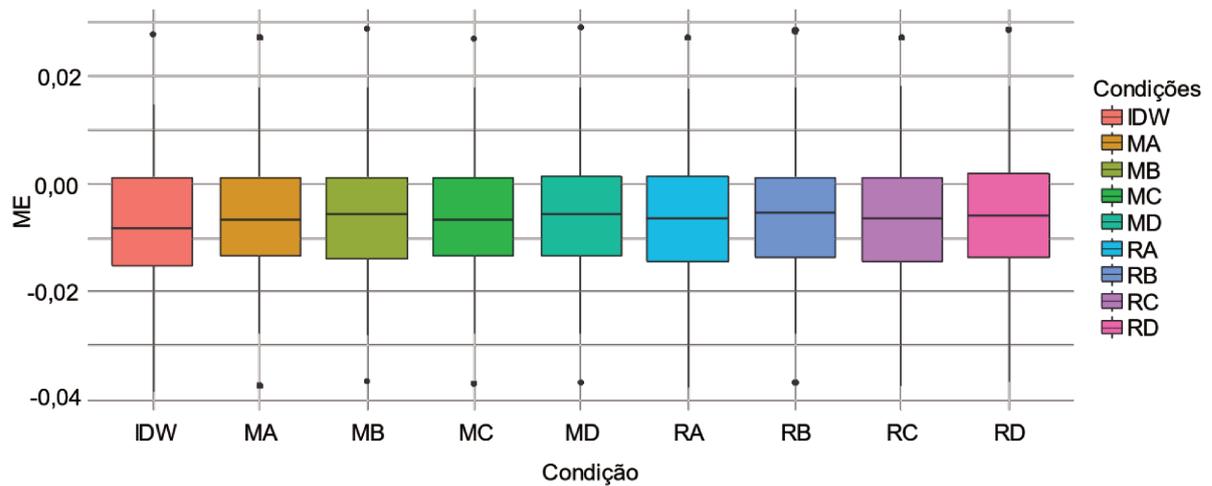
Os resultados de RMSE das 9 condições dos 100 VFs na amostragem densa:



Os resultados de RMSE das 9 condições dos 100 VFs na amostragem esparsa:



Os resultados de ME das 9 condições dos 100 VFs na amostragem densa:



Os resultados de ME das 9 condições dos 100 VFs na amostragem esparsa:

