**UNIVERSIDADE FEDERAL DE PELOTAS**

**Centro de Desenvolvimento Tecnológico**

**Programa de Pós-Graduação em Recursos Hídricos**


**Tese de Doutorado**

# Avaliação da co-simulação sequencial gaussiana na geração de campos aleatórios da condutividade hidráulica do solo saturado

**Mauricio Fornalski Soares**

**Pelotas, 2021**

**Mauricio Fornalski Soares**

# Avaliação da co-simulação sequencial gaussiana na geração de campos aleatórios da condutividade hidráulica do solo saturado

Tese apresentada ao Programa de Pós-Graduação em Recursos Hídricos do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de doutor em Recursos Hídricos.

Orientador: Prof. Dr. Luís Carlos Timm

Coorientador: Profª. Drª. Tirzah Moreira Siqueira

Pelotas, 2021

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

S676a Soares, Mauricio Fornalski

Avaliação da co-simulação sequencial gaussiana na geração de campos aleatórios da condutividade hidráulica do solo saturado / Mauricio Fornalski Soares ; Luís Carlos Timm, orientador ; Tirzah Moreira Siqueira, coorientadora. — Pelotas, 2021.

102 f. : il.

Tese (Doutorado) — Programa de Pós-Graduação em Recursos Hídricos, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2021.

1. Simulação geoestatística. 2. Variáveis secundárias. 3. Componentes principais. 4. Análise de incertezas. 5. Pedoestatística. I. Timm, Luís Carlos, orient. II. Siqueira, Tirzah Moreira, coorient. III. Título.

CDD : 627

Elaborada por Maria Inez Figueiredo Figas Machado CRB: 10/1612

## Agradecimentos

Primeiramente à minha família, minha mãe Sofia Verônica Fornalski Soares, meu pai José Nélson Soares, minha irmã Isaura Helena Fornalski Soares e meu irmão Marcelo Fornalski Soares por todo apoio e incentivo.

À Universidade Federal de Pelotas, e especialmente ao Programa de Pós-Graduação em Recursos Hídricos, por ter proporcionado minha formação acadêmica.

Ao meu orientador, Professor Doutor Luís Carlos Timm, pela orientação, apoio, confiança, empenho dedicado à elaboração deste trabalho e por todo auxílio prestado a mim, durante o período de realização deste curso de pós-graduação.

Agradeço a todos os professores e técnicos por me proporcionarem um grande aprimoramento técnico, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

Meus agradecimentos aos amigos e colegas do Programa de Pós-Graduação em Recursos Hídricos. Agradeço também aos colegas do Programa de Pós-Graduação em Manejo e Conservação do Solo e da Água da Universidade Federal de Pelotas.

Aos produtores na qual concederam as áreas para realização das coletas de solo.

A todos, aqui citados ou não, que contribuíram de alguma forma, direta ou indiretamente, para a realização deste trabalho.

## Resumo

Soares, Mauricio Fornalski. **Avaliação da co-simulação sequencial gaussiana na geração de campos aleatórios da condutividade hidráulica do solo saturado.** 2021. 102f. Tese (Doutorado em Recursos Hídricos) – Programa de Pós-Graduação em Recursos Hídricos, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2021.

A dinâmica da água no solo é um ponto central na regulação dos ciclos hidrológicos em bacias hidrográficas, no entanto, a representação espacial dos processos ligados ao movimento da água no solo em modelos matemáticos ainda representa um grande desafio, dada a complexidade dessas interações. A condutividade hidráulica do solo saturado ($K_{Sat}$) é um parâmetro de entrada chave para modelar o fluxo de água no solo e aplicável a projetos de irrigação e drenagem, para a modelagem do escoamento superficial e para a contenção da perda de solo e de água por erosão. Existem muitas formas para a determinação da $K_{Sat}$, com métodos de mensuração in situ e de laboratório. No entanto, estas análises são caras e demoradas, principalmente em escala de bacias hidrográficas, devido à magnitude das áreas e também a grande heterogeneidade das características do solo e da paisagem que controlam este fenômeno. O presente trabalho tem como objetivo compreender as interações de atributos hidrológicos do solo e características da paisagem com a $K_{Sat}$ a partir de análises estatísticas multivariadas e, posteriormente, aplicar estes atributos/características como variáveis secundarias na espacialização da $K_{Sat}$ utilizando métodos geoestatísticos, aprimorando os resultados obtidos e reduzindo as incertezas associadas. A tese foi dividida em dois artigos complementares, porem independentes. O primeiro artigo é intitulado "Examining the spatial relationship between saturated soil hydraulic conductivity and co-driven factors at watershed scale using multivariate and geostatistical methods" e o segundo artigo é intitulado "Investigating the spatial uncertainties of saturated soil hydraulic conductivity and co-regionalized variables at the watershed scale using multivariate and Gaussian co-simulation analyses". A pesquisa para elaboração de ambos os artigos foi conduzida na bacia hidrográfica da Sanga Ellert, que possui cerca de 70 hectares onde o uso e ocupação do solo combinam áreas de produção agrícola e florestas nativas. Foram coletadas informações do solo e da paisagem em 178 pontos amostrais distribuídos na área da bacia a partir de levantamento terrestre e remoto. Para o primeiro artigo proposto, os mapas das duas componentes principais demonstra uma relação muito próxima com o uso do solo e com as classes texturais, permitindo afirmar que estas variáveis são sistemáticas para a $K_{Sat}$ na área de estudo. No segundo artigo, entre atributos hidrológicos do solo, características da paisagem e as componentes principais, os resultados demonstram que a densidade do solo e a macroporosidade são as variáveis secundárias mais confiáveis para aprimorar a avaliação espacial de $K_{Sat}$ por co-simulação sequencial Gaussiana.



**Palavras Chaves:** Simulação geoestatística; Variáveis secundárias; Componentes principais; Análise de Incertezas; Pedoestatística.

## Abstract

Soares, Mauricio Fornalski. **Sequential Gaussian co-simulation evaluation for the generation of random fields of the saurated soil hydraulic conductivity.** 2021. 102p. Thesis (Doctor's Degree in Water Resources) – Graduate Program in Water Resources, Center of Technological Development, Federal University of Pelotas, Pelotas, 2021.

The dynamics of soil water is a central point for hydrological cycles in watersheds; however, the spatial representation of processes related to soil water movement in mathematical models still represents a major challenge, given the complexity of these interactions. Saturated soil hydraulic conductivity ($K_{Sat}$) is a key input parameter for modeling soil water flow and applicable to irrigation and drainage projects, for modeling runoff, and for containing soil and water loss by erosion. There are many ways to determine $K_{Sat}$, with in situ and laboratory measurement methods. However, these analyzes are expensive and time-consuming, especially at the watershed scale, due to the magnitude of the areas and also the great heterogeneity of the soil and landscape characteristics that control this phenomenon. The present work aims to understand the interactions of soil hydrological attributes and landscape characteristics with $K_{Sat}$ from multivariate statistical analysis and, subsequently, apply these attributes/characteristics as secondary variables for spatial modeling of $K_{Sat}$ using geostatistical methods, improving the results obtained and reducing the associated uncertainties. This thesis was divided into two complementary but independent original papers. The first paper is entitled "Examining the spatial relationship between saturated soil hydraulic conductivity and co-driven factors at a watershed scale using multivariate and geostatistical methods" and the second paper is entitled "Investigating the spatial uncertainty of saturated soil hydraulic conductivity and co-regionalized variables at the watershed scale using multivariate and Gaussian co-simulation analyses". The research for the elaboration of both original papers was conducted in the Ellert Creek Watershed, which has about 70 hectares where land use and combine agricultural production areas and native forests. Soil and landscape information was collected at 178 sampling points distributed iat the watershed area from a terrestrial and remote survey. On the first paper, the maps of the two principal components showed a very close relationship with land use and soil textural classes, allowing to state that these variables are systematic for $K_{Sat}$ in the study area. For the second paper, among soil hydrological attributes, landscape characteristics and principal components, the results demonstrate that soil density and macroporosity are the most reliable secondary variables to improve the spatial evaluation of $K_{Sat}$ by Sequential Gaussian Co-simulation.



**Key words:** Geostatistic simulation; Secondary variables; Principal components; Uncertainty; Pedostatstics.

Sumário

## 1. Introdução

A condutividade hidráulica saturada do solo ($K_{Sat}$) é um parâmetro essencial para a compreensão do movimento da água no solo e, consequentemente, um dado de entrada importante para modelos hidrológicos (BAGARELLO *et al*., 2020). A $K_{Sat}$ é definido pela taxa de infiltração em solos em condição de estado estacionário (NAGANNA *et al*. , 2017). Determinar valores bem representativos de $K_{Sat}$ e sua distribuição no espaço é muito importante em várias circunstâncias porque ela controla vários processos hidrológicos, por exemplo, o escoamento superficial direto, recarga de água subterrânea, erosão e perda de solo e movimento de contaminantes do solo, afetando todo o meio ambiente em uma bacia hidrográfica (BECKER *et al*., 2018). As estimativas de $K_{Sat}$ podem apoiar decisões de gestão de recursos relacionadas à conservação de água, sistemas de irrigação, aplicação de fertilizantes, drenagem, mitigação de solutos e manejo de cultivos agrícolas (GOOTMAN *et al*., 2020). Esses desafios sugerem a importância de métodos para estimar $K_{Sat}$ que sejam eficientes para determinar com precisão as melhores práticas de manejo do solo para inclusive melhorar os rendimentos agrícolas e ao mesmo tempo ser sustentável (BAGARELLO *et al*., 2020; ZHANG; SCHAAP, 2019).

Os atributos do solo se comportam de maneira diferente no ambiente devido a fatores como material de origem do solo, clima, topografia, cobertura vegetal e usos do solo. Eles são o principal motor de heterogeneidade do solo. É amplamente reconhecido que a heterogeneidade é maior para atributos hidrológicos do solo e especialmente para condutividade hidráulica do solo saturado que é excessivamente variável no espaço (BAIAMONTE *et al*., 2017; CENTENO *et al*., 2020; YONG *et al*., 2020). Os fatores que levam à variabilidade espacial do $K_{Sat}$ são mais complexos na escala da bacia

hidrográfica e uma razão para a complexidade são os diferentes usos do solo, que levam a diferenças significativas na condutividade hidráulica do solo (GOOTMAN *et al*., 2020; TRINH *et al*., 2018).

Métodos geoestatísticos têm sido aplicados em ciências do solo e ambientais com o objetivo de visualizar ou simular a estrutura espacial variabilidade dos atributos do solo (JIN; LV, 2020; SOARES *et al*., 2020). Dentre os métodos geoestatísticos, destacam-se a krigagem e a simulação estocástica, por serem amplamente utilizadas para modelar a variabilidade de atributos do solo. A simulação estocástica foi projetada para evitar o efeito de suavização da krigagem (SANTOS *et al*., 2021). Além disso, a simulação estocástica tem a capacidade de não apenas prever a distribuição espacial de uma propriedade em questão, mas também avaliar a incerteza local e espacial sobre as estimativas, que devem ser incorporadas nos processos de tomada de decisão (GOOVAERTS, 2001; ZIEGEL *et al*., 1998). A simulação sequencial gaussiana (SSG) é amplamente difundida no mapeamento de solo por ser relativamente simples de implementar. A co-simulação sequencial gaussiana (CSSG) é uma extensão bivariada do SSG proposta como uma ferramenta para usar uma variável secundária a fim de melhorar os resultados da simulação da variável alvo (ou seja, aquela que será simulada) com base em um modelo linear de ajuste de co-regionalização (LMC) e algoritmo de co-krigagem simples (JIN; LV, 2020). Nesse sentido, o CSSG pode ser uma alternativa para atributos de solo com grande variabilidade espacial, como o $K_{Sat}$.

Para executar o algoritmo CSSG é necessário selecionar variáveis secundárias que contribuem para a variável principal. No caso do $K_{Sat}$, os atributos físicos e hidrológicos do solo, bem como as características ambientais, como o uso do solo, costumam ser bons candidatos a variáveis secundárias (GOOTMAN *et al* 2020; ZHAO LV, 2020). Às vezes, os atributos que poderiam contribuir para a precisão do CSSG não correspondem a essa expectativa. Quando existe um banco de dados com diversas variáveis regionalizadas em mãos, uma alternativa seria compactar essas informações em conjuntos adimensionais e usá-los como variáveis secundárias. A análise de componentes principais (PCA) é uma técnica que indica associações entre variáveis, reduzindo assim o tamanho do número de dados a partir do

agrupamento de variáveis (JOHN *et al*., 2021). Esses grupos de variáveis gerados pela PCA podem ser regionalizados e posteriormente utilizados como variáveis secundárias no processo CSSG se apresentarem correlação significativa com a variável principal, levando a uma simulação bivariada uma variável secundária que, embora não tenha magnitude física, representa o agrupamento de vários recursos, tornando o CSSG um processo multivariado.

Em países em desenvolvimento, como o Brasil, a falta de informações hidrológicas para bacias hidrográficas é uma das principais dificuldades para a gestão dos recursos hídricos. Os planejadores muitas vezes precisam tomar decisões sobre bacias hidrográficas não avaliadas ou aquelas que têm poucos dados espaciais e temporais disponíveis (BESKOW *et al*., 2016). A Bacia Hidrográfica da Sanga Ellert (BHSE), por exemplo, é uma bacia de cabeceira cujo principal curso d'água deságua diretamente na Bacia Hidrográfica do Arroio Pelotas (BHAP), uma grande bacia hidrográfica que impacta mais de meio milhão de habitantes da região para abastecimento direto de água potável e todas as outras atividades humanas dependentes dos recursos hídricos. Além disso, o BHSE está sujeito a impactos decorrentes do manejo intensivo do solo para a agricultura e pecuária, sendo vulnerável à erosão hídrica e inundações, causando perdas de solo e nutrientes resultando em danos econômicos e sociais (SOARES *et al*., 2020).

A compreensão dos processos hidrológicos é essencial para a adoção de práticas de manejo sustentável em bacias hidrográficas como a BHSE e, para isso, o conhecimento da distribuição espacial dos atributos hidrológicos do solo como a $K_{Sat}$, que é um parâmetro crítico para a aplicação de modelos hidrológicos (WANG *et al*. , 2013; HU *et al*., 2015), pode apoiar decisões de gestão de recursos hídricos.

## 2. Hipóteses

i. Há incorporação de atributos físicos e hidrológicos do solo bem como de feições do ambiente em processos estatísticos bivariados e multivariados contribuem para uma melhor compreensão da condutividade hidráulica do solo saturado em escala de bacia hidrográfica.

ii. A análise de componentes principais pode identificar e agrupar atributos físicos e hidrológicos do solo e características do ambiente, como o uso e cobertura do solo, reduzindo o tamanho do banco de dados, retornando grupos de variáveis que explicam a variância da condutividade hidráulica do solo saturado na área de estudo. Os autovalores disponibilizados pela análise de componentes principais podem ser estabelecidos como variáveis regionalizadas e posteriormente, a partir de ferramentas da geoestatística, interpolados por krigagem ordinária.

iii. O uso de algoritmos de Simulação Sequencial Gaussiana e co-simulação Sequencial Gaussiana são eficientes para estimar a condutividade hidráulica do solo saturado em locais não amostrados na área da bacia hidrográfica da Sanga Ellert. Para a co-simulação sequencial Gaussiana, o uso de atributos físicos hidrológicos do solo, características ambientais e de variáveis advindas da análise de componentes principais como variáveis secundarias irá incorporar eficiência ao processo, reduzindo as incertezas associadas à espacialização da condutividade hidráulica do solo saturado em escala de bacia hidrográfica.

# 3. Objetivos

## 3.1 Objetivo geral

O objetivo deste trabalho foi avaliar a variabilidade espacial da condutividade hidráulica do solo saturado na bacia hidrográfica Sanga Ellert, usando algoritmos de simulação e co-simulação geoestatística para a espacialização e análise das incertezas associadas.

## 3.2 Objetivos específicos

i. Examinar a influência dos atributos físico-hidrológicos do solo, do uso e ocupação do solo e do perfil topográficos na condutividade hidráulica do solo saturado em escala de bacia hidrográfica, por meio da análise de componentes principais, bem como avaliar a distribuição espacial dos componentes gerados por ferramentas geoestatísticas e suas associações com padrões não numéricos da paisagem.

ii. Estimar e avaliar a variabilidade espacial da condutividade hidráulica do solo saturado por Simulação Sequencial Gaussiana e co-simulação sequencial Gaussiana.

iii. Analisar diferentes atributos físicos e hidrológicos do solo e de uso do solo, bem como as variáveis geradas pela análise de componentes principais, como variáveis secundárias para a co-simulação Sequencial Gaussiana.

iv. Quantificar as incertezas globais e espaciais da Simulação Sequencial Gaussiana e para cada processo com diferentes variáveis secundárias da co-simulação sequencial Gaussiana.

# 4. Revisão da literatura

## 4.1 Condutividade hidráulica do solo saturado: Dinâmica da água no solo

A compreensão espacial da dinâmica da água no sistema solo-planta-atmosfera é fundamental para o manejo da água e do solo em bacias hidrográficas considerando as modificações antrópicas da superfície e a heterogeneidade natural dos solos e do terreno (REICHARDT; TIMM, 2020). A natureza das interações entre as características ecológicas, físicas e hidrológicas que determinam os efeitos da mudança da cobertura da terra na hidrologia superficial e subterrânea não é bem compreendida em ambientes naturais e humanos (WINCKLER *et al*., 2018). Diferentes tipos de solos e mudanças nos usos da terra agrícola são centrais para os estudos de mudanças climáticas porque estão situados na interface entre os ecossistemas e a sociedade.

A condutividade hidráulica do solo saturado ($K_{Sat}$) é um parâmetro essencial para a compreensão do movimento da água no solo e, consequentemente, um importante dado de entrada para modelos hidrológicos (BAGARELLO *et al*., 2020). A $K_{Sat}$ é definida pela taxa de infiltração no solo em condição de saturação, quando esta apresenta um estado estacionário (NAGANNA *et al*., 2017). A determinação de valores representativos de $K_{Sat}$ e sua distribuição no espaço é necessária em muitas circunstâncias, pois esta influencia vários processos hidrológicos, por exemplo, o escoamento superficial direto, recarga de água subterrânea, erosão e perda de solo e o movimento de contaminantes do solo, afetando todo o ambiente em uma bacia hidrográfica (BECKER *et al*., 2018). A espacialização da $K_{Sat}$ para uma determinada área

pode fornecer informações uteis para a tomada de decisões na gestão de recursos em bacias hidrográficas, relacionados à conservação de água, sistemas de irrigação, aplicação de fertilizantes, drenagem, mitigação de solutos e gerenciamento agrícola (GOOTMAN *et al*., 2020). A influencia da $K_{Sat}$ em todos os processos descritos sugere a importância de métodos para estimar e espacializar este atributo de forma eficiente, podendo assim ser essa uma importante ferramenta para determinar práticas sustentáveis de manejo do solo (ZHANG *et al*., 2019; BAGARELLO *et al*., 2020).

## 4.2 Condutividade hidráulica do solo saturado: Uma variável integradora

De acordo com Urso *et al*. (2019), uma componente sistemática é uma variável explicativa ou independente ($x_1$, ..., $x_n$), que descreve cada domínio $X_i$ de um conjunto de dados, onde i = 1, ..., p é o número total de domínios no conjunto de dados. Os valores das variáveis explicativas são tratados como fixos e não como variáveis aleatórias/integradoras. Muitos estudos demonstram que a condutividade hidráulica do solo saturado ($K_{Sat}$) se comporta como uma variável aleatória integradora. Principalmente nas camadas superficiais do solo, o $K_{Sat}$ aumenta gradativamente de solos degradados por uso intensivo, como pastagem e lavouras anuais, para solos sob condições florestais, nativas ou implantadas (SHI *et al.*, 2016; ARCHER *et al.*, 2013; BECKER *et al*., 2018). Em outro exemplo, solos sob condições naturais, como os de florestas nativas, estão associados a maiores taxas de infiltração de água e consequentemente, a uma $K_{Sat}$ mais elevada (KURNIANTO *et al*., 2018; ARCHER *et al*., 2013) bem como à uma menor geração de escoamento superficial direto (KURNIANTO *et al*., 2018; LIU *et al*., 2018) do que solos sob ocupação antrópica.

A $K_{Sat}$ pode ser considerada uma variável integradora, regida por diversos fatores ambientais considerados componentes sistemáticas. Muitos destes componentes que controlam a $K_{Sat}$ são relativamente difíceis para se quantificar, causando problemas para a modelagem a partir do uso de variáveis secundárias. O uso do solo é uma informação chave para entender o comportamento da $K_{Sat}$. Além do mais, existem diversas plataformas que

fornecem estes dados por ferramentas de sensoriamento remoto, o que representa uma grande facilidade de aquisição. Porém, a quantificação destes dados para aplicação em modelos matemáticos ainda é um desafio, uma vez que são estacionários (discretos) e qualitativos, abrindo diversos questionamentos para o uso eficiente destes dados de forma numérica (HASTIE, 2004).

Outra característica fundamental para a compreensão da $K_{Sat}$ se refere as frações texturais do solo. Embora as frações solo, de forma isolada, não sejam tão complexas para serem quantificadas, é necessário entender que argila, silte e areia atuam juntos no sistema de solo e, talvez, o vetor de classes texturais seja uma representação mais eficiente desse conjunto. Porem as classes texturais determinadas a partir do triangulo textural (HILLEL, 2003) retornam valores não numéricos e, tal qual o uso do solo, estacionários.

### 4.3 Condutividade hidráulica do solo saturado em modelos matemáticos

Os processos de infiltração de água no solo, projetos de irrigação e drenagem, perdas de fertilizantes e de solo por erosão e de substâncias químicas por lixiviação são geralmente relacionados ao fluxo de água, o qual influencia todo o processo de utilização dos recursos do solo e da água (Santos *et al*, 2021). Neste sentido, a condutividade hidráulica saturada é um atributo hidrológico imprescindível para a modelagem chuva-vazão, dimensionamento de sistemas de irrigação e drenagem, transporte de solutos no solo, recarga de aquíferos, escoamento superficial e transporte de sedimentos. A condutividade hidráulica do solo reflete a maior ou menor resistência que a matriz do solo oferece à passagem da água (Soares *et al*, 2020).

Dentre as variáveis relacionadas com a dinâmica da água no solo, a $K_{Sat}$ exerce influência direta em processos relacionados ao ciclo hidrológico como um importante parâmetro hidráulico, pois afeta o meio ambiente controlando a infiltração da água, a taxa de aplicação de água em sistemas de irrigação e, consequentemente, o movimento da água através do solo. Mapear e monitorar

com precisão a $K_{Sat}$ é fundamental para uma gestão eficiente dos recursos do solo e da água em relação a sustentabilidade agrícola, bem como evitar impactos ambientais negativos em bacias hidrográficas (Siqueira *et al.*, 2019).

Diversos pesquisadores (Siqueira *et al.*, 2019; Santos *et al*, 2021; Hu et al., 2007) tem utilizado ferramentas de espacialização para a $K_{Sat}$ com o objetivo de usar essas informações como dados de entrada para modelos hidrológicos de chuva-vazão, para avaliação de escoamento superficial direto, drenagem e irrigação e manejo de bacias hidrográficas. Siqueira et al., (2019) aplicaram a simulação sequencial gaussiana para geração de 100 campos aleatórios da $K_{Sat}$ em uma bacia hidrográfica de pequeno porte, identificando diversas escalas de heterogeneidade espacial e temporal, possivelmente relacionadas a fatores como clima, uso e cobertura do solo, adensamento da malha amostral e também ao grau de compactação do solo. Nesse sentido, é possível identificar a importância do conhecimento detalhado da condutividade hidráulica do solo saturado para simulação hidrológica, visto que tal propriedade está intimamente relacionada com a infiltração de água no solo, influenciando assim no balanço hídrico e consequentemente nas vazões máximas.

Santos et al., (2021) utilizaram o modelo hidrológico LISEM (Limburg Soil Erosion Model) para avaliar a influência da variabilidade espacial da condutividade hidráulica do solo saturado e o efeito da umidade do solo sobre hidrogramas de escoamento superficial direto originado de eventos extremos de chuva, com base em simulação em uma área de bacia hidrográfica. Os autores ainda relatam que a condutividade hidráulica do solo saturado sofre influência de atributos do solo tais como: estrutura, textura, homogeneidade, densidade do solo, densidade das partículas, porosidade total e macro e micro porosidade, e também de qualquer fator que exerça alguma influência no tamanho e na configuração dos poros.

## 4.4 Pedoestatística: Conceito e aplicações

O presente trabalho consolida o termo “Pedoestatística” como o conjunto de técnicas estatísticas, desenvolvidas para as mais diversas fontes de conhecimento, que são uteis e aplicáveis às ciências do solo. Esta transferência *inter* metodologias é hábito frequente entre pesquisadores de diversas instancias e busca aprimorar um determinado conhecimento a partir de técnicas pré-existentes advindas de outros campos da ciência (COBURN; TIMOTHY, 2004). Existem diversos exemplos onde metodologias estatísticas e matemáticas desenvolvidas para uso militar, medicina, econometria, mineração, administração publica, computação gráfica, etc., são transferidas entre si ou para outros campos do conhecimento, explicando fenômenos que ainda não eram inteiramente compreendidos pelos estudiosos dessas áreas especificas, aumentando o leque de possibilidades para a pesquisa e desenvolvimento (MIZUTA *et al*., 2021). Esta conceitualização nos permite afirmar que a Pedoestatística está presente em praticamente todos os estudos ligados a Ciência do Solo, fazendo parte desde o planejamento de um determinado experimento como coleta de dados e sua organização, passando pela interpretação, análise e representação matemática em um dado processo científico.

A abrangência da Pedoestatística como campo de estudo é dependente de dois eixos bem definidos e pode ser caracterizada como o produto da amplitude dos mesmos, onde um eixo é o das ciências do solo e outro eixo o da estatística (BERKHOUT *et al*., 2019). Colocando em perspectiva, a abrangência da Pedoestatística é efeito da convergência cartesiana destes eixos para aquilo que já está consolidado em ambas as ciências, para o que vem sendo desenvolvido atualmente e também das inúmeras possibilidades que possam ser vislumbradas para estas ciências no futuro. No eixo das ciências do solo podemos exemplificar a aplicação da Pedoestatística para a química e fertilidade do solo, física do solo, pedologia, irrigação e drenagem, hidrologia, conservação dos solos e diversas outras áreas que possam fornecer organização e armazenagem de informações sobre um domínio específico (NADKARNI *et al*., 1999). No eixo da estatística, é importante enfatizar que

este está intimamente ligado à evolução tecnológica visto que a aplicação de determinadas metodologias é precedida de uma grande capacidade de processamento de dados. A base da estatística moderna, como as medidas de posição, dispersão e frequência, embora habitualmente não exijam grande capacidade de processamento, são requisitos fundamentais para uma análise exploratória dos dados, fornecendo informações indispensáveis para os processos estatísticos subsequentes (COBURN; TIMOTHY, 2004). Como processos subsequentes, podemos citar desde métodos relativamente simples, como análise de correlação e regressão, passando pela estatística bivariada ou multivariada e posteriormente indo para onde a estatística subdivide-se em análises espaciais e temporais, explorando, por exemplo, ferramentas da geoestatística, simulação de dados, mineração de dados e inteligência artificial (VARIAN *et al*, 2014).

## 4.5 Ferramentas pedoestatísticas aplicadas

### 4.5.1 Estatística básica e análise exploratória

A Estatística Descritiva se preocupa com a organização, apresentação e sintetização de dados. Utilizam gráficos, tabelas e medidas descritivas como ferramentas. Utilizada na etapa inicial da análise, destinada a obter informações que indicam possíveis modelos a serem utilizados numa fase final que seria a chamada inferência estatística.

Dentre as medidas estatísticas mais comuns podemos citar as medidas de tendência central ou posição, que são utilizadas para resumir, em um único número, o conjunto de dados observados da variável em estudo. Usualmente emprega-se uma das seguintes medidas de posição: média, mediana ou moda; As medidas de dispersão, que servem para indicar o quanto os dados se apresentam dispersos, ou afastados, em relação ao seu valor médio, por exemplo, desvio padrão, variância, coeficiente de variação, etc.; e as medidas

que indicam o formato da distribuição, como assimetria e curtose (Conover, 1981).

Dentre os gráficos utilizados para análise exploratória dos dados o histograma é um dos mais utilizados. O histograma é um gráfico de barras que demonstra uma distribuição de frequências. No histograma, a base de cada uma das barras representa uma classe e a altura representa a quantidade ou frequência absoluta com que o valor de cada classe ocorre. Ao mesmo tempo, ele pode ser utilizado como um indicador de dispersão de processos (Kramer, 1988).

Uma alternativa para uma visualização resumida dos dados que tem sido bastante utilizada é a dos gráficos boxplots. Nesse gráfico, a barra transversal do gráfico do boxplot mostra a mediana para cada conjunto de dados. O comprimento da caixa reflete o intervalo interquartil, e as caudas verticais (ou barras) do boxplot são marcadas pelos extremos (o maior ou o menor dado observado); no entanto, eles não se qualificam como outliers. Outliers são definidos como dados ≥ 1,5 vezes maiores (cerca superior) ou ≥ 1,5 vezes mais baixos (cerca inferior) do que o intervalo interquartil (Kramer, 1988).

Um coeficiente de correlação mede o grau pelo qual duas variáveis tendem a mudar juntas. O coeficiente descreve a força e a direção da relação. Correlação de momento de produto de Pearson. A correlação de Pearson avalia a relação linear entre duas variáveis contínuas. Uma relação é linear quando a mudança em uma variável é associada a uma mudança proporcional na outra variável (Conover, 1981).

### 4.5.2 Análise de componentes principais

A análise de componentes principais (PCA) é um método estatístico multivariado e tem como finalidade modelar e interpretar grandes conjuntos de dados, reduzindo sua dimensionalidade e extraindo informações úteis para avaliação e gestão da qualidade do solo (HAIR *et al*., 2009). A análise de componentes principais tem como objetivo reorganizar diversas variáveis de

um conjunto de dados, com as novas variáveis correspondentes ao agrupamento de variáveis que exerçam influencia de forma conjunta. A PCA transforma muitas variáveis em alguns indicadores abrangentes, chamados de componentes principais. Cada componente principal pode refletir a maioria das informações da variável original.

A PCA é uma técnica que indica associações entre variáveis, reduzindo assim o tamanho do número de dados e agrupando aqueles com maior semelhança (MUSHINA *et al*., 2020). Os componentes principais (PCs) são selecionados com base no fato de que explicam pelo menos 70% da variância total de cada conjunto de dados e que a variância explicada mínima por PC é superior a 20%. Esta etapa fornece um número n de componentes para o PCA com todos os atributos que foram organizados como um banco de dados espacial e então usados na análise geoestatística (LIBOHOVA *et al*., 2018).

### 4.5.1 Ferramentas da geoestatística e espacialização de dados

A geoestatística é o ramo da estatística aplicada que desenvolve e aplica modelos para representar fenômenos naturais cujas propriedades variam em função da localização espacial dos pontos de observação (MATHERON, 1962). O desenvolvimento da geoestatística nos anos 60 foi resultado da necessidade de se ter um método para avaliação dos depósitos de reservas minerais (GOOVAERTS, 1997; VIEIRA, 2000).

Segundo Clark (1979), a geoestatística surgiu para realizar a estimação de ouro em reservas de minas e essa foi sua utilidade por um período considerável. Contudo, suas técnicas podem ser usadas em qualquer situação em que medidas amostrais sejam feitas em um espaço e de onde se espera que as observações da variável em estudo sejam função da sua posição no espaço e, ainda, que exista relação de dependência espacial entre os valores da variável nos diferentes pontos amostrados.

Métodos geoestatísticos têm sido aplicados em ciências do solo e ambientais com o objetivo de visualizar ou simular a estrutura espacial e a variabilidade dos atributos do solo (ZHAO; LV, 2020; SOARES *et al*., 2020).

Dentre os métodos geoestatísticos destacam-se a krigagem e a simulação estocástica amplamente utilizadas para modelar a variabilidade dos atributos do solo. A simulação estocástica foi projetada para evitar o efeito de suavização da krigagem (ISAAKS; SRIVASTAVA, 1988). Além disso, a simulação estocástica tem a capacidade de não apenas prever a distribuição espacial da propriedade em questão, mas também avaliar a incerteza local e espacial sobre as estimativas, que devem ser incorporadas nos processos de tomada de decisão (GOOVAERTS, 1997; DEUTSCH; JOURNEL , 1998). A simulação sequencial gaussiana (SGS) é amplamente difundida no mapeamento de solo devido à melhoria da velocidade de execução e simples de implementar a simulação. A co-simulação sequencial gaussiana (SGCS) é uma extensão bivariada do SGS proposta como uma ferramenta para usar uma variável secundária a fim de melhorar os resultados da simulação da variável alvo (ou seja, aquela que será simulada) com base em um modelo linear de co-regionalização (LMC) algoritmo de ajuste e co-krigagem simples. (LV; LIU, 2019). Nesse sentido, o SGCS pode ser uma alternativa para atributos de solo com grande variabilidade espacial, como para a $K_{Sat}$.

## 5. Organização metodológica e execução da pesquisa

O estudo foi realizado na Bacia hidrográfica da Sanga Ellert que é uma sub-bacia da Bacia Hidrográfica do Rio Pelotas, localizada no município de Canguçu, sul do estado do Rio Grande do Sul (RS), aproximadamente 50 km a do município de Pelotas. A BHSE tem uma área de aproximadamente 0,7 $km^2$, e a altitude varia de 310 a 419 m.

De acordo com a classificação climática de Köppen o clima da região é do tipo Cfa, caracterizado por ser um clima mesotérmico indicando condições subtropicais úmidas e com temperatura média anual de 18 ° C, com verões quentes e invernos frios (KUINCHTNER; BURIOL 2001). A precipitação é bem distribuída ao longo do ano, e seu valor médio anual gira em torno de 1350 mm (IBGE 1986). O relevo regional varia de ondulado a forte ondulante, com predominância de mata nativa ou arbustos.

A bacia hidrográfica compreende apenas uma classe de solo, os Neossolos (Sistema Brasileiro de Classificação de Solos), identificados por meio do mapa de solos desenvolvido pela Embrapa (1999), e atualizado pela última versão do Sistema Brasileiro de Classificação de Solos (EMBRAPA 2006). A BHSE foi escolhida por sua importância econômica e social para o município de Canguçu, área fortemente dominada pela agricultura familiar. O principal curso de água da BSHE deságua diretamente no arroio Pelotas, localizado no sul do estado do Rio Grande do Sul, com área total de aproximadamente 940 $km^2$, incluindo os municípios de Pelotas, Morro Redondo, Arroio do Padre e Canguçu.

Uma malha amostral de 106 pontos de amostragem espaçados 50 m no oeste por 75 m na direção sul foi estabelecida no início do trabalho de campo. Posteriormente, a fim de melhor modelar a estrutura de variabilidade espacial dos atributos de hidrologia do solo, 78 amostras de solo foram coletadas em uma área específica da microbacia do ECW espaçadas 25 m em ambas as direções, totalizando 184 pontos de amostragem. Esta região específica da bacia foi selecionada devido às suas diversas características, incluindo diferentes usos do solo, classes de textura do solo e características do perfil topográfico. A localização geográfica de cada ponto na área do ECW foi registrada usando equipamento de navegação GPS. O software ArcGIS foi usado para estabelecer a malha amostral e obter as coordenadas UTM de cada ponto (Environmental Systems Research Institute, Redlands, CA).

Amostras de solo com a estrutura preservada foram coletadas em todos os 184 pontos de amostragem da camada de 0,20m usando cilindros metálicos com altura de 5,0 cm e diâmetro interno de 4,8 cm. É importante destacar que as amostras foram coletadas da camada superficial do solo (0–20 cm), a área da bacia hidrográfica mais afetada por diferentes sistemas de uso do solo (ALVAREZ; STEINBACH 2009). Os seguintes atributos hidrologicos do solo foram determinados: densidade do solo (BD) (BLAKE; HARTGE 1986), porosidade total do solo (TP), macroporosidade (Mac) e microporosidade (Mic) (KLUTE 1986). A condutividade hidráulica saturada do solo ($K_{Sat}$) foi medida em cada ponto usando o método de carga constante (KLUTE; DIRKSEN 1986).

Os principais usos da terra identificados na BHSH foram floresta, silvicultura, cultivo anual e pastagem. Floresta aqui inclui sítios de floresta

nativa que representam cerca de 10% da área da BHSH; as áreas de silvicultura foram caracterizadas pela ocorrência de Pinus sp., Eucalyptus sp. e Acacia sp., equivalentes a 12% do BHSH. As áreas de cultivo anuais incluíram o cultivo de Glycine max, Zea mays e Nicotiana tabacum que ocuparam a maior área da BHSH (71%). As pastagens caracterizam-se principalmente por áreas de campo nativo ou gramíneas plantadas, representando 7% da área total da BHSH.

O presente trabalho esta articulado em dois artigos, onde foram aplicadas técnicas pedoestatísticas para um mesmo banco de dados com os atributos hidrológicos do solo e do ambiente acima mencionado. Para a realização dos artigos foram aplicadas metodologias estatísticas como as medidas de posição, dispersão e frequência, análise de correlação linear, estatística multivariada, análises espaciais, explorando ferramentas da geoestatística e da simulação de dados.

A figura 1 demonstra a organização e a ligação entre os dois artigos. No artigo 1 o banco de dados coletados na BHSE passa por um redimensionamento por PCA gerando novos grupos de dados. Esses dados são dispostos como variáveis regionalizadas e posteriormente são interpolados por krigagem ordinária e comparados a variáveis não numéricas por meio de validação cruzada. O artigo dois usa o mesmo banco de dados mais os grupos de dados gerados pela PCA para simular a $K_{Sat}$ na área da bacia. Os atributos hidrológicos do solo e ambientais mais os PC's gerados foram utilizados como variáveis secundárias para o processo de co-simulação sequencial Gaussiana. Os mapas são simulados e também as incertezas locais e globais são calculadas.

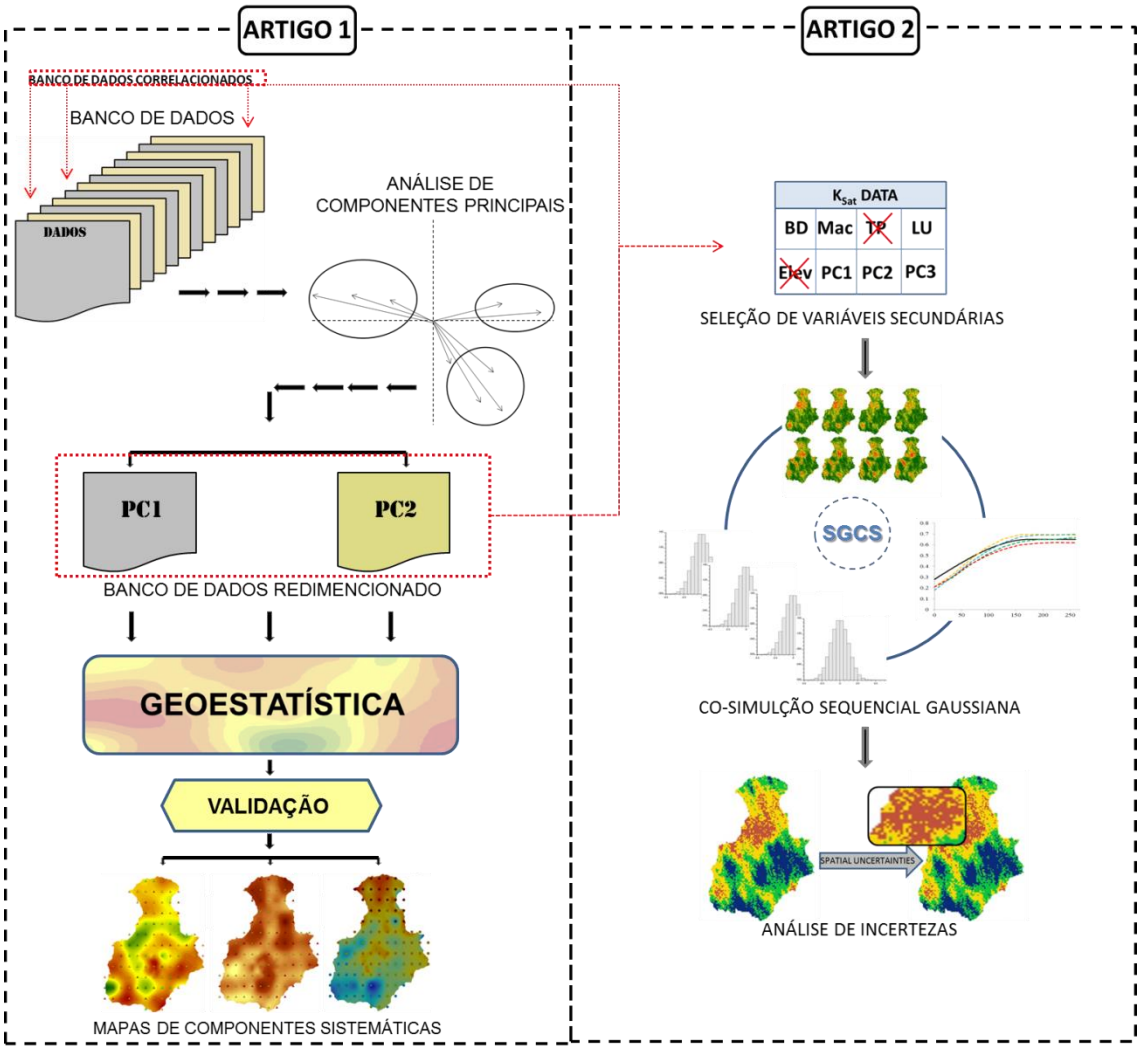


Figura 1. Representação gráfica da organização da pesquisa indicando a ligação entre os artigos 1 e 2 que compõem a tese.

## Artigo 1

# Examining the spatial relationship between saturated soil hydraulic conductivity and co-driven factors at watershed scale using multivariate and geostatistical methods

**Abstract:** The saturated soil hydraulic conductivity ($K_{Sat}$) presents great difficulties for mathematical modeling, mainly on a watershed scale, where the heterogeneity over several landscape characteristics controls this phenomenon. Many of these factors are relatively difficult to measure and quantify but can be stratified in non-numerical data. The study was conducted in the Ellert Creek watershed (ECW) of around 0.7 $km^2$, with mixed agricultural land use and native forest, with at least four different soil textural classes and strongly undulated relief (310 to 420 meters above sea level). A database with soil hydrological and topographic attributes (soil bulk density, total porosity, macroporosity, microporosity, organic carbon, elevation, slope) sampled in a grid of 178 points exactly at the same coordinates where $K_{Sat}$ was sampling, were reduced by Principal Component Analysis (PCA). The components generated by the PCA analysis were mapped by geostatistical tools and later compared with land use, textural classes and topographic profile. PCA resulted in three components () that together explain more than 75% of $K_{Sat}$ variance in the ECW watershed. None of the three components generated explains less than 15% of variance. The geostatistical process of the generated components was composed by experimental semivariograms constructed by the Matheron estimator for the PCA1 and PCA2 and Cressie and Hawkins estimator for the PCA3. The non-numerical factors plotted on the maps of the PCA's showed similarity patterns proven by the coefficient of agreement, and were classified as moderate agreement for the maps of PCA1 and PCA2 and substantial for PCA3. The results suggest that the reduction of the database based by principal components analysis is a suitable alternative for a better understanding of the $K_{Sat}$, mainly in a watershed scale, where the factors generated by the PCA present great spatial dependence, becoming a key input data for geostatistical modeling.

**Key words:** Principal component, kriging, land use, soil textural class, topography.

## 1. Introduction

Saturated soil hydraulic conductivity ($K_{Sat}$) is a soil hydrological attribute governed by several other attributes, so that it can be called an integrating, although random, attribute (Leij et al., 2004; Gupta et al., 2021). Many of these attributes that control $K_{Sat}$ are relatively difficult to be measured and quantified, causing problems in modeling procedures when they are needed. Land use has been a difficult variable to be expressed numerically for many soil scientists, although being a key information to understand $K_{Sat}$ variability in the field (Centeno et al., 2020a). Although soil particles are not so difficult to be quantified, it is necessary to perceive that clay, silt and sand act together in the soil system, and perhaps that the textural class vector is a more efficient representation of this set, however complex to be quantified too. To overcome this, the grouping of these variables can facilitate the task of understanding the relationship between the topographic profile and $K_{Sat}$. Therefore, if we have an integrating variable ($K_{Sat}$), it is essential to know the systematic components to better understand the phenomenon of $K_{Sat}$, even if these are stationary variables.

Urso et al. (2019) defined that a systematic component is an explanatory or independent variable ($x_1,\ldots,x_n$), which describe each instance $X_i$ of the data set, where i=1,…,n is the total number of instances in the data set. Values of the explanatory variables are treated as fixed and not as random/integrator variables. Many studies demonstrate that $K_{Sat}$ behaves like an integrating random variable. Mainly in the superficial layers of the soil, $K_{Sat}$ gradually increases from degraded soils by intensive use, such as pasture and annual crops, to soils under forest, native or implanted conditions (Archer et al., 2013; Shi et al., 2015; Becker et al., 2018). Forest soils are associated with higher rates of water infiltration (Archer et al., 2013; Kurnianto et al., 2018) and lower

generation of runoff (Kurnianto et al., 2018; Liu et al., 2018) than soils under other types of vegetation.

It is well known that most soil hydrological attributes are highly variable in space and time. This can be explained by the intense soil biological activity (Becker et al., 2018), frost processes (Wang et al., 2020), water, air and wind erosion, anthropogenic human activities (Soares et al., 2020), land-use systems (Centeno et al., 2020b), among other driving processes. Some soil hydrological attributes of special interest are those related to water and gas transport into soils, such as soil structure, pore volume, pore size distribution, bulk density, and other associated to the landscape. While the soil texture can be expected not to vary over time, soil structure can be influenced by land use types and soil management practices (Wang et al., 2020; Soares et al., 2020) and soil biological activity (Meurer et al., 2020). So, the soil land use is a key anthropogenic process to understand and quantify the spatial variability of soil hydrological attributes at the watershed scale (Shi et al., 2015; Centeno et al., 2020a).

Statistical tools (descriptive statistics, principal component analysis (PCA) and geostatistics) can potentially be used to assess the effects of land use on soil hydrological attributes at the watershed scale (Centeno et al., 2020a). PCA is a technique that indicates associations between variables, thus reducing the size of the number of data and grouping those with most similarity (Mushina et al., 2020). Geostatistics is a tool that considers the coordinates of the variable in space, which has been used to quantify and map the spatial variability structure of the variable under study as well as to quantify and map the spatial variability structure between two variables.

The present study aimed to examine the influence of soil attributes and the soil land use on the saturated soil hydraulic conductivity, at a watershed scale, by principal components analysis as well assess the spatial distribution of the components generated by geostatistical tools and its associations with non-numerical landscape patterns.

## 2. Material and methods

In this study, the Ellert Creek Watershed (ECW) was chosen due its economic and social importance for the southern region of the State of Rio Grande do Sul, Brazil, whereas its area is predominantly occupied by family farming. The development of the ECW is regionally strategic, as it is a watercourse that flows directly into the Pelotas River, which in turn provides drinking water for a population of almost half a million people (Beskow et al., 2016). Native forest covers about 10% of the watershed area and basically comprises the riparian forest. The other areas are occupied by silviculture, annual crops and pasture, therefore characterizing an intense anthropic and soil management activity in most of the ECW area.

The headwater sub-basin ECW of the Pelotas River Watershed is located in the municipality of Canguçu, Southern Rio Grande do Sul (RS) state, Brazil (Fig. 1). With an area of approximately 0.7 $km^2$, and an altitude varying from 310 to 419 m, it has a Cfa climate type according to the Köppen climate classification, a mesothermal climate indicating wet subtropical conditions, and characterized by an annual average temperature of 18 °C, with hot summers and cold winters (Kuinchtner and Buriol, 2001). The precipitation is well distributed throughout the year, and its mean annual value is around 1350 mm. The regional relief varies from undulating to strong undulating, with a predominance of native forest or sparse shrub, and with shallow soils.

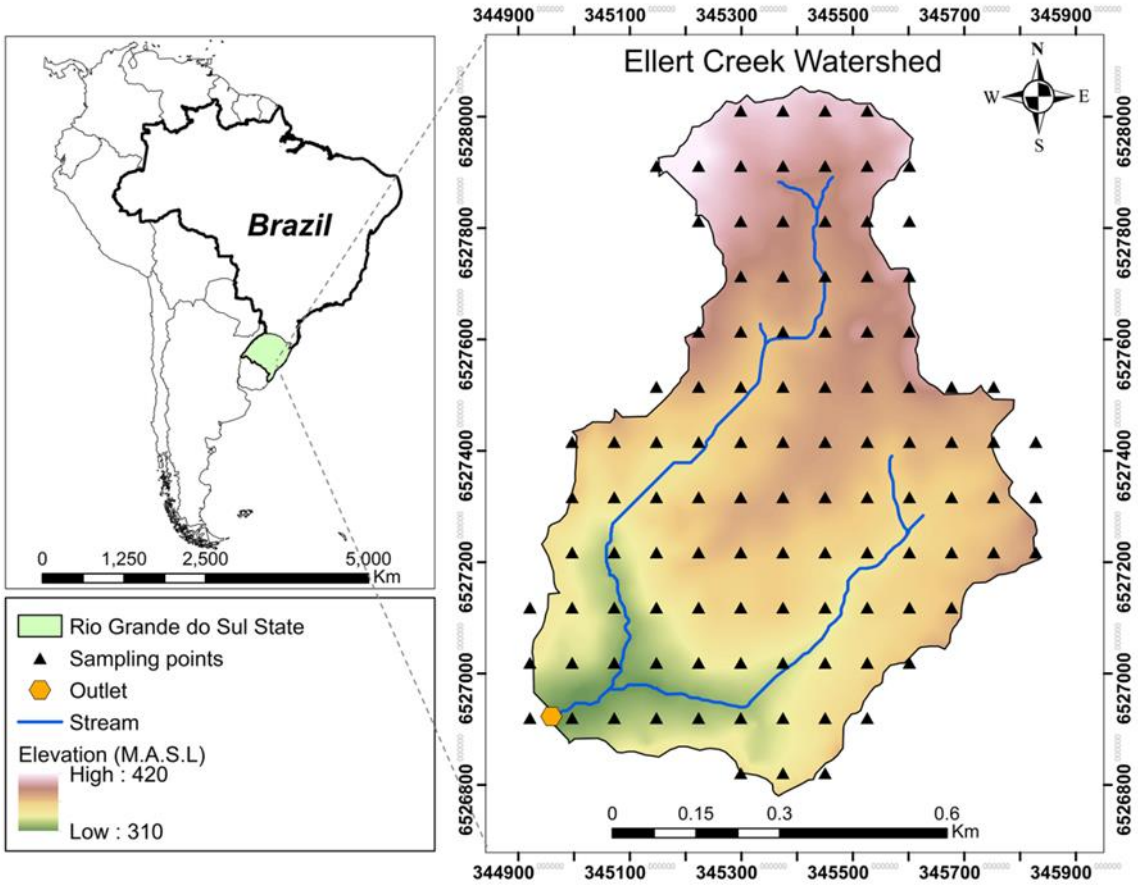


Fig. 1 Location, topography, hydrography, and sampling points of the Ellert Creek Watershed, Rio Grande do Sul, Brazil.

A sampling grid of 106 sampling points spaced 50 m in the west direction by 75 m in the south direction was established at the beginning of field work. Subsequently, in order to better model the spatial variability structure of soil hydrology attributes, 78 soil samples were collected in a specific area of the ECW spaced 25 m at both directions, totaling 184 sampling points (Fig. 1). At each sampling point, soil samples were collected from the top soil layer (0-0.20 m) to evaluate the following attributes: soil bulk density (BD) (Blake and Hartge, 1986); total porosity (TP), soil macroporosity (Mac), and soil microporosity (Mic) (Klute, 1986); soil organic carbon content (OC) (Walkley and *Black, 1934*), and saturated soil hydraulic conductivity ($K_{Sat}$) using the constant method (Klute and Dirksen 1986).

The main land uses were identified during the soil sampling field campaign describing the land use around each soil sampling point. The soil land uses at ECW were classified into forest, silviculture, annual cropping, and

pasture. Forest includes native forest sites which account for about 10% of the ECW's area; areas of silviculture are equivalent to 12% of the ECW watershed area. Annual cropping areas include the many types of crops, for example *Glycine max* and *Zea mays*, and occupy the largest area of the ECW (71%). Pasture areas account for 7% of the total area.

## 2.1 Exploratory data analysis

The mean and the median of stratified data for the four land use types identified in the ECW (forest, silviculture, annual cropping and pasture) were calculated in order to characterize each soil attributes. For the purpose of better visualization of the distribution of data within each land use, the data were arranged in boxplots identifying the mean, the mean and the outliers for each set. Also an exploratory analysis of the obtained eigenvalues from each PCA component was performed, calculating the skewness and octile skewness coefficients of its distribution. The center, spread, skewness, and the presence of outliers of each PCA component distribution were assessed by construction its histogram plot. This second step was used for better decision making in the subsequent geostatistical process.

## 2.2 Multivariate statistical methods

Multivariate statistical methods are useful for modeling and interpreting large datasets, reducing their dimensionality and extracting useful information for soil quality assessment and management (Hair et al., 2009). In the present study, PCA was employed in a dataset composed of 6 soil hydrological attributes. Principal components (PCs) were selected based on the fact that they explain at least 70% of the total variance of each data set and that the minimum explained variance per PC was higher than 20%. This step provided an n-number of components for the PCA with all attributes that were organized as a spatial database and then used in the geostatistical analysis (Libohova et al., 2018).

## 2.3 Geostatistical analysis

Geostatistical analysis included semivariograms to quantify spatial patterns of the regionalized variables and derive key input parameters for kriging spatial interpolation (Lark, 2003). The semivariograms were calculated from the set ($Z_u$ ($x_i$), i = 1, 2,…, $N$), usually by the Matheron classical estimator (Webster and Oliver 2001), defined by Eq. (1):

$$\gamma_{Z_u}^{M}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z_u(x_i) - z_u(x_i + h)]^2 \quad (1)$$

Where $\gamma_{Z_u}^{M}(h)$ is the semivariance value estimated by the Matheron estimator, $z_u(x_i)$ and $z_u(x_i + h)$ are values of $Z_u$ at locations $x_i$ and $x_i + h$, respectively, and $N(h)$ is the number of pairs [$z_u(x_i)$, $z_u(x_i+h)$] separated by the lag distance $h$.

Another used estimator is the robust semivariance estimator developed by Cressie and Hawkins (1980) (Lark, 2003), which deal better with the presence of outliers in the datasets. This estimator is presented in Eq. (2):

$$\gamma_{Z_u}^{CH}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \frac{\left\{\left[|z_u(x_i) - z_u(x_i + h)|^{\frac{1}{2}}\right]^4\right\}}{0.457 + \frac{0.494}{N(h)}} \quad (2)$$

where $\gamma_{Z_u}^{CH}(h)$ is the semivariance value calculated by the Cressie-Hawkins estimator. This estimator was designed to calculate the semivariogram of a primary process with differences normally distributed even in the presence of contaminants of a secondary process (Webster and Oliver 2001).

According to Lark and Bishop (2007), it is essential to check the histograms and the skewness coefficient of the data before running geostatistical analyses. Therefore, the primary parameter to select a

semivariogram (classical or robust) is the octile skewness coefficient (Eq. 3). The octile skewness uses more information from the tails of the distribution and is often more appropriate for detecting the symmetry and asymmetry in the data (Brys et al., 2004). The octile skewness coefficient (OSC) is defined as:

$$OSC = \frac{[(P_{0.875} - P_{0.5}) - (P_{0.5} - P_{0.125})]}{(P_{0.875} - P_{0.125})} \quad (3)$$

where $P_q$ is the value of the ordered datum such that the proportion $q$ of the data are smaller than $P_q$ (Brys et al., 2004). For the choice of a semivariogram estimator, if the conventional skewness coefficient is >1, but the octile coefficient is <0.2, there is a great possibility to see in the histogram a normal distribution, but with some outliers. In this case the robust estimator is likely to be suitable. If both skewness coefficients are large, and the histogram looks to have a positive skew, then some data transformation is likely to be necessary, and then, the same skewness analyses must be checked. If the skewness coefficients reach <1 and <0.2, respectively, for conventional and OSC coefficients, the Matheron classical estimator can be applied.

For geostatistical interpolation of the PCA's scores, the selected method was the ordinary kriging (OK). The OK is one of the most commonly used kriging techniques (Goovaerts, 1997). The spatial prediction of the unmeasured point $x_o$ is given by predicting the value $z^*(x_0)$, which equals the line sum of the known measured values. Isaaks and Srivastava (1989), provide a description of OK with the following equation:

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

Where $z^*(x_o)$ is the predicted value at the unmeasured position $x_o$, $z(x_i)$ are the measured values of $Z_u$ at positions $x_i$, $\lambda_i$ are the weighting coefficients and $n$ is the number of positions within the searched neighborhood. For each PC experimental semivariogram (Eq. 1 or 2), a mathematical model (theoretical semivariogram) was adjusted to obtain the parameters $C_0$ (nugget effect), C

(contribution) and A (spatial range). In this study, the OK method was performed using the R software with a framework introduced by Gräler et al. (2016).

## 2.4 Cross validation

To validate the comparison between non-numeric data of land use, textural class and topographic profile with kriged maps of the main PCs, the Kappa coefficient was used. The Kappa coefficient of agreement is used to describe the agreement between two or more qualitative attributes when they perform a nominal or ordinal evaluation of the same sample. The Kappa coefficient of agreement, suggested by Cohen in 1960, is calculated using the equation below:

$$\widehat{K} = \frac{\hat{p}_o - \hat{p}_e}{1 - \hat{p}_e} \quad (5)$$

where $\hat{p}_o$ is the proportion of cases correctly classified (i.e. overall accuracy) and $\hat{p}_e$ is the expected proportion of cases correctly classified by chance.

The value of the Kappa coefficient of agreement can vary from [-$\hat{p}_e$ / 1-$\hat{p}_e$] to 1. The closer to 1 its value, the greater the indication that there is an agreement between the judges and the closer to zero, greater is the indication that the agreement is purely random. Landis and Koch (1977) suggest a classification to interpret the Kappa value where less than zero is insignificant; between 0 and 0.2 is weak; between 0.21 and 0.4 is fair; between 0.41 and 0.6 is moderate; between 0.61 and 0.8 is substantial and between 0.81 and 1 is almost perfect.

The statistical parameters used to evaluate the performance of the ordinary kriging interpolation were the mean error (ME) (Eq. 6) and the root of the mean squared error (RMSE) (Eq. 7).

$$ME = \frac{1}{n}\sum_{i=1}^{n}(e_i - m_i) \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - m_i)^2} \quad (7)$$

where: $n$ = number of sampling points; $e_i$ = estimated value of; and $m_i$ = observed value. The ME is an indicator of the accuracy of the estimation, showing the tendency of the interpolator to overestimate the values if positive or to underestimate if negative and the RMSE quantifies the dispersion of the measured and estimated values around the 1:1 line.

## 3 Results and discussion

### 3.1 Exploratory data analysis

A decreasing trend for $K_{Sat}$ at the land uses forest, silviculture, annual cropping and pasture (Figure 2). The boxplots of Mac and TP follow the same $K_{Sat}$'s trend, while the BD boxplots presents an opposite behavior, increasing from forest to silviculture, annual crops and pasture. The Mic and OC boxplots show a short variation within each land use, except for the forest OC, which is much higher than the others.

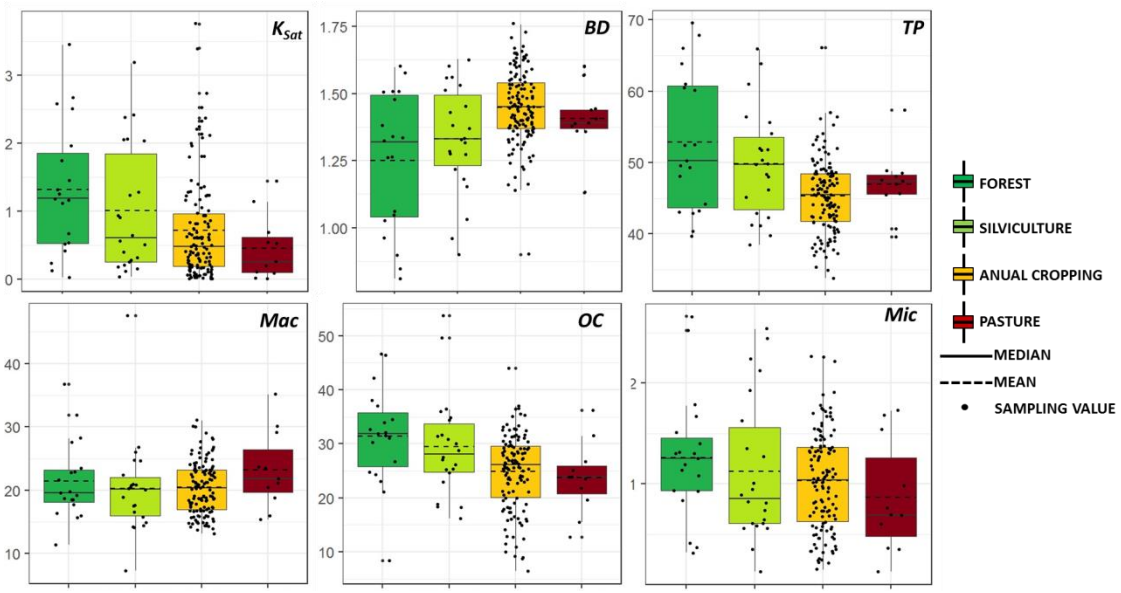


Fig. 2. Boxplots for soil attributes by type of land use. Attributtes; $K_{Sat}$: Saturated soil hydraulic conductivity (m $h^{-1}$); BD: Bulk density (g $cm^{-3}$); TP: Total porosity (%); Mic: Microporosity (%); Mac: Macroporosity (%); OC: Soil Organic Carbon (%).

Native forests provide higher mean values of $K_{sat}$ (1.32 m.$h^{-1}$) and Mac (32%) and lower values of BD (1.25 g.$cm^{-3}$) when compared to the other land use systems (Silviculture: $K_{sat}$ = 1.01 m.$h^{-1}$, Mac = 29%, and BD = 1.33 g.$cm^{-3}$; Annual cropping: $K_{sat}$ = 0.72 m.$h^{-1}$, Mac = 25%, and BD = 1.45 g.$cm^{-3}$; and Pasture: $K_{sat}$ = 0.46 m.$h^{-1}$, Mac = 24%, and BD = 1.41 g.$cm^{-3}$). On average, native forest soils had $K_{sat}$ values approximately 2.4 and 4.5 times greater than those found under annual cropping and pasture soils in our study, respectively. Cultivated soils under native vegetation (e.g., native forest) when compared with soils impacted by agricultural land-use (e.g., annual cropping) generally presented lower BD and higher $K_{sat}$ and Mac as a result of ample organic inputs and abundant burrowing fauna (Jarvis et al., 2013).

ECW areas where there is a predominance of the annual cropping land use (mainly in tobacco cultivated areas) in which commonly adequate soil conservation practices by farmers have been not adopted. The lack of soil conservation practices associated to the tobacco cultivation in areas with high environmental fragility have increased the soil susceptibility to degradation, decreasing its quality and increasing soil erosion risks, mainly because the

intensive use of the plow layer (0-0.20 m soil layer) by farmers. In the agricultural areas of the watershed, different farming systems with differing levels of mechanization, and technology such as conventional tillage, no-tillage and minimum tillage were observed, increasing the variability of $K_{sat}$. Several studies have shown that land use systems affect $K_{sat}$ values and therefore their spatial variation in a watershed (Hao et al., 2019; Li et al., 2019; Centeno et al., 2020a).

### 3.2 Multivariate analysis

Table 1 showed that the PCA performed for the set of six variables, where the first two PCA's explained almost 80% of the total variance. The first PC factor summarized the strong relationship among $K_{Sat}$, BD, TP and Mac, which showed the highest loadings. On the second PCA, Mic and OC were the highly and significantly weighted variables.

An overview on the variables and scores showed the inverse relationship between $K_{Sat}$ and BD and a directly relation with TP and Mac. This result is due to the preferential flow in the soil that controls the saturated soil hydraulic conductivity and occur through the pores of the soil, mainly the macropores, by the gravitational action. Soils with high bulk density have a reduced number of large pores, which prevents the passage of water, significantly reducing the movement of water in the soil.

Table 1. Principal component of $K_{Sat}$ with eigenvalue, individual variance and cumulative variance.

| | PC1 | PC2 |
|---|---|---|
| $K_{Sat}$ | 0.72 | 0.33 |
| BD | -0.92 | 0.32 |
| TP | 0.92 | -0.32 |
| Mic | -0.24 | -0.91 |
| Mac | 0.92 | 0.32 |
| OC | 0.27 | -0.58 |
| Eigenvalue | 3.18 | 1.59 |
| Variance | 53.00 | 26.52 |
| Cumulative | 53.00 | 79.53 |

$K_{Sat}$: Saturated soil hydraulic conductivity (m $h^{-1}$); BD: Bulk density (g $cm^{-3}$); TP: Total porosity (%); Mic: Microporosity (%); Mac: Macroporosity (%); OC: Soil Organic Carbon (%); Elev: Elevation (m); PC1: first principal component; PC2: second principal component

In the PCA performed of six soil hydrological attributes, a high association of BD, TP and Mac was observed (loading factor of ±0.92) for the first component, however, it explained 53% of the total variance. In the second PC, Mic showed a high inverse relationship with $K_{Sat}$ (loading factor of -0.85) following by the OC (-0.58) and it explained 26.5% of the total variance. This result may be due to the fact that soil microporosity is a variable that depends on several other soil attributes, mainly linked to the textural class (Pachepsky and Park, 2015) and even though it is the most influential variable in PCA2 followed by OC (loading factor of -0.61).

The results of PCA were also summarized in a biplot (Fig. 3) and show the rotations of the analyses divided into two major soil water dynamic trend. The first trend is indicated by Mac, BD and TP and likely reflects either soil land use due to these variables being quite close to the changes promoted by soil management in a watershed (Centeno et al., 2020a). A second trend includes the elements thought to be related to soil textural class, which here fall together because soil microporosity and organic carbon be active variables on different soil classes.

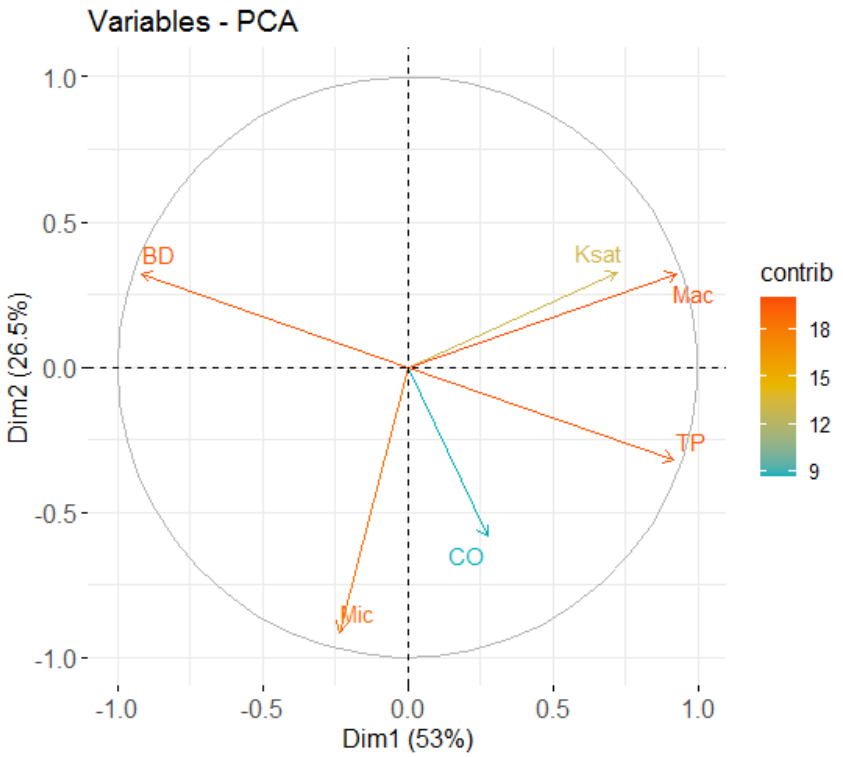


Fig. 3. Biplot of the $K_{Sat}$ components in the analysis carried out on the whole dataset.

Especially on the superficial soil layers, attributes associated to soil porosity and bulk density are quite susceptible to changes each year due to the movement of the soil carried out by agricultural management (Fei et al., 2019). In cultivation areas where plowing is frequent and machinery is used for planting practices or where livestock promotes the movement of animals in the field, affecting the movement of water in the soil. In natural forest and silviculture areas, where the movement of machines and animals is reduced, the opposite phenomenon is observed (Libohova et al., 2018). The PC2 is linked to soil textural class what is few or practically without temporal variation (Baskan et al., 2013).

### 3.3 Geostatistics analysis

The first parts of the application of geostatistics make up an exploratory analysis of the data to be analyzed. As the data were generated in a primary multivariate process of this same work (PCA), this exploratory analysis was carried out *a priori* from the subsequent geostatistical process.

The data distribution of the first PC (PC1) had a higher concentration of values in the center of the histogram with a slight positive skewness and the absence or very low number of outliers, i.e., its distribution was visually very close to normal. The data distribution of the second PC (PC2) showed a slightly more prominent negative asymmetry and the possible presence of some outliers. Some broken bins with ‘‘0’’ frequencies at the histograms of PC2 starting from the value of 2.5 and -3.8, respectively, which imply on the existence of global outliers for this data set.

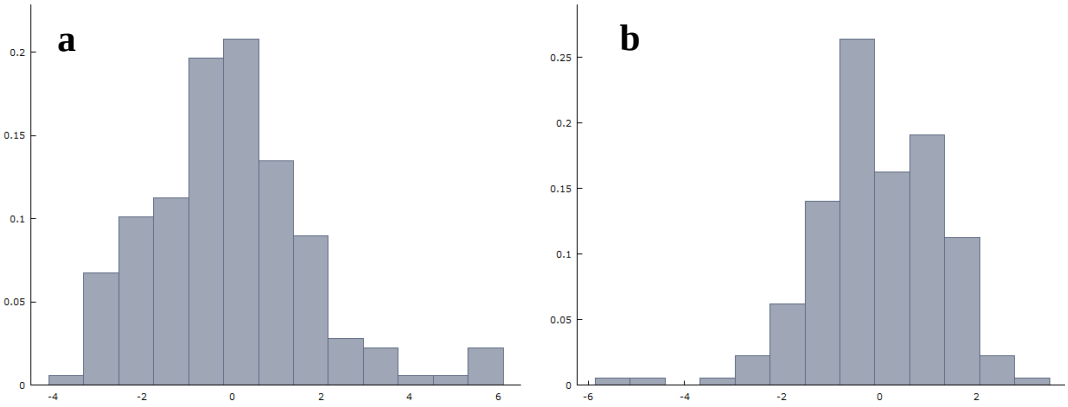


Fig. 4. Histograms of frequency distribution of PC1(a) and PC2(b).

Table 2 shows the results of the conventional skewness, the octile skewness, and the normality test which were used to select the semivariogram estimator for studying the spatial distribution of each PC. The PC2 data distribution showed conventional skewness out of the range of [-1, 1], but its octile skewness coefficient fell inside of the range of [−0.2, 0.2], which indicated that the presence of outliers affected the distribution of PC2 data. However, the outliers may not be unambiguously wrong and may represent an important phenomenon to mapping and data transformation is not suitable. Thus, the

robust semivariogram estimator can be used to calculate the semivariance values to drawback the effects of outliers in the data set.

Table 2 – Choice parameters for semivariograms estimator of PCA's

| Component | Skew. | Octile Skewness | K-S | Normal | Estimator |
|---|---|---|---|---|---|
| PC1 | 0.62 | 0.01 | 0.02 | No | Matheron |
| PC2 | -1.01 | 0.05 | 0.22 | Yes | Creesie and Hawkins |

PC1: eigenvalues for the first principal component; PC2: eigenvalues for the second principal component; K-S: P-valor for Kolmogorov-Smirnov

In this study, isotropic experimental semivariograms were estimated, assuming an identical spatial correlation in all directions and neglecting the influence of anisotropy on the semivariogram parameters. Isotropy is a feature of a natural process in which directional influence is considered insignificant and spatial dependence (autocorrelation) changes only with the distance between lags (Isaaks and Srivastava, 1989).

The spherical model best described the spatial variability structure for both PC's data sets (Table 3 and Fig. 5). The ranges of the theoretical semivariograms were 213 m for PC1 data, and 205 m for PC2.

Table 3 - Parameters of semivariograms of PCA's adjusted by theoretical models.

| Semivariogram | Model | $C_0$ | C | $S^2$ | $a$ (m) |
|---|---|---|---|---|---|
| $\hat{\gamma}(h)$PC1 | Spherical | 1.59 | 1.72 | 3.18 | 213 |
| $\hat{\gamma}(h)$PC1 | Spherical | 0.36 | 1.24 | 1.62 | 205 |

$\hat{\gamma}(h)$PC1: Semivariogram for first principal component; $\hat{\gamma}(h)$PC2: Semivariogram for second principal component $C_0$: nugget effect; $C_0$+$C$: sill; S2: variance; $a$: range.

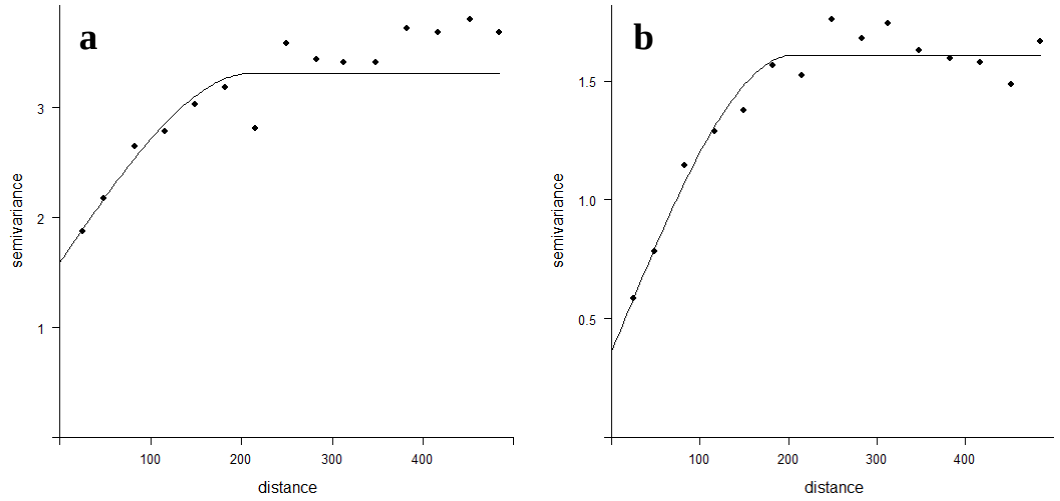


Fig. 5. Semivariograms standardized of PC1(a) and PC2(b).

For different components, the variograms generated by the classical and robust estimator and the fitted models are showing a high spatial dependence for the PC1 and PC2. This result indicates that the PC2 probably would have its variances overestimated by the standard estimator because of the effects of outliers in the variograms.

### 3.3 First principal component – Soil land use

Several studies related that the saturated soil hydraulic conductivity ($K_{Sat}$) presents higher values for forest compared to grassland and cropland (Bormann and Klaassen, 2008; Libohova et al., 2018, Godoy et al., 2018; Hao et al., 2019; Centeno et al., 2020a). With increasing bulk density, the hydraulic conductivity decreases, significantly. While in the forest soil the activity of the soil fauna is much higher (Fei et al., 2019), on the cropland, deep ploughing can lead to an increased hydraulic conductivity in the ploughed soil layer. On the grassland (pasture in this case), livestock movement compacts the soil, which can create denitrification with attendant loss of soil fauna and soil aeration, in a mild reducing condition. As a result of these processes is, a lower $K_{Sat}$ is observed in these layers.

Libohova et al. (2018) found that the pore size distribution established by soil tillage is partly degraded by intense rainfalls. Only the growth of plant roots

can reestablish a continuous pore system. Similarly, soil compaction by soil tillage must not be constant from year to year. Beside alternative soil tillage, climatic factors and soil biological activity can accelerate bulking of soils (Liu et al., 2018).

In the Fig. 6, the PC 1 krigged map is showing a pattern that follows different soil land uses in the ECW. Plotting the field observations data, overlaying the PC 1 krigged maps, it is observed that the pattern clearly follows the boxplot, showing a stratification of forest, silviculture, annual cropping and pasture. The first two land use types can be associated to high values of $K_{Sat}$ and the other two are characterized by low $K_{Sat}$ values. The Kappa coefficient for the kriged map of PC 1 with the data observed in the field, exactly at the sample points was 0.55. According to Landis and Koch's classification, crossing of krigged and observed data shows a moderate agreement.

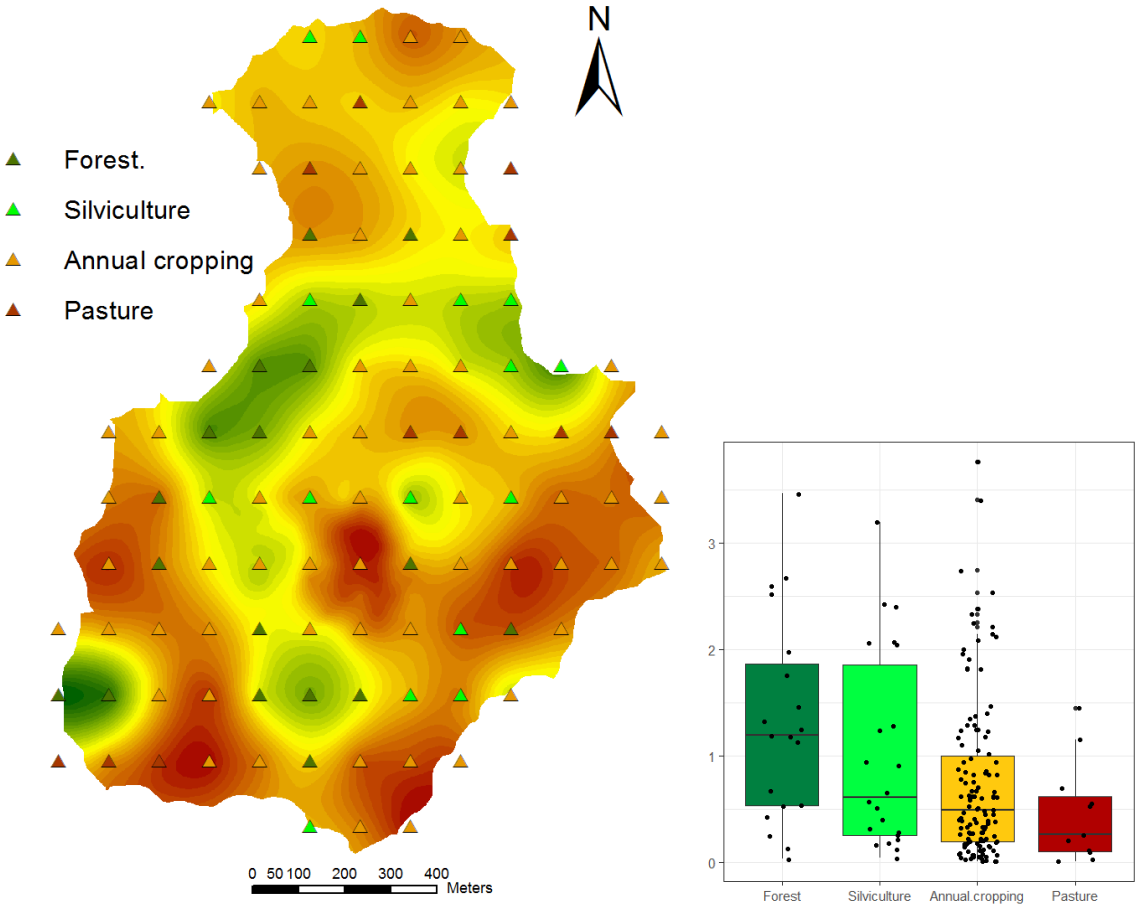


Fig. 6. (a) Krigged map of the first principal component and (b) boxplot of $K_{Sat}$ stratified on different soil land uses.

An increasing bulk density is detected from forest to grassland to cropland, which is in line with the findings of literature (Becker et al., 2018; Kurnianto et al., 2018). It is very widespread in the literature that forest soils are associated with higher rates of water infiltration (Archer et al., 2013; Gupta et al., 2021) and lower surface runoff generation (Germer et al., 2010; Wang et al., 2020), a reflection of the recovery of soil properties in the absence of grazing.

Marshall et al. (2014) performed a grazing exclusion experiment to investigate vegetation effects and found that infiltration rates in plots planted with broadleaf trees were significantly higher than in plots without trees. Chandler and Chappell (2008) also observed higher $K_{Sat}$ around isolated forest compared with surrounding parkland. However, all these studies corroborate the statement that land use would be the main systematic component of the spatial distribution of $K_{Sat}$, confirmed by the analysis of the main components of this research.

### 3.4 Second principal component – Textural class

The dependence of $K_{Sat}$ on soil texture has been well documented (Fei et al., 2019, Liu et al., 2018). Different parameterizations of particle size were suggested to relate $K_{Sat}$ and soil texture. The heterogeneity of particle size distributions appears to be an important factor affecting hydraulic parameters of soils, including the saturated soil hydraulic conductivity. Values of $K_{Sat}$ depend on both distribution of soil particle size (textural class) and the spatial arrangement of these particles (soil structure) (García-Gutiérrez et al., 2017). Soil structure can be, to some extent controlled by soil texture, since packing of particles is affected by particle size distributions (Gupta and Larson, 1979; Ajayi et al., 2016; Fei et al., 2019).

Figure 7a shows the PC 2 krigged map with the points of the four different textural classes plotted at each ECW sample point. The map coincides with a clay textural pattern decreasing to a sandier one in a dark brown to light brown color spectrum. The textural triangle (Fig. 7b) was divided into two colors, demonstrating the soil classes that occur in the basin and separating into a

more clayey pattern (dark brown) and a less clayey pattern (light brown). In this way, the points plotted on the PC 2 map demonstrate more clearly the systematic potential of the texture in relation to the spatial distribution of $K_{Sat}$. The Kappa coefficient for the krigged map of PC 2 with the data observed in the field exactly at the sample points was 0.51. According to Landis and Koch's classification, the crossing of krigged and observed data shows a moderate agreement.

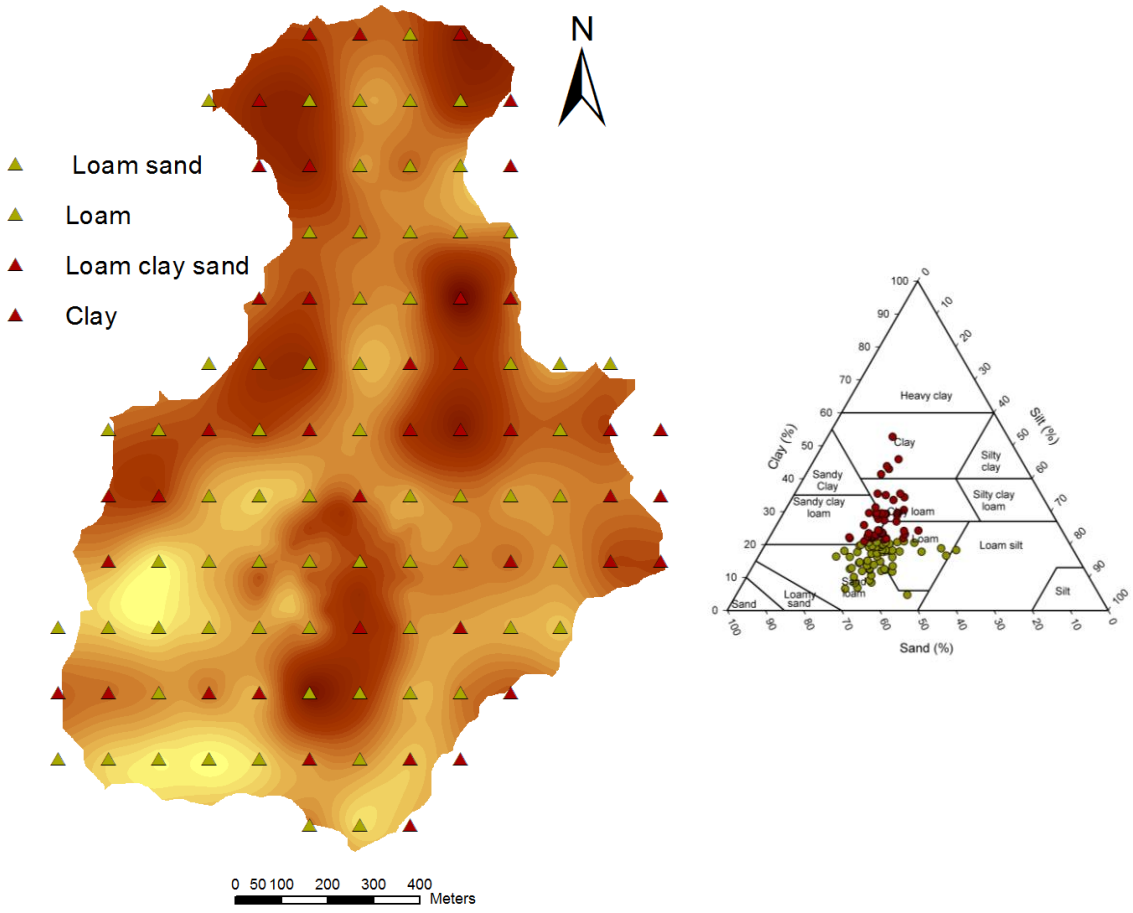


Fig. 7. (a) Krigged map of the second principal component and (b) and textural triangle of the soils of ECW

García-Gutiérrez et al. (2017) concluded that textural heterogeneity is an important factor of $K_{Sat}$, but it acts along many other ecological factors, as animal activity, root exudates, soil aggregation, etc. Alvarez et al. (2019), evaluating the surface $K_{Sat}$ values in relation to soil texture, observed that greater $K_{Sat}$ variation was found in soils with higher clay content, contrasting with lower variation in soils with higher sand content. Soils with higher sand

content also often did not show the higher $K_{Sat}$. The sandy loam texture highlighted the large differences between $K_{Sat}$ in different soil land uses. It clearly indicates that $K_{sat}$ and soil texture relationship is satisfactory in some textural classes, but it is not so clear when we analyze the $K_{Sat}$ relationship with land use because other textural classes could appear predominating in a same land use with fine particles, in which other processes as aggregation or weathering cannot be elucidated by the single textural data input. However, all these questions show that textural class is a systematic component of the $K_{Sat}$ spatial distribution, with less relevance than land use, mimetizing a second principal component.

### 3.5 Cross Validation

For the mapping process by ordinary kriging of de components generated by the PCA, the RMSE values, which indicate the data dispersion behavior, presented better performance for interpolation of PC2 data set (Table 5). The values of ME, which identify if the model underestimated (negative values) or overestimated (positive values) was quite satisfactory for all maps, presenting low and negative values, with a slight underestimating behavior.

Table 4. Summarized statistics of the cross-validation for observed vs. interpolated data

| Map | RMSE | ME |
|---|---|---|
| PC1 | 1.69 | -0.03 |
| PC2 | 1.28 | -0.01 |

PC: Principal component maps (1 and 2); RMSE: root of the mean squared error; ME: Mean error

## Conclusions

Principal component analysis had a successfully performance for soil hydrological attributes of the watershed. The first component was composed by soil hydrologic attributes that are sensitive due to soil land use mainly by agricultural management impact, like soil density, macroporosity and total

porosity, contributed with the highest value on the cumulative variance. The second principal component generated rated by the accumulated variance for the description of the saturated soil hydraulic conductivity was that soil hydrological attributes with a strong relationship with the soil texture: soil microporosity and soil organic carbon.

The geostatistical process for the first two components generated by the analysis of principal components was very efficient, allowing building maps by ordinary kriging with high accuracy. The relation of the maps of the principal components with the non-numeric attributes demonstrates a very close relationship with the data of land use and textural class, allowing us to confidently say that these are the most relevant systematic components for the saturated soil hydraulic conductivity on the Ellert Creek Watershed.

## Acknowledgments

The authors wish to thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the financial support and scholarships provided, and the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES), Finance Code 001, for scholarships.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**References:**

AJAYI, A.e.; HOLTHUSEN, D.; HORN, R.. Changes in microstructural behaviour and hydraulic functions of biochar amended soils. **Soil And Tillage Research**, [S.L.], v. 155, p. 166-175, jan. 2016. Elsevier BV. http://dx.doi.org/10.1016/j.still.2015.08.007

ALVAREZ, Edwin Esteban Reyes; PARRA-GONZÁLEZ, Sergio David; RODRIGUEZ-VALENZUELA, Jeisson. 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

ANDERSON, M.G.; KNEALE, P.e.. Topography and hillslope soil water relationships in a catchment of low relief. **Journal Of Hydrology**, [S.L.], v. 47, n. 1-2, p. 115-128, maio 1980. Elsevier BV. http://dx.doi.org/10.1016/0022-1694(80)90051-7

ARCHER, N.A.L.; BONELL, M.; COLES, N.; MACDONALD, A.M.; AUTON, C.A.; STEVENSON, R.. Soil characteristics and landcover relationships on soil hydraulic conductivity at a hillslope scale: a view towards local flood management. **Journal Of Hydrology**, [S.L.], v. 497, p. 208-222, ago. 2013. Elsevier BV. http://dx.doi.org/10.1016/j.jhydrol.2013.05.043

BASKAN, Oguz; KOSKER, Yakup; ERPUL, Gunay. Spatial and temporal variation of moisture content in the soil profiles of two different agricultural fields of semi-arid region. **Environmental Monitoring And Assessment**, [S.L.], v. 185, n. 12, p. 10441-10458, 31 jul. 2013. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s10661-013-3343-8

BECKER, R.; GEBREMICHAEL, M.; MÄRKER, M.. Impact of soil surface and subsurface properties on soil saturated hydraulic conductivity in the semi-arid Walnut Gulch Experimental Watershed, Arizona, USA. **Geoderma**, [S.L.], v. 322, p. 112-120, jul. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2018.02.023

BESKOW, Samuel; TIMM, Luis Carlos; TAVARES, Vitor Emanuel Quevedo; CALDEIRA, Tamara Leitzke; AQUINO, Leandro Sanzi. Potential of the LASH model for water resources management in data-scarce basins: a case study of the fragata river basin, southern brazil. **Hydrological Sciences Journal**, [S.L.], v. 61, n. 14, p. 2567-2578, 15 jul. 2016. Informa UK Limited. http://dx.doi.org/10.1080/02626667.2015.1133912

BITENCOURT, Dioni Glei Bonini; BARROS, Willian Silva; TIMM, Luís Carlos; SHE, Dongli; PENNING, Letiane Helwig; PARFITT, José Maria Barbat; REICHARDT, Klaus. 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

Blake, G.R., Hartge, K.H.. **Bulk density**. Em: A. Klute (ed) *Methods of soil analysis*. (pp 363–382). Agronomy Monograph, ASA-SSSA, Madison, 1986

BORMANN, H.; KLAASSEN, K.. Seasonal and land use dependent variability of soil hydraulic and soil hydrological properties of two Northern German soils. **Geoderma**, [S.L.], v. 145, n. 3-4, p. 295-302, jun. 2008. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2008.03.017

BRYS, G; HUBERT, M; A STRUYF,. A Robust Measure of Skewness. **Journal Of Computational And Graphical Statistics**, [S.L.], v. 13, n. 4, p. 996-1017, dez. 2004. Informa UK Limited. http://dx.doi.org/10.1198/106186004x12632

CENTENO, Luana Nunes; TIMM, Luís Carlos; REICHARDT, Klaus; BESKOW, Samuel; CALDEIRA, Tamara Leitzke; OLIVEIRA, Luciana Montebello de; WENDROTH, Ole. Identifying regionalized co-variate driving factors to assess spatial distributions of saturated soil hydraulic conductivity using multivariate and state-space analyses. **Catena**, [S.L.], v. 191, p. 104583, ago. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2020.104583

CENTENO, Luana Nunes; HU, Wei; TIMM, Luís Carlos; SHE, Dongli; FERREIRA, Arlan da Silva; BARROS, Willian Silva; BESKOW, Samuel; CALDEIRA, Tamara Leitzke. Dominant Control of Macroporosity on Saturated Soil Hydraulic Conductivity at Multiple Scales and Locations Revealed by Wavelet Analyses. **Journal Of Soil Science And Plant Nutrition**, [S.L.], v. 20, n. 4, p. 1686-1702, 6 maio 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s42729-020-00239-5

CHANDLER, Kathy R.; CHAPPELL, Nick A.. Influence of individual oak (Quercus robur) trees on saturated hydraulic conductivity. **Forest Ecology And Management**, [S.L.], v. 256, n. 5, p. 1222-1229, ago. 2008. Elsevier BV. http://dx.doi.org/10.1016/j.foreco.2008.06.033

CRESSIE, Noel; HAWKINS, Douglas M.. Robust estimation of the variogram: i. **Journal Of The International Association For Mathematical Geology**, [S.L.], v. 12, n. 2, p. 115-125, abr. 1980. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/bf01035243

FEI, Yuanhang; SHE, Dongli; GAO, Lei; XIN, Pei. Micro-CT assessment on the soil structure and hydraulic characteristics of saline/sodic soils subjected to short-term amendment. **Soil And Tillage Research**, [S.L.], v. 193, p. 59-70, out. 2019. Elsevier BV. http://dx.doi.org/10.1016/j.still.2019.05.024

GARCÍA-GUTIÉRREZ, Carlos; PACHEPSKY, Yakov; MARTÍN, Miguel Ángel. Technical note: saturated hydraulic conductivity and textural heterogeneity of soils. **Hydrology And Earth System Sciences**, [S.L.], v. 22, n. 7, p. 3923-3932, 20 jul. 2018. Copernicus GmbH. http://dx.doi.org/10.5194/hess-22-3923-2018

GERMER, Sonja; NEILL, Christopher; KRUSCHE, Alex V.; ELSENBEER, Helmut. Influence of land-use change on near-surface hydrological processes: undisturbed forest to pasture. **Journal Of Hydrology**, [S.L.], v. 380, n. 3-4, p. 473-480, jan. 2010. Elsevier BV. http://dx.doi.org/10.1016/j.jhydrol.2009.11.022

GRÄLER, Benedikt; PEBESMA, Edzer; HEUVELINK, Gerard. Spatio-Temporal Interpolation using gstat. **The R Journal**, [S.L.], v. 8, n. 1, p. 204, 2016. The R Foundation. http://dx.doi.org/10.32614/rj-2016-014

GODOY, Vanessa A.; ZUQUETTE, Lázaro Valentin; GÓMEZ-HERNÁNDEZ, J. Jaime. Spatial variability of hydraulic conductivity and solute transport parameters and their spatial correlations to soil properties. **Geoderma**, [S.L.], v. 339, p. 59-69, abr. 2019. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2018.12.015

Goovaerts, Pierre. **Geostatistics for Natural Resources Evaluation. Applied Geostatistics Series**. Oxford: Oxford University Press, p.483, 1997

GUPTA, S. C.; LARSON, W. E.. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. **Water Resources Research**, [S.L.], v. 15, n. 6, p. 1633-1635, dez. 1979. American Geophysical Union (AGU). http://dx.doi.org/10.1029/wr015i006p01633

GUPTA, Surya; HENGL, Tomislav; LEHMANN, Peter; BONETTI, Sara; OR, Dani. SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications. **Earth System Science Data**, [S.L.], v. 13, n. 4, p. 1593-1612, 15 abr. 2021. Copernicus GmbH. http://dx.doi.org/10.5194/essd-13-1593-2021

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L.. **Multivariate Data Analysis**. 5th ed. Prentice Hall, New York. 2009

Isaaks, E.H., Srivastava, R.M.. ***An Introduction to Applied Geostatistics*****.** Oxford University Press, New York. 1989

Klute, A., Dirksen, C.. **Hydraulic conductivity and diffusivity:** laboratory methods. Em: A Klute (ed.), *Methods of Soil Analysis*. (pp. 687–734) Agronomy Monograph, ASA-SSSA, Madison, 1986

KUINCHTNER, Angélica, BURIOL, Galileo Adeli. The climate of Rio Grande do Sul state according to Kóppen and Thornthwaite classification. **Disciplinarum Scientia** 2, 171-182. 2001. https://doi.org/10.37779/nt.v2i1.1136

KURNIANTO, Sofyan; SELKER, John; KAUFFMAN, J. Boone; MURDIYARSO, Daniel; PETERSON, James T.. The influence of land-cover changes on the variability of saturated hydraulic conductivity in tropical peatlands. **Mitigation And Adaptation Strategies For Global Change**, [S.L.], v. 24, n. 4, p. 535-555, 19 mar. 2018. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11027-018-9802-3

LANDIS, J. Richard; KOCH, Gary G.. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, [S.L.], v. 33, n. 1, p. 159, mar. 1977. JSTOR. http://dx.doi.org/10.2307/2529310

LARK, R. M.. Two robust estimators of the cross-variogram for multivariate geostatistical analysis of soil properties. **European Journal Of Soil Science**, [S.L.], v. 54, n. 1, p. 187-202, 21 fev. 2003. Wiley. http://dx.doi.org/10.1046/j.1365-2389.2003.00506.x

LARK, R. M.; BISHOP, T. F. A.. Cokriging particle size fractions of the soil. **European Journal Of Soil Science**, [S.L.], v. 58, n. 3, p. 763-774, jun. 2007. Wiley. http://dx.doi.org/10.1111/j.1365-2389.2006.00866.x

LEIJ, Feike J.; ROMANO, Nunzio; PALLADINO, Mario; SCHAAP, Marcel G.; COPPOLA, Antonio. Topographical attributes to predict soil hydraulic properties along a hillslope transect. **Water Resources Research**, [S.L.], v. 40, n. 2, p. 327-345, fev. 2004. American Geophysical Union (AGU). http://dx.doi.org/10.1029/2002wr001641

LIBOHOVA, Zamir; SCHOENEBERGER, Phil; BOWLING, Laura C.; OWENS, Phillip R.; WYSOCKI, Doug; WILLS, Skye; WILLIAMS, Candiss O.; SEYBOLD, Cathy. Soil Systems for Upscaling Saturated Hydraulic Conductivity for Hydrological Modeling in the Critical Zone. **Vadose Zone Journal**, [S.L.], v. 17, n. 1, p. 170051, 2018. Wiley. http://dx.doi.org/10.2136/vzj2017.03.0051

LIU, Zhipeng; MA, Donghao; HU, Wei; LI, Xuelin. Land use dependent variation of soil water infiltration characteristics and their scale-specific controls. **Soil And Tillage Research**, [S.L.], v. 178, p. 139-149, maio 2018. Elsevier BV. http://dx.doi.org/10.1016/j.still.2018.01.001

MARSHALL, M. R.; BALLARD, C. E.; FROGBROOK, Z. L.; SOLLOWAY, I.; MCINTYRE, N.; REYNOLDS, B.; WHEATER, H. S.. The impact of rural land management changes on soil hydraulic properties and runoff processes: results from experimental plots in upland uk. **Hydrological Processes**, [S.L.], v. 28, n. 4, p. 2617-2629, 19 abr. 2013. Wiley. http://dx.doi.org/10.1002/hyp.9826

MEURER, Katharina; BARRON, Jennie; CHENU, Claire; COUCHENEY, Elsa; FIELDING, Matthew; HALLETT, Paul; HERRMANN, Anke M.; KELLER, Thomas; KOESTEL, John; LARSBO, Mats. A framework for modelling soil structure dynamics induced by biological activity. **Global Change Biology**, [S.L.], v. 26, n. 10, p. 5382-5403, 23 ago. 2020. Wiley. http://dx.doi.org/10.1111/gcb.15289

MUHSINA, A.; JOSEPH, Brigit; KUMAR, Vijayaraghava. Principal Component Analysis on Soil Fertility Parameters of Vegetable Growing Panchayats/ Locations in Ernakulam District of Kerala. **International Research Journal Of Pure And Applied Chemistry**, [S.L.], p. 58-65, 4 jul. 2020. Sciencedomain International. http://dx.doi.org/10.9734/irjpac/2020/v21i1030208

PACHEPSKY, Yakov; PARK, Yongeun. Saturated Hydraulic Conductivity of US Soils Grouped According to Textural Class and Bulk Density. **Soil Science Society Of America Journal**, [S.L.], v. 79, n. 4, p. 1094-1100, jul. 2015. Wiley. http://dx.doi.org/10.2136/sssaj2015.02.0067

RIECKH, Helene; GERKE, Horst H.; SOMMER, Michael. Hydraulic properties of characteristic horizons depending on relief position and structure in a hummocky glacial soil landscape. **Soil And Tillage Research**, [S.L.], v. 125, p. 123-131, set. 2012. Elsevier BV. http://dx.doi.org/10.1016/j.still.2012.07.004

SHI, Yinguang; ZHAO, Xining; GAO, Xiaodong; ZHANG, Shulan; WU, Pute. The Effects of Long-term Fertiliser Applications on Soil Organic Carbon and Hydraulic Properties of a Loess Soil in China. **Land Degradation & Development**, [S.L.], v. 27, n. 1, p. 60-67, 15 maio 2015. Wiley. http://dx.doi.org/10.1002/ldr.2391

SOARES, Mauricio Fornalski; CENTENO, Luana Nunes; TIMM, Luís Carlos; MELLO, Carlos Rogério; KAISER, Douglas Rodrigo; BESKOW, Samuel. Identifying Covariates to Assess the Spatial Variability of Saturated Soil Hydraulic Conductivity Using Robust Cokriging at the Watershed Scale. **Journal Of Soil Science And Plant Nutrition**, [S.L.], v. 20, n. 3, p. 1491-1502, 27 abr. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s42729-020-00228-8

SHE, D; QIAN, Chen; TIMM, Luis Carlos; BESKOW, Samuel; WEI, Hu; CALDEIRA, Tamara Leitzke; OLIVEIRA, Luciana Montebello de. Multi-scale correlations between soil hydraulic properties and associated factors along a Brazilian watershed transect. **Geoderma**, [S.L.], v. 286, p. 15-24, jan. 2017. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2016.10.017

URSO, Alfonso; FIANNACA, Antonino; LAROSA, Massimo; RAVÌ, Valentina; RIZZO, Riccardo. Data Mining: prediction methods. **Encyclopedia Of Bioinformatics And Computational Biology**, [S.L.], p. 413-430, 2019. Elsevier. http://dx.doi.org/10.1016/b978-0-12-809633-8.20462-7

WANG, Shuang; ZHANG, Biao; XIE, Gao-Di; ZHAI, Xiu; SUN, Hai-Lian. Vegetation cover changes and sand-fixing service responses in the Beijing–Tianjin sandstorm source control project area. **Environmental Development**, [S.L.], v. 34, p. 100455, jun. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.envdev.2019.08.002

WALKLEY, A.; BLACK, I. Armstrong. AN EXAMINATION OF THE DEGTJAREFF METHOD FOR DETERMINING SOIL ORGANIC MATTER, AND A PROPOSED MODIFICATION OF THE CHROMIC ACID TITRATION METHOD. **Soil Science**, [S.L.], v. 37, n. 1, p. 29-38, jan. 1934. Ovid Technologies (Wolters Kluwer Health). http://dx.doi.org/10.1097/00010694-193401000-00003

## Artigo 2

# Gaussian co-simulation to access spatial variability of water dynamics in the soil at the watershed scale

**Abstract:** Accurate spatial characterization of saturated soil hydraulic conductivity ($K_{Sat}$) is vital for modeling hydrological processes. An effective representation of $K_{Sat}$ by field sampling is very difficult on a watershed scale, due the heterogeneity of landscape characteristics for wide land surfaces. To improve spatial distributions of $K_{Sat}$ at this scale, Sequential Gaussian Simulation (SGS) and Sequential Gaussian Co-Simulation (SGCS) algorithms were applied. The study was conducted in a 70 ha watershed, with 178 sampling points. In addition to $K_{Sat}$ the database is composed of Bulk Density (BD), Total Porosity (TP), Macroporosity (Mac), Microporosity (Mic) and Organic Carbon (OC). The Land Use (LU) was also quantified as a regionalized variable. A principal component analysis generated other two regionalized variables (PC1 and PC2). The simulations were evaluated by faithfully reproducing of the stochastic and spatial characteristics of the $K_{Sat}$ original data. To estimate the $K_{Sat}$, a univariate SGS was first run. The reproduction of the $K_{Sat}$ histograms was satisfactory, while the semivariograms of the simulated fields underestimated the sill. The SGCS was run using five different secondary variables: BD; Mac; LU; PC1 and PC2. All SGCS satisfactorily represented the original $K_{Sat}$ histogram, with a slight overestimation for some simulated fields. The semivariogram was very well reproduced for SGCS, being an indication that the process was quite efficient for all secondary variables. The smallest uncertainties were identified for the SGCS that used BD and Mac as secondary variables while the greatest uncertainties were associated with the SGS. The results of this study demonstrate that soil hydrological attributes as BD and Mac are a reliable auxiliary dataset to improve the spatial assessment of $K_{Sat}$ by SGCS at watershed scale.

**Key words:** Geostatistic simulation; Secondary variables; Principal components; Uncertainty.

## 1. Introduction

Saturated soil hydraulic conductivity ($K_{Sat}$) is an essential parameter for understanding soil water movement and consequently an important input for hydrologic models (Bagarello et al., 2020) $K_{Sat}$ is defined by the infiltration rate into soils at steady state condition (Naganna et al., 2017). Determining well representative values of $K_{Sat}$ and their distribution in space is a must in several circumstances, because $K_{Sat}$ often controls several hydrological processes, e.g. the amount of surface runoff, groundwater recharge, erosion and soil loss and movement of soil contaminants, affecting the whole environment in a watershed (Becker et al., 2018). $K_{Sat}$ estimates can support resource management decisions related to water conservation, irrigation systems, fertilizer application, drainage, solute mitigation, and plant growth (Gootman et al., 2020). These challenges suggest the importance of methods to estimate $K_{Sat}$ that are accurate and efficient to precisely determine the best soil management practices to improve crop quantity while being environmentally sustainable (Bagarello et al., 2020; Zhang and Schaap, 2019).

Soil attributes behave differently in the landscape due to factors such as parent soil material, climate, topography, vegetation cover and land uses. They are the main form of soil heterogeneity at many scales. It is widely known that the heterogeneity is greater for hydrological soil attributes and especially for saturated soil hydraulic conductivity that is overly spatially variable(Baiamonte et al., 2017; Centeno et al., 2020; Yong et al., 2020). The factors that drive to spatial variability of $K_{Sat}$ are more complex at the watershed scale and one reason for the complexity are the different land uses, which lead to significant differences in soil hydraulic conductivity (Gootman et al., 2020; Trinh et al., 2018). Studies conducted in small experimental plots, watersheds, especially when they present a rugged topography, have numerous different land uses and soil classes even in a very small range because of vertical zonality (Yong et al., 2020).

Geostatistical methods have been applied in soil and environmental science expecting to visualize or simulate the spatial structure and variability of soil attributes (Jin and Lv, 2020; Soares et al., 2020). Among geostatistical methods, kriging and stochastic simulation stand out since they are widely used

to model the variability of soil attributes. Stochastic simulation was designed to avoid the smoothing effect of kriging (dos Santos et al., 2021). In addition, stochastic simulation has the ability to not only predict the spatial distribution of a property at hand but also to assess both local and spatial uncertainty about the estimates, which should be incorporated in decision-making processes (Goovaerts, 2001; Ziegel et al., 1998). Sequential Gaussian simulation (SGS) is widespread in soil mapping due to improving running speed and being straightforward to implement simulation. Sequential Gaussian co-simulation (SGCS) is a bivariate extension of SGS proposed as a tool for using a secondary variable in order to improve the simulation results of the target variable (i.e, that one that will be simulated) based on a linear model of coregionalization (LMC) fitting and simple co-kriging algorithm (Jin and Lv, 2020). In this regard, SGCS can be an alternative for soil attributes with great spatial variability, such as $K_{Sat}$.

To run the GSCS algorithm it is necessary to select secondary variables that contribute to the main variable. In the case of $K_{Sat}$, soil physical and hydrological attributes as well as environmental characteristics such as the soil land use are usually good candidates to be secondary variables (Gootman et al 2020; Zhao and Lv, 2020). Sometimes those attributes that could contribute to the accuracy of the GSCS do not match this expectation. When there is a database with several regionalized variables at hand, an alternative would be to compress this information into dimensionless sets and use them as secondary variables. Principal component analysis (PCA) is a technique that indicates associations between variables, thus reducing the size of the number of data and grouping those with the most (John et al., 2021). These groups of variables generated by PCA can be regionalized and later used as secondary variables in the GSCS process if they present a significant correlation with the main variable, leading to a bivariate simulation a secondary variable that, although not having physical magnitude, represents the grouping of several features making the GSCS a multivariate process.

In developing countries, such as Brazil, the lack of hydrological information for watersheds is one of the main difficulties for the management of water resources. Planners often need to make decisions on ungauged watersheds or those that have few spatial and temporal data available (Beskow

et al., 2016). The Ellert Creek Watershed (ECW), for example, is a headwater watershed in which the main watercourse flows directly to the Pelotas River Watershed (PRW), a large watershed that somehow impacts more than half a million inhabitants in the region for direct supply of drinking water and all other human activities dependent on water resources. Furthermore, the ECW is subjected to impacts from intensive soil management for agriculture and livestock, being vulnerable to water erosion and floods, causing soil and nutrient losses resulting in economic and social damage (Soares et al., 2020).

Understanding hydrological processes is essential for the adoption of sustainable management practices in watersheds such as ECW and for this, knowledge of the spatial distribution of soil hydrological attributes such as $K_{Sat}$, which is a critical parameter for the application of hydrological models (Wang et al., 2013; Hu et al., 2015), can support decisions for management of water resources. Possibly, such local good practices will in the future reflect in the hydrological behavior of the PRW downstream.

The present study attempted to predict and evaluate spatial variability of saturated soil hydraulic conductivity by Gaussian sequential simulation and Gaussian sequential co-simulation, applying hydrological soil attributes and environmental characteristics as secondary variables. At a second stage resize the database by Principal Components Analysis and apply the generated components as secondary variables in the sequential Gaussian co-simulation to predict and evaluate the spatial variability of saturated soil hydraulic conductivity. Finally, verify the global and spatial uncertainty of Gaussian sequential simulation and of the Gaussian sequential co-simulation for each selected secondary variable.

## 2. Material and methods

### 2.1. Study and sampling area

This study was carried out in the above mentioned headwater watershed named Ellert Creek Watershed (ECW), located in the municipality of Canguçu, in the state of Rio Grande do Sul, south Brazil (Fig. 1). The ECW has an area of

approximately 0.7 km$^2$, and the altitude varies from 310 to 419 m. According to the Köppen climate classification, ECW is Cfa type, a mesothermal climate indicating wet subtropical conditions, characterized by an annual average temperature of 18 °C, with hot summers and cold winters (Kuinchtner and Buriol, 2001). The rainfall is well distributed throughout the year, and its annual mean is 1350 mm (INMET; available online atwww.inmet.gov.br). The regional relief varies from undulating to strong undulating, with a predominance of native forest or sparse shrub, and shallow soils.

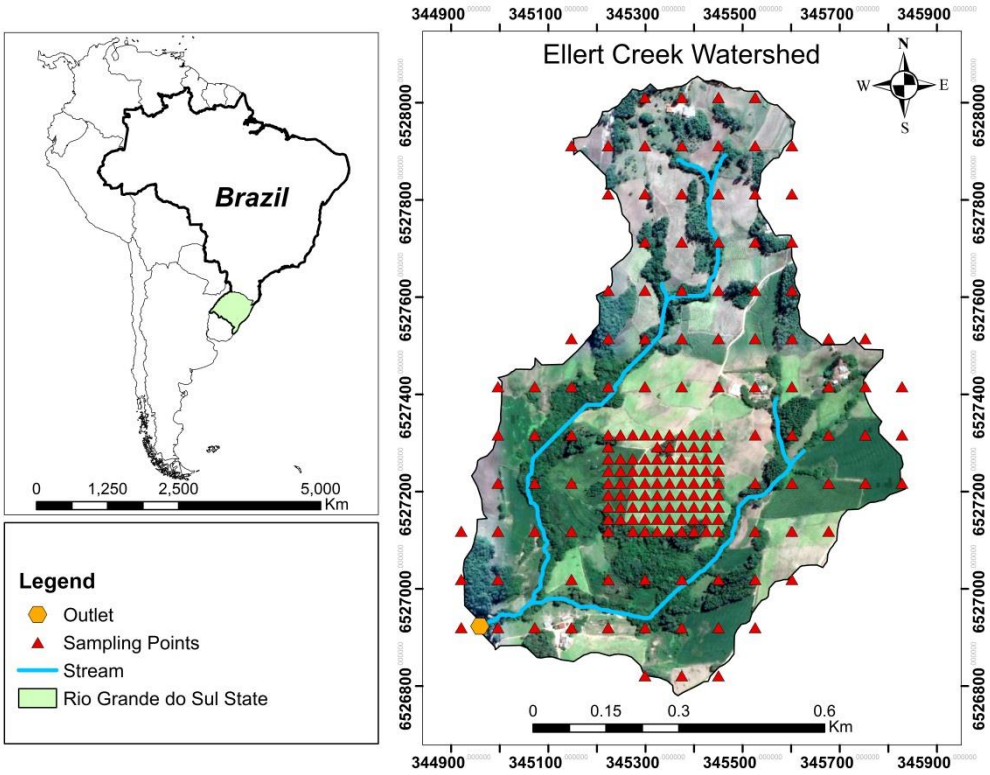


Figure 1. Location, hydrography, and sampling points of the Ellert Creek Watershed.

A sampling grid was established with points spaced 75 m in the east-west direction by 100 m in the north-south direction, accounting for 106 sample points. Subsequently, in order to better capture the spatial variability structure of the soil hydro-physical attributes, another grid with 78 soil sampling points was established, spaced by 25 m in both directions, totaling 184 sample points (Figure 1). At each sampling point, soil samples were collected with a preserved structure in the 0-0.20 m layer, and the following soil hydro-physical attributes of the soil were determined: Saturated soil hydraulic conductivity ($K_{Sat}$) soil bulk

density (BD), total porosity (TP), macroporosity (Mac), microporosity (Mic) and organic carbon (OC).

## 2.2. Exploratory data analysis

The exploratory statistical analysis was carried out in four stages. The first stage consisted of $K_{Sat}$ analysis, taken as the main variable to be modeled, data were compared before and after a declustering process. Mean, median, standard deviation (SD), maximum (Max) and minimum (Min.) values, skewness, kurtosis coefficient of variation (CV), and Kolmogorov Smirnov (K-S) normality test (5% significance level) were calculated.

The second stage consisted to spatialize the data (non-continuous bases) and mean, skewness, octile skewness and the normality test were calculated. These parameters were used to make decisions during the geostatistic process for secondary variables. After defining the secondary variables for the co-simulation, a third exploratory analysis was carried out for the densely sampled database of each of the secondary variables. The total number of data, the mean, the median, the variance, the CV, the skewness and the kurtosis where calculated.

The stability of the standard deviation is a measure to decide on the number of simulation runs to differentiate the effects from uncertainty. Thus, after simulations and co-simulations the accumulated standard deviation was graphed against the successively increasing number of runs until 50 realizations. Mean, median, SD and CV of the co-simulations were calculated and then compared with the main variable. Random field histograms were also constructed to be compared to that of the $K_{Sat}$ histogram. A Spearman correlation matrix of the secondary variable candidates with the $K_{Sat}$ was also carried out.

## 2.3. Declustering of data

The cell-declustering technique described by (Ziegel et al., 1998) was used due to the differences between the distances of the sampling points, which can cause clusterization of high/low means in certain areas. The equal-weighted mean (*w*) was, therefore, calculated for clustered and declustered data, as follows:

$$w = \frac{1}{B}\sum_{b=1}^{B} w_b \qquad \text{Eq.(1)}$$

where $B$ is the number of cells that contain at least one datum; $w$ is the equal-weighted mean of $z$-values within cell $b$. The declustered variance ( $\hat{\sigma}^2$ ) is then computed as Ziegel et al. (1998) suggested:

$$\hat{\sigma}^2 = \sum_{\alpha=1}^{n} \lambda_\alpha \left[z(u_\alpha) - w\right]^2 \qquad \text{Eq.(2)}$$

where $\lambda_\alpha$ are the declustering weights; $z(u_\alpha)$ is the value at a datum location.

Effective use of this declustering method depends on selection of an appropriate cell size for the calculation of weighting values $w$. A very fine grid size would result in all data points receiving a declustering weight of one. Conversely, a very coarse grid size would result in only a few highly weighted values.

## 2.4. Secondary variables

The two soil hydro-physical attributes chosen as secondary variables, soil bulk density (BD) and soil macroporosity (Mac), were originated from the same $K_{Sat}$ database. Therefore, for the use of BD and Mac, densely sampled in the co-simulation process, they were interpolated by ordinary kriging (Eq. 5). The secondary variable Land Use (LU) come from continuous bases and these

were densely extracted from these bases in GIS software for a regular 10x10m grid.

Principal component analysis (PCA) of five soil hydro-physical attributes (BD, TP, Mac, Mic and OC) was conducted to identify latent $K_{Sat}$ factors. This approach is useful for modeling and interpreting large datasets to reduce dimensionality of data and extract useful information for soil quality assessment and management (Hair Jr. et al., 2009). The chosen criterion was using the first generated components that explain 70 % of total variance of $K_{Sat}$ and that the minimum variance per component is not less than 20%. This step provides an "n" number of components for the PCA with all attributes that were organized as a spatial database and then used in geostatistical analysis (Libohova et al., 2018). The "factorextra" R code was used in the multivariate analysis (R Core Team, 2016). After being spatialized, the extracted factors were used as secondary variables too.

Geostatistics analyses were conducted using data obtained from GIS extraction and the multivariate statistical analysis. A semivariogram was first used to analyze the spatial structure of PC factors. The semivariogram, as defined below, is one of the most commonly used functions in assessing spatial autocorrelation (Goovaerts, 1999). For the spatialized secondary variables, two distinct semivariogram estimators were applied: The classic estimator of Matheron (Eq. 3) and the robust estimator of Cressie and Hawkins (Eq. 4).

$$\gamma_{Z_u}^{M}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z_u(\mathrm{x}_i) - z_u(\mathrm{x}_i + \mathrm{h})]^2 \tag{3}$$

$$\gamma_{Z_u}^{CH}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \frac{\left\{\left[|z_u(\mathrm{x}_i) - z_u(\mathrm{x}_i + \mathrm{h})|^{\frac{1}{2}}\right]^4\right\}}{0.457 + \frac{0.494}{N(h)}} \tag{4}$$

where $\gamma_{Z_u}^{M}(h)$ is the semivariance value using the Matheron estimator and $\gamma_{Z_u}^{CH}(h)$ is the semivariance by Cressie an Hawkins, $z_u(x_i)$ and $z_u(x_i + h)$ are values of $Z_u$ at locations $x_i$ and $x_i + h$, respectively, and $N(h)$ is the number of pairs $[z_u(x_i)$,

$z_u(x_i+h)]$ separated by the lag distance $h$. The term $0.457+(0.494/N(h)$ (Eq. 4) contributes to the bias correction, particularly if $N(h)$ is to large. Robust estimators are used to estimate variogram parameters for data including outliers (Lark, 2003). The selection of which estimator was applied followed the proposal of Lark and Bishop (2007) taking into account the conventional skewness and the octile skewness (Brys et al., 2004).

The method of ordinary kriging (OK) was applied for geostatistical interpolation. The spatial prediction of the unmeasured point $x_o$ is given by predicting the value $Z^*(x_0)$, which equals the line sum of the known measured values, following equation:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda Z(x_i) \quad (5)$$

where $Z^*(x_o)$ is the predicted value at the unmeasured position $x_o$, $Z(x_i)$ is the measured value at position $x_i$, $\lambda_i$ is the weighting coefficient from the measured position to $x_o$ and $n$ is the number of positions within the searched neighborhood. The OK method was calculated using the R software with a framework introduced by Gräler et al. (2016).

For the co-simulation process, all data identified as possible secondary variables were subjected to a correlation matrix analysis with the main variable $K_{Sat}$. Sperman's correlation is considered a non-parametric process and was chosen for this step due to the fact that of the main variable and many of the possible secondary variables do not follow normal distribution. Only the secondary variables that showed a significant correlation with $K_{Sat}$ were used in the subsequent co-simulation processes.

## 2.5. Sequential Gaussian co-simulations

The sequential stochastic simulations, including sequential Gaussian co-simulation (SGCS), aim at building alternative, equally probable, and high-resolution models of the spatial distribution of $z^*(u_0)$ by generating many

alternative outcomes, named as realizations (dos Santos et al., 2021). Unlike kriging, the sequential stochastic simulations do not focus on minimizing the kriging variance $\sigma^2(u)$ but aim at reproducing a sample histogram (Pachepsky and Acock, 1998). Moreover, measured data $z(u_i)$ are preserved. The procedure of SGCS is shown in Fig. 2 and detailed as follows:

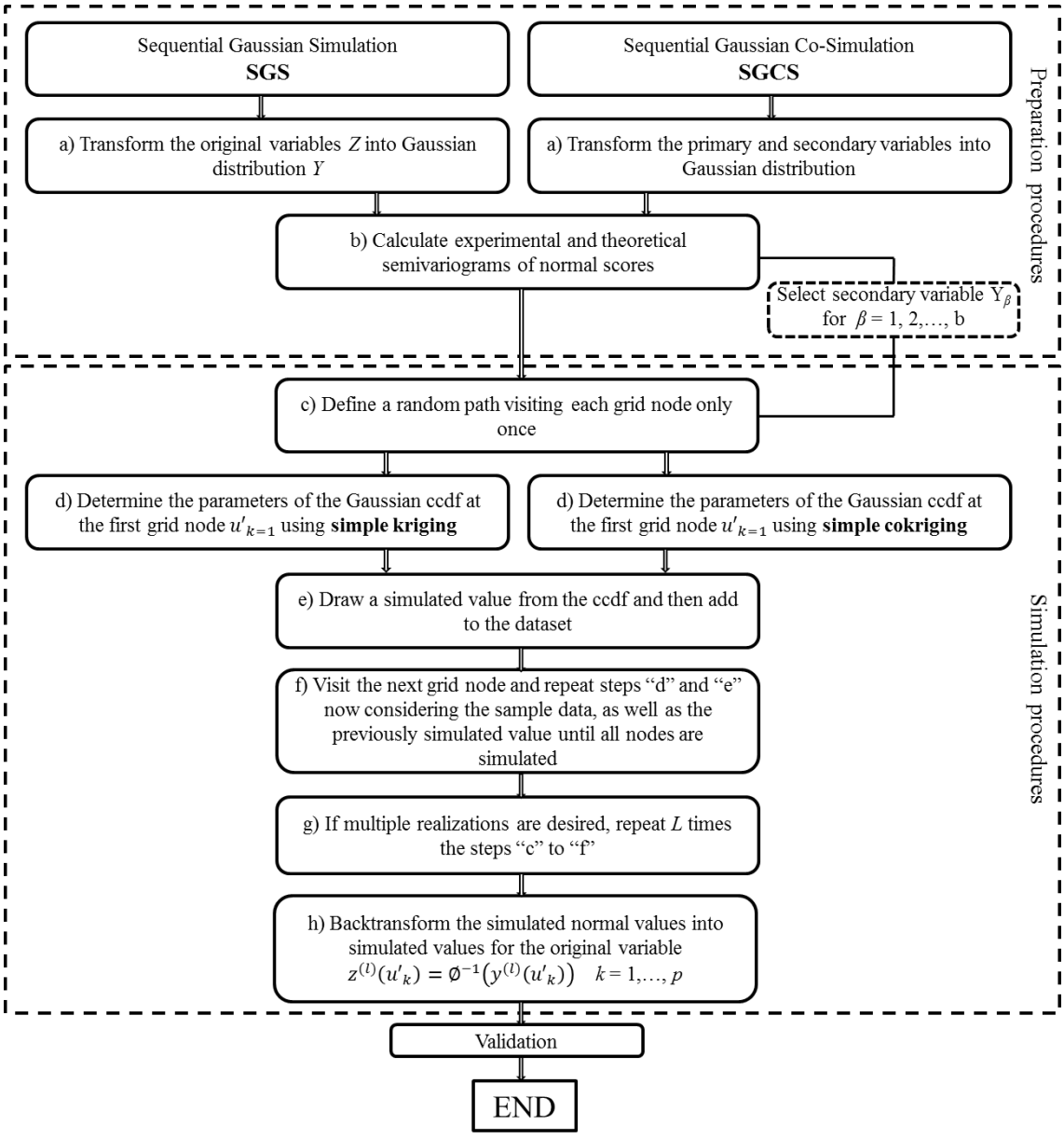


Figure 2. General procedure of sequential Gaussian simulation (SGS) and sequential Gaussian co-simulation (SGCS).

SGCS is a type of sequential stochastic simulation and incorporates different types of information, e.g., primary ($K_{Sat}$ data) and secondary (soil and

terrain data) variables used for this study. The initial step is to standardly normalize $z(u_i)$ into $y(u_i)$

$$y(u) = \emptyset(z(u)) = G^{-1}\left[F\left(z(u)\right)\right] \quad \text{Eq.(6)}$$

I. Define a random path visiting each node of unsampled locations $u_0$ only once. Let the number of unsampled nodes be $p$. Then we need to repeat Stage II until Stage V for *p* times.

II. At the $k$th node $u'_k$ where $k$ ranges from 1 to $p$, the conditioning data consist of neighboring primary $y_1(u_i)$, with $i = 1$ to $n$, and auxiliary $y_2(u_j)$, with $i = 1$ to $m$, as well as $y_1(u'_{k-1})$ within the range ($a$) of $u'_k$ ($n$ is the number of samples of the primary variable and $m$ is the number of data of the secondary variable). Here $y_1(u'_{k-1})$ contains simulated primary values until the $k-1$ step ($u$ represents nodes with samples and $u'$ represents nodes to be simulated).

III. Estimate the local Gaussian conditional cumulative distribution function (ccdf) for the primary variable at $u'_k$, with the mean equal to the OCK estimate $z_1{}^*_{OCK}(u'_k)$ and the variance equal to the OCK variance $\sigma^2_{OCK}(u'_k)$.

IV. Draw a value $y_1(u'_k)$ from the Gaussian ccdf using a Monte Carlo type simulation at $u'_k$ and add it to $y_1$.

V. Go back to Stage II if $k$ is less than $p$. If $k$ is equal to $p$, the sampling data ($y_1(u_i)$) and the simulated values ($y_1(u'_k)$) are back-transformed to $z_1(u_i)$ and $z_1{}^*(u'_k)$.

VI. If multiple realizations are desired $\{z^{(l)}(u),\ u \in A\},\ l = 1, \ldots, L$, the previous algorithm is repeated $L$ times.

Different implementations of sequential simulations in SGeMS can be used for different purposes (Olea et al., 2011; Remy et al., 2010). For this work, the sequential Gaussian co-simulation with Markov-model-1 (MM1) was selected. The absence of need to generate cross-correlation, while still maintaining the ability to produce realistic results (Karacan and Goodman,

2012; Olea et al., 2011), is an advantage of this method in face of especially limited data points.

Sequential Gaussian co-simulation allows for simulation of a Gaussian variable while accounting for the secondary information to which it is correlated too(Remy et al., 2010). Due to the nature of the simulation method, the variables should either be Gaussian or should be transformed to normal scores. The latter was followed in the study since the variables were not Gaussian.

MM1 considers the following Markov-type screening hypothesis during simulations: the dependence of the secondary variable on the primary is limited to the co-located primary variable. The cross-covariance is then proportional to the auto-covariance of the primary variable (Remy et al., 2010), which can be shown as:

$$\gamma_{12}(h) = \sqrt{\frac{C_{22}(0)}{C_{11}(0)}}\rho_{12}(0)\gamma_{11}(h) \tag{7}$$

where the cross semivariogram $\gamma_{12}(h)$ is proportional to the direct semivariogram $\gamma_{11}(h)$ of the primary variable $Z_1(u)$, $\rho_{12}(0)$ is the correlation coefficient between the variables, $C_{11}(0)$ is the variance of the primary variable $Z_1(u)$ and $C_{22}(0)$ is the variance of the secondary variable $Z_2(u)$. Thus, solving the co-kriging algorithm with MM1 requires the knowledge of correlation between primary and secondary variables, as well as the semivariogram(s) of the primary variable(s). These requirements, as implemented in face of limited amount of data for primary variables, were addressed by determining the range of correlation coefficients instead of using a single value. The MM1 model requires the secondary variable to be densely packed in the simulation space. For this reason, the soil hydro-physical attributes were kriged to later be resampled in the adopted simulation grid (10x10m) (Goovaerts, 1999).

## 2.6. Global and spatial uncertainty analysis

The global uncertainty was calculated using cumulative probability which determines the reliability of delineating areas with a high probability of exceeding a certain critical value and, thus, represents an indication of spatial uncertainty (Goovaerts, 2001). To assess the quality of the local uncertainty models obtained by the two simulation algorithms (SGS and SGCS), the procedure first proposed by Ziegel et al. (1998) and then applied by other researchers (Bourennane et al., 2007; Goovaerts, 2001) was used. The set of simulated $K_{Sat}$ fields enable the estimation of the conditional probability distribution at each validation point. This probability distribution can be expressed by the cumulative distribution function (CDF). To determine the uncertainties of the simulated $K_{Sat}$ by the CDF, a confidence interval of 90% was defined, with the percentile width within this interval being the uncertainty value associated with each simulation process.

To analyze the spatial distribution of uncertainties in the $K_{Sat}$ simulation, standard deviation maps were built for each of the processes. The standard deviation of all simulated fields was calculated pixel by pixel for all simulations applied in this work. The distribution map of standard deviation can be used to evaluate the spatial prediction uncertainty regarding the performance of the model. While ensuring the accuracy of spatial prediction, a smaller standard deviation value indicates that the spatial prediction uncertainty is smaller. The maps of standard deviation were used to assess local spatial simulated uncertainty of different sampling densities under the same simulated method (Shiwen et al., 2017).

# 3. Results and Discussion

## 3.1. Exploratory analysis and data declustering

The descriptive statistics of the declustered $K_{Sat}$ data is summarized in Table1. The declustering weights were calculated using the cell declustering procedure described by (Deutsch, 1989) and calculated in GSLIB (Ziegel et al.,

1998). Cells of various sizes were tested and, finally, square cells of 30m on each side were chosen. Table 1 summarizes the descriptive statistics of the $K_{Sat}$ data before and after the declustering process. Clearly, a large part of the mean fluctuation was dampened by cell declustering, which showed a 17% reduction in the average value (0.81 to 0.67 $m.h^{-1}$). After the declustering process, the sampling standard deviation of $K_{Sat}$ also presented a relative reduction greater than the mean, around 22% of the value of $K_{Sat}$ only quantified in the laboratory (0.89 to 0.70). The reduction of the mean and standard deviation values, in different relative scales, impacts the variation coefficient, resulting in a slight increase in the declustered data (99% to 104%).

Table 1 - Descriptive statistics and the non-parametric Kolmogorov-Smirnov test of saturated soil hydraulic conductivity data sets, before and after the declustering process, measured in the Ellert Creek Watershed, Southern Brazil

| $K_{Sat}$ | Clustered | Declustered |
|---|---|---|
| Mean | 0.81 | 0.67 |
| SD | 0.89 | 0.70 |
| Median | 0.52 | 0.45 |
| C.V. | 99.4 | 104.0 |
| Max. | 3.76 | 3.76 |
| Min | 0.01 | 0.01 |
| Skewness | 1.38 | 1.37 |
| Kurtosis | 1.46 | 1.38 |
| K-S Test | 0.175 | 0.172 |

$K_{Sat}$: Saturated soil hydraulic conductivity ($m\ h^{-1}$); SD: Standard deviation; C.V.: Coefficient of variation (%); Max and Min: maximum and minimum values, respectively, of the data set; K-S: Kolmogorov-Smirnov test at 5% of significance level (K-S critical value is 0.086)

The average and the median of declustered $K_{Sat}$ corresponded to 0.67 $m.h^{-1}$ and 0.45 $m.h^{-1}$, respectively. This contrast shows the great variability of this soil hydrology attribute, also demonstrated by the coefficient of variation (104%), being classified as having a high variability (CV > 35%) according to Wilding and Drees (1983). Soares et al. (2020) and (Dongli et al., 2017) found a

CV for $K_{Sat}$ of 99.0 and 50.8% in studies conducted in southern of Brazil and in the Loess Plateau region (China), respectively. Declustered $K_{Sat}$ data do not follow a normal distribution as determined by applying the K-S test, agreeing with results widely reported in the literature (Ahuja et al., 2010; Dongli et al., 2017; Godoy et al., 2019).

### 3.2. Secondary variables: PCA analysis

Table 2 shows that the PCA performed for the set of six variables, where the two PC's accounted cumulatively for total variance >79%. The first PC factor summarized the strong relationship among $K_{Sat}$, BD, TP and Mac, which showed the highest loadings. For the second PC, Mic and OC were the highly and significantly weighted variables.

Table 2 - Results of the principal component analysis applied to the soil attributes (saturated soil hydraulic conductivity, soil bulk density, total porosity, microporosity, macroporosity, soil organic carbon and elevation)

| | PC1 | PC2 |
|---|---|---|
| $K_{Sat}$ | 0.72 | 0.33 |
| BD | -0.92 | 0.32 |
| TP | 0.92 | -0.32 |
| Mic | -0.24 | -0.91 |
| Mac | 0.92 | 0.32 |
| OC | 0.27 | -0.58 |
| Eigenvalue | 3.18 | 1.59 |
| Variance | 53.00 | 26.52 |
| Cumulative | 53.00 | 79.53 |

$K_{Sat}$: Saturated soil hydraulic conductivity (m $h^{-1}$); BD: Soil bulk density (g $cm^{-3}$); TP: Total porosity (%); Mic: Microporosity (%); Mac: Macroporosity (%); OC: Soil organic carbon (%); Elev: Elevation (m); PC1: first principal component; PC2: second principal component.

In the PCA performed with six soil hydro-physical attributes, a high association of BD, TP and Mac was observed (±92) for the first principal component (PC), however, the total associated variance was about 53%. This may highlight that attributes did not completely overlap. For the second PC, Mic showed a high relationship with $K_{Sat}$ while CO demonstrated a moderate association, although the total variance corresponds to 26%. This result may be due to the fact that soil microporosity is a variable dependent on several other soil attributes, mainly linked to the textural class (Pachepsky and Park, 2015) and even though it is the most influential variable in PC2 (-91) followed by OC (-58). Mic is influenced by the contents of clay, silt sand and, widely variable in the ECW area.

### 3.4. Secondary variables: Correlation selection criteria

The strength of correlations between $K_{Sat}$ data and secondary variables was examined by a Spearman correlation matrix (Table 3). The correlation between the main variable and the secondary variables for the bivariate geostatistical processes is an important criterion, because to run the SGCS algorithm (SGSim) the correlation coefficient is applied. By the way, these analyses are only exploratory because the variables may be spatially correlated and hence not independent (Gamie and De Smedt, 2018).

Table 3 - Spearman's correlation coefficient values between saturated soil hydraulic conductivity and the other soil attributes, land use types, PC1 and PC2 principal components

| | $K_{Sat}$ | BD | Mac | LU | PCA1 |
|---|---|---|---|---|---|
| BD | -0.51* | | | | |
| Mac | 0.73* | -0.75* | | | |
| LU | 0.28* | -0.25** | 0.32* | | |
| PC1 | 0.76* | -0.90** | 0.91* | -0.11 | |
| PC2 | 0.39* | 0.21** | 0.36 * | -0.25* | $0.07^{NS}$ |

$K_{Sat}$: Saturated soil hydraulic conductivity (m $h^{-1}$); BD: Soil bulk density (g $cm^{-3}$); Mac: Macroporosity (%); LU: land use; PC: Principal component data sets (PC1 and PC2).

* Significant at 0.01 probability level

** Significant at 0.05 probability level

$^{NS}$ No significant Spearman correlation coefficient

Table 3 shows that $K_{Sat}$ is strongly correlated with Mac, negative and moderate correlated with soil BD and weakly correlated with LU. The high correlation between $K_{Sat}$ and Mac is expected due to the fact that macropores represent the preferred downward gravitational flow of water in the soil, a phenomenon widely reported in the literature (Centeno et al., 2020; Dongli et al., 2017; Soares et al., 2020). The negative correlation of $K_{Sat}$ with BD, or inversely proportional, was also an expected phenomenon. The increase in soil density implies in reducing soil porosity, therefore reducing saturated hydraulic conductivity. By the criterion of the correlation between the main variable and the second variable, Mac and BD were selected as covariates to run the GSCS algorithm. Also according to Table 3, PC1 and PC2 showed a positive correlation with KSat, being also selected as secondary variables.

### 3.5. Secondary variables: Geostatistical process

The exploratory analysis of possible candidates for secondary variables (non-continuous data) was carried out and is shown in Table 4. Only PC2

showed conventional skewness out of the range of [-1, 1], but for octile skewness, the PC2 is inside the range of [−0.2, 0.2], which indicates that the outliers affected the distribution. However, the outliers may not be unambiguously wrong and may represent an important phenomenon for the mapping and indicating that data transformation is not suitable. Thus, the robust semivariogram estimator is applied to the interpolators to weaken the effects of these outliers. According to the Kolmogorov-Smirnov test (significant at the level of 0.05), the data extracted from PC2 and Mac follow the normal distribution while the data from BD, Mac and PC1 do not follow the normal distribution.

Table 4 – Skewness and Octile Skewness coefficients and results for the K-S test used to select the semivariogram estimator for the secondary variables

| SV | Skewness | Octile Skewness | K-S (p-value) | Normal | SE |
|---|---|---|---|---|---|
| BD | -0.98 | 0.02 | 0.01 | No | M |
| Mac | 0.15 | -0.17 | 0.05 | Yes | M |
| PC1 | 0.62 | 0.01 | 0.02 | No | M |
| PC2 | -1.00 | 0.05 | 0.22 | Yes | CH |

SV: secondary variable; BD: bulk density (g $cm^{-3}$); Mac: macroporosity (%); PC1 and PC2: first and second principal components, respectively; K-S (p-value): calculated p-value applying the Kolmogorov-Smirnov test (p-value = 0.05 was used as a reference value); SE: Semivariance estimator (M: Matheron classical estimator; CH: Cressie and Hawkins robust estimator)

The variographic analysis (Tab. 5) demonstrates that all semivariograms were fitted using the spherical theoretical model. The great nugget effect of Mac, much higher than for the other attributes, is due to the effect of the magnitude of the variable, given in percentage, but does not indicate that the macroporosity in the study area has a less stronger spatial autocorrelation or has problems with geostatistical modeling. Indeed, the other parameters of the semivariograms generated for Mac demonstrate excellent fit and spatial dependence, as well as for all other databases that will be secondary variables in the $K_{Sat}$ co-simulation process. The nugget coefficients of the other four secondary variables were between 1.59 and 0.01, considered low, according to the magnitude and amplitude of each database and also had their spatial

dependences weal modeling. The range values of the hydro-physics attributes that will be used as secondary variables were 405 m for BD and 136 m for Mac. The ranges for the semivariograms of PC's eigenvalues datasets were 213 m for PC1 and 187m for PC2.

Table 5 - Theoretical semivariogram models and their adjusted parameters for the secondary variables

| Semivariogram | Model | $C_0$ | C | $s^2$ | *a* (m) |
|---|---|---|---|---|---|
| BD | Spherical | 0.02 | 0.01 | 0.03 | 405 |
| Mac | Spherical | 32.26 | 31.21 | 61.05 | 136 |
| PC1 | Spherical | 1.59 | 1.72 | 3.18 | 213 |
| PC2 | Spherical | 0.36 | 1.24 | 1.62 | 205 |

BD: soil bulk density (g $cm^{-3}$); Mac: macroporosity (%); PC1 and PC2: first and second principal components, respectively; $C_0$: nugget effect*; C*: structural variance; $s^2$: sample variance; *a*: range.

Fig. 3 represents the spatial distribution of secondary variables applied later in the co-simulation process. We have applied attribute maps linked to $K_{Sat}$ as the second variable for co-simulation. The maps of BD (Fig 3b) and Mac (Fig. 3c) were interpolated from the ordinary kriging of sampled points (Fig 3a) with the SGeMS software. For the Land Use map (Fig. 3d) second imagery source was Landsat 7 (03/29/2003) and Landsat 8 (04/17/2013) satellites with 30 m resolution (path 220, row 79), from which we used the full set of spectral bands (USGS 2015) to extract the satellite images provided there for preparing the land use map for the region of interest. An advantage of Google Earth is that it provides images taken at different time periods which will be very useful for soil land use change detection studies. The area used for the land use map as the secondary variable (Fig 3d) was stratified into: Forest (dark green); Annual cropping (light green); Silviculture (Gray) and pasture (Yellow).

The secondary variables corresponding to maps 3(e) and 3(f) are products of the kriging of the eigenvalues extracted from the PCA, representing the first principal component PC1 and the second principal component PC2.

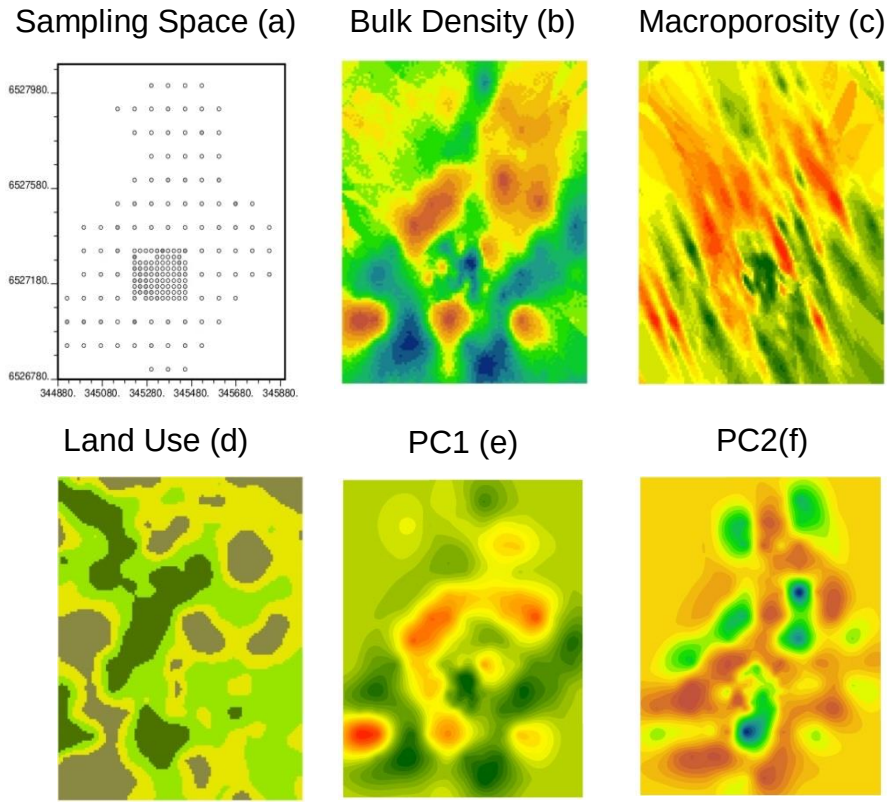


Figure 3. Local where the secondary variables have been kriging or digital sampled whit the (a) field sample points; (b) map of bulk density; (c) macroporosity; (e) land use; (e) PC1and (f) PC2.

All maps of the secondary variables were gridded in pixels corresponding to 10m x 10m, through the "extract by point" toolbox of the ArcGIS software. The numerical values of each pixel were extracted resulting in a database with 13770 points for each secondary variable. This work routine was used to meet the assumptions of the Markov1 model (MM1), where the variable of interest has a reduced number of samples (178) compared to a high number of points (13770) extracted for the secondary variable maps.

The statistical summaries of secondary variables are presented in Table 3. All secondary variables have the same data number 13770. The difference between the means and the medians was practically null for all data sets, indicating a low asymmetry in their frequency distributions. Probably this condition of symmetry of the data is linked to the large number of data for each series, according to the central limit theorem the distribution of the sample means will be approximately normally distributed (Hoeffding and Robbins, 1948). The coefficient of variation (CV) for all data sets was also quite low.

According to the classification of Wilding and Drees (1983), the coefficients of variation (CV) for all secondary variables were considered low (CV ≤ 15%). For the secondary variables PC1 and PC2, the coefficients of skewness and kurtosis were quite low, classifying all the other secondary variables as "symmetric" and "mesokurtic".

Table 6 - Exploratory analysis of the secondary variables densely sampled in the simulation space

| Variables | $\overline{X}$ | Med. | s | CV% | Skew | Kurt |
|---|---|---|---|---|---|---|
| BD | 1.39 | 1.38 | 0.09 | 7.0 | -0.35 | 0.65 |
| Mac | 26.87 | 26.86 | 2.33 | 9.0 | 0.23 | 2.53 |
| LU | 4.61 | 4.60 | 0.26 | 6.0 | 0.20 | -0.33 |
| PCA1 | 0.10 | 0.09 | 0.70 | 7.1 | 0.54 | 1.18 |
| PCA2 | 0.10 | 0.10 | 0.68 | 7.2 | -0.54 | 1.35 |

BD: soil bulk density (g $cm^{-3}$); Mac: macroporosity (%); LU: Land use; PC1 and PC2: first and second principal components, respectively; $\overline{X}$ and Med.: mean and median values, respectively; s: standard deviation; CV (%): coefficient of variation; Skew: coefficient of skewness; Kurt: coefficient of kurtosis.

### 3.6. Sequential Gaussian simulation

Fig. 4 shows the graph of the accumulated standard deviation (ASD) considering all random fields generated by SGS. The standard deviation was calculated by joining the first with the second field, and then to the first three fields, and so on until obtaining the standard deviation for at least 50 realizations. All of the standard deviations these fields were calculated together at the end of the SD dataset. It is possible to identify the number of necessary fields to capture the variability of the studied phenomenon through the stabilization of the standard deviation.

For this work it was assumed that 50 random fields would be sufficient for the SGS, once the ASD of simulated values became stabilized with a lower number of random fields. Generating more fields than this number would not

add more information about the $K_{Sat}$ variability and would be time consuming beyond useless machine processing. Fig. 4 represents the ASD for simulated $K_{Sat}$ random fields generated by SGS. When the value of “0.726” was repeated for the fourth time, the ASD was considered to be stable, which occurred in the simulated field number 39.

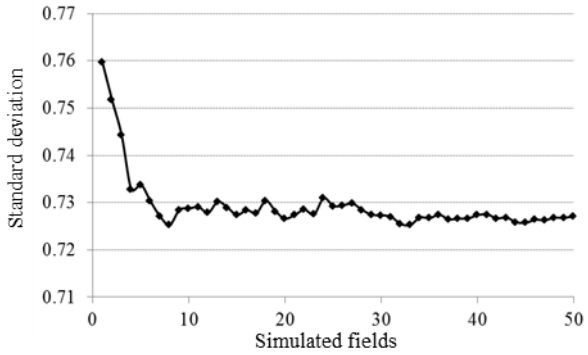


Figure 4 - Determination number of fields to run geostatistical simulation by cumulate standard deviation

Four randomly selected realizations are shown in Fig. 4. Each realization represents a realistic spatial distribution of $K_{Sat}$ without a smoothing effect represented by sudden color changes from one block to another. Although maps are not smoothed out, it is possible to observe a pattern in which the highest values of $K_{Sat}$ are located near the areas of native forest and watercourses. This behavior is widely reported (Baiamonte et al., 2017; Becker et al., 2018; Gootman et al., 2020) and it is due to soil characteristics in forest areas where soil management is reduced or absent, mainly preserving macroporosity and soil organic carbon, maintaining soil density at natural levels (Centeno et al., 2020; Gootman et al., 2020).

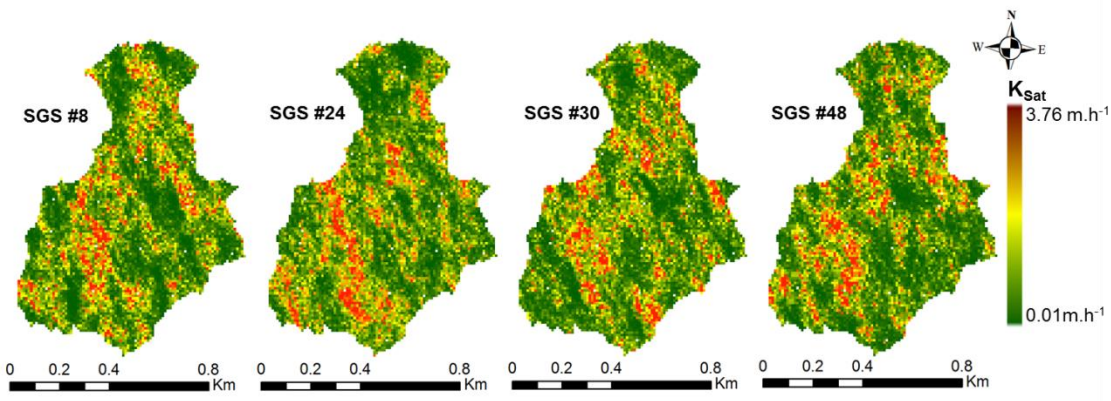


Figure 5. Maps of four random fields for $K_{Sat}$ simulations by GSG

However, simulated realizations must be checked carefully before use. A reliable simulated model should reproduce the observed values of $K_{Sat}$ at their locations, the sample histogram and the semivariogram model (Madenoglu et al., 2020). This is not practical for examining the reproduction of sample statistics by all generated realizations; therefore, for the same four realizations selected randomly for the reproduction of the maps, the statistics, the histograms and the semivariograms were also verified.

Fig. 6 shows the histogram of the declustered $K_{Sat}$ and four histograms of the same realizations for GSS. This figure clearly demonstrates the proximity of the original histogram of the primary variable with the histograms of the achievements generated for realizations by the simulation process. The descriptive statistics including mean, standard deviation and coefficient of variation (CV) of the declustered $K_{Sat}$ and the four simulated random field is shown at each histogram too, for comparison purposes. The mean values of original data were not reproduced well except for random field number 24 (SGS #24) in which the mean ($\dot{X}$) and (SD) the standard deviation of declustered $K_{Sat}$ has been overestimated. The CV for SGS #24 was somewhat smaller for all other realizations and also from declustered $K_{Sat}$.

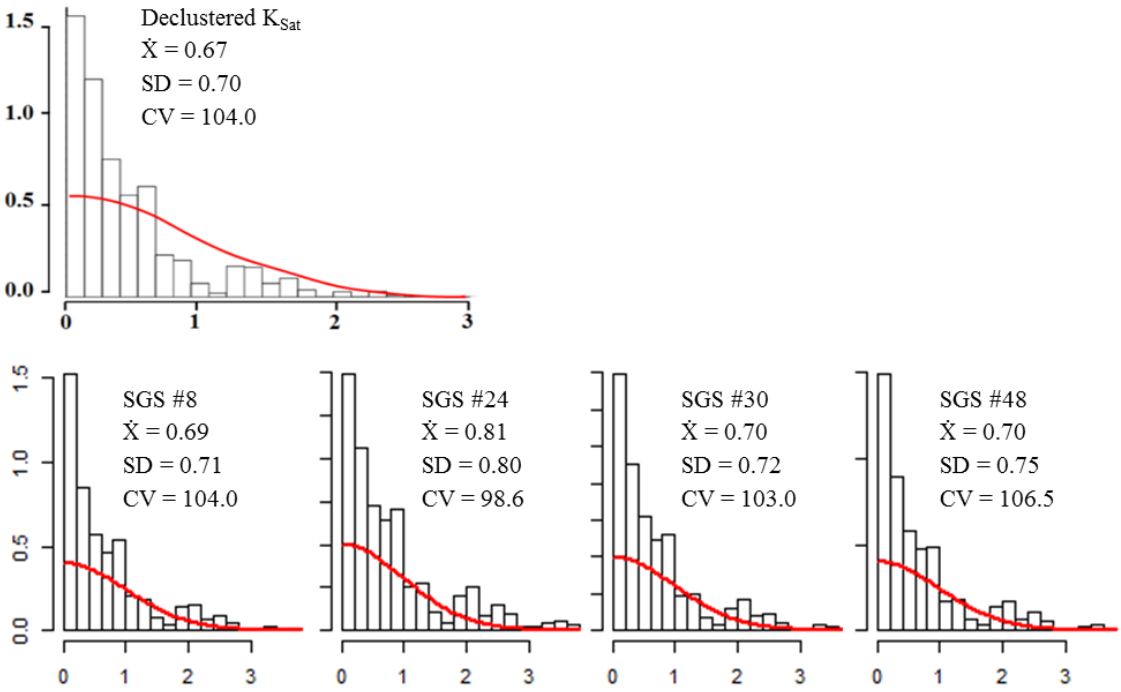


Figure 6. Histograms and statistical parameters of declustering $K_{Sat}$ and four randomly realizations for SGS process

The reproduction of the sample semivariogram model by some randomly selected simulated maps is represented in Fig. 7, which shows the longest range for $K_{Sat}$ represented by the black line. The semivariogram is in azimuth of 157.5 ° and the range is 180m. The nugget effect (C0) was 0.28 and the sill (C0 + c) was 0.65 given by the spherical model calculated for the declustered $K_{Sat}$ data. The colored lines represent the semivariograms of simulated fields chosen at random from the SGS processes.

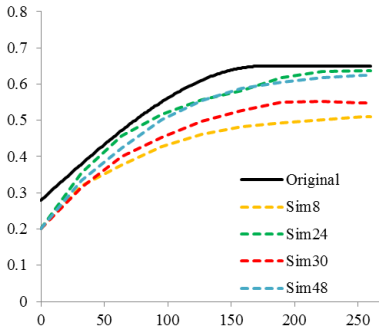


Figure 7. The sample theoretical semivariogram model of $K_{Sat}$ data and those generated by four random sequential Gaussian simulation realizations (SGS-8$^{th}$, SGS-24$^{th}$, SGS-30$^{th}$ , and SGS-48$^{th}$).

It is possible to observe that the semivariograms of the simulated fields underestimated the sill in relation to the semivariogram of the declustered $K_{Sat}$. Indeed, some discrepancies are expected between realizations and the original model called ergodic fluctuations (Goovaerts, 2001). These fluctuations may have different reasons, including the algorithm used for the simulation, the amount of data to be used for the simulation, and the semivariogram model parameters (Goovaerts, 2001). In the case of the SGS algorithm, the statistics (histogram/semivariogram model) reproduced over a number of realizations should be, on average, equal to the sample statistics (Madenoglu et al., 2020).

### 3.7. Sequential Gaussian co-simulations

Fig. 8 shows the graph of the accumulated standard deviation (ASD) considering all random fields, generated with different auxiliary variables. For this work, as shown in Fig. 4, it was assumed that 50 fields would be enough for the SGCS, once the ASD of simulated values stabilized with a lower number of random fields.

Fig. 8a represents the ASD for $K_{Sat}$ random fields that used BD as secondary variable ($K_{Sat}$ - BD). When the value of “0.803” was repeated for the fourth time, the ASD was considered to be stable, which occurred in the simulated field number 38. Fig. 8b represents the ASD of $K_{Sat}$ random fields considering Mac as secondary variable ($K_{Sat}$ - Mac). The SD of this simulation stabilized when it reached the value of "0.786" in the simulated field of number 40. For the composition "$K_{Sat}$ - LU", the ASD of $K_{Sat}$ random fields stabilized at 0.828 in the simulated field number 48. Manchuk and Deutsch (2012) defined the number of 50 realizations as sufficient for the estimation of stochastic variables, since, from a set of 50 simulations, the standard deviation presented stability. Qi et al. (2018), studying soil salinization, generated 100 realizations to obtain a stable result. Therefore, the results presented here are in agreement with other studies that generated $K_{Sat}$ random fields. Fig. 8(d; e; f) showed the ASD for SGCS processes using PC's as secondary variables. For the co-simulation that used PC1 as a secondary variable, the ASD stabilized at36 when it reached the standard deviation value of 0.783. As for the co-simulation

that used PC2 as a secondary variable, the ASD stabilized in field 44 at the standard deviation value of 0.795.

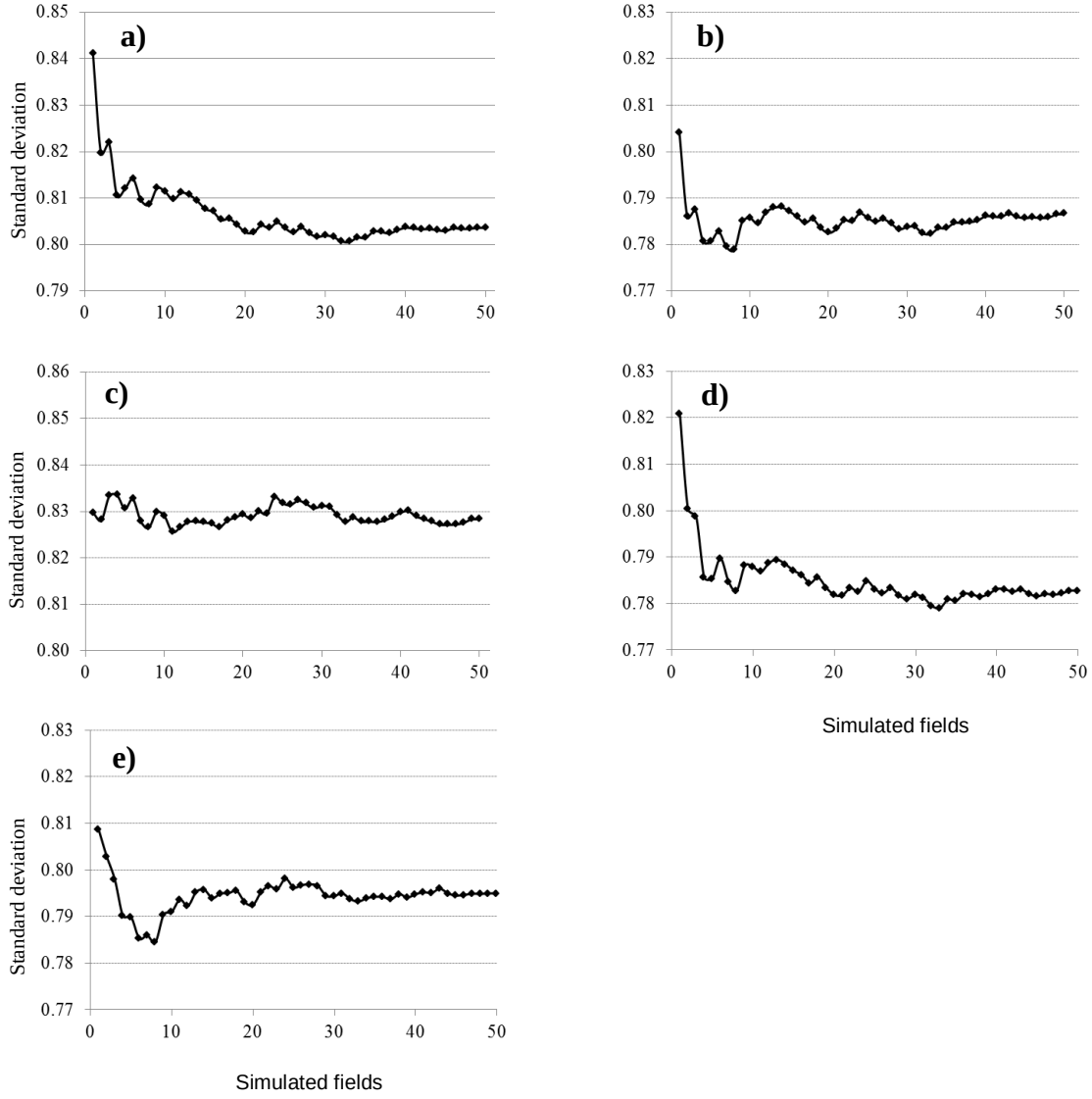


Figure 8. Determination number of fields to run geostatistical simulation by cumulate standard deviation; (a) $K_{Sat}$XBD; (b) $K_{Sat}$XMac; (c) $K_{Sat}$XLU; (d) $K_{Sat}$XPC1; (e) $K_{Sat}$XPC2.

It is important to note that the $K_{Sat}$ principal component analysis stratified in the first component (PC1) soil attributes that are highly impacted by the type of land use, such as Mac and BD (Fei et al., 2019; Libohova et al., 2018), explained almost 40% of the $K_{Sat}$'s variance in the study area. The second main component (PC2) combined attributes linked to the soil textural classes, such

as Mic and OC (Alvarez et al., 2019) and corresponds to 20% of the total variance.

Fig. 9 shows four random field maps for each SGCS process. Stochastic simulation algorithms have been used in field studies, especially in the situation where the spatial variation of the measured field must be preserved (Goovaerts, 1999). This technique avoids the smoothing effects of Kriging estimators (Ziegel et al., 1998) and provides more reliable simulated results through computing plenty of possible realizations of the unknown spatial distribution (Delbari et al., 2009). The set of individual realizations provided exhaustive descriptions of the spatial distribution of $K_{Sat}$ over the study area. The general pattern of distribution of $K_{Sat}$ by the GSCS method is the same as observed by SGS, with a higher accumulation of high values in areas close to watercourses and native forest, and lower on areas under intensive soil cultivation. However, for each group of SGCS it is possible to observe details in the spatial distribution pattern of $K_{Sat}$ that are due to the nature of the secondary variable. For the random field maps generated by the GSCS when the secondary variables were BD and Mac, the spatial distribution pattern is very similar, as well for the PC1, which is essentially a principal component generated by these two variables. For the SGCS that uses land use as a secondary variable, it is possible to observe in the four maps a region to the south of the watershed, where there is an area of preserved forest, with many points with high values of $K_{Sat}$.

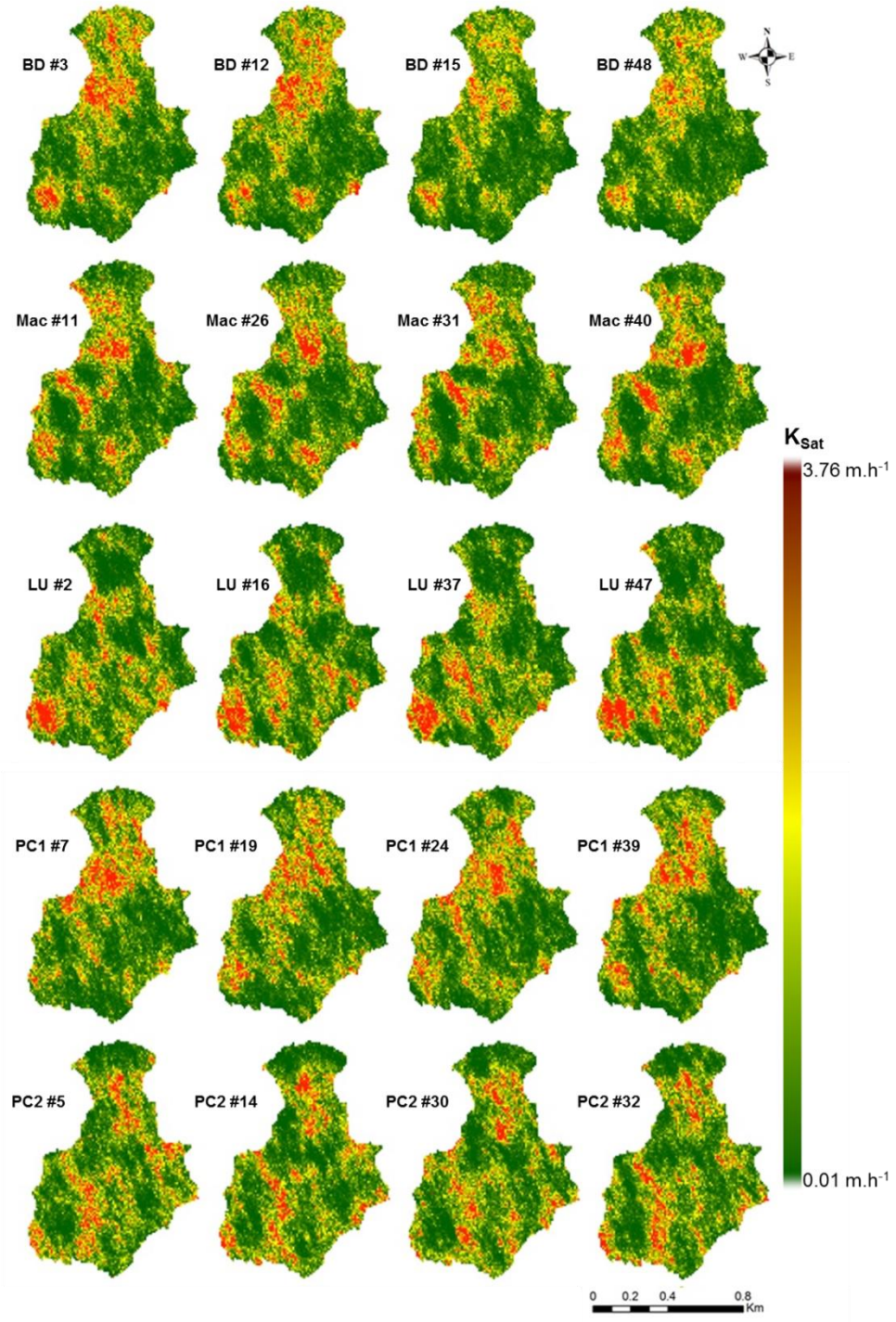


Figure 9. Maps of four random fields for KSat simulations by GSCG

Fig. 10 shows the histogram of the declustered $K_{Sat}$, which is the modeled primary variable, and four realizations histograms randomly selected for each process of the Gaussian Sequential Co-Simulation (GSCS), considering each secondary variable (BD, Mac, LU, PC1 and PC2). The

proximity of the original histogram of the primary variable with the histograms of the achievements generated by GSCS is clearly shown in fig.10. For the same realizations used to carry out a histogram, those parameters were also calculated as standards for comparison. The mean values of original data were well reproduced and the standard deviation of most of realizations has a slightly increase compared to original $K_{Sat}$ data. The CV's were a little smaller for all the randomly chosen realizations for this parallel. The histograms of the simulated fields have followed very closed those of the original variable as well.

Still in relation to the stochastic characteristics of the fields simulated by the GSCS, the standard deviation showed a greater variation than the other statistics already mentioned, such as for the co-simulations that used BD and LU as secondary variables, showing a variation of up to 15% in SD for some simulated fields. The standard deviation is a measure of the amount of dispersion of a set of values and is strongly impacted by outliers. For each field generated by GSCS, 8711 values are calculated for each simulated cell. This value is multiplied by the number of fields required for the effectiveness of the method and can reach millions of simulated values. The calculation of this large amount of values can present outliers, or more likely can suppress the presence of outliers of the original $K_{Sat}$ data series, that counts only 178 data. These differences, such as the reduction of values considered outliers, directly impact on the standard deviation of the fields simulated by the GSCS. These small to moderate changes in the means and the SD of the results end up having a more pronounced impact on the values of the coefficient of variation (CV) that is the standardized measure of dispersion defined as the ratio of the SD to the mean.

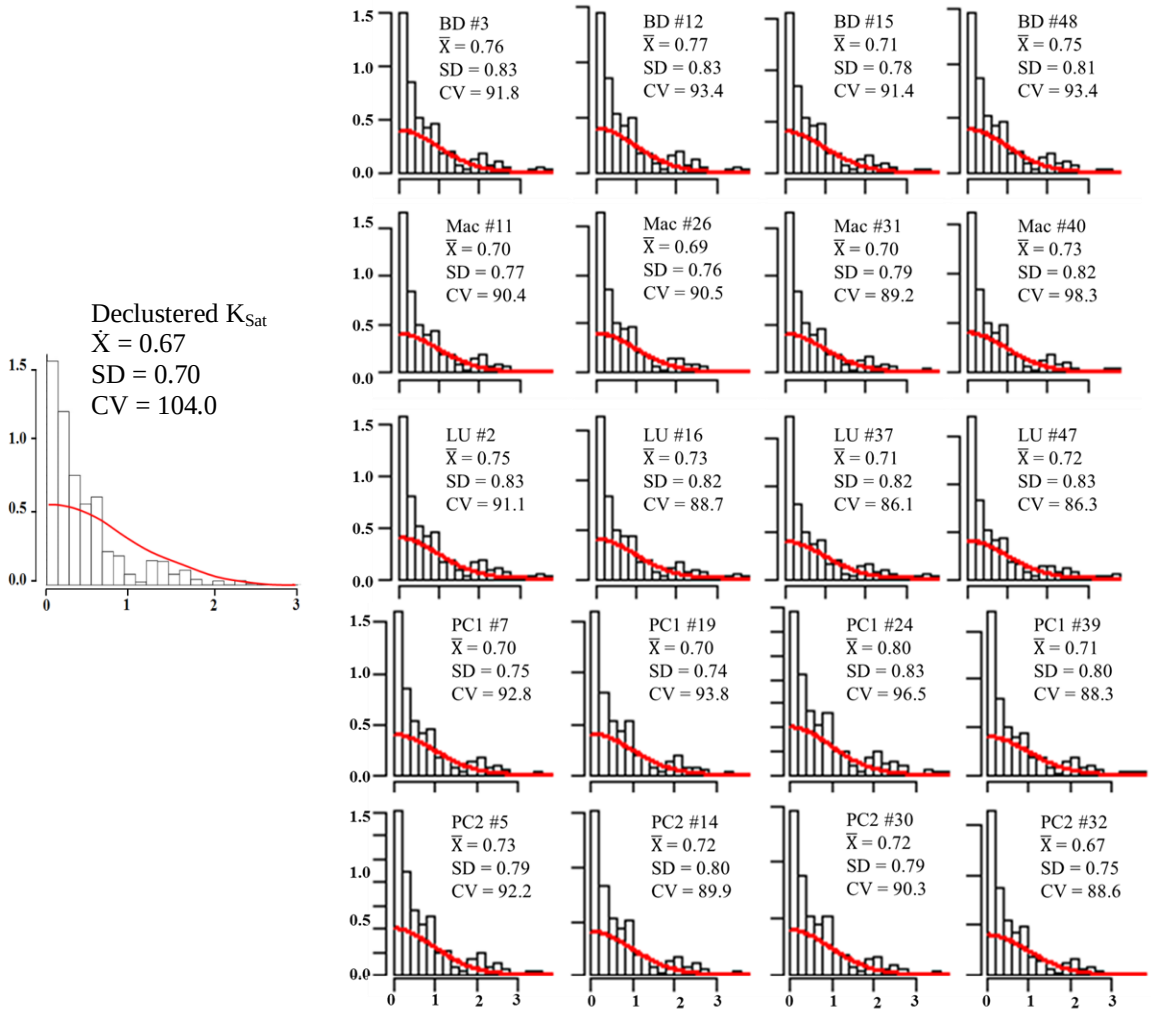


Figure 10. Histograms and statistical parameters of declustering $K_{Sat}$ and of 4 randomly selected realizations for each SGCS process.

Fig. 11 shows the semivariograms for the greater continuity direction in the five graphs (a, b, c, d, and e) represented by the black line. The semivariogram is in the direction of 157.5 ° and the range is 180m. The nugget effect ($C_0$) was 0.28 and the sill ($C_0$ + c) was 0.65. This is the spherical model calculated for the declustered $K_{Sat}$ data. The colored lines represent the semivariograms of the same simulated fields presented in Fig. 10. Fig. 11a is the SGCS using BD as a secondary variable; 11b is for the process that used the Mac as a secondary variable; Fig. 11c for Land Use and; the figures 11d and 11e are the semivariograms for SGCS that used principal components as secondary variable (PC1 and PC2 respectively).

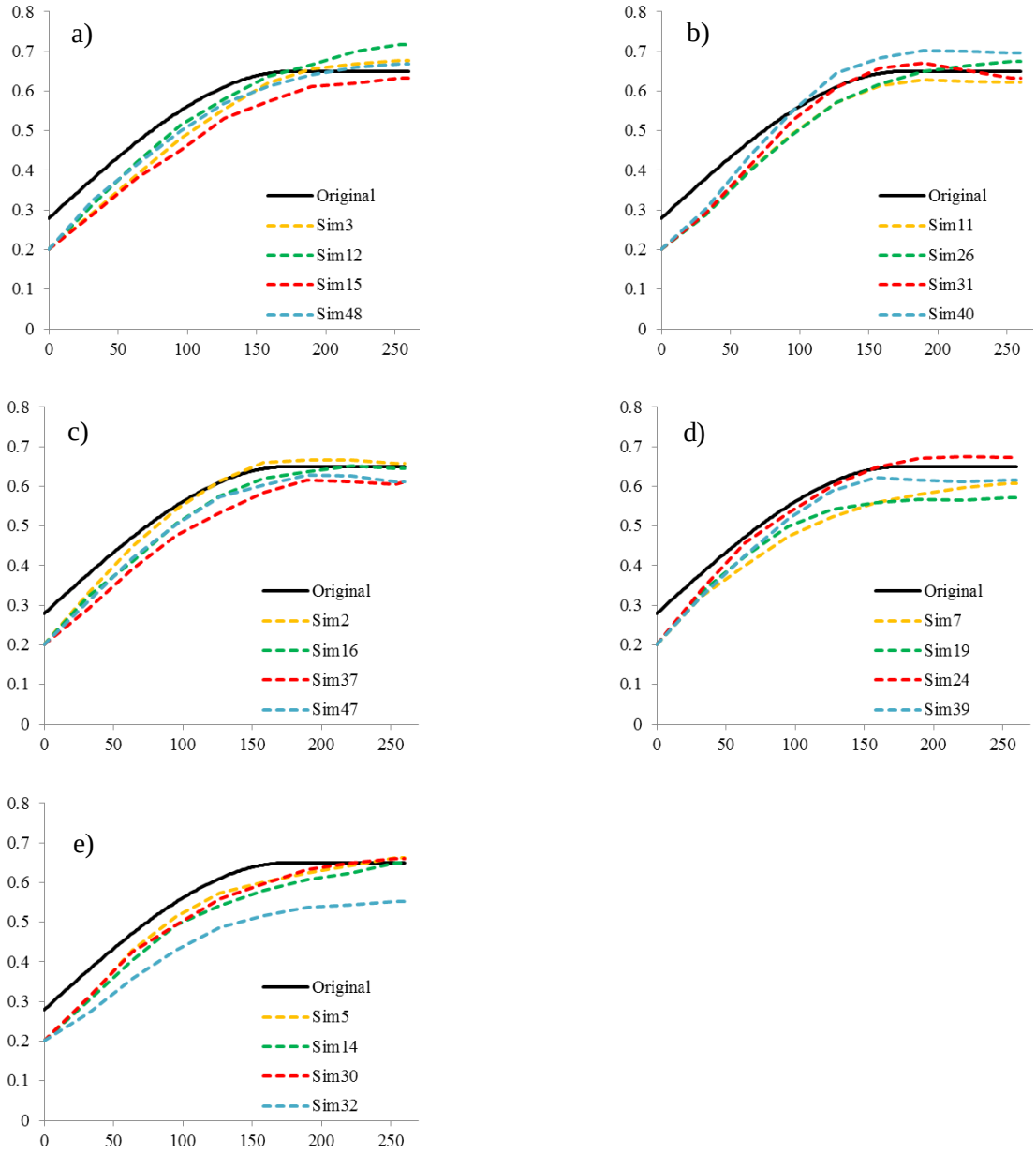


Figure 11 - The sample theoretical semivariogram model of $K_{Sat}$ data and those generated by four random sequential Gaussian co-simulation realizations for each secondary variable: (a) bulk density (BD), (b) macroporosity (Mac), (c) land use (LU), (d) and (e) PC1 and PC2 components, respectively.

Slight differences that appear in large lags are typically impacted by the extensiveness of the underlying area, which is bounded regionally, termed as ergodic fluctuations (Matheron et al., 1987). In geostatistical simulation, the discrepancy between the simulated values over a set of realizations and the corresponding model parameters is referred as to ergodic fluctuations (Nussbaumer et al., 2018). Ergodicity allows inference of the statistical

parameters of stationary random functions from realization statistics. Simulated values tend towards normality as the size of the model increases as regarding the range of correlation. Fluctuations in the simulated statistics are expected in any Gaussian simulation algorithm. Along these lines, the randomly generated semivariograms by the SGCS process show that when compared to the variograms of the original data (black line), it can be seen that the spatial continuity is maintained, especially when the secondary variables applied are BD and LU. It is observed that the simulations reproduced well the behavior of the average spatial relationship between pairs of samples, although in the case of declustered $K_{Sat}$ some simulations have underestimated the original data variance, which is in agreement with a smaller variance of all the simulations.

### 3.8. Uncertainty analysis

The parameters to determine the uncertainties of the saturated soil hydraulic conductivity range between 5% and 95% probability of not exceeding the chosen simulated values, representing, in this way, a more favorable scenario and another more unfavorable scenario with respect to the flow of water in the soil. For a better visualization, the means of all simulated fields for all simulations, univariate and bivariate, were plotted from the lowest to the highest value in the same graph, identifying the excluded and included probability percentiles according to the adopted confidence interval constituting the cumulative distribution function (CDF) for each process. Figure 12 shows the CDF's for the SGS and for the five SGCS processes identified by the applied secondary variable. Here, the "drop" of the CDF indicates the width of uncertainty associated with each simulation process. The greater this width, the greater will be the global uncertainty of $K_{Sat}$. The width of each process is described in the caption of fig. 12, as well as the proportional magnitude, shown by lines, where each color represents a certain process.

Fig.12 shows that the overall uncertainties for all processes were relatively low, ranging 23% from smallest to largest uncertainty. Even so, the smallest global uncertainties were identified for the SGCS that used BD and Mac as secondary variables, with values of 0.068 $m.h^{-1}$ of uncertainty. The

greatest uncertainties are associated with the univariate process SGS and for the SGCS that used the LU with secondary variable, presenting respectively values of 0.088 $m.h^{-1}$ and 0.085 $m.h^{-1}$ of uncertainties to determine the $K_{Sat}$.

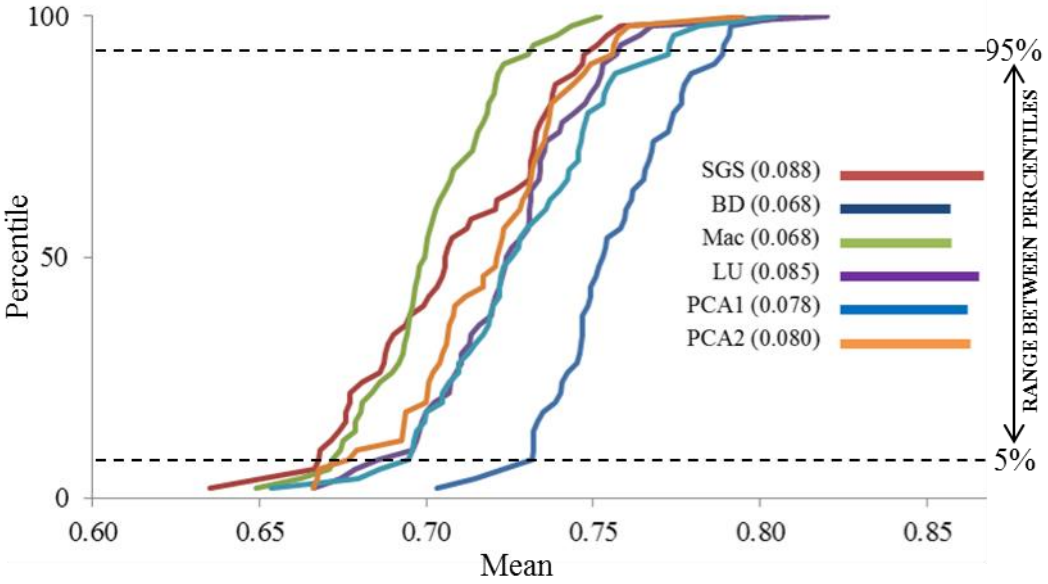


Figure 12: Cumulative distribution functions for all $K_{Sat}$ random fields generated by SGS and SGCS considering each secondary variable and the range between 5 and 95% percentiles.

The standard deviation (SD) maps represent the point variation in all simulated realizations; in other words, they represent the local uncertainty associated with predictions based on observed values and the sampling grid (Grunwald et al., 2007). Fig. 13 presents the standard deviation maps for the SGS and for the SGCS identified by the secondary variable used. In these maps we can see the areas of greatest and lowest uncertainty associated with $K_{Sat}$ in the watershed area.

SD was scaled in the same range for all processes in 0 to 0.3, 0.3 to 0.6, a.6 to 0.9 and above 0.9. In this way, we can compare the uncertainties located for each map and infer how each simulation process interfered in the occurrence of these uncertainties. In fig.13 it is possible to observe that for the SGS, the uncertainty distribution for the $K_{Sat}$ was homogeneous in the watershed area, with intermediate values of SD. This may have occurred due to the declustering method applied on the $K_{Sat}$ dataset, which removed the

clustering effect in the region of the watershed in which the sampling was densified. SGCS considering BD and Mac as secondary variables had similar behavior, presenting an extensive area of low uncertainty to the south and relatively higher values of SD in the northern part of the watershed for $K_{Sat}$ simulations. On the other hand SGCS for BD and Mac, when the secondary variable was LU, the lowest uncertainties are observed in the northern portion of the ECW and the highest in the southern portion. Still in relation to SGCS, the processes that used PC's as secondary variables showed different behaviors from each other, showing that each of the components represents a distinct natural phenomenon that is systematically related to $K_{Sat}$. Even though the associated uncertainty of $K_{Sat}$ presents different distribution patterns in the ECW area for each applied process, the global uncertainty is very close.

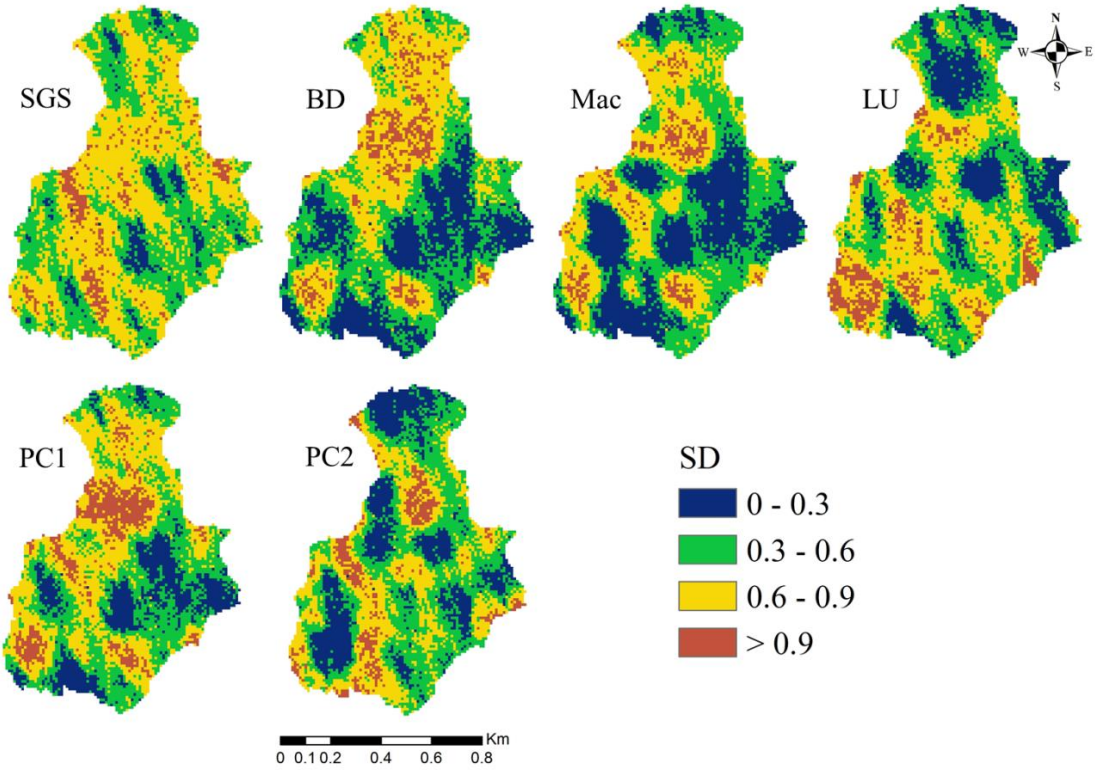


Figure 13: uncertainty maps associated with SGS and SGCS by secondary variable applied

It can therefore be said that the SGS was efficient to predict the $K_{Sat}$ in the ECW area, reproducing satisfactorily the stochastic parameters, such as histogram, mean and standard deviation and the spatial parameters, described

by the semivariogram. For this, no more than fifty realizations are required. The obtained maps show an expected pattern, where the highest values of $K_{Sat}$ are in the forest areas and close to the watercourses.

Among the unitary secondary variables, those selected for the GSCS are soil bulk density (BD), macroporosity (Mac) and land use (LU). Principal component analysis provided two more secondary variables for the GSCS process called PC1 and PC2, which together explain more than 79% of the $K_{Sat}$ variance in the watershed area. The first's two formed components showed high correlation and therefore were used as secondary variables.

For all SGCS, no more than fifty realizations were required. The stochastic parameters present in the histogram of the original $K_{Sat}$ data were reproduced satisfactorily for all processes, each with distinct secondary variables. Although all processes have slightly overestimated these parameters, this discrepancy is already expected in simulation results, due to ergodic fluctuations. The reproduction of semivariograms indicates that the GSCS processes had a spatial behavior very close to the target variable, concluding that the applied tool had an excellent performance.

The uncertainties of all processes, including SGS and SGCS, were evaluated using the cumulative density function. Although all processes have shown few uncertainties in the global distribution of $K_{Sat}$ in the ECW area, comparing the performances, SGS demonstrated a higher level of uncertainty. It is also concluded from the analysis of uncertainties that the SGCS that used Mac and BD as secondary variables performed better than the others. The standard deviation maps corroborate the conclusion that the SGCS that had BD and Mac as secondary variables presented superior performance in terms of uncertainty analysis.

## 4. Conclusions

It is concluded that stochastic simulation tools, mainly Gaussian sequential co-simulation using as secondary variable hydrological soil attributes, such as bulk density and soil macroporosity, are highly efficient to reproduce, predict and evaluate global and spatial uncertainties in small

watersheds, eg the Ellert Creck watershed, which presents highly heterogeneous environmental characteristics, a factor that promotes a great spatial variability of the saturated soil hydraulic conductivity, these methodologies being recommended for further studies and for use as a decision-making tool in watershed management.

**Acknowledgments**

The authors wish to thank the Brazilian National Council for Scientific and Technological Development (CNPq) for the financial support and scholarships provided, and the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES), Finance Code 001, for scholarships.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**REFERENCES:**

AHUJA, L. R.; MA, L.; GREEN, T. R. Effective Soil Properties of Heterogeneous Areas For Modeling Infiltration and Redistribution. **Soil Science Society of America Journal**, 2010. v. 74, n. 5, p. 1469–1482.

ALVAREZ, E. E. R.; PARRA-GONZÁLEZ, S. D.; RODRIGUEZ-VALENZUELA, J. 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**, 2019. n. October, p. 1–11.

BAGARELLO, V. *et al.* Determining short-term changes in the hydraulic properties of a sandy-loam soil by a three-run infiltration experiment. **Hydrological Sciences Journal**, 2020. v. 65, n. 7, p. 1191–1203.

BAIAMONTE, G. *et al.* Factors Influencing Point Measurement of Near-surface Saturated Soil Hydraulic Conductivity in a Small Sicilian Basin. **Land Degradation and Development**, 2017. v. 28, n. 3, p. 970–982.

BECKER, R.; GEBREMICHAEL, M.; MÄRKER, M. Impact of soil surface and subsurface properties on soil saturated hydraulic conductivity in the semi-arid Walnut Gulch Experimental Watershed, Arizona, USA. **Geoderma**, 2018. v. 322, n. October 2017, p. 112–120. Disponível em: <https://doi.org/10.1016/j.geoderma.2018.02.023>.

BESKOW, S. *et al.* Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil. **Hydrological Sciences Journal**, 2016. v. 61, n. 14, p. 2567–2578. Disponível em: <http://dx.doi.org/10.1080/02626667.2015.1133912>.

BISHOP, T. F. A. Cokriging particle size fractions of the soil. **European Journal of Soil Science**, 2007. v. 58, n. 3, p. 763–774.

BOURENNANE, H. *et al.* Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. **Ecological Modelling**, 2007. v. 205, n. 3–4, p. 323–335.

BRYS, G.; HUBERT, M.; STRUYF, A. A robust measure of skewness. **Journal of Computational and Graphical Statistics**, 2004. v. 13, n. 4, p. 996–1017.

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**, 2020. v. 191, n. March, p. 104583. Disponível em: <https://doi.org/10.1016/j.catena.2020.104583>.

DELBARI, M.; AFRASIAB, P.; LOISKANDL, W. Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. **Catena**, 2009. v. 79, n. 2, p. 163–169. Disponível em: <http://dx.doi.org/10.1016/j.catena.2009.08.001>.

DEUTSCH, C. DECLUS: a fortran 77 program for determining optimum spatial declustering weights. **Computers and Geosciences**, 1989. v. 15, n. 3, p. 325–332.

DONGLI, S. *et al.* Multi-scale correlations between soil hydraulic properties and associated factors along a Brazilian watershed transect. **Geoderma**, 2017. v. 286, p. 15–24. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2016.10.017>.

FEI, Y. *et al.* Micro-CT assessment on the soil structure and hydraulic characteristics of saline/sodic soils subjected to short-term amendment. **Soil and Tillage Research**, 2019. v. 193, n. April, p. 59–70. Disponível em: <https://doi.org/10.1016/j.still.2019.05.024>.

GAMIE, R.; SMEDT, F. DE. Experimental and statistical study of saturated hydraulic conductivity and relations with other soil properties of a desert soil. **European Journal of Soil Science**, 2018. v. 69, n. 2, p. 256–264.

GODOY, V. A.; ZUQUETTE, L. V.; GÓMEZ-HERNÁNDEZ, J. J. Spatial variability of hydraulic conductivity and solute transport parameters and their spatial correlations to soil properties. **Geoderma**, 2019. v. 339, n. December 2017, p. 59–69. Disponível em: <https://doi.org/10.1016/j.geoderma.2018.12.015>.

GOOTMAN, K. S.; KELLNER, E.; HUBBART, J. A. A comparison and validation of saturated hydraulic conductivity models. **Water (Switzerland)**, 2020. v. 12, n. 7, p. 1–19.

GOOVAERTS, P. Using elevation to aid the geostatistical mapping of rainfall erosivity. **Catena**, 1999. v. 34, n. 3–4, p. 227–242. . Geostatistical modelling of uncertainty in soil science. **Geoderma**, 2001. v. 103, n. 1–2, p. 3–26.

GRÄLER, B.; PEBESMA, E.; HEUVELINK, G. Spatio-temporal interpolation using gstat. **R Journal**, 2016. v. 8, n. 1, p. 204–218.

GRUNWALD, S. *et al.* Modeling of the spatial variability of biogeochemical soil properties in a freshwater ecosystem. **Ecological Modelling**, 2007. v. 201, n. 3–4, p. 521–535.

HAIR JR., J. F. *et al.* **Multivariate Data Analysis**. [S.l.]: [s.n.], 2009.

HOEFFDING, W.; ROBBINS, H. The central limit theorem for dependent random variables. **Duke Mathematical Journal**, 1948. v. 15, n. 3, p. 773–780.

JIN, Z.; LV, J. Integrated receptor models and multivariate geostatistical simulation for source apportionment of potentially toxic elements in soils. **Catena**, 2020. v. 194, n. April, p. 104638. Disponível em: <https://doi.org/10.1016/j.catena.2020.104638>.

JOHN, K. *et al.* Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistics. **International Journal of Environmental Science and Technology**, 2021. v. 18, n. 11, p. 3327–3342. Disponível em: <https://doi.org/10.1007/s13762-020-03089-x>.

KARACAN, C. Ö.; GOODMAN, G. V. R. Analyses of geological and hydrodynamic controls on methane emissions experienced in a Lower Kittanning coal mine. **International Journal of Coal Geology**, 2012. v. 98, p. 110–127. Disponível em: <http://dx.doi.org/10.1016/j.coal.2012.04.002>.

KUINCHTNER, A.; BURIOL, Ga. A. Clima Do Estado Do Rio Grande Do Sul Segundo a Classificação Climática De the Climate of Rio Grande Do Sul State According To Kóppen and Thornthwaite. **Disciplinarum Scientia. Série: Ciências Exatas**, 2001. v. 2, n. 1, p. 171–182. Disponível em: <http://sites.unifra.br/Portals/36/tecnologicas/2001/clima.pdf>.

LARK, R. M. Two robust estimators of the cross-variogram for multivariate geostatistical analysis of soil properties. **European Journal of Soil Science**, 2003. v. 54, n. 1, p. 187–202.

LIBOHOVA, Z. *et al.* Soil Systems for Upscaling Saturated Hydraulic Conductivity for Hydrological Modeling in the Critical Zone. **Vadose Zone Journal**, 2018. v. 17, n. 1, p. 170051.

MADENOGLU, S.; ATALAY, F.; ERPUL, G. Uncertainty assessment of soil erodibility by direct sequential Gaussian simulation (DSIM) in semiarid land uses. **Soil and Tillage Research**, 2020. v. 204, n. March, p. 104731. Disponível em: <https://doi.org/10.1016/j.still.2020.104731>.

MANCHUK, J. G.; DEUTSCH, C. V. Implementation aspects of sequential Gaussian simulation on irregular points. **Computational Geosciences**, 2012. v. 16, n. 3, p. 625–637.

MATHERON, G. *et al.* Condition simulation of the geometry of fluvio-deltaic reservoirs. 1987.

NAGANNA, S. R. *et al.* Factors influencing streambed hydraulic conductivity and their implications on stream–aquifer interaction: a conceptual review. **Environmental Science and Pollution Research**, 2017. v. 24, n. 32, p. 24765–24789.

NUSSBAUMER, R. *et al.* Which Path to Choose in Sequential Gaussian Simulation. **Mathematical Geosciences**, 2018. v. 50, n. 1, p. 97–120.

OLEA, R. A.; LUPPENS, J. A.; TEWALT, S. J. Methodology for quantifying uncertainty in coal assessments with an application to a Texas lignite deposit. **International Journal of Coal Geology**, 2011. v. 85, n. 1, p. 78–90. Disponível em: <http://dx.doi.org/10.1016/j.coal.2010.10.001>.

PACHEPSKY, Y.; ACOCK, B. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. **Geoderma**, 1998. v. 85, n. 2–3, p. 213–229.

PARK, Y. Saturated Hydraulic Conductivity of US Soils Grouped According to Textural Class and Bulk Density. **Soil Science Society of America Journal**, 2015. v. 79, n. 4, p. 1094–1100.

QI, Z. *et al.* Spatial distribution and simulation of soil moisture and salinity under mulched drip irrigation combined with tillage in an arid saline irrigation district, northwest China. **Agricultural Water Management**, 2018. v. 201, p. 219–231. Disponível em: <http://dx.doi.org/10.1016/j.agwat.2017.12.032>.

REMY, N.; BOUCHER, A.; WU, J. **Cambridge Books Online**. [S.l.]: [s.n.], 2010. V. 48.

SANTOS, R. C. V. DOS *et al.* Examining the implications of spatial variability of saturated soil hydraulic conductivity on direct surface runoff hydrographs. **Catena**, 2021. v. 207, n. August.

SHIWEN, Z. *et al.* Three-dimensional stochastic simulations of soil clay and its response to sampling density. **Computers and Electronics in Agriculture**, 2017. v. 142, p. 273–282. Disponível em: <https://doi.org/10.1016/j.compag.2017.08.031>.

SOARES, M. F. *et al.* Identifying Covariates to Assess the Spatial Variability of Saturated Soil Hydraulic Conductivity Using Robust Cokriging at the Watershed Scale. **Journal of Soil Science and Plant Nutrition**, 2020. v. 20, n. 3, p. 1491–1502.

TRINH, T. *et al.* Integrating global land-cover and soil datasets to update saturated hydraulic conductivity parameterization in hydrologic modeling. **Science of the Total Environment**, 2018. v. 631–632, p. 279–288.

YONG, Z. *et al.* Analysis of stability factors and interaction rules of soil slope under heavy rainfall. **IOP Conference Series: Earth and Environmental Science**, 2020. v. 546, n. 3.

WILDING L.P., DREES L.R.. Spatial variability and pedology. In: Wilding LP, Drees LR (eds) **Pedogenesis and soil taxonomy**: concepts and interactions. Elsevier, New York. 1983, pp. 83-116

ZHANG, Y., SCHAAP, M.G.. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. **J. Hydrol.** 575, 1011–1030. 2019. https://doi.org/10.1016/j.jhydrol.2019.05.058

ZIEGEL, E.R., DEUTSCH, C. V., Journel, A.G.. Geostatistical Software Library and User's Guide. **Technometrics** 40, 357. 1998.. https://doi.org/10.2307/1270548

## 6. Considerações Finais.

O presente trabalho foi dividido em dois artigos. A ligação destes dois artigos teve como elo a importância do uso de variáveis secundarias para uma melhor compreensão da distribuição espacial da condutividade hidráulica do solo saturado em escala de bacia hidrográfica.

O uso da análise de componentes principais para a composição de um banco de dados adimensional, formado por atributos hidrológicos do solo demonstrou ser uma alternativa eficiente para modelagem da $K_{Sat}$ a partir do uso de variáveis secundárias. O primeiro componente formado (PC1) é composto por atributos hidrológicos do solo que são sensíveis ao uso e ocupação do solo, principalmente ao impacto do manejo agrícola, como a densidade do solo, a macroporosidade e a porosidade total, contribuindo com o maior valor na variância acumulada. O segundo componente principal, que apresentou uma variância acumulada relativamente mais baixa que o primeiro, está ligado aos atributos hidrológicos do solo que possuem forte relação com a textura do solo: microporosidade do solo e carbono orgânico do solo.

O processo geoestatístico para os dois primeiros componentes gerados pela análise dos componentes principais foi muito eficiente, permitindo a construção de mapas por krigagem ordinária com alta precisão. A relação dos mapas dos componentes principais com os atributos não numéricos demonstra uma relação muito próxima com os dados de uso do solo e classe textural, permitindo-nos afirmar com segurança que estes são as componentes sistemáticas mais relevantes para a $K_{Sat}$ na área da bacia hidrográfica Sanga Ellert.

Conclui-se que as ferramentas de simulação estocástica, principalmente a co-simulação sequencial gaussiana usando como variáveis secundárias atributos hidrológicos do solo, como densidade do solo e macroporosidade do solo, são altamente eficientes para reproduzir, prever e avaliar as incertezas globais e espaciais em pequenas bacias hidrográficas, por exemplo, a bacia hidrográfica Sanga Ellert, que apresenta características ambientais altamente heterogêneas, fator que promove grande variabilidade espacial da condutividade hidráulica do solo saturado, sendo essas metodologias

recomendadas para estudos posteriores e para utilização como ferramenta de tomada de decisão no manejo de bacias hidrográficas.

A metodologia aqui apresentada, com o intuito de caracterizar a variabilidade espacial dos parâmetros hidrodinâmicos do solo, necessários ao dimensionamento do sistema de drenagem subterrânea, é recomendada como um modelo de procedimento em projetos de drenagem. Ela visa a um manejo mais adequado do sistema solo-água-planta, independentemente das dimensões da área estudada.

**REFERÊNCIAS:**

ARCHER, N.A.L.; BONELL, M.; COLES, N.; MACDONALD, A.M.; AUTON, C.A.; STEVENSON, R.. Soil characteristics and landcover relationships on soil hydraulic conductivity at a hillslope scale: a view towards local flood management. **Journal Of Hydrology**, [S.L.], v. 497, p. 208-222, ago. 2013. Elsevier BV. http://dx.doi.org/10.1016/j.jhydrol.2013.05.043.

BAGARELLO, V.; CECERE, N.; DAVID, S. M.; PRIMA, S. di. Determining short-term changes in the hydraulic properties of a sandy-loam soil by a three-run infiltration experiment. **Hydrological Sciences Journal**, [S.L.], v. 65, n. 7, p. 1191-1203, 12 mar. 2020. Informa UK Limited. http://dx.doi.org/10.1080/02626667.2020.1735637.

BECKER, R.; GEBREMICHAEL, M.; MÄRKER, M.. Impact of soil surface and subsurface properties on soil saturated hydraulic conductivity in the semi-arid Walnut Gulch Experimental Watershed, Arizona, USA. **Geoderma**, [S.L.], v. 322, p. 112-120, jul. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2018.02.023.

BERKHOUT, Ezra D.; MALAN, Mandy; KRAM, Tom. Better soils for healthier lives? An econometric assessment of the link between soil nutrients and malnutrition in Sub-Saharan Africa. **Plos One**, [S.L.], v. 14, n. 1, p. 210-220, 17 jan. 2019. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pone.0210642.

COBURN, Timothy C. Statistical and Econometric Methods for Transportation Data Analysis. **Technometrics**, [S.L.], v. 46, n. 4, p. 492-493, nov. 2004. Informa UK Limited. http://dx.doi.org/10.1198/tech.2004.s238.

CONOVER WJ, IMAN RL**. Rank transformations as a bridge between parametric and nonparametric statistics**. 35: 124-32. Am Stat 1981

GOOTMAN, Kaylyn; KELLNER, Elliott; HUBBART, Jason. A Comparison and Validation of Saturated Hydraulic Conductivity Models. **Water**, [S.L.], v. 12, n. 7, p. 2040, 18 jul. 2020. MDPI AG. http://dx.doi.org/10.3390/w12072040.

HASTIE, M. **Assessment of Samples from Crathes, Warren Field**. Unpublished report. Headland Archaeology. Jul. 2004.

HILLEL, D. **Introduction to environmental soil physics**. 2003. Elsevier.

HU, Kelin; WHITE, Robert; CHEN, Deli; LI, Baoguo; LI, Weidong. Stochastic simulation of water drainage at the field scale and its application to irrigation management. **Agricultural Water Management**, [S.L.], v. 89, n. 1-2, p. 123-130, abr. 2007. Elsevier BV. http://dx.doi.org/10.1016/j.agwat.2006.12.010.

KRAMER MS. **Clinical epidemiology and biostatistics**. Berlin: Springer-Verlag, 1988.

KURNIANTO, Sofyan; SELKER, John; KAUFFMAN, J. Boone; MURDIYARSO, Daniel; PETERSON, James T.. The influence of land-cover changes on the variability of saturated hydraulic conductivity in tropical peatlands. **Mitigation And Adaptation Strategies For Global Change**, [S.L.], v. 24, n. 4, p. 535-555, 19 mar. 2018. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11027-018-9802-3.

LIU, Zhipeng; MA, Donghao; HU, Wei; LI, Xuelin. Land use dependent variation of soil water infiltration characteristics and their scale-specific controls. **Soil And Tillage Research**, [S.L.], v. 178, p. 139-149, maio 2018. Elsevier BV. http://dx.doi.org/10.1016/j.still.2018.01.001.

MIZUTA, Katsutoshi; GRUNWALD, Sabine; PHILLIPS, Michelle A.; BACON, Allan R.; CROPPER, Wendell P.; MOSS, Charles B.. Emergence of the Pedo-Econometric Approach. **Frontiers In Soil Science**, [S.L.], v. 1, p. 327-345, 9 abr. 2021. Frontiers Media SA. http://dx.doi.org/10.3389/fsoil.2021.656591.

NAGANNA, Sujay Raghavendra; DEKA, Paresh Chandra; CH, Sudheer; HANSEN, William F.. Factors influencing streambed hydraulic conductivity and their implications on stream–aquifer interaction: a conceptual review. **Environmental Science And Pollution Research**, [S.L.], v. 24, n. 32, p. 24765-24789, 7 out. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s11356-017-0393-4.

NADKARNI, P. M.; MARENCO, L.; CHEN, R.; SKOUFOS, E.; SHEPHERD, G.; MILLER, P.. Organization of Heterogeneous Scientific Data Using the EAV/CR Representation. **Journal Of The American Medical Informatics Association**, [S.L.], v. 6, n. 6, p. 478-493, 1 nov. 1999. Oxford University Press (OUP). http://dx.doi.org/10.1136/jamia.1999.0060478.

REICHARDT K, TIMM LC. **Soil, plant and atmosphere:** concepts, processes and applications. Springer, Basel. 2020

SANTOS, Rodrigo César Vasconcelos dos; VARGAS, Marcelle Martins; TIMM, Luís Carlos; BESKOW, Samuel; SIQUEIRA, Tirzah Moreira; MELLO, Carlos Rogério; SOARES, Mauricio Fornalski; MOURA, Maíra Martim de; REICHARDT, Klaus. Examining the implications of spatial variability of saturated soil hydraulic conductivity on direct surface runoff hydrographs. **Catena**, [S.L.], v. 207, p. 105693, dez. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.catena.2021.105693.

SIQUEIRA, Tirzah M.; LOUZADA, José A. S.; PEDROLLO, Olavo C.; CASTRO, Nilza M. dos R.; OLIVEIRA, Marquis H. C. de. Soil physical and hydraulic properties in the Donato stream basin, RS, Brazil. Part 2: geostatistical simulation. **Revista Brasileira de Engenharia Agrícola e Ambiental**, [S.L.], v. 23, n. 9, p. 675-680, set. 2019. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/1807-1929/agriambi.v23n9p675-680.

SHI, Yinguang; ZHAO, Xining; GAO, Xiaodong; ZHANG, Shulan; WU, Pute. The Effects of Long-term Fertiliser Applications on Soil Organic Carbon and Hydraulic Properties of a Loess Soil in China. **Land Degradation & Development**, [S.L.], v. 27, n. 1, p. 60-67, 15 maio 2015. Wiley. http://dx.doi.org/10.1002/ldr.2391.

SOARES, Mauricio Fornalski; CENTENO, Luana Nunes; TIMM, Luís Carlos; MELLO, Carlos Rogério; KAISER, Douglas Rodrigo; BESKOW, Samuel. Identifying Covariates to Assess the Spatial Variability of Saturated Soil Hydraulic Conductivity Using Robust Cokriging at the Watershed Scale. **Journal Of Soil Science And Plant Nutrition**, [S.L.], v. 20, n. 3, p. 1491-1502, 27 abr. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s42729-020-00228-8.

URSO, Alfonso; FIANNACA, Antonino; LAROSA, Massimo; RAVÌ, Valentina; RIZZO, Riccardo. Data Mining: prediction methods. **Encyclopedia Of Bioinformatics And Computational Biology**, [S.L.], p. 413-430, 2019. Elsevier. http://dx.doi.org/10.1016/b978-0-12-809633-8.20462-7.

VARIAN, Hal R.. Big Data: new tricks for econometrics. **Journal Of Economic Perspectives**, [S.L.], v. 28, n. 2, p. 3-28, 1 maio 2014. American Economic Association. http://dx.doi.org/10.1257/jep.28.2.3.

WINCKLER, J.; REICK, C. H.; BRIGHT, R. M.; PONGRATZ, J.. Importance of Surface Roughness for the Local Biogeophysical Effects of Deforestation. **Journal Of Geophysical Research**: Atmospheres, [S.L.], v. 124, n. 15, p. 8605-8618, 14 ago. 2019. American Geophysical Union (AGU). http://dx.doi.org/10.1029/2018jd030127.

ZHANG, Qingyin; SHAO, Mingan; JIA, Xiaoxu; WEI, Xiaorong. Changes in soil physical and chemical properties after short drought stress in semi-humid forests. **Geoderma**, [S.L.], v. 338, p. 170-177, mar. 2019. Elsevier BV. http://dx.doi.org/10.1016/j.geoderma.2018.11.051.