UNIVERSIDADE FEDERAL DE PELOTAS
Faculdade De Agronomia Eliseu Maciel
Programa De Pós-Graduação em Manejo e Conservação do Solo e da Água

Tese

Modelagem da variabilidade de atributos do solo em diferentes escalas espaciais usando geoestatística baseada em modelos lineares mistos e Wavelets bidimensionais

Luana Nunes Centeno

Pelotas, 2024.

Luana Nunes Centeno

Modelagem da variabilidade de atributos do solo em diferentes escalas espaciais usando geoestatística baseada em modelos lineares mistos e Wavelets bidimensionais

Tese apresentada ao Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutora em Ciências.

Orientador: Prof. Dr. Luís Carlos Timm

Co orientadores: Profª. Drª. Alessandra dos Santos

Prof. Dr. Dioni Glei Bonini Bitencourt

Pelotas, 2024.

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação da Publicação

C397m Centeno, Luana Nunes

Modelagem da variabilidade de atributos do solo em diferentes escalas espaciais usando geoestatística baseada em modelos lineares mistos e Wavelets bidimensionais [recurso eletrônico] / Luana Nunes Centeno; Luís Carlos Timm, orientador; Alessandra dos Santos, Dioni Glei Bonini Bitencourt, coorientadores. — Pelotas, 2024. 151 f.: il.

Tese (Doutorado) — Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, 2024.

1. Dependência espacial. 2. Simulação hidrológica. 3. Espaçamento de amostragem. 4. Práticas conservacionistas. 5. Uso intensivo do solo. I. Timm, Luís Carlos, orient. II. Santos, Alessandra dos, coorient. III. Bitencourt, Dioni Glei Bonini, coorient. IV. Título.

CDD 631.6

Elaborada por Ubirajara Buddin Cruz CRB: 10/901

Luana Nunes Centeno

Modelagem da variabilidade de atributos do solo em diferentes escalas espaciais usando geoestatística baseada em modelos lineares mistos e Wavelets bidimensionais

Tese apresentada ao Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutora em Ciências.

Data da Defesa: 04/03/2024

Banca examinadora:

Prof. Dr. Luís Carlos Timm (Orientador)
Doutor em Agronomia pela Escola Superior de Agricultura Luiz de Queiroz/USP

Prof.ª Dr.ª Tirzah Moreira Siqueira
Doutora em Recursos Hídricos e Saneamento Ambiental pela Universidade Federal do Rio Grande do Sul

Prof.ª Dr.ª Maria Cândida Moitinho Nunes
Doutora em Ciência do Solo pela Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul

Prof. Dr. Willian Silva Barros
Doutor em Genética e Melhoramento pela Universidade Federal de Viçosa

Agradecimentos

Agradeço a Deus pela dádiva da vida e por cada detalhe que Ele coloca em meu caminho diariamente, sejam pessoas, objetos ou situações, através das quais posso crescer e compreender mais sobre eu mesma, buscando assim constantemente evoluir. Ao meu mentor, expresso minha mais profunda gratidão. Sem a sua orientação, nada disso seria possível. Tua presença é minha âncora nos momentos de desespero. Agradeço a Deus por tê-lo colocado em minha vida.

A Samanta, minha fonte de apoio inesgotável. Agradeço por acreditar em mim, mesmo quando eu mesma duvido, por sempre me incentivar a seguir em frente, a crescer. Você é um dos meus exemplos, e ter você ao meu lado nessa jornada é um privilégio. Agradeço também aos meus queridos amores - Kiara, Shopya, Branco, Tigrinho, Pequeno, Lara e Horús - por me ensinarem diariamente sobre amor, gratidão, dedicação, e por mostrarem que palavras nem sempre são necessárias para expressar o que sentimos.

Aos meus pais, Adriana e Jair, minha eterna gratidão. Obrigada por priorizarem minha educação, por seu apoio incondicional e pela paciência nos momentos em que não pude estar presente. Vocês me ensinaram que títulos não se comparam à vastidão do conhecimento, e por isso sou imensamente grata. Não posso deixar de mencionar minha irmãzinha Alana e meu afilhado sobrinho Pedro Henrique, que trazem luz e esperança à minha vida, mostrando-me as preciosidades que devemos valorizar neste mundo. À Julia, por trazer momentos de leveza e alegria em meio ao rigor deste doutorado. À Luciane, Sâmara, Duda e Jorge, meu sincero agradecimento por cada palavra de carinho, por ouvirem minhas aflições e por todo o apoio e acolhimento.

Expresso minha gratidão eterna ao meu orientador, o Prof. Dr. Luís Carlos Timm. Agradeço profundamente por ter acreditado em mim ao longo dos últimos dez anos e por me encorajar a superar meus desafios. Sua orientação sábia e ética, juntamente com sua amizade sincera, foram fundamentais para mim. Agradeço por estar ao meu lado em cada fase deste trabalho, não apenas me orientando, mas também colaborando e construindo este estudo comigo. Aos meus coorientadores, Profª. Drª. Alessandra dos Santos e Prof. Dr. Dioni Glei Bonini Bitencourt, minha profunda gratidão por sua ajuda na parte estatística e matemática, por abraçarem o tema comigo e por me ajudarem a superar obstáculos. Sou imensamente grata a este comitê de orientação por todo apoio, carinho e atenção dedicados a este trabalho.

A todos os colegas que compartilharam comigo suas experiências e conhecimentos, minha sincera gratidão pela troca enriquecedora. Ao Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água, e a todos os professores envolvidos, meu profundo agradecimento por contribuírem diretamente ou indiretamente para o enriquecimento deste trabalho.

Agradeço aos membros da banca examinadora - Professores Tirzah Moreira Siqueira, Willian Silva Barros, Danielle de Almeida Bressiani e Maria Cândida Moitinho Nunes - por aceitarem com entusiasmo o convite para avaliar este trabalho. Antecipo minha gratidão pelas valiosas contribuições que sei que oferecerão.

Por fim, expresso minha gratidão a todos os funcionários da UFPel, dos terceirizados à reitora, por sua dedicação incansável, pois sem vocês, a Universidade Federal de Pelotas não seria o que é. A todos os meus parentes e amigos, meu mais sincero agradecimento por torcerem por mim durante este período de doutorado. Sou eternamente grata a cada um de vocês.

Epígrafe

“O ideal da educação não é aprender ao máximo,
maximizar os resultados, mas é antes de tudo
aprender a aprender, é aprender a se desenvolver e
aprender a continuar se desenvolvendo...”

*Jean Piaget

SUMÁRIO

Resumo

CENTENO, Luana Nunes. **Modelagem da variabilidade de atributos do solo em diferentes escalas espaciais usando geoestatística baseada em modelos lineares mistos e Wavelets bidimensionais.** Tese (Doutorado em Ciências). Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água, Universidade Federal de Pelotas, Pelotas 2024.

Analisar a variabilidade espacial, tanto em parcelas agrícolas quanto em bacia hidrográfica, demanda de ferramentas estatísticas robustas que considerem a posição e a possível dependência entre as observações de variáveis no campo, uma vez que os atributos físicos, químicos e biológicos do solo variam amplamente no espaço. Sendo assim a identificação e quantificação da variabilidade espacial e temporal constitui-se em uma ferramenta valiosa, haja vista que permite definir o espaçamento adequado de amostragem, reduzindo os custos do monitoramento de dados de campo, sobretudo em estudos associados à simulação hidrológica. Frente ao exposto, este estudo tem como objetivo principal analisar a variabilidade espacial dos atributos do solo em escala de amostragem de áreas agrícolas e de bacia hidrográfica, por meio de modelos lineares mistos e de transformada bidimensional de Wavelet. Foi utilizada uma área agrícola localizada no sul do Rio Grande do Sul, uma vez que esta área é representativa do uso e manejo de solos adotados no sul do estado, ou seja, centrado no uso intensivo do solo, não associado ao uso de práticas conservacionistas. A saber: a área em questão é descrita como Granja Bretanhas (Estancinha). Destaca-se que as áreas agrícolas foram descritas baseadas nos estudos de Bitencourt et al. (2016) e Timm et al. (2020). Também será utilizada como área de estudo a bacia hidrográfica Sanga Ellert (BHSE) que possui uma área de 70 ha e tem uma malha amostral de 178 amostras. Nestes dados das áreas agrícolas serão aplicados os modelos lineares mistos, para analisar o efeito das classes ou do uso do solo no domínio espacial, sob diferentes estruturas de matriz de variâncias e covariâncias e incorporando o efeito anisotrópico. Já na BHSE buscar-se-á caracterizar e quantificar as correlações multi-espaciais, dentro de escalas específicas, entre os atributos físicos, químicos e hídricos do solo, na bacia hidrográfica Sanga Ellert, usando as transformadas bidimensionais de Wavelets. Sendo assim espera-se por meio deste estudo compreender a variabilidade espacial dos atributos do solo, através das ferramentas estatísticas supracitadas e definir as escalas de amostragem de acordo com as áreas estudadas.

Palavras-chave: Dependência espacial; Simulação hidrológica; Espaçamento de amostragem; Uso intensivo do solo e Práticas conservacionistas.

Abstract

CENTENO, Luana Nunes**. Modeling the variability of soil attributes at different spatial scales using geostatistics based on linear mixed models and two-dimensional Wavelets.** Thesis (Doctorate in Science). Postgraduate Program in Soil and Water Management and Conservation, Federal University of Pelotas, Pelotas 2024.

Analyzing spatial variability, both in agricultural plots and in Watershed, requires robust statistical tools that consider the position and possible dependence between observations of variables in the field since the physical, chemical and biological attributes of the soil diversify widely in space. Therefore, the identification and quantification of spatial and temporal variability constitutes a valuable tool, as it allows defining the appropriate sampling spacing, reducing the costs of monitoring field data, especially in studies associated with hydrological simulation. In view of the above, this study's main objective is to analyze spatial variability on a sampling scale of agricultural areas and river basins, using linear mixed models and two-dimensional Wavelet transform. For this, two agricultural areas located in the south of Rio Grande do Sul will be used, since these areas are representative of the use and management of soils adopted in the south of the state, that is, centered on the intensive use of soil, not associated with the use of conservation practices. Namely: the areas in question are described as Granja Bretania (Estancinha) and Systematized area of the Terras Baixas experimental station. It is noteworthy that the agricultural areas were described based on studies by Bitencourt et al. (2016) and Timm et al. (2020). The Sanga Ellert Watershed (SEW) will also be used as a study area, which has an area of 70 ha and has a sampling grid of 178 samples. Linear mixed models will be applied to these data from agricultural areas to analyze the effect of classes or land use in the spatial domain, under different variance and covariance matrix structures and incorporating the anisotropic effect. At SEW, we will seek to characterize and quantify multi-spatial correlations, within specific scales, between the physical, chemical and water attributes of the soil, in the Sanga Ellert watershed, using two-dimensional Wavelet transforms. Therefore, through this study, it is expected to understand the spatial variability of soil attributes, through the aforementioned statistical tools and to define the sampling scales according to the areas studied.

Keywords**:** Spatial dependence; Hydrological simulation; Sampling spacing; Intensive land use and conservation practices.

1. INTRODUÇÃO

O manejo do solo em áreas agrícolas no Sul do Brasil é, em geral, baseado em recomendações convencionais, não levando em consideração a variabilidade do ambiente físico onde as plantas se desenvolvem (González-Ávila et al., 2023; Parfitt et al., 2014; Suzuki et al., 2023; Nalin et al., 2023). Isso pode resultar em áreas com altos ou baixos níveis de nutrientes, impactando em uma maior heterogeneidade na produção e menor rentabilidade aos agricultores (Bitencourt et al., 2016; Babu et al., 2023; Lv et al., 2023). Sendo assim, informações sobre a variabilidade dos atributos do solo são pré-requisitos para a avaliação de terras agrícolas e também para apoiar as decisões apropriadas (Rosemary et al., 2017; Azizi; Wang; et al., 2023).

Já quando relacionado à escala de bacias hidrográficas, tem-se a necessidade de se compreender o funcionamento do balanço hídrico, dos processos que controlam o movimento, a quantidade e a qualidade da água (Ayoubi; Demattê, 2023; Centeno et al., 2020; Soares et al., 2023). Neste intuito, a identificação e quantificação da estrutura de variabilidade espacial e temporal dos atributos do solo tornam-se importantes para um melhor entendimento da dinâmica da água no sistema solo-planta-atmosfera.

A intensidade dessa variabilidade espacial em uma área depende dos fatores de formação do solo e, como estes atuam sobre os atributos e o manejo do solo ao longo do tempo (Tang et al., 2023; Tian et al., 2023; Vieira; Dechen, 2010). Ademais, os atributos do solo variam em diferentes escalas espaciais (Reichardt; Timm, 2022; Soares et al., 2023), principalmente devido à heterogeneidade de fatores intrínsecos e extrínsecos, como os tipos de solo, manejo e o uso do solo (Centeno et al., 2020).

Contudo, as mudanças nos padrões de distribuição espacial raramente ocorrem isotropicamente (Trangmar; Yost; Uehara, 1986), ou seja, elas se alteram conforme a direção adotada. Neste intuito, as ferramentas geoestatísticas podem quantificar as estruturas de dependência espacial de uma ou mais variáveis (Reichardt; Timm, 2022; Lu; Cavieres; Moraga, 2023; Sahu et al., 2023).

Ao propormos um modelo, com base no semivariograma, é pertinente considerar a existência de fatores fixos correlacionados diretamente com os dados de interesse. Dentre esses fatores podemos destacar o uso do solo, as unidades de mapeamento em escalas espaciais adequadas, a cobertura do solo, dentre outros, a natureza do atributo de solo analisado. Geralmente os atributos apresentam a estrutura de covariância direcional e dependência de localização, ou seja, apresenta anisotropia (Koch; Lele; Lewis, 2020).

Os modelos lineares mistos (MLMs) podem ser uma alternativa na modelagem desses fatores dada a possibilidade de incorporarmos efeitos fixos e aleatórios, bem como a dependência espacial simultaneamente, por meio da análise de variância e de covariância residual dos dados (Deiss et al., 2019; Heiskanen et al., 2018). Já a anisotropia deve ser considerada quando o semivariograma não apresenta estacionariedade de segunda ordem, ou seja, apresenta mudança na direção (Koch; Lele; Lewis, 2020). Por esse motivo, ao se considerar a influência dos fenômenos anisotrópicos na variabilidade espacial dos atributos do solo, os mapas krigados tenderão a apresentar maior acurácia.

Ademais, quando analisando diretamente os impactos nos atributos do solo, destaca-se como mais afetadas as frações texturais e a porosidade do solo, bem como as propriedades hidrológicas que controlam a infiltração de água no solo (Timm et al., 2020). A matéria orgânica é reduzida, podendo representar diminuição na produtividade e aumento da densidade, causando assim aumento na impermeabilização dos solos (Burst et al., 2020).

Em suma, torna-se essencial quantificar as mudanças espaciais na camada superficial do solo e seus impactos nas características de retenção de água e nas propriedades físico químicas, a fim de melhorar estratégias de gestão agrícolas (Bitencourt et al., 2016; Centeno et al., 2020).

Contudo, as mudanças nos padrões de distribuição espacial raramente ocorrem isotropicamente (Trangmar; Yost; Uehara, 1986), ou seja, elas se alteram conforme a direção adotada. Neste intuito, as ferramentas geoestatísticas podem quantificar as estruturas de dependência espacial de uma ou mais variáveis e em diferentes direções (Reichardt; Timm, 2022; Lu; Cavieres; Moraga, 2023; Sahu et al., 2023).

Ao propormos um modelo, com base no semivariograma, é pertinente considerar a existência de fatores fixos correlacionados diretamente com os dados de interesse. Dentre esses fatores podemos destacar o uso do solo, as unidades de mapeamento em escalas espaciais adequadas, a cobertura do solo, dentre outros, a natureza do atributo de solo analisado. Geralmente os atributos apresentam a estrutura de covariância direcional e dependência de localização, ou seja, apresenta anisotropia (Koch; Lele; Lewis, 2020).

Os modelos lineares mistos (MLMs) podem ser uma alternativa na modelagem desses fatores dada a possibilidade de incorporarmos efeitos fixos e aleatórios, bem como a dependência espacial simultaneamente, por meio da análise de variância e de covariância residual dos dados (Deiss et al., 2019; Heiskanen et al., 2018). Já a anisotropia deve ser considerada quando o semivariograma não apresenta estacionariedade de

segunda ordem, ou seja, apresenta mudança na direção (Koch; Lele; Lewis, 2020). Por esse motivo, ao se considerar a influência dos fenômenos anisotrópicos na variabilidade espacial dos atributos do solo, os mapas krigados tenderão a apresentar maior acurácia.

Ademais, são inexistentes os estudos publicados na literatura sobre avaliação do efeito da anisotropia na krigagem por meio de modelos lineares mistos. Por outro lado, modelos isotrópicos têm sido utilizados para espacializar atributos do solo relacionados à fertilidade, fornecendo ferramentas importantes para os agricultores quanto às áreas que necessitam de reposição de nutrientes (Behera et al., 2018; Burst et al., 2020; Deiss et al., 2019; Peralta et al., 2015). Por fim, destacam-se as transformadas bidimensionais de Wavelet, uma vez que possibilitam analisar a variabilidade espacial e suas relações dentro de escalas de amostragens definidas, através de uma decomposição de Wavelet de matrizes de dados (Moretin, 2014).

Frente ao exposto, o objetivo principal analisar e compreender a variabilidade de atributos do solo em diversas escalas espaciais, utilizando técnicas avançadas de geoestatística, modelos lineares mistos e Wavelets bidimensionais.

2. OBJETIVOS E HIPÓTESES

2.1 Objetivo geral

O objetivo geral deste estudo foi analisar e compreender a variabilidade de atributos do solo em diversas escalas espaciais, utilizando técnicas avançadas de geoestatística, modelos lineares mistos e Wavelets bidimensionais.

2.2 Objetivos específicos

- Avaliar o desempenho do modelo Matérn na análise da variabilidade espacial, levando em consideração fatores como tamanho da área, densidade amostral, distribuição espacial da amostra e métodos de coleta e análise de dados.

- Caracterizar e quantificar as correlações multiespaciais, dentro de escalas específicas, entre os atributos físicos, químicos e hídricos do solo, na bacia hidrográfica Sanga Ellert, usando as transformadas bidimensionais de Wavelets.

2.3. Hipóteses

- Identificar dentre os parâmetros simulados, por meio de modelos lineares mistos, quais parâmetros mais afetam fatores como tamanho da área, densidade amostral, distribuição espacial da amostra e métodos de coleta e análise de dados.

- As wavelets bidimensionais em escalas maiores, representam mais próximo da realidade a variabilidade espacial da Ksat.

3. REVISÃO DE LITERATURA

3.1 Variabilidade espacial em áreas agrícolas

Os atributos físicos, químicos e hídricos do solo e os fatores de formação do solo variam devido às suas características intrínsecas e posição na paisagem. A intensidade dessa variação dependerá das condições ambientais e como estas atuam sobre os atributos ao longo do tempo (Vieira; Dechen, 2010; Santos, 2022).

A variabilidade espacial dos atributos do solo tem sido alvo de estudos desde o início do século XX. Inicialmente, avaliou-se utilizando apenas ferramentas da estatística clássica, tais como média, variância e coeficiente de variação, exigindo-se que a variável sob investigação tivesse distribuição normal e fosse espacialmente independente (Reichardt; Timm, 2020). Nessas análises, a posição das observações no espaço não era considerada, ignorando a premissa de que unidades de observação próximas tendem a ter resultados mais parecidos do que unidades distantes.

Ferramentas estatísticas da Análise de Séries Temporais/Espaciais (autocorrelogramas, crosscorrelogramas, modelos de espaço de estados etc.) e da Geoestatística (semivariogramas, semivariogramas cruzados, krigagem etc.), que consideram a posição no espaço (ou no tempo) e a possível dependência entre as observações da variável em estudo, têm sido aplicadas no intuito de estudar a variabilidade das diferentes variáveis do sistema Solo-Planta-Atmosfera (Reichardt; Timm, 2020). Nesse intuito, Nielsen e Wendroth (2003) sugerem uma amostragem ao longo de uma transecção e/ou de uma malha para avaliar a estrutura de variabilidade espacial ou temporal e as relações entre os atributos do solo.

A identificação e quantificação da variabilidade espacial e temporal constitui-se em uma valiosa ferramenta, haja vista que permite definir o intervalo adequado de amostragem, reduzindo os custos do monitoramento de dados de campo, sobretudo em estudos associados, dentre outros, à simulação hidrológica (Hupet; Vanclooster, 2004; Schneider *et al.*, 2008).

Em resumo, os solos são altamente variáveis devido à interação complexa de processos físicos, químicos, biológicos e hídricos que operam em diversas intensidades e escalas (Goovaerts, 1997). Assim, o emprego de ferramentas estatísticas espaciais capazes de considerar essa variabilidade no campo mostra-se promissor para o manejo agrícola.

Dentre as ferramentas mencionadas anteriormente, é relevante destacar que as transformadas de Wavelets têm sido utilizadas para identificar escalas específicas dos atributos do solo. Elas são empregadas na análise da variabilidade espacial dos atributos do solo, proporcionando uma abordagem para estudar séries espaciais no domínio do tempo (ou espaço) e da frequência (Centeno et al., 2020).

3.2 Modelagem da variabilidade espacial em diferentes escalas

Os estudos com definições espaciais de análise geográfica são fundamentais para delimitação de bacias hidrográficas, além da avaliação dos níveis de degradação ambiental, pois buscam melhorar e realizar diagnósticos em relação ao planejamento e gestão (Souza; Silva, 2022). Nesse contexto, as modelagens ambientais costumam utilizar métodos estatísticos que combinem variáveis para geração de indicadores naturais (Costa, 2021). Dessa forma, a modelagem é um instrumento que permite identificar lacunas e incógnitas, bem como analisar e investigar espacialmente as modificações no ambiente como qualidade, levantamento de uso e cobertura do solo e seus atributos, permitindo caracterizar fisicamente o local de estudo (Soligo, 2021; Souza; Silva, 2022).

Contudo, a seleção de quantidade de atributos estudados e a relação entre escala de trabalho e quantidade de pontos de amostragem é sempre fonte de dúvida entres os pesquisadores (Borges, 2019). Por isso, a escolha da escala de amostragem dos dados é imprescindível para análises espaciais, uma vez que busca representar adequadamente o comportamento individual e prezar pela qualidade dos modelos espaciais (Rocha, 2019).

A representação da variabilidade do meio físico é um desafio no processo de modelagem, sendo fundamental saber como são obtidos e tratados os atributos, iniciando pelo processo de amostragem. Para essa etapa, a análise estatística de dados multivariados e a análise espacial de dados servem como ferramenta útil quando se deseja comparar a variabilidade das amostras e dos atributos (Borges, 2019).

Dessa forma, tem-se diversas pesquisas sobre a influência das escalas no estudo de solo. Por exemplo, Santos (2019) em seu estudo objetivou modelar atributos químicos e espectrais presentes nos solos de Feira de Santana, na Bahia, bem como modelar as características químicas naturais dos macronutrientes dos solos dessa mesma região. Foram realizadas coletas amostrais em diversas partes da área de estudo e foi concluído que para uma melhor resolução espacial e maior representatividade da variabilidade dos elementos maiores nos solos seria necessário um maior número de amostras. Além disso,

fatores como adubações e calagens podem ser os principais responsáveis pela variabilidade espacial dos macronutrientes dos solos.

Pinheiro (2018), embora com um objetivo distinto, buscou mapear qualitativamente os parâmetros para classificação dos solos superficiais em Palmas–TO, com objetivo final de elaborar uma carta de descrição dos solos, uma vez que o município carece de estudos desenvolvidos sobre a caracterização e classificação dos solos. E o autor concluiu que a análise da escala espacial foi um dos fatores de maior influência em seu estudo.

Corrêa (2018) corrobora com a afirmação acima ao relatar que um dos fatores limitantes para a agricultura de precisão é a noção do número de amostras necessárias para representar espacialmente a distribuição do nutriente ou insumo aplicado de forma variada. Por isso, neste estudo foi avaliado, com base na análise geoestatística, o número ideal de subamostras para melhor determinar as zonas específicas de manejo e para ratificar a existência de um atributo (solo e/ou planta) que possa, espacialmente, funcionar como indicador de qualidade, isto é, caracterizar as zonas específicas de manejo, com o intuito de aumentar a produtividade do *Eucalyptus camaldulensis*.

Outrossim, Barbosa (2020) conduziu um estudo em Tabapuã, município localizado no noroeste do estado São Paulo, que possui um cultivo histórico de cana-de-açúcar, em que foi realizada a caracterização da variabilidade espacial nas diferentes direções foi por meio de uma grade de 50x50 m com intervalos regulares de 0,50 m, sendo 270 na pedoforma côncava e 353 na pedoforma convexa, totalizando 623 pontos amostrais numa área de 200 ha. O estudo da anisotropia em diferentes formas de relevo demonstrou ser eficiente na identificação e mapeamento, podendo servir de base para a gestão e execução de programas que visem a conservação do solo e da água tendo como base a escala amostral.

Ferreira (2020) teve como área de estudo a Fazenda São Miguel, localizada no município de Barra do Choça, região cafeeira do planalto de Vitória da Conquista, estado da Bahia. A lavoura possuía aproximadamente 12,6 hectares no espaçamento de 3,8 m entre fileiras e 0,7 m entre plantas, totalizando 3.760 plantas ha^{-1}, sob regime de sequeiro e cultivado a pleno sol e solo classificado como Latossolo Amarelo distrófico. Os autores, ao empregar técnicas geoestatísticas, conseguiram inferir que a densidade amostral influenciou diretamente na identificação da estrutura e magnitude da distribuição espacial dos atributos químicos do solo e nutricionais, vegetativos e produtivos do cafeeiro arábica avaliados.

Segundo Barbosa (2020), a variabilidade espacial de um atributo está associada dentre outros fatores, dentro da modelagem matemática, à distância espacial da amostragem, que pode variar com a direção. Assim, existe uma relação de que, quanto maior o número de amostras de uma área, mais próximas da realidade estarão as estimativas das propriedades estudadas, o que também se aplica à grade de amostragem que auxilia no estabelecimento de ferramentas de agricultura de precisão (Ferreira, 2020).

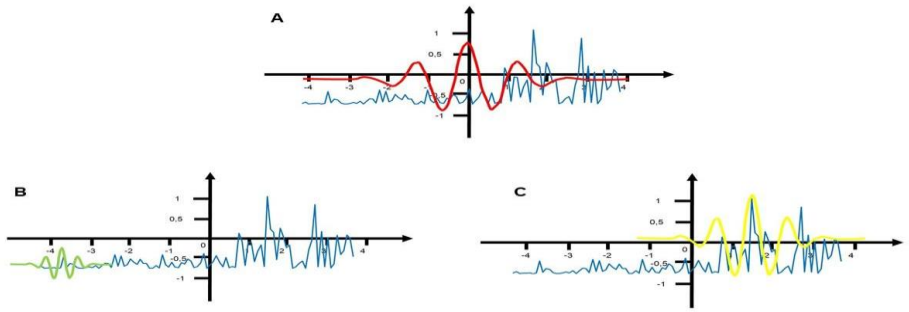
Com base no exposto acima, tendo como referência os objetivos deste estudo, a seguir foi realizada uma revisão bibliográfica sobre as principais ferramentas empregadas neste estudo para estudar as escalas de amostragem.

3.3 Transformada de Wavelet

A Wavelet é uma função capaz de decompor e, assim, descrever ou representar outra função (ou uma série de dados) originalmente descrita no domínio do tempo (ou espaço), de forma a analisar essa função em diferentes escalas de frequência e tempo (Centeno *et al.*, 2020). A coerência Wavelet, ou espectro de Wavelet pode ser utilizada para compreender relações específicas de escala e localizadas para processos transitórios (Biswas, 2018). Quando se decompõe uma função por meio das Wavelets, tem-se a transformada de Wavelet (sigla em inglês WT), bem como suas variantes contínuas e discretas.

Matematicamente falando, a transformada Wavelet é uma convolução, ou seja, uma multiplicação da função chamada Wavelet com o sinal a ser estudado. Essas decomposições são propriedades de translação e escalonamento (figura 1), contidas em um único protótipo de uma função Wavelet geradora $\psi_{j,k}(t)$, denominada do inglês "*Wavelet mother"*, de onde se obtêm as denominadas funções "Wavelets mães", pela variação de k e j ($j > 0$)) (Biswas; Si, 2011).

Figura 1 – Representação de um sinal da Wavelet Morlet contendo uma série de dados. Legenda: (A) série de dados e a representação gráfica do formato da Wavelet; (B) Wavelet se contraindo para capturar pequenas variações e (C) a Wavelet se dilatando procurando capturar picos existentes na série de dados. Fonte: Elaborado por Centeno (2023).



Portanto, de acordo com Biswas (2018), pode-se compreender as Wavelets como um conjunto de funções com forma de pequenas ondas geradas por translações $\psi(t) \rightarrow \psi(t+k)$ e escalonamento $\psi(t) \rightarrow \psi(t/j)$ de uma função Wavelet simples (Wavelet *mother)* que resulta na Equação 1.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{j}} \Psi_0 \left(\frac{t-k}{j}\right) \quad (1)$$

em que "j" é denominado fator de escalonamento (contração $j > 1$ e dilatação $j < 1$) (figura 1) sendo "k" o fator de translação. A função $\psi_{j,k}(t)$ é chamada de Wavelet filha, gerada a partir da Wavelet mãe Ψ_0. O fator $\frac{1}{\sqrt{j}}$ é chamado de constante de normalização da energia de cada Wavelet filha de forma a manter a energia da Wavelet mãe, ou seja, a soma das energias de todas as Wavelets filhas, oriundas da Equação (1), resulta na energia total da transformada. Contudo, segundo James (2011), para que uma função possa ser candidata a uma Wavelet mãe (representada por Ψ_0), ela deve satisfazer a duas propriedades distintas, a saber:

I) A integral dessa função deve ser zero (Equação 2).

$$\int_{-\infty}^{+\infty} \Psi_0\,(t)dt = 0 \quad (2)$$

Essa condição é conhecida como condição de "admissibilidade" a qual garante a transformada Inversa da função.

II) A função Wavelet mãe deve possuir energia finita (Equação 3).

$$\int_{-\infty}^{+\infty} |\Psi_0(t)|^2 dt \ < \infty \qquad (3)$$

Esta propriedade é equivalente a dizer que $\Psi_0(t)$ e é quadraticamente integrável. As funções bases senos e cossenos satisfazem apenas a propriedade (I).

A transformada Wavelet foi descrita como uma forma de analisar os sinais aperiódicos, ruidosos, intermitentes, transitórios, capaz de examinar o sinal simultaneamente em relação ao tempo e frequência ou a escala e posição, independentemente do tamanho e da intensidade da frequência (Hu; Si, 2016). Conclui-se assim que a análise Wavelet é realizada por meio da aplicação sucessiva da transformada Wavelet a diversos valores de k e j, o que representa a decomposição do sinal original da série temporal em diversos componentes localizados no tempo (parâmetro de translação k) e frequência (parâmetro de escala j). De certa forma, cada tipo de Wavelet mãe possui uma melhor ou pior localização nos domínios da frequência e/ou do tempo; dessa maneira, torna-se necessária a escolha de uma Wavelet mãe conveniente, de acordo com o sinal estudado e os objetivos almejados (Reichardt; Timm, 2020). Tendo ciência de que esta pesquisa abarca o interesse em mudanças de amplitude e fase, a teoria de Wavelets de Morlet foi a escolhida para ser aplicada nas análises realizadas neste trabalho. Sua definição é dada pela Equação 4.

$$\Psi_0(t) = \pi^{-\frac{1}{4}} \left(e^{i\omega_0 t} - e^{\frac{(\omega_0)^2}{2}} \right) e^{-\frac{t^2}{2}}, \qquad (4)$$

Em que: ω_0 é a frequência central (ou fundamental) da Wavelet mãe. O fator $\pi^{-\frac{1}{4}}$ é a constante de normalização. O termo $e^{-\frac{t^2}{2}}$ é o pacote gaussiano que confina as Wavelets de Morlet.

O segundo termo entre parênteses é conhecido como termo de correção, pois corrige a média quando esta não resulta em zero do senoidal complexo do primeiro termo. Na prática, esse valor torna-se insignificante para valores de f_0 não muito grandes e pode ser ignorado, gerando a Equação 5.

$$\Psi_0(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-\frac{t^2}{2}} \qquad (5)$$

3.3.1 Transformada bidimensional de Wavelet (CWT2D)

O texto a seguir está baseado em: Morettin (2014) e Jean-Pierre e Romain (2008).

A transformada de Wavelet em uma dimensão equivale a projetar o sinal na Wavelet que foi obtida pela translação e dilatação da Wavelet mãe. Assim, a transformada é totalmente determinada por essas operações elementares da linha. Contudo, quando se pensa em Wavelets bidimensionais consideram-se primeiras as operações elementares que queremos aplicar aos sinais, pois dessa forma é possível estender a transformada de Wavelet para situações gerais, como Wavelets em dimensões maiores, Wavelets dependentes da escala, Wavelets dependentes do tempo, dentre outras.

Dessa forma, é possível realizar uma decomposição de Wavelet de matrizes de dados, seja para compactar os dados ou para realizar uma caracterização paramétrica baseada em Wavelets dos dados. Para realizar uma decomposição de Wavelet de uma matriz 2D, é necessário usar transformadas de Wavelet 2D. Estas podem ser geradas por meio de produtos tensoriais de suas contrapartes ortonormais 1D. O arranjo mais comum e o mais simples empregado é usar a mesma escala nas direções horizontal e vertical, que ocasionam formas trans quadradas. Outras maneiras possíveis, por exemplo, são transformadas retangulares, nas quais as escalas horizontal e vertical variam independentemente e transformadas que não são simplesmente produtos tensoriais.

Sendo assim dado um sinal bidimensional (2D), por exemplo, uma imagem f(x,y) a sua CWT2D com relação a Wavelet ψ é:

Nesse sentido, pela simplicidade do método quando comparada a outras escalas bidimensionais, a de *Morlet* e as funções Wavelet são $função\ de\ escala\ 2D$, onde $\phi(t_1, t_2) = \phi(t_1)\phi(t_2)$ e a $Wavelet\ horizonta\ 2D$ é $\Psi^h(t_1, t_2) = \phi(t_1)\Psi(t_2)$. Por conseguinte, a $Wavelet\ vertical\ 2D$ é $\Psi^v(t_1, t_2) = \phi(t_1)\Psi(t_2)$ e a $Wavelet\ diagonal\ 2D$ é $\Psi^d(t_1, t_2) = \psi(t_1)\Psi(t_2)$. Sendo t a nossa variável independente, seja ela espacial ou temporal, nesse exemplo representando as coordenadas espaciais.

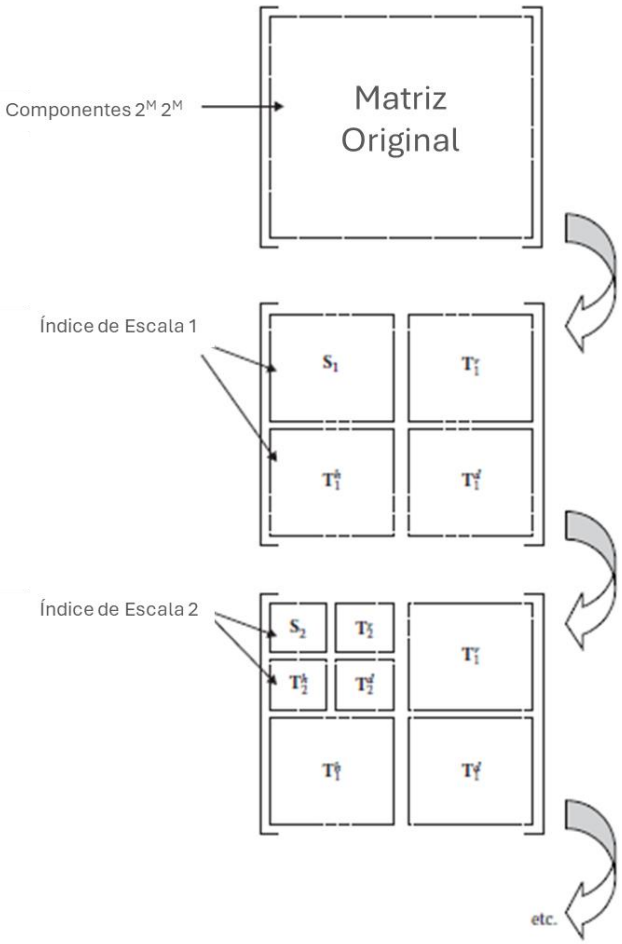
Destaca-se que a decomposição multirresolução das matrizes de coeficientes 2D pode ser expressa como $S_{m+1(n_1,n_2)} = \frac{1}{2}\sum_{k_1}\sum_{k_2} c_{k_1}\, c_{k_2} S_{m(2n_1 + k_1, 2n_2 + k_2)}$.

Posteriormente, $T^h{}_{m+1(n_1,n_2)} = \frac{1}{2}\sum_{k_1}\sum_{k_2} b_{k_1}\, c_{k_2} S_{m(2n_1 + k_1, 2n_2 + k_2)}$. E o $T^v{}_{m+1(n_1,n_2)} = \frac{1}{2}\sum_{k_1}\sum_{k_2} c_{k_1}\, b_{k_2} S_{m(2n_1 + k_1, 2n_2 + k_2)}$.e por fim, $T^d{}_{m+1(n_1,n_2)} =$

$\frac{1}{2}\sum_{k_1}\sum_{k_2} b_{k_1}\, b_{k_2} S_{m(2n_1+k_1,2n_2+k_2)}$. onde k_1, k_2 são os índices de coeficiente de escala e n_1, n_2 são os índices de localização na escala $m+1$. Dessa forma, pode-se utilizar as versões das Wavelets 2D no índice de escala 1 para realizar a análise multirresolução do arranjo. Dessa forma, a escala 2D de *Morlet* e as funções Wavelet na forma de matriz no índice de escala $m = 1$.

Como eles estão no índice de escala 1, são simplesmente produtos tensoriais dos coeficientes da escala e do filtro da Wavelet $\frac{c_k}{\sqrt{2}}$ e $\frac{b_k}{\sqrt{2}}$. Nas matrizes demonstradas é possível observar que o fator $\frac{1}{2}$ antes de cada matriz é simplesmente o quadrado do fator $\frac{1}{\sqrt{2}}$, que precede as funções unidimensionais correspondentes. Para escalonamento e funções Wavelet em escalas maiores, esse fator de normalização torna-se $\frac{1}{2^m}$. Essas quatro matrizes $2X2$ são necessárias para a decomposição da Wavelet de *Morlet* da matriz 2D, uma vez que a função geral da decomposição de Wavelet 2D é descrita de acordo com a figura 2 a seguir.

Figura 2 – Diagrama esquemático da manipulação de matriz necessária para executar a decomposição de Wavelet em uma grade bidimensional.



Nessa matriz a entrada original é representada por X_0 definida no índice de escala $m = 0$. Assim como no caso unidimensional, seus componentes são inseridos como os coeficientes de aproximação no índice de escala 0, ou seja, a matriz S_0. Após a primeira decomposição Wavelet, uma matriz de decomposição é formada no índice de escala 1 que é dividida em quatro submatrizes distintas, a saber: os componentes verticais detalhados T_1^v, os componentes horizontais detalhados T_1^h, os componentes diagonais detalhados T_1^d e a componentes de aproximação restantes S_1, assim como ocorre com a decomposição de Wavelets de sinais 1D, em que os coeficientes detalhados são inalterados, sendo que na próxima interação se decompõe ainda mais os componentes de aproximação na submatriz S_1.

Isso resulta em coeficientes detalhados contidos nas submatrizes T_2^v, T_2^h e T_2^d no índice de escala 2 e coeficientes de aproximação na submatriz S_2. Esse procedimento pode ser repetido inúmeras vezes para uma matriz de ordem $2^M X 2^M$ para obter um

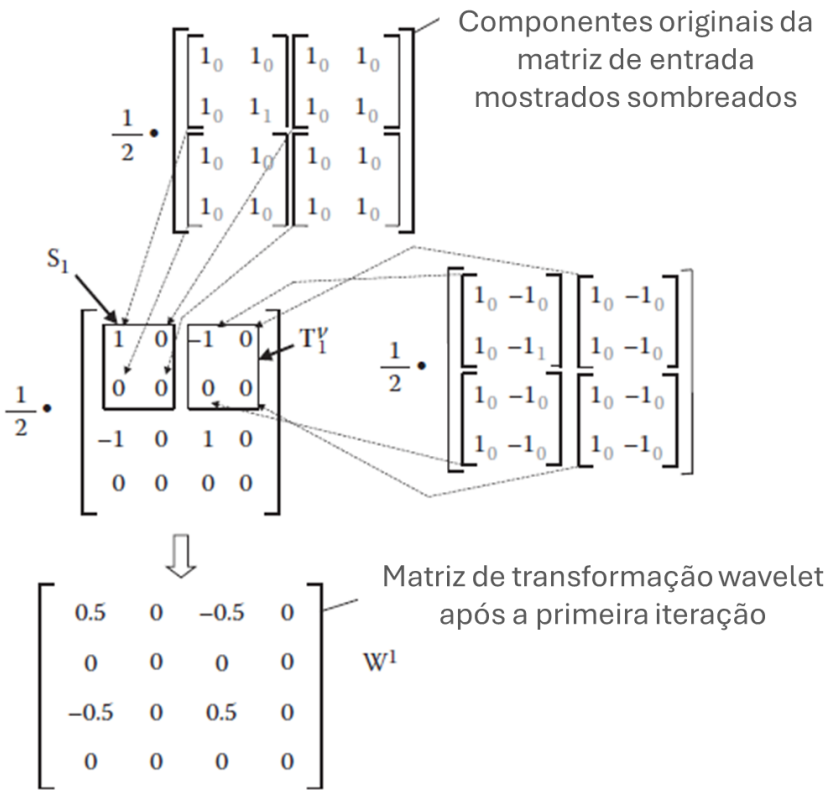
número de submatrizes de coeficiente T_m^v, T_m^h e T_m^d, de tamanho $2^{M-m}X2^{M-m}$, onde $m = 1,2,\ldots,M$.

Se pensarmos em uma matriz com um único componente diferente de zero:

$$x_0 = \begin{vmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

Essa matriz pode representar um único pixel ativado em uma tela ou até mesmo uma parte saliente de uma topossequência. A decomposição multirresolução 2D de *Morlet* dessa matriz é realizada por varredura sobre ela com cada uma das quatro matrizes de Wavelets uma por vez (figura 3).

Figura 3 – Diagrama esquemático da manipulação da matriz necessária para realizar a decomposição da Wavelet de *Morlet* do sinal de pico.



Considerando que na primeira interação (da escala $m = 0\ a\ 1$), a matriz de função de escala $2X2$ são varridas sobre a matriz de entrada, são necessárias quatro matrizes para

cobrir essa matriz $4X4$ demonstrada. Os componentes da matriz original são multiplicados por sua vez pela matriz de funções de escala para fornecer o produto da matriz resultante S_1, que é colocado no quadrante superior direito da primeira interação da matriz.

Da mesma forma, a matriz Wavelet vertical $2X2$ é varrida sobre a matriz e cada valor é colocado no quadrante superior direito da primeira matriz de interação. O procedimento é repetido para as componentes Wavelet horizontal e diagonal, colocando os produtos no canto inferior esquerdo e quadrantes inferiores à direita, respectivamente. O fator de normalização de $\frac{1}{2}$ é deixado de fora da manipulação da matriz até o final para ajudar na clareza do cálculo. Assim, após a primeira interação, a matriz de transformada Wavelet resultante é:

$$W^{(1)} = \begin{vmatrix} 0{,}5 & 0 & -0{,}5 & 0 \\ 0 & 0 & 0 & 0 \\ -0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

A próxima interação usa apenas os componentes de aproximação no quadrante superior direito. Os quatro valores de componentes dessa matriz são então “interrogados” por cada uma das matrizes de escala e Wavelet por sua vez. Isso produz a segunda matriz de interação:

$$W^{(2)} = \begin{vmatrix} 0{,}25 & 0{,}25 & -0{,}5 & 0 \\ 0{,}25 & 0{,}25 & 0 & 0 \\ -0{,}5 & 0 & 0{,}5 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

Essa é a segunda e última interação que se pode realizar na decomposição da matriz $4X4$, ou seja $M = 2$. Essa matriz é composta por $T_1^v, T_1^h, T_1^d, T_2^v, T_2^h, T_2^d$, e, no canto superior esquerdo, a componente de aproximação S_2 que está relacionada com a média da matriz original. Sendo assim, a cada interação a matriz resultante tem a mesma energia da matriz original, em que a energia é a soma dos componentes ao quadrado da matriz, ou seja: $E = \sum_{i=0}^{2^M-1}\sum_{j=0}^{2^M-1}(X_{0,i,j})^2 = \sum_{i=0}^{2^M-1}\sum_{j=0}^{2^M-1}(W_{i,j}^{(m)})^2$. Onde $X_{0,i,j}$ e $W_{i,j}^{(m)}$ são, respectivamente, os elementos das matrizes de entrada e de decomposição de

Wavelets localizados na linha i e na coluna j. Tal energia é igual à 1, tanto para a matriz de decomposição intermediária quanto para a matriz de decomposição completa.

3.4 Geoestatística clássica

A variabilidade espacial de atributos do solo, da planta e da atmosfera tem sido estudada por meio da geoestatística clássica e é função da estimativa de um modelo empírico de semivariograma e em um ajuste de um modelo matemático ao semivariograma experimental, para posterior elaboração dos mapas de variabilidade por meio do interpolador geoestatístico de krigagem. Essa ferramenta é extremamente difundida e para mais detalhes sobre ela pode-se consultar o estudo de Reichardt e Timm (2022).

3.5 Geoestatísca baseada em modelos

A geoestatística baseada em modelos, objeto deste estudo, foi proposta por Diggle *et al.* (1998), com a intenção de apresentar uma abordagem para problemas geoestatísticos com base na aplicação de métodos estatísticos "formais" e não empíricos assumindo explicitamente um modelo estocástico (Diggle; Ribeiro Júnior, 2007).

Nesse sentido, destaca-se que Cressie (1993) foi quem propôs um modelo estatístico que decompunha os dados espacializados em um componente determinístico ou de tendência, denominado de variação de larga escala (componente não estocástico), e um erro aleatório com média zero (ε). A componente determinística $[\mu(x)]$ é comumente função da posição no espaço e provavelmente de outras variáveis explicativas. A componente aleatória $(\varepsilon(x))$ pode incluir um processo aleatório $Z(x)$ espacialmente correlacionado com média zero, o qual Cressie (1993) denomina de variação em pequena escala, e um processo ruído branco $[\eta(x) - resíduos\ da\ variável\ Y]$, espacialmente não correlacionado com média zero matematicamente. O modelo proposto por Cressie (1993) pode ser representado pela equação 6.

$$Y(x) = \mu(x) + \varepsilon(x) = \mu(x) + Z(x) + \eta(x) \qquad \text{Equação (6)}$$

3.5.1 Modelos lineares mistos

A geoestatística baseada em modelos, quando relacionada a estudos de variabilidade espacial de atributos do solo, se baseia primordialmente na classe de

modelos gaussianos lineares mistos, que se utilizam de dados distribuídos no espaço de forma regionalizada, ou seja, com coordenadas conhecidas. Nessa abordagem geoestatística destaca-se uma classe particular de modelos lineares generalizados, definidos como Modelos Lineares Mistos (MLMs), que contêm efeitos fixos e aleatórios (Isik *et al.*, 2017).

Sendo assim, os MLMs podem ser definidos como um modelo que possui componentes que permanecem constantes sob uma amostragem repetida (efeitos fixos) e outros componentes que variam randomicamente seguindo alguma distribuição de probabilidade (efeitos aleatórios). A distinção entre efeitos fixos e aleatórios frequentemente depende do interesse do pesquisador (Reichardt; Timm, 2021). Ou seja, por meio dos MLMs é possível modelar a variabilidade espacial dos dados.

Quando definimos um efeito como aleatório, significa que ele representa uma amostra de uma população maior. Assim, o interesse principal, na maioria das vezes, não é no valor em si, ou seja, no seu efeito médio, mas sim na variância contida no conjunto de dados em análise (Ghosh; Ramchandran, 2020).

Em suma, os efeitos fixos são os componentes estruturais e os efeitos aleatórios capturam as partes de variância e covariâncias do modelo. Nesse sentido, Du e Wang (2020), destacam que por meio dos componentes de variância do modelo é possível ajustar/modelar a heterogeneidade entre os dados em análise. Esse é um processo interessante, uma vez que influencia nas estimativas dos parâmetros e, de acordo com a estrutura assumida, pode resultar na não convergência do modelo.

Esses modelos são empregados para se compreender a relação existente entre uma variável resposta e uma ou mais covariáveis. Nesse contexto, de acordo com Du e Wang (2020), pode-se analisar a relação de componentes de variâncias dentro de cada conjunto de dados ou ainda analisar simultaneamente a relação entre os conjuntos de dados.

3.5.1.1 Estruturas de covariâncias

A modelagem da parte aleatória dos MLMs é realizada por meio da inclusão da matriz de variância e covariância (Du; Wang, 2020). Sendo que, neste estudo, a necessidade de se adicionarem parâmetros de variância e covariância está abarcada na premissa de que nos atributos físicos, hídricos e químicos do solo existem estruturas de dependência espacial.

De acordo com Diggle *et al.* (2002), para a escolha da matriz de variância e covariância adotada no estudo, deve-se analisar a variabilidade devido aos efeitos

aleatórios, ou seja, quando os dados em análise representam uma amostra aleatória de uma porção maior. A variabilidade espacial é significante, uma vez que se acredita que observações mais próximas medidas a campo estejam mais correlacionadas do que medidas mais distantes. Por fim, deve-se analisar a variabilidade atrelada aos erros de mensuração dos dados em análises. A seguir serão descritas, conforme Dessotti (2014), algumas das estruturas que podem ser adotadas na matriz de variância e covariância, sendo que para a sua escolha deve-se conhecer as características das variáveis de interesse no estudo.

Na análise não estruturada, as variâncias e covariâncias podem ser distintas. As variâncias possuem apenas valores positivos e não há restrições em relação às covariâncias. Por sua vez, os componentes de variância são descritos por variâncias iguais e covariâncias nulas, ou seja, assume-se que as variáveis são independentes. Em contrapartida, no componente de variância com heterogeneidade a estrutura mantém a suposição de independência, contudo, o mesmo não ocorre com a homogeneidade de variâncias. Na de simetria composta assume-se variâncias e covariâncias iguais, entre todas as observações, que possuem uma mesma unidade, devido a erros igualmente correlacionados.

3.5.1.2 Seleção dos modelos

Para definir o MLM adequado, seleciona-se os efeitos fixos, aleatórios e suas estimativas, bem como realiza-se a comparação dos modelos. A comparação dos MLM pode ser realizada por meio dos critérios de informações de Akaike (AIC) e bayesiano (BIC). Estes baseiam-se no valor do logaritmo natural da função de verossimilhança restrita do modelo.

O critério de AIC baseia-se na definição de entropia, fornecendo uma medida relativa das informações perdidas, de quando um determinado modelo é usado para descrever a realidade. Ele se baseia no fato de que o viés tende ao número de parâmetros a serem estimados no modelo (Equação 7).

$$AIC = -2\, logL(\hat{\theta}) + 2\,(k) \qquad (7)$$

Em que, o $logL(\hat{\theta})$ representa a função suporte maximizada e K o número de parâmetros envolvido no modelo. Por conseguinte, o critério de informação bayesiano é um critério de avaliação de modelos definido em termos da probabilidade. Descritos

matematicamente de acordo com a Equação 8, em que $f\left(\frac{xn}{\theta}\right)$ é o modelo escolhido, p é o número de parâmetros a serem estimados e n é o número de observações da amostra.

$$BIC = -2\ log\ f\left(\frac{xn}{\theta}\right) + p\ logn \quad (8)$$

Os modelos são dependentes do número de observações e do número de parâmetros do modelo (fixos e aleatórios) (Pereira *et al.*, 2020). Akaike (1974) destaca que se dois modelos representam dados de forma distinta e satisfatória, a seleção do melhor modelo é representada pelo modelo mais simples. Por conseguinte, o critério bayesiano é similar ao de AIC, uma vez que busca escolher o modelo com menor número de parâmetros (Ariyo *et al.*, 2020). Sendo assim, busca-se em ambas as métricas encontrar valores baixos para se ter uma melhor modelagem.

3.5.1.3 Estimação dos componentes de variâncias

Dentre os métodos empregados para a estimativa da variância tem-se: a análise de variância (ANOVA), a máxima verossimilhança (ML), método da estimação quadrática não viesada da variância mínima (MIVQUE) e o método da máxima verossimilhança restrita (REML). Nesse sentido, destaca-se que no método da máxima verossimilhança, maximiza-se a função de verossimilhança, com o respectivo de cada variável utilizando os dados observados e o modelo especificado. Cabe destacar que nesse método assume-se que os efeitos fixos são conhecidos sem erros associados.

Já o REML leva em consideração os graus de liberdade envolvidos na estimativa dos parâmetros fixos do modelo, o que diminui o enviesamento da estimativa e permite a imposição de restrições de não negatividade (Ghosh; Ramchandran, 2020). O REML possibilita a maximização da função de verossimilhança de todos os contrastes de erros ou resíduos, ou seja, a parte que é invariante aos efeitos fixos (Torabi; Jiang, 2020).

3.5.1.4 Estimação dos efeitos fixos e predição dos efeitos aleatórios

As estimativas dos efeitos fixos e aleatórios são obtidas por meio do uso de funções baseadas nas funções de verossimilhança dos dados. Nos MLMs os efeitos fixos serão calculados por meio do BLUE (*Best Linear Unbiased Estimator*), que é o melhor estimador não viesado desenvolvido a partir de dados referentes a uma variável de natureza regionalizada. O estimador é dito linear, pois a estimação é obtida combinando

os valores amostrados e ponderando os pontos amostrais para que se obtenha o menor erro quadrático médio. Os efeitos aleatórios serão calculados por meio do BLUP (*Best Linear Unbiased Predictor*), que é o melhor preditor linear não viesado (Torabi; Jiang, 2020). Esses são os "melhores estimadores", pois minimizam a variância da amostra linear (Dessotti, 2014).

3.5.1.5 Inferência para parâmetros de efeitos fixos e aleatórios

Em termos gerais quando se analisam os parâmetros de efeitos fixos do MLM, tem-se o uso dos testes de: Wald, t e wald-F. Esses testes são utilizados para analisar a significância dos termos fixos, ou de uma combinação linear deles. Os testes de t e wald-F têm a incorporação da variabilidade introduzida pela estimação dos componentes de variâncias (Stram; Lee, 1995), ao contrário do teste de Wald. Já para testar a permanência ou não de um efeito aleatório no modelo, pode-se utilizar, dentre outros, o teste da razão de verossimilhança (*likelihood ratio test – LRT*) (Self; Liang, 2009).

Destaca-se ainda a importância fundamental da análise dos resíduos, uma vez que os resíduos são utilizados para examinar as suposições do modelo linear misto criado e detectar a presença de valores atípicos e possíveis observações errôneas (Torabi; Jiang, 2020).

Maiores detalhes sobre o assunto podem ser encontrados, por exemplo, em Haskard (2007), Diggle e Ribeiro Júnior (2007), Isik *et al.* (2017), Slaets *et al.* (2021), dentre outros textos.

3.5.1.6 Estado da arte dos modelos lineares mistos

Atualmente, são escassos os estudos envolvendo os MLMs quando relacionado à análise espacial de áreas agrícolas. Entretanto, essa ferramenta estatística tem sido amplamente difundida em diversas áreas do saber. Aqui destacamos os estudos de Romić *et al.* (2020), que utilizaram os MLMs para modelar espaço-temporalmente a qualidade da água de diferentes estações. Por conseguinte, na engenharia de automação Heuberger, Bains e Mac Dowell (2020) desenvolveram um modelo de otimização espaço-temporal, denominado ESONE, para investigar o impacto da implantação de veículos elétricos. Na área da saúde, Gasparini *et al.* (2020) geraram um modelo para monitoramento à distância de pacientes, com base em seus sintomas. Os autores buscaram monitorar de forma mais precisa a saúde dos pacientes em piores condições.

Outro estudo interessante foi o realizado por de Faveri *et al.* (2017), que modelaram a variância e covariância residual, em dados anteriormente incorporados à análise de correlação canônica. Esses dados representavam a seleção de variedades de uma determinada cultura, na qual os autores buscaram analisar os efeitos genéticos. Com esse modelo, conseguiram analisar múltiplos dados simultaneamente, bem como identificar as múltiplas características. Apesar de ser extremamente importante, abarca resultados voltados ao melhoramento genético de uma determinada cultura e não à análise de dados físicos, químicos e hídricos do solo.

Outro estudo interessante foi devolvido por Charles *et al.* (2022), no qual avaliaram diferentes estruturas de covariância para análises de multicolheita e multisite para otimizar o melhoramento de plantas perenes, analisando dois conjuntos de dados. Os modelos forneceram uma melhor visão sobre a dinâmica das variações ao longo dos anos/sítios de colheita. Apesar dos resultados semelhantes, o modelo padrão, modelado com estruturas de covariância que assumem homogeneidade de covariâncias, não foi o modelo estatisticamente mais adequado para D2 de acordo com os critérios de informação. Portanto, a modelagem de estruturas de covariância pode e deve ser utilizada na avaliação genética de plantas perenes.

Quando relacionado aos MLMs envolvendo geoestatística aplicada a ciência do solo, Slaets, Boeddinghaus e Piepho (2020) buscaram introduzir as vantagens da integração da análise de variância e covariância com a geoestatística. Os autores buscaram correlacionar as terminologias das ferramentas estatísticas citadas, bem como encontrar pontos semelhantes entre essas definições, visando a incorporação desses métodos simultaneamente.

Rosemary *et al.* (2017) utilizaram o MLM para compreender o padrão de distribuição espacial dos atributos do solo frente aos efeitos da variável categórica, uso do solo, levando em consideração a autocorrelação dos resíduos padronizados. Os resíduos padronizados foram usados para calcular o semivariograma experimental, sendo o melhor modelo definido por meio do critério de AIC e BIC. Com isso, os autores encontraram um efeito significativo do uso do solo na variabilidade espacial da condutividade elétrica. Ademais, o uso do solo nessa área é destinado para o cultivo do arroz, que pode estar relacionado, de acordo com os autores, à qualidade da água utilizada para a irrigação, em virtude das altas concentrações de sais solúveis nas áreas de cultivo.

He *et al.* (2022) aplicaram um modelo híbrido combinando Modelagem de Regressão Múltipla (MRM) e Modelos Lineares Mistos (MLM) para a previsão rápida de

temperaturas interiores afetadas por diodos térmicos de tubo de calor e cavidades solares com base em dados experimentais. No MRM, as fórmulas multivariadas lineares foram geradas de acordo com as duas condições de trabalho diferentes do diodo térmico, no caso de MLM, foi desenvolvido um modelo usando os dados experimentais. A semelhança entre a previsão e os resultados experimentais demonstra claramente a precisão e eficiência do modelo. Assim, a pesquisa foi uma tentativa de integrar ferramentas computacionais emergentes e fornecer um meio para realizar análises quantitativas altamente eficientes de ambientes térmicos internos para estudos ambientais e projetos sustentáveis nos estágios iniciais.

Heiskanen *et al.* (2018) empregaram os MLMs para estudar a relação de algumas propriedades físico-hídricas (curva de retenção de água no solo) do solo, com a cobertura vegetal do solo e puderam observar que o efeito fixo, descrito por meio de um índice de qualidade (composto pelas alturas das árvores e das frações finas do solo), tem relação direta com a textura do solo, pH e a capacidade de campo, o que propiciou a geração de MLMs de previsão apropriados para o crescimento da mata nativa, a hidrologia e a trafegabilidade local. Contudo, os autores destacam que a quantidade amostral de pontos foi um dos fatores limitantes em seu estudo, deixando evidente que o efeito fixo adotado pode sim descrever os atributos analisados. Porém, para isso seria necessária a incorporação de dados principalmente relacionada à parte oeste e sul da área em estudo na Finlândia, nesse caso, em áreas costeiras.

Deiss *et al.* (2019) avaliaram como a distribuição espacial e as características das propriedades químicas em solo sob sistema de integração lavoura-pecuária afetam os sistemas de plantio direto. Esse experimento foi um dos poucos encontrados em nível nacional, sendo realizado na região Sul do Rio Grande do Sul. Nesse estudo, as profundidades analisadas foram consideradas como efeitos fixos e as propriedades químicas do solo os efeitos aleatórios. Os autores constataram que o sistema de integração aumentou a fertilidade do solo, o que resultou em uma maior produtividade na safra.

Burst *et al.* (2020) objetivaram responder, por meio dos modelos lineares mistos, como o uso do solo (floresta e pastagem) influencia na camada superficial a distribuição das propriedades do solo no nordeste da França. Ademais, analisaram até que ponto a mata nativa e o desmatamento modificam as propriedades do solo. Dentre os principais resultados encontrados, o pH diminuiu conforme a profundidade em áreas de mata nativa e aumentou em áreas de pastagem. O carbono orgânico e alguns nutrientes na pastagem diminuíram, conforme penetra as camadas do perfil do solo. Em contraponto, na mata

nativa houve um aumento desses elementos. Sendo assim, os autores constataram que de fato o uso da terra, a vegetação e as práticas de manejo têm forte impacto nas propriedades do solo.

Barca, De Benedetto e Stellacci (2019) analisaram a acurácia de dois sensores geofísicos, a saber: GPR (*Ground Penetratin Radar*) e EMI (*Electromacnetic Induction*) na análise do conteúdo de água no solo. Posteriormente, os autores modelaram os dados de umidade do solo por meio dos MLMs e da krigagem ordinária, realizando um comparativo entre métodos.

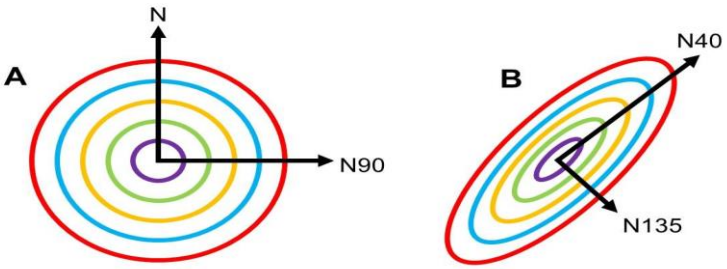
Destaca-se que nos estudos supracitados, envolvendo dados físico-hídricos e químicos do solo, mesmo realizando a análise de variância e covariância e incorporando os efeitos, nenhum considerou a anisotropia dos dados. Sendo assim, são inexistentes, estudos que descrevam e expliquem a incorporação da variabilidade direcional de dados relacionados à física do solo em MLM (anisotropia), ou seja todos os estudos são isotrópicos.

3.5.1.7 Fenômeno anisotrópico

Frequentemente os dados físicos-hídricos e químicos do solo, tem que ser transformados para um cenário parcimonioso de covariância isotrópica e estacionária, ou seja, onde a estrutura de covariância carece de direcionalidade e de dependência de localização (Koch; Lele; Lewis, 2020), conforme supracitado nos modelos lineares mistos. Contudo, essa modificação nos dados, pode trazer perdas de informações e ainda tornar-se um processo gerador de erros.

Nesse contexto, é importante destacar que um determinado dado espacial apresenta anisotropia quando o semivariograma direcional muda conforme a orientação; quando ele não se altera com a direção, tem-se um fenômeno isotrópico (Yamamoto; Landim, 2013). Se for função apenas da distância euclidiana entre as observações da variável, o semivariograma apresenta estacionariedade de segunda ordem e o fenômeno é classificado como isotrópico (Figura 3A). Caso contrário, é considerado anisotrópico (Figura 4B) (Zimmerman, 1993).

Figura 4 – Representação esquemática de fenômenos isotrópicos.

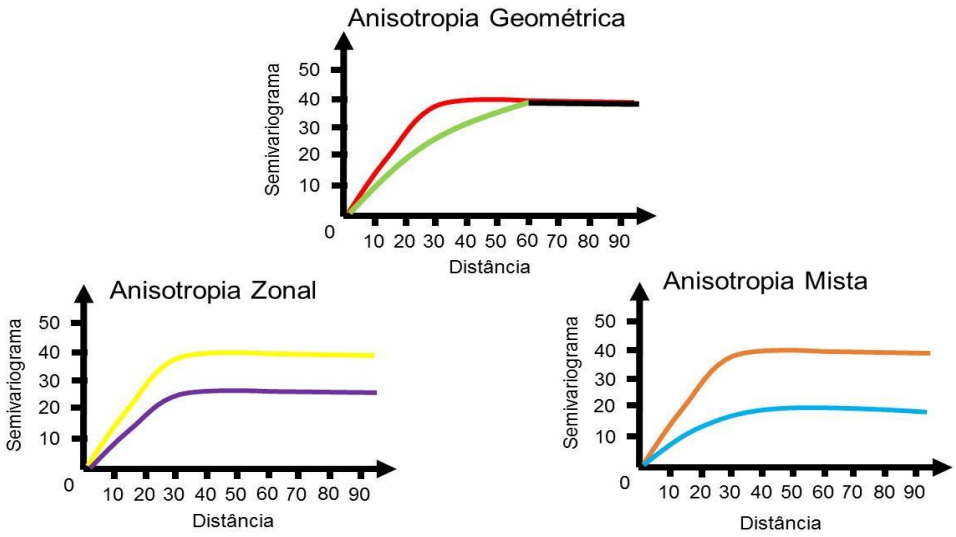


Legenda: (A) e anisotrópicos (B), onde o espessamento das linhas indica aumento do valor da variável. Fonte: Adaptado de Yamamoto (2020).

Existem diferentes tipos de anisotropia na natureza, a saber: a anisotropia geométrica, que se caracteriza pela existência de um único patamar e dois alcances distintos (Isaaks; Srivastava, 1989). A anisotropia zonal, que possui patamares distintos, conforme a direção analisada, mas todos possuem o mesmo alcance e a anisotropia mista, onde tanto a amplitude como o patamar variam conforme a direção (Yamamoto; Landim, 2020). Em contrapartida Journel e Huijbregts (1978), corroborados por Olea (2009), definem que anisotropia zonal é qualquer tipo de anisotropia que não é geométrica, explicitando o exemplo de anisotropia mista definido por Yamamoto e Landim (2013). Isaaks e Srivastava (1989) apresentam a mesma definição para ansinotropia zonal de Yamamoto e Landim (2013).

Em suma, à definição da existência de anisotropia procede-se determinando as direções correspondentes às amplitudes mínima e máxima, quando se tratando de anisotropia geométrica ou aos patamares máximo e mínimo no caso da anisotropia zonal (Isaaks; Sirivatava, 1989). Na figura 5 encontram-se alguns dos modelos de anisotropia existentes na natureza.

Figura 5 – Modelos comuns de anisotropia encontrados na natureza.



Fonte: Adaptado de Yamamoto (2020).

Portanto, a identificação e o uso da anisotropia em modelos espaciais são de grande importância para a estimativa de locais não amostrados. Sendo assim, se um modelo isotrópico for incorporado em um determinado estudo, a magnitude das dispersões aleatória das observações é espacialmente relacionada e considerada a mesma em todas as direções. Se um modelo anisotrópico é usado, a magnitude dessas flutuações, bem como a extensão das relações espaciais entre as observações pode variar conforme a direção (Crawford; Hergert, 1997).

A incorporação denominada correção da anisotropia é a obtenção de um semivariograma “isotrópico” para o modelo de correlação espacial, ou seja, busca-se um modelo com parâmetros comuns (efeito pepita, variância espacial e amplitude) em todas as direções. A seguir é apresentado a correção da anisotropia geométrica, de acordo com Yamamoto (2020), visto que é o objeto deste estudo.

3.6 Simulação sequencial gaussiana

Baseado em Reichard e Timm (2022), a simulação geoestatística objetiva gerar um conjunto de representações equiprováveis da distribuição espacial dos valores do atributo em estudo e usar as diferenças entre os mapas simulados como uma medida de incerteza. Esta avaliação da incerteza acerca dos valores do atributo em estudo é uma etapa preliminar para avaliar o risco envolvido em qualquer processo de tomada de decisão. Desta forma a simulação sequencial consiste em condicionar a estimativa de uma variável Z em uma posição x a toda informação disponível na vizinhança de x, incluindo

valores observados e valores previamente simulados. Portanto, o termo sequencial deriva do fato de que o conjunto de dados condicionante é progressivamente atualizado conforme os valores são simulados. De acordo com Goovaerts (1997), os seguintes passos são adotados para a simulação sequencial:

1. definição de um caminho aleatório entre os nós;

2. para cada nó, encontrar os dados vizinhos mais próximos, incluindo os dados originais e valores de nós previamente simulados;

3. usar a krigagem simples para estimar a média da distribuição gaussiana e o desvio padrão da krigagem simples para calcular o erro. Relembrando que a krigagem simples pressupõe que a média do conjunto de dados é conhecida e considerada constante em todo o domínio amostral (Yamamoto, 2020);

4. adicionar o valor simulado ao conjunto de dados;

5. ir ao próximo nó e repetir até que todos os nós tenham sido visitados.

Dentre os métodos de simulação sequencial mais utilizados, tem-se a simulação sequencial gaussiana (SSG), a qual utiliza a média e o desvio padrão da distribuição normal da krigagem para gerar a forma da distribuição das incertezas associadas à estimativa em um determinado ponto amostral.

O método SSG é baseado no fato de que, no i-ésimo nó, a distribuição de frequências acumulada condicional (ccdf) é modelada com o conjunto de dados original e todos os valores simulados nos i-1 nós visitados anteriormente, podendo ser usada para a simulação de nós subsequentes. A SSG é baseada na aplicação do modelo de funções aleatórias multigaussianas, que é o modelo paramétrico mais amplamente usado porque possui propriedades extremamente convenientes, tornando simples a inferência dos parâmetros da distribuição de frequências acumulada condicional (ccdf). O método SSG começa com uma identificação da distribuição normal da variável em estudo, sendo esta imprescindível para realização dos próximos passos.

Os dados originais (ou transformados) da variável $Z(x)\{[z(xi), i = 1,..,n]\}$ são convertidos em valores $y(x)\{[y(xi), i = 1,\ldots,n]\}$, que são os escores da distribuição normal, com média zero $\{E[Y(x)] = 0\}$ e variância unitária $\{Var[Y(x)] = 1\}$. Logo após, o semivariograma experimental da variável transformada $Y(x)$ é calculado e um modelo de semivariograma teórico $\gamma Y(h)$ será usado na simulação sequencial gaussiana realizada para a variável $Y(x)$, como é visto a seguir:

i. um caminho aleatório visitando cada nó x' é definido, contudo cada nó da malha é visitado somente uma vez;

ii. em cada nó $x'j$ ($j = 1, 2, \ldots, N$ nós) da malha simulada, os parâmetros gaussianos (média e variância), que definem a função de distribuição acumulada condicional, são determinados usando a krigagem simples com o modelo de semivariograma dos escores da distribuição normal. As informações de condicionamento (n) consistem em um número especificado $n(x')$ dos escores normais $y(xi)$ e dos valores $y(l)(x'j)$ simulados nos nós da malha visitados anteriormente ("l" indica o l-ésima realização). Cada nó $x'j$ da malha simulada possui uma ccdf gaussiana associada a ele, condicionada ao número de dados amostrados e aos dados simulados anteriormente usando krigagem simples;

iii. extrai-se aleatoriamente um valor simulado $y(l)(x'j)$ da ccdf gaussiana determinada no nó x'_j e esse valor simulado é adicionado ao conjunto de dados;

iv. prossegue-se para o próximo nó da malha simulada ao longo do caminho aleatório e repete-se os passos ii e iii até que todos os N nós da malha sejamos simulados;

v. ao final da simulação sequencial gaussiana são obtidos os conjuntos dos valores simulados $\{y(l)(x'j), j = 1, \ldots, N\}$ que estão no domínio da distribuição Gaussiana. Desse modo, os valores simulados devem ser retrotransformados para a escala original da variável $Z(x)$, aplicando a função inversa de transformação dos dados gaussianos que depende do tipo de transformação usada, ou seja, se, por exemplo, a função logaritmo neperiano foi adotada, tem-se que aplicar a função inversa que é a função exponencial e;

vi. o conjunto resultante de valores simulados $[z(l)(x'j), j = 1,2, \ldots, N]$ representa uma realização l ($l = 1, 2, \ldots, L$) da função aleatória $[Z(xj)]$ nos N nós $x'j$. Qualquer número L de realizações $[z(l)(x'j), j = 1,2, \ldots, N], l = 1, 2, \ldots, L]$ pode ser obtido repetindo L vezes todo o processo sequencial com possíveis caminhos diferentes para visitar os N nós. O procedimento de simulação gera, em cada nó $x'j$, uma distribuição de L valores os quais pode ser usada para uma aproximação numérica da ccdf.

Ao invés de um mapa das melhores estimativas locais (krigagem), o SSG gera um mapa ou uma realização l de valores z, ou seja, $[z(l)(x'j), j = 1, \ldots, N, xi \in \mathcal{R}d, d = 2\ dimensões]$, que reproduz estatísticas consideradas mais importantes para o problema em questão. Os requisitos típicos para esse mapa simulado são os seguintes:

i. os valores simulados $z(l)(xi)$ devem honrar os valores amostrados $z(xi)$;

ii. o histograma de valores simulados em cada realização l reproduz satisfatoriamente o histograma dos dados amostrais; e

iii. o modelo teórico de semivariograma do conjunto de dados original é reproduzido por cada modelo simulado de semivariograma gerado por cada realização l.

Com relação a avaliação das incertezas**,** o intuito central é obter um conjunto de L representações equiprováveis da distribuição espacial da variável Z em estudo e usar as diferenças entre os L mapas simulados como uma medida de incerteza (Goovaerts, 2001), ou seja, o conjunto de L realizações $[z(l)(x'j), j = 1,2, \dots, N], l = 1, 2, \dots., L,$ fornecerá uma medida visual e quantitativa da incerteza espacial da variável em estudo.

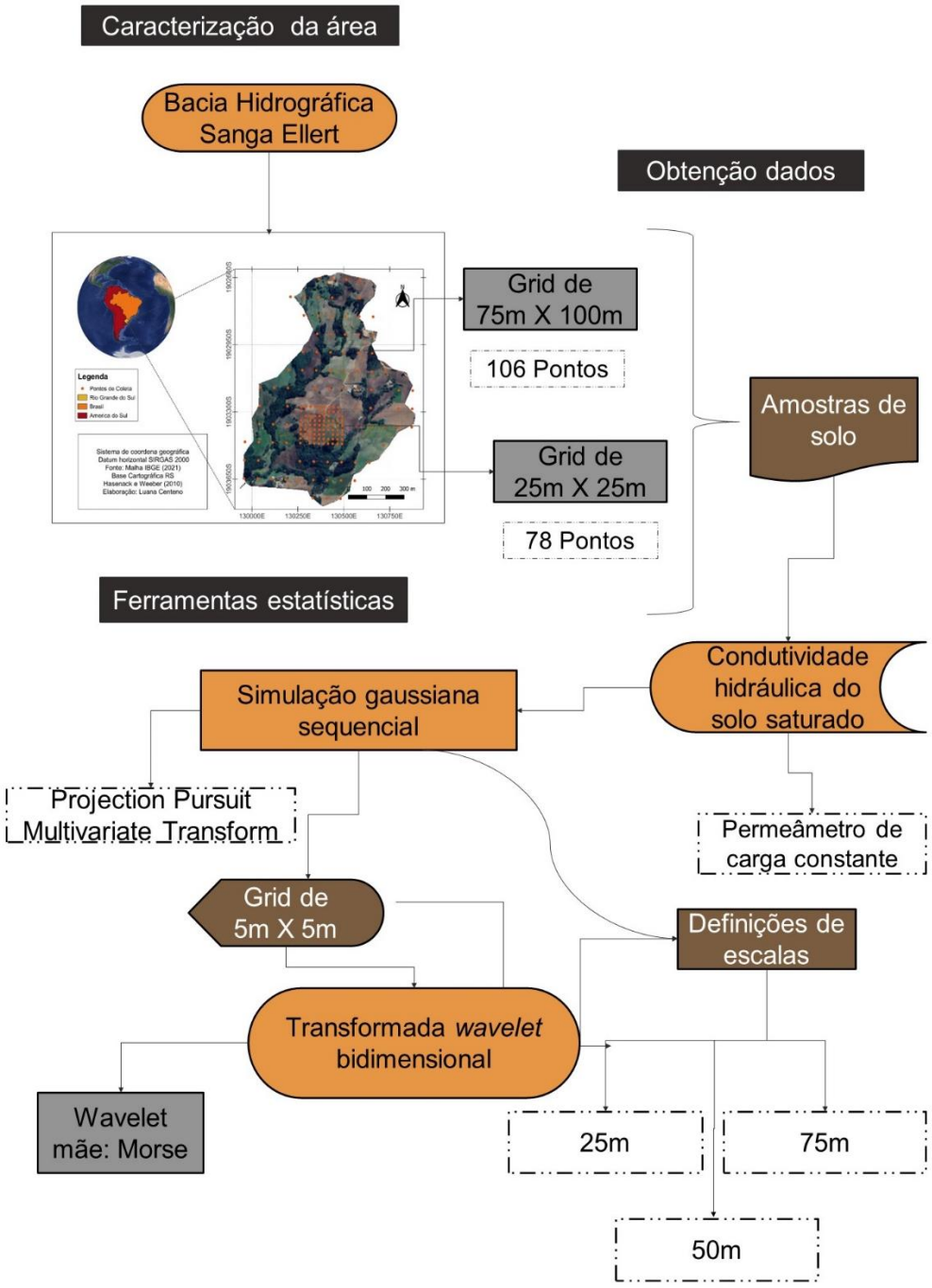
4. ESTRUTURA DOS ARTIGOS

A seguir serão apresentados os resultados que estão divididos em dois artigos. Sendo o artigo 1 relacionada à variabilidade espacial, analisada por meio dos modelos lineares mistos em áreas agrícolas descrito resumidamente através do fluxograma 1.

Fluxograma 1: Descrição da metodologia empregada na análise de simulação dos modelos lineares mistos.

E o artigo 2 está relacionado à variabilidade espacial em escala de bacia hidrográfica, estudada por meio de transformadas bidimensionais de Wavelet, descrito resumidamente através do fluxograma 2.

Fluxograma 2: Descrição da metodologia empregada no estudo em escala de bacia hidrográfica aplicando transformadas bidimensionais de wavelet.



4.1 CAPÍTULO 1

UNDERSTANDING THE PERFORMANCE OF THE MÁTERN MODEL FOR SPATIAL SOIL VARIABILITY THROUGH A SIMULATION STUDY

A modified version of this manuscript is in the review process of a peer-review in European Journal of Soil Science.

ABSTRACT: This study consists of a data simulation and data analysis to examine the spatial variability of soil attributes, a crucial aspect for effective soil management in agriculture. The goal was to better understand this variability to support decisions such as fertilizer application, crop selection, and identification of areas with high production potential. The study specifically focused on assessing the performance of the Matérn model in analyzing spatial variability, taking into account factors like area size, sample density, spatial sample distribution, and data collection methods and analysis. Two packages in R language, Asreml-R and GeoR, were used to compare the spatial data analysis. The results revealed limitations in using the Matérn model, especially in scenarios with greater distances between sampling points. and, consequently, more extensive ranges. This led to convergence issues, particularly evident in the Asreml-R. The choice of initial parameter values influenced the convergence rate in the GeoR package. The study highlights the importance of careful model settings and package selection when analyzing spatial variability. The flexibility of the Matérn model is an advantage, but its performance can be affected by greater distances between sampling points. Transformations in distance metrics may be necessary to improve the analysis. Future work will include incorporating nugget and anisotropy effects and applying the findings to soil data collected in the field. These findings are crucial for enhancing sustainable soil management and decision-making in agriculture.

KEYWORDS: Spatial variability; soil attributes; experimental design; packages comparison.

INTRODUCTION

Studying the spatial variability of soil attributes is a key subject for making rational and adequate soil and water management decisions in agricultural lands. These

include applying fertilizers, choosing suitable crops, and identifying areas with greater production potential (KALANTARI et al., 2023). Analyzing the spatial variability of soil attributes allows finding patterns and trends that would not be easily observed in a random sampling approach (AYELE et al., 2022; CHEN et al., 2022). In this regard, soil attributes are affected by spatial factors such as topography, vegetation, drainage, and impermeable layers (YU et al., 2023; KALANTARI et al., 2023). Collecting information about these attributes at various locations in space assists in a better understanding of the processes occurring in soil and their interrelationships.

However, it is essential to construct the experimental design of the sampling grids to obtain correct and representative information on the spatial variability of soil attributes (ZHOU et al., 2023). A well-planned experimental strategy considers the size of the study area, the sampling density, the grid sampling arrangement, and the data collection method (SOLANKI et al., 2022; MAIER et al., 2022). This consideration is of utmost importance, as an insufficiently planned design can result in skewed or unreliable outcomes, ultimately hindering data interpretation and effective decision-making (SLAESTS; BOEDDINGHAUS; PIEPHO, 2020).

A commonly employed approach to assess this variation is a geostatistical model, which focuses on describing the spatial dependence of data. This type of model uses a variogram to describe how the variability between points varies according to distance (SLAETS; BOEDDINGHAUS; PIEPHO, 2019). However, it is important to note that geostatistical model can be considered as a specific case of a broader framework known as the linear mixed model (MLM). Within the geospatial context, the linear mixed model offers a more comprehensive framework for understanding data variation.

Within the linear mixed model, data is treated as a combination of fixed and random components. Fixed effects represent deterministic factors, such as the overall mean or coefficients from a spatial trend model. On the other hand, random effects capture variation that is spatially dependent, reflecting variation not explained by fixed effects and modeling the spatial dependence of the data (LARK; CULLIS and WELHAM, 2006). Furthermore, in the geospatial context, there also exists an independent random error (nugget). This component represents the residual variability that is explained neither by fixed effects nor by spatial random effects. Essentially, it is the unrestricted variation that can exist in the data (HASKARD, 2007).

In this context, the Matérn geostatistical model, also known as the Matérn family, plays a crucial role in the analysis of the spatial variability of soil attributes. It is a

mathematical model used to describe spatial dependence in geostatistics based on models considering the distance and direction between sample points (GILMOUR; CULLIS; VERBYLA, 1997; HASKARD; CULLIS; VERBYLA, 2007).

An advantage of the Matérn Model is its flexibility. It is possible to adjust up to five parameters: ϕ – range; ν – smoothing; δ – anisotropy rate; α - anisotropy angle; λ - distance metric. This versatility enables it to accommodate a wide range of adjustment options and effectively capture diverse behaviors similar to the original semivariogram. The fact of Matérn's versatility is ignored when using the semivariogram of the spherical and exponential models, to account for the uncertainties of the semivariogram (MARCHANTY; LARK, 2007). In addition, those models assume that all the parameters are aligned precisely, which leads to uncertainty in their estimates (WEBSTER; OLIVER, 2001).

Several authors discuss the use of the Matérn family model by Diggle and Ribeiro (2003), Minasny and McBratney (2005), and Haskard, Cullis, and Verbyla (2007). The smoothness parameter (v) describes spatial processes with different local behaviors. Like other semivariogram parameters, ν can be estimated from the observed data set, with the flexibility to test different ν values in the Matérn model. When ν is 0.5, the model is exponential; when it is 1, the model is called Whittle model (WEBSTER; OLIVER, 2001). Therefore, if the smoothness parameter is large, i.e. unlimited, it approaches the power function ($v=\infty$), also called Gaussian. In this case, it can be considered a generalization of various theoretical models (MINASNY; MCBRATNEY, 2005).

Another crucial factor when analyzing the spatial variability of soil attributes is the support of suitable statistical packages. Two packages used in this context are Asreml-R and GeoR, both in R packages. Asreml-R is a commercial statistical package, but specific for mixed model analysis, which allows complex models to be adjusted considering the spatial-dependence structure of the data (BUTLER, 2022). It offers advanced features for estimating parameters and making statistical inferences. GeoR is a free statistical package developed for spatial analysis of georeferenced data. It supplies tools for modeling and analyzing spatial dependence, allowing the creation of maps and visualization of results more intuitively (DIGGLE; RIBEIRO, 2007). Besides linear models, GeoR includes a variety of spatial statistical methods, such as kriging, interpolation, and spatial simulation, which are useful for analyzing the spatial variability of soil attributes (RIBEIRO et al., 2020).

Simulation plays a significant role in adequately estimating the parameters related to the spatial variability of soil attributes before statistical modeling. This technique allows synthetic data sets to be generated based on mathematical and statistical models. By simulating data, it is possible to evaluate the sensitivity of results to different scenarios and conditions and thus understand the statistical models used. Simulation can also be used to estimate the uncertainty associated with the results of spatial variability analyses. By generating multiple sets of simulated data, it is possible to obtain statistical distributions of the estimated parameters and to calculate confidence intervals.

In this sense, several soil spatial variability studies have been conducted using the Matérn family. For example, Stefanova, Smith, and Cullis (2009) analyzed a series of uniformity field trials using the technique proposed by Gilmour, Cullis, and Verbyla (1997), and presented a variety of spatial analyses to aid the spatial modeling process. On the other hand, Minasny and McBratney (2005) studied the influence of the variogram on the description and quantification of soil spatial variability using a simulation study. McCullagh and Clifford (2005) used the Matérn model to investigate the nature of spatial correlation in crop yields, focusing on natural variation through modifications of the smoothing parameter. Marchant and Lark (2007) went beyond simulation by using the Matérn model, exploring the optimization of sample planning and reducing the uncertainty associated with estimating soil attributes.

Furthermore, the justification for carrying out this research derives from the current gap in comparative analysis between packages tools available for spatial variability studies. Although there is an extensive literature that supports the use of Matérn family models, there is a lack of work that directly compares the effectiveness of these models in different statistical packages. The absence of studies investigating the performance of GeoR, a free package, compared to Asreml-R, a commercial package, to analyze spatial variability, represents a critical gap. Considering the increasing accessibility and popularity of open-source tools, understanding the potential advantages or limitations of GeoR in contrast to Asreml-R becomes critical. Furthermore, addressing issues of convergence problems and other limitations is essential, as they can significantly impact the accuracy and reliability of results when related to real data. The goal was to better understand this variability to support decisions such as fertilizer application, crop selection, and identification of areas with high production potential. The study specifically focused on assessing the performance of the Matérn model in analyzing

spatial variability, taking into account factors like area size, sample density, spatial sample distribution, and data collection methods and analysis.

MATERIAL AND METHODS

The simulation study considered a rectangular area with a combination of 16 observational units in the vertical direction and 25 in the horizontal direction, totaling 400 observational units. The response variable, $y(s)$, can be modeled using Equation 1 according to Haskard, Cullis, and Verbyla (2007), as follows:

$$y(s) = X_s\tau_s + Z_s u_s(s) + e \quad (1)$$

where $y(s)$ is spatially evaluated in all 400 observational units ($n = 400$); X_s is a $n \times p$ matrix of polynomials of the spatial location, being represented by a linear function (1st degree); τ_s is the vector of coefficients of the polynomial regression with dimension $p \times 1$– at this point of the study $p = 1$, since it was only assumed $\tau_s = \mu$ (mean of the process under study) that $X_s = 1_n$ is an unitary vector; Z_s is an indicator matrix for the random effects at different locations, typically it is an identity matrix of dimension $n \times n$, because all observations are considered distinct; $u_s(s)$ is the spatial random effect, i.e., the spatially correlated residual error, such that $u_s \sim N(0, \sigma_s^2 G_s)$. $e = (e_1, \ldots, e_n)^T$ is the independent error vector (called nugget effect), in which $e \sim N(0, \sigma^2 I_n)$. σ^2 and σ_s^2 are the variances of the nugget effect and the spatially correlated process, respectively.

Adopting γ as the ratio between σ^2 and σ_s^2, i.e., $\gamma = \frac{\sigma^2}{\sigma_s^2} \Rightarrow \sigma^2 = \sigma_s^2 . \gamma$. Therefore, $y(s) \sim N(1_n\mu, \sigma_s^2(\gamma I_n + G_s))$. The elements of G_S are given by $\rho(s_i - s_j, \varphi)$, being $\rho(\cdot)$ a correlation function with a parameter vector φ which is dependent on the spatial separation vector $h_{ij} = s_i - s_j$. The matrix G_S is considered positive definite. The i and j subscripts of s indicate that the elements of G_S are specifically related to spatial effects (Haskard, 2007). In this study, the isotropic spatial correlation function ($\rho(\cdot)$) of the Matérn family was used (Equation 2).

$$\rho(.) = \rho M(d; \phi, \nu) = \{2^{\nu-1}\,\Gamma(\nu)\}^{-1} \left(\frac{d}{\phi}\right)^{\nu} K_\nu \left(\frac{d}{\phi}\right) \quad (2)$$

where $d > 0$ is the separation distance between observational units; $\phi > 0$ is a range or distance parameter, $\nu > 0$ is a smoothness parameter, $\Gamma(\cdot)$ is the gamma function, and $K\nu(.)$ is the modified Bessel function of the third order type ν.

To evaluate the complexity of the model, three options for spacing between observational units were defined: 1×1, 10×10 and 100×100. Thus, different ϕ values were simulated based on data sets from Bitencourt *et al.* (2015). Here, we used ϕ values of $3, 5$ and 10, which are multiple values of the assumed three grid spacings. Table 1 shows nine simulated scenarios constructed varying the ϕ - Matérn model parameter, assuming a regular and an irregular experimental grid.

Table 1: The nine simulated scenarios varying ϕ - Matérn model parameter values and grid spacings.

	ϕ		
Grid spacing	Multiple of 3	Multiple of 5	Multiple of 10
1×1	Scenario 1 ($\phi = 3$)	Scenario 2 ($\phi = 5$)	Scenario 3 ($\phi = 10$)
10×10	Scenario 4 ($\phi = 30$)	Scenario 5 ($\phi = 50$)	Scenario 6 ($\phi = 100$)
100×100	Scenario 7 ($\phi = 300$)	Scenario 8 ($\phi = 500$)	Scenario 9 ($\phi = 1000$)

Three different values for the smoothing parameter ($\nu = 0.5, 1$ and 2) were also used, totalizing 27 simulated scenarios. For each scenario, 100 simulated data sets were generated using the *grf* function of the GeoR package (RIBEIRO Jr et al., 2020). From these generated data, four studies were performed as follows:

First study – isotropic correlation function without nugget effect

All ϕ and ν combinations were used considering regular and irregular grids. The construction of the irregular grid occurred in a random way according to the characteristics of the effects of the Matérn model. Figure 1 shows an example of this study, in which $\nu = 0.5$ and $\phi = 3$ at (a) and (b); $\nu = 0.5$ and $\phi = 30$ at (c) and (d); $\nu = 0.5$ and $\phi = 300$ at (e) and (f).

Second study – isotropic correlation function with nugget effect

All ϕ and ν combinations proposed in the first study were also used here. A nugget effect $\sigma^2 = 0.4$ was assumed (Haskard, 2007).

Third study – rescheduling grid spacing schemes.

Here, the x- and y-coordinates in scenarios 4 to 9 (Table 1) were resized, making the distance between the sample points close to 1. Then, where the scale was 10×10, the x and y coordinates were divided by 10; and where it was 100 × 100, it was divided by 100.

Fourth study – considering the anisotropic effect

New data sets were simulated using 1, 3, 4 and 7 scenarios taking $\nu = 0.5$ and 1 and 3 scenarios taking $\nu = 1.0$ considering the anisotropic effect on the correlation function. It was used the anisotropy angle $(\alpha) = 75° = 5.\frac{\pi}{12}$ and the anisotropy ratio $(\delta) = 2$.

Analysis of the results

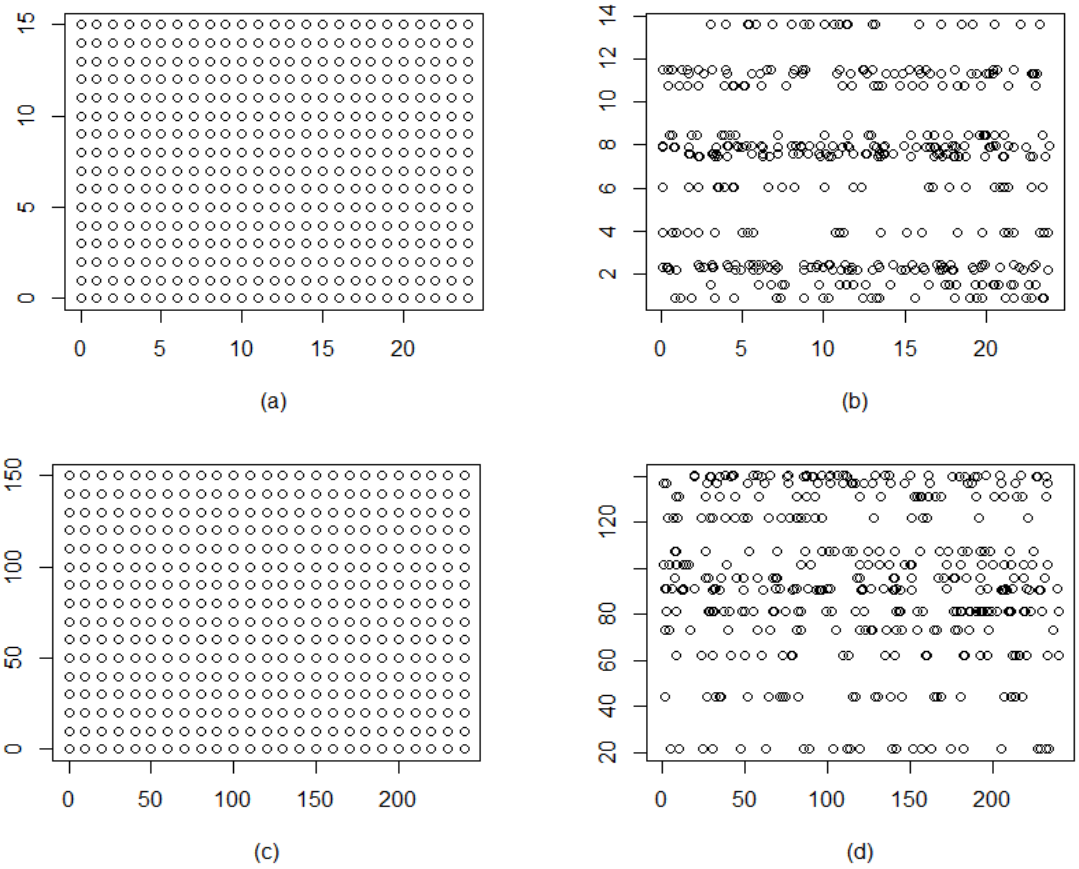
For the evaluation of the results, two approaches were taken to set smoothing effect (ν): 1) considering ν as a fixed and 2) as a random effect in the model (Equation 2). The initial value of ν in the model was assumed as 40% of its true simulated value. In order to understand advantages and disadvantages of each statistical package in the analysis of experimental data with spatial dependence effects, we employed both Geo-R (RIBEIRO Jr et al., 2020) and Asreml-R (BUTLER, 2022) for the simulation process.

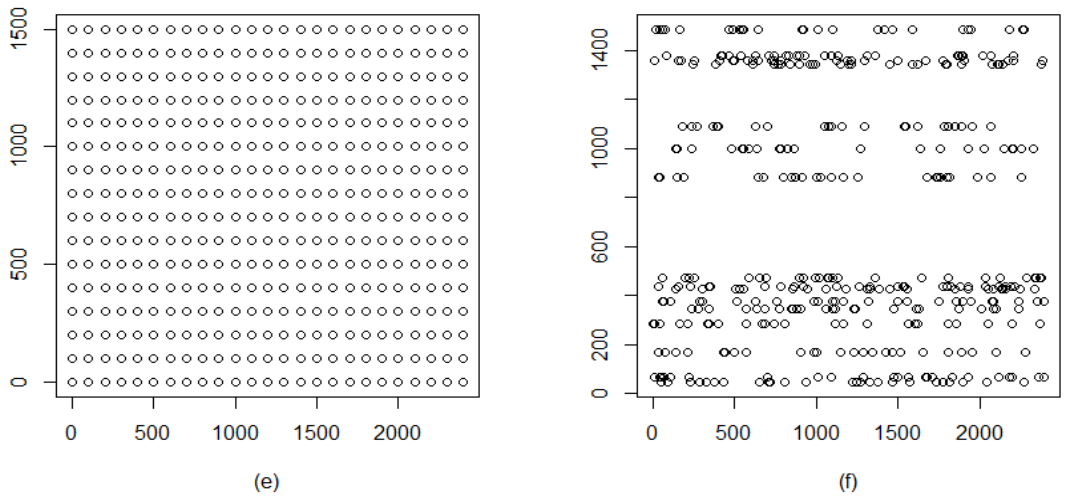
Both packages allow for using the restricted maximum likelihood (REML) estimation process, which was fixed to ensure comparison between generated results. However, in the GeoR package, the user needs to choose initial values to begin the process of estimating the model parameters using the *GeoR* package. In practice, this can be considered one of the bottlenecks in using it, since experience is needed to choose suitable values. On the other hand, the commercial *Asreml-R* package does not require this step, which at first appears to be a great advantage using the default values. However, as Asreml-R convergence can be difficult to obtain (HASKARD, 2007), the iterative process was extended to up to 100 iterations. Descriptive measures such as median, maximum, minimum, quartiles and inter-quartiles values of the estimated parameters and the converged analyses were used as criteria to assess the performance of each simulated scenario for each model and study. We also calculated the relative error (RE) for each estimated parameter in each case as ($\hat{\theta}$ was used as generic symbol): $E_\theta = \frac{|\hat{\theta}-\theta|}{\theta}$. Then,

we calculated RE as the maximum difference value between the random parameters for each fitted model. The complete model is:

$$RE\ (\ \hat{v},\ \hat{\phi}\ ,\hat{\delta},\hat{\alpha},\hat{\sigma_s}^2,\hat{\sigma}^2) = \max\{RE_v\ ,RE_\phi\ ,RE_\delta\ ,RE_\alpha,RE_{\sigma_s^2},RE_{\sigma^2}\ \}$$

Figure 1 - Regular ((a), (c) and (e)) and irregular ((b), (d) and (f)) simulated grids. (a) and (b) represents the 1×1 grid spacing and $\phi\ = 3$; (c) and (d) represents the $10\ X\ 10$ grid spacing and $\phi\ = 30$, and (e) and (f) represents the $100\ X\ 100$ grid spacing and $\phi\ = 300$.





RESULTS

WITHOUT REGULAR NUGGET EFFECT

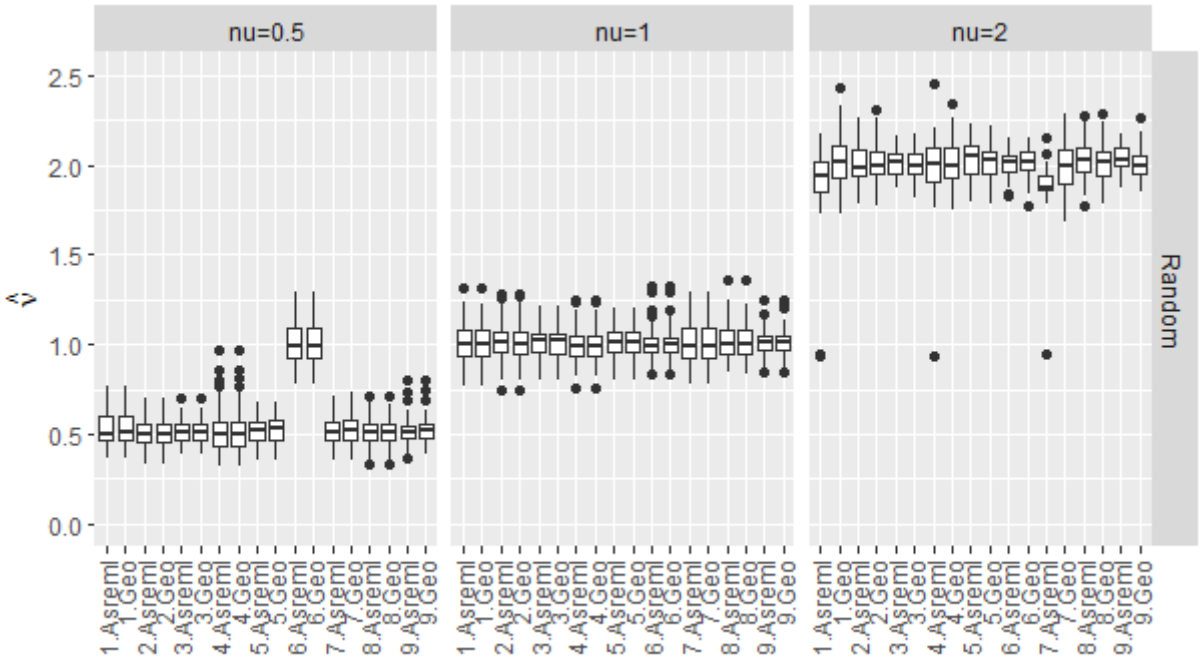
The different configurations, containing: the fixed effect smoothing parameter (ν), the software used (Asreml or GeoR), the convergence of the models and the estimated values for different variables, will be discussed below. In the analysis carried out considering a regular grid and the fixed-effect smoothing parameter $\nu = 0{,}5$ of the Asreml-R software, only the first three simulated scenarios converged. The convergence rate was 100% in the first two and 97% in the third scenario. However, using the GeoR software, the convergence rate was 100% for all scenarios. In both software, the median values of the estimated parameters were close to those simulated in all scenarios.

When analyzing the variance with regular grid and fixed effect smoothing parameter $\nu = 1$, it was observed that GeoR and Asreml-R showed similar behavior in terms of convergence rate. In scenario 3, both reached 100% convergence. However, when considering $\nu = 2$, only scenarios 1, 2 and 3 showed convergence, all reaching 100%, but only in the Asreml -R software. However, it is important to highlight that scenario 6 remained one of the worst results in terms of convergence in both models throughout the entire study, regardless of the presence or absence of the nugget effect.

This shows that, for the fixed effect, GeoR was unable to estimate the maximum intervals, that is, 300, 500 and 1000, thus showing that in the fixed effect $\hat{v}$ there is a direct influence on the irregularity of spatial fluctuations, that is, abrupt variations between points. When analyzing the random effect, it was observed that all scenarios presented results, regardless of $\hat{v}$.

In relation to the results obtained for the estimated value ($\hat{v}$) as shown in Figure 2, when adopting $v = 0.5$, it appears that, with the exception of scenario 6 in both models, all scenarios exhibited a median close to the expected value. However, it $\hat{v}$ overestimated the expected result in scenario 6 for both models. A joint analysis reveals that the best performing scenarios were 3 and, later, 8, for both models. When considering $v = 1$, it is noted that all scenarios in both models presented medians close to v, with scenarios 6 and 9 being the most prominent for both.

Figure 2: estimated values without nuggets for $\hat{v}$scenarios 1 to 9, considering regular mesh and modeled with random effect.

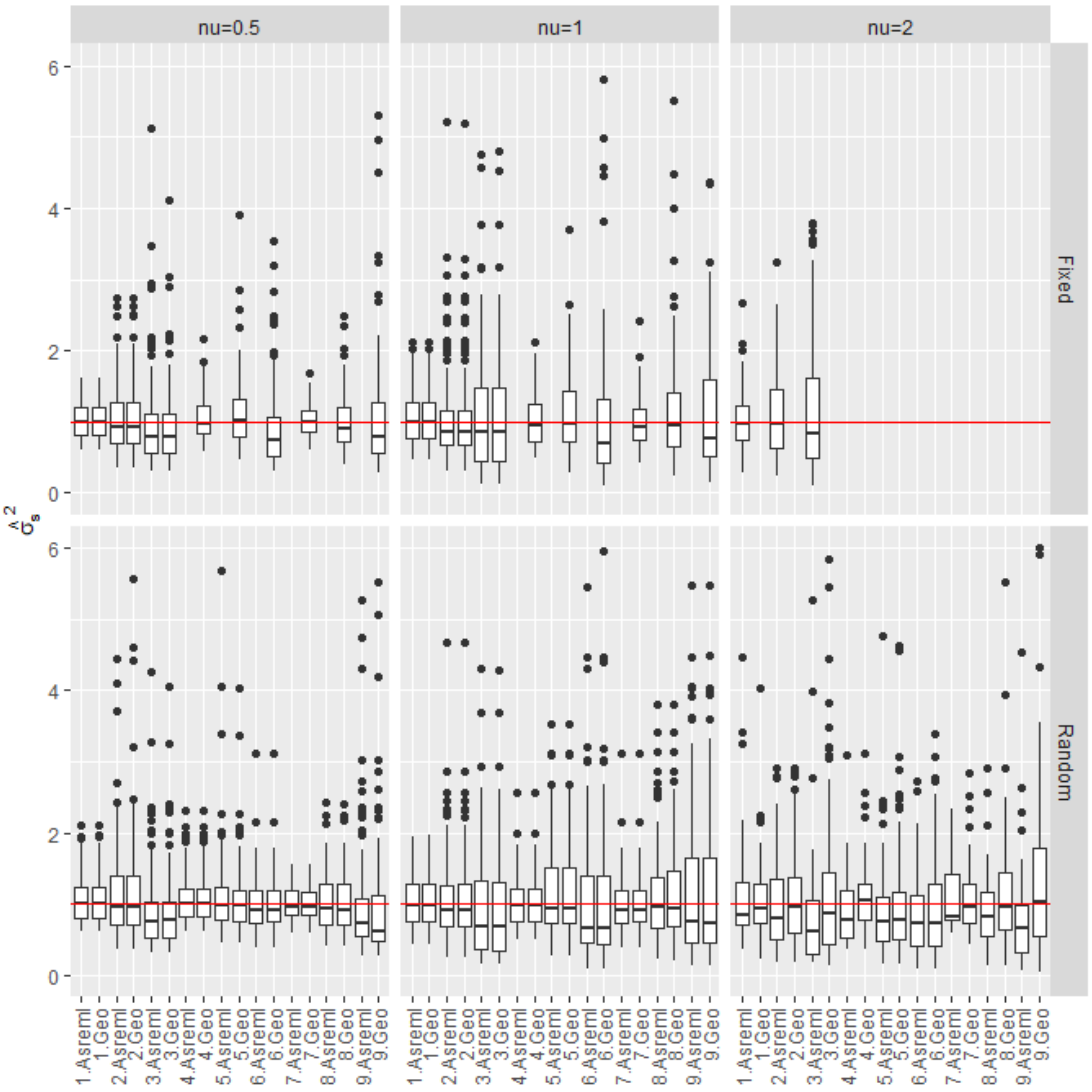


However, when estimating $v = 2$, there was a greater dispersion in the median results. Scenario 7 in Asreml-R proved to be the worst among all scenarios evaluated, showing a change in the behavior of the scenarios when comparing Asreml-R and GeoR. The best results were found in GeoR, highlighting scenarios 3 and 6 as the most favorable. However, when analyzing GeoR with $v = 0.5$, scenario 3 stands out as the best.

When related to variance, the lowest among all random and fixed models occurred in GeoR, in scenario 9, ($\widehat{\sigma}_s^2$ median = 0.65), in the fixed effect, scenario 9 presented $\widehat{\sigma}_s^2$ a little closer to the real (median = 0.85). Therefore, Asreml-R was unable to estimate a σ_s^2

in scenario 9 of the fixed effect, and in the random effect, it presented a median of $\hat{\sigma}_s^2$ = 0.81 (Figure 3).

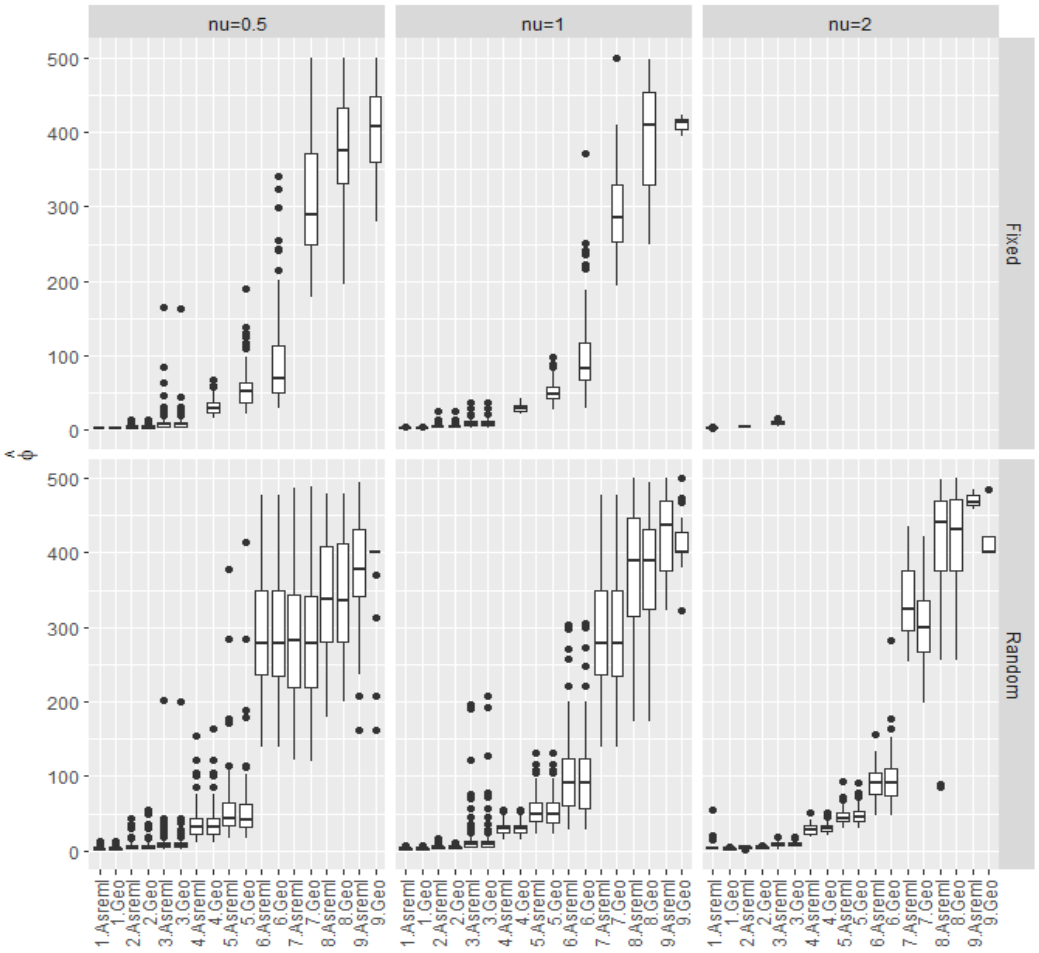
Figure 3: Estimated values without nuggets for $\hat{\sigma}_s^2$ scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



The convergence rate with $\nu = 1$ random effect increased in both software, remaining between 99% and 100%. However, there was an underestimation of the median $\hat{\sigma}$ in relation to σ (median = 1). Analyzing $\nu = 2$, the convergence rate was 100% in GeoR and ranged from 98% to 40% in Asreml-R, with 40% in the circumflex scenario, the median was the third lowest among the regular scenarios $\hat{\sigma}_s^2$(median = 0.66). However, in general, both software presented similar values to each $\hat{\sigma}_s^2$ other, in the scenarios in which there was convergence, except in the random scenario 9 with$\nu = 2$.

As for the intervals, the greatest difficulties in estimating the regular fixed effect were in the largest intervals of each set of scenarios, that is, ϕ= 10, 100 and 1000, and there was underestimation in the estimates (Figure 4).

Figure 4: Estimated values without nuggets for $\hat{\phi}$ scenarios 1 to 9, considering a regular grid and modeled with fixed and random effects.



This underestimation remained in all other fixed effect scenarios. With a random effect, $\hat{\phi}$ it increased at greater distances, except for $\hat{v} = 0{,}5$ GeoR, scenario 9, where the median $\hat{\phi}$ increased to 470.63m.

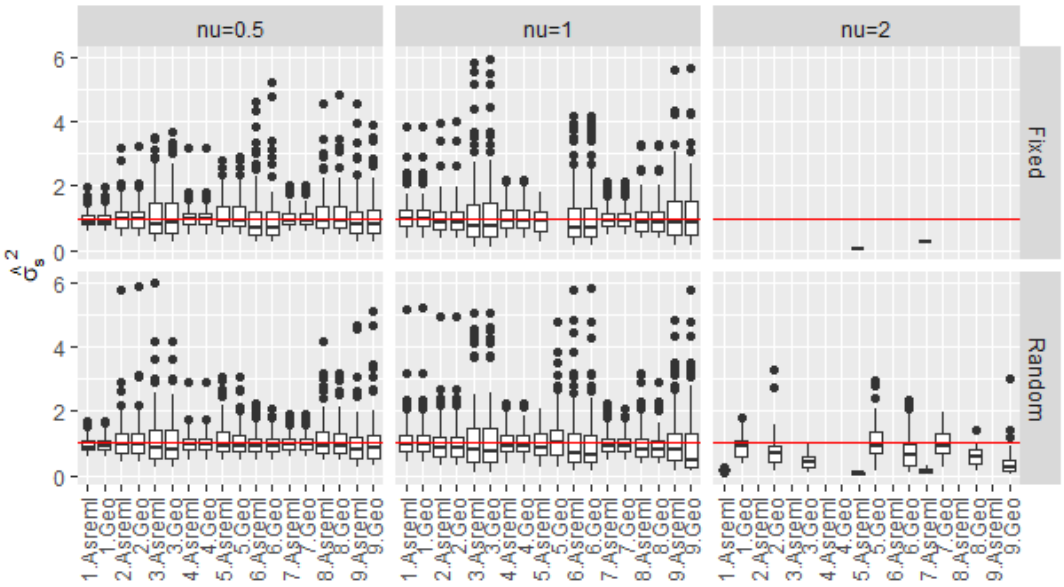
WITHOUT IRREGULAR NUGGET EFFECT

When analyzing the fixed irregular mesh model, $v = 0{,}5$it was observed that the convergence rates were consistent at 100% for most scenarios in both software (Asreml -R and GeoR), except for scenarios 3 and 6 of Asreml-R, which reached 99%. This irregular approach showed an improvement in convergence, especially notable in Asreml -R compared to the regular grid. There was a slight underestimation in the medians of the parameters $\hat{\sigma}_s^2$ (Figure 5). When analyzing the convergence rate with $v = 1$ the fixed effect, it was observed that all except scenario 5 of GeoR presented a convergence rate of 100%, different from that found in the regular mesh, where in this same scenario GeoR presented satisfactory results for the simulated parameters. .

However, when analyzing $v = 0.5$ both fixed and random contexts, the results in the generated scenarios showed very similar behaviors, reflecting consistency in both models. However, in relation to the limited dispersion of results around the median, it was observed that the estimated values $\hat{\sigma}_s^2$ were higher due to the random effect. The scenarios that achieved the best results in $\hat{\sigma}_s^2$ both models were identified in scenario 7 of the random effect. Next, in $v = 1$, there was a general tendency for superior performance in the fixed effect compared to the random effect, as the median was closer to the expected value. However, in this configuration, it was not possible to estimate $\hat{\sigma}_s^2$ GeoR scenario 5. In the random effect estimates in this configuration, the medians $\hat{\sigma}_s^2$ were, for the most part, underestimated. Scenario 7 was identified as the best in both the fixed and random effects, but in the context of the fixed effect, scenario 7 proved to be superior to the random effect.

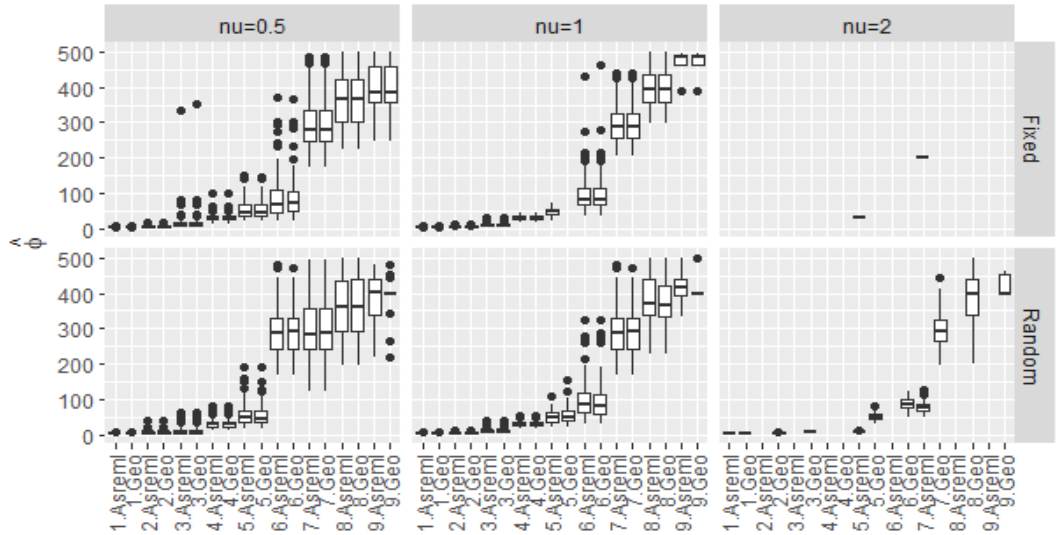
Finally, when considering $v = 2$, it was found that in the fixed effect, the results were the least accurate in estimating σ_s^2, generating values only in scenarios 5 and 7 of Asreml-R. However, in the random effect in the same configuration as Geo-R , unlike Asreml-R, it managed to estimate most of the results, although in most estimates the median was underestimated. These estimates were identified as the second least accurate in $\hat{\sigma}_s^2$ the irregular grid, while the best results were found in $v = 0.5$ scenario 7 of the random effect in both models.

Figure 5: Estimated values without nuggets from $\hat{\sigma}_s^2$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



When analyzing $\hat{\phi}$ (Figure 6) for Asreml-R, closer results were noticed when compared to the regular grid. When examining $\nu = 0.5$, from the irregular grid, it was noticed that, in terms of median, the scenarios with a smaller range behaved more satisfactorily than those with a greater range, which underestimated the ϕ, both in the fixed and random contexts.

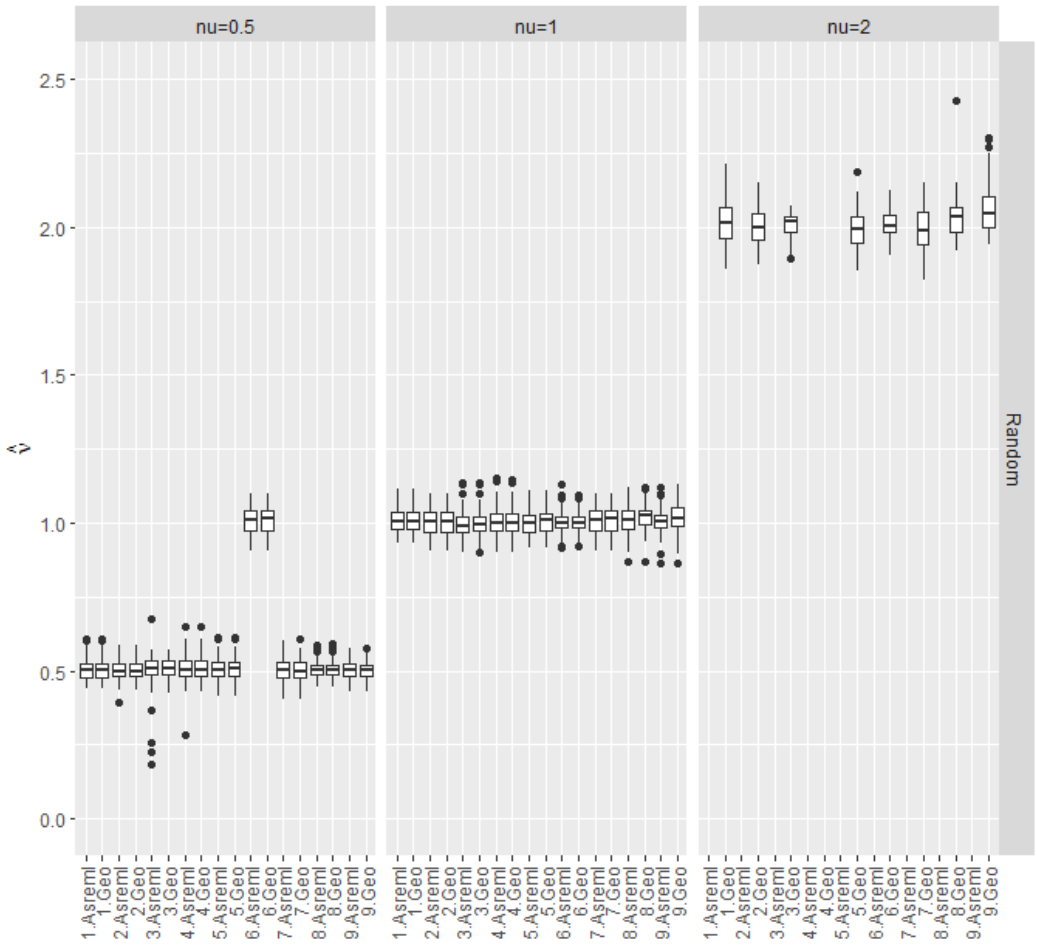
Figure 6: Estimated values without nuggets for $\hat{\phi}$scenarios 1 to 9, considering an irregular grid and modeled with fixed and random effects.



However, in the exponential model, the fixed effect revealed more consistent results for greater distances compared to the random effect. When considering $v = 1$ the fixed effect, despite the lack of estimate in GeoR scenario 5, there was a general tendency for results closer to expected for the estimated values of $\hat{\phi}$. However, with $v = 2$, both in the fixed and random contexts, the results of $\hat{\phi}$ were less satisfactory.

In the analysis of the random effect with $v = 0{,}5$ the irregular grid, a behavior similar to that of the regular grid was observed, except in Asreml-R scenario 1, which recorded a significant increase in the convergence rate, going from 3% to 99% (Figure 7). Furthermore, the pattern of overestimation in scenario 6 of both models with $v = 0{,}5$. Although all other scenarios had medians very close to v, the best results for both models were found in scenario 2.

Figure 7: Estimated values without nugget for $\hat{v}$scenarios 1 to 9, considering irregular mesh and modeled with random effect.

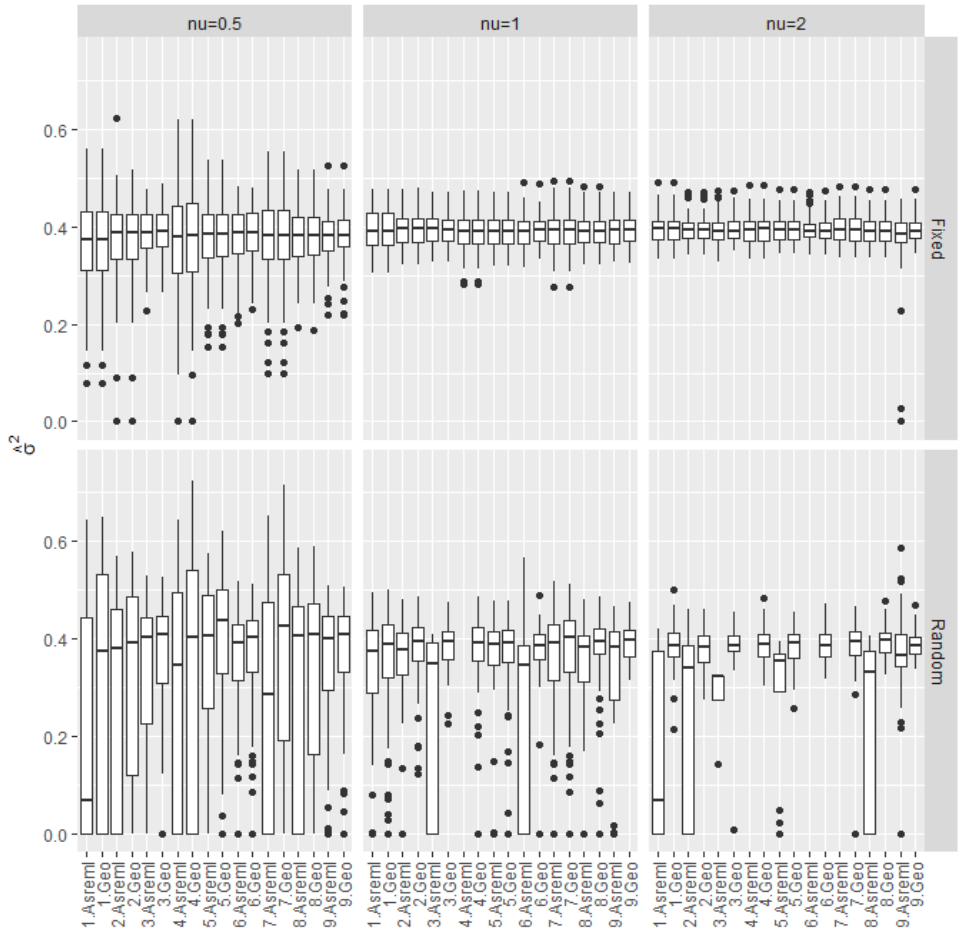


When evaluating $v = 1$, all scenarios showed medians very close to the expected value; however, due to homogeneity and the lower presence of outliers, the best results were observed in scenario 1 for both models. When analyzed $v = 2$, a greater variation in results in relation to the expected value is noticed v, with most scenarios presenting overestimations. In Asreml-R, in this configuration, no results were generated for any of the scenarios, making the scenarios with the smallest discrepancy both in $v = 0{,}5$ and in $v = 1$ the locations where the best results were obtained.

WITH REGULAR NUGGET EFFECT

When examining the nugget effect models, it was observed that with $v = 0{,}5$, GeoR achieved a convergence rate of 100%, while Asreml-R varied between 97% and 100% in different scenarios, in this configuration, it varied at the median of 0.37 to 0.39 (Figure 8), thus being very close to σ^2(= 0.4). However, when $v = 0.5$, analyzing $\hat{\sigma}^2$ in the fixed effect, a wide variation was observed in relation to the medians in the scenarios, with scenario 9 being the one with the smallest variation. For $v = 1$, the dispersion of the results $\hat{\sigma}^2$ was more consistent in relation to the median, proving to be very similar in all scenarios. On the other hand, when using $v = 2$ for $\hat{\sigma}^2$, a greater presence of outliers was observed, although the data were even more homogeneous than in $v = 1$.

Figure 8: Estimated Nugget values for $\hat{\sigma}^2$scenarios 1 to 9, considering regular mesh and modeled with fixed and random effects.



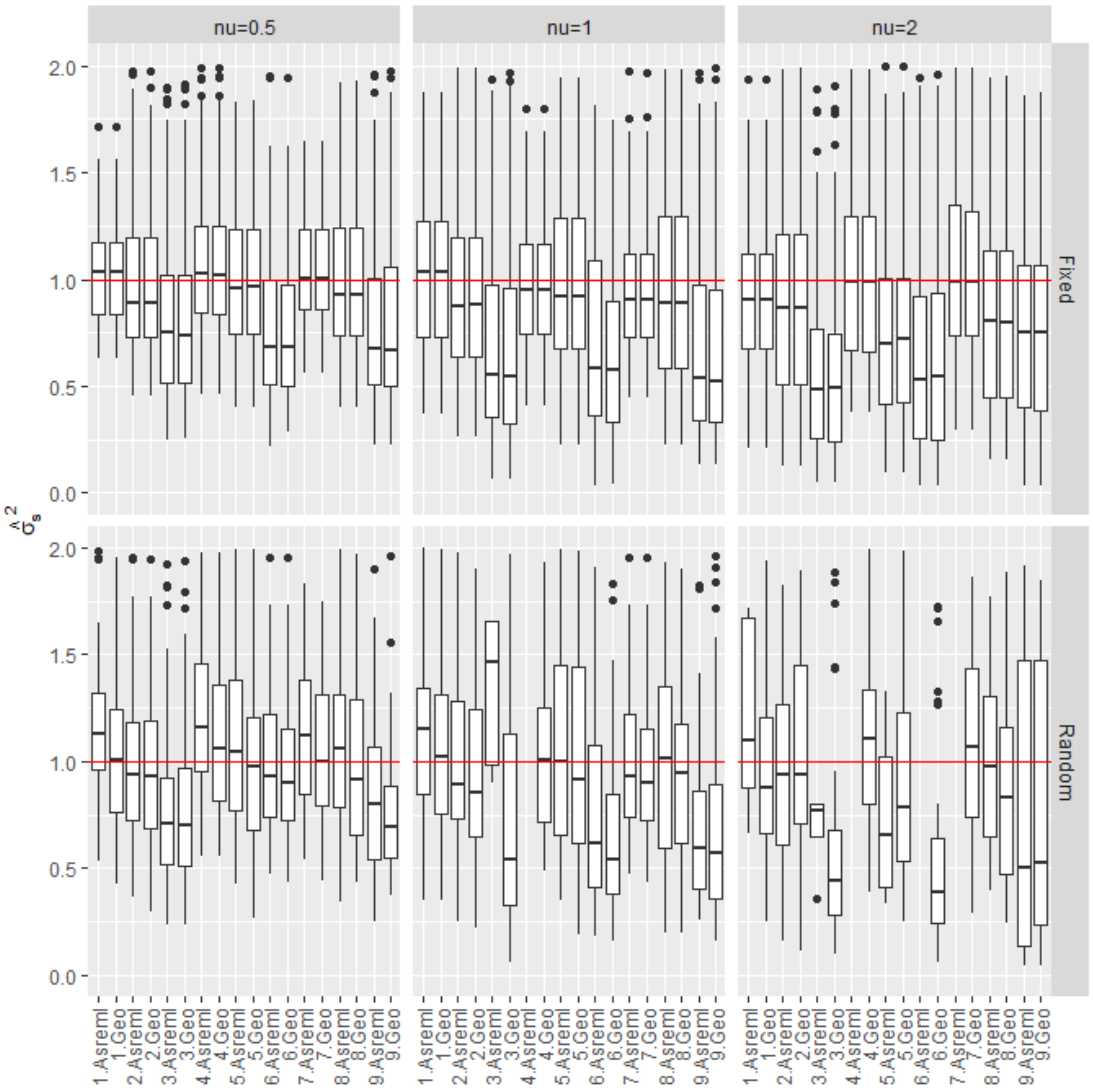
In the context of random effects, there was greater heterogeneity in the results in general when compared to fixed ones. With $v = 0.5$, GeoR demonstrated, in general, a superior performance to Asreml-R, although both presented $\hat{\sigma}^2$ very close to zero. Scenario 6 stood out as the best in both models. With $v = 1$, the estimate $\hat{\sigma}^2$, the heterogeneity of the results decreased in relation to $v = 0.5$, however, in this configuration, it was not possible to obtain scenario 4 of Asreml-R. This difficulty increased in $v = 2$, especially in the Asreml-R model, presenting dispersed values in relation to what was expected.

In general, the estimates of $\hat{\sigma}_s^2$ were considerably closer to those of σ_s^2 (Figure 9), while there was a tendency of underestimation in the estimates of $\hat{\phi}$ (Figure 10), even though they were better than in the models without the nugget effect. When analyzing $\hat{\sigma}_s^2$ with $v = 0.5$, it was noticed, in the fixed effect, a greater dispersion in the results compared to the exponential random effect. In scenarios 1, 4 and 7, there was a slight overestimation of the results, while in the others there was an underestimation. However,

in the random effect, despite significant fluctuations in the results in relation to the medians, in GeoR scenario 7, this was close to σ_s^2.

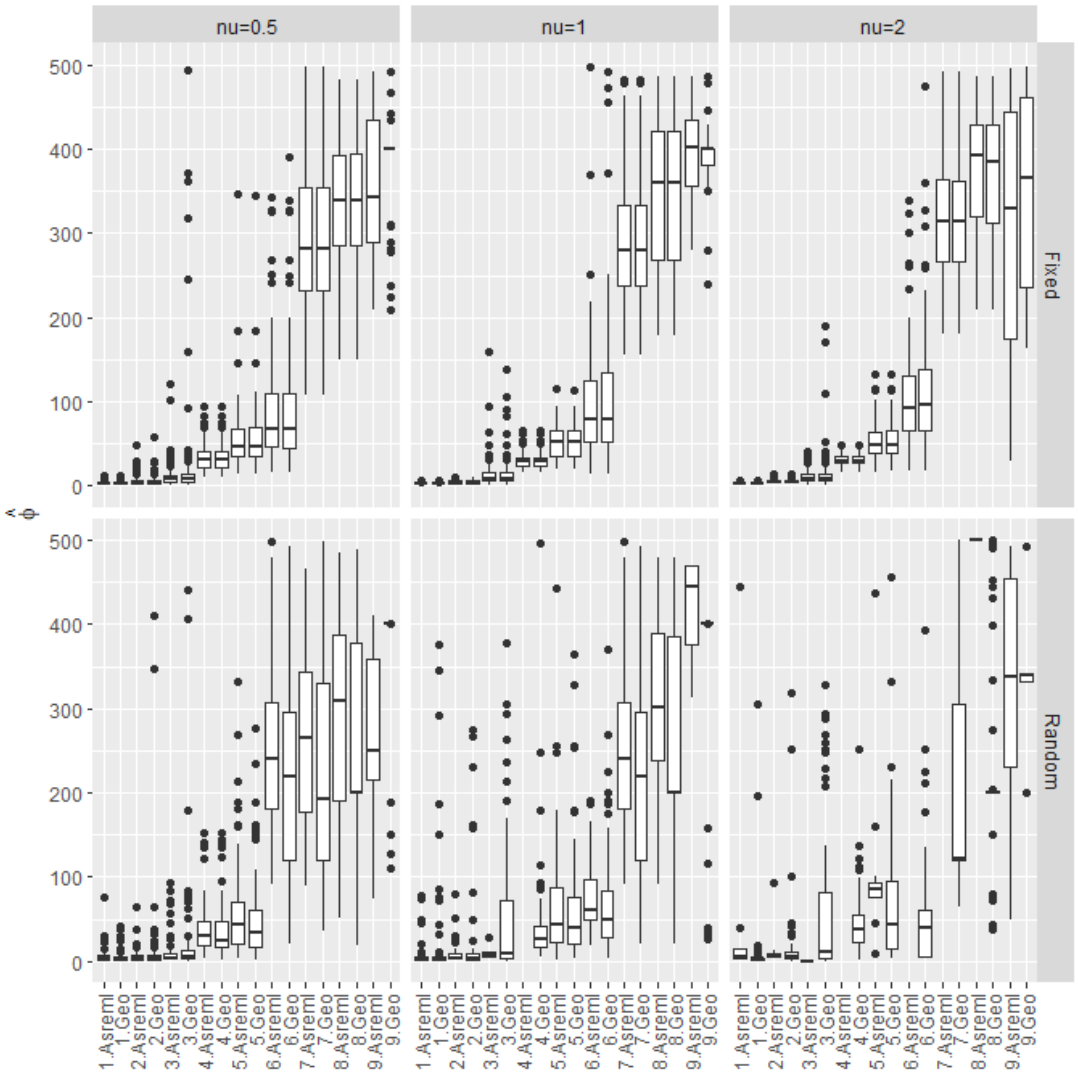
With $v = 1$, both in fixed and random effects, there was an increase in the variability of the medians of $\hat{\sigma}_s^2$. However, in the fixed effects, most scenarios showed underestimation, while in the random ones there was no clear pattern, although Asreml-R's random scenario 5 came closer to σ_s^2. When analyzing $v = 2$, even though there were significant disparities between the results of the scenarios of both models, in scenarios 4 and 7 of the fixed effect, the median was equal to σ_s^2. However, in the random effect, in addition to the inability to estimate some Asreml-R scenarios, the worst results were observed.

Figure 9: Estimated values with nugget for $\hat{\sigma}_s^2$scenarios 1 to 9, considering regular mesh and modeled with fixed and random effects.



When analyzing the estimates $\hat{\phi}$(Figure 10), the tendency of underestimation in the largest ranges is once again noticeable, while the more precise estimate is closer to the smallest ranges. It is important to highlight that, regardless of the values analyzed v , the estimates $\hat{\phi}$ for the fixed effect are more precise than those for the random effect. Additionally, there was an inability to generate models in Asreml-R with $v = 2$.

Figure 10: Nugget estimated values for $\hat{\phi}$scenarios 1 to 9, considering a regular mesh and modeled with fixed and random effects.



In $v = 1$, GeoR achieved a convergence rate of 100% in all scenarios, while Asreml-R varied between 96% and 100%, except in scenario 6 (98%) and scenario 9 (96%). There was greater variation $\hat{\sigma}_s^2$ compared to $v = 0{,}5$, with the lowest values in

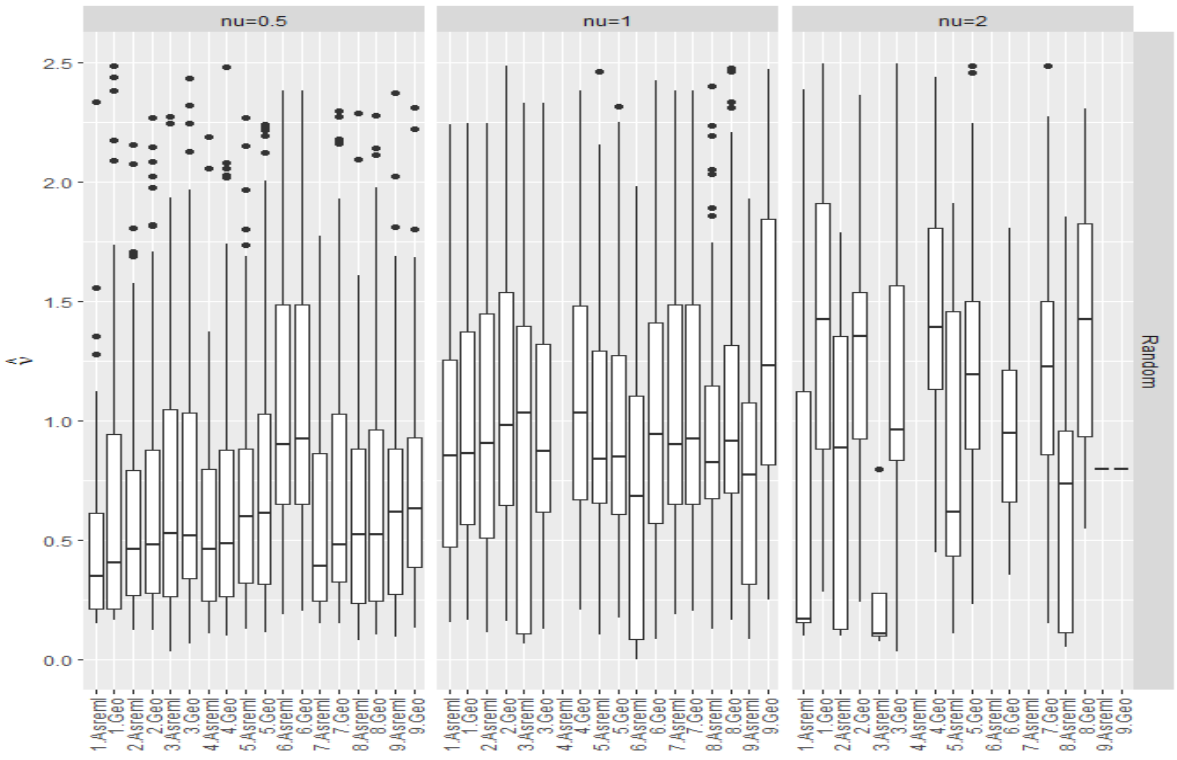
scenario 9 in both softwares. The median of the nugget effect remained between 0.39 and 0.40, and the reaches increased by approximately $v = 0{,}5$.

In the fixed effect with $v = 2$, GeoR continued with 100% convergence, while Asreml-R obtained 99% in scenarios 3 and 6 and 98% in scenario 5, remaining 100% in the others. In this configuration, the variance was overestimated, being more pronounced than in the others v, and the nugget effect remained at 0.39 at the median, with the intervals underestimated.

As for the random effect with $v = 0{,}5$ GeoR, maintained a 100% convergence rate across all scenarios, while Asreml-R ranged from 81% to 99%, with the lowest rate in scenario 9 (81%) (Figure 11).

Analyzing the estimates of v, the results were scattered and none of the configurations adopted achieved median values $\hat{v}$ similar to those v expected. In $v = 0.5$, scenarios 3 and 8 they were closer to v expected, while in scenarios with $v = 1$, scenario 2 with GeoR was the closest to v. On the other hand, in $v = 2$, in addition to the inability to generate some scenarios in Asreml, the largest underestimations in the results were observed.

Figure 11: estimated values with nugget for $\hat{v}$scenarios 1 to 9, considering regular mesh and modeled with random effect.

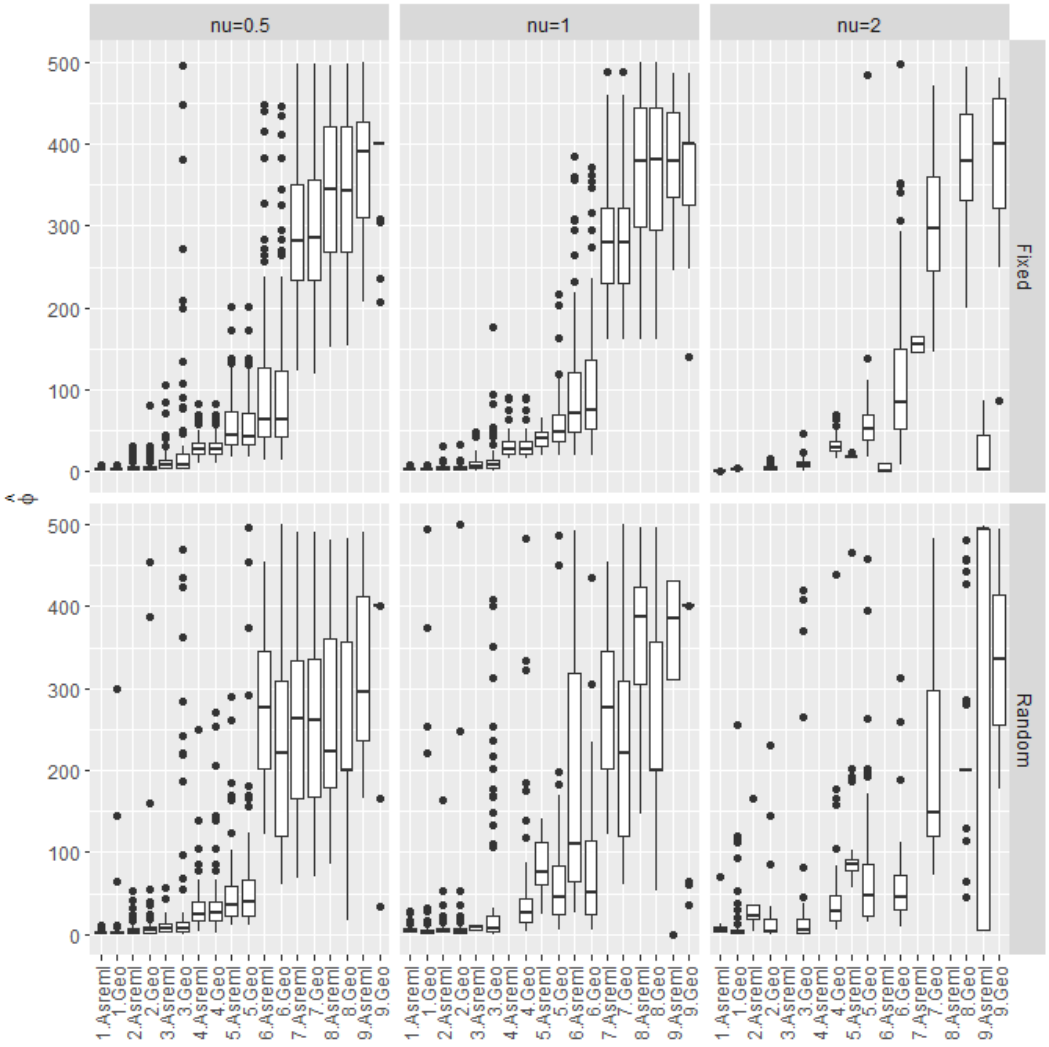


WITH IRREGULAR NUGGET EFFECT

Analyzing the irregular grid with fixed effect and $v = 0{,}5$, we observed that GeoR achieved a convergence of 100%, while Asreml-R varied between 95% and 100%, with the lowest rate in scenario 9. Both models underestimated $\hat{\phi}$ (Figure 12) and $\hat{\sigma}_s^2$ (Figure 13) in the nine scenarios, and the nugget effect fluctuated between 0.38 and 0.4 on the median, thus being close to the simulated nugget effect.

When analyzing $\hat{\phi}$, it is noted that, with $v = 0.5$, the fixed effect, despite underestimating the expected results, exhibited a behavior closer to ϕ. At greater distances, the random effect presented significantly lower medians than the fixed effect. When $v = 1$, although scenarios with greater distances increased the range, it still remained below that ϕ for $\hat{\phi}$. In the fixed effect with $v = 2$, most Asreml-R models could not be estimated; however, in the generated scenarios, the values $\hat{\phi}$ increased slightly. However, the same heterogeneous pattern found in the random effect with $v = 0.5$ also manifested itself with $v = 1$ and $v = 2$. In scenario 9 with $v = 2$, the median reached higher results for reach, but the results in general were highly dispersed, ranging from 0 to 500.

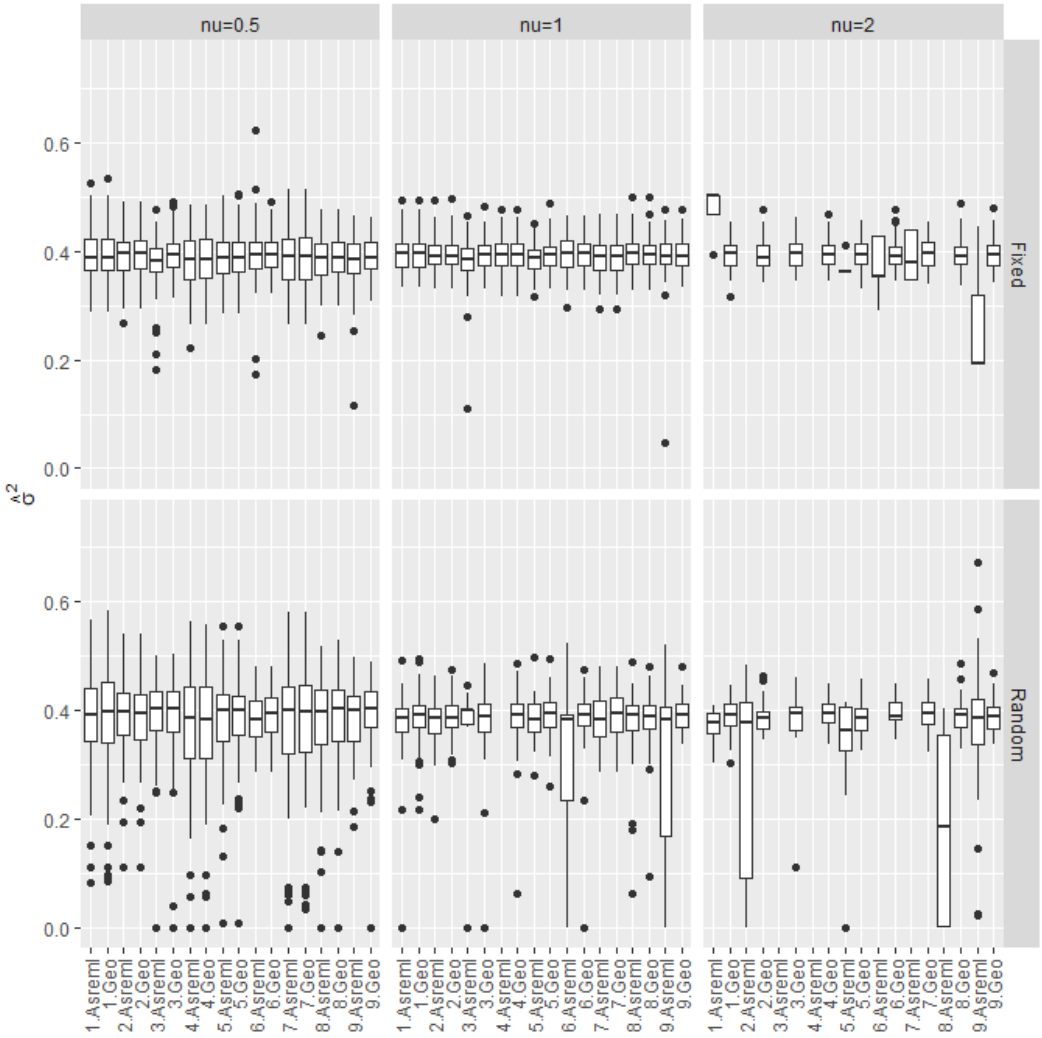
Figure 12: Nugget estimated values for $\hat{\phi}$scenarios 1 to 9, considering an irregular mesh and modeled with fixed and random effects.



Regarding the $\hat{\sigma}^2$ fixed effect, $v = 0.5$, it revealed results with averages close to σ^2, with GeoR scenario 3 being the most homogeneous among the data (Figure 13). When analyzing $v = 1$, we observe homogeneity in the results of each scenario in both models, greater than in $v = 0.5$, and in general, as in scenarios with $v = 0.5$, there is greater dispersion of data in relation to the median in the smaller ranges. In the case of $v = 2$a fixed effect, some results were not generated in Asreml-R, showing greater difficulty in estimating the nugget effect compared to GeoR. In the random effect with $v = 0.5$, the medians are $\hat{\sigma}^2$ close to σ^2, however, compared to the fixed effect, there is greater heterogeneity in the results generated. This same heterogeneity is observed in $v = 1$. Just

like in the regular grid with the nugget effect, in the irregular grid in the random effect, Asreml-R was not able to estimate in some configurations.

Figure 13: Estimated Nugget values for $\hat{\sigma}^2$scenarios 1 to 9, considering irregular mesh and modeled with fixed and random effects.



With $v = 1$, except in scenario 1, both models underestimated $\hat{\sigma}_s^2$(Figure 14). The $\hat{\sigma}^2$ remained in the range $v = 0{,}5$ and the ranges increased, especially in scenario 9, which stood out as the best range among the scenarios generated in both grids, despite still being underestimated. With $v = 2$, GeoR maintained a 100% convergence rate, but Asreml-R varied, not reaching 100% in some scenarios, such as 4 and 8 (no convergence) and 7% in 9. Differences in median values between $\hat{\sigma}_s^2$ software were observed, predominantly underestimating the observed values, with emphasis on Asreml-R scenario 6, which recorded the lowest value $\hat{\sigma}_s^2$ of 0.05. Analyzing the random effect, in $v = 0{,}5$,

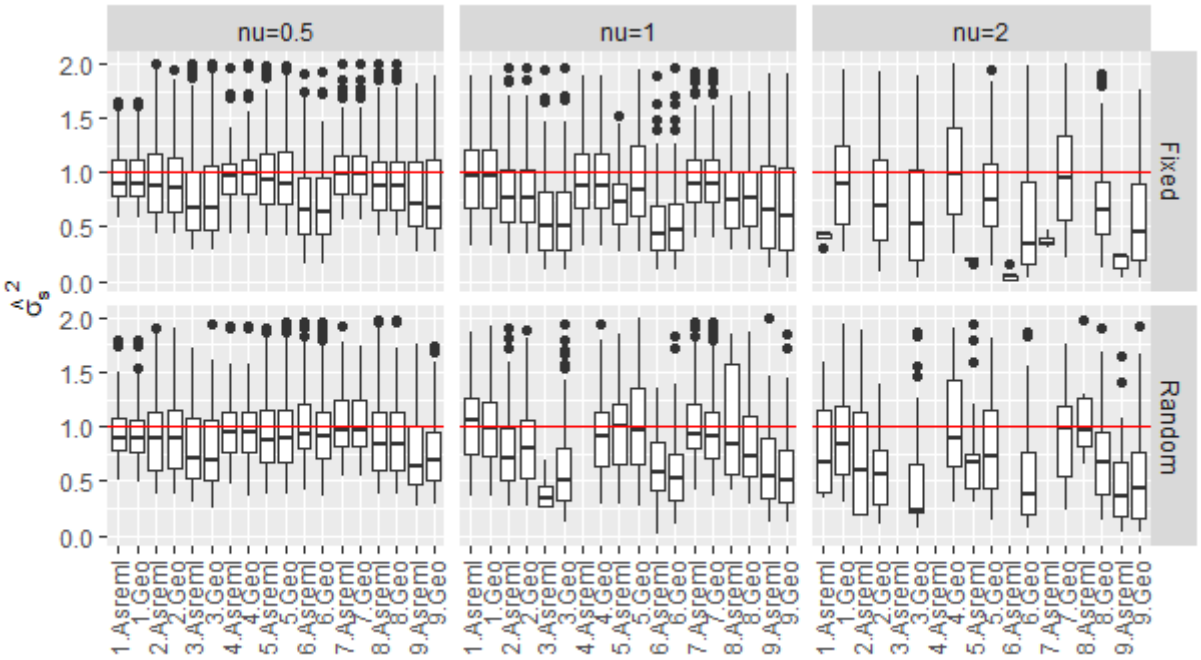
GeoR showed 99% convergence in scenario 1, while Asreml-R varied between 83% and 99%, with the lowest rate in scenario 9. In all scenarios, there was an underestimation of $\hat{\sigma}_s^2$.

However, when analyzing $v = 0.5$ the fixed effect, it is observed that, in general, all scenarios showed an underestimation of the median $\hat{\sigma}_s^2$, with the exception of scenario 7. In the same context with a random effect, although scenario 7 came considerably closer to $\sigma_s^2$the median, all scenarios underestimated σ_s^2 (Figure 14).

When analyzing $v = 1$, despite the underestimations, scenario 1 in the fixed effect was the closest to σ_s^2, while in the random effect, both scenario 1 and scenario 5, respectively from GeoR and Asreml-R, presented medians close to the expected value of σ_s^2, but with greater heterogeneity in the results of each scenario.

Finally, when analyzing $v = 2$, both fixed and random effects, there were difficulties in estimating using Asreml-R. In the fixed effect, scenario 5, and in the random effect, GeoR scenario 7, were those that presented median results closest to σ_s^2.

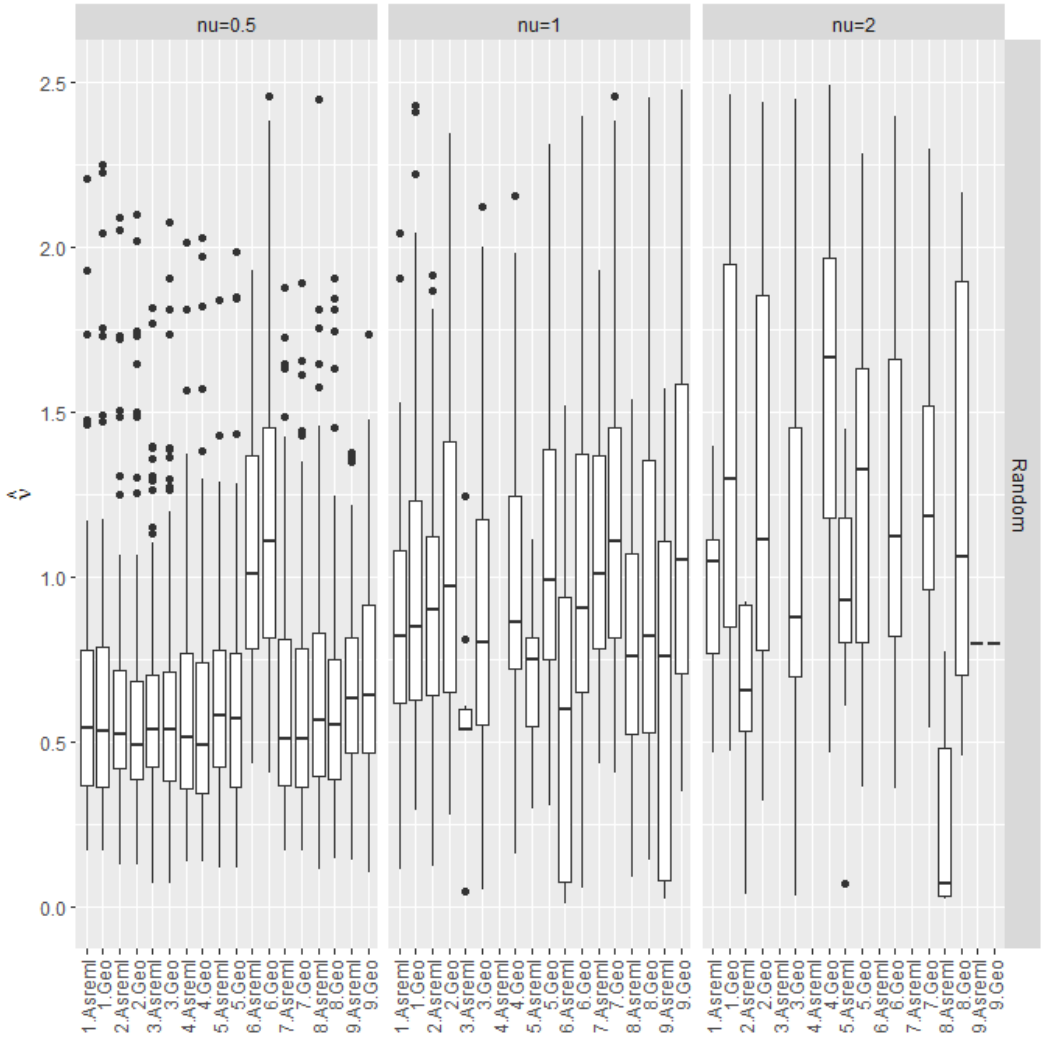
Figure 14: Nugget estimated values for $\hat{\sigma}_s^2$scenarios 1 to 9, considering irregular mesh and modeled with fixed and random effects.



In the analysis of v (Figure 15), it is noted that, with $v = 0.5$, the estimates were distant, especially for scenario 6, as was the case with the other values of v in previous configurations. The best estimates were observed, in both models, in scenarios 3 and 7.

Furthermore, GeoR scenario 2 also presented median results close to v. With $v = 2$, there is greater difficulty in estimating compared to $v = 0.5$; in most models, the results were underestimated, with GeoR scenario 5 being the one with the best estimate of v. For $v = 2$, the underestimates were even more pronounced than in $v = 1$, since none of the scenarios generated in any of the models were able to reach the median expected result.

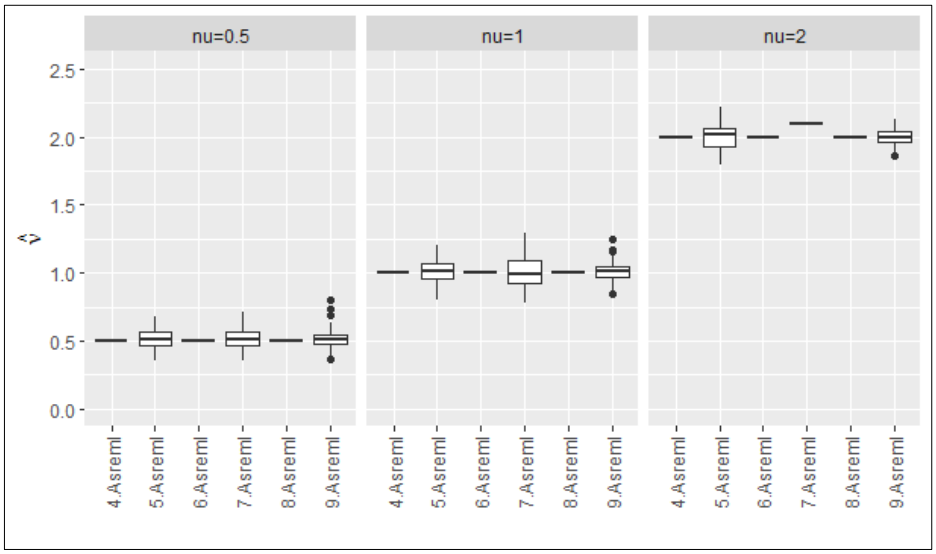
Figure 15: estimated values with nugget for $\hat{v}$scenarios 1 to 9, considering irregular mesh and modeled with random effect.



RESCHEDULING WITHOUT EFFECT REGULAR NUGGET

After applying rescaling, the models without a regular nugget effect, but with a random effect, are shown in Figure 16. In general, the models obtained medians coinciding with the indicated values, except in the cases of Asreml-R 5 and 7 models.

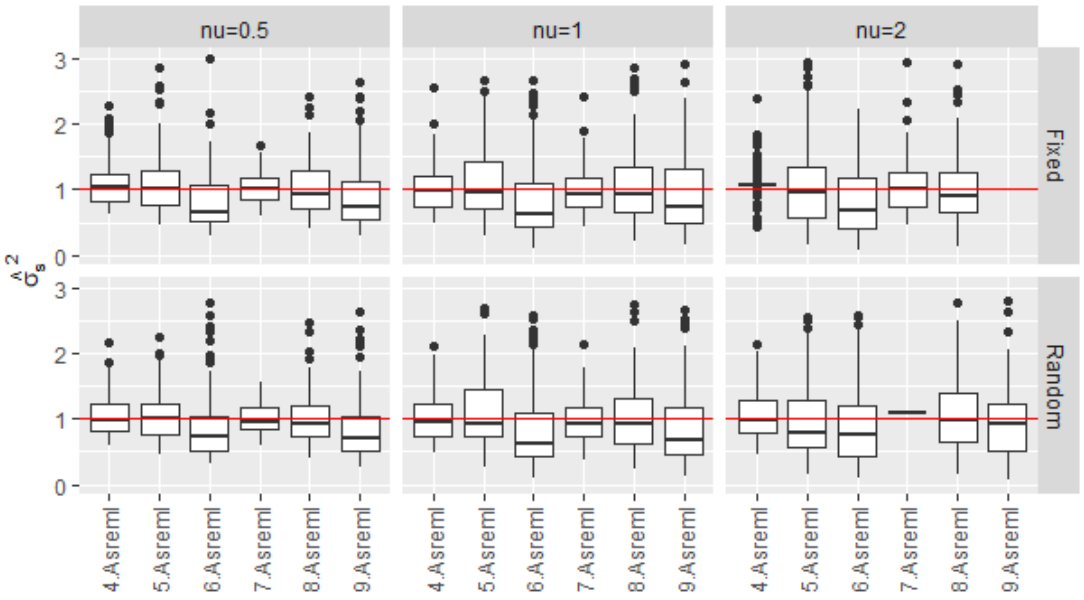
Figure 16: Estimated rescaling values, without nuggets for $\hat{v}$scenarios 1 to 9, considering a regular grid and modeled with fixed and random effects.



These models exhibited problems in estimation before rescaling, however, even with an underestimation of the median, after rescaling, they came closer to the simulated observed values. These characteristics of reducing underestimation of the median were observed in the parameter $\hat{v} = 0.5\ e\ 1$, for all scenarios tested. With $\hat{v} = 1$, all models showed a convergence rate of 100%. On the other hand, when observing $\hat{v} = 1$, the worst convergence rate was observed, reaching 57% in scenario 7 and 59% in scenario 9. However, when compared to data without rescaling, there was a general increase in the convergence rate in all models.

When analyzing the estimated values $\hat{\sigma}_s^2$ after rescaling (see Figure 17), it was observed that, with $\hat{v} = 0.5$, scenarios 5 and 7, considered fixed effects, and scenarios 4 and 5, random effects, approached the expected medians ($\sigma_s^2 = 1$). On the other hand, the other scenarios, still using the exponential model, showed an underestimation of the medians, except scenario 4, with a fixed effect, which overestimated the median. The most significant underestimations were evident in the most comprehensive scenarios, that is, in scenarios 6 and 9.

Figure 17: Estimated rescaling values without nuggets$\hat{\sigma}_s^2$ scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



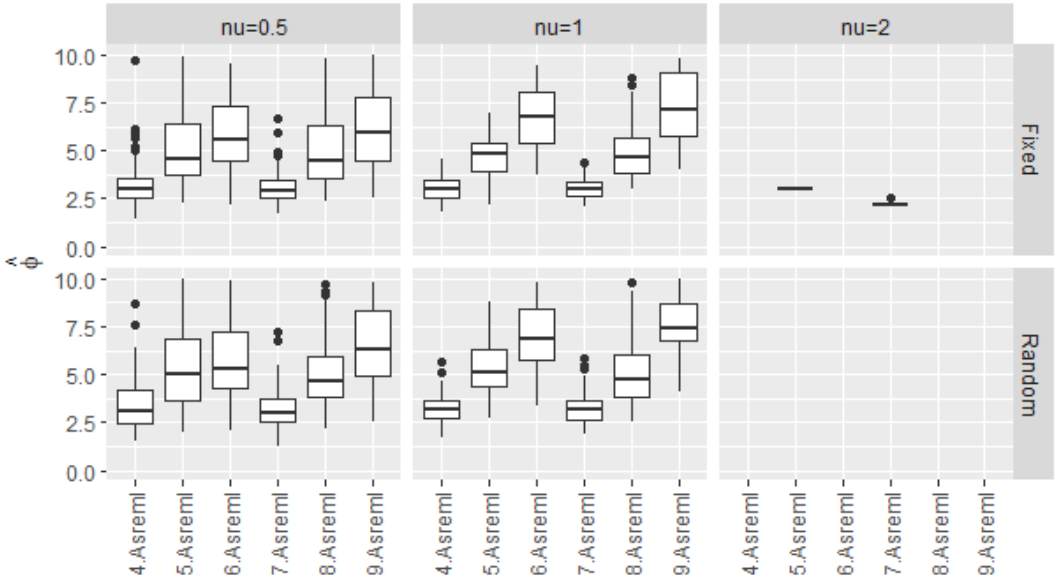
When adopting $\hat{v} = 1$, all scenarios presented underestimation of the medians of $\hat{\sigma}_s^2$, except the lowest-range scenario (scenario 4), where in both the fixed and random effects, the median was exactly equal to 1. Subsequently, when considering $\hat{v} = 2$, it was observed that only the scenario 7, with a fixed effect, presented $\sigma_s^2 = 1$, the same occurred in the random effect, scenarios 1 and 8. In the fixed effect context, the model was not able to estimate scenario 9 and demonstrated overestimation in the median of scenario 4. This underestimation did not random effect context was also observed, but in scenario 7.

When comparing the rescaled results (Figure 17) with the data in Figure 3, a substantial gain can be seen in the scenarios, especially in the fixed effect scenarios. In Figure 3, without rescaling, estimates were not possible for any scenario from 4 to 9. However, with rescaling, only fixed-effect scenario 9 could not be estimated. In relation to the random effect scenarios, although the range of the medians had $\hat{\sigma}_s^2$ been reduced, the same problems of overestimation and underestimation as presented in the scenarios represented in Figure 3 persisted.

After rescaling, the ranges were now 3, 5 and 10 meters (Figure 18). It is observed that scenarios 4 and 7 of each smoothing effect reached median values very close to expected ($\hat{\phi} = 3$), with scenarios 7 being more homogeneous in both effects compared to scenarios 4. Next, when analyzing scenarios 5 and 8, it should be noted that both in the fixed and random effects there was an underestimation of the median of $\hat{\phi}$. However, in

the fixed effect, these values, overall, were closer to the median of the simulated range ($\phi = 5$) than there was without a random effect.

Figure 18: Estimated rescaling values without nuggets$\hat{\phi}$ scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



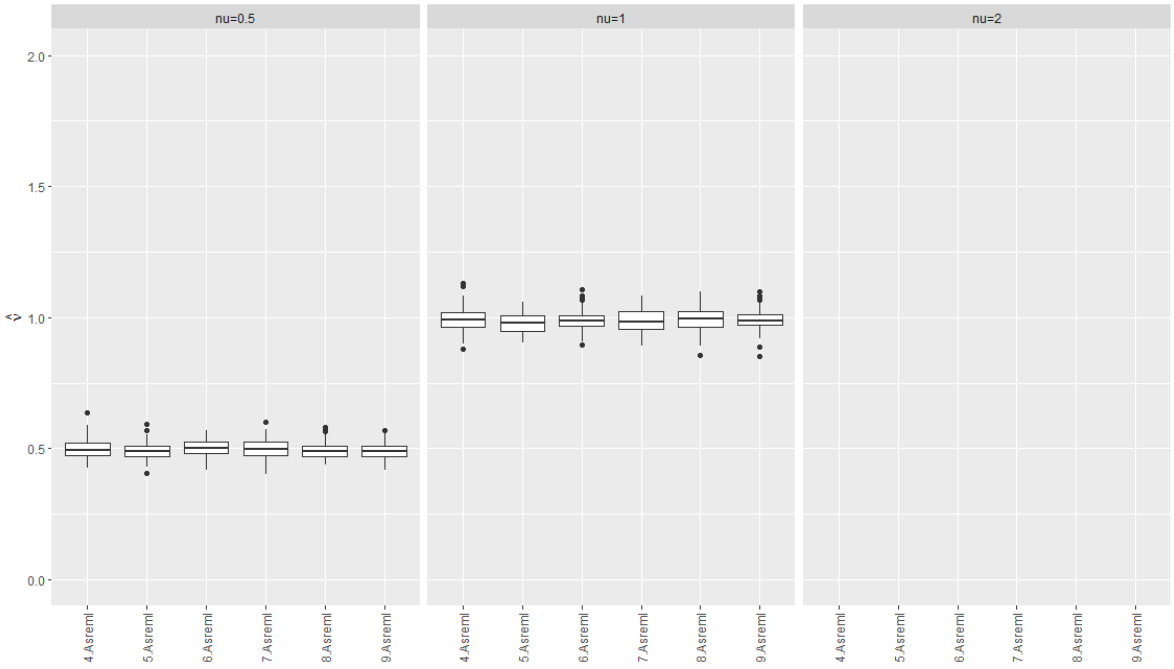
Finally, in relation to the $\hat{\phi}$ larger rescaled ($\phi = 10$) presented in scenarios 6 and 9, it is observed that these results demonstrated the worst underestimated results, in all smoothing parameters considered, in both effects. When comparing the results after rescaling $\hat{\phi}$ with those prior to it (as represented in Figure 4), not only the progress in scenarios with fixed effect convergence rates stands out, but also the proximity of the $\hat{\phi}$ for all smoothing parameters in relative to their medians. In the context of the random effect, it is shown that scenarios 4 and 7 showed much more detailed results of the medians $\hat{\phi}$ after rescaling. However, in the largest $\hat{\phi}$, both in rescaled and non-rescaled results, there was an underestimation in ϕ.

RESCHEDULING WITHOUT EFFECT IRREGULAR NUGGET

When analyzing the rescaled results, without the nugget effect, not randomly irregular, the $\hat{v} = 2$ was unable to estimate any of the scenarios (Figure 19). In contrast, the non-rescheduled results (as shown in Figure 7) showed results for scenarios 1, 5 and

7, but these results are considerably far from the reality of the simulated data, being extremely overestimated.

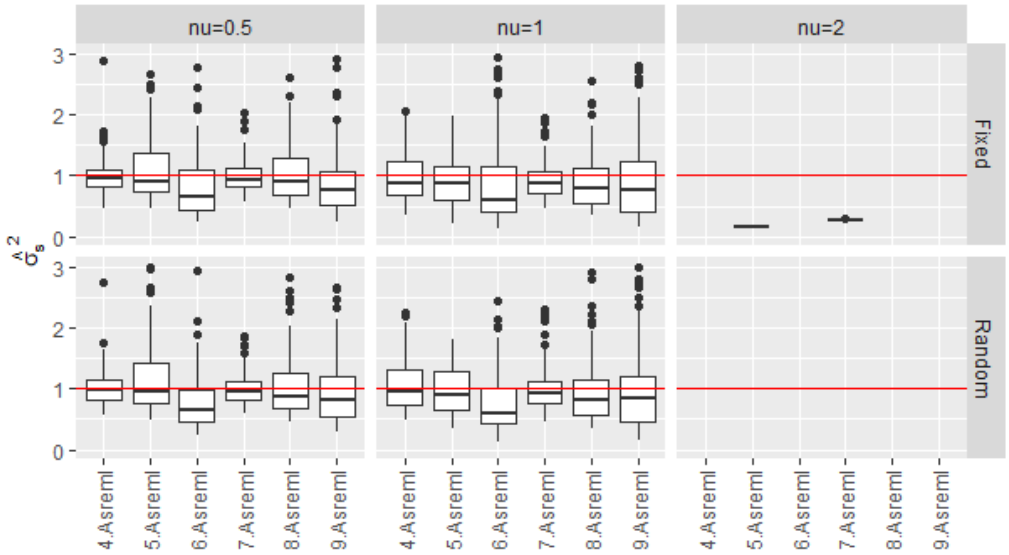
Figure 19: Estimated rescaling values without nuggets for $\hat{v}$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



However, during the rescaling the median of the generated scenarios changed these values in $\hat{v} = 0.5\ e\ 1$. Note that this median did not change significantly compared to the results without rescaling (Figure 7), however, there was a reduction in relation to outliers and the convergence rate increased considerably.

The results $\hat{\sigma}_s^2$ indicate an underestimation in all scenarios, as illustrated in Figure 20, in relation to the simulated variance ($\sigma_s^2 = 1$). In the fixed effect scenarios, where the results were observed only in scenarios 5 and 7, more pronounced underestimations were identified, which were similar to the results obtained without rescaling.

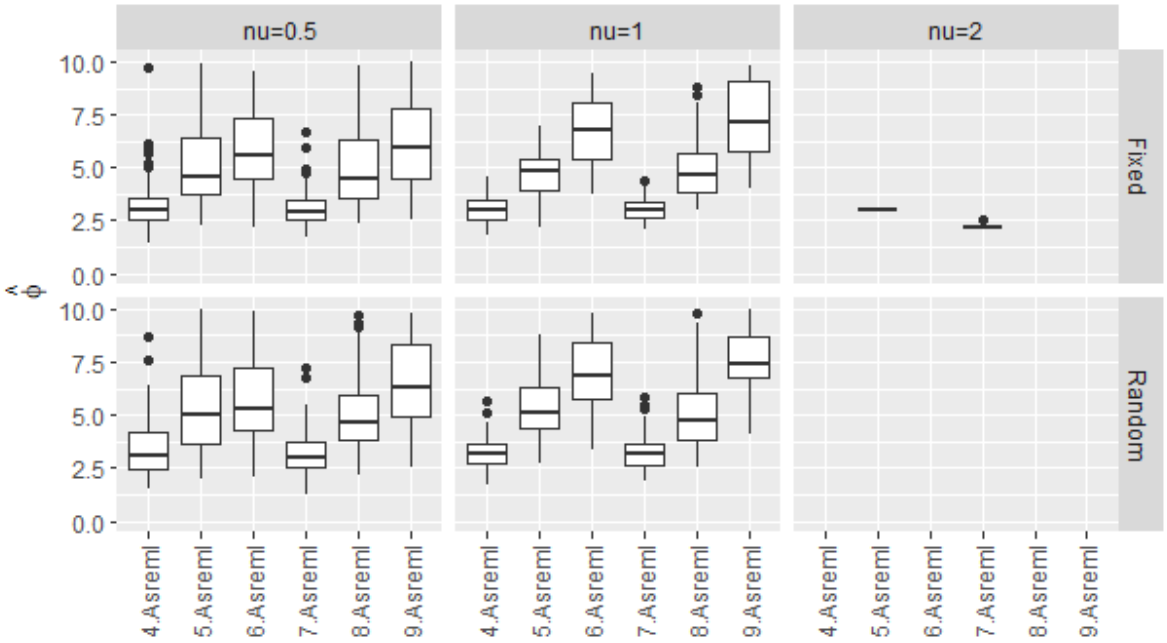
Figure 20: Estimated nugget-free rescaling values for $\hat{\sigma}_s^2$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



However, when analyzing the random effect, it is observed that in the non-rescaled results (Figure 5), extreme underestimations occurred in scenarios 1, 5 and 7, which were not replicated after rescaling. When examining the rescaled results, it appears that the median $\hat{\sigma}_s^2$ coincides exactly with σ_s^2, but only occurs in scenario 4 for both fixed and random effects, and only in the case $\hat{v} = 0{,}5$. However, the most significant underestimations were observed in the broadest scopes in both effects, that is, in scenarios 6 and 9.

In the context of $\hat{\phi}$ rescaling, the smallest ranges represented in scenarios 4 and 7 exhibited medians very close to $\phi = 3$ (Figure 21). However, in scenarios 6 and 8, where $\phi = 3$, underestimations were identified in $\hat{\phi}$, except in scenario 5 with a random effect with $\hat{v} = 1$, where the median corresponded exactly to the simulated value. This underestimation trend has also been observed in broader scopes. These behaviors $\hat{\phi}$were quite similar to the $\hat{\phi}$ regular grid without nugget effect, considering the rescaling.

Figure 21: Estimated nugget-free rescaling values for $\hat{\phi}$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.

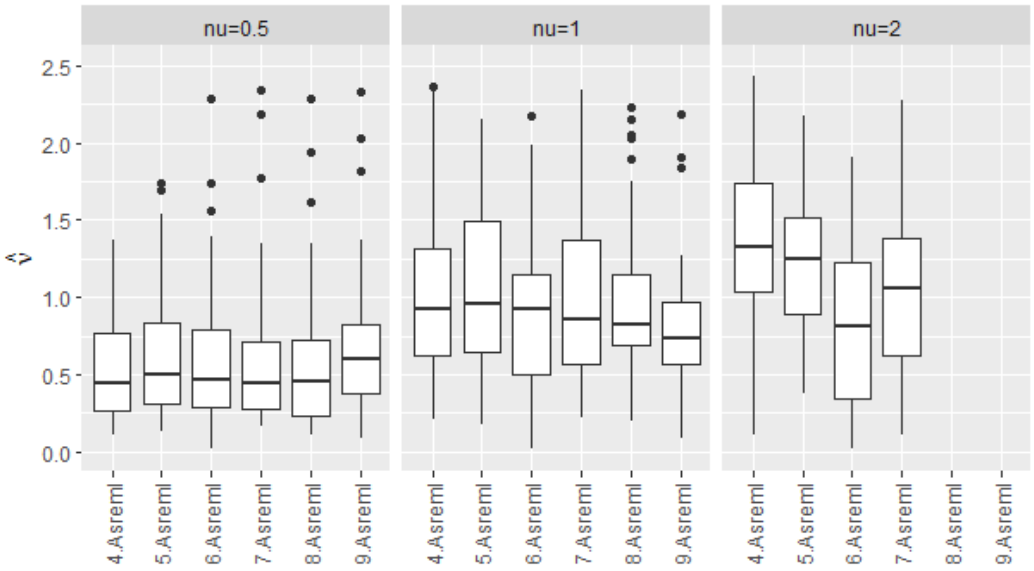


When analyzing the results before and after rescaling, a similar pattern is observed. With the exception of $\hat{v} = 2$, there was no convergence in any of the scenarios tested after rescheduling, unlike the results obtained without rescheduling, which presented results in scenarios 1, 5 and 7, all extremely underestimated. In the context of the fixed effect, although results were found, they were also extremely underestimated in both the rescaled and non-rescaled results. However, in general, the estimates of $\hat{\phi}$ the rescaling, as well as the regular grid (Figure 6), were observed at the greatest distances.

RESCHEDULING WITH REGULAR NUGGET EFFECT

After analyzing the effect of smoothing with a random regular nugget effect (Figure 22), it is clear that in scenario 5, when considering $\hat{v}$, the simulated smoothing effect was exactly replicated ($\nu = 1$).

Figure 22: Estimated nugget rescaling values for $\hat{v}$scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.

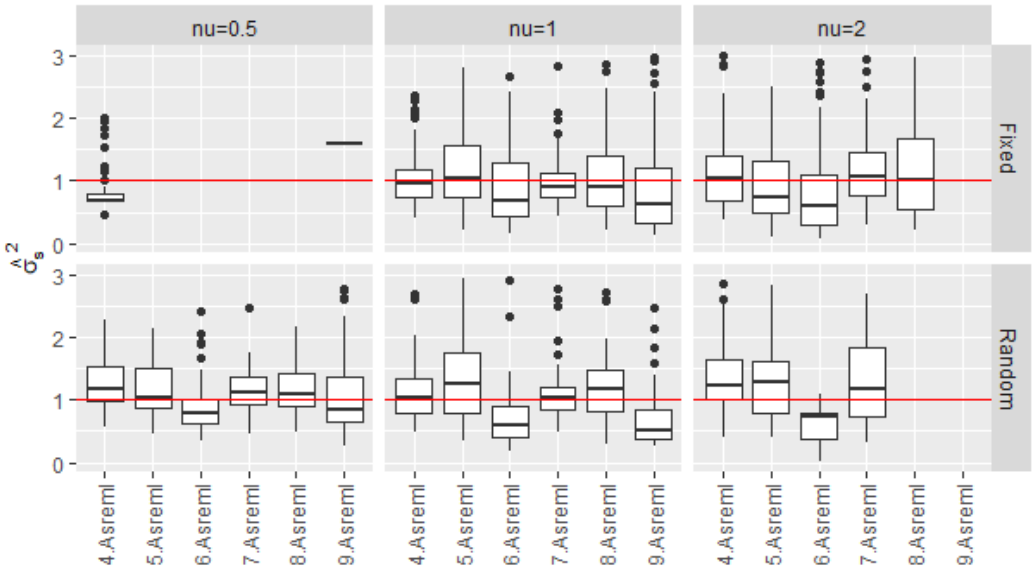


In the other scenarios, there was a slight underestimation in the median of the exponential results, with the exception of scenario 9, where the median demonstration $\hat{v}$ was overestimated. When subsequently adopting Whitte's approach, we observed that, in general, the scenarios resulted in underestimated medians $v = 1$, with the exception of scenario 9, whose median coincided exactly with the expected value. However, the less predominant results occurred when adopting $v = 2$, as the largest underestimations were valid, with their medians below $\hat{v} = 0.5$. Furthermore, in this Gaussian estimate, no convergences were found for the largest ranges, that is, in scenarios 8 and 9.

When contrasting the results without rescaling (Figure 11) and with rescaling (Figure 22), greater uniformity in the data is evident, with a reduction in the number of discrepant points after this transformation. However, as in previous analyses, the smoothing effects for $v = 1$ and $v = 2$ resulted in the least overwhelming performances, all of which underestimated the simulated values. It is important to note that, without rescaling, scenarios 6 and 7, with $v = 2$, did not produce results, but after this transformation, results were obtained. But these results underestimated the medians. It is observable, however, that despite the inability to accurately estimate the simulated values, there was a reduction in the relationship to the medians after the rescaling intervention.

After generating the results in $\hat{\sigma}_s^2$ (Figure 23), it was noted that, keeping it $v = 0.5$ fixed, there was less convergence of the scenarios, and underestimation of the median in scenarios 4 and overestimation of the median in scenario 9, the latter maintaining a constant value around 1.6 da $\hat{\sigma}_s^2$.

Figure 23: Estimated nugget rescaling values for $\hat{\sigma}_s^2$scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



When considering the random effect while still exponential, scenario 5 stood out as the most favorable, but still exhibited underestimation in $\hat{\sigma}_s^2$. The worst-performing scenario in this configuration was 6, with underestimation of the median, repeated in scenario 9, while the others are considered overestimation.

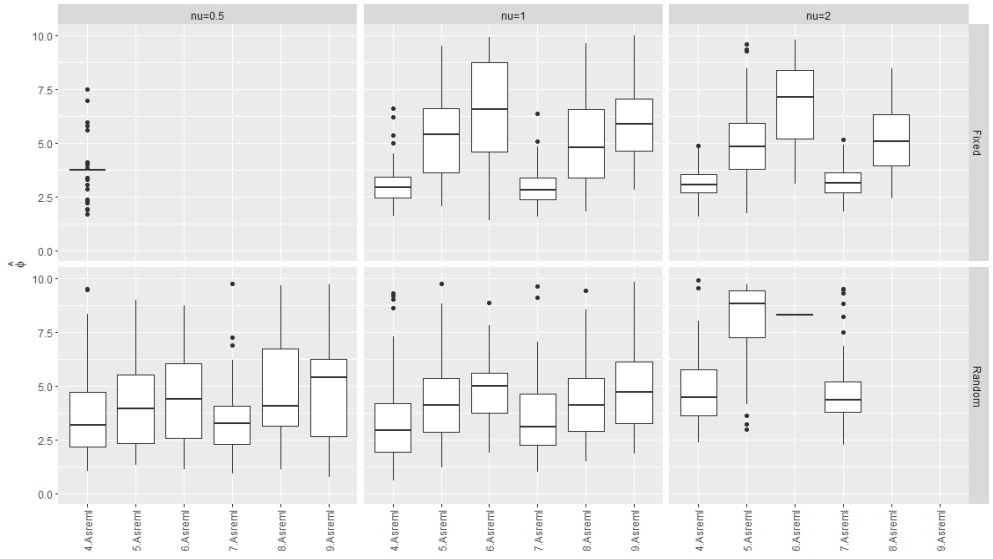
In the case of $v = 1$, in the fixed effect, scenario 4 coincided exactly with what was σ_s^2 expected, but in the others there was an underestimation of the median, except in scenario 5. For the random effect with $v = 1$, the most protective results were exposed in smaller-range locations, 4 and 7, although they didn't exactly succeed $\sigma_s^2 = 1$. Analyzing $v = 2$ the fixed effect, scenario 8 presented the best result, with its median coinciding exactly with the estimated value. There was also performance on scenarios 4 and 7, but without the ability to estimate scenario 9. In the case of the random effect, in addition to scenario 9, this configuration was unable to estimate scenario 8, and most other scenarios overestimated σ_s^2, except for scenario 6, which underestimated it, this being the worst configuration in this estimate.

In general, in larger-range scenarios, we observed a tendency towards underestimation of σ_s^2, these being the results found in this configuration. When comparing the results with and without rescaling, a loss in the estimate was noted in the $v = 0.5$fixed-effect configuration. Without rescheduling, although some scenarios were underestimated σ_s^2, they were able to generate results. The inability to generate results

after rescheduling harmed scenario 9, which was previously presented $\hat{\sigma}_s^2$close to the simulated value. In the case of the random effect, in general, the results were more homogeneous and slightly closer to the median after rescaling, except in scenarios 8 and 9, which fail to converge to $\hat{\sigma}_s^2$. It is notable that, before rescheduling, scenarios 4 with $v = 1$ and $v = 2$ there were no results obtained for $\hat{\sigma}_s^2$, but after rescheduling, results were found, as in scenario 7 with$v = 2$.

When analyzing $\hat{\phi}$, we notice a difficulty similar to other scenarios after the rescheduling, presented previously, especially from the perspective of ϕgreater distances, which are consistently underestimated. Again, $v = 0.5$, it revealed the most unfavorable results, estimating only scenario 4 with a median of $\hat{\phi} = 3.75$, that is, overestimating (Figure 24).

Figure 24: Estimated nugget rescaling values for $\hat{\phi}$scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



In scenarios 4 and 7, with lower $\hat{\phi}$, it was possible to estimate exactly the expected range $\phi = 3$, while in the fixed scenario with $v = 1$, scenario 5 overestimated $\hat{\phi}$ and scenario 8 underestimated, although close to the expected value. However, the largest $\hat{\phi}$ represented by scenarios 6 and 9 were a significant underestimation of ϕ, being considerably below expectations. When analyzed $v = 2$ in the fixed effect, scenarios 5 and 8 presented exactly the expected value, with the results for this $\hat{\phi}$being the best after rescheduling, although the others presented close, did not exactly reach the simulated

value of $\phi = 5$. It is noteworthy that, for the greatest range of $\phi = 10$com $v = 2$, it was not possible to estimate it.

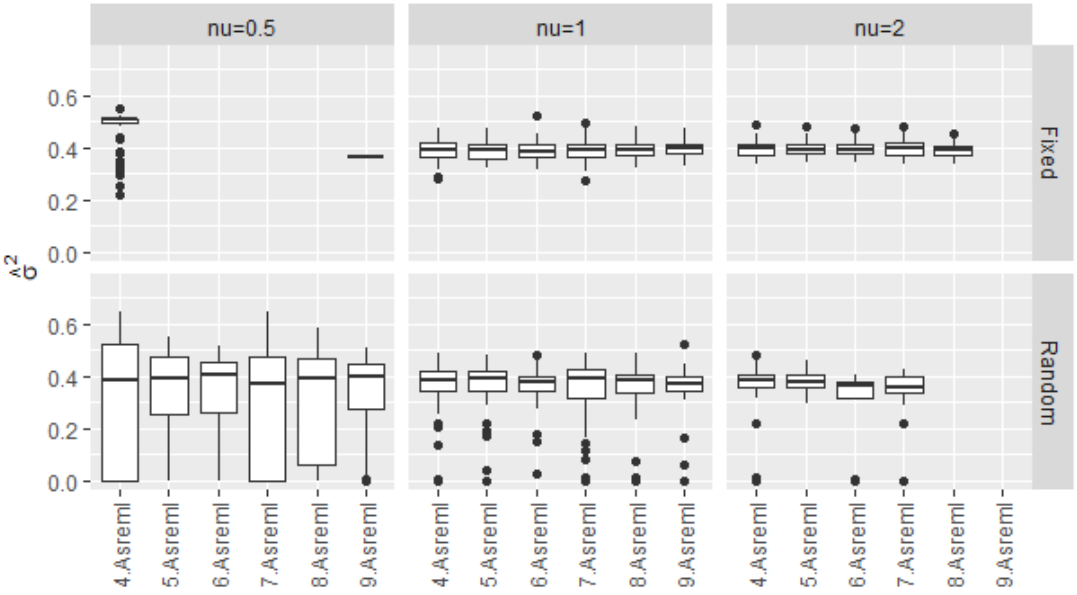
Without a random effect, the scenarios with the smallest range (4 and 7) were the closest to ϕ, while the others were considered an underestimation of $\hat{\phi}$. The same pattern occurred with $v = 1$and $v = 2$, where the lowest scenarios $\hat{\phi}$had adequate estimates, but the others resulted in underestimates. It is worth mentioning that, in the random effect with $v = 2$, scenarios 8 and 9 could not be estimated, while scenario 8 without a fixed effect, in this same configuration, presented the best result. In general, the random effect results were inferior, that is, worse than the fixed effect results for greater distances.

When comparing the results before and after rescaling, note that with $v = 0.5$ the fixed effect, before rescaling, there were results, although underestimated for ϕ.later, with $v = 1$and $v = 2$, scenario 4 presented better results in relation to its medians after rescaling. In the scenarios representing $\phi = 5$, both before and after rescaling, they underestimated ϕat greater distances. Furthermore, although it was underestimated, $v = 2$in scenario 9 without rescheduling it showed results, while after rescheduling it was not possible to obtain results.

In the random effect, there was an improvement in the first two $\hat{\phi}$after rescaling, but, for the largest range, as with rescaling, there was an underestimation of the median. Note that scenario 4 with $v = 1$and $v = 2$did not present results before rescheduling, but after, the results were close to expected, especially for $v = 1$. However, the model was not able to estimate scenarios 8 and 9 without a random effect, unlike the one without rescaling, which was capable, but generated underestimated and inconsistent results.

Regarding the nugget effect, when considering $v = 0.5$fixed, note that only scenarios 4 and 9 present results. The first overestimated σ^2, while scenario 9 underestimated. In the case of the random effect with $v = 0.5$, all results were medians equal to or very close to σ^2(=0.4). However, the vast majority of results in this configuration revealed underestimates in the results of $\hat{\sigma}^2$(Figure 25).

Figure 25: Estimated nugget rescaling values for $\hat{\sigma}^2$scenarios 1 to 9, considering a regular mesh modeled with fixed and random effects.



When considering $v = 1$fixed, we notice a smaller amplitude and greater homogeneity in the results of $\hat{\sigma}^2$, which are close to σ^2. This consistency remains in the case of $v = 2$no fixed effect, although in this scenario it was not possible to estimate scenario 9. However, in the random effect with $v = 2$, in addition to scenario 9, the model was not able to estimate scenario 8 after rescaling. Furthermore, scenarios 6 and 7 considered underestimates in relation to σ^2.

Comparing the results before and after rescaling, $v = 0.5$fixed, after rescaling, performed worse than the results without rescaling, resulting in the loss of ability to estimate $\hat{\sigma}^2$satisfactorily in all scenarios. Only scenarios 4 and 9 were estimated, but with greater discrepancy in relation to σ^2. In the case of the random effect with $v = 0.5$, although there was a reduction in outliers after rescaling, the amplitude of the results increased, revealing even more underestimations of σ^2.

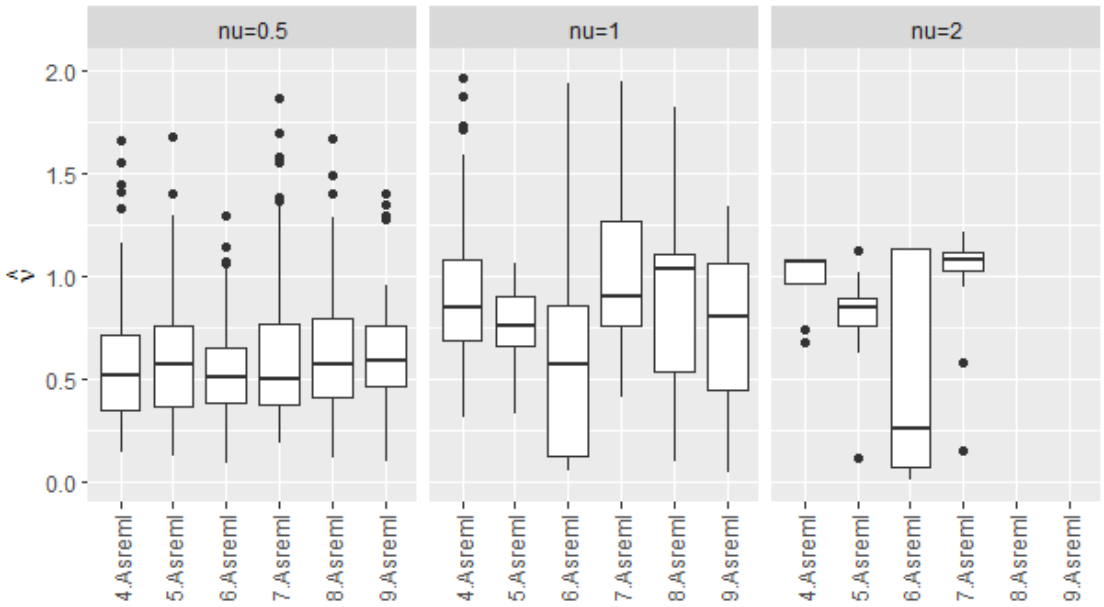
However, with $v = 1$, both fixed and random, there was a reduction in both outliers and the range of results after rescaling, making them more homogeneous and closer to the median, aligning better with σ^2. This pattern was also observed to $v = 2$,decrease that rescheduling brought no benefits $\hat{\sigma}^2$.

RESCHEDULING WITH IRREGULAR NUGGET EFFECT

After rescaling with the irregular grid without random effect, when analyzing $\hat{v}$, observe in figure 26, that with $v = 0.5$, all scenarios showed values very close to v, but

with a slight underestimation in the median. Scenario 7 presented the median exactly at the expected value.

Figure 26: Estimated nugget rescaling values for $\hat{v}$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.

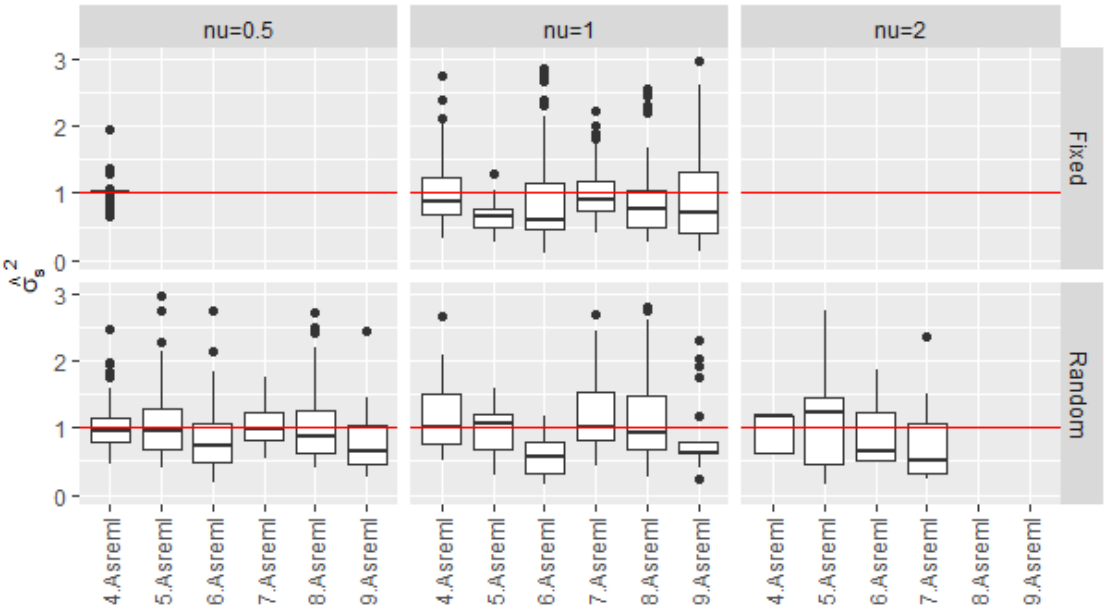


When considering $v = 1$, all scenarios were underestimated, except scenario 8, which, despite offering the best estimate in terms of median, presented a small overestimation. In this configuration, the worst result in terms of median was scenario 6. In the case of $v = 2$, scenario 6 also showed the worst result, with all scenarios being extremely underestimated. In scenarios 8 and 9 in this configuration, it was not possible to obtain results.

When we compared the results of rescaled data with non-rescaled data (Figure 15), with $v = 0.5$, there was a small reduction in the median, and the range of simulated results later after rescaling. For $v = 1$, before rescheduling, it was not possible to obtain scenario 4, but after rescheduling, scenario 4 was obtained, but with underestimation. In the other scenarios in this configuration, there were no improvements, except in scenario 8, whose median also moved more from the expected results to $\hat{v}$. Finally, $v = 2$scenarios 4, 6 and 7 were able to generate results after rescaling, something not possible in data without rescaling, however, the results were underestimated.

After examining$\hat{\sigma}_s^2$ with $v = 0.5$, only scenario 4 was generated in the fixed effect, revealing the median exactly the expected result ($\sigma_s^2 = 1$). However, in the random effect with the same configuration, all scenarios were generated, with scenario 7 being the highlight, followed by scenarios 4 and 5, while the others are considered small underestimates.

Figure 27: Estimated nugget rescaling values for $\hat{\sigma}_s^2$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



Fast forward to $v = 1$, when comparing fixed and random, all fixed scenarios showed underestimated medians, with scenario 7 being the closest to σ_s^2its median. This scenario 7 was consistently the best result across all configurations on the irregular nugget grid. In the random effect with $v = 1$, scenario 7 exactly achieved the expected value of $\hat{\sigma}_s^2$, while scenario 4 also exactly reflected the simulated value of $\hat{\sigma}_s^2$. This $v = 1$non-random configuration revealed the best estimates for $\hat{\sigma}_s^2$. It is also noteworthy that once again, the worst results seem to occur in more far-reaching scenarios.

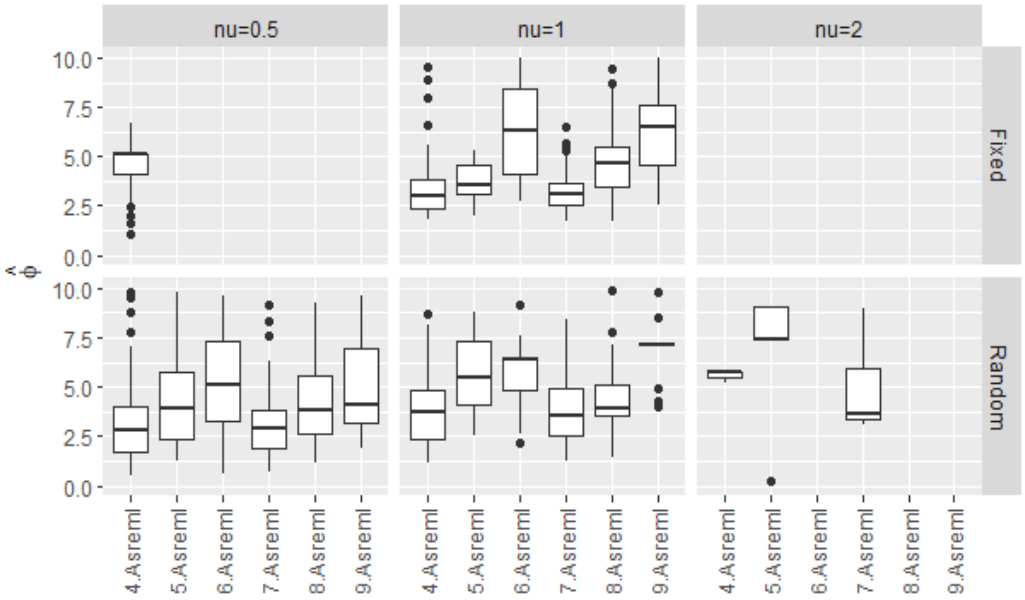
For $v = 2$the fixed effect, no results were obtained, while for the random effect, scenarios 4 and 5 underestimated σ_s^2, while scenarios 6 and 7 also considered underestimation, and the others did not converge in this configuration. Comparing the rescaled results with the non-rescheduled ones (Figure 14), a loss in the estimate is observed with $v = 0.5$, because after the rescheduling, only scenario 4 was generated,

whereas before the rescheduling, there were promising results in scenarios 5 and 7, albeit with underestimations. With $v = 1$, the results before and after rescaling behaved similarly, with only a reduction in outliers after rescaling. However, with $v = 2$, before rescaling, the values were extremely outliers, and after rescaling, the model did not generate any results.

In the random effects with $v = 0.5$, as well as in the $v = 1$fixed effect, the results after rescaling were quite similar to the previous results. With $v = 1$, there was an improvement after rescheduling, with scenario 4 generating excellent results. However, with $v = 2$no rescaling, the model was not able to estimate scenarios 4, 6 and 7, while after rescaling, these scenarios were generated, although the first two overestimated the median value and the last one underestimated it σ_s^2. Finally, without rescheduling, scenarios 8 and 9 were generated, with scenario 8 standing out as the best for this configuration.

When analyzing the results of$\hat{\phi}$ in Figure 28, it is observed that the $v = 0.5$fixed presented the median of scenario 4 overestimated, reflecting the trend observed in previous parameters. In the random effect, the scenarios with the smallest range (4 and 7) showed moderately underestimated medians, although close to ϕ; however, the other scenarios were underestimated in relation to the median.

Figure 27: Estimated nugget rescaling values for $\hat{\phi}$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



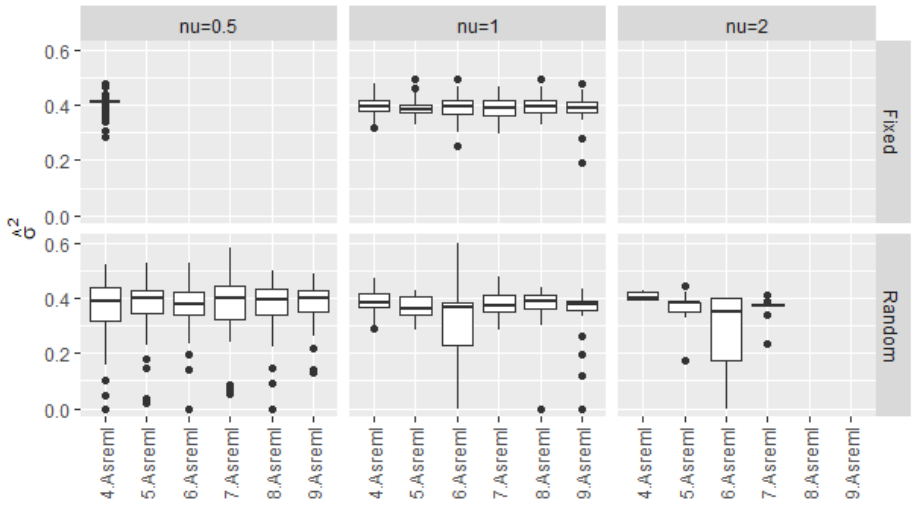
When considering $v = 1$both fixed and random effects, the smallest distances revealed the most accurate estimates. In other scenarios,$\hat{\phi}$ was consistently underestimated. It is important to note that the configurations planned with $v = 1$and $v = 2$produced the results, even for the smallest distances. In the fixed mode scenario $v = 2$, there were no results, but without random effect, only three scenarios were generated: scenarios 4 and 5 overestimated ϕ, while scenario 7 presented the median $\hat{\phi}$exactly at the value presented. However, most results in this setting were underestimated.

When analyzing these results together with the results prior to rescheduling (Figure 12) for $\hat{\phi}$, with $v = 0.5$no fixed, there was a loss, as only scenario 4 was generated after rescheduling. Without rescheduling, all scenarios considered results, but with underestimations. No random effect,$\hat{\phi}$ showed fewer outliers after rescaling, with a reduced range in the results, resulting in values closer to the median. However, both before and after rescaling, there was underestimation for the largest ranges.

to $v = 1$mention, despite the greater homogeneity, the behavior of the results was very similar before and after rescaling. With the exception that not randomly, after rescaling, it was possible to estimate scenario 4, but with an overestimate in the median. Not $v = 2$in the fixed, there were no results after the rescaling, but before the rescaling, the results were extremely underestimated and discrepant. No random effects, after rescheduling, scenarios 4, 5 and 7 were estimated, which did not occur before rescheduling. Finally, it is worth highlighting that despite the specific variability in the results, before rescheduling it was possible to estimate scenario 9.

When examining the results of $\hat{\sigma}^2$, it is noted that in the fixed scenario with $v = 0.5$, scenario 4 presented an overestimate, even though its median was very close to the real value of σ^2. On the other hand, in the random scenario, all results were exactly as expected for $\hat{\sigma}^2$($\sigma^2 = 0.4$), except for scenarios 4 and 6, which presented a significantly underestimated mean. These satisfactory estimates $\hat{\sigma}^2$were replicated in the fixed scenario $v = 1$, demonstrating the best results due to the small dispersion of values, very close to the median (Figure 28).

Figure 28: Estimated nugget rescaling values for $\hat{\sigma}^2$scenarios 1 to 9, considering an irregular mesh modeled with fixed and random effects.



However, in the random scenario with $v = 1$, despite the medians $\hat{\sigma}^2$representing or being very close to the expected value, the results obtained a greater dispersion compared to the homogeneity observed in the scenario with $v = 1$fixed. Still, in the random scenario with $v = 1$, the results were better than in the random scenario with $v = 0.5$, with the exception of scenario 6, which presented underestimated estimates.

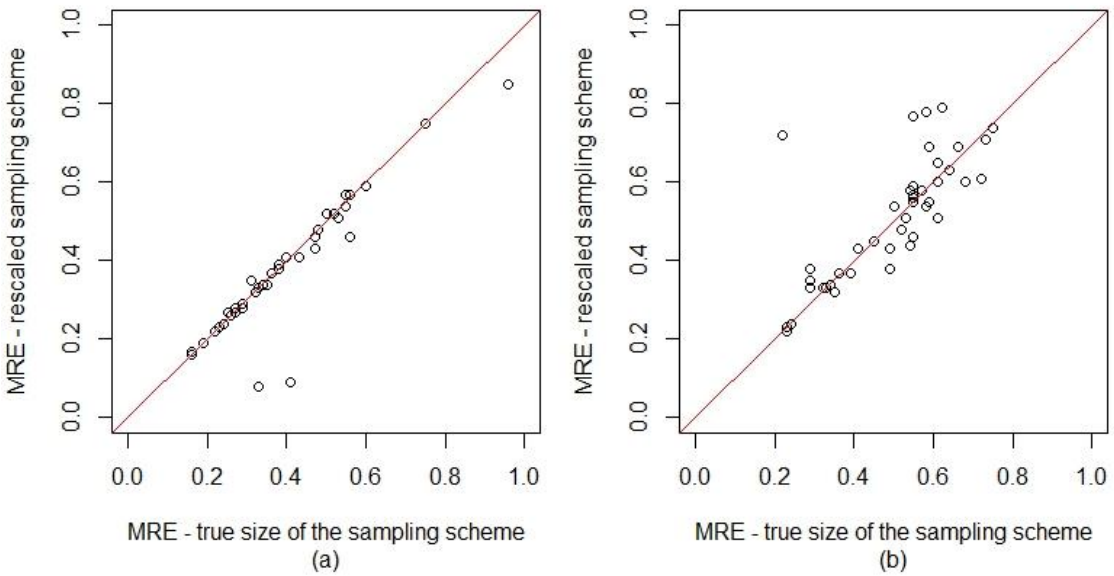
In the fixed configuration $v = 2$, no results were obtained after rescheduling, while in random mode, scenarios 4 and 5 were the best, followed by scenario 7, when compared to each other. However, scenario 6 $v = 2$presented the poorest estimates for $\hat{\sigma}^2$. When comparing the rescaled and non-rescheduled results, it can be seen that in the fixed scenario with $v = 0.5$, before rescheduling, all scenarios were satisfactorily estimated (Figure 13). In the random scenario, there was a decrease in outliers and greater uniformity in post-rescaling results, although, in general, the behavior of the data remained consistent before and after rescaling.

Then, in the scenario with $v = 1$, there was also a reduction in outliers, and the range of results for $\hat{\sigma}^2$reduction, providing the most consistent estimates both before and after rescaling, when analyzing all configurations together. This reduction in outliers was also observed in the random scenario with $v = 1$, along with the ability to estimate $\hat{\sigma}^2$in scenario 4 after rescaling, which was not possible before. Furthermore, we observed a slight improvement in the uniformity of results for the scenarios, especially for scenario 8. Finally, in the fixed scenario with $v = 2$after rescheduling, there was a loss, as none of the scenarios were estimated, although before rescheduling they were obtained viable results for scenarios 5, 6 and 7, although with little uniformity in results. No random effect

with $\nu = 2$, after rescaling, it was possible to estimate the $\hat{\sigma}^2$close to σ^2in scenarios 4 to 7, except for scenario 5 which, without rescaling, generated overwhelming results. However, the other scenarios were not discovered after rescheduling, despite showing results before rescheduling. Finally, scenario 8 generated extremely underestimated estimates, while scenario 9 showed a median with a good estimate for $\hat{\sigma}^2$.

Comparing the parameter estimates, the reach obtained was proportionally similar to its scaled result; example: $\phi = 30$ implied $\hat{\phi} \approx 3$ with the exception of scenario 6 with nugget effect, when $\nu = 2$ of random effect. The other parameters are close to the originally simulated values and similar to those obtained in the analysis without rescaling – when there was convergence. Figure 29 presents the scatterplots of the median relative error (MRE) referring to the results obtained in cases with true grid size versus rescaling, in scenarios in which there was convergence. Note that in cases without nugget effect the median presented cases with lower ER.

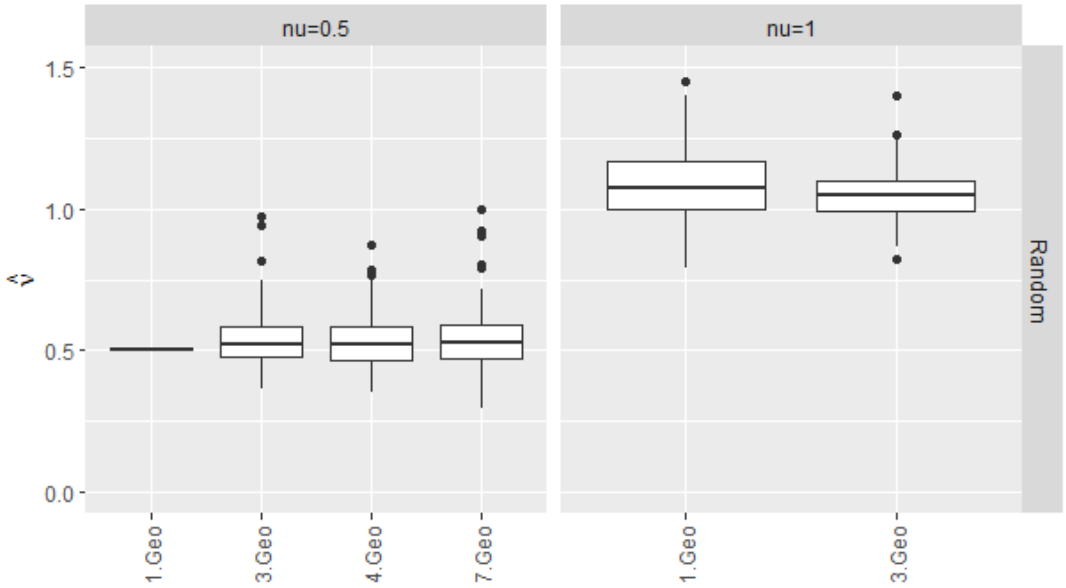
Figure 29: Dispersion of the median relative error from each combination of scenario and model analysed with the true size versus reescaled sampling schemes. All data were isotropic, however layout (a) is without nugget effect and (b) has nugget effect. The diagonal line represent the similarity results in x- and y- axis.



ANISOTROPY WITHOUT REGULAR NUGGET EFFECT

The results obtained for anisotropy without nugget effect on a regular grid are represented in Figure 30. The planes included are the best results produced before rescaling. However, it is important to highlight that the Asreml-R model was not able to estimate any of the scenarios with anisotropy. Therefore, the results discussed below refer to GeoR. In these results, Figure 30, it is noted that when adopting a random value of $v = 0.5$, all tested scenarios exhibited medians very close to v, reducing small overestimates, with the exception of scenario 1, where all results were exactly the same as expected ($v = 0.5$).

Figure 30: Estimated values without nuggets with anisotropy for the scenarios $\hat{v}$chosen as best before rescaling, considering a regular grid, and modeled with random effects.

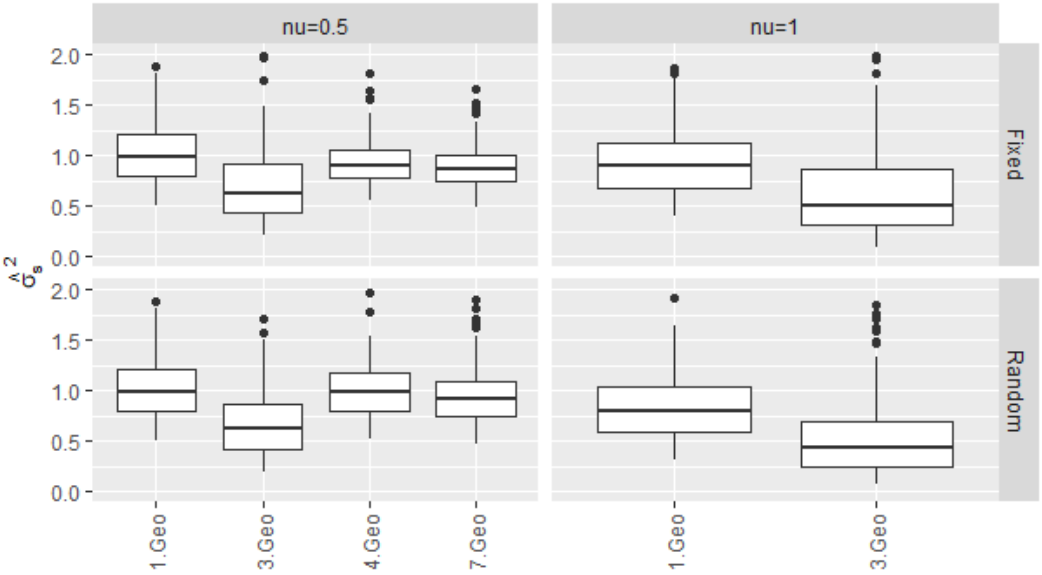


It is important to mention that, in both scenarios $v = 0.5$and those with $v = 1$, the convergence rate was 100%. However, when adopting $v = 1$, it is observed that the results not only show a greater dispersion in relation to the $v$estimated median, but also show overestimated estimates. Between the two scenarios tested, scenario 3 demonstrated is the most favorable, as it is closer to the value of v, in addition to presenting greater uniformity in the results.

Regarding the estimate of $\hat{\sigma}_s^2$anisotropy, it was shown that with $v = 0.5$the fixed effect, scenario 1 presented the most developed results, evidenced by the median close to

v. On the other hand, scenario 3 in this configuration showed the greatest underestimation, while scenarios 4 and 7 also exhibited an underestimation of $\hat{\sigma}_s^2$, although their medians were closer than σ_s^2 in scenario 3 (Figure 31).

Figure 31: Estimated values without nuggets with anisotropy for the scenarios $\hat{\sigma}_s^2$chosen as best before rescaling, considering a regular grid, and modeled with random effects.



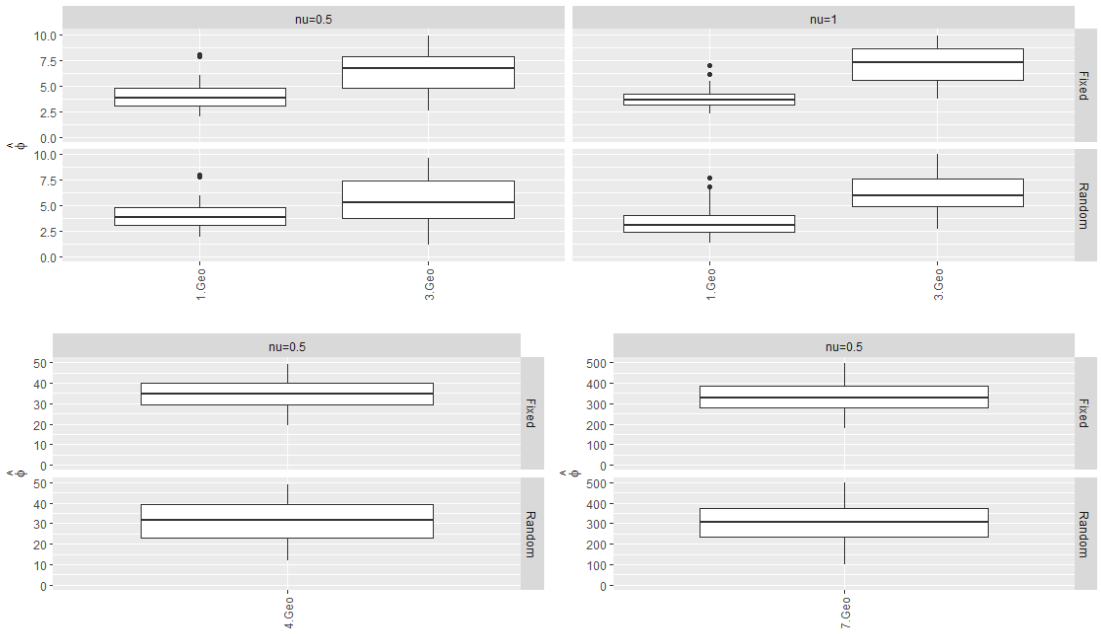
When analyzing the random effect in the same exponential context, it should be noted that scenario 1 behaves similarly to the fixed effect, but with a small dispersion in relation to its median. The same pattern was noted in scenario 3. In the case of scenarios 4 and 7, these were also underestimated, as in the fixed effect, however, their medians in the random context were closer to σ_s^2. When considering $v = 1$with fixed effect, both scenarios 1 and 3 exhibited underestimations with respect to σ_s^2, with scenario 1 showing a better improved performance compared to scenario 3.

These same characteristics were observed in the random effect, however, in the random context, in addition to the increase in outliers, there was a greater underestimation $\sigma_s^2$when compared to the results found in the fixed context. In general, $v = 0.5$in both fixed and random contexts, scenario 1 represented the most accurate estimates of $\hat{\sigma}_s^2$.

To evaluate the estimate $\hat{\phi}$in relation to distances, separate graphs were generated for each scenario. When considering scenario 1 with $v = 0.5$, the estimate ($\hat{\phi}$) was 3. For both the fixed and random effects, the medians of the results coincided with the estimated value (Figure 32). However, a slight greater homogeneity was noted in the

results of the random effect. In the case of $\nu = 1$scenario 1, both effects (fixed and random) had median results close to the expected value for $\hat{\phi}$. However, in the fixed model, scenario 1 proved to be more homogeneous than in the random model, both with $\nu = 1$and when compared to$\nu = 0.5$.

Figure 32: Estimated values without nuggets with anisotropy for the scenarios $\hat{\phi}$chosen as best before rescaling, considering a regular grid, and modeled with random effects.

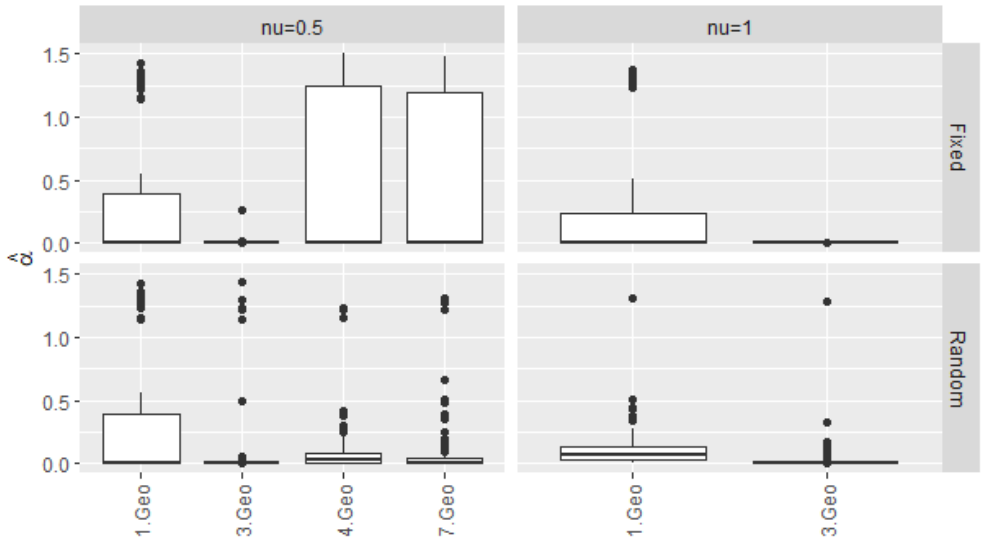


Analyzing $\hat{\phi}$ scenario 3 with $\nu = 0.5$, an underestimation is suggested, since the expected result was $\phi = 10$. This underestimation was also observed in the random model for the same configuration and in the fixed model with $\nu = 1$. However, it $\nu = 1$presented superior results compared to $\nu = 0.5$, with fixed being the most accurate. When examining scenario 4 with $\nu = 0.5$, both the fixed and random models overestimated the expected result $\phi = 30$. However, between the two, the best result was not $\nu = 0.5$with the random effect, as the median changed more than ϕ. Finally, when analyzing model 7 with $\nu = 0.5$, where $\phi = 300$, the same occurrences that occurred in scenario 4 were observed. Both models overestimated ϕ, but scenario 7 with a random effect called a median closer to ϕ.

The estimate from this $\hat{\alpha}$ presentation in figure 33 highlights that the anisotropy orientation angle used in this study is $\alpha \cong 1{,}31$ radians, which is equivalent to 75° degrees. Therefore, none of the scenarios generated were able to accurately estimate the

α. With $\nu = 0.5$, a small part of the results of scenarios 4 and 7 came close to the expected result; however, like the others, their medians were very close to zero.

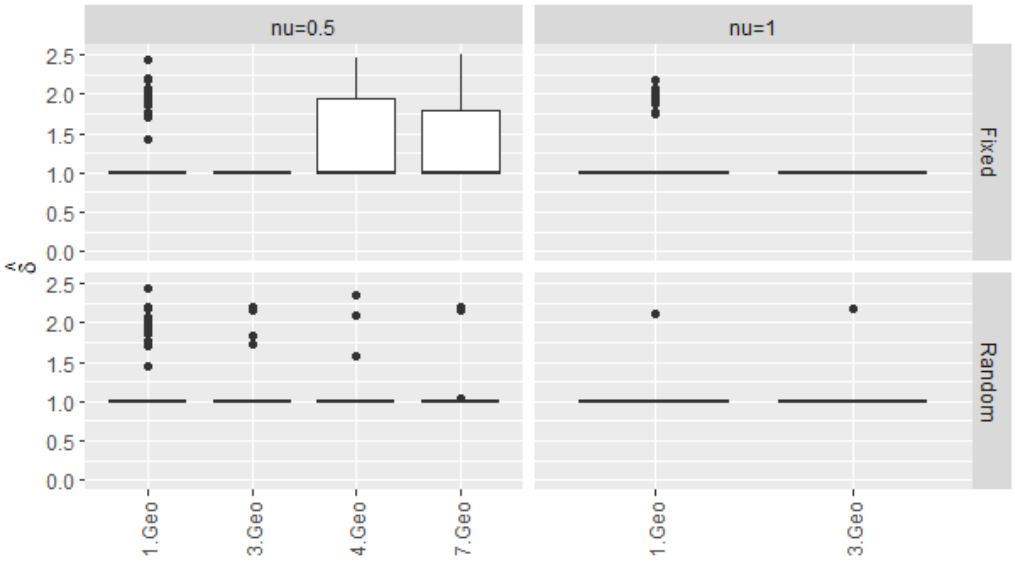
Figure 33: Estimated values without nuggets with anisotropy for the scenarios $\hat{\alpha}$chosen as best before rescaling, considering a regular grid, and modeled with random effects.



In other words, all generated scenarios, regardless of settings, were underestimated α. The most unfavorable result, where there was no variation in the results and the value was consistently zero, occurred in scenario 3 with ν =1 at fixed. In addition to this, only with outliers, but it also remained with all results close to zero in scenarios 3 with $\nu = 0.5$both the fixed and random models, and in scenario 3 with $\nu = 1$the random effect.

The ratio between the magnitudes of variation in different variations, represented by the rules δ adopted in this study, was established as 2 (Figure 34). Note that, similarly to $\hat{\alpha}$, none of the results will be presented in the median close to $\hat{\delta}$. The scenarios that came closest to this value were scenarios 4 and 7 in the context of the fixed effect with $\nu = 0.5$. However, these scenarios, like others, underestimatedδ.

Figure 34: Estimated values without nuggets with anisotropy for the scenarios $\hat{\delta}$chosen as best before rescaling, considering a regular grid, and modeled with random effects.

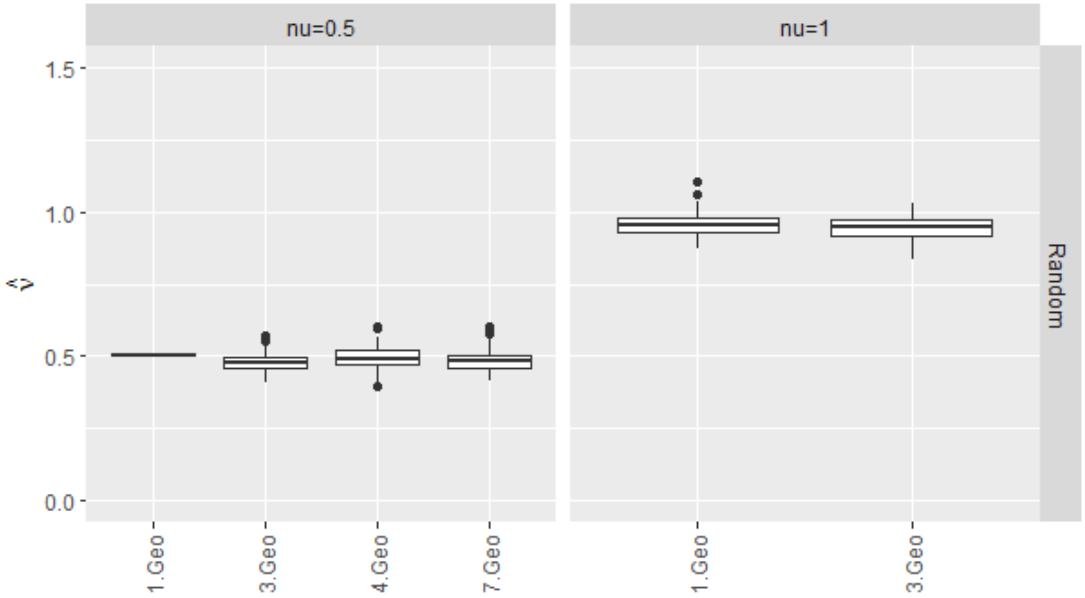


In the other scenarios, all medians of the $\hat{\delta}$result estimates were 1, with the presence of outliers above the median. An exception was observed in scenario 3 of the fixed effect, as well $\nu = 0.5$as with $\nu = 1$, in which all results concentrated on 1, without the presence of outliers.

WITH ANISOTROPY NUGGET EFFECT

An analysis of the anisotropy results without the nugget effect, with an irregular and random grid, reveals similarities with the results obtained in the same configuration, with a regular grid (Figure 30). Scenario 1 presented all results in $\hat{v}$accordance with v(Figure 35).

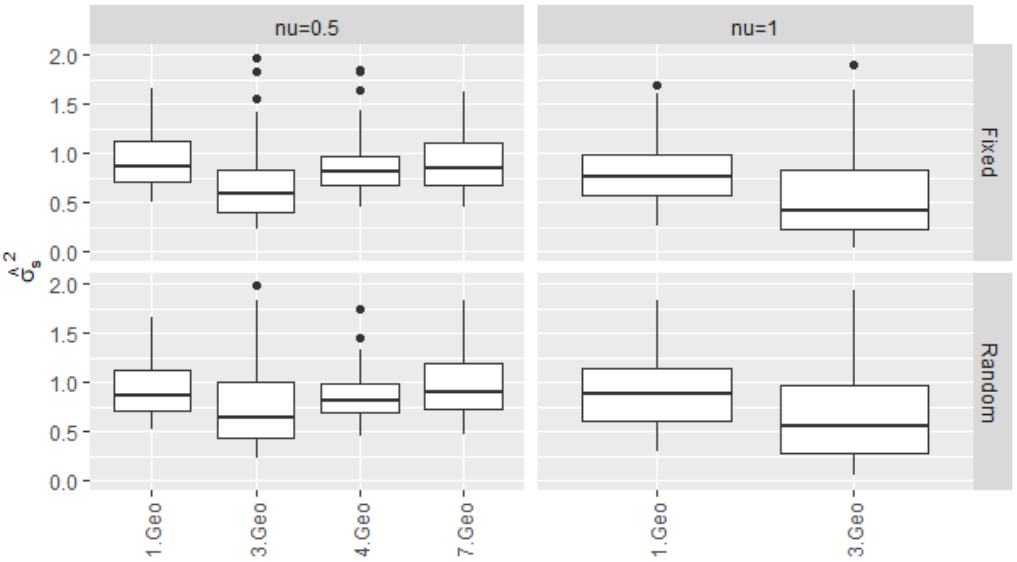
Figure 35: Estimated values without nuggets with anisotropy for the scenarios $\hat{v}$chosen as best before rescaling, considering an irregular grid, and modeled with random effects.



In scenario 4, the median results also aligned with the expected value of v. However, there was a variation in the results, while scenarios 3 and 7 showed a small underestimation in relation to v. When analyzing $v = 1$, both generated scenarios considered underestimations of v ($v = 1$). Scenario 1, however, was closest to the expected value. Overall, when comparing the results, the most dominant scenario was scenario 1 where o $\hat{v}$= 0.5. It is important to highlight that all scenarios analyzed achieved a convergence rate of 100%.

When examining the estimate $\hat{\sigma}_s^2$ with $v = 0.5$in the context of the fixed effect, all results indicated an underestimation of σ_s^2, with scenario 3 standing out as the least accurate (Figure 36). Regarding the random effect, the results were similar to the fixed effect scenario, with scenario 3 presenting the most unfavorable performance.

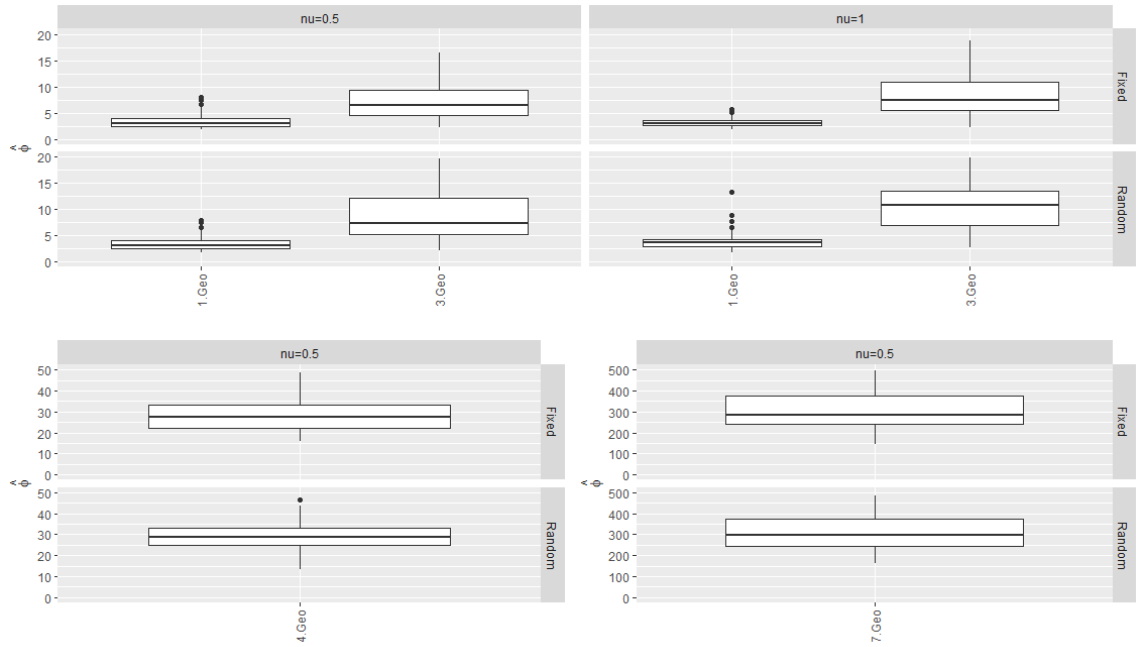
Figure 36: Estimated values without nuggets with anisotropy for the scenarios $\hat{\sigma}_s^2$chosen as best before rescaling, considering an irregular grid, and modeled with random effects.



All scenarios showed underestimation of σ_s^2. However, scenario 7, as well as scenario 1 of the fixed effect, came closer to σ_s^2. When considering $v = 1$, both in the context of fixed and random effects, underestimations of σ_s^2. In both cases (fixed and random), scenario 3 displayed the least accurate results.

Regarding $\hat{\phi}$, it should be noted that scenario 1 of the fixed effect, as well $v = 0.5$as with $v = 1$, presented medians exactly the same as $\phi = 3$ (Figure 37). However, when comparing both configurations, the $v = 1$ presented the most accurate results for this scenario, both in the fixed and random effects, exhibiting less dispersion in relation to the median.

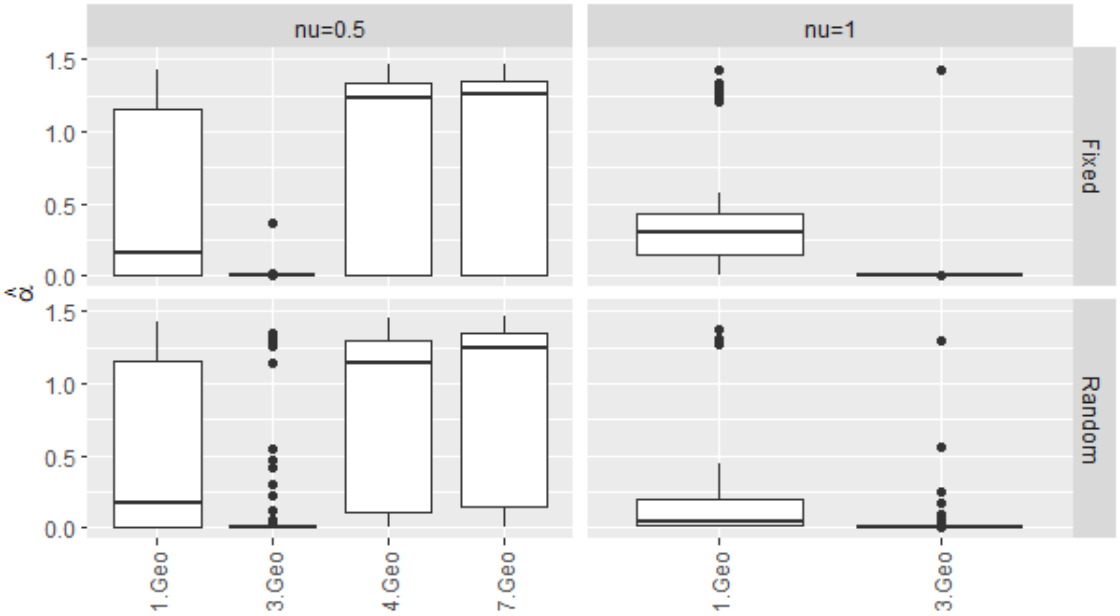
Figure 37: Estimated values without nuggets with anisotropy for the scenarios $\hat{\phi}$chosen as best before rescaling, considering an irregular grid, and modeled with random effects.



when $v = 0.5$was observed ϕ, both in the fixed and random effects, as it was expected to $\hat{\phi}$ be equal to 10 meters. With $v = 1$, there was also an underestimation of the estimate $\hat{\phi}$ in the fixed effect, however, in the random effect, despite underestimation occurring, this was the one that presented the closest median among ϕall the configurations that included this scenario. Analyzing scenario 4 with, where $\phi = 30$ meters, it is observed that although both effects were underestimated ϕ, this one was closer in terms of random effect. Finally, in scenario 7, where meters are expected $\phi = 300$, the median of the random effect caused $\hat{\phi}$exactly the expected result, although with variation in the results in general. Therefore, you can see that, just like scenario 1 (fixed and random), scenario 7 in the random context caused overwhelming results for range, even though scenario 1 had the smallest range and scenario 7 had the largest range.

The $\hat{\alpha}$irregular grid, represented in figure 38, was not able to reach a median close to $\alpha = 1.31$radians. However, compared to the regular grid, the results were improved. In scenarios 4 and 7 with $v = 0.5$, the median, although overestimated α, was closer to the real value. This trend was observed in both the fixed and random effects, with the results in the random context being more promising. However, in both scenarios (4 and 7), the results showed specificity in relation to the median.

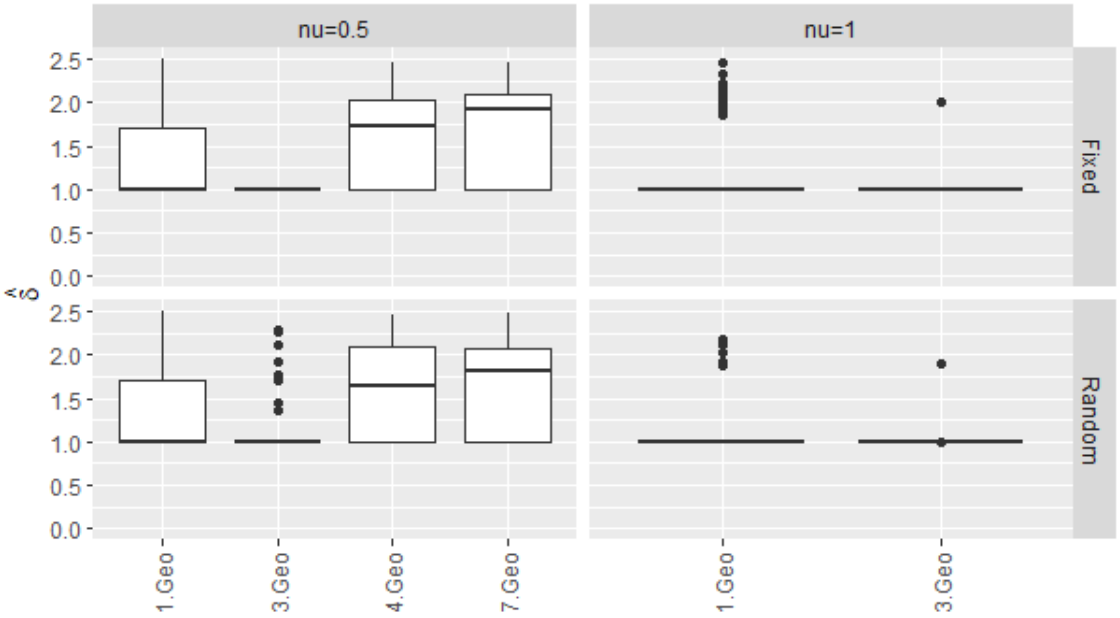
Figure 38: Estimated values without nuggets with anisotropy for the scenarios $\hat{\alpha}$chosen as best before rescaling, considering an irregular grid, and modeled with random effects.



In scenario 1 with $v = 0.5$, both in the fixed and random effects, the results were better compared to the regular grid. A part approaches the α. With $v = 1$, despite the median not being zero, the set of results was not able to reach $\hat{\alpha}$close to 0.5. However, the worst results remained in scenario 3 in all configurations, generally focusing on zero.

In $\hat{\delta}$this irregular grid, as well as $\hat{\alpha}$irregular, it presented better results than those obtained in the regular grid (Figure 39). However, in none of the scenarios was the median $\hat{\delta}$achieved $\delta = 2$. Note that scenarios 4 and 7, both in the fixed and random effect with $v = 0.5$, were the ones that came closest to the expected value for δ, highlighting the results in the context of the random effect as the most promising. The results of scenario 1 were positioned as the second best, but still demonstrated underestimation in relation to δ, while scenarios 3 in all configurations appeared to be the worst results, following a trend observed in the regular grid. Finally in scenario 1 with $v = 1$, proven in the worst result, since all results were concentrated in 1.

Figure 39: Estimated values without nuggets with anisotropy for the scenarios $\hat{\delta}$chosen as best before rescaling, considering an irregular grid, and modeled with random effects.



With the simulated data with anisotropic effect, we obtained convergence only in the GeoR software. This showed 100% convergence in all situations studied. However, note in Table 2 that the anisotropy angle was not always correctly estimated, with 50% of cases or more resulting in isotropic estimates ($\hat{\alpha} = 0$).

Table 2: The ERM for each scenario and model combination in GeoR package from the fouth study (anisotropic without nugget effect), given 100 data sets.

	True ν	Scenario	ν model as fixed				ν model as random				
			ERM				ERM				
			σ_s^2	ϕ	α	δ	σ_s^2	ϕ	α	δ	ν
Grid regular	0.5	1.	0.21	0.33	1.00	0.00	0.21	0.33	1.00	0.00	0.00
		3.	0.50	0.47	1.00	0.00	0.52	0.52	1.00	0.00	0.11
		4.	0.17	0.24	1.00	0.00	0.21	0.40	0.98	0.00	0.13
		7.	0.20	0.20	1.00	0.00	0.19	0.29	1.00	0.00	0.12
	1.0	1.	0.25	0.22	1.00	0.00	0.32	0.24	0.95	0.00	0.09
		3.	0.64	0.31	1.00	0.00	0.67	0.41	1.00	0.00	0.07
Grid irregular	0.5	1.	0.26	0.21	0.88	0.00	0.26	0.21	0.88	0.00	0.00
		3.	0.54	0.50	1.00	0.00	0.55	0.48	1.00	0.00	0.06
		4.	0.24	0.20	**0.11**	0.77	0.23	0.18	0.13	0.67	0.05
		7.	0.26	0.24	0.08	0.94	0.28	0.26	0.35	0.82	0.05
	1.0	1.	0.33	0.16	0.77	0.00	0.31	0.30	0.97	0.00	0.05
		3.	0,64	0,33	1,00	0,00	0,61	0,37	1,00	0,00	0,05

The comparison of simulations using GeoR and Asreml-R packages for mixed linear modeling, analyzing different scenarios with and without nugget effect on regular

and irregular grids, with fixed and random effects, revealed different patterns of behavior in relation to convergence, $\hat{v}$, $\hat{\sigma}_s^2$and $\hat{\phi}$. Overall, GeoR demonstrated a 100% convergence rate across all fixed-effects scenarios, while Asreml-R exhibited varying rates of convergence in these cases. With random effects, both packages presented varying convergence rates, but GeoR consistently maintained a rate of 100% in most cases.

A relevant point is the influence of changes in $\hat{v}$ on the heterogeneity of spatial parameters, especially with fixed effects. In a joint analysis of all parameters, the most favorable scenario with fixed and random effects and convergence rate was scenario 1 in both packages. When analyzing the scenarios generated for the irregular grid, without nugget effect, GeoR stood out for presenting a higher convergence rate and results closer to the observed (expected) values compared to Asreml-R. For a regular grid, $\hat{v}$ >= 1 had a negative impact on parameter estimation, especially when $v = 2$ and positive impacts when $v = 0.5$. The best-simulated scenarios were found in the smallest $\hat{\phi}$, mainly in scenario 1 for all estimated $\hat{v}$ values. As it $\hat{\phi}$ increased, a more pronounced underestimation was observed.

The comparison of the results of the models with regular and irregular grids without nugget effect revealed distinct patterns. The regular grid at $v = 2$ showed a greater number of successfully simulated scenarios than the irregular grid. With regard to $\hat{\sigma}_s^2$, all scenarios had medians close to the true (expected) σ_s^2 values in the regular grid, however, with regular grid a greater number of scenarios was successfully simulated compared to irregular grids. $\hat{\phi}$ was considerably underestimated in both regular and irregular grids. Therefore, the regular grid in scenario 1 was the best simulation generated from both packages.

Considering the nugget effect in a regular grid, GeoR stood out for its impressive convergence rate, especially for the results considering $v = 0.5$ as a fixed effect. With the nugget effect in an irregular grid, scenario 3 for $v = 0.5$ demonstrated the most solid results for GeoR with random effects. However, scenario 1 in both packages excelled due to more favorable median results, i. e., median results with smaller intervals.

The joint analysis of both configurations reveals that, in relation to $\hat{v}$, the scenarios in both the regular and irregular grids showed similar behavior, although outliers decreased in the regular grid. As for $\hat{\sigma}_s^2$, the most notable difference is that the irregular grid allowed more scenarios to reach the median of σ_s^2, especially at smaller ranges, while the regular grid facilitated estimation in more Asreml-R scenarios. In relation to $\hat{\phi}$, both

grids still underestimated, but the regular grid generated more scenarios in this condition than the irregular one.

Comparing the results with and without nugget effect in the different configurations, it is first observed the greater potential in the estimates using GeoR when compared to Asreml-R, considering the convergence rate, where GeoR demonstrated a notable consistency, reaching a rate of 100% in many scenarios, especially when it came to fixed effects. In contrast, Asreml-R exhibited varying rates of convergence, highlighting the importance of considering different packages when choosing the most appropriate model for a given analysis.

GeoR and Asreml -R packages for spatial analysis, GeoR stood out for several reasons. Firstly, its remarkable 100% convergence rate in many scenarios, especially when dealing with fixed effects, is a strong point. Meanwhile, Asreml -R showed varying rates of convergence, indicating different sensitivity across different model configurations.

GeoR stand out was the way it handled the parameter v, which directly influences the smoothing of the results. Variations in this parameter affected the dynamics of the estimates, and GeoR proved to be more stable and consistent when v it was smaller, producing results closer to the observed values. On the other hand, when v reached 2, GeoR was still able to maintain more regular estimates than Asreml-R. This superiority of GeoR can be attributed to specific algorithms, optimization strategies, or approaches best suited to dealing with linear mixed models and spatial analyses. It is important to highlight that the choice between packages may depend on the specific context of the analysis, the data structures, and the researcher's needs, but the results of this study highlight the effectiveness of GeoR in different scenarios.

The smoothing parameter v directly affects the variability of the spatial results, especially with fixed effects, because v determines the shape of the spatial correlation function. The correlation function describes how the similarity between two points in space varies according to the distance between them. In the present study, results closer to the real values were obtained when v it was smaller, while $v = 2$ resulted in less accurate estimates. When v is smaller, the correlation function is more irregular and capable of capturing abrupt variations in the spatial patterns in the data. This ability to adjust to sudden changes results in estimates that more accurately reflect the true patterns observed in spatial data.

On the other hand, when v = 2, the correlation function becomes smoother which can cause estimates to underestimate or miss the complexity in real spatial patterns, resulting in softer and sometimes less accurate estimates. Variations in v directly affect how the model interprets and models the spatial relationship between points. In conclusion, smaller values of v are more probable to result in estimates closer to the real values.

In experiments with regular and irregular grids, the introduction of the nugget effect resulted in differences in the observed improvements, for variance $\hat{\sigma}_s^2$ and $\hat{\phi}$. This discrepancy may be associated with the arrangement of the points and the presence of groupings in the grids. On the regular grid, where the points are arranged in a uniform and predictable way, the nugget effect seemed to be more relevant. Its ability to capture unstructured local variation was most evident in this homogeneous grid structure and contributed to more accurate estimates of unstructured variability ($\hat{\sigma}_s^2$). On the other hand, in the irregular grid, the uneven layout and the possible formation of clusters may have limited the corrective power of the nugget effect. Apparently, the nugget effect's ability to capture unmodeled variation in the data, adjust for residual error, and reduce bias in estimates is challenged more on the irregular grid, where the structure of the data is less predictable.

Another important aspect is the range, i.e., the distance at which points in space are correlated with each other. The range significantly influences how smoothly or quickly the correlation decreases as the distance between points increases. When the ranges are smaller, as in scenarios 1, 2 and 3 (3 m, 5 m and 10 m, respectively), the influence of neighboring points is only considered at shorter distances. This means that the model can capture more detailed and local spatial variations, as it is more sensitive to short-range fluctuations. On the other hand, in scenarios with greater ranges, such as 300, 500 and 1000 m (scenarios 7, 8 and 9), the influence of neighboring points extends over greater distances. This can lead to a smoothing of spatial patterns, where local variations are not as well represented, leading to an underestimation of model parameters.

However, the optimum range strongly depends on the modeling objective: to capture local fluctuations and specific spatial details, smaller ranges are recommended, while larger ranges can be useful to describe broader and smoother patterns in the spatial distribution of data.

The post-rescheduling scenarios were exclusively generated from Asreml-R, as most scenarios from this model faced convergence issues. Without the nugget effect,

compared to $\hat{v}$ after rescaling, although underestimated results still occur, these values were closer to v, especially in $v = 2$, where the main estimation challenges were encountered in Asreml-R before rescaling. This coincided with an increase in the rate of convergence. The $\hat{\sigma}_s^2$ also showed improvements with rescaling, mainly in the fixed effect, while in the random effect there was a reduction in the amplitude of the means of $\hat{\sigma}_s^2$, although the over- and underestimation behavior persisted. Underestimations $\hat{\phi}$ in some scenarios, such as 4 and 7, have decreased, but still persist. When analyzing the rescaling of the irregular grid, we observed that in relation to $\hat{v}$, the improvements in the median results were not significant, but there was a reduction in discrepant results. It then $\hat{\sigma}_s^2$ improved in homogeneity in some smaller range cases, but still faces problems at larger ranges, since underestimations remained at the larger ranges. Furthermore, $\hat{\sigma}_s^2$ showed a loss of information after rescheduling to $v = 0.5$, but surprisingly, $v = 1$ $v = 2$ showed improvements, $\hat{\phi}$despite still presenting underestimates in general for the generated scenarios, showed better results after rescheduling.

Regarding the analysis of the grid with irregular nugget after rescaling, there was a small overall reduction in the median of $\hat{v}$. However, when analyzing the scenarios, there was a performance gain in scenario 8. $\hat{\sigma}_s^2$, as well as in the rescheduling of the regular grid, suffered losses in the estimation of this parameter, especially in scenarios 5 and 7. In relation to $\hat{\phi}$, there was a loss in results after rescaling, with no change in underestimations in the largest $\hat{\phi}$.

The rescaling applied in scenarios 4 to 9, with different ranges, positively impacted the model parameter estimates, especially in the regular grid. Specifically, the more accurate approximation of $\hat{v}$, especially in situations where previous estimates were more problematic (as in $v = 2$), can be attributed to this technique. Because by adjusting the ranges to 3, 5 and 10 meters in the scenarios, the rescaling may have allowed a more adequate representation of spatial fluctuations in the models. This may have helped the model to more accurately capture the smoothness or roughness of the spatial correlation function, reflected by the parameter $\hat{v}$ as well as $\hat{\sigma}_s^2$ especially the fixed effects. Furthermore, the improvement in the model convergence rate after rescaling suggests that adjustments in ranges contributed to greater compatibility of the data with the model structure. This facilitated the convergence of the algorithms, allowing a better adaptation of the model to the spatial patterns present in the data.

However, the difficulties observed in the amplitudes of the means in the random effects after rescaling may indicate that some challenges in representing the patterns of variability in the random effects persisted. Mainly at distances adjusted to 5 and 10 meters, even after rescaling, where the largest underestimates in the model parameters remained. This may be due to persistent complexity in the spatial patterns present at these wider distances. The rescaling contributed to improving the model's adaptation to spatial patterns, favoring the convergence of the algorithms. However, some reasons may explain the persistence of underestimates. Such as larger distances that may contain more complex spatial patterns that were not fully represented even after rescaling. This complexity may be generating underestimates in the model parameters, indicating the need for more sophisticated modeling for these distances.

Another factor may be that range adjustments may not have captured in detail all spatial patterns present at greater distances. These patterns may require more granular analysis or finer adjustments to be correctly represented by the model. And also despite rescaling, there may still be specific local variations that have not been properly modeled. These variations may be influencing underestimations in model parameters at wider distances.

However, in the case of the irregular grid, the improvements after rescaling were more subtle. There was a small reduction in the median estimates of v, but without significant improvements. The estimates $\hat{\sigma}_s^2$suffered losses in scenarios with greater scope, maintaining underestimations, and the parameter$\hat{\phi}$ also faced challenges after rescheduling, with no improvements at higher values.

Furthermore, in some cases, rescheduling resulted in loss of information, especially in estimating $\hat{\sigma}_s^2$in specific scenarios. Surprisingly, there were improvements after rescaling in $v = 1$and $v = 2$, but less accuracy in $v = 0.5$. Even after rescheduling, certain scenarios, such as 5 and 7 in the irregular grid, continued to face persistent challenges in estimating $\hat{\sigma}_s^2$e $\hat{\phi}$. As previously mentioned in data without irregular grid rescaling, here in rescaling variations in spatial complexity and the distribution of points in the irregular grid may be affecting the effectiveness of rescaling in correcting underestimates and improving the accuracy of model parameter estimates. These implications highlight the complexity of this linear mixed model approach and the need to consider running simulations on their specific effects on spatial model results before applying to real field data.

To complete the anisotropy analysis without the nugget effect, it was generated in the GeoR model on a regular and irregular grid, since the Asreml-R model presented a singularity error. When considering only the best scenarios without rescaling, with a regular grid, it was observed that the parameters $\hat{v}$both with $v = 0.5$and $v = 1$showed improvements in relation to the absence of anisotropy. The same pattern was noticed for a $\hat{\sigma}_s^2$, however only in scenario 1. Subsequently, those $\hat{\sigma}_s^2$with $v = 0.5$showed better performance than without anisotropy in scenarios 1, 4 and 7, while in scenario 3 underestimations were obtained. Both do $\hat{a}$and a $\hat{\gamma}$, in relation to the values α and γ, were distant. The irregular grid, the results of $\hat{v}$were more uniform compared to the regular grid, while the $\hat{\sigma}_s^2$and $\hat{\phi}$showed little change. Surprisingly, o $\hat{a}$and a $\hat{\gamma}$, although still underestimated, came closer to the values α and γ compared to the regular grid

In spatial anisotropy models, the $\hat{a}$and a $\hat{\gamma}$, of anisotropy are fundamental parameters that play important roles in describing the direction and extent of spatial variations in the data. When these values do not correspond to those expected, as in the case of a regular grid, it may indicate inadequacies in the representation of spatial patterns. Because O $\hat{a}$ of anisotropy, which in this case would be 1.31 radians or 75 degrees, is a measure of the preferred direction of spatial variation. And since the results do not match the expected value, this may suggest that the predominant direction of variations in the data is not being captured well by the model. This can lead to distortions in the representation of spatial patterns, resulting in inaccurate estimates of model parameters.

Likewise, the anisotropy ratio, in this case equal to 2, refers to the extent of variation in one direction compared to the perpendicular direction. As the results differ from the expected value, they indicate that there is a distortion in the relationship between the extent of these variations. This resulted in underestimation of the true extent of spatial patterns in the data. As for $\hat{\sigma}_s^2$and $\hat{v}$mainly present results close to expected in scenario 1, even with the angle and anisotropy ratio presenting results different from those expected, this may indicate a compensation in the modeling effects. Because these parameters in terms of linear mixed models are fundamental in describing fluctuations and the smoothness of spatial variations. Therefore, even if the anisotropy $\hat{a}$and a $\hat{\gamma}$do not correspond to the expected values, the model may be adjusting to compensate for these discrepancies. This may happen due to complexities in spatial patterns.

Finally, the intriguing discovery about anisotropy in the irregular grid is surprising because, despite the difficulties faced in other aspects of modeling in this specific type of point configuration, the model demonstrated a better ability to adjust or get closer to the directions and extensions real spatial variations. This situation may be related to different factors, such as the unique sensitivity of the model to non-uniform patterns present in the irregular grid. Perhaps the model has greater adaptability to capture the nuances of anisotropy, even in the face of an irregular configuration of sample points.

Another possible explanation is that the distribution of points on the irregular grid may have favored a more accurate representation of anisotropy, aligning in a more compatible way with the model structure for this specific characteristic. It is important to consider that anisotropy, in some cases, may be less complex to model than other forms of spatial variation. This relative simplicity in the representation of anisotropy may have allowed a closer adjustment to the real values, even in a challenging context such as the irregular grid. So, these possibilities suggest that the model may have a special ability to deal with certain aspects of anisotropy, even when other areas of spatial modeling face difficulties in irregular grid contexts.

Among others, Cullis and Gleeson (1991), Zimmerman and Harville (1991), Martin (1990) and Cressie and Hartfield (1996), discussed spatial modeling methods in field experiments, through the evaluation of different approaches to the analysis of variation range in agricultural experiments. However, the only study directly comparable to this study is that of Haskard (2007). Comparing that study with the presented results (without the nugget effect), frequent convergence difficulties were identified in both, especially with the lowest value of ν. The study by Clifford (2005), which aimed to analyze the nature of the spatial correlation of agricultural crops, corroborate our results, since the authors consistently indicated that the smoothness parameter tends to be small in the best estimates. Dutta (2016) found reasonable estimates (close to the true values), when v is close to the true value.

Minasny and McBratney (2005) highlight that for values of v between 0.25 and 0.50 the spatial behavior of the soil data is less smooth and more unstable, because the differences between neighboring areas are more pronounced, as opposed to a more uniform or gradual pattern.

Haskard (2007) also reported difficulties in relation to the convergence rate and the underestimation of certain parameters, such as $\hat{\sigma}_s^2$and $\hat{\phi}$, indicating challenges in

modeling more complex spatial processes. Gilmour, Cullis and Verbyla (1997) conclude that although there is no single model that adequately adjusts to all field experiments, the autoregressive model is predominant. However, they frequently noted the presence of additional identifiable variation, suggesting that components other than the main model could be influencing results in specific experiments. These nugget-free scenarios provide a more robust and more precise understanding of the analyzed phenomenon although being the simplest, as they allow the incorporation of uncertainty in the estimation of the covariance structure according to Handcock and Wallis (2012) through parameters such as range and spatial smoothing.

For Haskard (2007), the insertion of the nugget effect caused significant convergence difficulties in the analyzed models. In this study, the introduction of the nugget effect also had an impact on the estimates, especially for parameters such as $\hat{\sigma}_s^2$ and $\hat{\phi}$, of the irregular grid. Haskard (2007) also inferred that the arrangement of points on the grids, whether regular or irregular, influenced the effectiveness of the nugget effect.

For this reason, Goovaerts (1998) suggested adjusting a small nugget effect, as the model becomes less sensitive to extreme fluctuations between observation points, making estimates more stable. Furthermore, according to SLAETS et al. (2019), in mixed models, which are often used in spatial data analysis, the inclusion of this nugget effect helps to capture variation not explained by the model's fixed or random factors. This improves the accuracy of estimates and reduces the influence of extreme points on the results. However, it tends to increase the complexity of the model.

Both in this study and in that of Haskard (2007), regular sampling schemes with smaller spacing tend to improve the success rate in model convergence, in addition to providing less biased estimates. Furthermore, a fact that was not raised in this study, but was proposed in the study by Marchant and Lark (2007), is the importance of developing sampling schemes testing different variogram models, however in the study of these authors the Matérn model in a regular grid it surpasses the spherical and exponential. This is relevant because by designing sampling schemes that better align with the flexible nature of the variogram, it is possible to capture a greater variety of spatial variability behaviors. Minasny and McBratney (2005) also explore the application of the Matérn function as a general model for variograms of soil attributes and make it clear that it is a suitable model as it allows describing several isotropic spatial processes and using appropriate techniques to estimate their parameters.

Guttorp and Gnetting (2006) address another interesting point also found in this study, although the authors focused on a more historical part of the Matérn family of spatial correlation functions. They discuss specifying the correlation between the values of a random spatial function at locations separated by a lag vector. In other words, as the distance between points increases, a decrease in the accuracy of the prediction model may occur. In the same sense, the practical study by Clark and Stefanova (2011) further highlights the issue of experimental rearrangement as they suggest that it is not advisable to use a high frequency of control plots (for example, more than one in every five plots) unless the size of the experiment can be increased proportionately.

Clark and Stefanova (2011) as well as Cullis et al. (2006) highlight the ability of linear models to consider both fixed and random trends in the field, going beyond the simple average model. The substantial advantage of these models is that it is not limited to just declaring the variety terms as random or fixed. Instead, it is not just a matter of labeling variables as random or fixed, but rather the overall approach to how the model encompasses and handles the complexities and variations within the data.

The reported studies point out to further aspects not included in the present study: Stefanova Smith and Cullis (2009) carried out formal tests to identify discrepant points (outliers) in linear mixed models that directly affect the estimate, however, identifying outliers in statistical models is a challenge, especially in linear mixed models, due to the complexity of dealing with multiple levels of randomness and correlation between data.

The simulation study by Haskard (2007) was applied to real potassium data by Haskard, Rawlins and Lark (2010), who were able to verify that random models proved to be more effective in capturing the complexity of variation of the potassium content in the soil, especially when compared to fixed ones. But Haskard, Rawlins and Lark (2010) point out that accurate selection of adequate and model settings is still a challenging area and requires more robust evaluation and comparison methods to determine the most appropriate model to predict variation in validation data.

CONCLUSIONS

This simulation study concluded that GeoR performed superiorly in several situations, showing a consistent rate of convergence, especially with fixed effects. In contrast, Asreml-R demonstrated greater variability and challenges in certain scenarios. Furthermore, the influence of the smoothing parameter v on the model estimates was evident. GeoR proved to be more stable and consistent, especially when v was smaller,

producing estimates closer to the observed values. This sensitivity of the model to variation in v is essential to understanding spatial modeling and its implications for the interpretation of observed patterns.

The introduction of the nugget effect, besides considerably increased data processing time, resulted in differences in observed gains, especially in relation to the ability to capture unstructured variation. In regular grids, the nugget effect appeared more effective due to the uniform arrangement of sampling points, better capturing unstructured variation. In irregular grids, the uneven arrangement and formation of clusters limited the effectiveness of the nugget effect, compromising the accuracy of the estimation. The practical application of these models is crucial, as the estimates obtained not only influence decision-makers in various areas, but also serve as a basis for the development of more accurate models in the future. Furthermore, future research suggests exploring other models and packages, investigating anisotropy in regular and irregular contexts and with a nugget effect, and validating these results using real data sets. Because the estimates derived from these models are essential not only for understanding spatial variability, but also for guiding practical decisions and improving the accuracy of future models and forecasts, the Matérn family represents a significant advance in understanding the complexity of spatial patterns.

REFERENCES

AYELE, Gebiaw T. *et al*. Relationship of Attributes of Soil and Topography with Land Cover Change in the Rift Valley Basin of Ethiopia. **Remote Sensing**, [S.L.], v. 14, n. 14, p. 1-21, 6 jul. 2022. MDPI AG. http://dx.doi.org/10.3390/rs14143257.

BITENCOURT, Dioni Glei Bonini et al. Multivariate and geostatistical analyses to evaluate lowland soil levelling effects on physico-chemical properties. **Soil And Tillage Research**, [S.L.], v. 156, p. 63-73, mar. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.still.2015.10.004.

BITENCOURT, Dioni Glei Bonini et al. Spatial variability structure of the surface layer attributes of gleysols from the coastal plain of Rio Grande do Sul. **Bioscience Journal**, [S.L.], v. 31, n. 6, p. 1711-1721, 2015. EDUFU - Editora da Universidade Federal de Uberlandia. http://dx.doi.org/10.14393/bj-v31n6a2015-29457.

BROWN, Richard D.; BARDSLEY, Johnathan M.; CUI, Tiangang. Semivariogram methods for modeling Whittle-Matérn priors in Bayesian inverse problems. **Arxiv**, [S.L.], v. 1, n. 1, p. 1-30, 2018. ArXiv. http://dx.doi.org/10.48550/ARXIV.1811.09446.

BUTLER D (2022). _asreml: Fits the Linear Mixed Model_. R package. version 4.1.0.176, Disponível em: <www.vsni.co.uk>. Acessado em: 19 de dez. 2022.

CHEN, Longlong *et al*. Probabilistic assessment of slope failure considering anisotropic spatial variability of soil properties. **Geoscience Frontiers**, [S.L.], v. 13, n. 3, p. 1-21, maio 2022. Elsevier BV. http://dx.doi.org/10.1016/j.gsf.2022.101371.

CLARK, G. P. Y.; STEFANOVA, K. T. Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines: Unreplicated and partially replicated designs. **Australian & New Zealand journal of statistics**, v. 53, n. 4, p. 461–480, 2011. https://doi.org/10.1111/j.1467-842X.2011.00642.x

CLARK, G. Peter Y.; STEFANOVA, Katia T. Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines**. Australian & New Zealand Journal of Statistics**, [S.L.], v. 53, n. 4, p. 461-480, dez. 2011. Wiley. http://dx.doi.org/10.1111/j.1467-842x.2011.00642.x.

CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of agricultural, biological, and environmental statistics**, v. 11, n. 4, p. 381–393, 2006. https://doi.org/10.1198/108571106x154443

DIGGLE, Peter J., RIBEIRO, Paulo J. Model-based Geostatistics. Springer New York, NY. 2007. p. 237. Doi: https://doi.org/10.1007/978-0-387-48536-2.

DUTTA, Somak; MONDAL, Debashis. REML estimation with intrinsic Matérn dependence in the spatial linear mixed model. Electronic Journal of Statistics, [S.L.], v. 10, n. 2, p. 1-38, 1 jan. 2016. Institute of Mathematical Statistics. http://dx.doi.org/10.1214/16-ejs1125.

GILMOUR, Arthur R.; CULLIS, Brian R.; VERBYLA, Arūnas P.; VERBYLA, Arunas P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. **Journal Of Agricultural, Biological, And Environmental Statistics**, [S.L.], v. 2, n. 3, p. 269-293, set. 1997. Springer Science and Business Media LLC. http://dx.doi.org/10.2307/1400446.

GOOVAERTS, P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. **Biology and fertility of soils**, v. 27, n. 4, p. 315–334, 1998. https://doi.org/10.1007/s003740050439

GUTTORP, P.; GNEITING, T. Studies in the history of probability and statistics XLIX on the Matérn correlation family. **Biometrika**, v. 93, n. 4, p. 989–995, 2006.

HASKARD, K. A.; RAWLINS, B. G.; LARK, R. M. A linear mixed model, with non-stationary mean and covariance, for soil potassium based on gamma radiometry. **Biogeosciences**, [S.L.], v. 7, n. 7, p. 2081-2089, 2 jul. 2010. Copernicus GmbH. http://dx.doi.org/10.5194/bg-7-2081-2010.

HASKARD, K.A. **An anisotropic Matern spatial covariance model**: REML estimation and properties. Adelaide, 2007. Ph.D. Thesis. 197f. School of Agriculture and Wine Biometrics. The University of Adelaide. Adelaide, 2007. Australia.

HASKARD, Kathryn A.; CULLIS, Brian R.; VERBYLA, Arūnas P.. Anisotropic Matérn correlation and spatial prediction using REML. **Journal Of Agricultural, Biological, And Environmental Statistics**, [S.L.], v. 12, n. 2, p. 147-160, jun. 2007. Springer Science and Business Media LLC. http://dx.doi.org/10.1198/108571107x196004.

KALANTARI, A.R. *et al*. Effect of spatial variability of soil properties and geostatistical conditional simulation on reliability characteristics and critical slip surfaces of soil

slopes. **Transportation Geotechnics**, [S.L.], v. 39, n. 1, p. 1-13, mar. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.trgeo.2023.100933.

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. v. 57, n. 6, p. 787–799, 1 dez. 2006. http://dx.doi.org/10.1111/j.1365-2389.2005.00768.x

MAIER, Martin *et al*. Introduction of a guideline for measurements of greenhouse gas fluxes from soils using non-steady-state chambers. **Journal of Plant Nutrition and Soil Science**, [S.L.], v. 185, n. 4, p. 447-461, 29 jun. 2022. Wiley. http://dx.doi.org/10.1002/jpln.202200199.

MARCHANT, B.P.; LARK, R.M. The Matérn variogram model: implications for uncertainty propagation and sampling in geostatistical surveys. **Geoderma**, [S.L.], v. 140, n. 4, p. 337-345, ago. 2007. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2007.04.016.

MARTIN, R. J. The use of time-series models and methods in the analysis of agricultural field trials. **Communications in Statistics - Theory and Methods**, v. 19, n. 1, p. 55–81, jan. 1990. http://dx.doi.org/10.1080/03610929008830187

MCCULLAGH, P.; CLIFFORD, D. Evidence for conformal invariance of crop yields. **Proceedings. Mathematical, physical, and engineering sciences**, v. 462, n. 2071, p. 2119–2143, 2006. https://doi.org/10.1098/rspa.2006.1667

MCCULLAGH, Peter; CLIFFORD, David. Evidence for conformal invariance of crop yields. **Proceedings of the Royal Society A**: Mathematical, Physical and Engineering Sciences, [S.L.], v. 462, n. 2071, p. 2119-2143, 28 fev. 2006. The Royal Society. http://dx.doi.org/10.1098/rspa.2006.1667.

MINASNY, Budiman; MCBRATNEY, Alex. B. The Matérn function as a general model for soil variograms. **Geoderma**, [S.L.], v. 128, n. 3-4, p. 192-207, out. 2005. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2005.04.003.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: URL https://www.R-project.org/. Acessado em: 19 de dez. 2022.

RIBEIRO JR PJ, DIGGLE PJ, SCHLATHER M, BIVAND R, RIPLEY B (2020). geoR: Analysis of Geostatistical Data_. R package version 1.8-1, Disponível em: <https://CRAN.R-project.org/package=geoR>. Acessado em: 19 de dez. 2022.

SLAETS, J. I. F.; BOEDDINGHAUS, R. S.; PIEPHO, H.-P. Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin? **European journal of soil science**, v. 72, n. 1, p. 47–68, 2021. https://doi.org/10.1111/ejss.12976

SOLANKI, Bobby; *et al.* Structure and Electronic Properties of Some Transition Metal Sulfides Nanomaterials: a density functional theory study. **Engineered Science**, [S.L.], v. 19, n. 1, p. 83-88, 2022. Engineered Science Publisher. http://dx.doi.org/10.30919/es8d684.

STEFANOVA, Katia T.; SMITH, Alison B.; CULLIS, Brian R.. Enhanced diagnostics for the spatial analysis of field trials. **Journal Of Agricultural, Biological, And Environmental Statistics**, [S.L.], v. 14, n. 4, p. 392-410, dez. 2009. Springer Science and Business Media LLC. http://dx.doi.org/10.1198/jabes.2009.07098.

WEBSTER, R., OLIVER, MA, **Geoestatística para Meio Ambiente Cientistas**. John Wiley & Sons, Chichester. 2001.

YU, Wu *et al*. High-resolution mapping and driving factors of soil erodibility in southeastern Tibet. **Catena**, [S.L.], v. 220, n. 1, p. 1-11, jan. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2022.106725.

ZHOU, Haizuo *et al*. Quantitative bearing capacity assessment of strip footings adjacent to two-layered slopes considering spatial soil variability. **Acta Geotechnica**, [S.L.], v. 1,

n. 1, p. 1-12, 14 abr. 2023. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11440-023-01875-8.

ZIMMERMAN, D. L.; HARVILLE, D. A. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. **Biometrics**, v. 47, n. 1, p. 223, mar. 1991. http://dx.doi.org/10.2307/2532508.

4.2 CAPÍTULO 2

ASSESSING THE VARIABILITY OF HYDRAULIC CONDUCTIVITY OF SATURATED SOIL AND ENVIRONMENTAL DRIVERS AT MULTIPLE SCALES THROUGH 2D-WAVELET ANALYSIS

A modified version of this manuscript is in the review process of a peer-review in Catena Journal.

Abstract: The study focuses on soil hydrology properties, particularly saturated soil hydraulic conductivity (Ksat), which is crucial for understanding various hydrological processes, supporting decision-making in water resource conservation, and enhancing hydrological models. Ksat is known for its high variability in the field, leading to uncertainties in hydrological model outputs. The research aims to assess and quantify the spatial relationships between Ksat and environmental attributes at the watershed scale using CWT-2D and geostatistical methods. The study area is the Ellert Creek Watershed (ECW) in Southern Brazil, with a mesothermic climate and various soil types. Soil attributes were collected through sampling, including Ksat, soil water retention curve, soil bulk density, and macroporosity. Land-use maps were created using satellite images, and Random Forest classification was applied. Descriptive statistics, geostatistical modeling (Sequential Gaussian Simulation), and CWT-2D were employed to analyze the spatial variability of Ksat and its co-driven factors. Results indicated high variability in Ksat, with the 75m scale being the most representative according to the Wavelet transform. The study also considered the influence of land use on soil properties, noting the impact of different land-use classes on Ksat, macroporosity, and soil bulk density. The research emphasizes the importance of selecting appropriate scales for analysis and understanding the spatial relationships between soil properties and environmental factors. The findings contribute to improving the characterization of soil hydrological properties, essential for effective water resource management in watersheds.

Keywords: Saturated Hydraulic Conductivity (Ksat). Spatial Variability. CWT-2D Analysis. Scale. Land use.

1. INTRODUCTION

Soil hydrology properties, mainly saturated soil hydraulic conductivity (Ksat), are key soil properties for the understanding of many superficial hydrological processes (runoff, infiltration, among others), (BERTHELIN et al., 2023; GAJIć et al., 2023; WELDEGEBRIEL et al., 2023), for the use of hydrological models and for supporting decision makers on water resources conservation and management at the watershed scale. Ksat is a highly variable soil hydrology property in the field and its measurements (or values) can produce high uncertainties about the outputs of hydrological models (MORRIS et al., 2022; MBAYAKI; KARUKU, 2022; PÁEZ-BIMOS et al., 2022).

Ksat is a highly variable soil hydrological property in the field and its measurements (or values) can produce large uncertainties on the results of hydrological models (HU et al., 2015). Therefore, many studies have focused on quantifying the Ksat spatial variability under different climate and hydrological conditions and at different spatial scales. (CENTENO et al., 2020; SOARES et al., 2023; TANG et al., 2023; TENG et al., 2023; ZHANG et al., 2023). The acquisition of detailed information about the Ksat spatial variability in the field is time-consuming and expensive, at the watershed scale (CENTENO et al., 2020; SOARES et al., 2023).

In this sense, pedotransfer functions (PTFs) have been developed to estimate Ksat by using less time-consuming and expensive attributes which are more easily available in soil databases (ALBALASMEH; et al., 2022; WELDEGEBRIEL et al., 2023; BASSET et al., 2023). However, most of the developed PTFs are constructed using simple or multiple linear models which assume that the data sets are randomly distributed over the study area (as examples, SCHAAP et al., 2001; OTTONI et al., 2019), ignoring the spatial covariance of Ksat and its co-regionalized underlying influence factors.

Geostatistical methods have been used to quantify and characterize the spatial variability structure of Ksat and its co-driven factors at the field (WANG et al., 2023; ZAHEDIFAR et al., 2023) and watershed (SOARES et al., 2023; KE et al., 2023; FLAMMINI et al., 2023) scales. In addition. geostatistical simulations allow us to quantify the uncertainty associated with these variables, which often consist of multiple related variables that have complex relationships (HADJIPETROU; MARIETHOZ; KYRIAKIDIS, 2023; KARACAN; ERTEN; MARTÍN-FERNÁNDEZ, 2023). Nevertheless, the most used geostatistical methods require stationary data sets, i.e., the mean and variance of the data set to over space (LI et al., 2023; LIU; PRODANOVIć; PYRCZ, 2022)

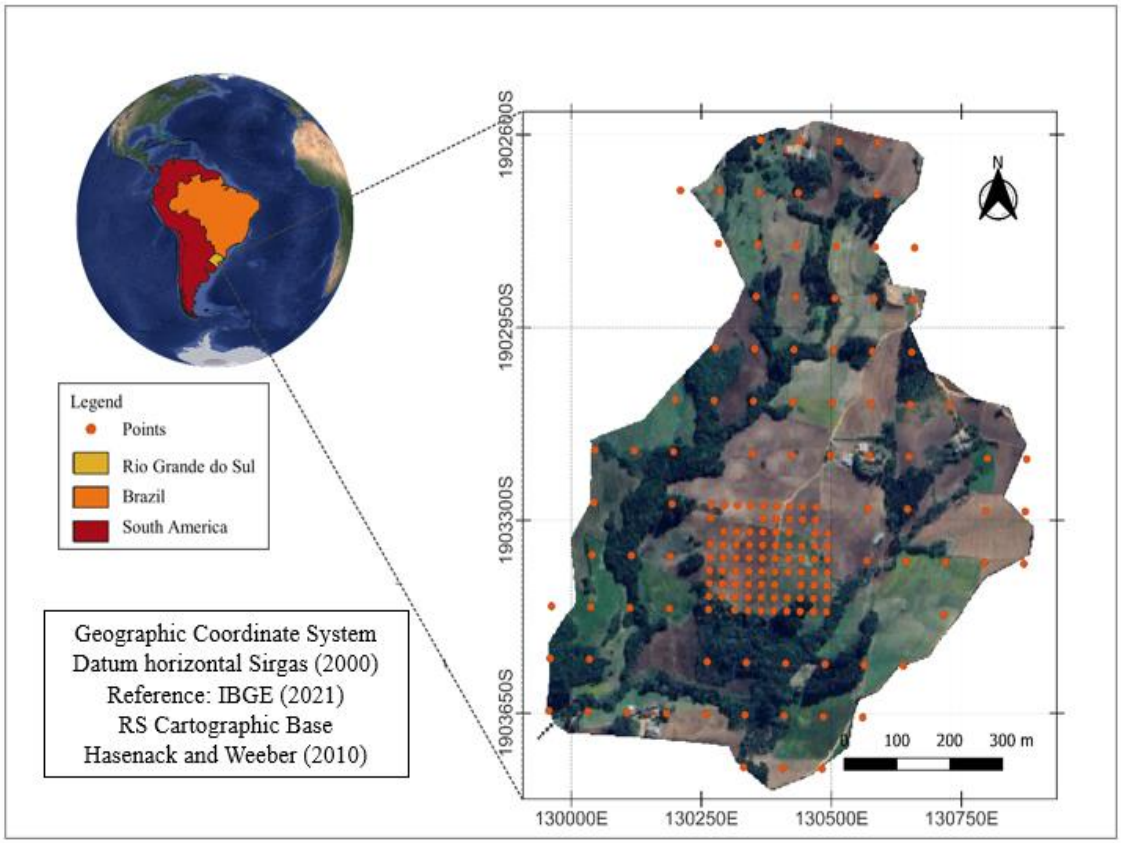
To overcome this drawback, statistical methods such as 2D-Wavelet analysis have been used to examine both stationary and non-stationary natural trends of the soil hydro-physical attributes (SOUZA; SÁ; DOMINGUES, 2013; DOMINGUES et al., 2016), and in conjunction with geostatistical tools can reveal important localized or transient features of the spatial variability of Ksat and its co-driven regionalized factors. Therefore, the objective of the study was to assess and quantify the spatial relationships between Ksat and its co-driven environmental attributes at watershed scale using the CWT-2D and geostatistical methods.

2 MATERIALS AND METHODS

2.1 Study area

The study was carried out in a sub-basin called Ellert Creek Watershed (ECW), located in the municipality of Canguçu, Southern Brazil (Fig. 1), which has an area of approximately 80 ha. The ECW's main watercourse flows directly into the Pelotas River Watershed, which has a total area of approximately 940 km^2 and is strategic for the economic and social development of the Southern Brazil. The soils are shallow, and the relief varies from undulated to strongly undulated (Soares et al., 2020). The main soil type in the ECW is the Entisol according to the US Soil Taxonomy (Soil Survey Staff, 2010.

Figure 1: ECW location map, containing the sampling points.



The climate of the study area is mesothermic, indicating subtropical humid conditions and characterized by a mean annual temperature of 18°C with hot summers and cold winters, which is classified as Cfa according to the Köppen climate classification (Alvares et al., 2014). The annual precipitation varies between 1300 and 1600 mm, being uniformly distributed throughout the year (Alvares et al., 2014).

2.2. Soil Attributes

Soil sampling was carried out by Soares et al. (2020) under a sampling grid of 75 m x 100 m, covering the entire watershed and totaling 106 sampling points (Fig. 1). Subsequently, a second sampling campaign was performed, collecting at other 78 sampling points in a grid of 25 m x 25 m to achieve a higher sampling resolution. Therefore, 184 experimental points were sampled (Fig. 1).

At each sampling point, undisturbed soil samples were collected in the 0-0.20 m soil layer using volumetric rings (5.0 cm in height and 4.8 cm in diameter) to determine the saturated soil hydraulic conductivity (Ksat) using the head constant permeameter (Klute and Dirksen, 1986), soil water retention curve (soil water content values hold at

matric potentials of -1 kPa, -2 kPa, -6 kPa, -10 kPa, -33 kPa, -100 kPa, -300 kPa and -1,500 kPa) and soil bulk density (BD) (Klute, 1986). Soil macroporosity (MP) at each experimental point was calculated considering the soil water content at the matric potential of -6 kPa.

2.3 Land-use map

To construct the land use map, an image was obtained from 05/08/2020, referring to the CBERS 4A/WPM satellite/sensor, through the official website of the National Institute for Space Research (INPE, 2023). All samples collected were polygons, in *shapefile forma*, created using the QGIS.22 *software* based on the analysis of the acquired image, using *Google Earth images* and previous knowledge of the area as references. A supervised classification was carried out using the Dzetsaka: *Classification Tool add-on, using the Random Forest* classifier in the QGIS 3.22.16 *software*. The *Random Fores- t* classifier is a special tool in machine learning scenarios as it is based on the iterative and random creation of decision trees that use a set of class-defining rules and conditions to randomly create multiple decision trees based in the created model. These trees are used to classify all pixels (ZERMANE et al., 2023).

Subsequently, with the resulting classifications, after the transformation from *Raster* to vector, analyzes were carried out, with the aim of calculating and obtaining the areas of each pattern of the different classes of land use and cover. This procedure was done with QGIS 3.22.16 using a *software algorithm* that returns the area and count of each unique value in layer data, that is, the algorithm calculates the number of pixels for each class and subsequently multiplies it by the area from each pixel, the area, in m^2, for each land use class is obtained. It is important to highlight that all mappings produced followed the coordinate reference system with the Universal Transverse Mercator projection (UTM) and the SIRGAS 2000 Datum – Zone 22S.

The identification of land use and cover classes was carried out based on the analysis of the features present in the study area. Considering that the images used for interpretation come from *Wide sensors Scan Multispectral and Panchromatic Camera* (WPM) of the CBERS 4A Satellite, with a spatial resolution of 2 meters, using bands 2, 3 and 4 with 10m color composition 342. Therefore, resulting from the needs of this study and as previously carried out by Soares et al. (2020), ECW was classified into permanent cultivation, forestry (*Pinus sp., Eucalyptus Sp. and Acácia sp*.), native forest and pasture.

Finally, in the last stage, the accuracy of the supervised classification was verified, using the *Acatama complement.* In this tool, in the *Stratified Random Sampling tab*, after selecting the classified image, 396 random points were generated proportional to the size of the study area, to be checked and analyzed by comparing the points with the aid of the Google Satellite image and the points collected in the field and even the RGB 321 color composition. Once the verification is done, the plugin defines an accuracy value, determining whether the supervised classification has a good level of accuracy or not, with the closer to 1 the better (CONGALTON, 1991; STEHMAN, 2014; CONGALTON; GREEN, 2018).

2.4 Descriptive statistics

Mean, standard deviation, maximum and minimum values, and coefficient of variation (CV) were calculated for each variable of the data set. The CVs were classified according to Wilding and Drees (1983): CV≤ 15% -low variability of the data set around their mean; 15% < CV ≤ 35% -moderate variability; and CV > 35% -high variability. All sets of data were submitted to the Kolmogorov-Smirnov (K-S) test (Massey, 1951), at the 5% significance level. A lognormal transformation was applied to those data sets that did not show normality. The Spearman rank correlation coefficient was calculated to evaluate the relationship among Ksat and the other variables. The R statistical software (R core Team, 2016) was used for these calculations.

2.5 Geostatistical modeling

Sequential Gaussian simulation is a well-established geostatistical approach to quantifying spatial uncertainty of sparse data sampled from the subsurface. However, to simulate multivariate datasets with complex relationships, prior steps must be taken to preserve those correlations as well as their spatial characteristics.

For multivariate datasets, most geostatistical simulation techniques rely on the assumption of a multivariate Gaussian distribution. For this reason, it is recommended to use the Projection Pursuit Multivariate Transform (PPMT) approach (Barnett et al., 2014, 2016; Manchuk et al., 2017) to model spatial variables that have complex relationships (Bassani et al., 2018, Kloeckner et at. 2021). The method is briefly explained as follows:

(i) The method forces the variables to a univariate standard Gaussian distribution via a normal score transformation (mean equal zero and standard deviation equal one) (Deutsch and Journel, 1998).

(ii) The new data is transformed using a method called data sphering, which consists of rotating the normal score variable units to principal components (with a standardized variance). These scaled principal components are rotated back to the original axes of the normal score variables. Despite these variables are uncorrelated and univariate Gaussian, they remain presenting multivariate relationships.

(iii) The method searches for the most non-Gaussian data's projection using a projection index measurement.

(iv) A normal score transformation is applied to the most non-Gaussian projection.

(v) Iterate steps 3 and 4 until the data presents a multi-Gaussian distribution and variables are independent using stopping criteria (bootstrap sampling).

In summary, the workflow consists of transforming the original variables from the dataset into PPMT variables, simulating them independently, and back transforming them to the original space. Because the variables are independent and uncorrelated, we can simulate each transformed PPMT variable using sequential Gaussian simulation (sGs). 40 realizations (simulated models which reproduce the spatial characteristics of the data) were obtained through sGs.

2.6 Two-Dimensional Wavelet Transform

Two-Dimensional Wavelet Transform was applied, adopting the Morse *wavelet on the simulated grid with a scale of 5m.* To identify the best scale for estimating the variability of the hydraulic conductivity of saturated soil, the 2-D Continuous Wavelet Transform (CWT 2-D) of Morlet was applied, which is a representation of 2-D data in 4 variables: dilation, rotation, and position. Dilation and rotation are real-valued scalars and position is a 2-D vector with real-valued elements. Let x be a two-element vector of real numbers. If $f(x) \in L2(\mathbb{R}2)$, is square-integrable in the plane, the CWT-2D is defined as:

$$\mathrm{WT}_f(a,b,\theta) = \int_{\mathbb{R}^2} f(x)\frac{1}{a}\overline{\psi}(r_{-\theta}(\frac{x-b}{a}))\,dx \quad a \in \mathbb{R}^+,\ x,b \in \mathbb{R}^2$$

where the slash denotes the complex conjugate and rθ is the 2-D rotation matrix,

$$r_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad \theta \in [0, 2\pi)$$

.

The CWT-2D is a spatially scaled representation of data. In this sense, one can view the inverse of the scale and the rotation angle together as a spatial frequency variable, which gives the CWT-2D an interpretation as a spatial frequency representation. For all admissible 2-D Wavelets, the CWT-2D acts as a local filter for a scaled and positioned image. If the wavelet is isotropic, there is no angle dependence in the analysis.

The Morlet wavelet is an example of an isotropic wavelet. Isotropic wavelets are suitable for point analysis of images. If the wavelet is anisotropic, there is an angle dependence in the analysis, and the CWT-2D acts as a local filter for an image in scale, position, and angle. To calculate these transforms, the MATLAB R2022B *software* was used. More information about the methodology can be found in the study by Domingues et al. (2016) and Souza; Sa; Domingues (2013).

3. RESULTS AND DISCUSSION

3.1 Exploratory Analyzes

The summary statistics of soil saturated hydraulic conductivity (Ksat) and the other two attributes associated with macroporosity and soil density, for 178 points, are presented in Table 1. According to the classification of Wilding and Drees (1983), the coefficients of variation (CV) for Ds can be considered low (CV ≤ 15%), while the CV value for Mac is considered moderate (15% < CV ≤ 35%). And Ksat varied across the study area with a CV of 99% indicating high variability (CV > 35%). The high spatial variability of Ksat is well known, regardless of measurement methodology, geographic location, land use, soil type or scale (CENTENO et al., 2020; SOARES et al., 2023).

Table 1: Descriptive statistical analysis of ECW

Attribute	Samples	Average	Median	Standard deviation	CV (%)	Asymmetry	Kurtosis
Ksat	179	0.81	0.52	0.8	99.4	1.38	1.46
Macro	179	20.6	20.4	5.05	24.3	1.17	4.1
SBD	179	1.41	1.42	0.17	12.4	1.61	1.61

* Ksat: Hydraulic Conductivity of Saturated Soil; Macro: Macroporosity and SBD: Soil Apparent Density. CV: Coefficient of variation.

3.2 Sequential Gaussian Simulation (SGS)

The geostatistical models were validated through a visual check, correlation matrices, scatter plots with kernel density for bivariate relationships, variograms, and histograms. The models presented satisfactory results. Figure 2 shows the first realizations map generated by SGS for all variables on a regular grid.

Figure 2: First realization maps of the simulated variables. (a) ksat, (b) Macro and (c) BDS.

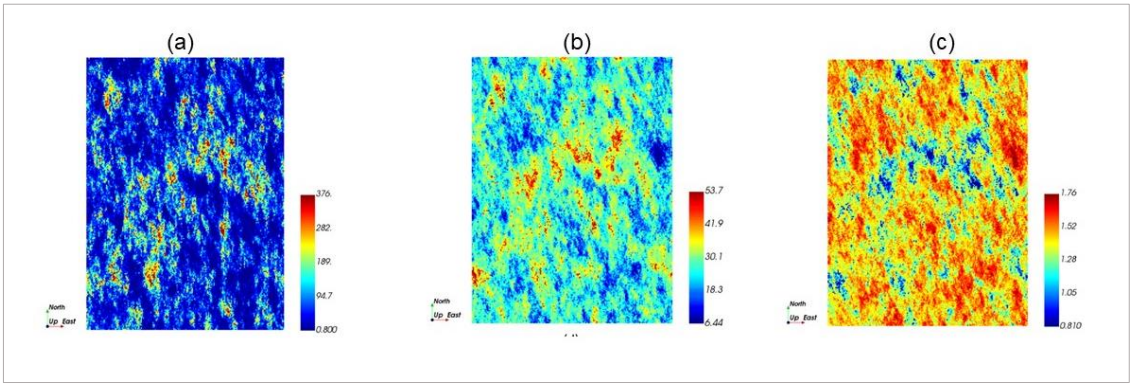


Table 2 shows two correlation coefficient matrices. The first one, matrix a, shows the correlation coefficient between each variable from the dataset. The second matrix b shows the correlation coefficient between each variable from the first realization. The method should produce similar values for both matrices, as we verified. In addition, the similarity between matrices is verified for all realizations.

Table 2: Correlation coefficient matrices between all variables from the dataset and from the first realization generated by SGS. The. correlation matrix of the data before the simulation. B. correlation matrix of the zero realization of the Gaussian sequence simulation.

a.

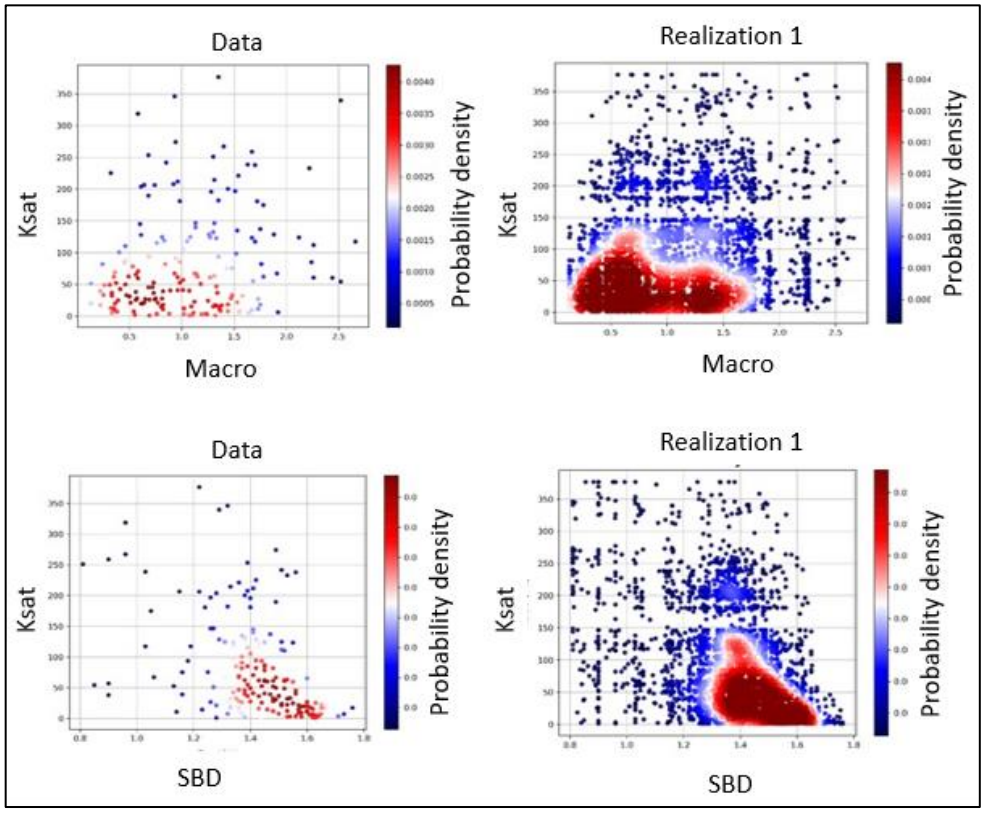
	Ksat	Macro	BDS
Ksat	1.00		
Macro	0.64	1.00	
BDS	-0.45	-0.77	1.00

b.

	Ksat_Real0	Macro_Real0	BDS_Real0
Ksat_Real0	1.00		
Macro_Real0	0.65	1.00	
BDS_Real0	-0.45	-0.72	1.00

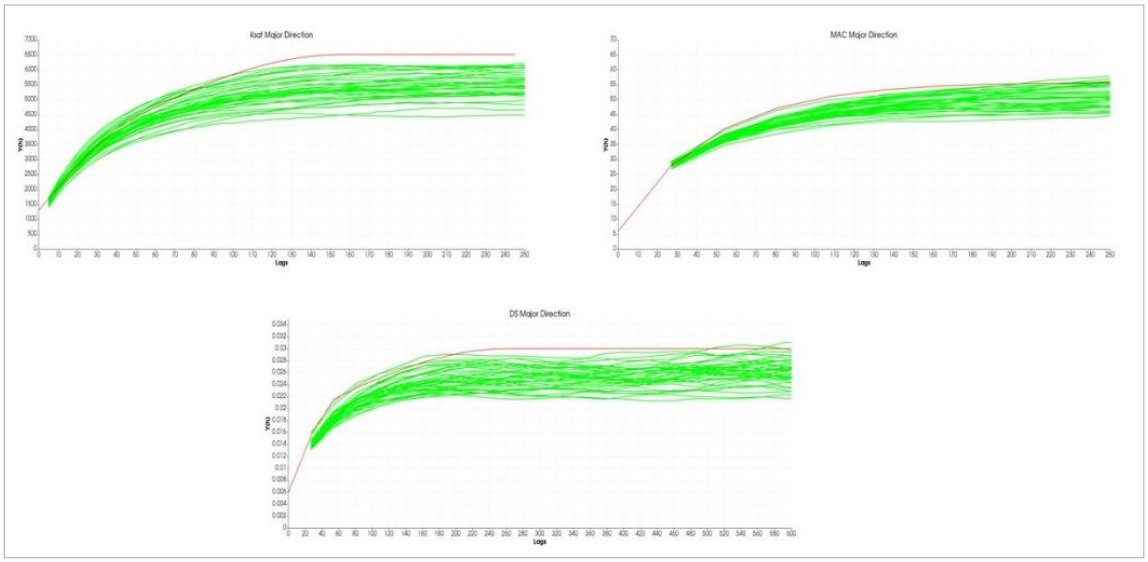
Figure 3 shows the scatterplot between the most important variables. Once again, we compared the scatterplot from the dataset to the scatterplot from the first realizations generated by SGS. We used kernel density estimation to color the scatterplot. We verified that all scatterplots between all variables present similar distributions as the dataset.

Figure 3: Scatter plot of the original data and the first realization for ksat x Macro and ksat x BDS.



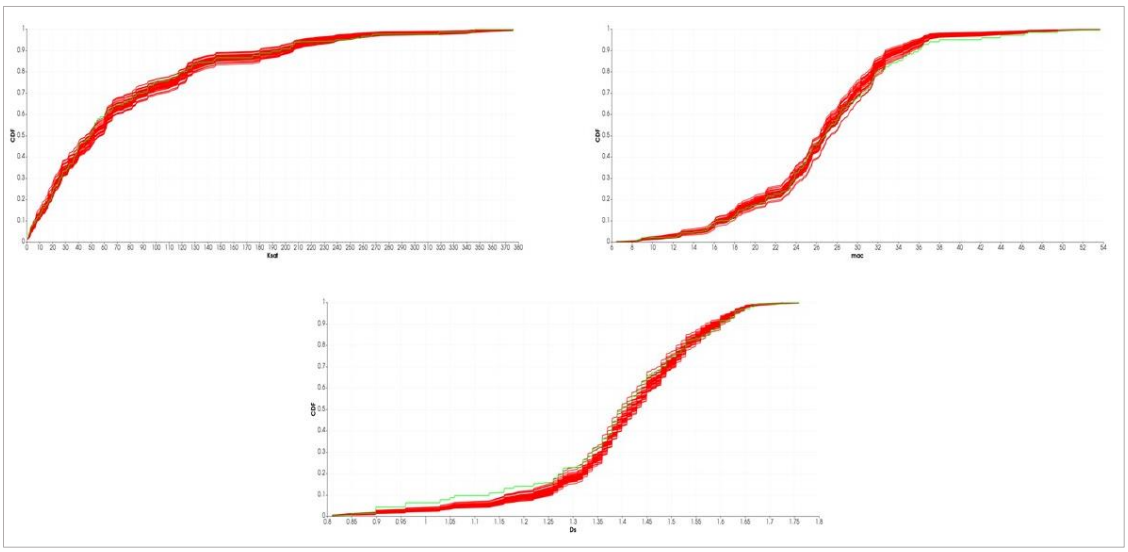
The variograms were also verified. Figure 4 shows the major direction of continuity of each variable. The red model shows the variogram of the original variable from the dataset and the green model shows the variograms of all realizations. The results were satisfactory, given the complexity of the multivariate case study.

Figure 4: Variogram model of the variables in red and realizations in green.



Finally, the cumulative probability function (CDF) was verified, comparing the original CDF to all realizations CDFs for all variables (Figure 5)

Figure 5: Cumulative histogram of the original data (green) and of the realizations (red).



3.3 Two-dimensional wavelet transform - CWT-2D.

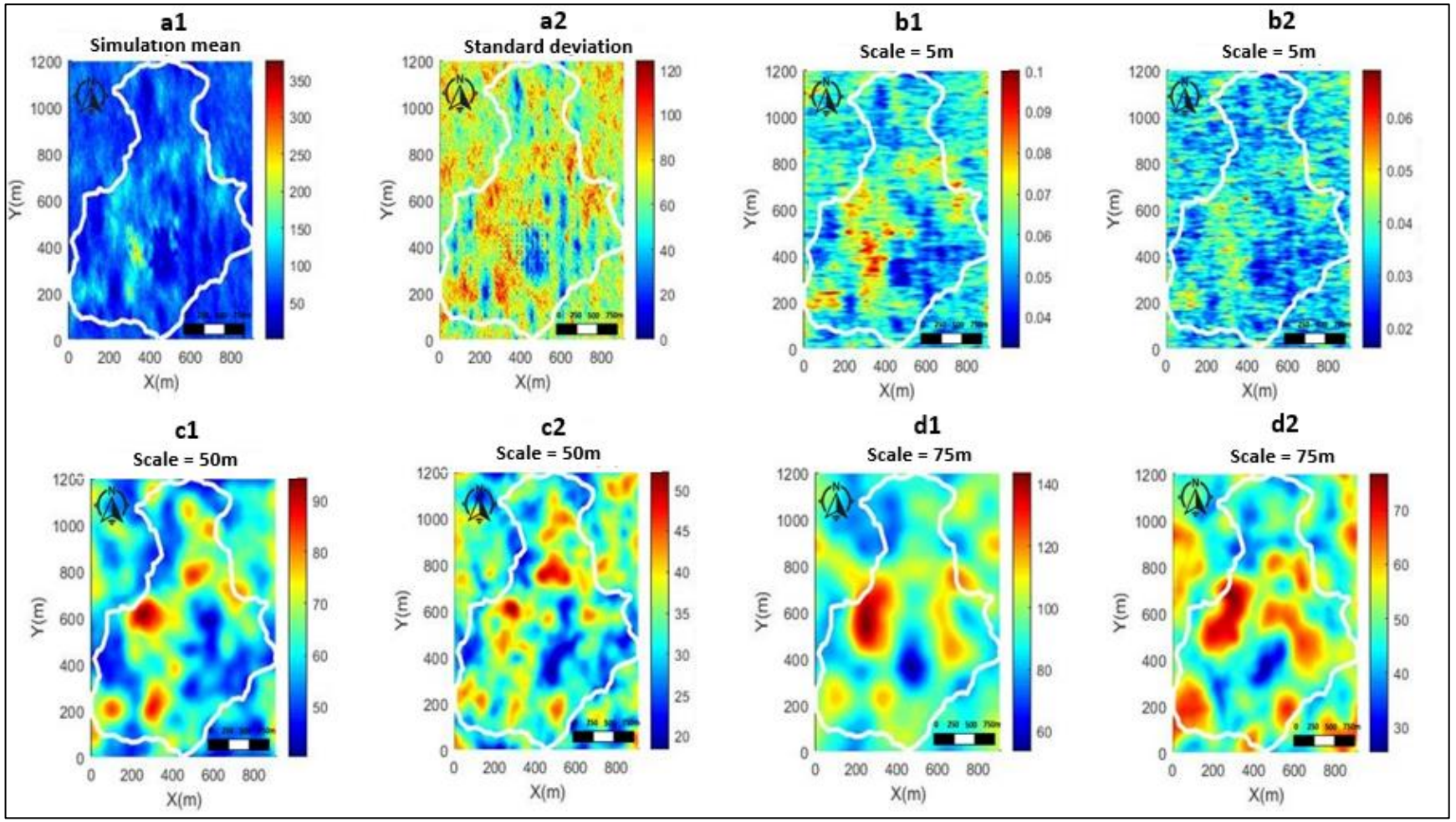
Analyzing the average variations of the Ksat coefficients (Figure 6), obtained through the two-dimensional Wavelet transform, on the three scales of 5, 50 and 75 meters (Figure 6b, c, d) it is observed that there is a similarity between them in the central part of the images. However, at the 5m scale (Figure 6 b1) it isobserved that these variations are more heterogeneous, when compared to the other two scales of 50 and 75m (Figure 6 c1 and d1).

It is also observed that as the scale increases, the amplitude in the average values of the Ksat coefficient increases. On the 5m scale, this amplitude varied from 0.04 to 0.1, with the areas that showed the greatest variability occurring from 0 to 450m on the x-axis and from 200 to 700m on the y-axis, after which high to moderate variability occurs. on the scale of 600 to 800m on the x-axis and 600 to 800m on the y-axis (Figure 6 b1). Figure 1b1, the formation of small areas with high and moderate variability is also observed throughout.

When analyzing Figure 1 c1, referring to Ksat on the 50m scale. it is observed that the average Ksat coefficients now assume average values of 40 to 100. Although the areas with greater variability remain the same on the 5m scale, on the 50m scale more extensive areas are observed. Therefore, the high variability is concentrated only in these regions. This fact also occurs on the scale from 50 to 75m, however in this area of 75m the variability of the coefficient becomes greater and with more homogeneity, appearing more concentrated in the x coordinates from 200 to 350m and y coordinates from 500 to 800 (Figure 6 d1).

When analyzing these scales (5, 50 and 75m), together with the simulated Ksat values (Figure 6 a1), it is observed that the tested scales sought to maximize the variability in Ksat values, mainly around 200 to 400m on the x-axis and around 400m on the y-axis. However, in general, considering the average coefficient values obtained by the wavelet transform and the average simulation values, it is observed that the 75m scale was the most representative of the simulated values. Regarding the standard deviation of the Ksat coefficients, Figure 6, it is observed that the 5m scale presents more uniform values in its coefficients, thus having a set of more uniform coefficients, however the 75m scale (Figure 6 d2) presented a standard deviation closer to the behavior of the simulated data (Figure 6 a2).

Figure 6: Mean simulation of the hydraulic conductivity of saturated soil (Ksat), and the Ksat coefficients, after the CWT-2D, at scales of 5, 50 and 75m.



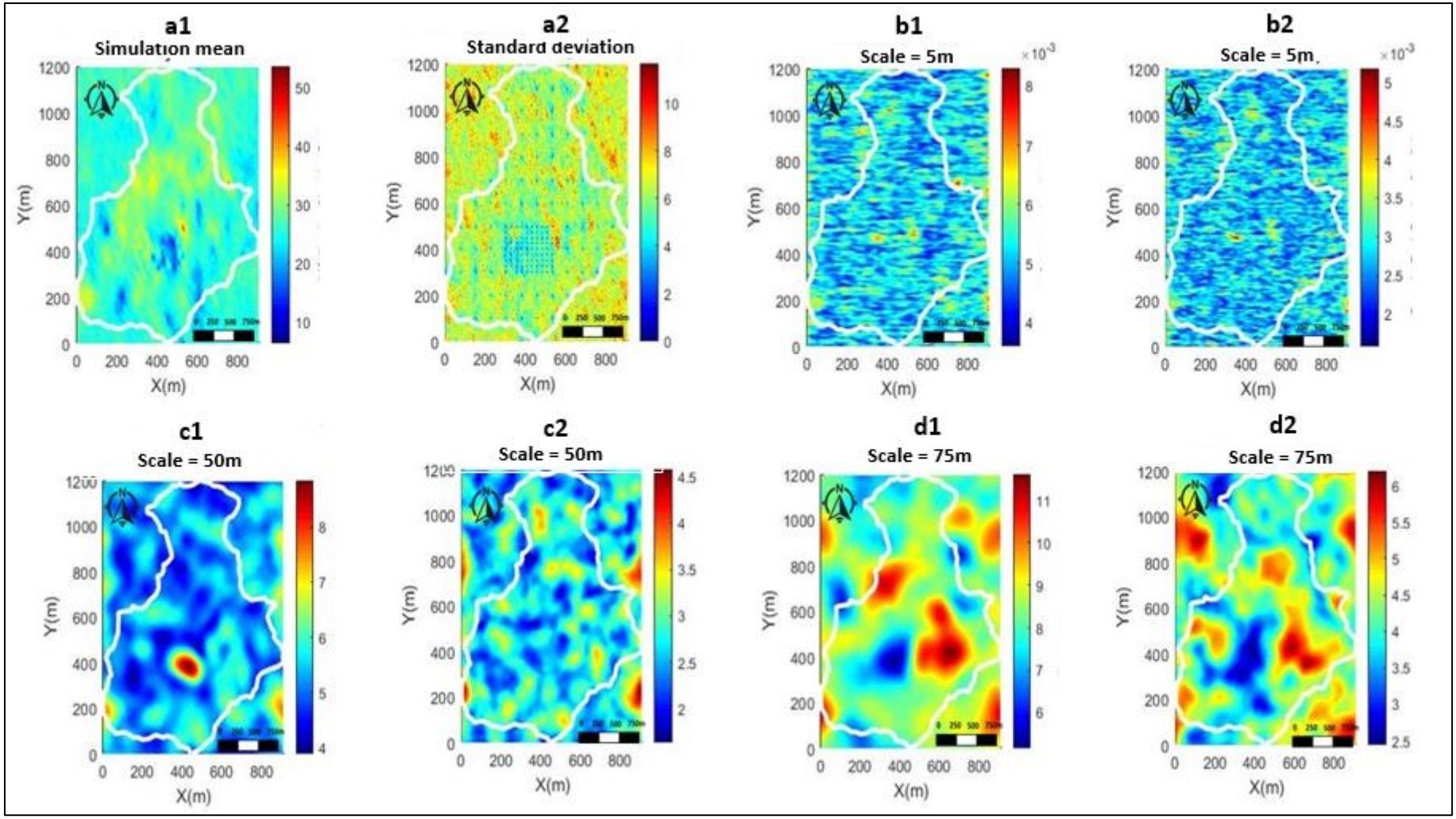
Analyzing Macroporosity at its average simulation values (Figure 7 a1), it is observed that the greatest variability, as well as Ksat, is found more towards the center of the figure, however this presents an area of lower high variability when compared to Ksat and are concentrated around 500m on the x-axis and 400 to 600m on the y-axis. When comparing the scales of the wavelet transform, it is observed that it presents a similar pattern to that of Ksat within its scales. Namely, at the 5m scale (Figure 7b1), greater heterogeneity is observed throughout the area, since there is no way to define areas with high variability, except for a small central area similar to that found in Figure 7a1. At the 50m scale (Figure 7c1), the data as well as Ksat already present greater homogeneity, seeking to form larger areas with high variability, these being close to the medium and high areas of the simulation (Figure 7a1). However, although the 75m scale (Figure 7 d1) maximized the high values of Macroporosity variability, this, like Ksat, better represents the data used as input data in the Wavelet transform. And its areas of high variability match Ksat's areas of high variability.

Another interesting fact is that, as with Ksat, as the scale increases, the average values of the Macroporosity coefficient increase. Therefore, in relation to the standard deviation, it is observed that the 75m scale (Figure 7d2) is the one that presents the greatest dispersion in its data set, however being more homogeneous than the simulated data (Figure 7 a2) used as data of input to the transform. This similarity relationship in the areas between Ksat and Macro was already expected since there is a tendency for the hydraulic conductivity of saturated soils to increase as macroporosity increases (CENTENO et al., 2020).

Because Ksat describes the functionality of the porous system, encompassing properties interconnected with porosity, such as quantity, size, morphology, continuity, and orientation of pores (REICHARDT; TIMM, 2022). However, another relevant factor is in relation to average variations in Macroporosity, Figure 7 a1 (yellow color), this variation is maximized on the 75m scale (Figure 7 d1), and this behavior is followed in Ksat (Figure 7 a1 and d1). This fact can be explained by Bouma, Belmans and Dekker (1982), who described those small pores can conduct more when they are continuous, while larger ones in a given section may not contribute to the flow when they present discontinuity in the soil profile. Therefore, the classification by size does not necessarily reflect the important pattern of continuity of the soil pores, a fact observed in the largest variations in Ksat on the 75m scale.

.

Figure 7: Mean macroporosity simulation (Macro), and macro coefficients, after the CWT-2D, at scales of 5, 50 and 75m.



Finally, regarding soil density, it is observed that the average values of the simulation (Figure 8 a1), as well as the analyzed scales of soil density (Figure 8 b1, c1, d1) showed high variability in areas inverse to those found in Ksat. And although the 75m scale (Figure 8 d1) brought it to 600 to 800m on the x-axis and 200 to 400m on the y-axis, this maximized the simulation average found at around 200m on the x-axis and 400 to 600m on the y-axis (Figure 8 a1). However, as mentioned above, the standard deviation showed the greatest heterogeneity around the average values on the 75m scale (Figure 8 d2), which are very similar to the dispersion found in the simulation standard deviation (Figure 8 a2).

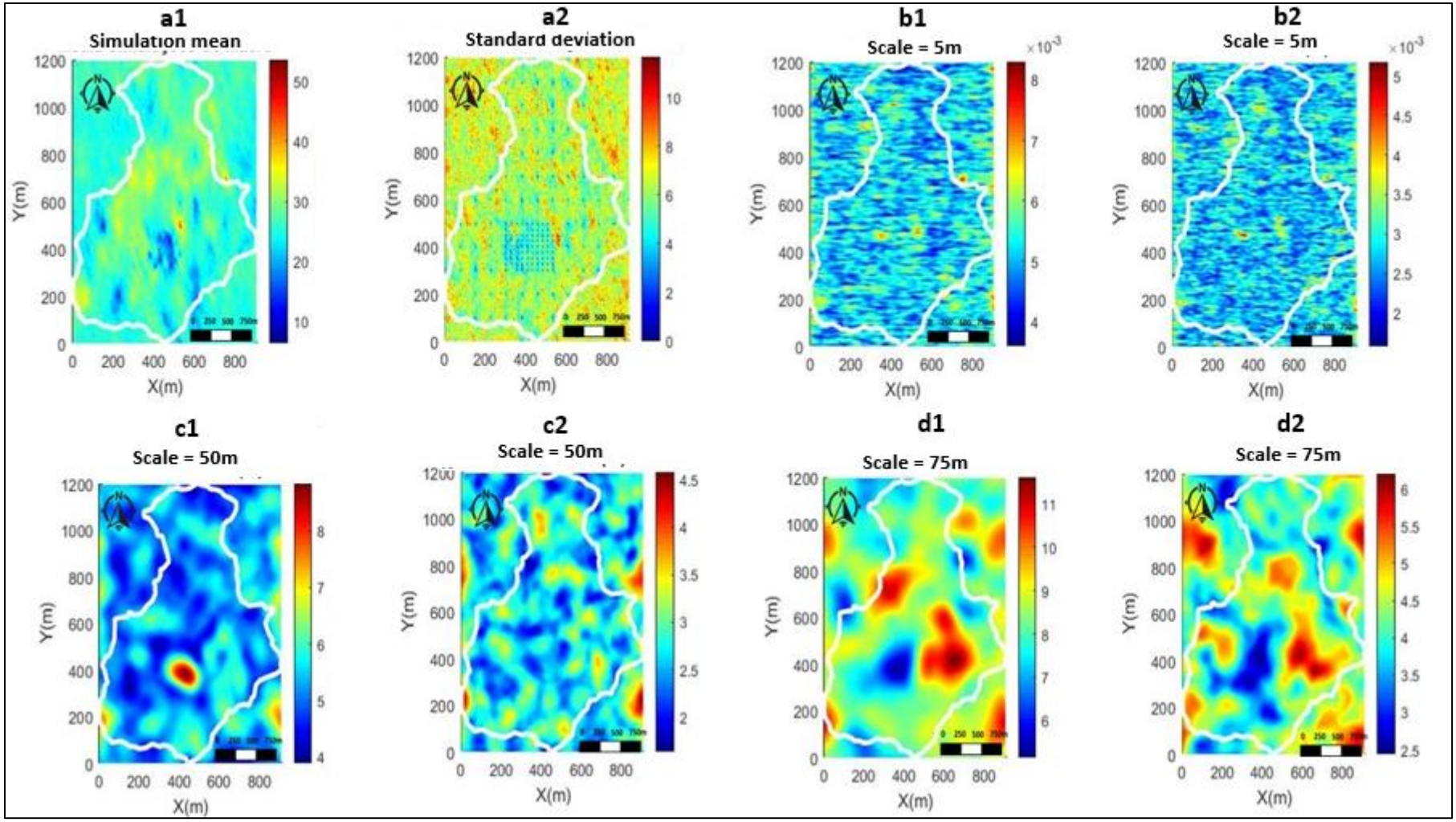
This is the inverse dynamic between Ksat and soil density. It may be related to the fact that Ksat, which infers the soil's ability to allow water to pass through, and soil density, is associated with soil compaction and reduced porosity. Denser soils tend to exhibit lower porosity, which negatively impacts Ksat. The influence of soil porosity emerges as a critical factor, explaining, in part, the observed inverse relationship. More porous soils, characterized by lower density, provide higher Ksat. However, the differential behavior at different scales points to the presence of regions where soil density favors saturated hydraulic conductivity. When examining the average variations of the Ksat, Macro and Ds coefficients, on the 5m, 50m and 75m scales, it is evident that each scale has distinct characteristics.

And the enlargement of the scale results in an increase in the amplitudes of the coefficients, indicating greater spatial variability. Notably, the 75m scale stands out for maximizing this variability, however, this scale also introduces greater heterogeneity, revealing complex patterns that are more difficult to measure, especially when dealing with attributes that do not have a normal distribution. The relationship between macroporosity and Ksat is coherent since hydraulic conductivity is closely linked to porosity. The 75m scale highlights patterns of pore continuity and discontinuity that influence Ksat variability.

In relation to soil density, an inverse dynamic is observed with Ksat. Areas of high variability in Ksat coincide with areas of lower soil density. The 75m scale once again stands out in maximizing soil density variability, indicating complex patterns of compaction and porosity. Choosing the best observation scale is crucial and depends on the specific objectives of the analysis. The joint analysis of the properties suggests that the 75m scale is the most representative of the simulated values, but this choice implies

dealing with greater heterogeneity and spatial complexity. Smaller scales reveal fine details, while larger scales highlight broader patterns.

Figure 8: Simulation mean Soil Bulk density, and BDS coefficients, after the CWT-2D, at scales of 5, 50 and 75m.



3.4 Land Use

Another extremely relevant factor when analyzing variations in Ksat, Macro and BDS is land use and occupation, as this can contribute to both soil degradation and recovery. Therefore, after generating the land use map in table 2, it is possible to observe the confusion matrix, of the established soil classes, as well as the area covered by a given class, where it is possible to observe that around 20% of the area it is occupied by annual cultivation, 16% by forestry, 24% native forest and 10% pasture. However, each of these classes of land use results in unequal conditions of soil physical balance, because they have on the formation and stabilization of aggregates, which are responsible for the dynamics of the water-soil-plant system (SHARMA; DATTA; SHARMA, 2022; SILVA, et al., 2021).

Physical degradation of the soil due to management, mainly in agricultural cultivation systems that intensively use machinery and equipment. It results from the loss of structural stability, increased soil compaction, decreased total porosity and altered pore diameter distribution, creating an unfavorable physical environment for plant growth (POLICH et al., 2023; RAJ et al., 2023; GAJIć et al., 2023).

Table 3: containing ECW metrics and areas, land use classification.

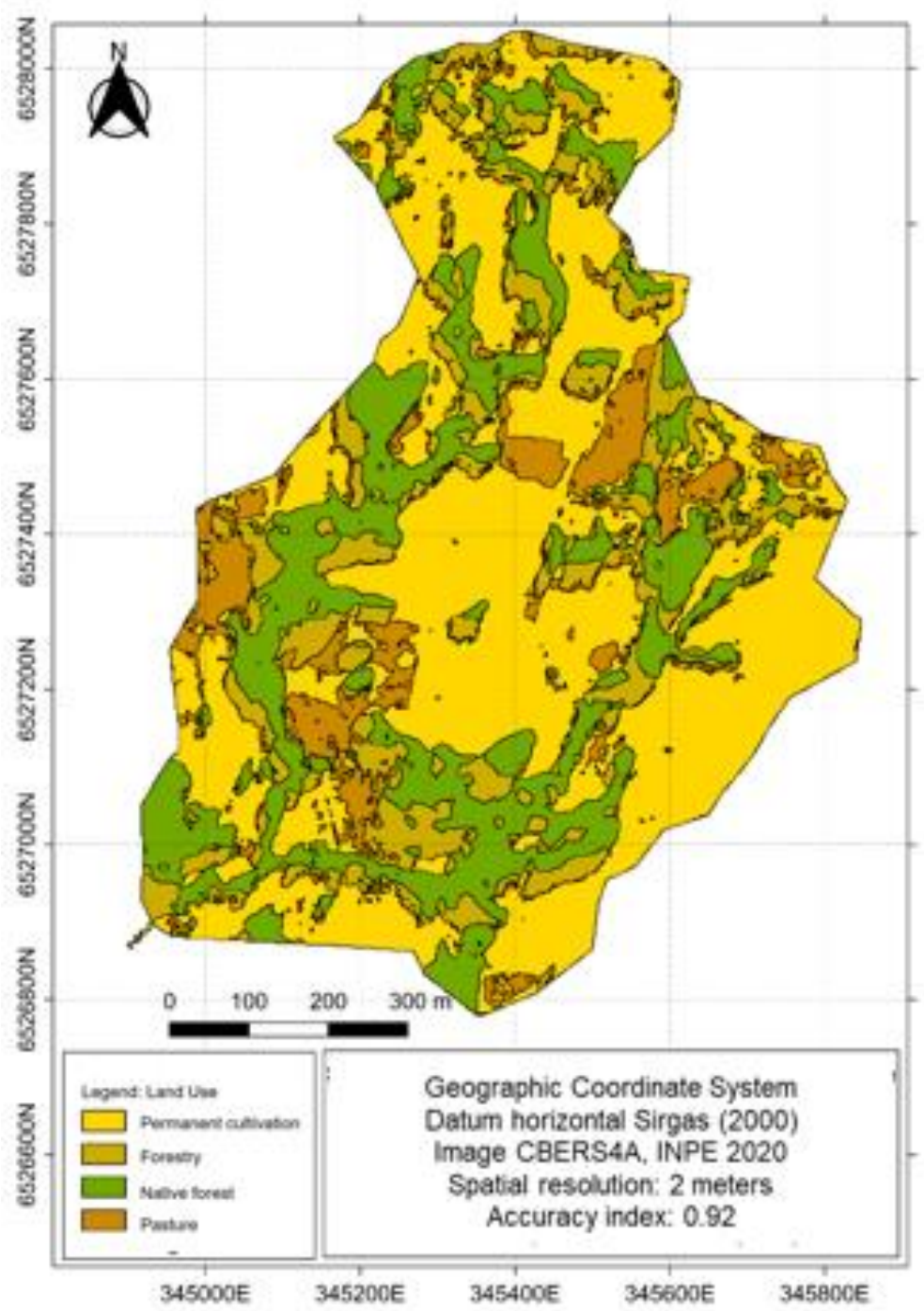
Classes	Total samples	Accuracy by class	Total area (m²)	Total area (%)	Global Accuracy	Standard deviation
Annual cultivation	199	0.980	347044.904	50	0.914	0.013
Forestry	63	0.778	110625.409	16		
Native forest	93	0.935	162702.530	24		
Pasture	41	0.756	70747.182	10		
total	396		691120.025			

This fact is related to Ksat being reduced through changes in the soil structure, such as increased soil density, changes in aggregation, reduction in total porosity and changes in the distribution of pore diameter, its management when modifying these attributes, consequently, interferes with Ksat. Oliveira et al. (2004) observed that the reduction in macroporosity led to an increase in microporosity in cultivated soils, which directly reflected in water retention and availability. Corroborating this study, Beutler et al. (2001), observed in their study a reduction in total porosity that reflected in the reduction of macroporosity in the management systems analyzed when compared to native forest. This is due to the use of conventional tillage, with excessive soil disturbance, using aerator harrows and disc plows, which promotes soil degradation in the surface layer, mainly through the formation of compacted layers, leading to increased soil losses, water and nutrients and reduced crop productivity. However, soil management also alters several processes, with influence on plant growth, as it relates to aspects such as: absorption of nutrients by the plant, production, and supply of organic matter to the soil, microbial life, plant turgidity, diffusion and, among others, soil temperature (GHORBANI et al., 2008; MAMATHA; KAVITHA, 2023).

Regarding soil density, although it behaves contrary to Ksat, density is also directly influenced by the cultivation and management system, which alter the structure and pore space associated with it, interfering in the total porosity, distribution of pores by size, aeration capacity, amount of available water, permeability, and infiltration rate (TALUKDER et al; 2023). In short, it is an attribute that allows the level of densification or compaction of a soil to be assessed. And soil compaction creates an unfavorable

environment for plant growth and can impose severe restrictions on productivity and longevity, even under ideal nutrient supply conditions. Therefore, through BDS, it is possible to understand how much land use influences the sustainability of the exploitation to which the soil is subjected.

Figure 9: Land use map referring to ECW.



4 CONCLUSIONS

Through this study it is possible to conclude the use of geostatistical models, specifically the Sequential Gaussian Simulation (SSG), allowed the generation of a regular data matrix spaced equidistant every 5m, through maps of realizations for the variables studied (Ksat, Macroporosity, and Soil Apparent Density), validating through various methods such as visual verification, correlation matrices, scatter plots and variograms. The satisfactory results of these models indicate the effectiveness of the approach used, which made it possible to apply CWT-2D.

When applying the two-dimensional Wavelet transform at different scales (5m, 50m and 75m), it was observed that the choice of scale is crucial and depends on the specific objectives of the analysis. The 75m scale emerged as the most representative of the simulated values, despite introducing smaller spatial heterogeneity. Joint analysis of soil properties revealed that while smaller scales provide fine details, larger scales highlight broader patterns.

These scales adequately represent what occurs in the nature of the experimental area and can be validated through the land use map, which revealed that areas with a predominance of native forest presented the greatest variations in Ksat, especially towards the West and Northeast. In contrast, areas destined for permanent cultivation exhibited the smallest variations in the average Ksat and macroporosity coefficients, indicating less complexity in the soil structure due to agricultural practices.

REFERENCES

ALBALASMEH, Ammar; et al. Artificial neural network optimization to predict saturated hydraulic conductivity in arid and semi-arid regions. Catena, [S.l.], v. 217, p. 1-20, out. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2022.106459.

ALVAREZ, Edwin Esteban Reyes et al. Assessing Functions of Pedotransference to Determine Microporosity for a Soil (Typic Hapludox) under Two Conditions of Use in the Orinoco Region Colombia. Journal Of Agriculture and Ecology Research International, [S.l.], p. 1-11, 27 set. 2019. Sciencedomain International. http://dx.doi.org/10.9734/jaeri/2019/v19i430089.

BARNETT, R. M.; MANCHUK, J. G.; DEUTSCH, C. V. The Projection-Pursuit Multivariate Transform for Improved Continuous Variable Modeling. Spe Journal, [S.l.], v. 21, n. 06, p. 2010-2026, 20 out. 2016. Society of Petroleum Engineers (SPE). http://dx.doi.org/10.2118/184388-pa.

BARNETT, R.M., MANCHUK, J.G., DEUTSCH, C.V. Projection pursuit multivariate transform. Math. Geosci. v. 46, p. 337–359. 2014. https://doi.org/10.1007/s11004-013-9497-7.

BASSANI, Marcel Antonio Arcari; COSTA, João Felipe Coimbra Leite; DEUTSCH, Clayton Vernon. Multivariate geostatistical simulation with sum and fraction constraints. Applied Earth Science, [S.l.], v. 127, n. 3, p. 83-93, 19 jun. 2018. Informa UK Limited. http://dx.doi.org/10.1080/25726838.2018.1468145.

BERTHELIN, Romane; et al. Estimating karst groundwater recharge from soil moisture observations – a new method tested at the Swabian Alb, southwest Germany. Hydrology And Earth System Sciences, [S.l.], v. 27, n. 2, p. 385-400, 19 jan. 2023. Copernicus GmbH. http://dx.doi.org/10.5194/hess-27-385-2023.

BEUTLER, A. N. et al. Resistência à penetração e permeabilidade de latossolo vermelho distrófico típico sob sistemas de manejo na região dos cerrados. Revista Brasileira de Ciência do Solo, [S.l.], v. 25, n. 1, p. 167-177, mar. 2001. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s0100-06832001000100018.

BOUMA, J.; BELMANS, C. F. M.; DEKKER, L. W. Water Infiltration and Redistribution in a Silt Loam Subsoil with Vertical Worm Channels. Soil Science Society of America Journal, [S.l.], v. 46, n. 5, p. 917-921, set. 1982. Wiley. http://dx.doi.org/10.2136/sssaj1982.03615995004600050006x.

CENTENO, L.N., et al. Identifying regionalized co-variate driving factors to assess spatial distributions of saturated soil hydraulic conductivity using multivariate and state-space analyses. Catena, v.191, p.1-12. 2020. https://doi.org/10.1016/j.catena.2020.104583.

CONGALTON, Russell G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, [S.l.], v. 37, n. 1, p. 35-46, jul. 1991. Elsevier BV. http://dx.doi.org/10.1016/0034-4257(91)90048-b.
CONGALTON, Russell G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, [S.L.], v. 37, n. 1, p. 35-46, jul. 1991. Elsevier BV. doi:10.1016/0034-4257(91)90048-b.

CONGALTON, Russell G.; GREEN, Kass. Assessing the Accuracy of Remotely Sensed Data. 3. ed. Si: Principles and Practices, 2018. 346 p.

DEUTSCH, C.V.; JOURNEL, A.G. Gslib. In: Geostatistical Software Library and User's Guide, vol. 369. Oxford University Press New York, 1998. URL: http://claytonvdeutsch.com/wp-content/uploads/2019/03/GSLIB-Book-Second-Edition.pdf

DOMINGUES, M.O.; et al. Explorando a transformada Wavelet contínua. Revista Brasileira de Ensino de Física, [S.L.], v. 38, n. 3, p. 1-12, set. 2016. FapUNIFESP (SciELO). doi:10.1590/1806-9126-rbef-2016-0019

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Sistema brasileiro de classificação de solos. Rio de Janeiro, 2006. 306p.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Manual de métodos de análises de solo. 2.ed. Rio de Janeiro, Ministério da Agricultura e do Abastecimento, 1997. 212p.

FLAMMINI, A.; et al. A statistical approach for the assessment of the saturated hydraulic conductivity applied to an Austrian region. Journal Of Hydrology: Regional Studies, [S.L.], v. 45, p. 1-10, fev. 2023. Elsevier BV. doi:10.1016/j.ejrh.2022.101310

GAJIć, Katarina; et al. Hydraulic properties of fine-textured soils in lowland ecosystems of Western Serbia vary depending on land use. Geoderma Regional, [S.L.], v. 32, p. 1-10, mar. 2023. Elsevier BV. doi:10.1016/j.geodrs.2022.e00603

GHORBANI, R. et al. Soil management for sustainable crop disease control: a review. Environmental Chemistry Letters, [S.L.], v. 6, n. 3, p. 149-162, 10 abr. 2008. Springer Science and Business Media LLC. doi:10.1007/s10311-008-0147-0

HADJIPETROU, Stylianos; MARIETHOZ, Gregoire; KYRIAKIDIS, Phaedon. Gap-Filling Sentinel-1 Offshore Wind Speed Image Time Series Using Multiple-Point Geostatistical Simulation and Reanalysis Data. Remote Sensing, [S.L.], v. 15, n. 2, p. 409-421, 9 jan. 2023. MDPI AG. doi:10.3390/rs15020409

IBGE. Instituto Brasileiro de Geografia e Estatística. Geologia, geomorfologia, pedologia, vegetação, uso potencial da terra. Rio de Janeiro: IBGE, 1986.

INPE. Instituto Nacional de Pesquisas Espaciais. Satélite Sino-Brasileiro de Recursos Terrestres CBERS. 20232. http://www.cbers.inpe.br/. Acesso em: 13 jan. 2024.

KARACAN, C. Özgen; ERTEN, Oktay; MARTÍN-FERNÁNDEZ, Josep Antoni. Assessment of resource potential from mine tailings using geostatistical modeling for compositions: a methodology and application to katherine mine site, arizona, usa. Journal Of Geochemical Exploration, [S.L.], v. 245, p. 1-15, fev. 2023. Elsevier BV. doi:10.1016/j.gexplo.2022.107142.

KE, Xianmin et al. Spatial variability of the vertical saturated hydraulic conductivity of sediments around typical thermokarst lakes. Geoderma, [S.L.], v. 429, p. 1-15, jan. 2023. Elsevier BV. doi:10.1016/j.geoderma.2022.116230.

KLOECKNER, Jonas; et al. Application of risk assessment to improve sustainability in bauxite mining. Resources Policy, [S.L.], v. 74, p. 1-14, dez. 2021. Elsevier BV. doi:10.1016/j.resourpol.2021.102328.

KLUTE, A., DIRKSEN, C. Hydraulic conductivity and diffusivity: laboratory methods. In: Klute, A. (Ed.), Methods of Soil Analysis. Agronomy Monograph, ASA-SSSA, Madison, pp. 687–734. 1986.

KLUTE, A. Water retention: laboratory methods. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1, 2nd ed. Agronomy Monograph vol. 9. ASA-SSSA, Madison, WI, pp. 635–662. 1986.

KUINCHTNER, A., BURIOL, G.A. Clima do estado do Rio Grande do Sul segundo a classificação climática de Köppen e Thornthwaite. Disciplinarum Scientia - Serie: Ciências Exatas 2, 171–182. 2001.

LI, Zhi et al. Design and Analysis of an Effective Multi-Barriers Model Based on Non-Stationary Gaussian Random Fields. Electronics, [S.L.], v. 12, n. 2, p. 345-351, 9 jan. 2023. MDPI AG. doi:10.3390/electronics12020345.

LIU, Lei; PRODANOVIć, Maša; PYRCZ, Michael J. Impact of geostatistical nonstationarity on convolutional neural network predictions. Computational Geosciences, [S.L.], v. 27, n. 1, p. 35-44, 17 nov. 2022. Springer Science and Business Media LLC. doi:10.1007/s10596-022-10181-3.

MAMATHA, V.; KAVITHA, J.C. Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques. Measurement: Sensors, [S.L.], v. 25, p. 1-17, fev. 2023. Elsevier BV. doi:10.1016/j.measen.2023.100665.

MORRIS, P. J. et al. Saturated Hydraulic Conductivity in Northern Peats Inferred From Other Measurements. Water Resources Research, [S.L.], v. 58, n. 11, p. 1-18, nov. 2022. American Geophysical Union (AGU). doi:10.1029/2022wr033181.

OLIVEIRA, G. C.; et al. Caracterização química e físico-hídrica de um Latossolo Vermelho após vinte anos de manejo e cultivo do solo. Revista Brasileira de Ciência do Solo, [S.L.], v. 28, n. 2, p. 327-336, abr. 2004. FapUNIFESP (SciELO). doi:10.1590/s0100-06832004000200011.

PÁEZ-BIMOS, Sebastián et al. Vegetation effects on soil pore structure and hydraulic properties in volcanic ash soils of the high Andes. Hydrological Processes, [S.L.], v. 36, n. 9, p. 1-15, set. 2022. Wiley. doi:10.1002/hyp.14678.

POLICH, Nicolás Guillermo; et al. Cover crops effects on anisotropy of unsaturated soil hydraulic properties. Soil And Tillage Research, [S.L.], v. 227, p. 1-16, mar. 2023. Elsevier BV. doi:10.1016/j.still.2022.105601.

RAJ, Rishi; et al. Soil physical environment and active carbon pool in rice–wheat system of South Asia: impact of long-term conservation agriculture practices. Environmental Technology & Innovation, [S.L.], v. 29, p. 1-17, fev. 2023. Elsevier BV. doi:10.1016/j.eti.2022.102966.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2016.

REICHARDT, Klaus; TIMM, Luís Carlos. Solo, planta e atmosfera: conceitos, processos e aplicações. 4. ed. Baurueri [Sp]: Manole, 2022. 508 p.

SHARMA, U. C.; DATTA, M.; SHARMA, Vikas. Land Use and Management. Soils In The Hindu Kush Himalayas, [S.L.], v. 1, n. 1, p. 295-462, 2022. Springer International Publishing. doi:10.1007/978-3-031-11458-8_7.

SILVA, Thais Palumbo; et al. Investigating spatial relationships of soil friability and driving factors through co-regionalization with state-space analysis in a subtropical

watershed. Soil And Tillage Research, [S.L.], v. 212, p. 1-12, ago. 2021. Elsevier BV. doi:10.1016/j.still.2021.105028.

SOARES, M. F., et al. Identifying Covariates to Assess the Spatial Variability of Saturated Soil Hydraulic Conductivity Using Robust Cokriging at the Watershed Scale. J. Soil Sci. Plant Nutr. V. 20, p. 1491–1502. 2020. doi:10.1007/s42729-020-00228-8.

SOARES, Mauricio Fornalski; et al. Assessing the spatial variability of saturated soil hydraulic conductivity at the watershed scale using the sequential Gaussian co-simulation method. Catena, [S.L.], v. 221, p. 1-12, fev. 2023. Elsevier BV. doi:10.1016/j.catena.2022.106756.

SOUZA, c. m.; SÁ, L. D. A.; DOMINGUES, M. O. Variabilidade em Escala do Dossel da Floresta Amazônica: resultados para a reserva rebio jarú-rondônia. Tema (São Carlos), [S.L.], v. 14, n. 3, p. 1-14, 24 nov. 2013. Brazilian Society for Computational and Applied Mathematics (SBMAC). doi:10.5540/tema.2013.014.03.0415.

STEHMAN, S. V. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. International Journal of Remote Sensing, [S.L.], v. 35, n. 13, p. 4923-4939, 3 jul. 2014. Informa UK Limited. doi:10.1080/01431161.2014.930207.

TALUKDER, Rasendra; et al. Soil hydraulic properties and pore dynamics under different tillage and irrigated crop sequences. Geoderma, [S.L.], v. 430, n. 5, p. 1-11, fev. 2023. Elsevier BV. doi:10.1016/j.geoderma.2022.116293.

WANG, X. et al. Mapping conduits in two-dimensional heterogeneous karst aquifers using hydraulic tomography. Journal Of Hydrology, [S.L.], v. 617, p. 1-17, fev. 2023. Elsevier BV. doi:10.1016/j.jhydrol.2022.129018.

WELDEGEBRIEL, L.; et al. Organization of the soil profile controls the risks of runoff in the humid Ethiopian Highlands. Journal Of Hydrology, [S.L.], v. 617, p. 1-12, fev. 2023. Elsevier BV. doi:10.1016/j.jhydrol.2022.129031.

WILDING, L. P.; DREES, L. R. Spatial variability and pedology. In: WILDING, L. P.; DREES, L. R. (Eds.). Pedogenesis and soil taxonomy: concepts and interactions. New York: Elsevier, 1983. p. 83–116.

ZAHEDIFAR, M. et al. Assessing alteration of soil quality, degradation, and resistance indices under different land uses through network and factor analysis. Catena, [S.L.], v. 222, p. 1-16, mar. 2023. Elsevier BV. doi:10.1016/j.catena.2022.106807.

ZERMANE, A.; et al. Predicting fatal fall from heights accidents using random forest classification machine learning model. Safety Science, [S.L.], v. 159, p. 1-10, mar. 2023. Elsevier BV. doi:10.1016/j.ssci.2022.106023.

5. CONSIDERAÇÕES FINAIS

Com base neste estudo em relação aos modelos lineares mistos, os resultados da simulação enfatizam que o software GeoR demonstrou uma performance superior em várias situações, evidenciando uma taxa consistente de convergência, especialmente nos efeitos fixos. Por outro lado, o Asreml-R enfrentou desafios adicionais, como maior variabilidade e dificuldades em determinados cenários. A sensibilidade do modelo à variação do parâmetro v foi evidente, com o GeoR apresentando maior estabilidade e consistência, especialmente em valores menores de v, o que resultou em estimativas mais próximas dos valores observados. A introdução do efeito nugget, embora tenha aumentado significativamente o tempo de processamento de dados, produziu diferenças nos ganhos observados, principalmente em relação à capacidade de capturar variação não estruturada. Além disso, a análise de anisotropia indicou um melhor desempenho do GeoR em comparação com o Asreml. Essas discrepâncias de desempenho entre os softwares destacam a importância de escolher o mais adequado para uma análise específica, levando em consideração as características dos dados e as necessidades do pesquisador em relação a distância existente na sua área de estudo.

No contexto do artigo 2, a simulação sequencial gaussiana permitiu a geração de uma matriz de dados regular e equidistante, o que possibilitou a utilização das Wavelets bidimensionais. A escolha criteriosa da escala revelou-se crucial, dependendo dos objetivos específicos da análise. A escala maior, de 75m foi a mais representativa, apesar de introduzir menor heterogeneidade espacial. As diferentes escalas forneceram auxiliaram na compreensão e detalhamento dos padrões finos e grosseiros, destacando assim a complexidade da estrutura do solo em diversas áreas da bacia hidrográfica. A análise conjunta das propriedades do solo revelou padrões variados, influenciados pela cobertura vegetal e práticas agrícolas presentes na região.

Como sugestões para pesquisas futuras, recomenda-se a exploração de outros modelos e softwares, além da investigação mais aprofundada da anisotropia em contextos irregulares e com efeito nugget. Além disso, é importante validar os resultados obtidos usando conjuntos de dados reais, a fim de confirmar a robustez das conclusões aqui apresentadas. Sugere-se também a aplicação da técnica de crosswavelets para uma análise mais abrangente das escalas e da variabilidade espacial dos dados simultaneamente.

Destaca-se por fim, que estas considerações oferecem resultados relevantes, especialmente para pesquisadores que requerem dados em escalas variadas, desde as mais

detalhadas até as mais amplas. Além disso, é crucial para aqueles que buscam compreender o comportamento da distância entre os pontos amostrais, visando utilizá-los em modelagens posteriores. Considerando que as atividades de campo são dispendiosas e nem sempre viáveis, este estudo proporciona uma oportunidade para uma melhor compreensão da variabilidade espacial em diversos contextos agrícolas e de bacias hidrográficas. Assim, representa um avanço significativo no entendimento dos padrões espaciais, subsidiando decisões mais informadas em várias áreas de estudo relacionadas ao solo e à hidrografia.

REFERÊNCIAS

AKAIKE, HIROTUGU. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, [S. l.], v. 19, n. 6, p. 716–723, 1974. DOI: 10.1109/TAC.1974.1100705.

AQUINO, Leandro Sanzi *et al.* State-space approach to evaluate effects of land levelling on the spatial relationships of soil properties of a lowland area. **Soil And Tillage Research**, [S.l.], v. 145, p. 135–147, 2015. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.still.2014.09.007.

ARIYO, Oludare *et al.* Bayesian model selection in linear mixed models for longitudinal data. **Journal of Applied Statistics,** [S.l.], v. 47, n. 5, p. 890–913, 22 ago. 2020. Informa UK Limited. DOI: http://dx.doi.org/10.1080/02664763.2019.1657814.

AZIZI, Kamran; AYOUBI, Shamsollah; DEMATTÊ, José A.M. Controlling factors in the variability of soil magnetic measures by machine learning and variable importance analysis. **Journal Of Applied Geophysics**, [S.L.], v. 210, p. 1-16, mar. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.jappgeo.2023.104944.

BABU, Subhash *et al*. Soil carbon dynamics under organic farming: impact of tillage and cropping diversity. **Ecological Indicators**, [S.L.], v. 147, p. 1-18, mar. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.ecolind.2023.109940.

BARBOSA, Kátia Noronha. **Anisotropia espacial da perda de solo e atributos do solo em diferentes pedoformas com cultivo de cana-de-açúcar**. 2020. 47 f. Dissertação (Mestrado) - Curso de Agronomia (Ciência do Solo), Universidade Estadual Paulista, Jaboticabal, 2020.

BARCA, Emanuele; BENEDETTO, Daniela de; STELLACCI, Anna Maria. Contribution of EMI and GPR proximal sensing data in soil water content assessment by using linear mixed effects models and geostatistical approaches. **Geoderma**, [S.l.], v. 343, n. 1, p. 280–293, 2019. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.geoderma.2019.01.030.

BATISTA, A. P. B. *et al*. Geoestatística No R: Um Estudo Comparativo Entre Dois Scripts. *In:* IV SIMPÓSIO DE GEOESTATÍSTICA APLICADA EM Ciências AGRÁRIAS, 4., 2015, Botucatu, São Paulo. **Anais...**, Botucatu, São Paulo, 2015. p. 1-5.

BEHERA, Sanjib K.; et al. Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. **Catena**, [S. l.], v. 165, n. February, p. 251–259, 2018. DOI: 10.1016/j.catena.2018.02.008. Disponível em: https://doi.org/10.1016/j.catena.2018.02.008.

BISWAS, Asim. Scale–location specific soil spatial variability: a comparison of continuous Wavelet transform and hilber-huang transform: A comparison of continuous Wavelet transform and Hilbert–Huang transform. **Catena**, [s.l.], v. 160, p. 24–31, jan. 2018. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.catena.2017.08.019.

BISWAS, Asim; SI, Bing Cheng. Identifying scale specific controls of soil water storage in a hummocky landscape using Wavelet coherency. **Geoderma**, [S.l.], v. 165, n. 1, p. 50–59, out. 2011. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.geoderma.2011.07.002.

BITENCOURT, Dioni Glei Bonini et al. Multivariate and geostatistical analyses to evaluate lowland soil levelling effects on physico-chemical properties. **Soil And Tillage Research**, [S.L.], v. 156, p. 63-73, mar. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.still.2015.10.004.

BITENCOURT, Dioni Glei Bonini. **Variabilidade espacial da camada superficial de um gleissolo da planície costeira do estado do Rio Grande do Sul**. Dissertação (Mestrado em Agronomia). 2007. Universidade Federal de Pelotas, [S. l.], 2007.

BORGES, Marcelo de Vasconcelos. **Cartografia geotécnica da cidade de Rio Branco – Acre**: uso de estatística descritiva, geoestatística e estatística multivariada. 2019. 319 f. Tese (Doutorado) - Curso de Geotecnia, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília/DF, 2019.

BRIEN, C. 2016. asremlPlus: **Augments the Use of "ASReml-R" in Fitting Mixed Models**. Disponível em: https://cran.r-project.org/package=asremlPlus. Acesso em: 10 maio. 2020.

BURST, Maxime et al. Distribution of soil properties along forest-grassland interfaces: influence of permanent environmental factors or land-use after-effects? **Agriculture, Ecosystems & Environment**, [S.L.], v. 289, p. 106739, fev. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.agee.2019.106739.

CENTENO, L.N., et al. Identifying regionalized co-variate driving factors to assess spatial distributions of saturated soil hydraulic conductivity using multivariate and state-space analyses. Catena v.191, p1-12. 2020.
https://doi.org/10.1016/j.catena.2020.104583

CENTENO, Luana Nunes *et al.* Dominant Control of Macroporosity on Saturated Soil Hydraulic Conductivity at Multiple Scales and Locations Revealed by Wavelet Analyses. **Journal Soil Science Plant Nutrition**, [S.l.], v. 31, n. 6, p. 1–17, 2020. DOI: https://doi.org/10.1007/s42729-020-00239-5.

CORRÊA, Adriany Rodrigues. **Densidade de amostragem de dados para análise da correlação espacial da produtividade do eucalipto em função de atributos do solo**. 2018. 79 f. Tese (Doutorado) - Curso de Agronomia, Faculdade de Engenharia - Unesp, Universidade Estadual Paulista "Júlio de Mesquita Filho", Ilha Solteira, 2018.

COSTA, Vagner Nascimento *et al.* Contribuciones a las Ciencias Sociales: Modelagem espacial e sistema de abastecimento de água na Amazônia. **Revista Contribuciones a Las Ciencias Sociales**, [S.l.], v. 1, n. 5, p. 24–45, 11 maio. 2021. Servicios Academicos Intercontinentales. DOI: http://dx.doi.org/10.51896/ccs.

CRAWFORD, Carol A. Gotway; HERGERT, Gary W. Incorporating Spatial Trends and Anisotropy in Geostatistical Mapping of Soil Properties. **Division S-8-Nutrient Management & Soil & Plant Analysis**, [S.I.], v. 61, n. 1, p. 298–309, 1997.

DE FAVERI, Joanne *et al.* Residual Variance–Covariance Modelling in Analysis of Multivariate Data from Variety Selection Trials. **Journal of Agricultural, Biological, and Environmental Statistics**, [S. l.], v. 22, n. 1, p. 1–22, 2017. DOI: 10.1007/s13253-016-0267-0.

DEISS, Leonardo et al. Soil chemical properties under no-tillage as affected by agricultural trophic complexity. **European Journal of Soil Science**, [S.L.], v. 7, n. 8, p. 1-16, 20 ago. 2019. Wiley. http://dx.doi.org/10.1111/ejss.12869.

DESSOTTI, Cássio. **Modelos lineares mistos para explicar a variabilidade espacial na análise conjunta de experimentos agronômicos**. 2014. Universidade de São Paulo, [S. l.], 2014.

DIGGLE, P. J.; RIBEIRO JR., P. J. Bayesian inference in Gaussian model-based geo-statistics. Graphical and Environmental Modelling 6, 129–146. 2002.

DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. G**eoR:** Analysis of geostatistical data. R package version 1.7–5.2.1. 2018. Disponível em: https://CRAN.R-project.org/ package=geoR.

DIGGLE, P.; *et al.* **Analysis of longitudinal data**. 2. ed. New York: Oxford University Press, 2002. 396 p. DESSOTTI, Cassio. **Modelos lineares mistos para explicar a variabilidade espacial Análise conjunta de experimentos agronômicos**. 2014. 102 f. Tese (Doutorado) - Curso de Licenciatura em Matemática, Estatística e Experimentação Agronômica, Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2014.

DU, Han; WANG, Lijuan. Testing Variance Components in Linear Mixed Modeling Using Permutation. **Multivariate Behavioral Research**, [S.L.], v. 55, n. 1, p. 120-136, 27 jun. 2019. Informa UK Limited. http://dx.doi.org/10.1080/00273171.2019.1627513.

ERIKSSON, Marian; SISKA, Peter P. Understanding anisotropy computations. **Mathematical Geology**, [S. l.], v. 32, n. 6, p. 683–700, 2000. DOI: 10.1023/A:1007590322263.

FARGE, M. Wavelet Transforms and their Applications to Turbulence. **Annual Review Of Fluid Mechanics**, [s.l.], v. 24, n. 1, p. 395-458, jan. 1992. Annual Reviews. http://dx.doi.org/10.1146/annurev.fl.24.010192.002143.

FARIAS FILHO, Marcelino Silva; BUENO, Célia Regina Paes; VALLADARES, Gustavo Souza. Caracterização e classificação de solos hidromórficos sobre os aluviões fluviomarinhos no município de Arari – MA. **Raega**, Curitiba/pr, v. 47, n. 1, p. 85-98, jul. 2020.

FAVERI, Joanne de *et al.* Residual Variance–Covariance Modelling in Analysis of Multivariate Data from Variety Selection Trials. **Journal of Agricultural, Biological and Environmental Statistics**, [S.L.], v. 22, n. 1, p. 1–22, 16 dez. 2016. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s13253-016-0267-0.

FERREIRA, Gabriel Fernandes Pinto. **Espacialização de atributos do solo e do cafeeiro arábica em densidades amostrais no planalto de Vitória da Conquista - BA**. 2020. 111 f. Dissertação (Mestrado) - Curso de Agronomia, Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2020.

GASPARINI, Alessandro *et al.* Mixed-effects models for health care longitudinal data with an informative visiting process: a monte carlo simulation study. **Statistica Neerlandica,** [S.l.], v. 74, n. 1, p. 5–23, 5 set. 2020. Wiley. DOI: http://dx.doi.org/10.1111/stan.12188.

GHOSH, Avishek; RAMCHANDRAN, Kannan. Alternating Minimization Converges Super-Linearly for Mixed Linear Regression. **Statistics Machine Learning**: University Cornell, [S.I.], v. 1, n. 11, p. 1-12, set. 2020.

GOLKARIAN, Ali; et al. Spatial variability of soil water erosion: comparing empirical and intelligent techniques. **Geoscience Frontiers**, [S.L.], v. 14, n. 1, p. 1-15, jan. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.gsf.2022.101456.

GONZÁLEZ-ÁVILA, Itzayana *et al*. Sociogeomorphological Analysis in a Headwater Basin in Southern Brazil with Emphasis on Land Use and Land Cover Change. **Land**, [S.L.], v. 12, n. 2, p. 306-321, 21 jan. 2023. MDPI AG. http://dx.doi.org/10.3390/land12020306.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York: Oxford University Press, 1997.

GRINGARTEN, Emmanuel; DEUTSCH, Clayton V. Teacher's aide: Variogram interpretation and modeling. **Mathematical Geology**, [S. l.], v. 33, n. 4, p. 507–534, 2001. DOI: 10.1023/A:1011093014141.

GUAN, Yongtao *et al.* A Nonparametric Test for Spatial Isotropy Using Subsampling. **Journal of the American Statistical Association**, [S.l.], v. 99, n. 467, p. 810–821, set. 2004. Informa UK Limited. DOI: http://dx.doi.org/10.1198/016214504000001150.

GUEDES, Luciana Pagliosa Carvalho *et al.* Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soi. SciELO Agencia Nacional de Investigacion y Desarrollo (ANID). **Chilean Journal of Agricultural Research**, [S.l.], v. 73, n. 4, p. 414–423, dez. 2013. DOI: http://dx.doi.org/10.4067/s0718-58392013000400013.

HASKARD, Kathryn A.; CULLIS, Brian R.; VERBYLA, Arũnas P. Anisotropic matérn correlation and spatial prediction using REML. **Journal of Agricultural, Biological, and Environmental Statistics**, [S. l.], v. 12, n. 2, p. 147–160, 2007. DOI: 10.1198/108571107X196004.

HASKARD, Kathryn Anne. An anisotropic Matérn spatial covariance model: REML estimation and properties. **Philosophy**, [S. l.], n. November, p. 1–197, 2007.

HEISKANEN, Juha *et al.* Co-variation relations of physical soil properties and site characteristics of Finnish upland forests. **Silva Fennica**, [S.l.], v. 52, n. 3, p. 1–18, 2018. Finnish Society of Forest Science. DOI: http://dx.doi.org/10.14214/sf.9948.

HEUBERGER, Clara F.; BAINS, Praveen K.; DOWELL, Niall Mac. The EV-olution of the power system: a spatio-temporal optimisation model to investigate the impact of electric vehicle deployment. **Applied Energy**, [S.l.], v. 257, n. 1, p. 1–18, 2020. DOI: http://dx.doi.org/10.1016/j.apenergy.2019.113715.

HU, Wei; SI, Bing Cheng. Technical note: multiple Wavelet coherence for untangling scale-specific andlocalized multivariate relationships in geosciences: Multiple Wavelet coherence for untangling scale-specific andlocalized multivariate relationships in geosciences. **Hydrology and Earth System Sciences**, [S.l.], v. 20, n. 8, p. 3183–3191, 8 ago. 2016. Copernicus GmbH. DOI: http://dx.doi.org/10.5194/hess-20-3183-2016.

HUANG, Jingyi *et al.* Unravelling scale- and location-specific variations in soil properties using the 2-dimensional empirical mode decomposition. **Geoderma**, [S. l.], v. 307, p. 139–149, 2017. Disponível em: http://dx.doi.org/10.1016/j.geoderma.2017.07.024.

HUANG, Norden E. *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. **Proceedings of the Royal Society of London. Series A:** Mathematical, Physical and Engineering Sciences, [S. l.], v. 454, n. 1971, p. 903–995, 1998. DOI: 10.1098/rspa.1998.0193. DOI: https://royalsocietypublishing.org/doi/10.1098/rspa.1998.0193.

HUANG, Norden E.; WU, Zhaohua. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. **Reviews of Geophysics**, [S. l.], v. 46, n. 2, p. RG2006, 2008. DOI: 10.1029/2007RG000228.

HUPET, F.; VANCLOOSTER, M. Sampling strategies to estimate field areal evapotranspiration fluxes with a soil water balance approach. **Journal of Hydrology**, [S.l.], v. 292, n. 1–4, p. 262–280, 2004. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.jhydrol.2004.01.006.

ISAAKS, Edward H.; SRIVASTAVA, R. Mohan. **Applied Geostatistics**. Nova York: Oxford University Press, 1989.

JAMES, J. F. **A Student's Guide to Fourier Transforms:** With Applications in Physics and Engineering. 3. ed. New York: Cambridge University Press, 2011. 156 p. ISBN-13: 978-0521176835.

JEAN-PIERRE, Antoine; ROMAIN, Murenzi. **Two-dimensional Wavelets**. Si: Cambridge University Press, 2008. 480 p.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**: Academic Press, London. 1978.

KOCH, DEAN; LELE, SUBHASH; LEWIS, MARK A. Computationally simple anisotropic lattice covariograms. **Environmental and Ecological Statistics**, [S. l.], 2020. DOI: 10.1007/s10651-020-00456-2. Disponível em: https://doi.org/10.1007/s10651-020-00456-2.

LAUDICINA, VITO ARMANDO; et al. Soil profile dismantlement by land levelling and deep tillage damages soil functioning but not quality. **Applied Soil Ecology**, [S.L.], v. 107, p. 298-306, nov. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.apsoil.2016.07.002.

LI, XIAOXIAO; et al. Identifying the Relationship between Soil Properties and Rice Growth for Improving Consolidated Land in the Yangtze River Delta, China. **Sustainability**, [S.L.], v. 10, n. 9, p. 3072-3087, 29 ago. 2018. MDPI AG. http://dx.doi.org/10.3390/su10093072.

LU, Meng; CAVIERES, Joaquin; MORAGA, Paula. A Comparison of Spatial and Nonspatial Methods in Statistical Modeling of NO2: prediction accuracy, uncertainty quantification, and model interpretation. **Geographical Analysis**, [S.L.], v. 1, n. 1, p. 1-17, 17 jan. 2023. Wiley. http://dx.doi.org/10.1111/gean.12356.

LV, Ligang *et al*. Impact of conservation tillage on the distribution of soil nutrients with depth. **Soil And Tillage Research**, [S.L.], v. 225, p. 1-18, jan. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.still.2022.105527.

MARTÍNEZ-CASASNOVAS, J. A.; RAMOS, M. Concepción. Soil alteration due to erosion, ploughing and levelling of vineyards in north east Spain. **Soil Use and Management**, [S.L.], v. 25, n. 2, p. 183-192, jun. 2009. Wiley. http://dx.doi.org/10.1111/j.1475-2743.2009.00215.x.

MARTÍNEZ-CASASNOVAS, José A.; ESCOLÀ, Alexandre; ARNÓ, Jaume. Use of farmer knowledge in the delineation of potential management zones in precision agriculture: A case study in maize (Zea mays L.). **Agriculture (Switzerland)**, [S. l.], v. 8, n. 6, 2018. DOI: 10.3390/agriculture8060084.

MENTGES, Marcelo Ivan *et al.* Alterações estruturais e mecânicas de solo de várzea cultivado com arroz irrigado por inundação. **Revista Brasileira de Ciência do Solo**, [S.l.], v. 37, n. 1, p. 221–231, fev. 2013. FapUNIFESP (SciELO). DOI: http://dx.doi.org/10.1590/s0100-06832013000100023.

MORETTIN, Pedro A. **Ondas e Ondaletas**: da análise de fourier à análise de ondaletas de séries temporais. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2014. 326.p.

NALIN, Renan Storno; et al. Accounting for the spatial variation of phosphorus available explained by environmental covariates. **Geoderma Regional**, [S.L.], v. 32, p. 1-12, mar. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.geodrs.2022.e00594.

NIELSEN, D. R.; WENDROTH, O. **Spatial and temporal statistics:** Sampling field soil and their vegetation. Reiskirchen: Catena Verlag, 2003. 398 p. ISBN-13: 978-1593262594.

OLEA, R. A. A. **Practical primer on geostatistics**. U.S.: Geological Survey, 2009.

PARFITT, José Maria Barbat et al. Impacts of land leveling on lowland soil physical properties. **Revista Brasileira de Ciência do Solo**, [S.L.], v. 38, n. 1, p. 315-326, fev. 2014. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s0100-06832014000100032.

PERALTA, Nahuel Raúl; et al. Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, [S. l.], v. 110, p. 103–113, 2015. DOI: 10.1016/j.compag.2014.10.017.

PEREIRA, Roney Peterson *et al.* Linear mixed model for weight analysis in mice infected by Trypanosoma cruzi. **Acta Scientiarum. Health Sciences**, [S.l.], v. 42, n. 1, p. 1–14, 20 jul. 2020. Universidade Estadual de Maringá. DOI: http://dx.doi.org/10.4025/actascihealthsci.v42i1.49916.

PINHEIRO, Saulo Tavares. **Cartografia geotécnica para a cidade de Palmas-TO: descrição táctil visual do solo da região Sudeste, entre as avenidas LO-03 e LO-27.** 2018. 94 f. TCC (Graduação) - Curso de Engenharia Civil, Universidade Federal do Tocantins, Palmas – TO, 2018.

PINTO, L. F. S. *et al.* Solos de várzea do Sul do Brasil cultivados com arroz irrigado. *In:* GOMES, A. da S.; MAGALHÃES JR. A. M. (eds.). **Arroz irrigado no sul do Brasil**. Pelotas: Embrapa – Clima Temperado, 2004. p. 75-95.

REICHARDT, K., TIMM, L. C. **Soil, Plant and Atmosphere**: concepts, processes and applications. 1. ed. Springer, Basel. 2020. ISBN 978-3-030-19322-5.

ROMIĆ, D. C. *et al.* Modelling spatial and temporal variability of water quality from different monitoring stations using mixed effects model theory. **Science of The Total Environment**, [S.l.], v. 704, n. 1, p. 1–10, fev. 2020. Elsevier BV. DOI: http://dx.doi.org/10.1016/j.scitotenv.2019.135875.

ROSEMARY, F. et al. Exploring the spatial variability of soil properties in an Alfisol soil catena. **Catena**, [S.L.], v. 150, p. 53-61, mar. 2017. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2016.10.017.

SAHU, Sanjeev Kumar; et al. A Geostatistical Framework Predicting Zooplankton Abundance in a Large River: management implications towards potamoplankton sustainability. **Environmental Management**, [S.L.], v. 2, n. 1, p. 10-31, 7 jan. 2023.

Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s00267-023-01784-2.

SANTOS, Jonathas Jesus dos. **Modelagem de atributos químicos e espectrais dos solos de Feira de Santana – BA**. 2019. 121 f. Dissertação (Mestrado) - Curso de Modelagem em Ciências da Terra e do Ambiente, Universidade Estadual de Feira de Santana – UEFS, Feira de Santana – BA, 2019.

SCHNEIDER, K. *et al.* Temporal stability of soil moisture in various semi-arid steppe ecosystems and its application in remote sensing. **Journal of Hydrology**, [S. l.], v. 359, n. 1–2, p. 16–29, 2008. DOI: 10.1016/j.jhydrol.2008.06.016.

SELF, S. G.; LIANG, K. Y. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Test Under Nonstandard Conditions. **Journal of the American Statistical Association**, Alexandria, v. 82, n. 398, p. 605–610, 1987.

SHARMA, Priyanka; KAUR, Jaskaran; KATNORIA, Jatinder Kaur. Assessment of spatial variations in pollution load of agricultural soil samples of Ludhiana district, Punjab. **Environmental Monitoring and Assessment**, [S.L.], v. 195, n. 1, p. 1-13, 22 dez. 2022. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s10661-022-10816-z.

SLAETS, Johanna I. F.; BOEDDINGHAUS, Runa S.; PIEPHO, Hans Peter. Linear mixed models and geostatistics for designed experiments in soil science: Two entirely different methods or two sides of the same coin? **European Journal of Soil Science**, [S. l.], p. 1–22, 2020. DOI: 10.1111/ejss.12976.

SOARES, Amanda Nunes. **Adoção do software livre koha por um corpus de bibliotecas do Distrito Federal**: percepções bibliotecárias. 2022. 71 f. Monografia (Especialização) - Curso de Biblioteconomia, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília - DF, 2022.

SOARES, M. F., et al. Identifying Covariates to Assess the Spatial Variability of Saturated Soil Hydraulic Conductivity Using Robust Cokriging at the Watershed Scale.

J. Soil Sci. Plant Nutr. V. 20, p. 1491–1502. 2020. **https://doi.org/10.1007/s42729-020-00228-8**

SOARES, Mauricio Fornalski; et al. Assessing the spatial variability of saturated soil hydraulic conductivity at the watershed scale using the sequential Gaussian co-simulation method. Catena, [S.L.], v. 221, p. 1-12, fev. 2023. Elsevier BV. **http://dx.doi.org/10.1016/j.catena.2022.106756**

SOLIGO, Matheus Flesch. **Métodos de amostragem para a modelagem espacial de fósforo disponível no solo**. 2021. 64 f. Dissertação (Mestrado) - Curso de Ciência do Solo, Universidade Federal de Santa Maria, Santa Maria, 2021.

SOSBAI. **Recomendações técnicas da pesquisa para o Sul do Brasil – XXIX Reunião Técnica da Cultura do Arroz Irrigado**, 2018.

STEIN, Michael L. **Interpolation of spatial data:** some theory for kriging. Nova York: Springer-Verlag, 1999.

STRAM, Daniel O.; LEE, Jae Won. Corrections to “Variance components testing in the longitudinal mixed effects model” by D. O. Stram and J. W. Lee; 50, 1171-1177, 1994. **Biometrics**, [S.l.], v. 51, n. 3, p. 1196, 1995. DOI: 10.2307/2533038.

SUZUKI, Luis Eduardo Akiyoshi Sanches *et al*. Challenges in the Management of Environmentally Fragile Sandy Soils in Southern Brazil. **Soil Systems**, [S.L.], v. 7, n. 1, p. 9-24, 2 fev. 2023. MDPI AG. http://dx.doi.org/10.3390/soilsystems7010009.

TANG, Qiuyue; et al. Variability and driving factors of the soil saturated hydraulic conductivity along the horizontal and vertical directions in the upper catchment of Benggang. **Catena**, [S.L.], v. 222, p. 1-17, mar. 2023. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2022.106810.

TIAN, Mi; et al. Spatial-temporal variability and influence factors of Cd in soils of Guangxi, China. **Plos One**, [S.L.], v. 18, n. 1, p. 1-14, 10 jan. 2023. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pone.0279980.

TIMM, Luís Carlos et al. Assessment of land levelling effects on lowland soil quality indicators and water retention evaluated by multivariate and geostatistical analyses. **Land Degradation & Development**, [S.L.], v. 31, n. 8, p. 959-974, 7 jan. 2020. Wiley. http://dx.doi.org/10.1002/ldr.3529.

TORABI, Mahmoud; JIANG, Jiming. Estimation of mean squared prediction error of empirically spatial predictor of small area means under a linear mixed model. **Journal of Statistical Planning and Inference**, [S.l.], v. 208, n. 8, p. 82–93, set. 2020. Elsevier, BV. DOI: http://dx.doi.org/10.1016/j.jspi.2020.02.001.

TRANGMAR, B. B. et al. Spatial Dependence and Interpolation of Soil Properties in West Sumatra, Indonesia: i. anisotropic variation. **Soil Science Society Of America Journal**, [S.L.], v. 50, n. 6, p. 1391-1395, nov. 1986. Wiley. http://dx.doi.org/10.2136/sssaj1986.03615995005000060004x.

VIEIRA, S. R.; DECHEN, S. C. F. Spatial variability studies in São Paulo, Brazil along the last twenty five years. **Bragantia**, *[S. l.]*, v. 69, n. suppl, p. 53–66, 2010. Disponível em: https://doi.org/10.1590/s0006-87052010000500007.

YAMAMOTO, Jorge Kazuo. **Estatística, Análise e interpolação de dados geoespaciais**. São Paulo. 2020, 308p.

YAMAMOTO, Jorge Kazuo; LANDIM, Paulo Milton Barbosa. **Geoestatística**: conceitos e aplicações. São Paulo, 2013. 215p.

ZIMMERMAN, Dale L. Another look at anisotropy in geostatistics. **Mathematical Geology**, [S. l.], v. 25, n. 4, p. 453–470, 1993. DOI: 10.1007/BF00894779.