UNIVERSIDADE FEDERAL DO OESTE DO PARÁ

INSTITUTO DE BIODIVERSIDADE E FLORESTAS

PROGRAMA DE PÓS-GRADUAÇÃO EM SOCIEDADE, NATUREZA E DESENVOLVIMENTO

CAUAN FERREIRA ARAÚJO

MAPEAMENTO DE SOLOS COM ÊNFASE EM ANÁLISES MULTIESCALARES PARA APLICAÇÃO NA AGRICULTURA FAMILIAR

Santarém, Pará

2022

CAUAN FERREIRA ARAÚJO

MAPEAMENTO DE SOLOS COM ÊNFASE EM ANÁLISES MULTIESCALARES PARA APLICAÇÃO NA AGRICULTURA FAMILIAR

Tese submetida ao Programa de Pós-graduação em Sociedade, Natureza e Desenvolvimento da Universidade Federal do Oeste do Pará para obter o título de Doutor em Sociedade, Natureza e Desenvolvimento, Área de concentração: Impactos ambientais e sociais da mudança do uso da terra na Amazônia.

Orientador: Troy Patrick Beldini

Coorientador: Raimundo Cosme de Oliveira Jr.

Santarém, Pará

2022

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema Integrado de Bibliotecas – SIBI/UFOPA

A663m Araújo, Cauan Ferreira

Mapeamento de solos com ênfase em análises multiescalares para aplicação na agricultura familiar./ Cauan Ferreira Araújo . – Santarém, 2022.

85 p. : il.

Inclui bibliografias.

Orientador: Troy Patrick Beldini.

Coorientador: Raimundo Cosme de Oliveira Jr.

Tese (Doutorado) – Universidade Federal do Oeste do Pará, Instituto de Biodiversidade e Florestas, Programa de Pós-Graduação em Sociedade, Natureza e Desenvolvimento.

1. Mapeamento digital – Solo. 2. Geomorfometria. 3. Pedometria. 4. Escala. I. Beldini, Troy Patrick, *orient.*. II. Oliveira Jr., Raimundo Cosme de, *coorient*. III. Título.

CDD: 23 ed. 596.09811

Bibliotecária - Documentalista: Cátia Alvarez – CRB/2 843

À minha companheira,

Naiana Marinho de Souza,

e nossos filhos Benício e Hugo,

por trilharmos juntos os dias mais felizes de nossas vidas.

AGRADECIMENTOS

Agradeço ao suporte e ao apoio da minha família, em especial meus pais e minha companheira.

Agradeço aos meus orientadores, Troy Beldini e Raimundo Cosme, pelas valorosas contribuições, pela confiança depositada e pela compreensão quanto aos percalços.

Agradeço a Ufopa, por me possibilitar cursar este Doutorado e pelas inúmeras oportunidades de aprendizagem e crescimento profissional.

Aos colegas de instituição, que colaboraram direta ou indiretamente para a execução desta Tese. Em especial, aos meus colegas do Campus de Oriximiná.

Aos meus orientandos, discentes da Ufopa e de iniciação científica do ensino médio da EETEPA, pelas contribuições e pela oportunidade de aprender ensinando.

Agradeço, *in memoriam*, ao professor Domingos Diniz, por inspirações e ensinamentos que apenas alguém com sua distinta mentalidade visionária e professoral poderia prover.

RESUMO

A informação sobre os solos é fundamental para a intensificação sustentável dos sistemas produtivos familiares, sendo recomendável o mapeamento de classes de solos até o nível dos subgrupos, além do mapeamento das classes texturais. No contexto amazônico, as informações dos solos necessitam estar em escala compatível com pequenas propriedades rurais, apesar da grande extensão da região, para que possam ser úteis para agricultores familiares e empresas de assistência técnica e extensão rural. Para tal, faz-se necessário a adoção de estratégias multiescalares para o mapeamento digital de solos. No presente estudo, o mapeamento de classes texturais foi realizado na escala 1:25.000 e o mapeamento de classes de solo na escala 1:100.000. Por outro lado, a influência multiescalar da topografia na distribuição do solo tem um padrão complexo que está relacionado à sobreposição de processos pedológicos que ocorreram em diferentes épocas e forças motrizes que estão correlacionadas com diversas escalas. Nesse sentido, testou-se a hipótese de que covariáveis geomorfométricas generalizadas em múltiplas escalas podem melhorar a acurácia da modelagem pedométrica. No mapeamento de classes texturais, aplicou-se o algoritmo *Random Forest* a um banco de dados geomorfométrico multiescalar para prever os percentuais granulométricos na superfície do solo. As covariáveis generalizadas melhoraram a acurácia da classificação textural, com o Índice Kappa passando de 0,43 para 0,62. No mapeamento de classes de solo, modelos de conjuntos fuzzy foram aplicados usando covariáveis generalizadas para a definição de ambientes preferenciais para a ocorrência dos processos pedogenéticos de gleização, elutriação e translocação de argila, em oposição à ampla distribuição espaço-temporal do processo de ferralização. As regras de pertinência fuzzy e as escalas das covariáveis foram definidas para cada um dos processos, considerando o conhecimento pedológico obtido a partir de levantamentos de solo em transectos. As covariáveis generalizadas melhoraram a acurácia do mapeamento, com o Índice Kappa passando de 0,4 para 0,69. Os resultados demonstram que modelos baseados em conhecimento que observam a influência multiescalar da topografia nos processos de formação do solo podem melhorar a previsão de classes de solos em levantamentos de solo de nível intermediário. Logo, conclui-se que o uso de covariáveis geomorfométricas generalizadas em múltiplas escalas resultam em maior acurácia aos modelos, sendo um método flexível para utilização tanto em abordagens de aprendizagem de máquina quanto em modelagem baseada no conhecimento do pedólogo. Ainda, do ponto de vista geomorfométrico, faz-se necessário o teste de métodos de generalização, com foco na preservação de feições relevantes para a relação solo-paisagem.

Palavras chave: mapeamento digital de solos, geomorfometria, pedometria, escala.

ABSTRACT

Soil information is essential for the sustainable intensification of family production systems, and it is recommended that soil classes be mapped down to the subgroup level, in addition to the mapping of textural classes. In the Amazon context, soil information needs to be on a scale compatible with small rural properties, despite the large extension of the region, so that it can be useful for family farmers and technical assistance. To this end, it is necessary to adopt multiscale strategies for digital soil mapping. In the present study, the mapping of textural classes was performed on a scale of 1:25,000 and the mapping of soil classes on a scale of 1:100,000. On the other hand, the multiscale influence of topography on soil distribution has a complex pattern that is related to the overlapping of pedological processes that occurred at different times and driving forces that are correlated with different scales. In this sense, we tested the hypothesis that generalized geomorphometric covariates at multiple scales can improve the accuracy of pedometric modeling. In the mapping of textural classes, the Random Forest algorithm was applied in a multiscale geomorphometric database to predict the granulometric percentages on the soil surface. The generalized covariates improved the accuracy of the textural classification, with the Kappa Index increasing from 0.43 to 0.62. In the mapping of soil classes, fuzzy set models were applied using generalized covariates to define preferential environments for the occurrence of the pedogenetic processes of gleization, elutriation and clay translocation, as opposed to the wide spatio-temporal distribution of the ferralization process. Fuzzy membership rules and covariate scales were defined for each of the processes, considering the soil knowledge obtained from soil surveys in transects. The generalized covariates improved the accuracy of the mapping, with the Kappa Index going from 0.4 to 0.69. The results demonstrate that knowledge-based models that observe the multiscale influence of topography on soil formation processes can improve the prediction of soil classes in intermediate-level soil surveys. Therefore, it is concluded that the use of generalized geomorphometric covariates in multiple scales results in greater accuracy of the models, being a flexible method for use both in machine learning approaches and in modeling based on the pedologist's knowledge. Still, from the geomorphometric point of view, it is necessary to test generalization methods, focusing on the preservation of features relevant to the soil-landscape relationship.

Keyword: digital soil mapping, geomorphometry, pedometry, scale.

SUMÁRIO

‘

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

1. INTRODUÇÃO

Os solos tem importância crucial no funcionamento dos ecossistemas (ADHIKARI; HARTEMINK, 2016). Nesse sentido, alguns dos mais importantes desafios globais para o século XXI têm nos solos um elemento central, tais como as mudanças climáticas e a produção de alimentos e matérias primas para uma sociedade em crescimento. Logo, a segurança dos solos é crucial para as políticas ambientalistas no contexto atual.

O conceito de segurança dos solos abrange os aspectos edáficos relacionados à segurança alimentar, à segurança energética, aos serviços ecossistêmicos, à proteção da biodiversidade, à mitigação das mudanças climáticas e à segurança dos recursos hídricos (MCBRATNEY et al., 2017; MCBRATNEY; FIELD; KOCH, 2014). Ainda, o conceito de segurança dos solos apresenta cinco dimensões, que devem ser abordadas por cientistas e gestores para alcançar as metas de desenvolvimento sustentável, a saber: (a) a capacidade, relacionada ao potencial dos solos para exercer determinada função; (b) a condição, às mudanças do estado atual em relação a um estado de referência; (c) capital, relativo à valoração do solo e dos seus serviços; (d) conectividade, relacionada à dimensão social entre o solo e o tomador de decisão; (e) codificação, referente ao arcabouço legal (KOCH, 2017; MONTANARELLA, 2017).

Na Amazônia, as consequências ambientais do uso da terra em inobservância à segurança dos solos vão da perda de habitats para a fauna, redução da biodiversidade florestal, erosão e queda da fertilidade dos solos, assoreamento de igarapés, perda de nascentes e da qualidade dos recursos hídricos de maneira geral (BRONDÍZIO et al., 2009; FU et al., 2015). Adicionalmente, um dos efeitos nefastos da perda da fertilidade dos solos e degradação de áreas agropecuárias é a pressão econômica que resulta na invasão e desflorestamento de novas áreas, em um ciclo vicioso, degradando os recursos naturais, prejudicando os serviços ambientais e intensificando as pressões econômicas sobre o agricultor (RIBEIRO-FILHO; ADAMS; SERENI-MURRIETA, 2013).

No Brasil, a conservação e a proteção dos solos constam na legislação ambiental, na maior parte relacionada à regulamentação de atividades agropecuárias (TORNQUIST; BROETTO, 2017). Ainda, apenas três estados possuem leis suplementares relativas à proteção dos solos: Espirito Santo (6607/2001), São Paulo (6171/1988), Paraná (8014/1984). O desenvolvimento de políticas públicas, tanto regulamentações para a proteção, quanto subsídios em prol do uso sustentável, depende do conhecimento aprofundado dos solos (MONTANARELLA, 2017). Por outro lado, a falta de informação e de atenção para com esse recurso ocasiona processos de degradação das terras, potencialização de desastres naturais, emissão de gases de efeito estufa, entre outros.

Atualmente, o Brasil possui a totalidade de suas áreas mapeadas apenas em níveis exploratórios ou de reconhecimento de baixa intensidade (SANTOS et al., 2013). Os reconhecimentos de alta intensidade cobrem apenas 4,47% do território, e os níveis com maior detalhamento apenas 1,31%. Em face da necessidade urgente de ampliação do conhecimento sobre os recursos de solo do Brasil, foi formado um grupo de trabalho interinstitucional que delineou um planejamento com propostas de curto, médio e longo prazo (POLIDORO et al., 2016). O Programa Nacional de Solos (PronaSolos) objetiva realizar mapeamentos em escala 1:100.000, 1:50.000 e 1:25.000 em um período de 30 anos.

No contexto da agricultura familiar, a despeito da importância da temática, as políticas de acesso às informações sobre o solo são praticamente inexistentes (DELGADO; BERGAMASCO, 2017). Não havendo tecnologias para aquisição e comunicação de informações edáficas na escala necessária, com baixo custo ou ofertada gratuitamente por empresas públicas de assistência técnica e extensão rural, o agricultor familiar permanece à margem das políticas para o uso sustentável dos solos (BUCKNER, 2017). Nesse sentido, é preciso incentivar o desenvolvimento de soluções aplicáveis ao contexto dos agricultores familiares no mapeamento digital de solos. Os principais elementos funcionais de tais práticas são a aquisição de informações com baixo custo e a transferência de informações aos donos da terra (BARRIOS; TREJO, 2003; GALLAHER; WINKLERPRINS, 2016; ROSSITER et al., 2015).

No contexto amazônico, a necessidade de informações dos solos por pequenos produtores rurais e por empresas de assistência técnica e extensão rural apresentam as seguintes características gerais: (a) escala compatível com pequenas áreas; (b) precisão locacional e informacional intermediária; (c) adequação à sistemas de produção diversificados, convencionais ou agroecológicos; (d) disponibilidade prévia e sob demanda. Para atender tais características, faz-se necessário a adoção de estratégias multiescalares no levantamento, análise e representação dos mapas de solos.

O mapeamento multiescalar de solos está fundamentado na utilização de informações sobre o solo e a paisagem em diversos níveis de detalhamento, em um percurso metodológico que permita a análise consistente dessas bases de dado em conjunto, resultando em produtos nas escalas adequadas para os objetivos da avaliação (MALONE; MCBRATNEY; MINASNY, 2013).

As questões de escala das relações solo-paisagem estão relacionadas às interações complexas de ambos os elementos e como esses processos ocorrem e são percebidos. Estudos de geomorfologia relatam um acoplamento espaço-tempo entre escalas de tamanho do relevo e tempo de vida (SCHMIDT; ANDREW, 2005; WYSOCKI; SCHOENEBERGER, 2011). Do ponto de vista do solo, diferentes processos pedológicos manifestarão influência em curto, médio ou longo prazo (KÄMPF; CURI, 2012; TARGULIAN; KRASILNIKOV, 2007). Em contraste, estudos de topografia em física do solo demonstram uma complexa dinâmica da água relacionada à geometria aninhada das encostas, considerando padrões de relevo e microrrelevo, resultando em tendências no movimento de partículas e solutos e mudanças na textura e parâmetros químicos dos solos (FLORINSKY, 2016; HU; LI; YANG, 2020). Portanto, a topografia multiescalar influencia uma determinada distribuição do solo em dois aspectos gerais, a saber: sobreposição de processos pedológicos ocorridos em épocas diferentes; e forças motrizes, determinadas pela soma de forças melhor correlacionadas com diversas escalas geomorfológicas.

No presente estudo, o mapeamento de classes texturais foi realizado na escala 1:25.000 e o mapeamento de classes de solo na escala 1:100.000. Nesse sentido, testou-se a hipótese de que covariáveis geomorfométricas generalizadas em múltiplas escalas podem melhorar a acurácia da modelagem pedométrica.

2. ESCALA NO CONTEXTO DO MAPEAMENTO DIGITAL DE SOLOS

A presente seção é uma revisão de literatura que se destina a expor os fundamentos do mapeamento digital de solos e aspectos do conceito de escala relevantes para aplicação na ciência pedométrica. A revisão está dividida em quatro partes, a saber: (a) *Fundamentos do mapeamento digital de solos*, onde é apresentado o contexto geral desse campo do conhecimento; (b) *Métodos pedométricos*, onde consta uma revisão, não exaustiva, das técnicas e aplicações no contexto dos mapeamentos; (c) *Conceitos de escala*, que trata das dimensões, elementos e problemas deste fundamento da ciência geográfica; (d) *Escala no mapeamento digital de solos*, descrevendo métodos aplicados à transformações de escala e à análise multiescalar.

2.1. Fundamentos do mapeamento digital de solos

A prática do mapeamento de solos está inscrita no contexto dos paradigmas pedológicos, tecnologias geográficas e propósitos de mapeamento (MILLER; SCHAETZL, 2016). A história do mapeamento digital de solos é marcada pela adoção de novas ferramentas e técnicas de mapeamento, tais como: sistemas de gerenciamento de dados e métodos para analisar, integrar e visualizar conjuntos de dados do solo e do ambiente (GRUNWALD, 2009). Logo, a grande quantidade de dados gerados por tecnologias eletrônicas impulsionou o desenvolvimento de análises quantitativas dos solos, ou pedometria (BREVIK et al., 2016a; HARTEMINK; MCBRATNEY; CATTLE, 2001). As principais vantagens desta abordagem estão relacionadas à formalização dos modelos, permitindo a avaliação de incertezas, além da possibilidade de incremento das bases de dados, resultando na atualização dos mapeamentos.

No inicio do século XXI ocorreu o 'renascimento da ciência dos solos', principalmente em razão das funções ecológicas do solo relacionadas à temáticas globais de suma importância, tal como as mudanças climáticas e a disponibilidade de alimentos e matérias primas para uma sociedade em franco crescimento demográfico (HARTEMINK; MCBRATNEY, 2008; SANCHEZ et al., 2009). Fruto do aumento de investimentos na área, os programas de pesquisa em ciências do solo aceleraram o desenvolvendo de métodos para aquisição, tratamento e comunicação de informações (BURROUGH, 2007).

Podemos citar alguns dos novos desafios no mapeamento digital de solos (LAGACHERIE; MCBRATNEY, 2006; MINASNY; MCBRATNEY, 2016): trabalhar com grandes áreas de cobertura; lidar com a variação do solo em curta distancia; alcançar as demandas dos usuários; lidar com bases de dados em evolução. Nesse contexto, ainda são necessárias metodologias para a comunicação da informação do solo aos tomadores de decisão, incorporando informações do conhecimento local em estratégias interdisciplinares para o mapeamento de solos (BREVIK et al., 2016b).

O sensoriamento remoto, próximo ou orbital, tem grande aplicação na aquisição de dados para mapeamento digital de solos. Quanto ao sensoriamento remoto próximo, a morfometria digital do solo é definida como a aplicação de ferramentas e técnicas para medir e quantificar atributos do perfil do solo, geralmente por meio de espectrômetros portáteis (HARTEMINK; MINASNY, 2014; SILVA et al., 2016b). Uma alternativa de baixo custo em desenvolvimento é a possibilidade de morfometria digital com o uso de *smartphones*, principalmente colorimetria, apresentando boa capacidade de predição de atributos associados a este parâmetro (AITKENHEAD et al., 2016). Quanto ao sensoriamento remoto orbital, este vem sendo utilizado com sucesso na definição de unidades de mapeamento e na predição de propriedades do solo por meio de análises multiespectrais (ALVES; DEMATTÊ; BARROS, 2015; GRUNWALD; VASQUES; RIVERO, 2015).

No Brasil, a abordagem digital nos mapeamentos de solo, inicialmente incipiente, também vem ganhando importância (CATEN et al., 2012; DALMOLIN; CATEN, 2015). As características econômicas, a infraestrutura de transportes precária em algumas regiões, as dimensões continentais do país, com mapeamentos realizados apenas em escalas muito generalistas na maior parte do território (SANTOS et al., 2013), imprimem necessidades metodológicas no sentido da redução dos custos operacionais em atividades de levantamento dos solos.

2.2. Métodos pedométricos

O campo da pedometria está inserido em um contexto interdisciplinar, na interseção entre a ciência do solo, a estatística e os sistemas de informação geográfica, e por isso são possíveis diversos arranjos metodológicos (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003). A seguir, serão apresentadas algumas das metodologias atualmente empregadas no mapeamento digital de solos, organizadas segundo suas finalidades, a saber: (a) distribuição espacial de amostragens; (b) classificação geomorfométrica; (c) interpolação geoestatística; (d) técnicas de mapeamento por lógica *Fuzzy*; (e) métodos de aprendizagem de máquina.

A distribuição espacial dos pontos amostrais tem papel fundamental no mapeamento digital de solos, podendo inclusive alterar sobremaneira os resultados obtidos, e por isso deve ser considerada cuidadosamente (GESSLER et al., 1995). Via de regra, métodos de interpolação obtém melhores resultados utilizando grades regulares, ao menos quando não existem dados auxiliares disponíveis (HEUVELINK; BRUS; DE GRUIJTER, 2006). Por outro lado, amostragens em topossequências e/ou unidades significativas da paisagem podem ser realizadas quando se pretende utilizar abordagens pedométricas baseadas no conhecimento (SHI et al., 2009), tais como a modelagem por conjuntos *Fuzzy*, exigindo uma amostragem menos densa.

Nas etapas de validação dos mapeamentos é preciso adotar um esquema de amostragem não enviesado. Quando não são consideradas variáveis ambientais auxiliares, devido à indisponibilidade, incerteza quanto à qualidade ou quanto à correlação com os atributos do solo, o melhor a fazer é dispor os pontos no espaço, por exemplo, por meio de algoritmos de médias-K (BRUS; GRUIJTER; GROENIGEN, 2007). O hipercubo latino condicionado pode ser utilizado quando variáveis ambientais são consideradas, resultando em uma distribuição eficaz quanto à representação dos intervalos dos dados auxiliares (MINASNY; MCBRATNEY, 2006; MULDER; DE BRUIN; SCHAEPMAN, 2012). A abordagem por hipercubo latino condicionado limitado pelo custo, que utiliza uma camada extra representando os custos de acesso, apesar de produzir resultados um pouco inferiores, pode ser utilizada em um contexto de mapeamentos de baixo custo (SILVA et al., 2014, 2015). Nesses estudos, para a elaboração da camada de custos de acesso foram consideradas as variáveis distância da estrada, declividade e vegetação, como ilustrado na figura 2.1.

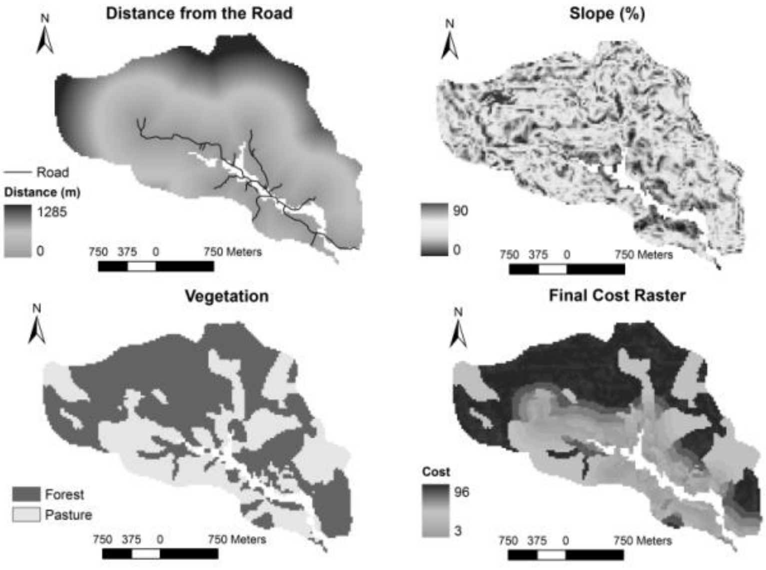


Figura 2.1. Variáveis utilizadas para a criação da camada de custo de acesso. Fonte: SILVA *et al*., 2014.

A geomorfometria compreende um conjunto de análises quantitativas da superfície do terreno, segundo duas abordagens (ZINCK, 2016a): a geomorfometria específica, que analisa as características de elementos discretos; e a geomorfometria geral, que trata das características de elementos contínuos. Atualmente a abordagem da geomorfometria geral tem recebido maior atenção, devido aos modelos digitais de elevação (MDE) serem superfícies continuas. Nessa linha de abordagem podemos distinguir ainda dois conjuntos de metodologias: a análise de parâmetros da superfície terrestre e os classificadores de relevo.

Os modelos digitais de elevação são matrizes georreferenciadas, cujas entradas representam dados altimétricos, podendo apresentar uma diversidade de fontes (NELSON;

REUTER; GESSLER, 2009). Os modelos digitais de elevação produzidos a partir do projeto "*Shuttle Radar Topography Mission*" são amplamente utilizados no mapeamento digital de solos por todo o mundo. No contexto brasileiro, os modelos digitais de elevação SRTM foram refinados e disponibilizados na base Topodata (VALERIANO; ROSSETTI, 2012).

Os MDE devem ser preparados antes da realização de análises geomorfométricas por apresentarem diferentes tipos de erros (REUTER et al., 2009), a saber: devido à presença de artefatos; enviesamento de sensores ou de métodos analíticos; ruídos ou erros ao acaso. Os MDE SRTM são sensíveis à presença de artefatos na superfície, especialmente no caso de vegetações densas e/ou de grande porte, afetando os resultados geomorfométricos principalmente em áreas de borda entre diferentes coberturas do solo. Nesse sentido, foram realizados estudos para minimizar o efeito do desflorestamento em MDEs da região amazônica por meio de diversos métodos de correção (BROCHADO, 2015), obtendo resultados satisfatórios.

Os parâmetros da superfície são organizados em parâmetros primários e secundários Wilson (WILSON, 2012). Os parâmetros primários, ou parâmetros básicos, da superfície terrestre são derivados diretamente dos DEMs, sem entradas adicionais. Por outro lado, existem dois conjuntos de parâmetros secundários: os parâmetros hidrológicos da superfície terrestre para quantificar o fluxo de água e processos de superfície relacionados, incluindo processos pedológicos; e modelos de radiação solar e métodos para quantificar as interações entre a superfície terrestre e a atmosfera.

Os classificadores de relevo abarcam técnicas de análise estatística que permitem delimitar unidades de relevo a partir de superfícies continuas, podendo ou não classifica-las segundo a taxonomia geomorfológica ou aproximações desta (ZINCK, 2016b). Entre os métodos que não resultam na classificação em *taxons* podemos citar a classificação automatizada de topografia por algoritmo de meios aninhados não supervisionados (IWAHASHI; PIKE, 2007). Por outro lado, podemos citar o método dos *geomorphons* (JASIEWICZ; STEPINSKI, 2013), que utiliza conceitos da visão computacional e da análise de padrões para delimitar e classificar o relevo em uma classificação com as 10 formas mais comuns de relevo, ilustrados na figura 2.2, semelhantes aos *taxa* geomorfológicos. As principais vantagens desse método são o custo computacional reduzido, a aplicabilidade em diversas escalas e a facilidade de interpretação dos resultados. Logo, diversos estudos tem aplicado essa técnica na segmentação da paisagem para mapeamentos de solo (ASHTEKAR et al., 2014; PINHEIRO et al., 2016; SILVA et al., 2016a).

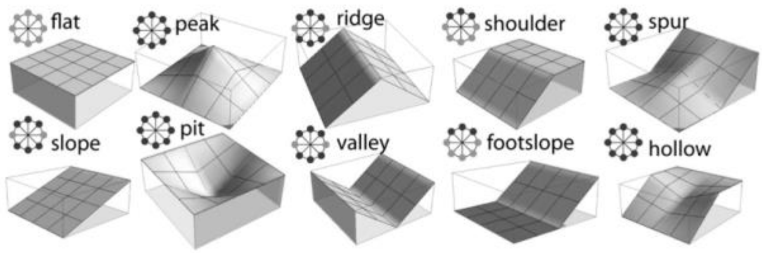


Figura 2.1. Morfologias 3D simbólicas e seus correspondentes em geomorphons. Fonte: JASIEWICZ; STEPINSKI, 2013

A interpolação geoestatística tem por objetivo a caracterização de uma variável regionalizada de interesse por meio do estudo de sua distribuição e variabilidade espacial, com determinação das incertezas associadas (GOOVAERTS, 1999). A inferência de uma variável regionalizada é representada por meio do semivariograma, figura 2.3, onde são expressos à aleatoriedade associada e a estrutura de dependência do fenômeno (YAMAMOTO; LANDIM, 2015).

Os métodos geoestatísticos, ou Krigagem, vem sendo utilizados no mapeamento digital de solos com sucesso, apesar de geralmente necessitarem de uma densa malha amostral para obter resultados ótimos. Entre as variações da Krigagem com aplicações na ciência dos solos (GRUNWALD; RIVERO; REDDY, 2007), podemos citar: krigagem simples; krigagem ordinária; krigagem universal (HEUVELINK; BRUS; DE GRUIJTER, 2006); krigagem indicativa (OBERTHÜR, 1999); *regression-kriging* (HENGL; HEUVELINK; STEIN, 2004). Adicionalmente, é possível integrar informações de mapas coropléticos com dados pontuais em interpolação geoestatística, por meio da *area-to-point kriging* (GOOVAERTS, 2011). Nesse sentido, tal método fornece uma estrutura para modelar a correlação espacial entre as propriedades do solo medidas sobre suportes geográficos irregulares, permitindo o mapeamento da distribuição de valores de atributos dentro de cada unidade de solo.

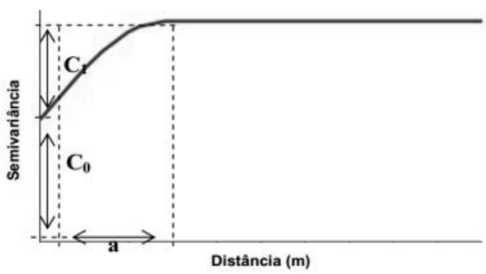


Figura 2.2. Semivariograma teórico, composto por: (C_0) efeito pepita, aleatoriedade à distância zero; (C_1) estrutura de dependência espacial; (a) alcance da dependência espacial. Fonte: YAMAMOTO; LANDIM, 2015.

As técnicas baseadas na lógica de conjuntos *Fuzzy* (NICOLETTI; CAMARGO, 2004) são utilizadas no mapeamento digital de solos por meio de sistemas especialistas, onde o conhecimento do pedólogo sobre a relação solo-paisagem é formalizado em curvas de pertinência, para predição de classes ou atributos (MCBRATNEY; ODEH, 1997; ZHU, 1997). Essa a abordagem utiliza os conceitos de modelo semântico e similaridade de vetores para integrar e relacionar dados espaciais de diversas fontes em entradas para análise (SCULL et al., 2003). Nesse sentido, a lógica de conjuntos *Fuzzy* mostra-se adequada por possibilitar a formalização da incerteza, representando melhor o *continuum* do solo e da paisagem, conforme ilustrado na figura 2.4.

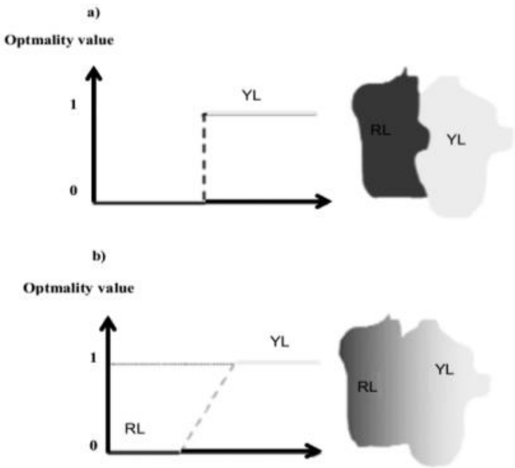


Figura 2.3. Distribuição de um valor ótimo sob a lógica booleana (a) e a lógica fuzzy (b) relacionada à distribuição do Latossolo Amarelo (YL) e do Latossolo Vermelho (RL) na paisagem. Fonte: MENEZES et al., 2013.

Os sistemas de inferência de solos baseados em lógica *Fuzzy*, tais como o SoLIM e o ArcSIE, permitem a construção de curvas de pertinência a partir de três abordagens (MENEZES et al., 2013): baseado em regras, *rule based reasoning*, quando o pedólogo aplica diretamente seu conhecimento na construção das curvas de pertinência; baseado em casos, *case based reasoning*, quando as curvas são extraídas a partir da análise estatística de pontos amostrais; e baseado em mapas legados, *knowledge discovery*, quando as curvas são extraídas a partir da análise estatística e interpretações de mapas vetoriais pretéritos.

O uso de técnicas de mapeamento de solos por lógica *Fuzzy* tem tido sucesso em predizer atributos e classes do solo, com baixa necessidade amostral em comparação com métodos geoestatísticos; além da possibilidade de extrapolar os resultados para novas áreas e incluir o conhecimento dos pedólogos (MENEZES et al., 2014, 2016; SILVA et al., 2016c).

As técnicas baseadas na abordagem da aprendizagem de maquina, ramo da inteligência artificial dedicado ao reconhecimento de padrões e à tomada de decisão a partir de um conjunto de dados, têm sido aplicadas com sucesso em diversos problemas do mapeamento digital de solos. Algoritmos de *deep learning*, a exemplo das redes neurais, foram aplicados com sucesso na predição de classes e atributos do solo, em diversas escalas espaciais e temporais

(PADARIAN; MINASNY; MCBRATNEY, 2019; REICHSTEIN et al., 2019). Os métodos baseados em árvores de decisão, a exemplo do *Random Forest* e *Cubist*, também têm sido amplamente aplicados no mapeamento digital de solos (BHERING et al., 2016; CASTRO-FRANCO et al., 2015; CHEN et al., 2017; MASSAWE et al., 2018; YANG et al., 2016). O algoritmo *Random Forest* possui a vantagem adicional de contar com uma série de métricas facilmente interpretáveis para a avaliação da importância das variáveis do modelo, ou seja, é possível analisar as relações solo-paisagem na perspectiva pedométrica.

Os métodos expostos ao longo dessa seção vêm sendo utilizados com sucesso no mapeamento digital de solos, cabendo ainda algumas observações no sentido da aplicação destes em bases de dados multiescalares. Nesse sentido, nas próximas seções serão expostos alguns fundamentos do conceito de Escala, dando especial destaque àqueles relevantes para a modelagem e mapeamento de solos.

2.3. Conceitos de escala

A escala diz respeito ao tamanho, relativo ou absoluto, e envolve um dos aspectos fundamentais da ciência geográfica. Podemos distinguir três dimensões primárias da escala (WU; LI, 2006): espaço, tempo e nível de organização, conforme ilustrado na figura 2.5. No escopo dessa revisão daremos foco à dimensão espacial da escala.

Adicionalmente, podemos classificar tipos de escala (MONTELLO, 2001): escala do fenômeno, intrínseca à entidade ou processo; escala de análise, abarcando da amostragem à interpretação dos resultados de modelos ou métodos estatísticos; escala cartográfica, ou escala de representação.

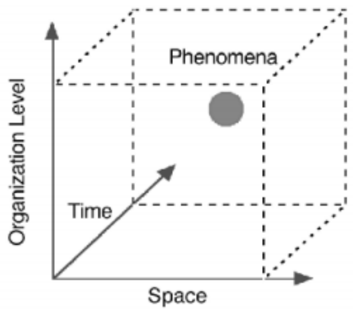


Figura 2.4. Dimensões de escala. Fonte: WU; LI, 2006.

Alguns conceitos centrais de escala merecem descrição, pois não são intercambiáveis, e são amplamente utilizados em pesquisas realizadas no contexto das análises estatísticas espaciais (DUNGAN et al., 2002) e dos sistemas de informação geográfica (GOODCHILD, 2011; WU; LI, 2009), a saber: extensão, *grain*, resolução, *lag* e suporte. A extensão é o comprimento, área ou volume daquilo que existe, é analisado ou representado. A unidade, ou *grain*, é o tamanho médio dos componentes daquilo que existe, ou o tamanho da unidade amostral ou da representação. A resolução é um conceito aplicado em relação as escalas analíticas e de representação, é composta pelo *grain* e também pelo grau de discernimento da mensuração, a exemplo da resolução radiométrica no sensoriamento remoto. O *lag* refere-se ao espaçamento ou intervalo entre unidades vizinhas nos fenômenos ou na análise. O conceito de suporte está relacionado à característica espacial-geométrica de uma variável analisada, abarcando *pixels*, pontos, linhas, polígonos ou formas tridimensionais.

O problema da unidade de área modificável tem especial importância nas análises estatísticas quando se pretende operar em diferentes escalas (DARK; BRAM, 2007; JELINSKI; WU, 1996). O problema compreende dois aspectos, conforme ilustrado na figura 2.6, a saber: o efeito de agregação, atribuído à variação dos resultados numéricos devido ao número de unidades utilizadas em uma mesma área; o efeito de zonação, atribuído à variação oriunda da maneira pela qual as unidades de área menores são reunidas em um número menor de unidades de área maiores.

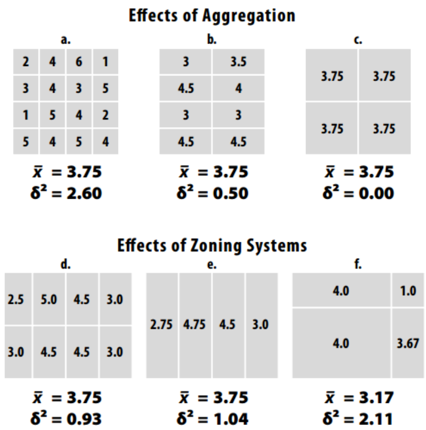


Figura 2.5. Efeitos de agregação (a,b,c) e zonação (d,e,f) sobre médias e variâncias. Fonte: JELINSKI; WU, 1996

2.4. Escala no mapeamento digital de solos

Historicamente as variações e evoluções na escala dos mapas de solo ocorreram em função dos mapas base disponíveis, do entendimento a respeito dos processos pedogenéticos, das tecnologias geográficas e dos propósitos de mapeamento (MILLER; SCHAETZL, 2014, 2016). A partir da segunda metade do século XX, no contexto do mapeamento digital de solos, os mapas passaram a apresentar uma diversidade de escalas e a operar em múltiplas escalas (MALONE; MCBRATNEY; MINASNY, 2013).

A variabilidade espacial dos solos constitui o aspecto fenomenológico prioritário a ser considerado na análise e representação multiescalar dos solos (BRAMMER; NACHTERGAELE, 2015; MILLER et al., 2015; O'ROURKE et al., 2015). Nesse sentido, diversos modelos estatísticos vêm sendo desenvolvidos para tratar esta questão (HEUVELINK; WEBSTER, 2001), conforme descrito na tabela 2.1.

Tabela 2.1. Modelos estatísticos da variabilidade dos solos. Fonte: (HEUVELINK; WEBSTER, 2001)

Tipo do modelo	**Variação Temporal**	**Variação Espacial**	**Fonte primária de informação**	**Adoção nas ciências do solo**
Classificação de solos	Ignorada	representações discretas	pedologia	1960s
Geoestatística	Ignorada	representações continuas	observações	1980s
Classificação+geoestatística	Ignorada	representações mistas	pedologia e observações	1990s
Analises de series temporais	Incluída	ignorada	observações	1990s
Abordagem *state-space*	Incluída	ignorada	leis da física e observações	1990s
Geoestatística espaço-temporal	Incluída	incluída	observações	1995+
Abordagem espacial *state-space*	Incluída	incluida	leis da física e observações	2000+

A resolução da informação utilizada para produzir modelos digitais de elevação afeta, e muitas vezes limita, a segmentação da paisagem com propósitos de mapeamento de solos (GAO, 1997; ROECKER; THOMPSON, 2010; THOMPSON; BELL; BUTLER, 2001). Ainda, é preciso considerar os limites de tamanho de *pixel*, com o limite superior relacionado à representação sem a perda de feições, e o limite inferior quando é alcançada a representação de 95% dos objetos topográficos (HENGL, 2006).

Logo, o *pixel* deve ser relacionado às escalas de mapeamento, tópico ainda omisso nos manuais de levantamento oficiais para o Brasil (IBGE, 2015). Nesse sentido, alguns estudos definiram recomendações de tamanho do *pixel* em função da escala de mapeamento, conforme descritos na tabela 2.2.

Tabela 2.1. Recomendações e estudos sobre tamanho de pixel no mapeamento de solos.

Estudo	**Tamanho de pixel (m x m)**	**Escala recomenda**	**Observações**
(CAVAZZI et al., 2013)	30 - 140		pixels menores produzem melhores resultados apenas para áreas de relevo movimentado
(MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003)	5 20 200 2000	1:5.000 1:20.000 - 1:200.000 1:200.000 - 1:2.000.000 < 1:2.000.000	

(HENGL, 2006)	12 - 53 2 - 13	1:50.000 1:5.000	
(SMITH et al., 2006)	3 - 27		A janela de vizinhança em relação ao tamanho do pixel tem papel crucial nas análises de terreno

A abordagem da calibração semântica da escala em análises digitais de terreno permite considerar a experiência de campo dos cientistas de solo para ajustar os parâmetros de escala do modelo estatístico utilizado (MILLER, 2014). A partir da comparação das derivadas de terreno com interpretações de campo, pode-se concluir que as análises do relevo por cientistas do solo têm resultados consistentes e que a metodologia de calibração conseguiu capturar a escala de análise utilizada nas observações de campo efetuadas.

As transformações de dados entre escalas exigem abordagens metodológicas específicas, em função do problema da unidade de área modificável. As transformações básicas são: o *upscalling*, quando a operação é realizada no sentido de escalas menores; *downscalling*, quando a operação é realizada no sentido de escalas maiores. Métodos geoestatísticos podem ser utilizados para ambas as finalidades com resultados satisfatórios, especialmente na presença de dados auxiliares (MALONE et al., 2012, 2017; STEIN; RILEY; HALBERG, 2001; VEREECKEN et al., 2007).

Quando são realizadas alterações no suporte durante a transformação de escalas é necessário atentar para os problemas de convolução, quando aumentamos o n-dimensional do suporte, e deconvolução, quanto reduzimos o n-dimensional. Nesse sentido, a interpolação por krigagem de bloco tem demonstrado ser eficaz em realizar tais transformações em mapeamentos de solos (MALONE; MCBRATNEY; MINASNY, 2013).

As abordagens de análise multiescalar podem ser agrupadas em dois tipos principais: análises univariadas, quando são consideradas variações no suporte ou resolução, sobre uma única variável; ou multivariadas, quando ocorre a análise conjunta de variáveis em diferentes escalas, conforme ilustrado na figura 2.7.

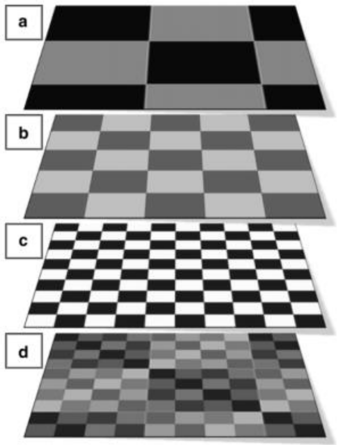


Figura 2.6. Conceito de representação em múltiplas escalas por dados *raster*. Um fenômeno detectado em uma resolução grosseira (a) combinado com fenômenos operando em resoluções mais finas (b e c), gerando uma representação complexa da paisagem (d). Fonte: (MILLER et al., 2015)

Quanto à abordagem univariada, na geomorfometria a aplicação de filtragem por médias locais é uma opção flexível e de aplicação simples em qualquer derivada do terreno (BEHRENS et al., 2010). Esse tipo de técnica é baseada na variação de *n* x *n pixels* nos parâmetros de vizinhança para o calculo de valores médios para os *pixels*, e apresentou resultados superiores quando comparada à abordagem de transformação do tamanho do *pixel*. Outra abordagem comum, porem mais laboriosa quanto a sua implementação computacional, é a utilização de janelas de vizinhança diferenciadas diretamente nos algoritmos de cálculo das derivadas de terreno (ROECKER; THOMPSON, 2010).

Quanto à abordagem multivariada, no mapeamento de solos uma técnica que pode ser utilizada é a do teste sistemático de diferentes conjuntos de variáveis para prever aquelas mais explicativas, por meio de técnicas de *data mining* (MILLER et al., 2015). Nesse estudo, foram avaliados grupos de variáveis por meio do algoritmo *Cubist*, apresentando resultados positivos na seleção de variáveis métricas ou categóricas. Outros algoritmos de aprendizagem de máquina, a exemplo do *Random Forest*, têm sido utilizados com sucesso para a predição de atributos do solo, para a seleção de variáveis nas escalas mais significativas e para a análise conjunta de covariaveis ambientais em diversas escalas (BEHRENS et al., 2018a, 2018b; NEWMAN; LINDSAY; COCKBURN, 2018).

2.5 Questões para pesquisa

A multiplicidade de ferramentas computacionais atualmente disponíveis, o crescente poder de processamento e a proliferação de bases de dados georreferenciados, têm possibilitado a ampla aplicação de métodos pedométricos e geomorfométricos no campo do mapeamento de solos. Fruto dessa mudança de paradigma, do analógico para o digital, o levantamento de solos alcança novas perspectivas quanto à sua flexibilidade, reprodutibilidade e avaliação de incertezas. Nesse sentido, cabe destaque aos estudos relativos a analise da relação solo-paisagem na perspectiva pedométrica, guardando o entendimento que tais relações se dão por meio de interações complexas e multiescalares.

O campo de estudo relacionado às análises geomorfométricas multiescalares têm avançado sobremaneira, como ficou demonstrado nas seções anteriores. No entanto, a despeito de tais avanços, a normatização brasileira ainda é omissa quanto às questões de escala no contexto digital para o levantamento de solos, a exemplo das resoluções dos modelos digitais de elevação.

Por fim, quanto à aplicação de métodos multiescalares para o mapeamento digital de solos no contexto da agricultura familiar amazônica, podemos citar três questões centrais, que merecem atenção e pesquisas futuras: (1) modelagem das relações pedométricas-geomorfométricas nos principais pedoambientes amazônicos; (2) metodologias capazes de fornecer informações compatíveis com pequenas propriedades rurais a partir de levantamentos com menor detalhamento, realizados nas grandes bacias hidrográficas da região, ou mesmo a partir de mapas de solo globais; (3) capacidade de previsão, e da avaliação de incertezas, para glebas, pois esta é a unidade de manejo do solo utilizada pelo agricultor familiar amazônico.

2.6. Referências

ADHIKARI, K.; HARTEMINK, A. E. Linking soils to ecosystem services - A global review. **Geoderma**, v. 262, p. 101–111, 2016.

AITKENHEAD, M. et al. Estimating Soil Properties with a Mobile Phone. Em: HARTEMINK, A. E.; MINASNY, B. (Eds.). **Digital Soil Morphometrics**. [s.l.] Springer International Publishing, 2016. p. 89–110.

ALVES, M. R.; DEMATTÊ, J. A. M.; BARROS, P. P. S. Multiple geotechnological tools applied to digital mapping of tropical soils. **Revista Brasileira de Ciência do Solo**, v. 39, n. 5, p. 1261–1274, 2015.

ASHTEKAR, J. M. et al. Digital mapping of soil properties and associated uncertainties in the Llanos Orientales, South America. Em: ARROUAYS, D. et al. (Eds.). **GlobalSoilMap: Basis of the global spatial soil information system**. [s.l.] CRC Press, 2014. p. 367–372.

BARRIOS, E.; TREJO, M. T. Implications of local soil knowledge for integrated soil management in Latin America. **Geoderma**, v. 111, n. 3–4, p. 217–231, 2003.

BEHRENS, T. et al. Multi-scale digital terrain analysis and feature selection for digital soil mapping. **Geoderma**, v. 155, n. 3–4, p. 175–185, 2010.

BEHRENS, T. et al. Multiscale contextual spatial modelling with the Gaussian scale space. **Geoderma**, v. 310, n. August 2017, p. 128–137, 2018a.

BEHRENS, T. et al. Multi-scale digital soil mapping with deep learning. **Scientific Reports**, v. 8, n. 1, p. 2–10, 2018b.

BHERING, S. B. et al. Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa Agropecuaria Brasileira**, v. 51, n. 9, p. 1359–1370, 2016.

BRAMMER, H.; NACHTERGAELE, F. O. Implications of soil complexity for environmental monitoring. **Journal of Environmental Studies**, v. 72, n. 1, p. 56–73, 2015.

BREVIK, E. C. et al. Soil mapping, classification, and pedologic modeling: History and future directions. **Geoderma**, v. 264, p. 256–274, 2016a.

BREVIK, E. C. et al. Geoderma Soil mapping , classi fi cation , and pedologic modeling : History and future directions. **Geoderma**, v. 264, p. 256–274, 2016b.

BROCHADO, G. T. **Atenuação do efeito do desflorestamento em dados SRTM por meio de diferentes técnicas de interpolação**. [s.l.] Instituto Nacional de Pesquisas Espaciais, 2015.

BRONDÍZIO, E. S. et al. Small farmers and deforestation in Amazonia. Em: [s.l: s.n.]. p. 117–143.

BRUS, D. J.; GRUIJTER, J. J. DE; GROENIGEN, J. W. VAN. DESIGNING SPATIAL COVERAGE SAMPLES USING THE K -MEANS CLUSTERING ALGORITHM. Em: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Eds.). **Developments in Soil Science**. [s.l.] Elsevier, 2007. v. 31p. 183–192.

BUCKNER, B. Soil Renaissance and the Connection to Land Managers. Em: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Eds.). **Global Soil Security**. [s.l.] Springer International Publishing, 2017. p. 255–260.

BURROUGH, P. A. THE DISPLAY OF DIGITAL SOIL DATA , 1976 – 2004. Em: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. (Eds.). **Developments in Soil Science**. [s.l.] Elsevier, 2007. v. 31.

CASTRO-FRANCO, M. et al. Prediction of Soil Properties at Farm Scale Using a Model-Based Soil Sampling Scheme and Random Forest. **Soil Science**, v. 180, n. 2, p. 74–85, 2015.

CATEN, A. TEN et al. Mapeamento digital de classes de solos : características da abordagem brasileira. **Ciência Rural**, v. 42, n. 11, p. 1989–1997, 2012.

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, v. 195–196, p. 111–121, 2013.

CHEN, W. et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. **Catena**, v. 151, p. 147–160, 2017.

DALMOLIN, R. S. D.; CATEN, A. TEN. Mapeamento Digital : nova abordagem em levantamento de solos. **Investigación Agraria**, v. 17, n. 2, p. 77–86, 2015.

DARK, S. J.; BRAM, D. The modifiable areal unit problem (MAUP) in physical geography. **Progress in Physical Geography**, v. 31, n. 5, p. 471–479, 2007.

DELGADO, G. C.; BERGAMASCO, S. M. P. P. **Agricultura Familiar Brasileira : Desafios e Perspectivas de Futuro**. Brasília: Secretaria Especial de Agricultura Familiar e do Desenvolvimento Agrário, 2017.

DUNGAN, J. et al. A balanced view of scale in spatial statistical analysis. **Ecography**, v. 25, n. NA, p. 626–640, 2002.

FLORINSKY, I. V. Influence of Topography on Soil Properties. Em: **Digital Terrain Analysis in Soil Science and Geology**. 2nd. ed. London: Elsevier, 2016. p. 482.

FU, B. et al. Ecosystem services in changing land use. **Journal of Soils and Sediments**, v. 15, n. 4, p. 833–843, 2015.

GALLAHER, C. M.; WINKLERPRINS, A. M. G. A. Effective use of mixed methods in African livelihoods research. **African Geographical Review**, v. 35, n. 1, p. 83–93, 2016.

GAO, J. Resolution and accuracy of terrain representation by grid DEMs at a micro-scale. **International Journal of Geographical Information Science**, v. 11, n. 2, p. 199–212, 1997.

GESSLER, P. E. et al. Soil-landscape modelling and spatial prediction of soil attributes. **International Journal of Geographical Information Systems**, v. 9, n. 4, p. 421–432, 1995.

GOODCHILD, M. F. Scale in GIS: An overview. **Geomorphology**, v. 130, n. 1–2, p. 5–9, 2011.

GOOVAERTS, P. Geostatistics in soil science: State-of-the-art and perspectives. **Geoderma**, v. 89, n. 1–2, p. 1–45, 1999.

GOOVAERTS, P. A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties. **European Journal of Soil Science**, v. 62, n. 3, p. 371–380, 2011.

GRUNWALD, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. **Geoderma**, v. 152, n. 3–4, p. 195–207, 2009.

GRUNWALD, S.; RIVERO, R. L.; REDDY, K. R. Understanding spatial variability and its application to biogeochemistry analysis. **Developments in Environmental Science**, v. 5, n. 07, p. 435–462, 2007.

GRUNWALD, S.; VASQUES, G. M.; RIVERO, R. G. **Fusion of soil and remote sensing data to model soil properties**. [s.l.] Elsevier Ltd, 2015. v. 131

HARTEMINK, A. E.; MCBRATNEY, A. A soil science renaissance. **Geoderma**, v. 148, n. 2, p. 123–129, 2008.

HARTEMINK, A. E.; MCBRATNEY, A. B.; CATTLE, J. A. Developments and trends in soil science: 100 volumes of Geoderma (1967-2001). **Geoderma**, v. 100, n. 3–4, p. 217–268, 2001.

HARTEMINK, A. E.; MINASNY, B. Towards digital soil morphometrics. **Geoderma**, v. 230–231, p. 305–317, 2014.

HENGL, T. Finding the right pixel size. **Computers and Geosciences**, v. 32, n. 9, p. 1283–1298, 2006.

HENGL, T.; HEUVELINK, G. B. M.; STEIN, A. A generic framework for spatial prediction of soil variables based on regression-kriging. **Geoderma**, v. 120, n. 1–2, p. 75–93, 2004.

HEUVELINK, G. B. M.; BRUS, D. J.; DE GRUIJTER, J. J. Optimization of sample configuration for digital mapping of soil properties with universal kriging. **Developments in Soil Science**, v. 31, p. 137–151, 2006.

HEUVELINK, G. B. M.; WEBSTER, R. Modelling soil variation: Past, present, and future. **Geoderma**, v. 100, n. 3–4, p. 269–301, 2001.

HU, G. R.; LI, X. Y.; YANG, X. F. The impact of micro-topography on the interplay of critical zone architecture and hydrological processes at the hillslope scale: Integrated geophysical and hydrological experiments on the Qinghai-Tibet Plateau. **Journal of Hydrology**, v. 583, n. January, p. 124618, 2020.

IBGE. **Manual Técnico de Pedologia**. 3ª ed. Rio de Janeiro: IBGE, 2015.

IWAHASHI, J.; PIKE, R. J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. **Geomorphology**, v. 86, n. 3–4, p. 409–440, 2007.

JASIEWICZ, J.; STEPINSKI, T. F. Geomorphons-a pattern recognition approach to classification and mapping of landforms. **Geomorphology**, v. 182, p. 147–156, 2013.

JELINSKI, D. E.; WU, J. The modifiable areal unit problem and implications for landscape ecology. **Landscape Ecology**, v. 11, n. 3, p. 129–140, 1996.

KÄMPF, N.; CURI, N. Formação e evolução do solo (pedogênese). Em: KER, J. C. et al. (Eds.). **Pedologia: fundamentos**. Viçosa, MG: SBCS, 2012. p. 207–302.

KOCH, A. Soil Security for Agricultural Productivity: The Policy Disconnect and a Promising Future. Em: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Eds.). **Global Soil Security**. [s.l.] Springer International Publishing, 2017. p. 425–435.

LAGACHERIE, P.; MCBRATNEY, A. B. **Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping**. Proceedings of the Global Workshop on Digital Soil Mapping. **Anais**...2006.

MALONE, B. P. et al. A general method for downscaling earth resource information. **Computers and Geosciences**, v. 41, p. 119–125, 2012.

MALONE, B. P. et al. Digital soil mapping of soil carbon at the farm scale: A spatial downscaling approach in consideration of measured and uncertain data. **Geoderma**, v. 290, p. 91–99, 2017.

MALONE, B. P.; MCBRATNEY, A. B.; MINASNY, B. Spatial Scaling for Digital Soil Mapping. **Soil Science Society of America Journal**, v. 77, n. November 2015, p. 890–902, 2013.

MASSAWE, B. H. J. et al. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. **Geoderma**, v. 311, p. 143–148, 2018.

MCBRATNEY, A. B. et al. Soil Security: A Rationale. Em: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Eds.). **Global Soil Security**. [s.l.] Springer International Publishing, 2017. p. 3–14.

MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1–2, p. 3–52, 2003.

MCBRATNEY, A. B.; ODEH, I. O. A. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. **Geoderma**, v. 77, n. 2–4, p. 85–113, 1997.

MCBRATNEY, A.; FIELD, D. J.; KOCH, A. The dimensions of soil security. **Geoderma**, v. 213, p. 203–213, 2014.

MENEZES, M. D. DE et al. Digital soil mapping approach based on fuzzy logic and field expert knowledge. **Ciência e Agrotecnologia**, v. 37, n. 4, p. 287–298, 2013.

MENEZES, M. D. DE et al. Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches. **Scientia Agricola**, v. 71, n. 4, p. 316–323, ago. 2014.

MENEZES, M. D. DE et al. Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. **Scientia Agricola**, v. 73, n. 3, p. 274–285, 2016.

MILLER, B. A. Semantic Calibration of digital terrain analysis scale. **Cartography and Geographic Information Science**, v. 41, n. 2, p. 166–176, 2014.

MILLER, B. A. et al. Impact of multi-scale predictor selection for modeling soil properties. **Geoderma**, v. 239, p. 97–106, 2015.

MILLER, B. A.; SCHAETZL, R. J. The historical role of base maps in soil geography. **Geoderma**, v. 230–231, p. 329–339, 2014.

MILLER, B. A.; SCHAETZL, R. J. History of soil geography in the context of scale. **Geoderma**, v. 264, p. 284–300, 2016.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, v. 32, n. 9, p. 1378–1388, 2006.

MINASNY, B.; MCBRATNEY, A. B. Digital soil mapping: A brief history and some lessons. **Geoderma**, v. 264, p. 301–311, 2016.

MONTANARELLA, L. Translating Soil Science Knowledge to Public Policy. Em: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Eds.). **Global Soil Security**. [s.l.] Springer International Publishing, 2017. p. 451–456.

MONTELLO, D. R. Scale in geography. **International Encyclopedia of the Social & Behavioral Sciences**, n. 1997, p. 13501–13504, 2001.

MULDER, V. L.; DE BRUIN, S.; SCHAEPMAN, M. E. Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. **International Journal of Applied Earth Observation and Geoinformation**, v. 21, n. 1, p. 301–310, 2012.

NELSON, A.; REUTER, H. I.; GESSLER, P. Dem Production methods and sources. **Developments in Soil Science**, v. 33, n. C, p. 65–85, 2009.

NEWMAN, D. R.; LINDSAY, J. B.; COCKBURN, J. M. H. Evaluating metrics of local topographic position for multiscale geomorphometric analysis. **Geomorphology**, v. 312, p. 40–50, 2018.

NICOLETTI, M. DO C.; CAMARGO, H. DE A. **Fundamentos da teoria de conjuntos fuzzy**. São Carlos, SP: EdUFSCar, 2004.

OBERTHÜR, T. Mapping soil texture classes using field texturing, particle size distribution and local knowledge by both conventional and geostatistical methods. **European Journal of Soil Science**, v. 50, n. 3, p. 457, 1999.

O'ROURKE, S. M. et al. **Soil organic carbon across scales**. **Global Change Biology**, 2015.

PADARIAN, J.; MINASNY, B.; MCBRATNEY, A. B. Using deep learning for digital soil mapping. **Soil**, v. 5, n. 1, p. 79–89, 2019.

PINHEIRO, H. S. K. et al. Applying Artificial Neural Networks Utilizing Geomorphons to Predict Soil Classes in a Brazilian Watershed. Em: **Digital Soil Mapping Across Paradigms , Scales and Boundaries**. [s.l.] Springer Science+Business Media Singapore, 2016. p. 89–102.

POLIDORO, J. C. et al. **Programa Nacional de Solos do Brasil (PronaSolos)**. Rio de Janeiro: Embrapa Solos, 2016.

REICHSTEIN, M. et al. Deep learning and process understanding for data-driven Earth system science. **Nature**, v. 566, n. 7743, p. 195–204, 2019.

REUTER, H. I. et al. Preparation of DEMs for geomorpometric analysis. **Developments in Soil Science**, v. 33, n. 2, p. 87–120, 2009.

RIBEIRO-FILHO, A. A.; ADAMS, C.; SERENI-MURRIETA, R. SERGIO. The impacts of shifting cultivation on tropical forest soil : a review. **Boletim do Museu Paraense Emilio Goeldi:Ciencias Humanas**, v. 8, n. 3, p. 693–727, 2013.

ROECKER, S. M.; THOMPSON, J. A. Scale Effects on Terrain Attribute Calculation and Their Use as Environmental Covariates for Digital Soil Mapping. Em: **Digital Soil Mapping**. Dordrecht: Springer Netherlands, 2010. p. 55–66.

ROSSITER, D. G. et al. Can citizen science assist digital soil mapping? **Geoderma**, v. 259–260, p. 71–80, 2015.

SANCHEZ, P. A. et al. Digital Soil Map of the World. **Africa**, n. August, p. 8–12, 2009.

SANTOS, H. G. et al. Distribuição espacial dos níveis de levantamento de solos no Brasil. **XXXIV Congresso Brasileiro de Ciência do Solo**, n. 2, p. 1–4, 2013.

SCHMIDT, J.; ANDREW, R. Multi-scale landform characterization. **Area**, v. 37, n. 3, p. 341–350, 2005.

SCULL, P. et al. Predictive soil mapping: a review. **Progress in Physical Geography**, v. 27, n. 2, p. 171–197, 1 jun. 2003.

SHI, X. et al. Integrating Different Types of Knowledge for Digital Soil Mapping. **Soil Sci Soc Am J**, v. 73, n. 5, p. 1682–1692, 2009.

SILVA, S. et al. Geomorphometric tool associated with soil types and properties spatial variability at watersheds under tropical conditions. **Scientia Agricola**, v. 73, n. 4, p. 363–370, 2016a.

SILVA, S. et al. Proximal Sensing and Digital Terrain Models Applied to Digital Soil Mapping and Modeling of Brazilian Latosols (Oxisols). **Remote Sensing**, v. 8, n. 8, p. 614, 25 jul. 2016b.

SILVA, S. H. G. et al. A Technique for Low Cost Soil Mapping and Validation Using Expert Knowledge on a Watershed in Minas Gerais, Brazil. **Soil Science Society of America Journal**, v. 78, n. 4, p. 1310, 2014.

SILVA, S. H. G. et al. Evaluation of Conditioned Latin Hypercube Sampling as a Support for Soil Mapping and Spatial Variability of Soil Properties. **Soil Science Society of America Journal**, v. 79, n. 2, p. 603, 2015.

SILVA, S. H. G. et al. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. **Geoderma**, v. 267, p. 65–77, 2016c.

SMITH, M. P. et al. The effects of DEM resolution and neighborhood size on digital soil survey. **Geoderma**, v. 137, n. 1–2, p. 58–69, 2006.

STEIN, A.; RILEY, J.; HALBERG, N. Issues of scale for environmental indicators. **Agriculture, Ecosystems and Environment**, v. 87, n. 2, p. 215–232, 2001.

TARGULIAN, V. O.; KRASILNIKOV, P. V. Soil system and pedogenic processes: Self-organization, time scales, and environmental significance. **Catena**, v. 71, n. 3, p. 373–381, 2007.

THOMPSON, J. A.; BELL, J. C.; BUTLER, C. A. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. **Geoderma**, v. 100, n. 1–2, p. 67–89, 2001.

TORNQUIST, C. G.; BROETTO, T. Protection of the Soil Resource in the Brazilian Environmental Legislation. Em: FIELD, D. J.; MORGAN, C. L. S.; MCBRATNEY, A. B. (Eds.). **Global Soil Security**. [s.l.] Springer International Publishing, 2017. p. 397–401.

VALERIANO, M. DE M.; ROSSETTI, D. DE F. Topodata: Brazilian full coverage refinement of SRTM data. **Applied Geography**, v. 32, n. 2, p. 300–309, 2012.

VEREECKEN, H. et al. Upscaling Hydraulic Properties and Soil Water Flow Processes in Heterogeneous Soils. **Vadose Zone Journal**, v. 6, n. 1, p. 1–28, 2007.

WILSON, J. P. Digital terrain modeling. **Geomorphology**, v. 137, n. 1, p. 107–121, 2012.

WU, H.; LI, Z. L. Scale issues in remote sensing: A review on analysis, processing and modeling. **Sensors**, v. 9, n. 3, p. 1768–1793, 2009.

WU, J.; LI, H. **SCALING AND UNCERTAINTY ANALYSIS IN ECOLOGY**. Dordrecht: Springer Netherlands, 2006.

WYSOCKI, D. A.; SCHOENEBERGER, P. J. Geomorphology of Soil Landscapes. Em: HUANG, P. M.; LI, Y.; SUMNER, M. E. (Eds.). **Handbook of Soil Science: Properties and Processes**. [s.l.] Chemical Rubber Company Press, Boca Raton, Florida, 2011. p. 1–26.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. [s.l.] Oficina de Textos, 2015.

YANG, R. M. et al. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. **Ecological Indicators**, v. 60, p. 870–878, 2016.

ZHU, A.-X. A similarity model for representig soil spatial information. **Geoderma**, v. 77, p. 217–242, 1997.

ZINCK, J. A. The Geomorphic Landscape: Criteria for Classifying Geoforms. Em: **Geopedology**. Cham: Springer International Publishing, 2016a. p. 77–99.

ZINCK, J. A. The Geomorphic Landscape: Classification of Geoforms. Em: **Geopedology**. Cham: Springer International Publishing, 2016b. p. 101–125.

CAPÍTULO I. MULTISCALAR GEOMORPHOMETRIC GENERALIZATION FOR SOIL-LANDSCAPE MODELING BY RANDOM FOREST: A CASE STUDY IN THE EASTERN AMAZON.

Abstract

Multiscalar topography influence on soil distribution has a complex pattern that is related to overlay of pedological processes which occurred at different times, and these driving forces are correlated with many geomorphologic scales. In this sense, the pre-sent study tested the hypothesis whether multiscale geomorphometric generalized covariables can improve pedometric modeling. To achieve this goal, this case study ap-plied the Random Forest algorithm to a multiscale geomorphometric database to predict soil surface attributes. The study area is in phanerozoic sedimentary basins, in the Alter do Chão geological formation, Eastern

Amazon, Brazil. The multiscale geomorphometric generalization was applied at general and specific geomorphometric covariables, producing groups for each scale combination. The modeling was run using Random Forest for A-horizon thickness, pH, silt and sand content. For model evaluation, visual analysis of digital maps, metrics of forest structures and effect of variables on prediction were used. For evaluation of soil textural classifications, the confusion matrix with a Kappa index, and the user's and producer's accuracies were employed. The geomorphometry generalization tends to smooth curvatures and produces identifiable geomorphic representations at sub-watershed and watershed levels. The forest structures and effect of variables on prediction are in agreement with pedological knowledge. The multiscale geomorphometric generalized covariables improved accuracy metrics of soil surface texture classification, with the Kappa Index going from 0.43 to 0.62. Therefore, it can be argued that topography influences soil distribution at combined coarser spatial scales and is able to predict soil particle size contents in the studied watershed. Future development of the multiscale geomorphometric generalization framework could include generalization methods concerning preservation of features, landform classification adaptable at multiple scales.

Keywords: Digital Soil Mapping, Upscaling, Machine Learning, Random Forest algorithm, multiscale geomorphometric generalization.

3.1. Introduction

Elements of the landscape control the processes acting on soils; therefore, the soil-landscape approach constitutes one of the most powerful conceptual tools in mapping activities, especially at scales with an intermediate or greater level of detail. The soil-landscape relationship is related to the concept of the catena, coined by Milne (MILNE, 1935). In a catena, variations in soils along a slope are attributed to the translocation of soluble elements and to erosive and depositional processes, not excluding different source mate-rials. Subsequently, the analyses of soil-landscape relations proposed by Hugget (HUGGETT, 1975), contemplated three-dimensional models of the slopes. In the context of digital soil map-ping, in the scorpan model paradigm (MCBRATNEY; MENDONÇA SANTOS; MINASNY, 2003), soil-landscape process modeling can be described as an interdisciplinary object of the interface between pedometry-geomorphometry (MA et al., 2019).

The scale issues of soil-landscape relationships are related to the complex interactions of both elements and how these processes occur and are perceived. Several geomorphology studies report a time–space coupling between landform size-scales and lifetime (SCHMIDT; ANDREW, 2005; WYSOCKI; SCHOENEBERGER, 2011). From a soil perspective, different pedological process will manifest influence at short, mid or long time-scales (KÄMPF; CURI, 2012; TARGULIAN; KRASILNIKOV, 2007). In contrast, topography studies in soil physics demonstrate a complex water dynamic related to the nested geometry of slopes, considering relief and micro-relief patterns, resulting in trends in the movement of particles and solutes and changes in texture and chemical parameters of soils (FLORINSKY, 2016; HU; LI; YANG, 2020). Therefore, multiscalar topography influences a particular soil distribution in two general aspects, overlay of pedological processes that occurred at different times, and driving forces in the present time, determined by the sum of forces better correlated with one, several, or many geomorphologic scales.

Some aspects of spatial scaling in digital soil mapping have been summarized in non-exhaustive reviews (MALONE; MCBRATNEY; MINASNY, 2013; PACHEPSKY; HILL, 2017). The hierarchical definition of scale can be used to understand the soil phenomena, from the soil region, passing through watersheds, catena, pedon, horizons and finally molecular interactions. The characteristics of measurement affect the results of analysis, and in this sense, the modifiable areal unit problem (MAUP) represents a key issue. Information transfer across scales could be classified in upscaling, in less detail, or downscaling, with greater detail, but both of these require must consider bias.

With respect to the scale of covariables for predictive soil mapping, highest DEM resolutions do not necessarily produce the highest accuracy (CAVAZZI et al., 2013; SAMUEL-ROSA et al., 2015). Despite the potential of machine learning to produce complex and nonlinear predictions, few studies have investigated multiscale perspective in covariables to account for physical process that are not predictable by finer scale environmental information (WADOUX; MINASNY; MCBRATNEY, 2020). Some studies propose data-driven techniques for selection of pixel size or neighborhood size for a particular landscape (HENGL, 2006; SMITH et al., 2006), but this approach produces a variety of results in different geomorphic units, which complicate the interpretation of scalar components in the soil-landscape relationship. For a friendly interpretation of scale relationships on soil-landscape models, this study proposed a cartographic-based criterion to formalize the scale correspondence to pixel size for geomorphometric covariables.

The present study tested the hypothesis whether multiscale geomorphic representation, obtained from cartographic generalization of a digital elevation model, can improve pedometric modeling. To achieve this goal, this case study applied the Random Forest algorithm to a multiscale geomorphometric database to predict soil surface attributes.

3.2. Material and methods

The procedures described in this section were performed using the open-source software QGIS 3.10; SAGA GIS 2.3; GRASS GIS 7.8; and R Programming 3.5 (CONRAD et al., 2015; GRASS, 2019; QGIS, 2019; R CORE TEAM, 2019).

3.2.1. Study area

The study was conducted in Iripixi Lake (ILW) and Caipuru Lake (CLW) watersheds, with an area of 27.137 and 28.315 ha respectively, located in the Trombetas basin in Oriximiná-Pará in the Eastern Amazon. Pilot areas, consisting of small farms, are distributed in the upper and lower courses in both basins and adjacent basin boundaries, as shown in Figure 3.1, totaling 697 ha, approximately 1% of the extent of the watersheds.

The study area is in phanerozoic sedimentary basins, in the Alter do Chão geological formation, whose local characteristics were evaluated from a faciological analysis near Óbidos (MENDES; TRUCKENBROD; RODRIGUES, 2012), a neighboring municipality. This region is classified as humid equatorial climate with 3 dry months (IBGE, 2002). The geomorphic units of these watersheds are classified as a homogeneous dissection with coarse drainage density and weak incision depth (IBGE, 2008). The pedoenvironment of this study area are in the upper lands of the lower Amazon basin and the most abundant soil classes in the area are *Latossolo Amarelo, Latossolo Vermelho-Amarelo, Argissolo Vermelho-Amarelo and Gleissolos Haplicos* (SCHAEFER et al., 2017).

3.2.2. Environmental covariates

In the proposed mapping scale, vegetation and topography factors are the main sources of soil variation. In this study, to represent such factors, the source multispectral images Landsat 8 and Shuttle Radar Topography Mission Digital Elevation Model 30m (SRTM DEM 30m), were used, respectively.

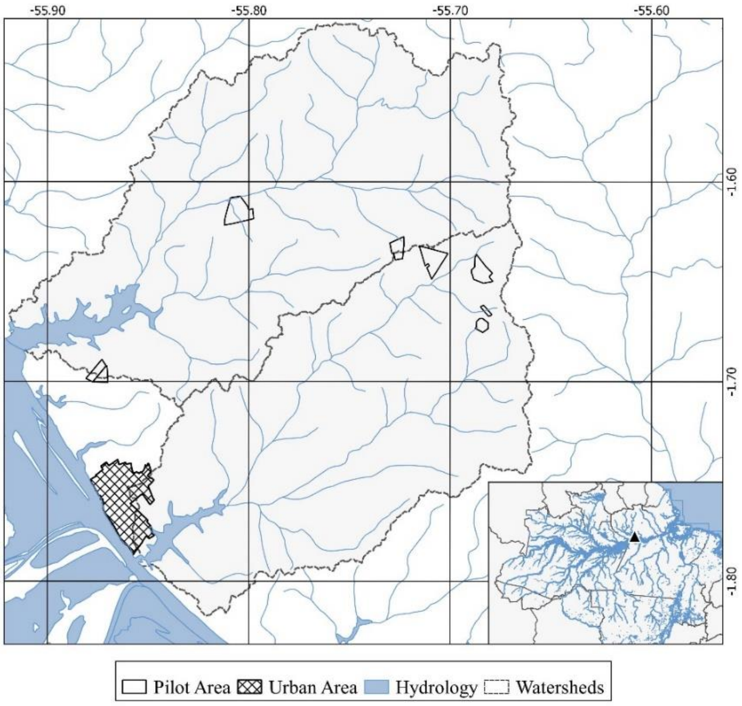


Figure 3.1. Study area location in the Eastern Amazon.

The Landsat 8 images are from September 11, 2017, corrected for surface reflectance with the LaSRC algorithm by USGS (U.S. GEOLOGICAL SURVEY, 2019) . The SRTM DEM is a digital elevation model based on stereoscopic radar survey, and has 30m pixels (FARR et al., 2007). The corrections made in SRTM DEM were filling in sinks, and reduction of deforestation effect by the estimated canopy addition method (BROCHADO, 2015).

The topography information was upscaled and organized into generalized multiscale geomorphometric variable groups, as detailed in the next section.

3.2.3. Multiscale geomorphometric generalization

The multiscale geomorphometric generalization (MGG), is an upscaling operation, based on cartographic concepts of generalization of digital maps (GUILBERT; BOGUSLAWSKI; ISIKDAG, 2019; LI; OPENSHAW, 1993). This approach can be applied for any geomorphic variable, including elevation models, land-forms units, and primary and secondary derivatives.

This operation results in variables at different scales, arranged in groups according to criteria required for the analysis. The framework of these upscaling methods brings to the pedometric perspective the understanding that the soil-landscape relationship occurs through complex and multiscale interactions. In this sense, the formalization of the desirable scales of analysis and modeling occurs both in their definition and in the group arrangements.

For the MGG operation, vector and raster representations demand different approaches for upscaling, because each of them has specific scale transformation problems due to their mathematical structures. Furthermore, it is necessary to have a unique reference for the scales for compatible representation of geomorphic features in both types of variables, thus allowing for joint interpretation. In this study, the concept of minimum mappable area for soil surveys (IBGE, 2015) was considered to define pixel sizes in relation to cartographic scale. The detailed descriptions for each of the four scales used are in Table 3.1. The area equivalence between raster and vector is calculated as a function of a 5x5 pixel grid, considered a conservative parameter to determine a geomorphic feature.

The MGG was applied to the following geomorphometric covariables: elevation (Elev), slope, relative slope position (RSP), topographic wetness index (TWI), plan curvature (PlanCurv), profile curvature (ProfCurv), topographic factor of water erosion (LS) and geomorphons.

These geomorphic variables, at different scales, were obtained from SRTM DEM from two upscaling methods, as illustrated in Figure 3.2. Using local averages on covariable elevation, in 2x2, 3x3, 4x4 windows, for resolutions 60m, 90m and 120m, respectively, and subsequent derivatives covariable calculation. Classification of geomorphons (JASIEWICZ; STEPINSKI, 2013) was followed by the exclusion of polygons smaller than the minimum mappable area for each scale. Such methods correspond to cartographic generalization applied to general geomorphometry and specific geomorphometry, respectively (ZINCK, 2016).

Table 3.1. Correspondence between scale and pixel size for Multiscale Geomorphometric Generalization (MGG), using the concept of minimum mappable area.

Scale	Minimal mappable area (m^2)	Pixel size (m)	Pixel area[a] (m^2)	pa/mma[b] (%)
1:25000	25000	30	22500	90
1:50000	100000	60	90000	90
1:75000	225000	90	202500	90
1:100000	400000	120	360000	90

a. For a 5x5 window. b. Ratio pixel area (pa) by minimal mappable area (mma), in percentage.

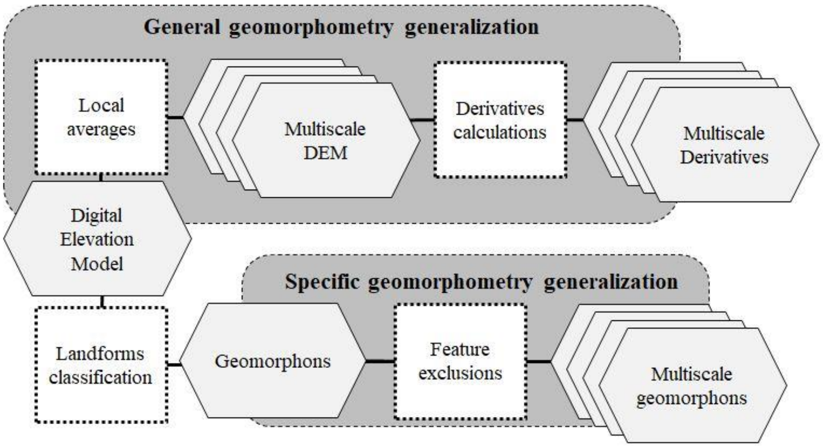


Figure 3.2. Methodology flowchart of MGG for the topography covariables.

In this case study, was used a machine learning approach to identify and select the optimum scales of variables for modeling. It was therefore necessary to provide a multiple database for training and evaluation each scale. In this sense, the variables were organized from the combination of the set of topography variables, arranged in all possible combinations.

3.2.4. Soil sampling and analysis

Soil sampling was performed in 9 pilot areas, considering covariables to evaluate the effect of topography related to variation in soil distributions. Each pilot area was sampled at 10 points, a sufficient density for semi-detailed soil surveys, compatible with the 1:25,000 scale soil maps (IBGE, 2015). The sample points were distributed according to a stratified random arrangement by the conditioned latin hypercube method (BISWAS; ZHANG, 2018; MINASNY; MCBRATNEY, 2006), with raster topography covariates, described in the previous section, at a 1:25.000 scale. At the total of 90 sample points, the morphological description of the A horizon was performed (SANTOS et al., 2015) and soil samples were collected in the 0-30cm depth for physical and chemical analyses (EMBRAPA, 2017; KETTLER; DORAN; GILBERT, 2001). To evaluate variances and patterns in the sample dataset, principal component analysis (ABDI; WILLIAMS, 2010) was conducted.

3.2.5. Modeling by Random forest

The modeling of the soil-landscape was done using Random Forest (BREIMAN, 2001), a machine learning algorithm frequently used to produce digital soil maps (LAMICHHANE; KUMAR; WILSON, 2019). Some characteristics of this algorithm that are worth mentioning are that it can handle categorical and continuous variables, it can do regression and classifications, it is robust for overfitting problems, and is feasible for interpretation of variable relationships, including linear and non-linear systems (MALONE; MINASNY; MCBRATNEY, 2017).

First, the set of training cases and those intended for validation were defined. The selection was made randomly, with proportions of 70% and 30%, respectively. The training was done for every group of geomorphometric covariables, one group at a time, for modeling of some soil attributes, namely, A horizon thickness, pH, silt and sand content. Next, for the prediction of the multiscale generalized geomorphometric groups, the groups with the best adjustment were evaluated and selected based on the highest values of % variation explained by the model. Finally, the modeling structure and results of original scale geomorphometric and generalized geomorphometrics groups were com-pared. For evaluation of this predictions, visual analysis of digital maps and multi-way plots of forest structures and effect of variables on prediction were used, calculated with Random Forest Explainer in the R package (PALUSZYNSKA; BIECEK; JIANG, 2019).

The prediction of soil texture was made by considering the silt and sand raster layers using the Brazilian soil classification system (SANTOS et al., 2018). For evaluation of this prediction, the confusion matrix was calculated with the Kappa index, and the user's and producer's accuracy (LIU; FRAZIER; KUMAR, 2007).

3.3. Results and discussion

3.3.1. Generalized Multiscale Geomorphometrics

The multiscale geomophometric generalization produces a dataset of covariables at each scale. For evaluation of that operation, in this section the tendency of geomorphometry distributions and a geomorphology interpretation of this generalized data will be described.

The application of MGG to the original DEM database resulted in 28 continuous geomorphometric variables, the distributions of which are illustrated in Figure 3.3. Some distributions had smaller changes, like TWI and LS, with a reduction of occurrence of the extreme values and some scatter reduction in the upscaling direction. This could be explained by the lowest representation of smaller geomorphic features that contributed to counting of upper and lower limits on dynamics of the water on the slope. Some variable distributions had highly significant changes, such as slope and geomorphic curvatures, which had a sharp and progressive reduction of scatter related to smoothest generalized surface. Considering the

PlanCurv, it is observable that the upscaled derivative better describes the features of the valleys and spikes in the study area, as illustrated in Figure 3.4 (b).

Some studies test upscaling effects on digital topographic information, with comparable results. The contextual spatial modeling, using gaussian space scale rates to produce a set of coarse resolution DEMs, had a similar result of smooth slope and geomorphic curvatures (BEHRENS et al., 2018). In contrast, other studies have proposed sophisticated calculations for generalization of DEM considering questions of feature preservation, and could be tested with the MGG framework. The Feature Preserving DEM Smoothing (FPDEMS) method reduces the complexity of the surface at the detailed spatial scales at which roughness dominates, while not significantly altering the topographic complexity at larger spatial scales (LINDSAY; FRANCIONI; COCKBURN, 2019). Other approaches use a multi-point algorithm to rapidly and accurately retrieve the critical points, using drainage-constrained TIN, to produce coarser-resolution DEMs (WU et al., 2019; ZHOU; CHEN, 2011).

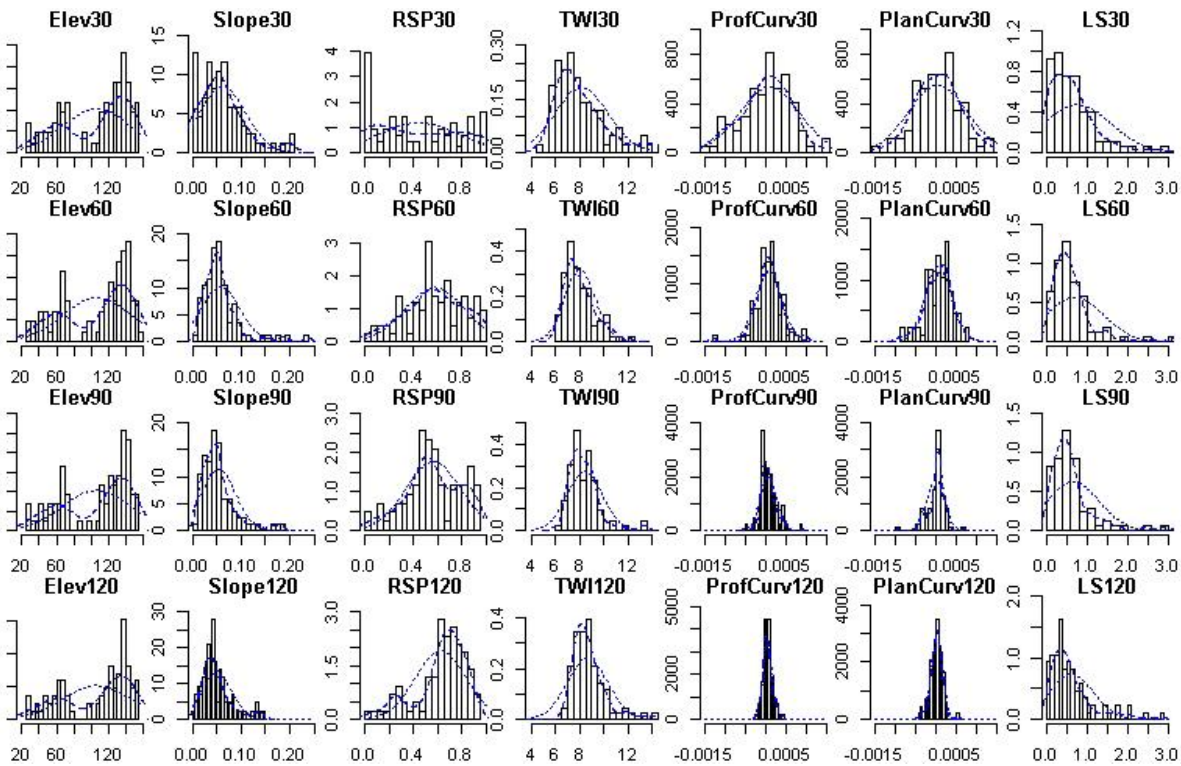


Figure 3.3. Distributions of geomorphometry raster variables at original and generalized scales.

The distribution of relative slope positions had the most complex variation, and was related to dissection process acting on a flattened surface of phanerozoic sedimentary basins, that results on organization of this type of landform, a homogeneous dissection with tabular top, coarse drainage density and weak incision depth (IBGE, 2008, 2009). A comparative 3D view of original and generalized relative slope positions is illustrated in Figure 4 (a). This variable at its original scale identifies the variation at a sub-watershed level, with 0-1 interval values that occur at each drainage division. In contrast, at 1:100.000, this variable shows variation at the watershed level, with high values only at the most elevated ridgeline section. In this perspective, the upscaled variable can identify upper and lower courses of basins, a relevant information for soil mapping, because it allows analysis of base level, exposition time to weathering and stratigraphic differentiation on sedimentary geological formation (WYSOCKI; SCHOENEBERGER, 2011).

The geomorphon scale transformations resulted in incorporation to the valley portions of some local landform segments which originally were classified in topo and slope, as illustrated in Figure 3.5 (a,b). Therefore, at a 1:100.000 scale, some geomorphic features were disconnected from the main patterns and were omitted and substituted by the local prevalent geomorphon type. Regarding the original scale geomorphons, the results show consistent valley-slope-top landforms sequences in the most of mapping site. However, at same scale of soil mapping in South Africa, the geomorphons were compared with expert manual delineation with a high degree of mismatch (57% of the area), despite reasonable prediction results (FLYNN et al., 2020). Alternatively, considering the low importance of geomorphons in decision trees, as we will see in the next section, it is possible to test other approaches to generalization, including the variation of search parameters in the definition of relief units.

Other approaches to classification of landforms, like the topographic position index (DE REU et al., 2013), and k-median clustering (SZYPUŁA; WIECZOREK, 2020), also had parameters that could be adjusted for the proposed correspondence of scales. Therefore, these techniques can be included and tested with the MGG framework in future research.

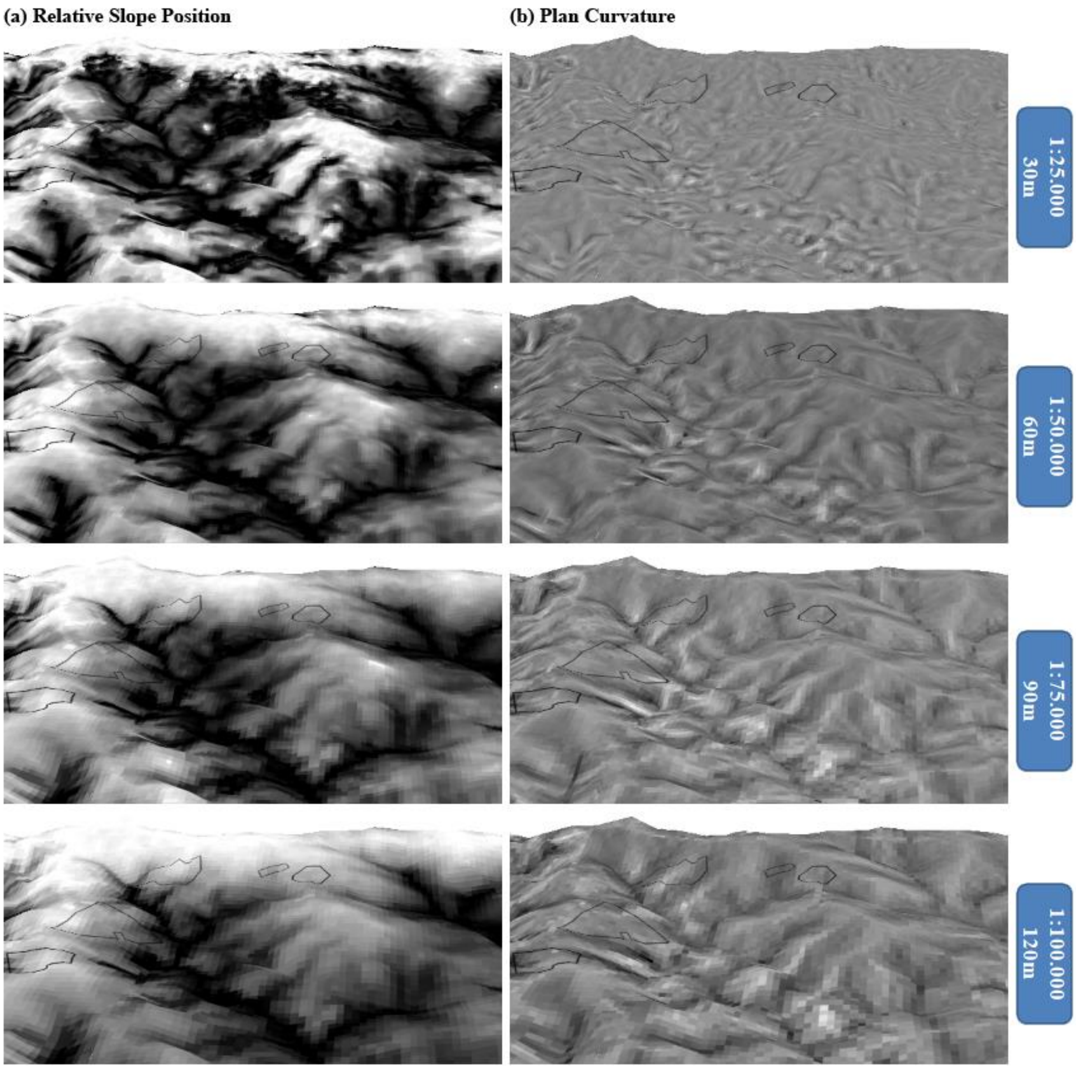


Figure 3.4. Comparative 3D view of original and generalized geomorphometric covariables: (a) Relative slope position; (b) Plan Curvature.

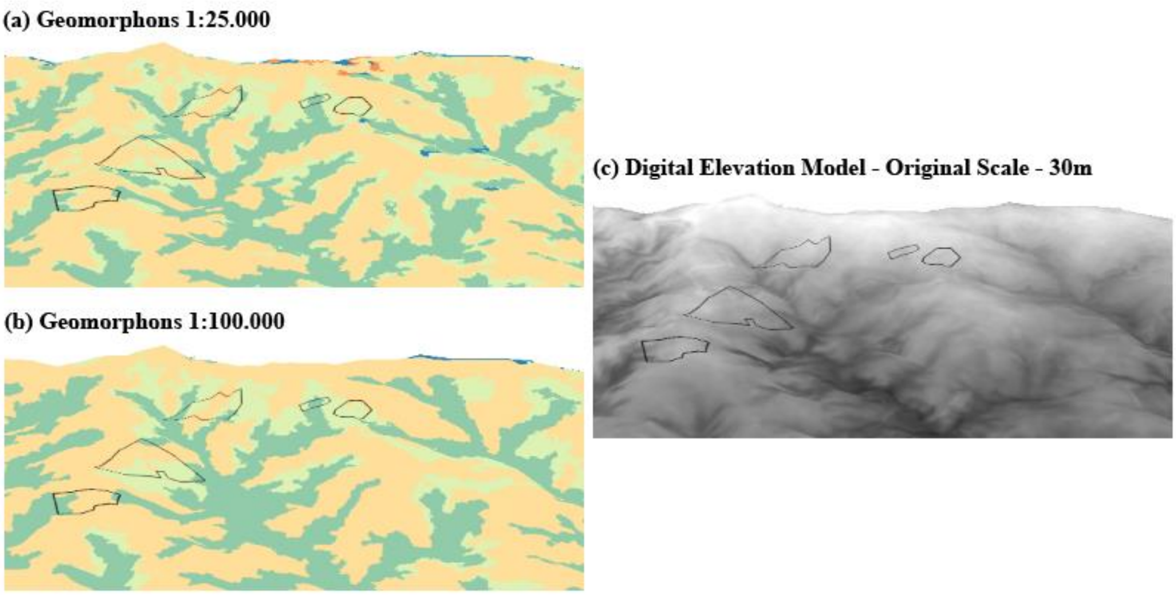


Figure 3.5. Comparative 3D view of original and generalized geomorphometric covariables: (a) Geomorphons 1:25.000; (b) Geomorphons 1:100.000; (c) Digital Elevation Model at original scale.

3.3.2. Evaluation of Soil-Landscape Models

This section presents the relationship of soil variables to geomorphometric covariables, in the context of MGG and will focus on whether the decision tree forest structures have an expected pedological meaning, and if a greater number of topographic variables could mask other factors.

The main soil variation is related to particle size distribution, as shown by the contribution of this variable to the first component of PCA, as illustrated in Figure 3.6. The A horizon thickness has an opposite tendency with respect to silt, and is only slightly in agreement with sand content, revealing the weathering dynamics related to translocation of organic matter in sandy rainforest soils. Organic matter has a relevant influence on soil differentiation and is directly related to pH, as expected in this pedoenvironment, a upper land of the lower Amazon basin (SCHAEFER et al., 2017).

Also, exchangeable aluminum has only a slight agreement with clay content, and this is probably associated with clay mineralogy dominated by kaolinite, common in that fraction in acid soils (CUNHA; DE ALMEIDA; BARBOZA, 2014). Studies indicate that predominance of kaolinite at the surface soil is explained by reduced Silica leaching by cycling of Si promoted by the forest vegetation, and the resulting soil properties patterns was found on Amazon soils located on sedimentary domains with high exchangeable aluminum content, like this study area (LUCAS et al., 1993; SOUZA et al., 2018).

The variation in the A horizon is strongly related to landcover and topographic factors (HARTEMINK et al., 2020). For modeling the A horizon, the MGG group 15 had the best adjustment, with 40.5% of variation explained, in contrast to 30.2% for the original scale DEM, indicating the influence of coarse variables. This could mean that there is prevalence of older pedological processes associated with larger and older geomorphic features.

The decision tree forest structures, as illustrated in Figure 3.7, clearly show the main controllers of tendencies, represented by Elev and RSP, such as the high portion of watersheds that have conserved and currently present long-term pedological processes of soil development that have resulted in an overall greater thickness, in contrast with the dissected portion. The water content and dynamics, represented by TWI, PlanCurv, ProfCurv, influence the variation in the A horizon thickness, and landcover also has a significant importance in the model, represented by surface reflectance variables.

These results show similarity with another study that reported increased accuracy of soil prediction by Random Forest on four datasets across the globe with addition of a coarser scale DEM (BEHRENS et al., 2019). These authors also proposed an approach with a variogram of soil properties to a priori approximations of the effective scale for modelling. In the context of MGG, this can be used for scale definition and group arrangement for soil attributes on specific pedoenvironment.

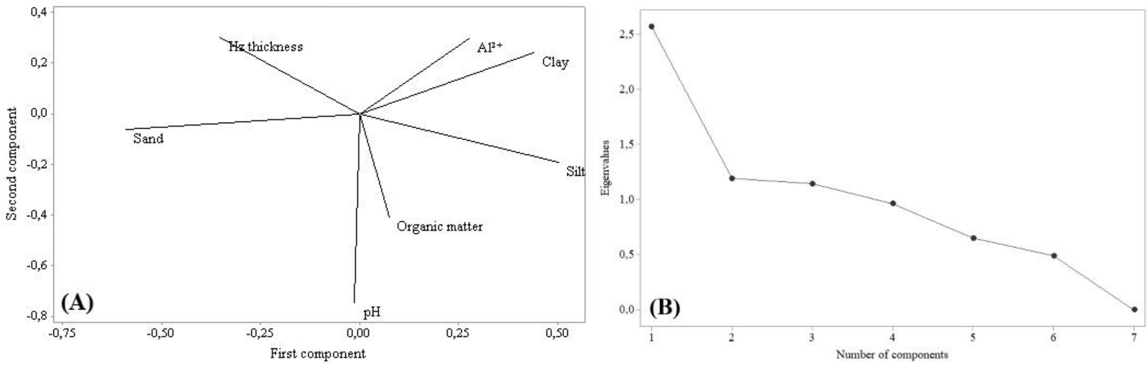


Figure 3.6. (a) Vectors and (b) variable loadings of principal components for soil attributes.

Soil pH influences microbiota, nutrient uptake, root growth and therefore plant development in general (NOVAIS, ROBERTO F.; MELLO, 2007). Soil pH is strongly related to landuse, or in scorpan, the organism factor. As expected, the most important variable for modeling of this soil attribute, at the original scale, is the surface reflectance, as shown in Figure 3.8 (a). When considering the importance plot of the MGG group 22, illustrated in Figure 3.8 (b), the same five variables have with highest importance index. In this sense, both models show topographic factors with less importance then vegetation factors, in the same proportion, despite twice the number of geomorphometric variables in the MGG.

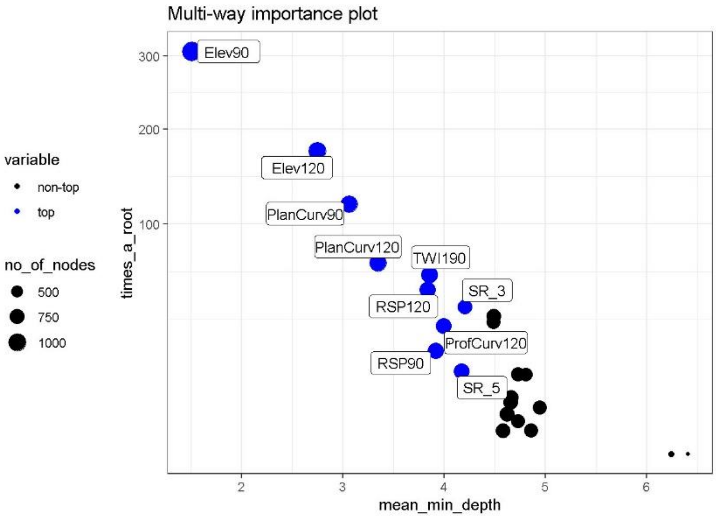


Figure 3.7. MGG decision tree forest structures for A horizon thickness, geomorphometric group 1:75000 plus 1:100000 scales.

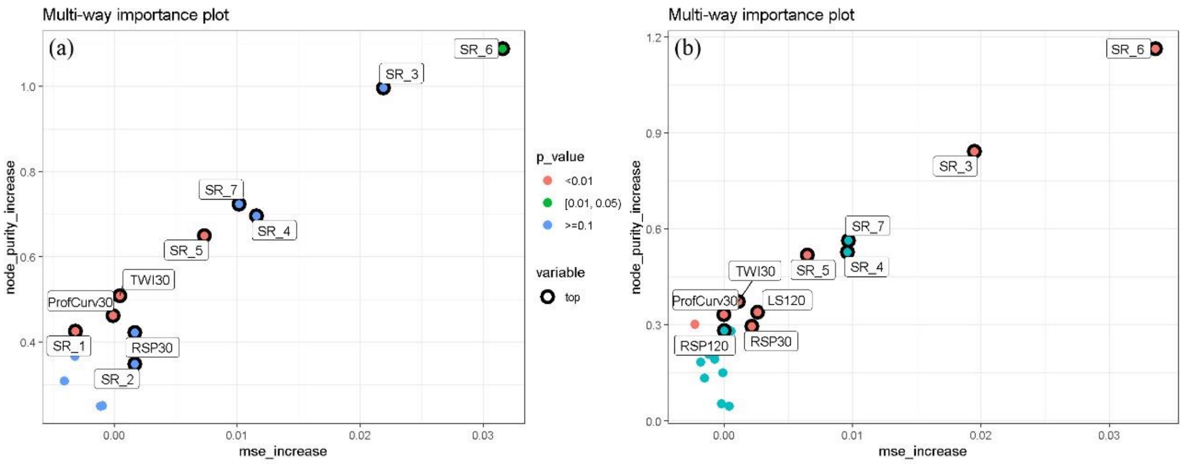


Figure 3.8. Variable prediction importance for pH at (a) original geomorphometric scale and (b) multiscale generalized geomorphometrics.

3.3.3. Prediction of superficial layer soil texture

This section discusses the Random Forest prediction of particle size, which is able to produce soil textural classifications for land users and stakeholders and will focus on the question of whether model adjustment can be improved by the MGG, and if the soil particle size maps result in a more accurate soil textural classification.

For prediction of particle size of surface layer, sand and silt fractions were used be-cause of the low contribution of clays in the Alter do Chão lithology. This could be explained by local characteristics of that geological formation, with an overall fine to medium grained sandstone content in the upper portion, and medium and coarse sand-stones with small contribution of red claystones in the lower portion (MENDES; TRUCKENBROD; RODRIGUES, 2012).

The Random Forest models for silt and sand, at original scale geomorphometrics, had a poor adjustment, as shown in Table 3.2. When considering the best adjusted MGG groups, despite the considerable portion of randomness related to the heterogeneity of soils, the % variation explained is reasonably higher and mean of squared residuals is reasonably lower. The model's predictors have a principal contribution from variables at 1:75.000 and 1:100.000 and can identify tendencies at the watershed scale.

Table 3.2. Model adjustment for silt and sand.

Particle Size	Predictors	% Var explained	Mean of squared residuals
Silt	Original scale	3.82	0.0007345
	MGG 1:75000+1:100000	31.73	0.0000674
Sand	Original Scale	6.9	0.0023651
	MGG 1:25000+1:75000	32.43	0.0002224

The most significant covariables of particle size prediction are shown in Figure 3.9 (a,b): Elev and RSP at coarser scales, associated with stratigraphy and long-term hillslope transportation; ProfCurv at coarser scales, and PlanCurv at original and coarser scales, related with accumulation, transit and dissipation zones (FLORINSKY et al., 2002).

The prediction of sand and silt content, at original and multiscale generalized geomorphometrics, is illustrated in Figure 10. In both variable groups, the MGG has produced maps with less noise and more recognizable patterns related to geomorphic features. These results corroborate the hypothesis that the topography has an influence, in a larger spatial context, and has prevalence on prediction of soil particle size contents in the tested basin. In contrast, a case study with Random Forest with 30m and 90m DEM did not achieve significant differences in prediction (BHERING et al., 2016). Despite some similarity with covariables importance, like Elev and RSP, the modeling is done on single scale datasets. In this sense, we can argue the importance of observing soil-landscape phenomena from a multiscale perspective.

The MGG was able to increase the accuracy of superficial layer soil texture classifications, as shown in Table 3.4. The most significant improvements occur in the franco-arenosa (MeAr) and areia franca (ArMe) soil texture classes, both with a smaller contribution area at the mapping site in relation to the more relevant areia (MAr) class. Also, the user´s accuracy has a considerably higher result, so the MGG increased the reliability of each mapped class. In the same way, the Kappa index also has higher values for MGG geomorphometric variables.

A case study done on a farm in China, with machine learning on single and multiple scales variables (SHI et al., 2018), also found better results with a range of appropriate scales, even using only local derivatives. In this sense, the MGG framework has greater potential because DEM transformation before derivate calculation allows for use of both local and regional derivatives in a multiscale group arrangement.

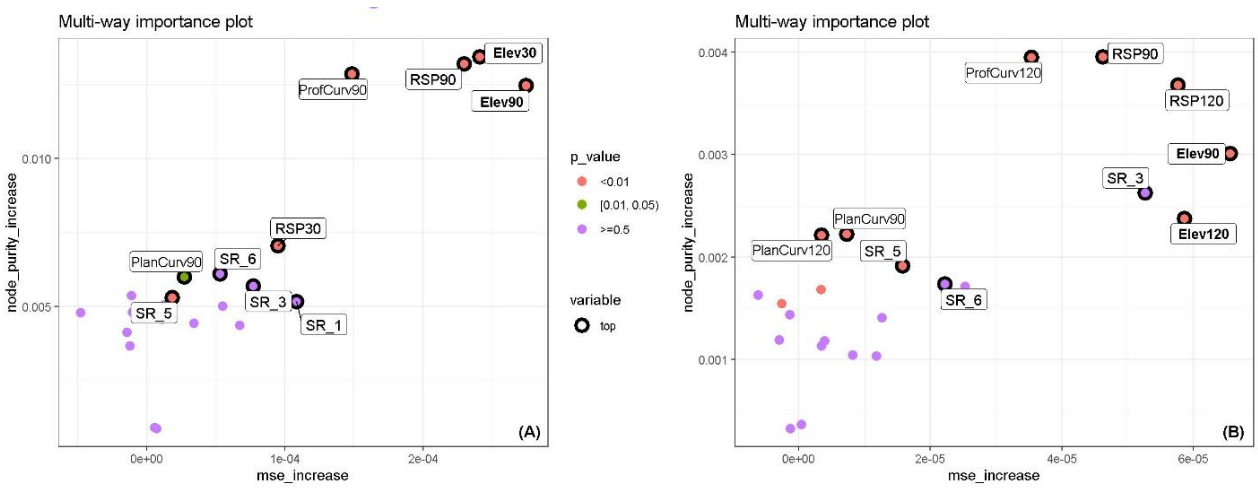


Figure 3.9. Variable prediction importance: (a) model of sand by geomorphometric group 1:25000 plus 1:75000; (b) model of silt by geomorphometric group 1:75000 plus 1:100000.

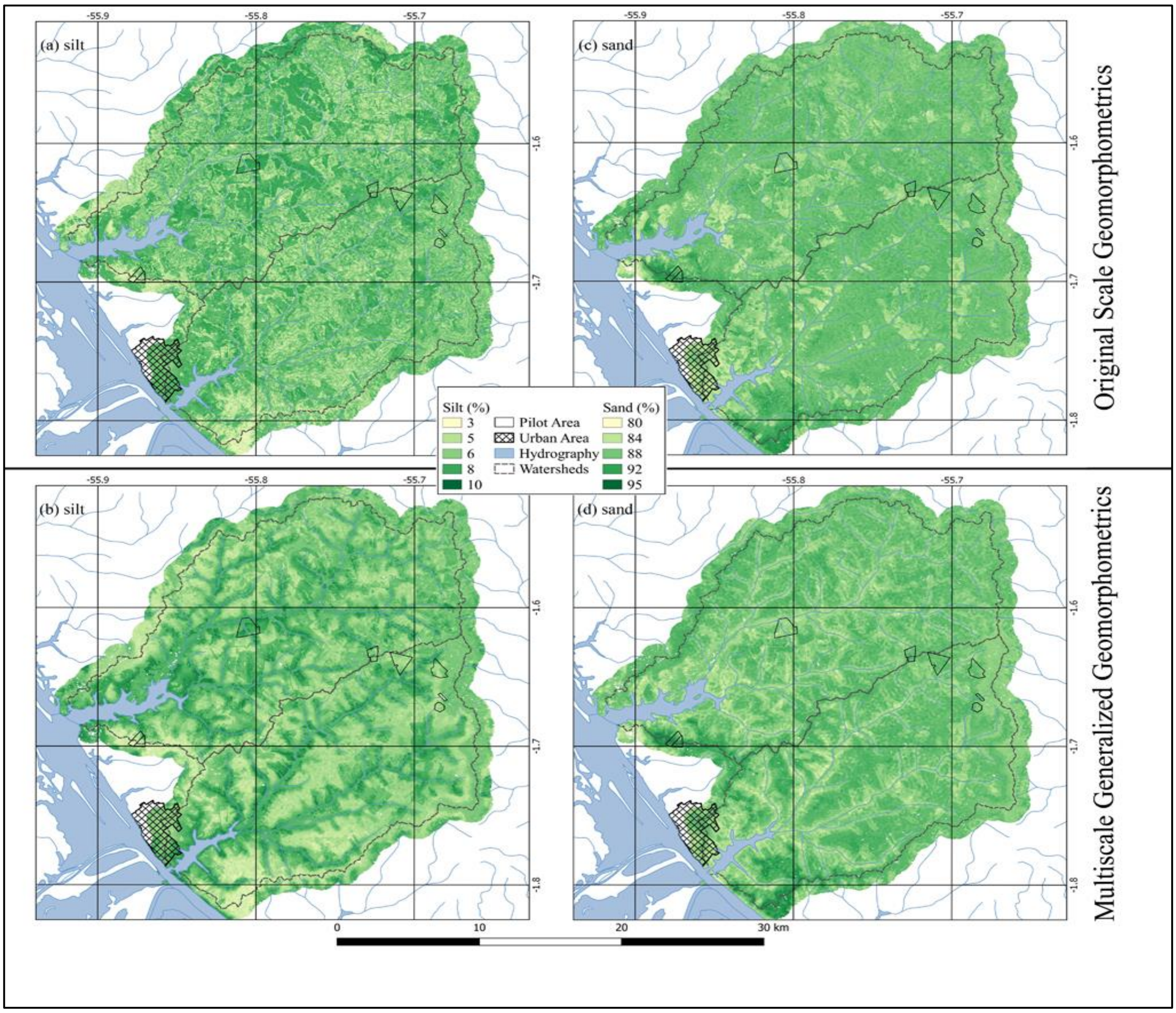


Figure 3.10. Predictive maps of silt (a,b) and sand (c,d) at original scale geomorphometrics and multiscale generalized geomorphometrics, respectively.

Table 3.3. Accuracy evaluation for soil texture classification.

Geomorphometric variables	User's / Producer's accuracy			Kappa Index
	MAr	ArMe	MeAr	All Classes
Original Scale	75% / 84%	72% / 58%	20% / 20%	0.43
MGG	81% / 88%	76% / 71%	100% / 67%	0.62

3.4. Conclusions

This study evaluates the multiscale geomorphometric generalization with the purpose of modeling the soil-landscape relationship. The comparison of original scale with multiscale generalized variables was based on Random Forest prediction of soil attributes. Regarding the proposed methodological framework, the following results and issues stand out from this case study:

The general geomorphometry generalization tends to smooth slopes and curvatures and produce identifiable representations of relative slope position at sub-watershed and watershed level. The specific geomorphometry generalization results in incorporation into valley portions some local landform segments which originally were classified in topo and slope.

Random Forest modeling with MGG variables did not mask greater importance of the vegetation factor for prediction of soil attributes related to landcover. The forest structures and effect of variables on prediction agree with pedological knowledge. Comparing these results with those from other studies, it could be argued that the soil-landscape phenomena from a multiscale perspective is highly relevant.

The MGG improved model adjustment for silt and sand particles and also improved the accuracy of metrics of soil texture classification of surface layer, especially for the most unusual classes, with the Kappa Index going from 0.43 to 0.62. Topography influences at a coarser spatial scale and has prevalence on prediction of soil particle size contents in the studied watershed.

Future development of the MGG framework should address generalization of DEM concerning feature preservation and comparison of landform classification adaptable at multiple scales.

3.5. References

ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433–459, 2010.

BEHRENS, T. et al. Multiscale contextual spatial modelling with the Gaussian scale space. **Geoderma**, v. 310, n. September 2017, p. 128–137, 2018.

BEHRENS, T. et al. The relevant range of scales for multi-scale contextual spatial modelling. **Scientific Reports**, v. 9, n. 1, p. 1–9, 2019.

BHERING, S. B. et al. Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa Agropecuaria Brasileira**, v. 51, n. 9, p. 1359–1370, 2016.

BISWAS, A.; ZHANG, Y. Sampling Designs for Validating Digital Soil Maps: A Review. **Pedosphere**, v. 28, n. 1, p. 1–15, 2018.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5–32, 2001.

BROCHADO, G. T. **Atenuação do efeito do desflorestamento em dados SRTM por meio de diferentes técnicas de interpolação**. [s.l.] Instituto Nacional de Pesquisas Espaciais, 2015.

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, v. 195–196, p. 111–121, 2013.

CONRAD, O. et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, v. 8, n. 7, p. 1991–2007, 2015.

CUNHA, G. O. DE M.; DE ALMEIDA, J. A.; BARBOZA, B. B. Relação entre o alumínio extraível com KCL e oxalato de amônio e a mineralogia da fração argila, em solos ácidos Brasileiros. **Revista Brasileira de Ciencia do Solo**, v. 38, n. 5, p. 1387–1401, 2014.

DE REU, J. et al. Application of the topographic position index to heterogeneous landscapes. **Geomorphology**, v. 186, p. 39–49, 2013.

EMBRAPA, C. N. DE P. DE S. **Manual de Métodos de Análise de Solo, 3a Edição**. Rio de Janeiro: EMBRAPA-CNPS, 2017.

FARR, T. G. et al. The Shuttle Radar Topography Mission. **Reviews of Geophysics**, v. 45, n. 2, p. RG2004, 19 maio 2007.

FLORINSKY, I. V. et al. Prediction of soil properties by digital terrain modelling. **Environmental Modelling & Software**, v. 17, n. 3, p. 295–311, 2002.

FLORINSKY, I. V. Influence of Topography on Soil Properties. Em: **Digital Terrain Analysis in Soil Science and Geology**. 2nd. ed. London: Elsevier, 2016. p. 482.

FLYNN, T. et al. Farm-scale soil patterns derived from automated terrain classification. **Catena**, v. 185, n. September 2019, p. 104311, 2020.

GRASS, D. T. **Geographic Resources Analysis Support System (GRASS) Software, Version 7.8**. Open Source Geospatial Foundation, , 2019.

GUILBERT, E.; BOGUSLAWSKI, P.; ISIKDAG, U. Multidimensional and multiscale GIS. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 6–8, 2019.

HARTEMINK, A. E. et al. Soil horizon variation: A review. Em: **Advances in Agronomy**. 1. ed. [s.l.] Elsevier Inc., 2020. v. 160p. 125–185.

HENGL, T. Finding the right pixel size. **Computers and Geosciences**, v. 32, n. 9, p. 1283–1298, 2006.

HU, G. R.; LI, X. Y.; YANG, X. F. The impact of micro-topography on the interplay of critical zone architecture and hydrological processes at the hillslope scale: Integrated geophysical and hydrological experiments on the Qinghai-Tibet Plateau. **Journal of Hydrology**, v. 583, n. January, p. 124618, 2020.

HUGGETT, R. J. Soil landscape systems: A model of soil Genesis. **Geoderma**, v. 13, n. 1, p. 1–22, 1975.

IBGE. **Mapa de Clima do Brasil**. Rio de JaneiroDiretoria de Geociências, , 2002.

IBGE. **Estado do Pará: Mapa de Geomorfologia**. IBGE, , 2008.

IBGE. **Manual técnico de geomorfologia**. 2ª ed. Rio de Janeiro: IBGE, 2009.

IBGE. **Manual Técnico de Pedologia**. 3ª ed. Rio de Janeiro: IBGE, 2015.

JASIEWICZ, J.; STEPINSKI, T. F. Geomorphons-a pattern recognition approach to classification and mapping of landforms. **Geomorphology**, v. 182, p. 147–156, 2013.

KÄMPF, N.; CURI, N. Formação e evolução do solo (pedogênese). Em: KER, J. C. et al. (Eds.). **Pedologia: fundamentos**. Viçosa, MG: SBCS, 2012. p. 207–302.

KETTLER, T. A.; DORAN, J. W.; GILBERT, T. L. Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses. **Soil Sci. Soc. Am. J**, v. 852, n. 13277, p. 849–852, 2001.

LAMICHHANE, S.; KUMAR, L.; WILSON, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. **Geoderma**, v. 352, n. January, p. 395–413, 2019.

LI, Z.; OPENSHAW, S. A Natural Principle for the Objective Generalization of Digital Maps. **Cartography and Geographic Information Systems**, v. 20, n. 1, p. 19–29, 1993.

LINDSAY, J. B.; FRANCIONI, A.; COCKBURN, J. M. H. LiDAR DEM smoothing and the preservation of drainage features. **Remote Sensing**, v. 11, n. 16, p. 17–19, 2019.

LIU, C.; FRAZIER, P.; KUMAR, L. Comparative assessment of the measures of thematic classification accuracy. **Remote Sensing of Environment**, v. 107, n. 4, p. 606–616, 2007.

LUCAS, Y. et al. The relation between biological activity of the rain forest and mineral composition of soils. **Science**, v. 260, n. 5107, p. 521–523, 1993.

MA, Y. et al. Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, v. 70, n. 2, p. 216–235, 2019.

MALONE, B. P.; MCBRATNEY, A. B.; MINASNY, B. Spatial Scaling for Digital Soil Mapping. **Soil Science Society of America Journal**, v. 77, n. 3, p. 890–902, 2013.

MALONE, B. P.; MINASNY, B.; MCBRATNEY, A. B. **Use R for Digital Soil Mapping**. [s.l.] Springer International Publishing, 2017.

MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, n. 1–2, p. 3–52, 2003.

MENDES, A. C.; TRUCKENBROD, W.; RODRIGUES, A. C. R. N. Análise faciológica da Formação Alter do Chão (Cretáceo, Bacia do Amazonas), próximo à cidade de Óbidos, Pará, Brasil. **Revista Brasileira de Geociências**, v. 42, n. 1, p. 39–57, 2012.

MILNE, G. Some suggested units of classification and mapping particularly for East African soils. **Soil Research**, v. 4, p. 183–198, 1935.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, v. 32, n. 9, p. 1378–1388, 2006.

NOVAIS, ROBERTO F.; MELLO, J. Relação Solo-Planta. Em: NOVAIS, R. F. et al. (Eds.). **Fertilidade do Solo**. [s.l.] SBCS, 2007.

PACHEPSKY, Y.; HILL, R. L. Scale and scaling in soils. **Geoderma**, v. 287, p. 4–30, 2017.

PALUSZYNSKA, A.; BIECEK, P.; JIANG, Y. **randomForestExplainer: Explaining and Visualizing**. R package, , 2019.

QGIS, D. T. **QGIS Geographic Information System**. Open Source Geospatial Foundation Project, , 2019. Disponível em: <http://qgis.osgeo.org>

R CORE TEAM, R. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, , 2019. Disponível em: <http://www.r-project.org/>

SAMUEL-ROSA, A. et al. Do more detailed environmental covariates deliver more accurate soil maps? **Geoderma**, v. 243–244, p. 214–227, 2015.

SANTOS, R. D. DOS et al. **Manual de descrição e coleta de solo no campo**. 7ª edição ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2015.

SANTOS, H. G. et al. **Sistema brasileiro de classificação de Solos**. 5ª ed. [s.l.] Sociedade Brasileira de Ciência do Solo, 2018.

SCHAEFER, C. E. G. R. et al. Solos da região amazônica. Em: CURI, N. et al. (Eds.). **Pedologia - Solos dos Biomas Brasileiros**. [s.l.] SBCS, 2017. p. 75–111.

SCHMIDT, J.; ANDREW, R. Multi-scale landform characterization. **Area**, v. 37, n. 3, p. 341–350, 2005.

SHI, J. et al. Machine-Learning Variables at Different Scales vs. Knowledge-based Variables for Mapping Multiple Soil Properties. **Soil Science Society of America Journal**, v. 82, n. 3, p. 645–656, 2018.

SMITH, M. P. et al. The effects of DEM resolution and neighborhood size on digital soil survey. **Geoderma**, v. 137, n. 1–2, p. 58–69, 2006.

SOUZA, C. M. P. DE et al. Multivariate Analysis and Machine Learning in Properties of Ultisols (Argissolos) of Brazilian Amazon. **Revista Brasileira de Ciência do Solo**, v. 42, n. 0, p. 1–20, 2018.

SZYPUŁA, B.; WIECZOREK, M. Geomorphometric relief classification with the k-median method in the Silesian Upland, southern Poland. **Frontiers of Earth Science**, v. 14, n. 1, p. 152–170, 2020.

TARGULIAN, V. O.; KRASILNIKOV, P. V. Soil system and pedogenic processes: Self-organization, time scales, and environmental significance. **Catena**, v. 71, n. 3, p. 373–381, 2007.

U.S. GEOLOGICAL SURVEY. Landsat 8 Surface Reflectance Code (LASRC) Poduct Guide. (No. LSDS-1368 Version 2.0). n. May, p. 40, 2019.

WADOUX, A. M. J. C.; MINASNY, B.; MCBRATNEY, A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. **Earth-Science Reviews**, v. 210, n. January, p. 103359, 2020.

WU, Q. et al. An effective parallelization algorithm for DEM generalization based on CUDA. **Environmental Modelling and Software**, v. 114, n. December 2018, p. 64–74, 2019.

WYSOCKI, D. A.; SCHOENEBERGER, P. J. Geomorphology of Soil Landscapes. Em: HUANG, P. M.; LI, Y.; SUMNER, M. E. (Eds.). **Handbook of Soil Science: Properties and Processes**. [s.l.] Chemical Rubber Company Press, Boca Raton, Florida, 2011. p. 1–26.

ZHOU, Q.; CHEN, Y. Generalization of DEM for terrain analysis using a compound method. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 1, p. 38–45, 2011.

ZINCK, J. A. The Geomorphic Landscape: Criteria for Classifying Geoforms. Em: **Geopedology**. Cham: Springer International Publishing, 2016. p. 77–99.

CAPÍTULO II. MAPPING PEDOGENETIC PROCESSES IN AN AMAZONIAN WATERSHED BY FUZZY LOGIC APPLIED TO MULTISCALE GENERALIZED GEOMORPHOMETRICS.

Highlights

Gleization occurs in a well-defined position at the bottom of valleys.

Elutriation is the prevalent process resulting in texture gradients, which is widespread on slopes.

Clay translocation has a minor effect and is restricted to lower concave positions.

Generalized geomorphometry variable selection improves accuracy of soil class maps.

Abstract

Soil formation factors determine the conditions of soil formation processes and their overlap over time results in soil properties in the pedon. These pedons can be described in terms of soil classes. However, the relationships are complex, on the one hand different processes can result in similar properties, and on the other hand the same process can result in different soil classes depending on their prevalence and the extent of changes in soil properties in the pedon. The multiscale influence of topography on soil distribution has two general aspects, overlay of pedological processes that occurred at different times, and driving forces in the present time, determined by the sum of forces better correlated with one, several, or many geomorphologic scales. The hypothesis tested in this study is whether the multiscale analysis of topography contributes to the definition of preferential environments for the occurrence of the pedogenetic processes of gleization, elutriation and clay translocation, as opposed to the spatiotemporally wide distribution of the ferralitization process. The objective of this research was to map the soil classes in two watersheds at a scale of 1:100,000. The basins of this study are homogeneous in terms of geological basement, with a predominance of sandstones, comprising the *Alter do Chão* formation. To achieve this objective, fuzzy set models were applied using a multiscale geomorphometric database, generalized from the SRTM digital elevation model, to predict the distribution of soil classes. Fuzzy membership rules were defined for each of the pedological processes, considering pedological knowledge obtained from soil surveys in transects, to define the membership curves and scales of the covariates. The maps were deffuzyfied, defining the soil class of each pixel. The maps were evaluated using confusion matrices, based on validation points defined by a conditioned Latin hypercube. With generalized covariates the overall accuracy was 84%, compared to 74% of the mapping considering only the covariates in the original scale. Considering the multiscale covariates, the following user and producer accuracies were obtained: *GXbd* (88%, 78%); undifferentiated group *PAd* and *LAdarg* (60%, 60%); undifferentiated group *LAd* and *LAdpsa* (88%, 92%). The results demonstrate that knowledge-based models that observe the multiscale influence of topography on soil formation processes can improve the prediction of soil classes in intermediate-level soil surveys.

Keywords: knowledge-based models; soil-landscape relationship; geomorphometric upscaling; soil forming process; soil survey.

4.1. Introduction

The soils of the Amazon basin are predominantly formed by strong chemical weathering processes typical of a warm and moist climate acting on a sedimentary geological basement over a long time of stability in environmental conditions (SCHAEFER et al., 2017). As a result, the long-term pedological processes are the ones that are most spatially distributed, with emphasis on ferralitization. However, other processes can become prevalent or significant and determine important changes in a given portion of soil under specific environmental conditions, like gleization, elutriation and clay translocation (KÄMPF; CURI, 2012). Therefore, in order to know the patterns of soil distribution, it is necessary to identify and delineate the preferred environments for the occurrence of the different pedological processes. In this sense, quantification methods have made considerable progress in predicting soil processes and related soil horizon variations across landscapes (HARTEMINK et al., 2020).

Topography is the main variable to define environments where different pedological processes can occur (FLORINSKY, 2016). The soil-landscape relationship is based on the concepts of the catena and hillslope models (HUGGETT, 1975; MILNE, 1935) and the essential concept of soil science established by Simonson (SIMONSON, 1959): soil forming factors, soil-forming processes, and resulting soil properties. When considering geomorphic development and the multiscale topography influences, two general aspects can be highlighted: overlay of pedological processes that occurred at different times (WYSOCKI; SCHOENEBERGER, 2011); and driving forces, determined by the sum of forces better correlated with one, several, or many geomorphologic scales (FLORINSKY, 2016). The classical mental models of pedologists generally include, even if tacitly, observations at different levels of relief generalization in the construction of soil-landscape rules, at various stages of knowledge development (BUI, 2004), which includes features resulting from geomorphic evolution and flexible selection of scales to determine terrain parameters.

In digital soil mapping applications, process-based modeling approaches have recently received attention, despite greater inherent complexity compared to models based on property prediction (LOZBENEV et al., 2021; MA et al., 2019a). The challenges that exist include site-specific dynamics, high uncertainty, or impossibility of direct observations under uncontrolled field conditions, need to infer pedological processes through soil properties, and ambiguity in the relationship between properties and processes. Despite this, process-based approaches can help in gaining new insights about soil processes by observing mapping sites, thus promoting feedback between digital soil mapping and pedology. The fuzzy set approach to knowledge-based modeling (MCBRATNEY; ODEH, 1997; ZHU, 1997) for process-based soil mapping is especially useful due to the handling of uncertainties, of qualitative or semi-quantitative information, and the use of multiple membership rules which can translate complex environmental conditions.

To achieve the task of producing digital soil maps of the Amazon region at a high level of resolution, the pedological processes in each pedoenviroment need to be modeled at a level of detail that could be useful for a low-cost soil survey approach, resulting in soil class maps at the subgroup categorical level. Furthermore, this pathway of digital soil mapping contributes to a better pedogenetic understanding of the soil in the landscape, which is a significant challenge to soil science in a pedometric context (ARROUAYS et al., 2020; WADOUX et al., 2021).

This study aimed to incorporate multiscale geomorphometry to a knowledge-based fuzzy modeling framework using a multiscale-generalized covariable database (ARAÚJO; JUNIOR; BELDINI, 2021). This methodology brings to the pedometric perspective the understanding that the soil-landscape relationship occurs through complex and multiscale interactions, allowing the formalization of the selection of scales according to the expert's knowledge. In this sense, the hypothesis tested in this study is whether the multiscale analysis of topography contributes to the definition of preferential environments for the occurrence of the pedogenetic processes of gleization, elutriation and clay translocation, as opposed to the spatiotemporally wide distribution of the ferralitization process.

4.2. Material and methods

The procedures described in this section were performed using the open-source software QGIS 3.16; SAGA GIS 2.3; GRASS GIS 7.8; and R Programming 3.5 (CONRAD et al., 2015; GRASS, 2019; QGIS, 2019; R CORE TEAM, 2019).

4.2.1. Study Area

The study was conducted in the Iripixi Lake (ILW) and Caipuru Lake (CLW) watersheds, with an area of 27,137 ha and 28,315 ha respectively, located in the Trombetas River basin in Oriximiná-Pará in the Eastern Amazon, as shown in **Figure 4.1**.

The Alter do Chão formation is a Cretaceous sedimentary deposit (CPRM, 2008). Locally, in western Pará, the formation process of Alter do Chão took place in a depositional environment of a sinuous fluvial system, resulting in the record of a succession of sandstones, conglomerates and mudstones (MENDES; TRUCKENBROD; RODRIGUES, 2012). It was formed approximately 135 million years ago and has evolved in accord with tectonic motion ever since (SOMOZA; GHIDELLA, 2012). Later, in the context of the beginning of the Andean uplift, another depositional formation buried the Alter do Chão formation in the Paleocene, approximately 55 million years ago. The movement resulting from the epeirogenesis of the South American continent and the variations in the global level of the seas caused the dissection and almost complete destruction of this Paleogenic Detritus-Lateritic Coverage in the Oligocene, 30 million years ago. In the final phase of geological evolution, from the Miocene, approximately 10 million years ago, continental movements related to the advanced stages of the Andean uplift and the variations in the global level of the ocean due to glaciations reduced the base level and promoted the dissection of the Alter do Chão formation.

From the Quaternary onwards, 1.4 million years ago, the region went through a period of geological stability (SOMOZA; GHIDELLA, 2012). In this sense, we can infer that geomorphological and pedological contemporary evolutionary dynamics have been developing since then. However, in this period there were at least three major reductions in the global ocean level that may have initiated erosive processes, soil rejuvenation and reorganization of drainage patterns (BRIDGLAND, 2021). The geomorphic units of these watersheds are classified as a homogeneous dissection with coarse drainage density and weak incision depth (IBGE, 2008, 2009).

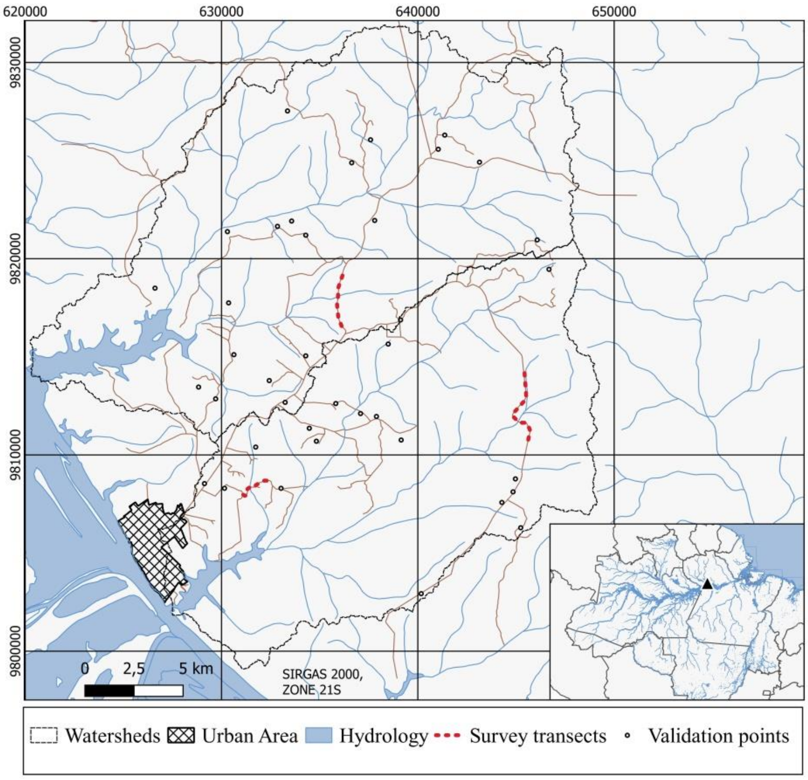


Figure 4.1. Study area location in the eastern Amazon.

4.2.2. Topography Covariates

In the proposed mapping scale, topography factors are the main sources of soil variation, and in this study, the Shuttle Radar Topography Mission Digital Elevation Model 30 m (SRTM DEM 30 m) (FARR et al., 2007) was used to represent such factors. The corrections made in SRTM DEM were filling in sinks, and reduction of deforestation effects by the estimated canopy addition method (BROCHADO, 2015). The topography information was upscaled and organized into generalized multiscale geomorphometric variable groups, as detailed in the next section.

4.2.3. Multiscale Geomorphometric Generalization

Multiscale geomorphometric generalization (MGG) is an upscaling operation based on cartographic concepts of generalization of digital maps (LI; OPENSHAW, 1993). This approach can be applied to any geomorphic variable, including elevation models, landforms units, and primary and secondary derivatives. This operation results in variables at different scales, arranged in groups according to criteria required for the analysis. In this sense, the formalization of the desirable scales of analysis and modeling occurs both in their definition and in the group arrangements.

For the MGG operation, vector and raster representations demand different approaches for upscaling, because each of them has specific scale transformation problems due to their mathematical structures. Furthermore, it is necessary to have a unique reference for the scales for compatible representation of geomorphic features in both types of variables, thus allowing for joint interpretation. In this study, the concept of minimum mappable area for soil surveys (IBGE, 2015) was considered to define pixel sizes in relation to cartographic scale. The detailed descriptions for each of the four scales used are in **Table 4.1**. The area equivalence between raster and vector is calculated as a function of a 5 × 5 pixel grid, considered to be a conservative parameter to determine a geomorphic feature (FLORINSKY, 2017). These geomorphic covariables were obtained by upscaling methods, as illustrated in **Figure 4.2**, wherein local averages of DEM and subsequent derivatives were used in the covariable calculation, and these were adjusted using the inner search radius for geomorphons classifications (JASIEWICZ; STEPINSKI, 2013). Such methods correspond to cartographic generalization applied to general geomorphometry and specific geomorphometry, respectively (ZINCK, 2016).

Table 4.1. Correspondence between scale and pixel size for Multiscale Geomorphometric Generalization (MGG), using the concept of minimum mappable area.

Scale	Minimal mappable area (m^2)	Pixel size (m)	Pixel area* (m^2)	pa/mma (%)
1:25000	25000	30	22500	90
1:50000	100000	60	90000	90
1:75000	225000	90	202500	90
1:100000	400000	120	360000	90

* For a 5x5 window

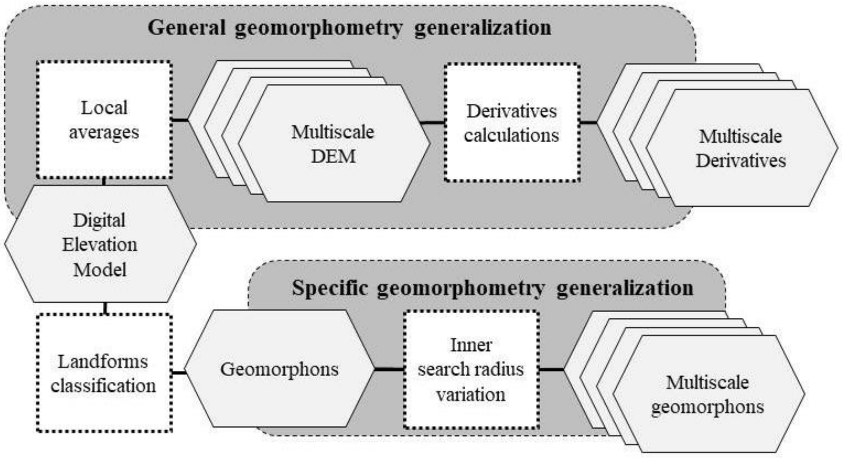


Figure 4.2. Methodology flowchart of MGG for the topography covariables.

4.2.4. Soil survey and classification

The soil survey was done to acquire knowledge for model construction, by empirical observation. Soil profile sampling was carried out in 3 transects, covering the lower, middle and upper portions of the hydrographic basins (**Figure 1**). An independent sampling dataset was used for validation, and extra observation points were analyzed in a random-stratified arrangement, distributed according to the terrain covariates in the original scale by a cost-constrained conditioned Latin hypercube method (ROUDIER; HEWITT; BEAUDETTE, 2012; SILVA et al., 2014). The watersheds were sampled at 18 and 20 points, respectively, a sufficient density for high resolution level soil surveys, compatible with 1:100,000 scale soil maps (IBGE, 2015).

The morphological description of the soil profile was performed up to 2 meters deep (SANTOS et al., 2015) in selected positions of the transects and in the extra validation points. Soil samples were collected from the horizons and their subdivisions for complementary physical and chemical analyses, necessary for the discrimination of soil classes (EMBRAPA, 2017).

The Brazilian Soil Classification System (SiBCS) is a hierarchical system designed to cover all soils present in the national territory (SANTOS et al., 2018). The categorical level structure determines the classification into orders, suborders, large groups, subgroups, families, and series. The prevalent pedological process determines the order, and minor environmental-pedological characteristics define lower categorical levels. In this sense, the soil classification of a pedon could vary despite the occurrence of the same pedological processes. The main criteria used for pedon classification are related to the level of evolution of soil characteristics and properties, produced by one or another soil processes. In this study the soil profiles were classified up to the 4th categorical level of the subgroups.

4.2.5. Fuzzy modeling and production-evaluation of soil maps

Fuzzy logic modeling is an approach capable of dealing with uncertainties, as the relationship between elements and fuzzy sets are expressed by degrees of membership (NICOLETTI; CAMARGO, 2004). Fuzzy logic modeling was used in several studies related to knowledge-driven soil mapping (MA et al., 2019b; MCBRATNEY; ODEH, 1997; MENEZES et al., 2013, 2014; SILVA et al., 2014).

In this study, the fuzzy modeling process and soil map production was carried out as shown in **Figure 4.3**. The definition of intervals, the shape of the pertinence curves, and the scales of the considered covariates was established according to the knowledge acquired in the soil survey along transects. The pertinence curves represent the favorable environmental conditions for the occurrence of the target pedogenetic processes of the study. The pedogenetic processes considered were gleization, elutriation, clay translocation and ferralitization. The defuzzification and production of soil class maps were performed selecting the highest pertinence value and observing the classification provided in the SiBCS. The soil class maps were evaluated in relation to validation points by a confusion matrix, considering the kappa index, global accuracy, and producer and user accuracy.

The main hypothesis of this study was that in dissection modeled on sedimentary geological formation, topography expressed through generalized multiscale covariates of slope, curvature, and moisture content (landscape factor) controls the redistribution of moisture via water flow along hillsopes (landscape process). This, in turn, controls the processes of clay translocation, elutriation and gleization (soil formation processes), resulting in the differentiation of pedons (soil properties) and in the distribution of soil classes.

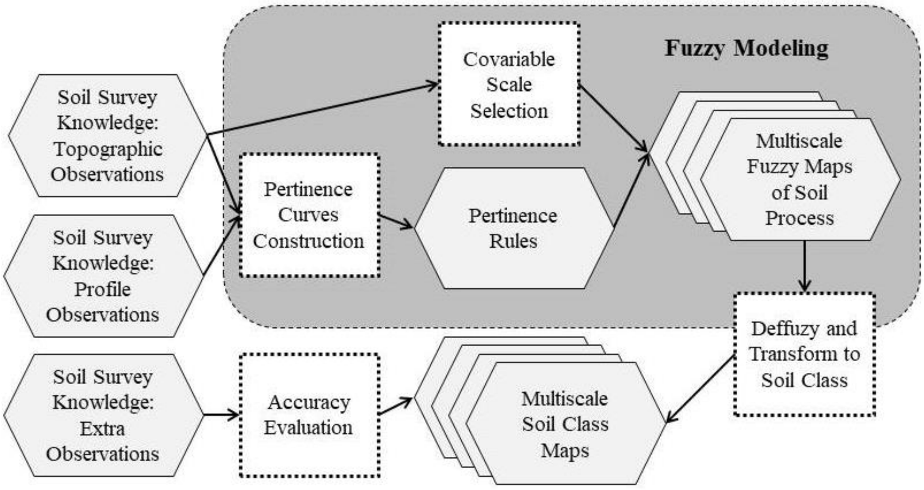


Figure 4.3. Methodology flowchart of Fuzzy Modeling with Covariable Scale Selection.

4.3. Results

The survey of soils in the transects allowed for the construction of a mental model of the soil-landscape relationship, explaining the link between topographic, landscape process, landscape properties, pedogenetic processes and resulting soil profiles, as shown in **Table 4.2**.

Ferralitization is the most widespread pedogenetic process in the landscape, due to the long period of weathering that has occurred in this area of geological stability. Considering the geomorphological evolution of a dissection model in this geological context, it is possible to affirm that the weathering mantle was formed prior to erosive dissection, resulting in a wide distribution pattern of *Oxisols*, eventually reorganized into differentiated profiles. It should be noted that the predominance of sandstones produced soils with high contents of the sand fraction, close to the threshold between *LATOSSOLO AMARELO Distrófico tipico* and *LATOSSOLO AMARELO Distrófico psamitico*. In this sense, both profiles were mapped as an undifferentiated group (LAd + LAdpsa).

Elutriation is the most relevant process for the reorganization of profiles in the mapped basins, occurring primarily under high runoff conditions. This selective erosion process, carrying mainly finer particles under conditions of forest vegetation cover, occurs in the high and convex portions of the slope, due to the greater potential energy of water in such positions. The elutriation process gradually removes clay from the upper layers, removing them from the profile, and eventually produces a textural gradient. Clay translocation is a relevant but localized process for the reorganization of profiles in the mapped basins, occurring primarily under high percolation conditions. This process of vertical movement of clays in the profile has a low distribution due to the reduced contents of clays in the geological basement. Therefore, the differentiation that results in textural gradients arising from the accumulation of clays in the lower layers is restricted to portions with intense water movement. In these places it was possible to notice the presence of weak clay films, which is the main indication of the predominance of clay translocation in such profiles. For both processes, the current stage of such evolution may or may not meet the requirements for classification as an *Argissolo*. In this sense, the *ARGISSOLO AMARELO Distrófico típico* and *LATOSSOLO AMARELO Distrófico argissólico* soils are mapped as an undifferentiated group (PAd + LAdarg).

The process of gleization takes place in a relatively quick period, unlike the three processes described above. In this sense, the reorganization of the profiles arising from the process of water saturation in a redoxymorphic environment is strongly correlated with the currently saturated zones. Therefore, gleization is predominant in the bottoms of flat valleys and in low areas under the influence of the water table. For this process, the soils are mapped as *GLEISSOLO HÁPLICO Tb Distrófico típico* (GXbd).

Table 4.2. Correspondence between Landscape and Soil according to the nested triad Factor-Process-Property.

Landscape Factor (Topographic)	**Landscape Process**	**Landscape Property / Soil Factor**	**Soil Process**	**Soil Property (SiBCS Soil Classes)**
Flat valley bottoms, low and wet areas		Most-time saturated soil	Gleization	GLEISSOLO HÁPLICO Tb Distrófico típico (GXbd)
Shoulder and upper convex positions	Water infiltration and redistribution	High runoff	Elutriation	ARGISSOLO AMARELO Distrófico típico (PAd); LATOSSOLO AMARELO Distrófico argissólico (LAdarg)
Lower concave positions		High percolation	Clay Translocation	ARGISSOLO AMARELO Distrófico típico (PAd); LATOSSOLO AMARELO Distrófico argissólico (LAdarg)

Widespread on top, slopes, and foothills	Long term unsaturated areas	Ferralitization	LATOSSOLO AMARELO Distrófico típico (LAd); LATOSSOLO AMARELO Distrófico psamítico (LAdpsa)

Soil fuzzy modeling was done for each soil process, resulting in one or more pertinence curves, and all of these curves are sigmoidal, with the adjustment points presented in **Table 4.3**. Additionally, covariates with different generalization scales were selected for application in the same curves, as shown in **Table 4.4**, and the generalization geomorphometric effects were exemplified in 3D view (**Figure 4.4)**. Therefore, this methodological structure allowed for the formalization of decisions based on the pedologist's knowledge regarding the selection of scales, relevance rules and considering a compartmentalization of the models according to the relief forms.

The soil wetness conditions (dry, wet or waterlogged), were principal aspects in fuzzy models, and therefore the topographic wetness index (TWI) was selected as the main geomorphometric covariable for distinguishing the occurrence conditions of soil processes. Other covariates were selected as important determinants of the dynamics of humidity in the landscape, including the potential energy compared to the base level; tendencies to gain or lose speed; and converging or diverging flows. In this sense, the covariates' slope, plan curvature, profile curvature and relative slope position were used.

Table 4.3. Landform segmentation and Fuzzy pertinence rules for gleization, elutriation, clay translocation and ferralitization process.

<table>
<tr><th>Pertinence Rules</th><th>Geomorphons masks</th><th>Pertinence Curves</th></tr>
<tr><td>R1. Gleization</td><td>Valley</td><td>TWI: sigmoidal; left-tailed; A=9; B=10</td></tr>
<tr><td>R2. Elutriation</td><td rowspan="2">Summit, Ridge; Shoulder; Spur; Slope</td><td>PlanCurv: two-tailed; A=-0.0005; B=0; C=0.001; D=0.0015
RSP: two-tailed; A=0.4; B=0.6; C=0.8; D=1
TWI: right-tailed; C=8; D=10</td></tr>
<tr><td>R3. Elutriation</td><td>Slope: two-tailed; A=0; B=0.07; C=0.15; D=0.18
TWI: left-tailed; A=2; B=12</td></tr>
</table>

R4. Clay Translocation	Slope, Hollow, Footslope, Valley	PlanCurv: right-tailed; C=-0.003; D=0 ProfCurv: right-tailed; C=-0.003; D=0 TWI: two-tailed; A=0; B=5; C=8; D=15
R5. Ferralitization	All (no mask)	TWI: right-tailed; C=7; D=12 Slope: right-tailed; C=0.1; D=0.2

Table 4.4. Generalized geomorphometric covariables selection for pertinence rules.

Pertinence Rules	**Covariables (Generalized Scales)**
R1	Geomorphons (1:75,000); TWI (1:50,000)
R2	Geomorphons, Plancurv, RSP, TWI (1:75,000)
R3	Geomorphons, Slope, TWI (1:75,000)
R4	Geomorphons, TWI (1:75,000); PlanCurv, ProfCurv (1:100,000)
R5	TWI, Slope (1:100,000)

The gleization process occurs in the valleys. In these segments, the process occurs at high levels of soil moisture, with TWI > 9, that is, conditions of accumulation of humidity and in a position in the landscape which makes drainage difficult. The adoption of generalized covariates for valley segmentation and moisture contents, at 1:75,000 and 1:50,000 scales respectively, produced maps where the probable occurrence of gleization in small valleys and headwaters was reduced. On the other hand, it favored the appearance of significant gleization spots in the broad valleys of the main channels, as seen in **Figure 4.5a**.

The elutriation process occurs on summits, ridges, shoulders, spurs and slopes. In these segments, two pertinence rules were constructed to encompass preference situations. In rule 2, the most likely areas are related to flat and convex areas; and with low and intermediate wetting rates; and on relative slope positions between the middle and upper thirds. In rule 3, the model expresses that the highest probability occurs in conditions of high slope, especially between 8% and 15%; and smooth-progressively increasing in areas with higher soil wetness. That is, conditions with water content and energy availability, promoting the occurrence of surface runoff. The adoption of generalized covariates for geomorphons segmentation, slope, TWI, and curvatures, at 1:75,000 and 1:100,000 scales respectively, produced maps where the probable occurrence of elutriation presented higher indices in upper convex portions, considering the large features in the basin, with emphasis on the interfluves of basins and subbasins, as see in Figure 5b.

The clay translocation process occurs in valleys, slopes, footslopes and hollows. In these segments, the process is more likely to occur when associated with high levels of humidity, without saturation; and in portions of convergence of flows, with concave plan curvatures; and in portions of flow deceleration, with concave profile curvatures. That is, conditions with significant water flow, converging laterally and losing kinetic energy, primarily promoting vertical flow. The adoption of generalized covariates for geomorphons segmentation, TWI, and curvatures, at 1:75,000 and 1:100,000 scales respectively, produced maps where the probable occurrence of clay translocation in footslopes and low concave positions appears in rounded spots with higher values on the central areas. With use only of geomorphometric covariables at the original scale, clay translocation does not occur in these zones, but only with a noise-like pixel distribution pattern (Figure 5c).

The ferralitization process is widespread across the landscape and is more likely to occur associated with lowest slope inclinations, above 10%; and dry or intermediate wetness conditions, especially with TWI < 7. That is, conditions that promote the deepening of the weathering mantle. The adoption of generalized covariates for slope and TWI, at a 1:100,000 scale, produced maps where the probable occurrence of ferralitization, considering the large features in the basin, with lowest values at dissection areas of interfluves of basins and sub-basins, as see in Figure 5d.

Figure 4.4. 3D View of covariables on original and generalized scales.

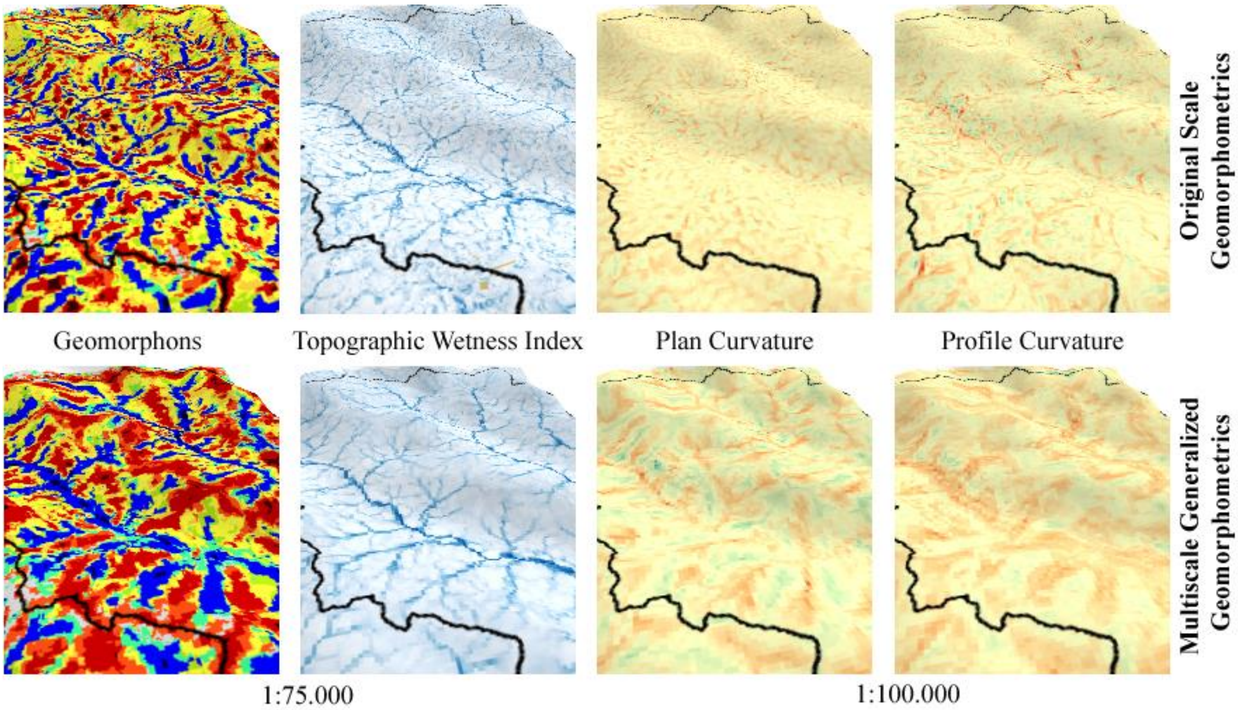
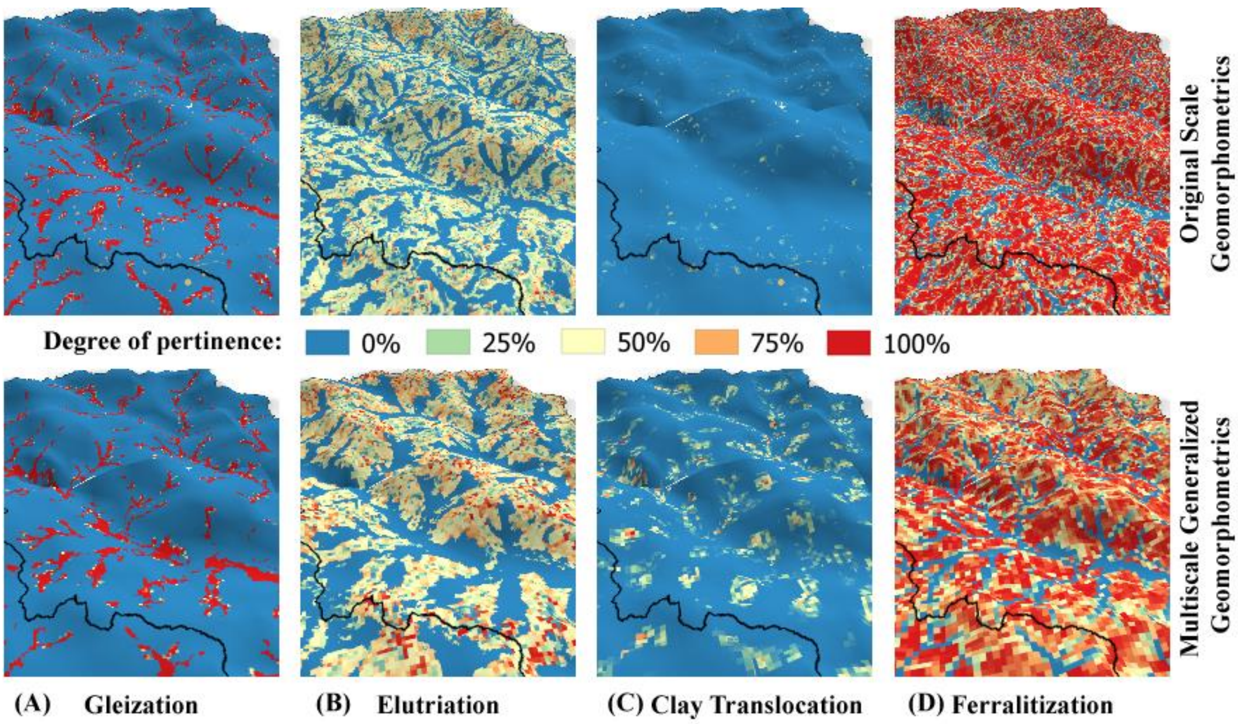


Figure 4.5. 3D View of fuzzy pertinence maps for: (A) Gleization, (B) Elutriation, (C) Clay Translocation, (D) Ferralitization.



The soil class maps were produced for original scale and multiscale generalized geomorphometrics, as seen in **Figure 4.6**, and resulted in one pure soil unit and two undifferentiated groups: *GXbd,* with a gley horizon; *PAd + LAdarg*, with a textural gradient; *LAd + LAdpsa*, with a Bw horizon. The adoption of generalized covariates in the fuzzy model rules produces more relevant patches with *PAd+LAdarg* and *GXbd*. In this sense, the prevalence of soil patches was noted at a large scale on ridges, or convex high geoformations, and footslopes or hollows. On the other hand, the GXbd soil patches appeared in flat and larger valleys.

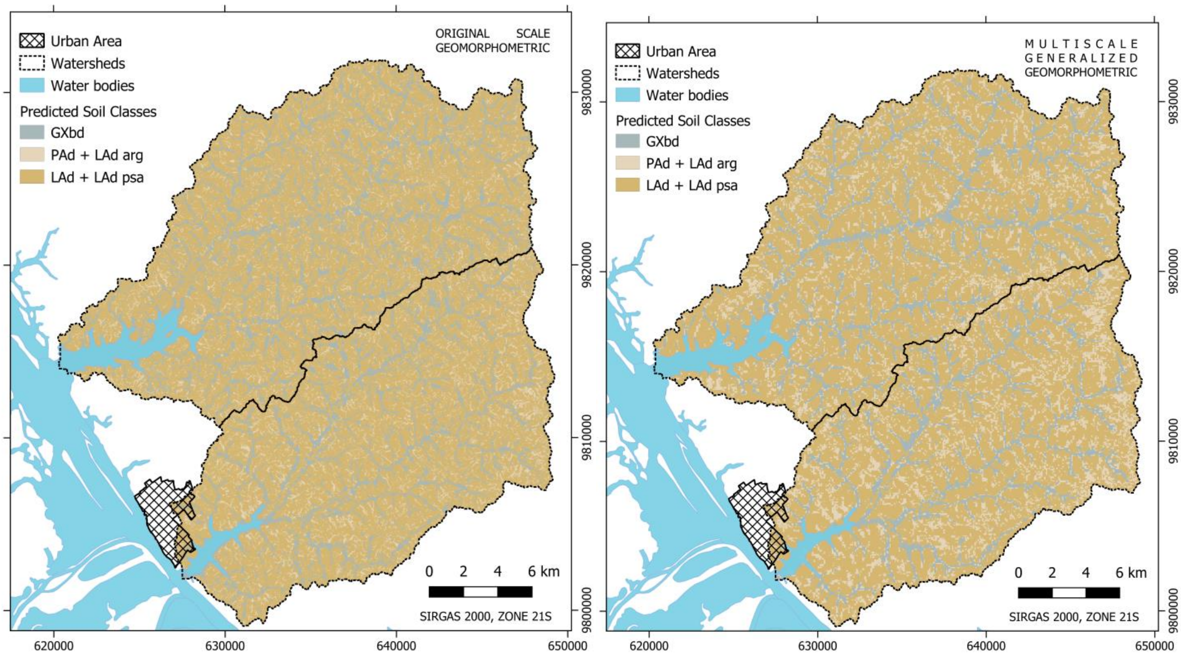


Figure 4.6. Soil class maps produced by fuzzy modeling with (a) original scale geomorphometric and (b) multiscale generalized geomorphometric covariates.

The evaluation of soil class maps reveals the increase in general accuracy and user and producer accuracies for the minority of soil classes with the use of generalized geomorphometric covariates, as shown in **Table 4.5**. The overall accuracy was higher in models that used generalized covariates. It is important to highlight the increase in the reliability of the maps in the less prevalent classes, represented by the highest values for the user’s accuracy of GXbd and PAd+LAdarg. Despite some restrictions on the use of the Kappa Index (PONTIUS; MILLONES, 2011), it is worth noting the significant increase in this measure, from fair agreement to substantial agreement. This pronounced effect is justified, in part, by the prevalence of *Latossolos* and *Latossolos psamiticos*, making these classes more prone to "random hits". When this characteristic is considered, through the Kappa index, the benefits obtained in the minority classes are highlighted.

Table 4.5. Accuracy evaluation for soil classification at the 4th categorical level of SiBCS.

Geomorphometric variables	User's/Producer's accuracy			Overall accuracy	Kappa Index
	GXbd	PAd+LAdarg	LAd+LAdpsa	All Classes	
Original Scale	50%/67%	20%/50%	92%/77%	74%	0.40
MGG	88%/78%	60%/60%	88%/92%	84%	0.69

4.4. Discussion

Using the high-intensity resolution 1:100,000 mapping scale proposed in this study, topography is the main source of variation. In this sense, digital soil mapping at this scale can collaborate with the understanding of soil-landscape relationships, specifically regarding the soil property patterns resulting from the hydrology of hillslopes.

4.4.1. Textural gradients

Textural differentiation, resulting in textural gradients in the soil profile, is a central aspect for mapping soils in watersheds located on the Alter-do-Chão geological formation. In this pedoenvironment, this characteristic may originate from two distinct pedological processes (KÄMPF; CURI, 2012), namely: clay translocation, or lessivage, and elutriation. Clay translocation refers to the suspended movement of particles of the clay fraction, mainly fine clays <0.002mm, oxides and organic compounds, inside the soils. Clays remain in suspension and are transported, through the macropores, where the flows are sufficiently rapid. Deposition occurs where dispersion and transport are less effective. Mobilization occurs by crumbling, separation of the aggregates, by wetting-drying processes, or by dispersion. Elutriation is a process of removing fine particles by selective erosion, common under tropical forests. Elutriation is caused by dripping, causing dispersion, and surface flow of water, causing removal of fine material into the drainage system. It may be associated with the removal of fine silty material and its absence in the B horizon.

The formation of soil textural gradients in the mapped watersheds does not have its predominant origin in the clay translocation process in most of the slope positions, except in the lower concave ones. Morphologically, in relation to the profiles sampled, this statement is supported by two direct and indirect examples, respectively: (a) the sub-surface horizons do not present clay films, denoting the small contribution of clay translocated from the upper horizons, and (b) the clay contents in the subsurface horizons in the *PAd+LAdarg* profiles are not significantly higher than those found in the subsurface horizons in the *LAd* profiles, denoting the absence of significant accumulation of clays in the context of the soil-landscape relationship present in this hydrographic basin. On the other hand, in the lower concave positions, soil profiles with textural gradients were surveyed that presented common and weak clay films. This evidence is supported by the tendency of high percolation in this segment of the landscape (WYSOCKI; SCHOENEBERGER, 2011), indicating the occurrence of clay translocation in these environments, even from a geological base poor in clays. Therefore, we can assume that the elutriation process is a widespread mechanism in different positions in the landscape, which would explain the significant occurrence of profiles with textural gradients in areas with higher runoff (MÖLLER; VOLK, 2015).

4.4.2. Glei horizon

Soil that is saturated, or with very high levels of moisture, creates an environment where the influx of O_2 from the atmosphere is very difficult, due to the absence of free pore space. The absence of O_2 favors reduction of complex organic compounds as well as anaerobic microorganisms that use metals as electron acceptors. In this way, Fe ions are reduced and released from oxides, migrating to the soil solution. Therefore, Fe loss zones turn grayish, eventually forming Glei horizons. These ions in solution migrate by diffusion to environments of lower redox potential, where they are oxidized and precipitate (KÄMPF; CURI, 2012).

The Fe accumulation zones, marked by concretions or mottles, require thousands of years for their formation. However, the time required to form a well-developed glei horizon is only a few years or decades(KÄMPF; CURI, 2012). Considering this geomorphological time scale, this time lapse can be understood as "present time". In this sense, it is expected that the patterns of occurrence of soils with a glei horizon are adjusted to the current disposition of the valleys, drainage and water table.

4.4.3. Scale in soil-landscape relationships

The investigation of soil-landscape relationships occurs indirectly in the mapping approach proposed in this study. The direct evidence is the soil profile, but the phenomenon studied is the spatial distribution of pedogenetic processes. Therefore, it is necessary to identify the relationship between such elements in a theoretical scheme that encompasses not only the factor-property relationship. In this study we analyze the results from the perspective of Simonson's theoretical model (SIMONSON, 1959), proposed as a basis for soil-landscape analysis through the nested triad factor-process-property (LOZBENEV et al., 2021).

The parameter variation tested in the model is the scale of the landscape factor. The generalization of landscape units, represented by geomorphons, occurred at a scale of 1:75,000 for all processes. This is because this level of detail presented the most consistent features with the expert's interpretation performed by field observations and in geographic information systems. For rougher and more heterogeneous reliefs, a 1:50,000 scale subdivision might better reflect the local dynamics of the hillslopes (MÖLLER; VOLK, 2015; PACHEPSKY; HILL, 2017).

The ideal prediction scale for soil properties, which in the end can be grouped into soil classes, varies according to the property in question. Therefore, this study built each one of the models from covariates at different scales, more suitable for the analysis of the hillslope in relation to the respective landscape property. Almost all the variables were selected from the generalized scales 1:75,000 and 1:100,000. This can be explained both by the prevalence of pedological phenomena at these scales and by the nature of the evolution of the local geomorphological landscape. This is because this landscape presents gentle undulations and is the result of a dissection process in an environment of geological stability. These results agree with a study carried out in plains and smooth valleys of western Romania (DORNIK et al., 2022). For slope gradient, general curvature, profile curvature, and convergence index, generally coarse and medium scales were identified as optimal, over 700 m. The short-range variation was only for slope gradient for clay and sand content, and by convergence index for porosity within the restrictive horizon and sand content. Still, the results of this study proved that more accurate maps could be obtained when predictors are optimally scaled, as compared to using unscaled or all multiscale predictors.

A study that mapped soils in Andalusia, Spain concluded that the relevant ranges of scales do not only differ in the horizontal domain, but also in the vertical domain across the soil profile (RENTSCHLER et al., 2022). They hypothesize that there are depth gradients in spatial and structural dependencies. The results showed that use of topsoil features at small to intermediate scales does not increase model accuracy. In contrast, subsoil models benefit from all scales, from small to large. For soil surface mapping realized at Krui River, Australia, Random Forest was used with DEM at 5 m, 25 m, 30 m, and 90 m resolution for prediction of SOC (GIBSON et al., 2021). They observed, using large-scale maps, that the predictions of SOC are not impacted by topography due to the effect of overriding climatic factors. The topography and DEM scale were found to have more influence on SOC at local scales, and modeling requires high resolution DEM and sampling data to capture hillslope scale variability.

The mapping of the texture of the surface layer (0-30cm) of the soil in the basins of the present study was carried out by Random Forest with generalized multiscale geomorphometric variables and indicated the prevalence of the coarser scales for the prediction of such properties (ARAÚJO; JUNIOR; BELDINI, 2021), namely: 1:75,000 and 1:100,000. The results differed from those found in the study done in Andalusia, and this is possibly explained by the longer-term characteristic of the phenomenon in question, associated with the period of geological stability of the Amazon craton and the intensity of weathering in the humid tropical region. Thus, the smaller-scale features probably did not have enough time for their influence to be noticed in terms of the soil profile. Although SOC mapping was not carried out in our study, it is possible that this property has a greater relationship with the finer scales and eventually with other factors, especially vegetation cover. This is mainly due to the rapid temporal dynamics of organic matter, so even small features could exert a noticeable influence on soils.

It is important to highlight that the elutriation process can be defined as a surface process, while the clay translocation and gleization processes are subsurface processes. In this sense, in the studied landscape, for both the coarser scales of terrain analysis, they presented greater correlation with the observed soil properties. We can infer that the use of covariates in multiple scales, generalized from a digital 30m elevation model, produced better predictive effects than the original scales or in a monoscale generalization. This result agrees with the conclusions found in soil mapping using the "mixed scaling" method and deep learning, carried out in Piracicaba-Brazil, Meuse-Netherlands and Rhine-Germany (BEHRENS et al., 2018). These authors observed that the proposed mixed scaling method shows that their use of multi-scale covariates returns the most accurate results in models and could allow more intuitive interpretation of the datasets.

In a later study on the same mapping areas, the concept of information horizon was proposed, and attention is drawn to the possibility of meaningless covariates being identified as important for the establishment of soil patterns (BEHRENS; VISCARRA ROSSEL, 2020). “The information horizon is the range (or scale) above which interpretations are no longer possible, because beyond the horizon, relationships lack structure”. They suggest several important recommendations for models to be interpretable: only domain-relevant and structurally related predictors must be used, the scales of the predictors must be aligned with the range of soil spatial dependence, and structural and spatial dependence should be analyzed separately. In this sense, the fuzzy modeling pertinence rules have a pedological meaning, keeping a structural relationship with the target variables of the prediction in the soil mapping carried out in the hydrographic basins on the Alter-do-Chão formation. On the other hand, further activities should be directed towards the analysis of the spatial dependence of soil properties.

4.5. Conclusions

As demonstrated, the analysis of scale in the soil-landscape relationship brings a series of complicating factors that deserve attention for methodological development, namely: (a) the establishment of unambiguous relationships between landscape properties, pedological processes and soil properties; (b) optimization of topography generalization methods, with emphasis on the preservation of relevant features; (c) the application of interpretable modeling methods capable of dealing with quantitative, semi-quantitative and qualitative information, whether driven by data or knowledge.

The mapping of pedogenetic processes and soil classes was carried out with a knowledge-oriented approach, considering soil survey observations for the establishment of fuzzy pertinence curves, as well as for the selection of the scale of the geomorphometric covariates. The use of generalized covariates, at different scales according to the pedogenetic process in question, resulted in maps of soil classes with greater accuracy: the general accuracy increased from 74% to 84%, user accuracy increased from 50% to 88% and from 20% to 60% for GXbd and PAd+LAd arg, respectively, and the Kappa index went from 0.4 to 0.69, signifying substantial agreement. Therefore, generalized geomorphometry covariable selection improves the accuracy of soil class maps.

The nested triad factor-process-property proved to be a useful theoretical model for modeling the soil-landscape relationship, even with the complex observation and association of pedogenetic process with related soil property in field conditions. Regarding the pattern of occurrence of pedogenetic processes associated with landscape properties, it is worth noting that (a) gleization has a well-defined position at the bottom of the valleys, (b) elutriation is the prevalent process resulting in texture gradients, widespread on slopes, and (c) clay translocation has a minor effect, restricted to some lower concave positions.

4.6. References

ARAÚJO, C. F.; JUNIOR, R. C. DE O.; BELDINI, T. P. Multiscalar Geomorphometric Generalization for Soil-Landscape Modeling by Random Forest: A Case Study in the Eastern Amazon. **Journal of Geographic Information System**, v. 13, n. 04, p. 434–451, 2021.

ARROUAYS, D. et al. Impressions of digital soil maps: The good, the not so good, and making them ever better. **Geoderma Regional**, v. 20, p. e00255, 2020.

BEHRENS, T. et al. Multi-scale digital soil mapping with deep learning. **Scientific Reports**, v. 8, n. 1, p. 2–10, 2018.

BEHRENS, T.; VISCARRA ROSSEL, R. A. On the interpretability of predictors in spatial data science: the information horizon. **Scientific Reports**, v. 10, n. 1, 2020.

BRIDGLAND, D. The role of geomorphology in the Quaternary. **Geological Society, London, Memoirs**, v. 16, n. 4, p. M58-2021–14, 28 set. 2021.

BROCHADO, G. T. **Atenuação do efeito do desflorestamento em dados SRTM por meio de diferentes técnicas de interpolação**. [s.l.] Instituto Nacional de Pesquisas Espaciais, 2015.

BUI, E. N. Soil survey as a knowledge system. **Geoderma**, v. 120, n. 1–2, p. 17–26, 2004.

CONRAD, O. et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. **Geoscientific Model Development**, v. 8, n. 7, p. 1991–2007, 2015.

CPRM. **Mapa geológico do estado do Pará**. [s.l: s.n.].

DORNIK, A. et al. Optimal scaling of predictors for digital mapping of soil properties. **Geoderma**, v. 405, 2022.

EMBRAPA, C. N. DE P. DE S. **Manual de Métodos de Análise de Solo, 3a Edição**. Rio de Janeiro: EMBRAPA-CNPS, 2017.

FARR, T. G. et al. The Shuttle Radar Topography Mission. **Reviews of Geophysics**, v. 45, n. 2, p. RG2004, 19 maio 2007.

FLORINSKY, I. V. Influence of Topography on Soil Properties. Em: **Digital Terrain Analysis in Soil Science and Geology**. 2nd. ed. London: Elsevier, 2016. p. 482.

FLORINSKY, I. V. An illustrated introduction to general geomorphometry. **Progress in Physical Geography**, v. 41, n. 6, p. 723–752, 1 dez. 2017.

GIBSON, A. J. et al. Assessing digital elevation model resolution for soil organic carbon prediction. **Geoderma**, v. 398, n. August 2020, p. 115106, 2021.

GRASS, D. T. **Geographic Resources Analysis Support System (GRASS) Software, Version 7.8**. Open Source Geospatial Foundation, , 2019.

HARTEMINK, A. E. et al. Soil horizon variation: A review. Em: **Advances in Agronomy**. 1. ed. [s.l.] Elsevier Inc., 2020. v. 160p. 125–185.

HUGGETT, R. J. Soil landscape systems: A model of soil Genesis. **Geoderma**, v. 13, n. 1, p. 1–22, 1975.

IBGE. **Estado do Pará: Mapa de Geomorfologia**. IBGE, , 2008.

IBGE. **Manual técnico de geomorfologia**. 2ª ed. Rio de Janeiro: IBGE, 2009.

IBGE. **Manual Técnico de Pedologia**. 3ª ed. Rio de Janeiro: IBGE, 2015.

JASIEWICZ, J.; STEPINSKI, T. F. Geomorphons-a pattern recognition approach to classification and mapping of landforms. **Geomorphology**, v. 182, p. 147–156, 2013.

KÄMPF, N.; CURI, N. Formação e evolução do solo (pedogênese). Em: KER, J. C. et al. (Eds.). **Pedologia: fundamentos**. Viçosa, MG: SBCS, 2012. p. 207–302.

LI, Z.; OPENSHAW, S. A Natural Principle for the Objective Generalization of Digital Maps. **Cartography and Geographic Information Systems**, v. 20, n. 1, p. 19–29, 1993.

LOZBENEV, N. et al. Incorporating process-based modeling into digital soil mapping: A case study in the virgin steppe of the Central Russian Upland. **Geoderma**, v. 383, n. January 2020, p. 114733, 2021.

MA, Y. et al. Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, v. 70, n. 2, p. 216–235, 2019a.

MA, Y. et al. Pedology and digital soil mapping (DSM). **European Journal of Soil Science**, v. 70, n. 2, p. 216–235, 2019b.

MCBRATNEY, A. B.; ODEH, I. O. A. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. **Geoderma**, v. 77, n. 2–4, p. 85–113, 1997.

MENDES, A. C.; TRUCKENBROD, W.; RODRIGUES, A. C. R. N. Análise faciológica da Formação Alter do Chão (Cretáceo, Bacia do Amazonas), próximo à cidade de Óbidos, Pará, Brasil. **Revista Brasileira de Geociências**, v. 42, n. 1, p. 39–57, 2012.

MENEZES, M. D. DE et al. Digital soil mapping approach based on fuzzy logic and field expert knowledge. **Ciência e Agrotecnologia**, v. 37, n. 4, p. 287–298, 2013.

MENEZES, M. D. DE et al. Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches. **Scientia Agricola**, v. 71, n. 4, p. 316–323, ago. 2014.

MILNE, G. Some suggested units of classification and mapping particularly for East African soils. **Soil Research**, v. 4, p. 183–198, 1935.

MÖLLER, M.; VOLK, M. Effective map scales for soil transport processes and related process domains - Statistical and spatial characterization of their scale-specific inaccuracies. **Geoderma**, v. 247–248, p. 151–160, 2015.

NICOLETTI, M. DO C.; CAMARGO, H. DE A. **Fundamentos da teoria de conjuntos fuzzy**. São Carlos, SP: EdUFSCar, 2004.

PACHEPSKY, Y.; HILL, R. L. Scale and scaling in soils. **Geoderma**, v. 287, p. 4–30, 2017.

PONTIUS, R. G.; MILLONES, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, p. 4407–4429, 2011.

QGIS, D. T. **QGIS Geographic Information System**. Open Source Geospatial Foundation Project, , 2019. Disponível em: <http://qgis.osgeo.org>

R CORE TEAM, R. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, , 2019. Disponível em: <http://www.r-project.org/>

RENTSCHLER, T. et al. Contextual spatial modelling in the horizontal and vertical domains. **Scientific Reports**, p. 1–12, 2022.

ROUDIER, P.; HEWITT, A. E.; BEAUDETTE, D. E. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. **Digital Soil Assessments and Beyond - Proceedings of the Fifth Global Workshop on Digital Soil Mapping**, n. March 2014, p. 227–231, 2012.

SANTOS, R. D. DOS et al. **Manual de descrição e coleta de solo no campo**. 7ª edição ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2015.

SANTOS, H. G. et al. **Sistema brasileiro de classificação de Solos**. 5ª ed. [s.l.] Sociedade Brasileira de Ciência do Solo, 2018.

SCHAEFER, C. E. G. R. et al. Solos da região amazônica. Em: CURI, N. et al. (Eds.). **Pedologia - Solos dos Biomas Brasileiros**. [s.l.] SBCS, 2017. p. 75–111.

SILVA, S. H. G. et al. A Technique for Low Cost Soil Mapping and Validation Using Expert Knowledge on a Watershed in Minas Gerais, Brazil. **Soil Science Society of America Journal**, v. 78, n. 4, p. 1310, 2014.

SIMONSON, R. W. Outline of a generalized theory of soil genesis. **Soil Science Society proceedings**, v. 23, n. 2, p. 152–156, 1959.

SOMOZA, R.; GHIDELLA, M. E. Late Cretaceous to recent plate motions in western South America revisited. **Earth and Planetary Science Letters**, v. 331–332, p. 152–163, 15 maio 2012.

WADOUX, A. M. J. C. et al. Ten challenges for the future of pedometrics. **Geoderma**, v. 401, n. May, p. 115155, 2021.

WYSOCKI, D. A.; SCHOENEBERGER, P. J. Geomorphology of Soil Landscapes. Em: HUANG, P. M.; LI, Y.; SUMNER, M. E. (Eds.). **Handbook of Soil Science: Properties and Processes**. [s.l.] Chemical Rubber Company Press, Boca Raton, Florida, 2011. p. 1–26.

ZHU, A.-X. A similarity model for representig soil spatial information. **Geoderma**, v. 77, p. 217–242, 1997.

ZINCK, J. A. The Geomorphic Landscape: Criteria for Classifying Geoforms. Em: **Geopedology**. Cham: Springer International Publishing, 2016. p. 77–99.

ANEXO – MATERIAL SUPLEMENTAR

PARTE I – DETALHAMENTOS METODOLOGICOS

1. Detalhamentos Metodológicos dos Levantamentos Pedológicos

a. Levantamento nas Áreas Piloto

Objetivo: levantar informações físico-quimicas sobre a camada superficial do solo e morfológicas sobre o horizonte A, na escala 1:25.000. Capitulo 1.

Distribuição dos pontos por hipercubo latino condicionado: Software R, pacote clhs (https://cran.r-project.org/web/packages/clhs/index.html); Rasters *ELEV, SLOPE, RSP, TWI, PROF.CURV, PLAN.CURV, LS.FACTOR*; 10 pontos; para o *input* os rasters foram recortados nos limites de cada área piloto; script executado para cada uma das 9 áreas piloto, totalizando 90 pontos amostrais.

Método de observação e amostragem: Tradagens e mini-trincheiras.

Foram realizadas descrições morfológicas do horizonte A e coletadas amostras de 0-30cm para análise-físico química em laboratório.

(a) 

(b) 

Fig1 . (a) Horizonte A de um pedon de Latossolo Amarelo distrófico típico, (b) Horizonte A de um pedon de Latossolo Amarelo distrófico psamítico.

Análises físico químicas: pH em água, alumínio trocável por extração em NaOH e titulação, matéria orgânica por combustão úmida, granulometria pelo método da sedimentação do silte. Análises realizadas no laboratório multidisciplinar do Campus de Oriximiná.

b. Levantamento de Perfís nas Topossequências

Objetivo: levantar informações físico-quimicas e morfológicas de perfis de solo representativos da variação da paisagem, até 2m de profundidade.

Distribuição dos pontos para amostragem por critério do pedólogo, considerando a representatividade na paisagem e a oportunidade/conveniência da observação. Nesse sentido, foram utilizados sempre que possível cortes de estradas para observação, após os devidos procedimentos de remoção da camada superficial do corte, em pelo menos 20cm, em função de efeitos da exposição atmosférica nas propriedades do pedon. Ainda assim, foram necessárias aberturas de trincheiras para observação no topo e mini-trincheiras, complementadas por tradagens, para observações nas porções inferiores, como sopé e vale.

Método de observação e amostragem: Trincheiras, Cortes de estrada, mini-trincheiras e tradagens.

Foram realizadas as descrições morfológicas do perfil até 200cm. Em nenhum dos casos se chegou ao horizonte C nessa profundidade. Nas posições baixas, sopé e vale, o lençol freático foi alcançado, limitando a profundidade da observação. Nesses casos, foram feitas observações complementares com trado, até os 200cm, para verificação da existência de horizonte C ou de deposições diferenciadas.

(a)


(b)


(c)


(d)


Fig 2. (a) Perfil de Gleissolo háplico típico, (b) Perfil de Latossolo Amarelo distrófico típico, (c) Perfil de Latossolo Amarelo distrófico argissólico, (d) Perfil de Argissolo Amarelo distrófico típico.

c. Levantamento de Observações extras

Objetivo: levantar informações físico-quimicas e morfológicas de perfis de solo, até 2m de profundidade. Foco na identificação dos atributos que discriminam as classes de solo identificadas no levantamento nas topossequencias.

Método de observação e amostragem: Tradagens e mini-trincheiras.

Foram realizadas as descrições morfológicas do perfil até 200cm. As observações ocorreram em mini-trincheiras de aprox. 80cm de profundidade, complementadas com tradagens. O critério para uso das mini-trincheiras com complementação por tradagens considera que os atributos diferenciadores das classes em avaliação são observáveis a partir de uma análise mais minuciosa dos atributos morfológicos no horizonte A e na porção superior do horizonte B. A saber, os atributos diferenciadores são: ocorrência de variação textural em algum nível, formando Bt ou se aproximando disso; ocorrência de cerosidade; ocorrência de gleização.

Fig3. Amostragem em uma mini-trincheira, em posição de fundo vale.

2. Detalhamentos Metodológicos da Generalização Geomorfométrica Multiescalar

a. Correção do efeito de desflorestamento no MDE SRTM 30 metros.

Objetivo: reduzir os efeitos do desflorestamento no MDE e nas derivadas do terreno.

QGIS: Criação de polígonos das áreas desflorestadas, referência em imagem Landsat de fevereiro de 2000, contemporânea a missão SRTM; Adição equivalente à altura de copa estimada para área, referência no raster do *Global Forest Canopy Height*; Delimitação de bordas, e suavização por IWD.

b. Generalização Multiescalar do MDE SRTM 30 metros.

Objetivo: realizar a operação de transformação de escalas nas covariáveis dos tipos geomorfometria geral e geomorfometria específica, conforme a tabela de equivalência proposta.

SAGA GIS: Média local em janelas de 2x2, 3x3, 4x4; Cálculo das derivadas de terreno para o raster original (30m) e para cada raster generalizado (60m, 90m, 120m).

No capitulo 1. GRASS GIS: Classificação de geomorphons.

No capitulo 1. QGIS: Agregação de geomorphons em 5 feições principais (topo, crista, encosta, sopé e vale); Filtragem (sieve) por área mínima mapeável (25.000m², 100.000m², 225.000m², 400.000m²)

No capitulo 2. GRASS GIS: Classificação de geomorphons com o raio de busca interno variável, em função da área mínima mapeável (raios 89m, 178m, 268m, 357m).

No capitulo 2. QGIS: Agregação de geomorphons em 5 feições principais (topo, crista, encosta, sopé e vale).

3. Detalhamentos Metodológicos da Modelagem por Random Forest

a. Base de pontos

Pontos em shapefile, considerando a amostragem realizada no levantamento em áreas piloto. Do total de 90 pontos, foram destinados 30% para validação e 70% para treinamento do modelo, por meio de seleção aleatória.

b. Organização da base de rasters

A base de dados de covariáveis generalizadas foi organizada considerando todas as combinações possíveis entre as escalas.

Table 2. Combinação de covariáveis generalizadas, em grupos uniescalares e multiescalares

Scales: (a) 1: 25,000, (b) 1: 50,000, (c) 1: 75,000, (d) 1: 100,000

	a	b	c	d	a,b	a,c	a,d	a,b,c	a,b,d	a,c,d	a,b,c,d	b,c	b,d	b,c,d	c,d
Raster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Raster + Vector	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

c. Predição de frações argila, silte, areia por Random Forest

A predição para as três frações foi realizada separadamente, por script Random Forest. A fração argila não obteve ajuste razoável e foi desconsiderada. Entende-se que não houve prejuízo a predição, considerando que duas frações são suficientes para a classificação granulométrica, podendo ainda obter-se a terceira por subtração simples.

O Script utilizado para a predição foi adaptado do livro "Using R for Soil Digital Mapping" (Malone *et al*, 2010), na seção 5.5 (Random forest, na seção 5. Continuous soil atribute modelling). Foi utilizado o software R e o pacote randomForest.

O melhor ajuste entre as covariáveis generalizadas foi o critério para seleção, considerando o maior % Var Explained, estatística baseada no R^2 do referido pacote.

d. Avaliação da acurácia

Foram então comparados, a predição pelo modelo Random Forest da escala original (raster 30m) com o grupo multiescalar generalizado de maior ajuste. Essa comparação foi realizada por matriz de confusão, considerando os 30% dos pontos separados para essa finalidade. Foram gerados os índices Kappa, acurácia geral, acurácia do produtor e acurácia do usuário. Essa operação foi realizada por meio de planilhas eletrônicas.

4. Detalhamentos Metodológicos da Modelagem por Lógica Fuzzy

a. Base de pontos

Pontos em shapefile, considerando a amostragem realizada no levantamento em transectos (9 perfis) e nas observações extras (38 perfis). Os pontos dos transectos foram utilizados indiretamente para definição das regras e curvas de pertinência; os pontos das observações extras foram utilizados diretamente para avaliação da acurácia do mapeamento.

b. Construção das Curvas de pertinência

As curvas foram construídas considerando as observações de campo, e tendo como referência os pontos georreferenciados e as informações topográficas em ambiente digital. Forma utilizados histogramas e variações nas escalas de cores das representações dos arquivos raster, para auxilio com a interpretação para construção das curvas.

SAGA GIS: Ferramenta Fuzzify. Parâmetros A (pertinência 0), B (pertinência 1), C (pertinência 1), D (pertinência 0). Os parâmetros foram ajustados considerando fenômenos com representação unicaudal ou bi-caudal. Todas as curvas foram definidas como do tipo sigmodal. Após a fuzificação das covariáveis, utiliza-se as ferramentas Fuzzy Intersection (AND) ou Fuzzy Union (OR) para definição da regra de pertinência.

c. Seleção das Escalas das Covariáveis

Os procedimentos para construção dos rasters fuzificados foi realizado para a escala original. Após, foi realizado a seleção das escalas generalizadas pertinentes para definição do pedoambiente preferencial de ocorrência de determinado processo. Essa seleção foi feita considerando as observações de campo, e tendo como referência os pontos georreferenciados e as informações topográficas em ambiente digital. Forma utilizados histogramas e variações nas escalas de cores das representações dos arquivos raster, para auxilio com a interpretação para construção das curvas. Nesse sentido, foram gerados rasters de pedoambiente preferencial para cada processo pedológico, na escala original e na escala considerando as covariáveis generalizadas em múltiplas escalas.

d. Defuzzificação e Transformação em Classes de Solos

A defuzificação com posterior transformação para classes de solo utilizou o critério de maior valor de pertinência, entre os pedoambientes mapeados em uma determinada escala. A definição das classes de solo (ou grupos indiferenciados) obedeceu os seguintes critérios: maior probabilidade de ocorrência predominante de gleização, classificado com GXbd; maior probabilidade de ocorrência predominante de elutriação, classificado como PAd+LA darg; maior probabilidade de ocorrência predominante de translocação de argila, classificado como PAd+ LAdarg; maior probabilidade de ocorrência predominante de ferralitização, classificado como LAd+LAd psa.

e. Avaliação da acurácia

Foram então comparados, a predição na escala original (raster 30m) com a predição realizada a partir da seleção de covariáveis em diversas escalas, de acordo com o conhecimento do pedólogo. Essa comparação foi realizada por matriz de confusão, considerando os 30% dos pontos separados para essa finalidade. Foram gerados os índices Kappa, acurácia geral, acurácia do produtor e acurácia do usuário. Essa operação foi realizada por meio de planilhas eletrônicas.