



**UNIVERSIDADE FEDERAL DO TOCANTINS  
CAMPUS UNIVERSITÁRIO DE GURUPI  
CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL**

**CARLOS HENRIQUE PINTO DA SILVA**

**MAPEAMENTO DIGITAL DO SOLO COM O USO DE REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM REGIÃO DE ORTIGUEIRA, PARANÁ**

**Gurupi (TO)  
2018**

**CARLOS HENRIQUE PINTO DA SILVA**

**MAPEAMENTO DIGITAL DO SOLO COM O USO DE REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM REGIÃO DE ORTIGUEIRA, PARANÁ**

Monografia apresentada à Universidade Federal do Tocantins (UFT), Campus Universitário de Gurupi para obtenção do título de bacharel em Engenharia Florestal, sob orientação do Prof. Dr. Marcos Giongo

**GURUPI (TO)**

**2018**

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Sistema de Bibliotecas da Universidade Federal do Tocantins**

---

S586m Silva, Carlos Henrique Pinto da.  
MAPEAMENTO DIGITAL DO SOLO COM O USO DE REGRESSÕES  
LOGISTICAS MULTIPLAS EM REGIÃO DE ORTIGUEIRA, PARANÁ. /  
Carlos Henrique Pinto da Silva. – Gurupi, TO, 2018.  
36 f.

Monografia Graduação - Universidade Federal do Tocantins – Câmpus  
Universitário de Gurupi - Curso de Engenharia Florestal, 2018.

Orientador: Marcos Vinicius Giongo Alves

1. Levantamento de solo. 2. Mapeamento digital do solo. 3. Pedometria. 4.  
Regressões logísticas. I. Título

**CDD 577.272**

---

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer  
forma ou por qualquer meio deste documento é autorizado desde que citada a fonte.  
A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184  
do Código Penal.

**Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os  
dados fornecidos pelo(a) autor(a).**

CARLOS HENRIQUE PINTO DA SILVA

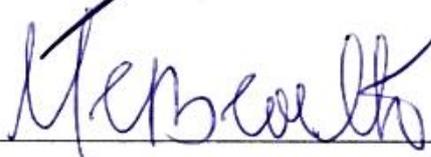
MAPEAMENTO DIGITAL DO SOLO COM O USO DE REGRESSÕES LOGÍSTICAS  
MÚLTIPLAS EM REGIÃO DE ORTIGUEIRA, PARANÁ.

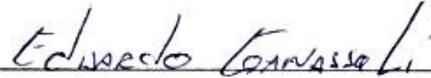
Monografia foi avaliada e apresentada à UFT – Universidade Federal do Tocantins – Campus Universitário de Gurupi, Curso de Engenharia Florestal para obtenção do título de Bacharel em Engenharia Florestal e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Data de aprovação: 28 / 11 / 18

Banca Examinadora

  
\_\_\_\_\_  
Prof. Dr. Marcos Vinicius Giongo Alves, Orientador, UFT.

  
\_\_\_\_\_  
Prof(a). Dra. Maria Cristina Bueno Coelho, Examinadora, UFT.

  
\_\_\_\_\_  
Eng. Florestal Eduardo Ganassoli Neto, Examinador, UFPR.

*"O homem é corda estendida entre o animal e o Super-homem: uma corda sobre um abismo; perigosa travessia, perigoso caminhar, perigoso olhar para trás, perigoso tremer e parar."*

*...Assim falava Zaratustra.*

## **AGRADECIMENTOS**

Aos pensadores novos e antigos.

Ao Professor Giongo por me incentivar a “engenhar”.

A Universidade Federal do Tocantins por se estruturar como uma instituição pública que dá oportunidades para jovens de origem pobre como a minha.

A Klabin S/A e meus supervisores de estágio James e Clewerson, que não me negaram conhecimento e incentivo.

Aos meus amigos verdinhos e pé-de-manguences, vocês são muitos e as páginas são poucas para citá-los, mas fica aqui em escrito meu obrigado e que vocês até não podem ser as pessoas da minha vida, mas eram as que eu precisava nela.

Aos meus parceiros de estágio Samuel e Alisson.

A minha irmã caçula. Cuidar de você me fez amadurecer.

A minha irmã de vida Tracy por segurar essas “marimbas” da vida adulta dentre outras coisas. “Needles and Pins” mana.

Ao Jamel, por ser grande homem e a figura paterna de apoio a essa minha família.

A minha mãezinha Lu, por acreditar em mim quando eu não acreditava, por me criar aos moldes do amor e carinho e por acreditar que a construção de uma pessoa de bem é pelo conhecimento.

## RESUMO

O presente estudo teve como avaliar o mapeamento digital de classes de solos com uso de regressões logísticas múltiplas em diferentes tamanhos de amostras, avaliando sua potencialidade como mapeamento acurado para classes de solo. O estudo foi conduzido em região de Ortigueira, Paraná, Brasil a partir de um Modelo Digital de Elevação, e seus derivados, com uma resolução de 4 m<sup>2</sup> e de amostras georreferenciadas de perfil de solo classificado. Foram avaliados diferentes *buffers* amostrais de pontos com 75, 100 e 125 metros que formaram diferentes bases de dados para as regressões logísticas. Os critérios avaliados foram a significância de variáveis de terreno, o ajuste das regressões logísticas aos dados amostrados, acurácias individuais de classe, acurácia geral e coeficiente Kappa Cohen. As variáveis de terreno que mais explicaram o comportamento das regressões foram altitude e declividade. Os coeficientes de determinação (R<sup>2</sup>) de Nagelkerke foram de 81,9 a 100% para CXa3, 42,5 a 51,1% para CXa1 e de 16,5 a 19,7 para LVd1. Os *buffers* amostrais analisados resultaram em valores de 65,2 a 69,6% para acurácia geral e 63,2 a 67,3% para coeficiente Kappa, em destaque os *buffers* amostrais de pontos de 75 e 100 metros obtiveram os melhores resultados com 69,6 e 67,3% para acurácia geral e coeficiente Kappa respectivamente, porém verificou-se que a definição do melhor *buffer* para o mapeamento está relacionada com a posição das amostras de solo classificadas em relação as variáveis de terreno.

**Palavras-chaves:** Levantamento de solo. Mapeamento digital do solo. Pedometria. Regressões logísticas.

## ABSTRACT

The present study aimed at the digital mapping of soil classes with the use of logistic regressions in several sample sizes, evaluating their potential as accurate mapping for soil classes. The study was conducted in a region of Ortigueira, Paraná, Brazil from a Digital Elevation Model, and its descendants, with a resolution of 4 m<sup>2</sup> and georeferenced samples of classified soil profile. We selected variable buffers of 75, 100 and 125 meters points that form databases for logistic regressions. The criteria used were the significance of field variables, the adjustment of logistic regressions to the sampled data, individual class accuracy, general accuracy and Kappa Cohen coefficient. The variables of greater extension explained the behavior of the regressions by altitude and declivity. The determination coefficients ( $R^2$ ) of Nagelkerke were 81.9 to 100% for CXa3, 42.5 to 51.1% for CXa1 and 16.5 to 19.7 for LVd1. State buffers resulted in values ranging from 65.2 to 69.6% for general accuracy and 63.2 to 67.3% for Kappa coefficients, in particular the sample buffers at points of 75 and 100 meters obtained the results of group results 69, 6 and 67.3% for the general set and Kappa coefficient respectively, but it was verified that a definition for the buffer for the mapping is related to the position of the soil samples classified in relation as terrain variables.

**Key-words:** Soil survey. Digital mapping of the soil. Pedometry. Logistic regressions.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Áreas de conhecimento da pedometria.....	15
Figura 2 - Área de estudo.....	18
Figura 3 - Fluxograma de procedimentos para mapeamento de solos por regressões logísticas múltiplas.....	20
Quadro 1 - Script de comandos em linguagem R para processamento dos dados do buffer amostral de 100 metros.....	21
Figura 4 - Tradagens de solo para avaliação.....	24
Gráfico 1- Coeficiente de determinação de Nagelkerke.....	26
Figura 5 - Declividade e altitude com pontos de amostras e buffer.....	27
Figura 6 - Mapas estimados de B100 e B125.....	28
Gráfico 2 - Acurácia do usuário.....	28
Gráfico 3 - Acurácia do mapeador.....	29
Gráfico 4 - Acurácia geral e coeficiente Kappa.....	30

## LISTA DE TABELAS

Tabela 1 - Legendas/Descrição das classes de solo apresentadas na Figura 2.....	19
Tabela 2 - Exemplo de matriz de erros de comparação entre pontos do mapa estimado e verdade de campo.....	22
Tabela 3 - Significância de variáveis de terreno em relação as classes e os buffers estudados e ajuste das regressões logísticas.....	25

## LISTA DE ABREVIATURAS E SIGLAS

AG	Acurácia Geral
AM	Acurácia do Mapeador
AU	Acurácia do Usuário
B57	Buffer Amostral de 75 Metros
B100	Buffer Amostral de 100 Metros
B125	Buffer Amostral de 125 Metros
LiDAR	Light Detection And Ranging
Kappa	Coefficiente Kappa de Cohen
MDE	Modelo Digital de Elevação
MDS	Modelo Digital do Solo
RML	Regressões Logísticas Múltiplas
SIG	Sistemas de Informações Geográficas

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
<b>1.1</b>	<b>Objetivos .....</b>	<b>13</b>
1.1.1	Objetivo geral .....	13
1.1.2	Objetivos específicos .....	13
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>14</b>
<b>2.1</b>	<b>Regressões logísticas múltiplas .....</b>	<b>14</b>
<b>2.2</b>	<b>Pedometria.....</b>	<b>14</b>
<b>2.3</b>	<b>Função scorpan .....</b>	<b>15</b>
<b>2.4</b>	<b>Variáveis de terreno .....</b>	<b>16</b>
<b>3</b>	<b>MATERIAL E MÉTODOS .....</b>	<b>18</b>
<b>3.1</b>	<b>Área de estudo.....</b>	<b>18</b>
<b>3.2</b>	<b>Procedimentos .....</b>	<b>19</b>
3.2.1	Variáveis de Terreno e Amostragem .....	19
<b>3.3</b>	<b>Regressões Logísticas Múltiplas .....</b>	<b>21</b>
<b>3.4</b>	<b>Avaliadores de mapeamento .....</b>	<b>21</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>25</b>
<b>4.1</b>	<b>Significância e ajuste de variáveis de terreno .....</b>	<b>25</b>
<b>4.2</b>	<b>Ajuste das regressões logísticas .....</b>	<b>25</b>
<b>4.3</b>	<b>Acurácia.....</b>	<b>27</b>
<b>4.4</b>	<b>Discussões gerais .....</b>	<b>30</b>
<b>5</b>	<b>CONCLUSÕES.....</b>	<b>31</b>
	<b>REFERÊNCIAS.....</b>	<b>32</b>

## 1 INTRODUÇÃO

Os conhecimentos sobre as propriedades do solo estão agregados a diversas finalidades de importância ambientais como práticas conservacionistas de auxílio na prevenção da erosão, contaminação de lençóis freáticos, e também florestal como a seleção de culturas, manejo, colheita e reflorestamento (MACMILLAN et al., 2004; ARRUDA et al., 2013). A demanda da sociedade por esse tipo de conhecimento tem aumentado devido ao interesse global na evolução da produção agrícola e nas questões de sustentabilidade ambiental (HARTEMINK e MCBRATNEY, 2008). Em nível operacional a existência de mapas que retratam as principais propriedades do solo em suas limitações e capacidades é uma das principais ferramentas de apoio das práticas supracitadas (MACMILLAN et al., 2004).

A produção dos primeiros mapas de propriedades do solo, por abordagem tradicional, tem qualidade definida pela capacidade do pedólogo em taxonomia para criar e distinguir as classes de solos e seus limites, baseada nas relações entre os diversos tipos de solos e seus ambientes predominantes (RANZANI, 1969). Tais relações estão apoiadas em observações de campo, foto interpretação e delimitação de polígonos em escala proporcional ao nível de detalhe esperado (EMBRAPA, 1995; NEUMANN, 2012).

Desde a década de 80 estudos vem debatendo sobre a qualidade da abordagem tradicional, criticando sua desconsideração a dependência espacial e seu caráter intuitivo e determinístico (WEBSTER, 1984; ODEH et al., 1990; HUDSON, 1992). Somando-se a isso o mapeamento tradicional é amplamente caracterizado por demandar muitos recursos humanos e temporais gastos em campo, que se contrapõe com a perspectiva e o desenvolvimento crescente da agricultura de precisão em produzir mapas de alta resolução e eficiência, com baixo custo e que atenda o número crescente de demandas (McBRATNEY et al., 2003; MACMILLAN et al., 2004; NEUMANN, 2012).

Sobre esse cenário se desenvolveu abordagens mais quantitativas com a capacidade de descrever, classificar e estudar os padrões de dependência espacial das propriedades do solo através de modelos numéricos ou estatísticos que relacionam as variáveis ambientais e atributos de solo a um banco de dados geográfico, garantindo o conhecimento da variabilidade espacial, sua precisão e qualidade, essas diversas técnicas quantitativas foram denominadas de Pedometria (McBRATNEY et al., 2000; MENDONÇA-SANTOS e SANTOS, 2003; SCULL et al., 2003). O avanço desse ramo científico se deve principalmente pelos avanços tecnológicos de Sistemas de Informação Geográfica relacionado a ferramentas de análise, processamento e disponibilidade de dados geográficos mais extensos capazes de cruzar informações do meio

físico (sobretudo informações de superfície) e classes de solos mapeadas (ARRUDA et al., 2013; MINASNY e McBRATNEY, 2016; MUÑOZ-ROJAS et al., 2017). Tal avanço colaborativo possibilitou a criação da subdisciplina da ciência do solo conhecida como Mapeamento Digital do Solo (MDS) (MINASNY e McBRATNEY, 2016).

Contudo a variação do solo no espaço é bastante complexa, e cada unidade de mapa de solo se difere de outras áreas de pesquisa em certa medida, é preferível que o tamanho dessas unidades deva ser mantido pequeno para uma melhor delimitação de resultados controláveis e que a decisão quanto as análises a serem feitas relacionadas a tamanho, número e delineamento de modelos para essas unidades sejam relacionadas a quantidade e tipo de dados disponíveis (MARK e CSILLAG, 1989; McBRATNEY e GRUIJTER, 1992; LORENZETTI et al., 2015). Devido essa complexidade é sempre necessário a produção de modelos aprimorados com a capacidade de relacionar as informações disponíveis e os processos analíticos relacionados (BREVIK et al., 2016).

Trabalhos anteriores já elaboraram análises significativas quanto a predição de tipologias de solo através da modelagem de regressões logísticas múltiplas (RLM) sobre diferentes perspectivas regionais e analíticas que se mostraram efetivas na inferência de variáveis categóricas como classes de solo (FIGUEIREDO, 2006; CATEN, 2018). Além de sua característica de predição de variáveis categóricas, as RLM se destacam na independência da distribuição de suas variáveis explicativas, já que as mesmas podem ser de caráter contínuo, discreto ou binário, conferindo assim a possibilidade de uma maior diversidade de variáveis explicativas para determinado evento (CHATTERJEE e HADI, 2006). Essa relação é definida como uma razão entre a probabilidade do evento ocorrer e a probabilidade dele não ocorrer, ou seja, em MDS de classes do solo determina a probabilidade de um pixel do mapeamento pertencer a uma classe de solo específica ou não (CHATTERJEE e HADI, 2006; FIGUEIREDO, 2006; CATEN, 2008).

Sabe-se que a topografia tem forte correlação com as propriedades do solo, controlando os fluxos de água em seu interior e determinando sua textura. Alguns métodos de classificação automática baseados em Mapeamento Digital do Solo (MDS) se viram capazes de incorporar tais critérios hidrológicos e geomorfológicos na modelagem e nas estimativas espaciais das propriedades do solo (MACMILLAN et al., 2004; SCHMIDT e HEWITT, 2004).

Contudo, mesmo que em vários países técnicas MDS tenham evoluído com o objetivo de produzir informações mais acuradas e precisas, os trabalhos recentes que fizeram o uso de RLM no Brasil para predição de tipologias foram feitos sobre bases de referência em escalas de 1:50000 e até então não se apresentou nada mais acurado originado de escalas menores ou

baseado em amostras de campo e variáveis explicatórias com resoluções de 4 m<sup>2</sup> (FIGUEIREDO, 2006; CATEN, 2008; GRUNWALD et al., 2011; ARROUAYS et al., 2017).

## **1.1 Objetivos**

### 1.1.1 Objetivo geral

Avaliar mapas de solo por regressões logísticas através do estabelecimento de relações matemáticas entre a distribuição e ocorrência das classes de solos com variáveis de terreno topográficas e hidrográficas, sobre diferentes tamanhos amostrais.

### 1.1.2 Objetivos específicos

1. Selecionar as variáveis significativas ao modelo dentro das diferentes perspectivas amostrais;
2. Definir qual a melhor tamanho amostral;
3. Avaliar os níveis de acurácia dos mapeamentos baseado nos tamanhos amostrais abordados.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Regressões logísticas múltiplas

As regressões logísticas múltiplas são funções logarítmicas da razão entre a probabilidade de um evento ocorrer e a probabilidade de ele não ocorrer, comumente chamado de razão de chance, a função logarítmica da razão de chance é chamada de Logit. Diferentemente das regressões lineares que tem comportamento gráfico linear as regressões logísticas têm um comportamento em “S” variando entre 0 e 1, ou transcrevendo em porcentagem, entre 0 e 100% de chance. Em mapeamentos digitais de classe de solo a razão de chance determina a probabilidade de um *pixel* do mapeamento pertencer a uma classe de solo específica ou não (CHATTERJEE e HADI, 2006; FIGUEIREDO, 2006; CATEN, 2008).

$$\text{Logit} = \text{Ln} \left[ \frac{P_i}{(1-P_i)} \right] = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k \quad (1)$$

$$P_i = \frac{1}{[1+e^{-(b_0+b_1*X_1+b_2*X_2+\dots+b_k*X_k)}]} \quad (2)$$

As equações 1 e 2 demonstram como calcular a função Logit e a probabilidade P de uma categoria predita baseado em um conjunto de variáveis teoricamente explicativas. O valor de  $b_0$  é o indicador do intercepto da regressão e os valores de  $b_k$  são os coeficientes que ajustam as variáveis explicativas  $X_k$ , estes parâmetros da regressão são estimados pelo método da verossimilhança (FIGUEIRA, 2006). Em regressões logísticas múltiplas as variáveis explicativas  $X_k$  não precisam ter distribuição normal, correlação linear, medidas em mesma escala ou homogeneidade de variância, também não se restringe a variáveis contínuas, tendo portabilidade para uma mescla destas com variáveis discretas e binárias (CHATTERJEE e HADI, 2006).

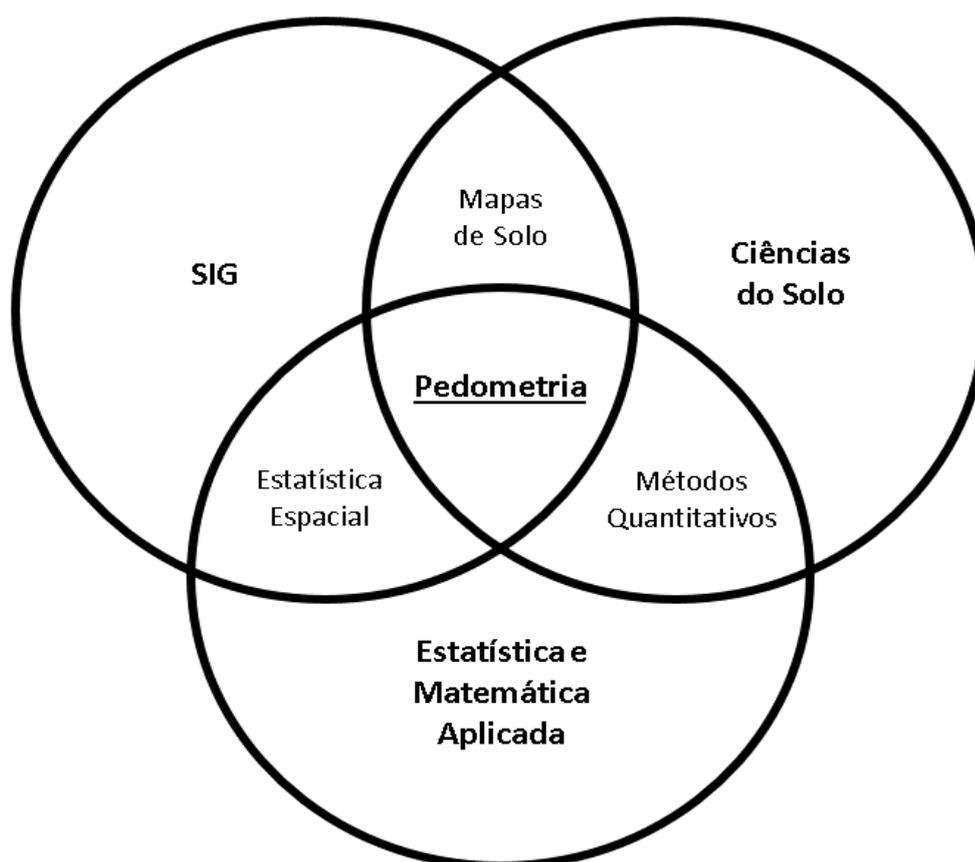
### 2.2 Pedometria

A Pedometria se baseia na junção de grandes áreas de conhecimento com objetivo geral de produzir informações de solos, por meio da aplicação de métodos matemáticos e estatísticos na modelagem de seus atributos, focado em sua gênese e distribuição. (CATEN, 2008; LOPÉZ, 2009; COELHO, 2010; SILVEIRA, 2010). Ela é capaz de definir os padrões de dependência

espacial das propriedades do solo através do relacionamento matemático e estatístico de dados georreferenciados de variáveis ambientais e atributos do solo de natureza quantitativa, com a garantia de precisão e qualidade nos resultados (McBRATNEY et al., 2000; MENDONÇA-SANTOS e SANTOS, 2003; SCULL et al., 2003; FIGUEIREDO, 2006).

Logo a Pedometria é uma ciência de domínio interdisciplinar entre ciência do solo, estatística-matemática aplicada e Sistemas de Informações Geográficas (SIG), como representado na figura 1 (HENGL, 2003; CATEN, 2008). E sua aplicação abrange a predição de diversas variáveis contínuas de solos que possuem distribuição de suave a brusca no espaço, como pH, cálcio, potássio, entre outros. Também abrange variáveis categóricas como estrutura e classes de solo (LOPÉZ, 2009).

Figura 1 - Áreas de conhecimento da pedometria



Fonte: Adaptado de Hengl (2003).

### 2.3 Função scorpan

Desde a década de 40 foram propostas funções que representassem as associações entre os solos e os fatores de formação (McBRATNEY et al., 2000). Com a evolução constante do

SIG e tecnologias relacionadas essas funções puderam evoluir para o modelo SCORPAN, conforme a Equação 3 (McBRATNEY et al., 2003).

$$Solo = f(s, c, o, r, p, a, n) \quad (3)$$

A Equação 3 e a representação da função Scorpan, onde os atributos de solos contínuos e categóricos podem ser previstos a partir de: (s): o próprio solo, em forma de amostras ou mapas de referência; (c): o clima, com sua macro e micro influência; (o): organismos vivos; (r): atributos da paisagem como relevo e características correlacionadas; (p): material de origem (litologia); (a): o tempo cronológico de exposição; e (n): as coordenadas geográficas dos fatores de formação (McBRATNEY et al., 2003). Onde todos ou parte desses atributos estão englobados em uma função ou metodologia preditiva de natureza matemática e estatística  $f(x...)$ .

O modelo SCORPAN pode ser visto como uma síntese dos fundamentos que regem as tecnologias de inferência espacial, principalmente consolidadas pela confecção de mapas digitais de solo (CATEN, 2008; LOPÉZ, 2009).

## 2.4 Variáveis de terreno

Como proposto no modelo Escorpan para predição de solos, as variáveis de Terreno se relacionam e controlam aspectos da pedogênese do solo e sua distribuição espacial. A distribuição espacial de variáveis como altitude, ondulações do terreno, acúmulo de água, concentração de matéria orgânica e derivadas podem ser constituintes fundamentais para a predição de classes de solos (MCBRATNEY et al., 2003; CATEN, 2008).

Uma das principais variáveis de terreno é a Altitude, que além de ser importante para o condicionamento do clima pode ser responsável por cerca de 72 variáveis de terreno topográficas e hidrográficas, como declividade, curvaturas, acúmulo de fluxo entre outras (WILSON e GALLANT, 2000; PARK et al., 2001).

A declividade é o principal regulador dos fluxos superficiais e subsuperficiais, do conteúdo de água no solo e conseqüentemente na formação do solo (WILSON e GALLANT, 2000). Em termos gerais ela é expressa como a mudança de elevação em uma determinada distância (BURROUGH, 1986).

As curvaturas representam as formas do relevo indicando a concentração ou dispersão da água sob o relevo e também a velocidade da água e os processos de transporte de sedimentos

relacionados. Tais fatores são expressos pela Curvatura Planar e Curvatura de Perfil (WILSON e GALLANT, 2000; SIRTOLI et al., 2008).

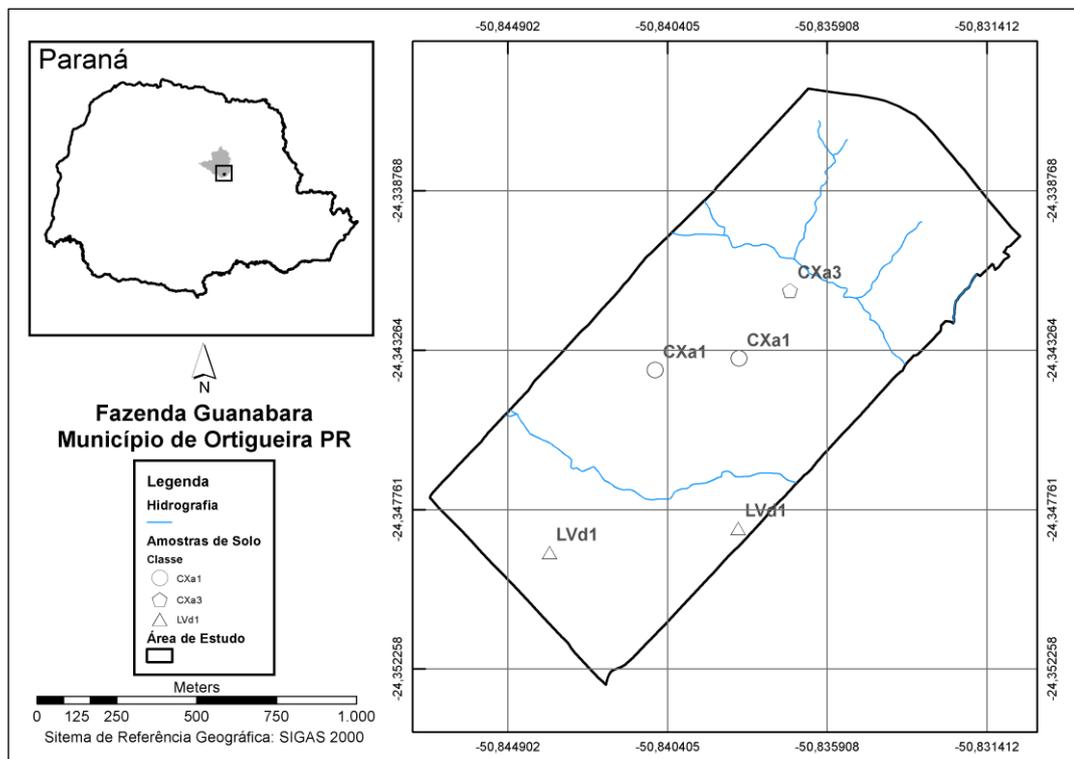
O acúmulo de fluxo é a quantidade de água acumulada na superfície em relação a quantidade dos fluxos que ocorreram sobre ela, se baseia nas células circundantes, porém é formado pelo acúmulo do peso de contribuição das células de cotas mais altas (JENSON e DOMINGUE, 1988).

### 3 MATERIAL E MÉTODOS

#### 3.1 Área de estudo

Este trabalho foi realizado para a Fazenda Guanabara pertencente a Klabin S/A, de aproximadamente 140,5 hectares destinada para cultura de espécies florestais, localizada no município de Ortigueira, Paraná, Brasil, que fica as margens da BR-376 com coordenadas centrais 24°20'38.0" Sul e 50°50'19.5" Oeste, abaixo na figura 2. De acordo com Köeppen o clima predominante, é do tipo Cfb, com precipitação média anual variando de 1.300 a 1.800 mm bem distribuído durante o ano (EMBRAPA, 2018).

Figura 2 - Área de estudo



Fonte: O Autor (2018).

Para este estudo havia dados de cinco pontos georreferenciados com informações dos perfis de solo classificados, constando três solos diferentes para área, sendo eles CXa1, CXa3 e LVd1 (tabela 1), classificados em dois períodos entre 2007 e 2008.

Tabela 1- Legendas/Descrição das classes de solo apresentadas na Figura 2

<b>Tipo de Solo</b>	<b>Legenda / Descrição</b>
CXa1	Cambissolo, Háplítico Alumínico típico ou léptico, Horiz. A moderado ou proeminente, relevo plano a ondulado, textura média a argilosa.
CXa3	Cambissolo, Háplítico Alumínico típico ou léptico, Horiz. A moderado ou proeminente, relevo ondulado a forte ondulado, textura argilosa a muito argilosa.
LVd1	Latossolo, Vermelho Distrófico típico, Horiz. A moderado ou proeminente, relevo plano a suave ondulado, textura argilosa a muito argilosa.

Fonte: Sistema Brasileiro de Classificação de Solos (SiBCS) (EMBRAPA, 2018).

## 3.2 Procedimentos

### 3.2.1 Variáveis de Terreno e Amostragem

O SIG empregado neste estudo foi o ArcMap® 10.5 (ESRI, 2018). E a extensão que possibilitou a geração imagens raster de variáveis hidrológicas e geomorfológicas do terreno foi a extensão de Analista Espacial (Spatial Analyst), que disponibiliza a Caixa de Ferramentas do Analista Espacial (Spatial Analyst Toolbox), além de criar as imagens raster essa caixa de ferramenta permite manipular e realizar cálculos algébricos com esses tipos de variável (ESRI, 2018).

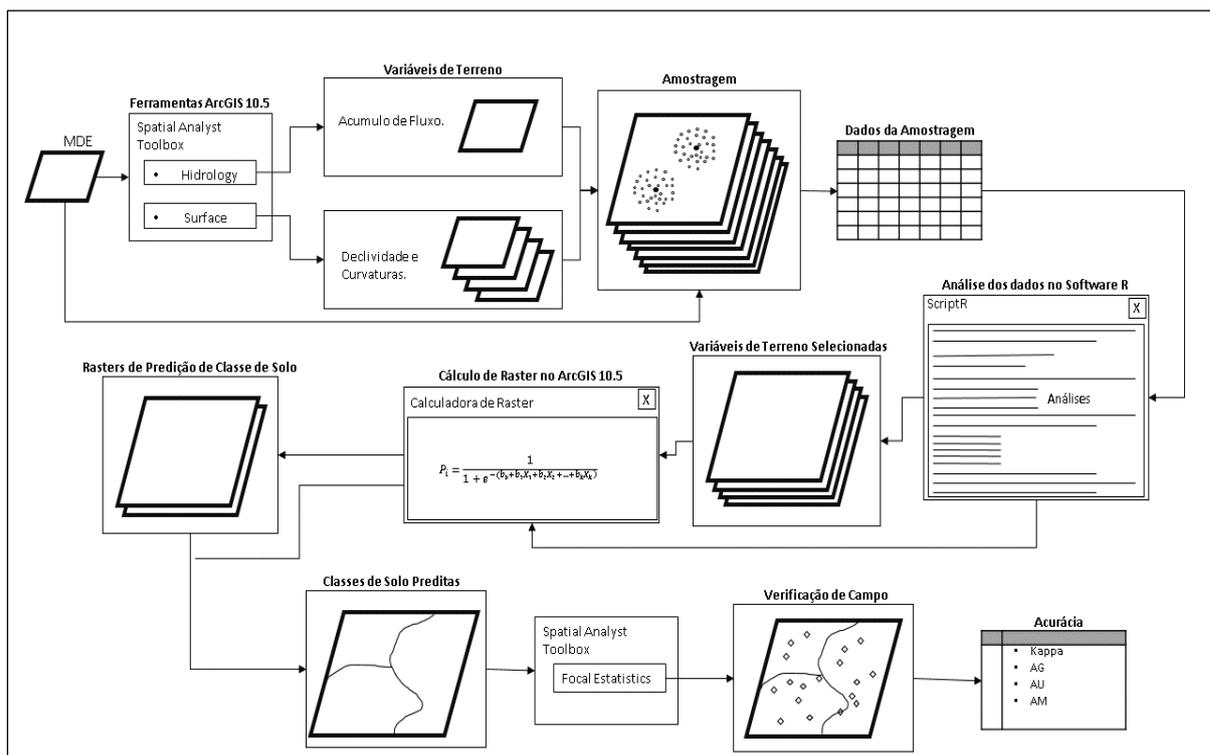
A principal fonte de dados usada neste trabalho foi a imagem de elevação da área, também conhecido como Modelo Digital de Elevação (MDE), obtido através do processamento de nuvem de pontos de Detecção de luz e alcance (Light Detection And Ranging – LiDAR). Com uma resolução de 4 m<sup>2</sup> por pixel no sistema de referência geográfica SIGAS 2000 22 Sul. Essa informação foi cedida pela empresa proprietária da área.

Como o objetivo deste estudo envolve a definição de uma tamanho amostral adequado para os dados disponíveis, a amostragem de dados das variáveis de superfície foi baseada em diferentes *buffers* amostrais de pontos aleatórios com raios de 75, 100 e 125 metros (B75, B100 e B125) em volta de cada um dos cinco perfis do solo classificado e georreferenciado, os pontos amostrais dentro de cada *buffer* tem uma distância mínima de 10 metros entre eles e cada *buffer* obteve uma quantidade igual de pontos de amostra. A decisão do uso desses *buffers* amostrais foi necessária devido ao nível de detalhamento das imagens e a falta de representatividade das variáveis explicativas correlacionadas ao solo local coletando apenas as informações do ponto perfil classificado.

As variáveis topográficas a serem amostradas, altitude, declividade, curvatura, plano de curvatura, perfil de curvatura e acúmulo de fluxo foram calculadas a partir do MDE, usando-se as ferramentas do grupo de ferramentas de Hidrologia (Hydrology) e Superfície (Surface). Essas imagens são calculadas com os coeficientes de regressão logística na ferramenta Calculadora de Raster (Raster Calculator) gerando rasters de predominância de cada classe de solo, com os valores de chance de ocorrência do solo distribuídos na área, esses rasters são comparados entre si e é definido para cada pixel a classe com a maior chance de ocorrência. As ferramentas de Hidrologia, Superfície e a Calculadora de Raster pertencem a Caixa de Ferramentas do Analista Espacial (ESRI, 2018).

Após a comparação de predominância de rasters de predição de classe de solo o resultado final com as classes já preditas é processado pela ferramenta Estatísticas Focais (Focal Statistics) da Caixa de Ferramentas do Analista Espacial, que neste caso calculou pixel a pixel o tipo de classe que ocorria com maior frequência em 10 m<sup>2</sup>. Isso teve o objetivo de minimizar os efeitos de microescala causados pela resolução de 4 m<sup>2</sup> como a presença de estradas, manejo de solo ou a possibilidade de amostrar as coordenadas de um pixel isolado de classe de solo nas análises de acurácia, como mostrado na figura 3.

Figura 3 - Fluxograma de procedimentos para mapeamento de solos por regressões logísticas múltiplas



Fonte: O Autor (2018).

### 3.3 Regressões Logísticas Múltiplas

Os dados das amostras dos *buffers* previamente tabelados, foram processados e analisados no sistema computacional estatístico R através de script de comandos e comandos dos pacotes “rms” e “DescTools” voltados neste estudo para o processamento de regressões logísticas, gerando estatísticas descritivas e os coeficientes de regressão usados para o cálculo de predição de solos com as variáveis de terreno (R Development Core Team, 2014), conforme o quadro 1.

Quadro 1 - Script de comandos em linguagem R para processamento dos dados do buffer amostral de 100 metros

```
#Importando a tabela exportada do ArcGIS.
Tabela<-read.csv2("D:/Estudo/Amostragens/B100/Amostragem_B100.csv")
#Realizando a regressões logisticas para cada Classe com as variáveis de
terreno amostradas
pCXa1<-glm(pCXa1~var1+var2+var3+var4+var5+var6,family=binomial
("logit"),data=Tabela)
pCXa3<-glm(pCXa3~var1+var2+var3+var4+var5+var6,family=binomial
("logit"),data=Tabela)
pLVd1<-glm(pLVd1~var1+var2+var3+var4+var5+var6,family=binomial
("logit"),data=Tabela)
#Teste Qui-Quadrado
anova(pCXa1,test="Chisq")
anova(pCXa3,test="Chisq")
anova(pLVd1,test="Chisq")
#Exportando os coeficientes de predição das classes para uso na
Calculadora de Raster no ArcGIS.
write.csv(pCXa1$coefficients,file="D:/Estudo/Amostragens/B100/pCXa1.csv")
write.csv(pCXa3$coefficients,file="D:/Estudo/Amostragens/B100/pCXa3.csv")
write.csv(pLVd1$coefficients,file="D:/Estudo/Amostragens/B100/pLVd1.csv")
#Exportando valores do Teste Qui-Quadrado para Análise
pCXa1_Anova<-anova(pCXa1,test="Chisq")
pCXa3_Anova<-anova(pCXa3,test="Chisq")
pLVd1_Anova<-anova(pLVd1,test="Chisq")
write.csv2(pCXa1_Anova,file="D:/Estudo/Amostragens/B100/pCXa1_Anova.csv")
write.csv2(pCXa3_Anova,file="D:/Estudo/Amostragens/B100/pCXa3_Anova.csv")
write.csv2(pLVd1_Anova,file="D:/Estudo/Amostragens/B100/pLVd1_Anova.csv")
#Avaliando Coeficientes de determinação de Nagelkerke
PseudoR2(pCXa1,c("Nagel"))
PseudoR2(pCXa3,c("Nagel"))
PseudoR2(pLVd1,c("Nagel"))
```

Nota: Na realização das regressões logísticas var1, var2, var3, var4, var5 e var6 são respectivamente os valores das variáveis de terreno amostradas, Altitude, Declividade, Curvatura, Plano de Curvatura, Perfil de Curvatura e Acumulo de Fluxo e pCXa1, pCXa3 e pLVd1 é a ocorrência da classe de solo em cada ponto amostral, sendo 1 onde está presente e 0 onde não está presente

### 3.4 Avaliadores de mapeamento

Avaliadores servem para medir a acurácia e verificar a concordância de um mapa com a realidade. Em estudos de acurácia é importante a quantificação do erro em sua frequência, natureza, magnitude e fonte, que respectivamente significam, qual a proporção de área concorda com a realidade, a que a não concorda, a que é atribuída e está sendo confundida, quão importante é o erro para o usuário e o porquê deste ocorrer (STEHMAN; CZAPLEWSKI, 1998).

Para quantificar esses erros existem quatro métodos estatísticos multinominais capazes de aferir quanto a acurácia de mapas de variáveis categóricas como os de classes de solos, estes métodos são a acurácia geral (AG), acurácia do usuário (AU), acurácia do mapeador (AM) e a acurácia pelo coeficiente Kappa de Cohen (Kappa) (ROSSITER, 2004).

A partir das amostras aleatórias comparativas avaliadas em campo se constrói uma matriz de erros colocando em colunas a quantidade de pontos para cada classe de solo encontrada em contraste com a quantidade destes pontos por classe foi indicado pelo mapa estimado nas linhas. Nessa conformação obtemos o total de pontos de amostragem obtidos para cada combinação de classes estimada e real.

Tabela 2 – Exemplo de matriz de erros de comparação entre pontos do mapa estimado e verdade de campo

		Verdade de Campo		
		Classe A	Classe B	Total
Mapa Estimado	Classe A	$X_{11}$	$X_{21}$	$\sum X_{i1}(i=1,2)$
	Classe B	$X_{21}$	$X_{22}$	$\sum X_{i2}(i=1,2)$
	Total	$\sum X_{1j}(j=1,2)$	$\sum X_{2j}(j=1,2)$	$\sum X_{ij}$

Fonte: O Autor (2018).

A acurácia geral (AG) é uma medida geral de adequação do mapa do preditivo. Uma razão entre o total de correspondências das classes com o total de observações, como na equação 4 (BANKO, 1998; ROSSITER, 2004).

$$AG = \frac{(X_{11}+X_{22}+\dots)}{\sum x_{ij}} \quad (4)$$

A acurácia do usuário (AU) de cada classe é uma medida de quanto um usuário ao observar o mapa encontrará em campo a classe mapeada. Uma razão entre o total de

correspondências de uma classe com o total de observações indicadas para essa classe, como equação 5 (BANKO, 1998; ROSSITER, 2004).

$$AU_A = \frac{X_{11}}{\sum X_{i1(i=1,2,3)}} \quad (5)$$

A acurácia do mapeador (AM) de cada classe determina os acertos na área mapeada. Uma razão entre o total de correspondências de uma classe com o total de observações onde essa classe realmente existe, como na equação 6 (BANKO, 1998; ROSSITER, 2004).

$$AM_A = \frac{X_{11}}{\sum X_{1j(j=1,2,3)}} \quad (6)$$

A acurácia pelo coeficiente Kappa de Cohen é uma estatística que leva em consideração as informações empregadas para se obter a AG, bem como a AU e à AM (ROSSITER, 2004). Esse valor é obtido através do conjunto equações 7, 8 e 9.

$$\sum cd = (X_{11} + X_{22} \dots) \quad (7)$$

$$\sum fc = \sum \left[ \frac{(\sum X_{1j(j=1,2)} * \sum X_{i1(i=1,2)})}{(\sum X_{2j(j=1,2)} * \sum X_{i2(i=1,2)}) + (\sum X_{2j(j=1,2)} * \sum X_{i2(i=1,2)} + \dots)} \right] \quad (8)$$

$$Coef\ Kappa = \frac{(\sum cd - \sum fc)}{(\sum X_{ij} - \sum fc)} \quad (9)$$

Para avaliação de acurácia neste trabalho foram coletadas informações de 23 pontos de avaliação de mapeamento dentro da área de estudo por caminhamento em regiões onde o acesso era possível. Um trado foi utilizado para perfurar e avaliar o solo, conforme a figura 4. O aplicativo Avenza® foi utilizado para obter as coordenadas decimais de cada ponto de verificação.

Figura 4 - Tradagens de solo para avaliação



Fonte: O Autor (2018).

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Significância e ajuste de variáveis de terreno

A Tabela 3 apresenta os valores “p” do teste Qui-quadrado para atestar a significância das variáveis de terreno às regressões logísticas, observa-se que predominantemente a altitude e a declividade se apresentam significantes na determinação de classes de solo. Na classe CXa1 o plano de curvatura foi significativo para os *buffers* amostrais de 125 metros. Na classe LVd1 o acúmulo de Fluxo é significativo para os *buffers* amostrais de 100 metros.

Tabela 3 – Significância de variáveis de terreno em relação as classes e os *buffers* estudados e ajuste das regressões logísticas

	CXa1			CXa3			LVd1		
	B75	B100	B125	B75	B100	B125	B75	B100	B125
Altitude	<b>8,51e-5</b>	<b>2,04e-7</b>	<b>8,77e-10</b>	<b>6,15e-34</b>	<b>1,75e-49</b>	<b>6,38e-68</b>	<b>4,83e-05</b>	<b>7,72e-08</b>	<b>5,95e-11</b>
Declividade	<b>7,51e-14</b>	<b>1,22e-18</b>	<b>4,63e-31</b>	<b>0,0004</b>	<b>0,0009</b>	<b>8,88e-06</b>	0,0818	<b>0,0097</b>	<b>0,0007</b>
Curvatura	0,4518	0,6793	0,2531	0,9997	0,7260	0,2571	0,6516	0,8604	0,6492
Plano de Curvatura	0,7087	0,6494	<b>0,0228</b>	0,9998	0,4060	0,9923	0,7003	0,7610	0,2286
Perfil de Curvatura	0,6253	0,2912	0,9275	1,0000	0,4417	0,9744	0,7198	0,4995	0,7342
Acúmulo de Fluxo	0,0552	0,2329	0,1063	0,9999	0,1647	0,5944	0,0805	<b>0,0031</b>	0,1867

Fonte: O Autor (2018).

Nota: valores em negrito foram significativos ao nível de 10% de probabilidade no teste Qui-quadrado. B75, B100 e B125 representam os *buffers* amostrais de 75, 100 e 125 metros respectivamente.

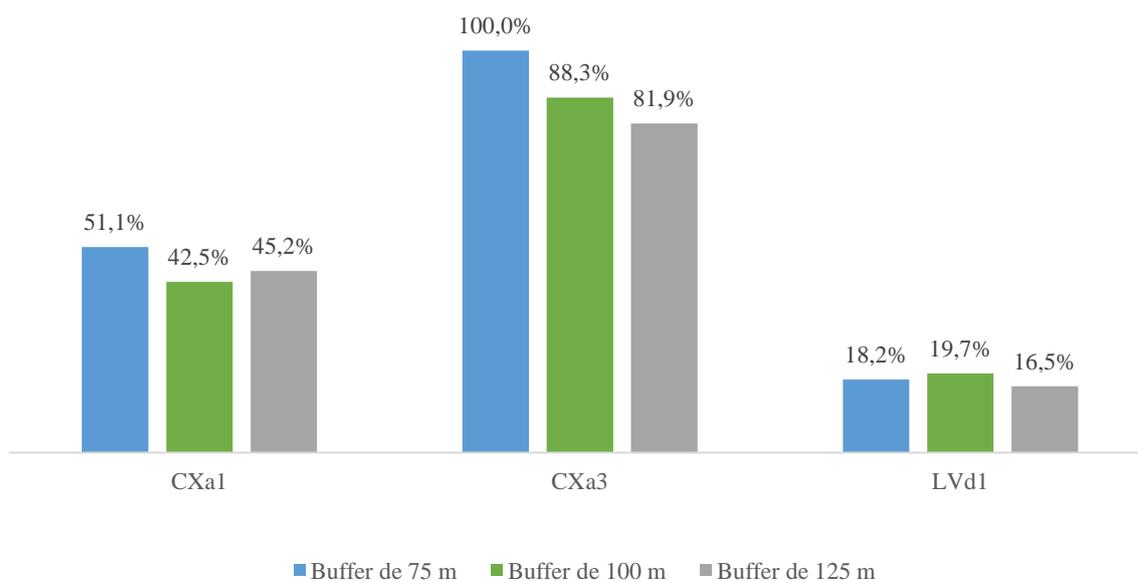
Figueiredo (2006) também obteve altitude e declividade como variáveis significativas em três áreas estudadas por ele no sul do Brasil. Caten (2008) também identificou estas variáveis como significativas para predição de classes de solo, especialmente para Cambissolos. Em um trabalho mais recente, Jeune (2018), também obteve como significativas essas mesmas variáveis.

### 4.2 Ajuste das regressões logísticas

O Gráfico 1 mostra os coeficientes de determinação das regressões logísticas realizadas para os diferentes cenários de tamanho amostral, observa-se que há uma tendência de ajustamento mais forte para o CXa3, seguido pelo CXa1 com valores intermediários e por último os coeficientes de ajustamento para LVd1 que apresentaram os menores valores. Apenas a classe de solo CXa3 apresenta um coeficiente de determinação isolado dos coeficientes dos outros *Buffers* Amostrais dentro da mesma classe, este foi o coeficiente de determinação do

*Buffer* de 75 metros que apresentou o valor de 100%, com mais de 15% de diferença dos demais coeficientes dentro da mesma classe de solo classificada.

Gráfico 1 - Coeficiente de determinação de Nagelkerke

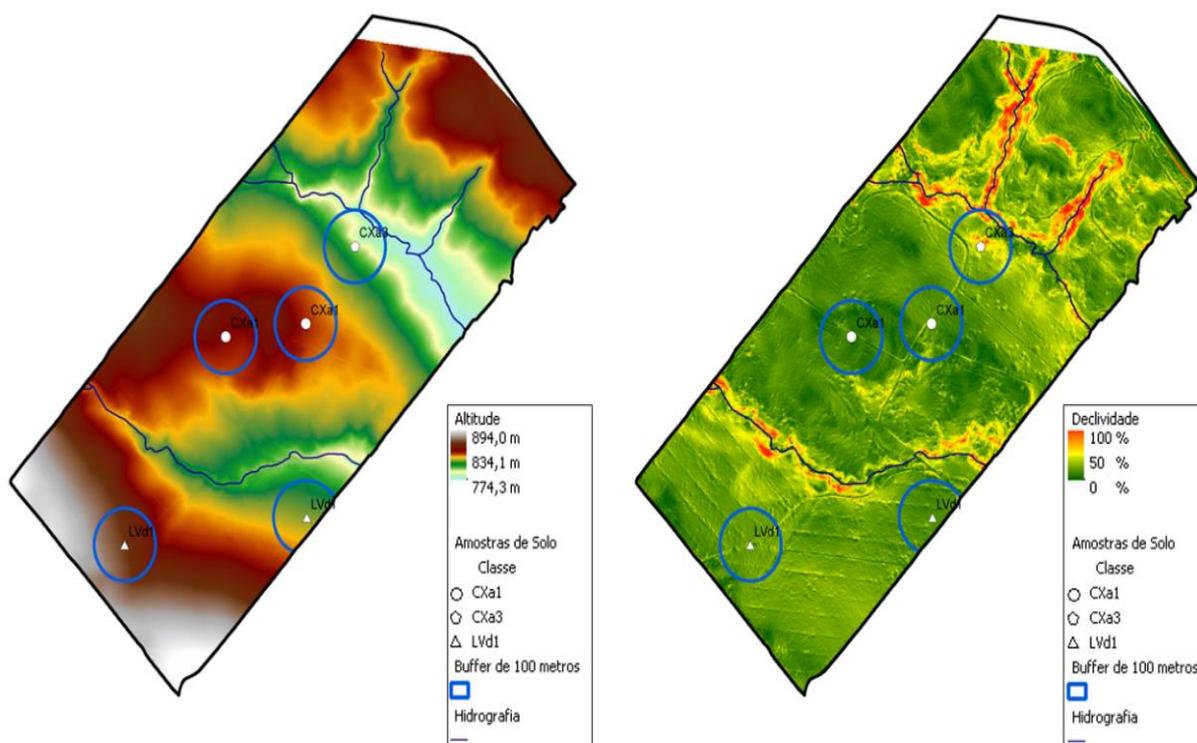


Fonte: O Autor (2018).

Nota: Coeficientes de determinação obtidos pelo método de Nagelkerke (NAGELKERKE, 1991), que tem um funcionamento análogo ao coeficiente de determinação para regressões lineares em representar o ajuste da função em relação aos dados

A tendência de ajuste das regressões para cada classe pode ser explicada pela variação espacial das variáveis de terreno em volta de cada ponto de solo classificado. Seguindo que quanto mais ajustado é a regressão mais aquelas determinadas variáveis de terreno explicam a presença da classe na região em específico. Porém o ajuste decai à medida que a amplitude de variação dos valores das variáveis independentes passa a convergir em sua amplitude entre mais de uma classe de solo, como são os casos respectivamente de CXa3 e LVd1 mostradas na figura 5. Este princípio também rege a forte significância das variáveis de terreno de altitude e declividade observados na tabela 3, visto que as classes concentradas em áreas de cota ou declividade distinta dos demais se correlacionam fortemente com essas amplitudes de variação, outro exemplo disso é o ponto da classe CXa3 que por ser o único ponto em área de menor altitude e de declividade acentuada, configurou a ele um melhor coeficiente de determinação, como observado na figura 5.

Figura 5 - Declividade e altitude com pontos de amostras e buffer



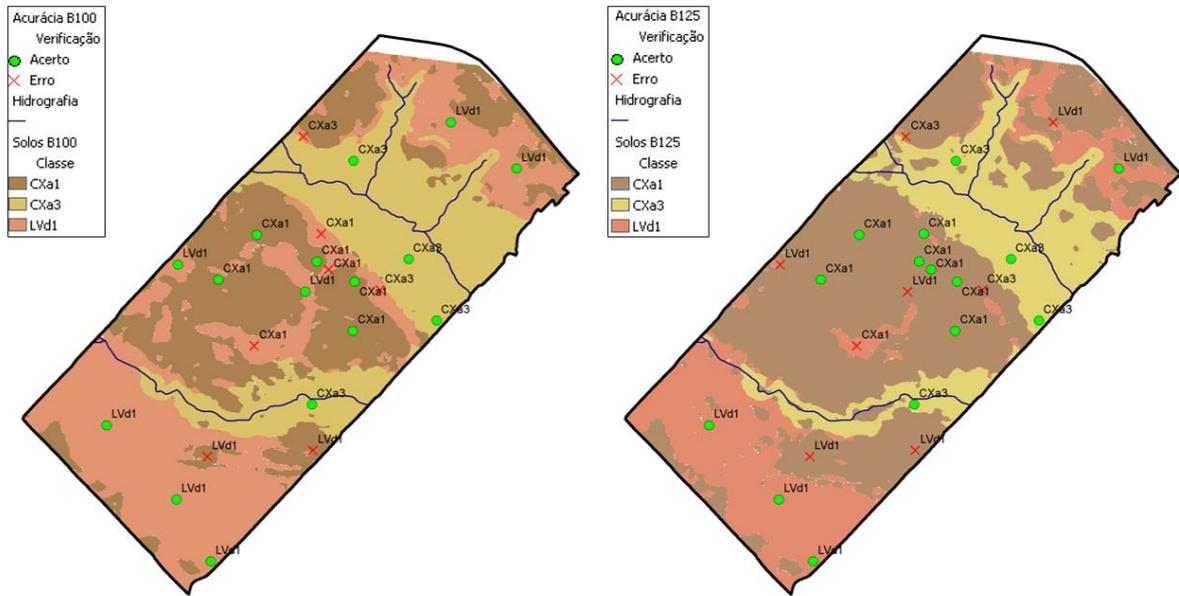
Fonte: O Autor (2018).

### 4.3 Acurácia

A acurácia foi avaliada através de 23 verificações de classe de solo na área de estudo. Os *buffers* de 100 e 75 obtiveram os mesmos valores de acurácia configurando assim mapeamentos bastante similares entre si, já esses tiveram resultados diferente do *buffer* amostral de 125 metros que obteve valores de acurácia distintos. Maiores diferenças de acurácia significam maiores diferenças entre mapas estimados, conforme é visto na figura 6.

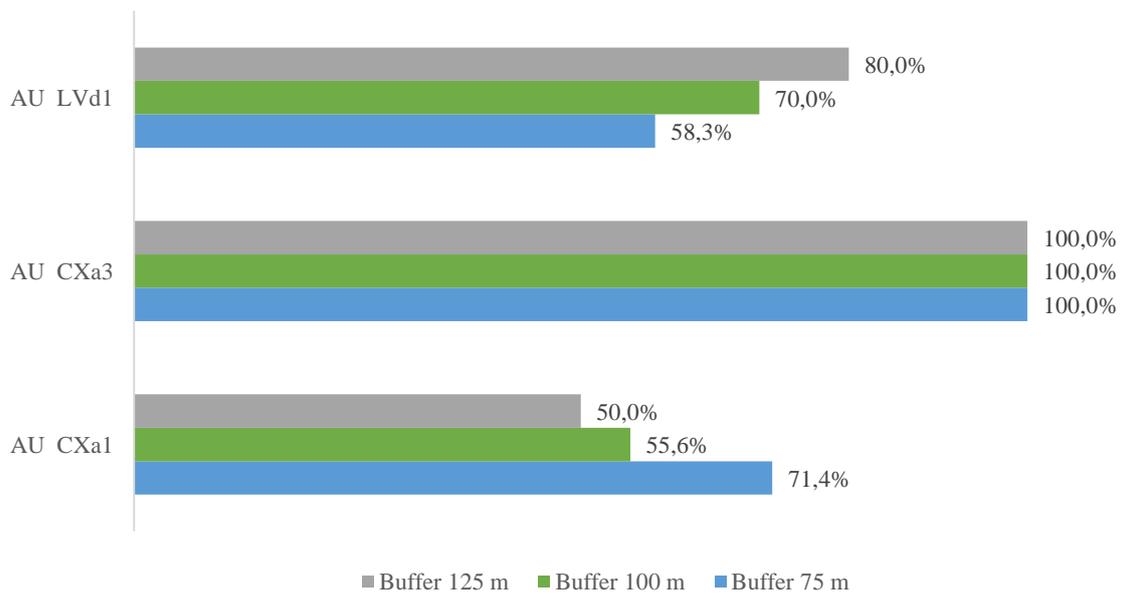
Na avaliação da acurácia do usuário no Gráfico 2, para a classe CXa3 todos os *buffers* amostrais tiveram 100% de acurácia. Isso significa que dentro dos pontos avaliados todos que foram indicados para a classe CXa3 corresponderam com o real. Na classe CXa1 a acurácia é maior para o *Buffer* Amostral de 75 metros, que tem no mínimo 15,8% mais acurácia que os demais *Buffers*. Na classe Lvd1 a acurácia é maior no *Buffer* amostral de 125 metros, que tem no mínimo 10% mais acurácia que os demais *buffers*.

Figura 6 - Mapas estimados de B100 e B125



Fonte: O Autor (2018).

Gráfico 2 - Acurácia do usuário

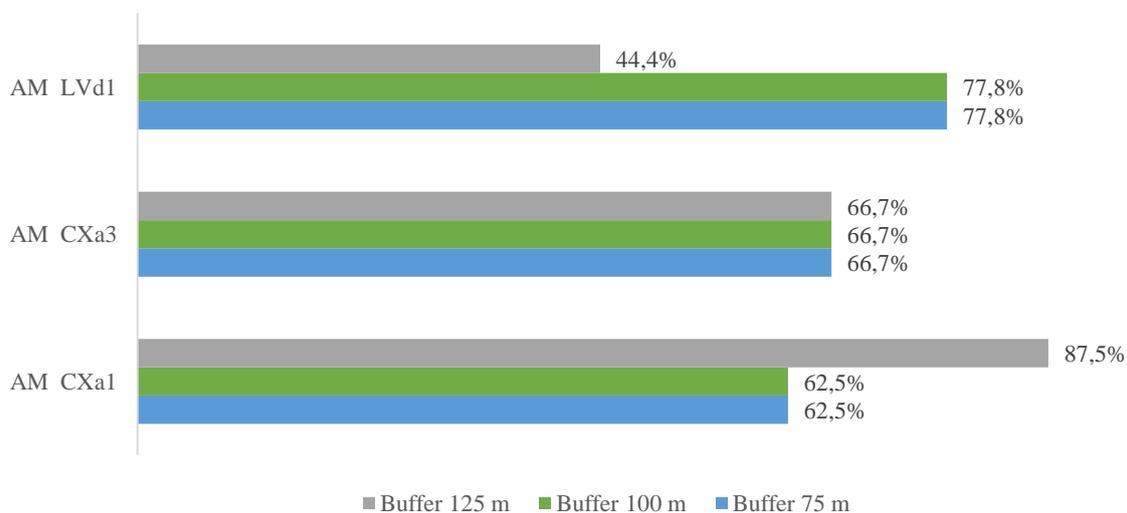


Fonte: O Autor (2018).

Na avaliação da acurácia do mapeador no Gráfico 3, a classe CXa3 tem o valor de 66,7% de acurácia em todos os *buffers* amostrais. Isso significa que dos pontos avaliados 66,7% que eram da classe CXa3 foram indicados pelo mapeamento estimado dos três *buffers* amostrais. Na Classe CXa1 o *Buffer* amostral de 125 metros possui a maior acurácia, com 87,5%, 25% a mais que os demais *buffers*. Na Classe LVd1 os *Buffers* Amostrais de 75 e 100 metros possuem

uma acurácia de 77,8%, 33,4% a mais de acurácia que o *buffer* amostral de 125 metros. Observa-se que a acurácia decresce na ordem de CXa1, CXa3 e LVd1 para o *Buffer* de 125 metros e cresce na mesma ordem para os demais *buffers*.

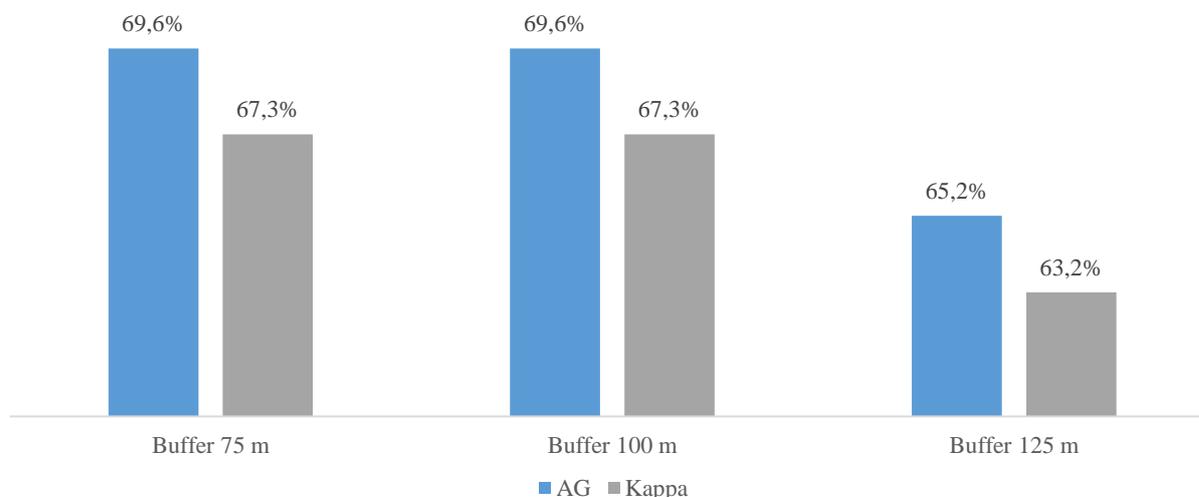
Gráfico 3 - Acurácia do mapeador



Fonte: O Autor (2018).

Para os avaliadores de acurácia generalizados no Gráfico 4, os *Buffers* Amostrais de 75 e 100 metros foram superiores e não tiveram diferenças entre si, com 69,6 % de Acurácia Geral e 67,3% de coeficiente Kappa. Segundo Zeilhofer (2006), valores entre 30 e 60% podem ser encontrados para modelos de preditivos para mapeamento de solo, e estes valores podem ser aumentados para algo maior que 60% com simplificação de legenda. Neste estudo os valores foram superiores a 60% e com a simplificação da legenda para o Nível de Ordem de Solo a Acurácia Geral não teve alteração e o coeficiente Kappa teve um pequeno aumento para 68,2%.

Gráfico 4 - Acurácia geral e coeficiente Kappa



Fonte: O Autor (2018).

Figueiredo (2006), usando uma legenda simplificada obteve AG de até 67,7% e coeficiente Kappa de até 50,3%, usando Regressões Logísticas Múltiplas baseadas em mapas de solos de escala 1:50000 e MDE de 90 m<sup>2</sup> de resolução.

Trabalhos recentes em diferentes metodologias de mapeamento digital de classes de solo, como de Rizzo (2016) e Xie (2018), encontraram respectivamente 41 e 43% em valores Kappa para legendas simplificadas para o nível de Ordem do Solo.

#### 4.4 Discussões gerais

As acurácias de CXa1 e LVd1 não seguem um padrão definido apresentando alternâncias com a mudança dos *buffers*, essas mesmas classes apresentam coeficientes de determinação baixos em relação a classe CXa3 que tem a tendência de conservar valores de acurácia fixos com a mudança de *buffer*, isso indica que os baixos ajustes destas regressões podem estar ligados a variação de suas acurácias. Considera-se também que os baixos ajustes por sua vez são influenciados pelo compartilhamento da amplitude dos valores das variáveis de terreno significativas como Altitude e Declividade e que mesmo que significativas ainda é necessário a exploração de variáveis significativas que possam se correlacionar com as classes de solo de modo a distingui-las com efetividade.

Em perspectiva geral quando duas ou mais referências amostrais de classe de solo estiverem compartilhando amplitude, suas variáveis de terreno, suas respectivas acurácias e ajustes tendem a diminuir.

## 5 CONCLUSÕES

As variáveis de altitude e declividade são as mais significativas em mapeamentos com essa metodologia.

A qualidade do MDS realizado para Fazenda Guanabara mostrou valores superiores aos resultados frequentemente encontrados para mapeamentos digitais de classes de solo para os tamanhos amostrais de *Buffer* de 75 e 100 metros.

O uso de pontos de amostra de classificação do solo com *buffers* amostrais para amostragem das variáveis de terreno tem influências dinâmicas na acurácia do MDS e dependem da posição das amostras de solo em relação as variáveis de terreno e na capacidade destas de predizer e distinguir as classes de solo.

## REFERÊNCIAS

ARRUDA, G.P. et al. Mapeamento digital de solos por redes neurais artificiais com base na relação solo-paisagem. **Revista Brasileira de Ciência do Solo**, v.37, n.2, p.327-338, 2013. DOI: <https://doi.org/10.1590/S0100-06832013000200004>.

BANKO, G. **A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory**. Laxenburg: International Institute for Applied Systems Analysis, 1998. 36 p.

BREVIK, E.C. et al. Historical perspectives and future needs in soil mapping, classification, and pedologic modeling. **Geoderma**, v.264, p.253–255, 2016. DOI: <https://doi.org/10.1016/j.geoderma.2015.09.022>.

BURROUGH, P.A. **Principles of Geographic Information Systems for Land Resource Assessment. Monographs on Soil and Resources Survey**. No. 12. New York: Oxford Science Publications, 1986. 193 p.

CATEN, A. **Aplicação de componentes principais e regressões logísticas múltiplas em sistema de informações geográficas para a predição e o mapeamento digital de solos**. 2008. 128 f. Dissertação (Mestrado) - Universidade Federal de Santa Maria, Santa Maria, 2008. Disponível em: <https://repositorio.ufsm.br/handle/1/5483>. Acesso em 30 jul. 2022.

CHATTERJEE, S.; HADI, A.S. **Regression analysis by example**. 4. ed. New York: John Wiley & Sons, 2006. 375 p.

COELHO, F.B. **Comparação de métodos de mapeamento digital de solos através de variáveis geométricas e sistemas de informações geográficas**. 2010. 95 f. Dissertação (Mestrado) - Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010. Disponível em: <https://lume.ufrgs.br/handle/10183/25062>. Acesso em 30 jul. 2022.

EMBRAPA. **Boletim de Pesquisa e Desenvolvimento 10** - Levantamento de reconhecimento dos solos da região central do estado do Paraná (área 8), Rio de Janeiro: Embrapa Solos, 2002. 156 p. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/159398/1/BPD-10-2002-Parana-Central-Area8-.pdf>. Acesso em 30 jul. 2022.

EMBRAPA. **Procedimentos normativos de levantamentos pedológicos**. Rio de Janeiro: Embrapa, 1995. 101 p.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 5 ed. Brasília: Embrapa, 2018. 353 p.

ESRI. **ArcMap 10.5**. Environmental Systems Research Institute, Inc. Redlands, Califórnia. 2017. Disponível em: <http://desktop.arcgis.com/en/arcmap/>. Acesso em 30 jul. 2022.

FIGUEIRA, C.V. **Modelos de Regressão Logística**. 2006. 149 f. Dissertação (Mestrado em matemática) Programa de Pós-graduação em matemática do Instituto de Matemática,

Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <https://www.lume.ufrgs.br/handle/10183/8192>. Acesso em 30 jul. 2022.

FIGUEIREDO, S.R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. 2006. 93 f. Dissertação (Mestrado)- Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <https://lume.ufrgs.br/handle/10183/15815>. Acesso em 30 jul. 2022.

HARTEMINK, A.E.; MCBRATNEY, A.B. A soil science renaissance. **Geoderma**, v.148, n.2, p.123-129, 2008. DOI: <https://doi.org/10.1016/j.geoderma.2008.10.006>.

HENGL, T. **Pedometric mapping: bringing the gaps between conventional and pedometric approaches**. Enschede: International Institute for Geo-Information Science and Earth Observation (ITC), 2003. 214 p.

HUDSON, B.D. The soil survey as paradigm-based science. **Soil Science Society of America Journal**, v.56, n.3, p.836-841, 1992. DOI: <https://doi.org/10.2136/sssaj1992.03615995005600030027x>.

JENSON S.K.; DOMINGUE, J.O. Extracting topographic structure from digital elevation data for geographic information system analysis. **Photogrammetric Engineering and Remote Sensing**, v.54, n.11, p.1593-1600, 1988.

JEUNE, W. et al. Multinomial logistic regression and random forest classifiers in digital mapping of soil classes in western Haiti. **Rev. Bras. Ciênc. Solo**, Viçosa, v.42, 2018. DOI: <https://doi.org/10.1590/18069657rbc20170133>.

LÓPEZ, L.R. **Pedologia Quantitativa: espectrometria VIS-NIR-SWIR e mapeamento digital de solos**. 2009. 171 f. Dissertação (Mestrado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2009. Disponível em: <https://www.teses.usp.br/teses/disponiveis/11/111140/tde-23062009-140151/pt-br.php>. Acesso em 30 jul. 2022.

LORENZETTI, R., et al. Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small scale maps. **Geoderma**, v.237-238, p.237-245, 2015. DOI: <https://doi.org/10.1016/j.geoderma.2014.09.006>.

MACMILLAN, R.A. et al. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). **Comput. Environ. Urban. Syst.**, v.28, n.3, p.175-200, 2004. DOI: [https://doi.org/10.1016/S0198-9715\(03\)00019-X](https://doi.org/10.1016/S0198-9715(03)00019-X).

MARK, D.M.; CSILLAG, F. The nature of boundaries on 'area-class' maps. **Cartographica**, v.26, n.1, p.65-78, 1989. DOI: <https://doi.org/10.3138/D235-3262-062X-4472>.

McBRATNEY, A.B. et al. An overview of pedometric techniques for use in soil survey. **Geoderma**, Amsterdam, v.97, n.3-4, p.293-327, 2000. DOI: [https://doi.org/10.1016/S0016-7061\(00\)00043-4](https://doi.org/10.1016/S0016-7061(00)00043-4).

McBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, n.1-2, p.3-52, 2003. DOI: [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).

McBRATNEY, A.B.; GRUIJTER, J.J. A continuum approach to soil classification by modified fuzzy k-means with extragrades. **Jornal Soil Science**, v.43, n.1, p.159–175, 1992. DOI: <https://doi.org/10.1111/j.1365-2389.1992.tb00127.x>.

MENDONÇA-SANTOS, M.L.; SANTOS, H.G. **Mapeamento digital de classes e atributos de solos: métodos, paradigmas e novas técnicas**. Rio de Janeiro: Embrapa, 2003. 19 p.

MINASNY B.; MCBRATNEY A.B. Digital soil mapping: a brief history and some lessons. **Geoderma**. v. 264, p. 301-311, 2016. DOI: <https://doi.org/10.1016/j.geoderma.2015.07.017>.

MUÑOZ-ROJAS M. et al. Soil mapping and processes models to support climate change mitigation and adaptation strategies: a review. In: 19th EGU General Assembly, EGU2017-10677. **Anais [...]**. Vienna, Austria: Geophysical Research Abstracts, 2017. p.10677.

NAGELKERKE, N.J.D. A note on a general definition of the coefficient of determination. **Biometrika**, v.78, n.3, p.691–692, 1991. DOI: <https://doi.org/10.1093/biomet/78.3.691>.

NEUMANN, M.R.B. **Mapeamento digital de solos no Distrito Federal**. 2012. 124 f. Tese (Doutorado) – Instituto de Geociências, Universidade de Brasília, Brasília, 2012. Disponível em: <https://repositorio.unb.br/handle/10482/11026>. Acesso em 30 jul. 2022.

ODEH, I.O.A.; McBRATNEY, A.B.; CHITTLEBOROUGH, D.J. Design of optimal sample spacings for mapping soil using fuzzy-k-means and regionalizes variable theory. **Geoderma**, Amsterdam, v.47, n.1-2, p.93-122, 1990. DOI: [https://doi.org/10.1016/0016-7061\(90\)90049-F](https://doi.org/10.1016/0016-7061(90)90049-F).

PARK, S.J.; MCSWEENEY, K.; LOWERY, B. Identification of the spatial distribution of soils using a process-based terrain characterization. **Geoderma**, v.103, n.3-4, p.249–272, 2001. DOI: [https://doi.org/10.1016/S0016-7061\(01\)00042-8](https://doi.org/10.1016/S0016-7061(01)00042-8).

R DEVELOPMENT CORE TEAM. R. **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <http://www.R-project.org>. Acesso em 30 jul. 2022.

RANZANI, G. **Manual de levantamento de solos**. São Paulo: Edgar Blucher, 1969. 167 p.

RIZZO R. et al. Digital soil mapping at local scale using a multi-depth Vis-NIR spectral library and terrain attributes. **Geoderma**, v.274, p.18-27, 2016. DOI: <https://doi.org/10.1016/j.geoderma.2016.03.019>.

ROSSITER, D.G. **Statistical methods for accuracy assessment of classified thematic maps**. Nota Técnica, Enschede (NL): International Institute for Geo-information Science & Earth Observation (ITC), 2004. 46 p.

SCULL, P. et al. Predictive soil mapping: a review. **Progress in Physical Geography: Earth and Environment**, v.27, n.2, p.171-197, 2003. DOI: <https://doi.org/10.1191/0309133303pp366ra>.

SILVEIRA, C.T. **Análise digital do relevo na predição de unidades preliminares de mapeamentos de solos: integração de atributos topográficos em sistemas de informações geográficas e redes neurais artificiais**. 2010. 154 f. Tese (Doutorado)- Universidade Federal do Paraná, 2010. Disponível em: <https://acervodigital.ufpr.br/handle/1884/24117>. Acesso em 30 jul. 2022.

SIRTOLI, A.V.; et al. Atributos do relevo derivados de modelo digital de elevação e suas relações com solos, **Scientia Agraria**, Curitiba, v.9, n.3, p.317-329, 2008. DOI: <http://dx.doi.org/10.5380/rsa.v9i3.11517>.

STEHMAN, S.V.; CZAPLEWSKI, R.L. Design and analysis of thematic map accuracy assessment: fundamental principles. **Remote Sensing of Environment**, v.64, n.3, p.331-344, 1998. DOI: [https://doi.org/10.1016/S0034-4257\(98\)00010-8](https://doi.org/10.1016/S0034-4257(98)00010-8).

WEBSTER, R. Quantitative spatial analysis of soil in the field. **Advances in soil Science**, New York, v.3, p.1-70, 1984. DOI: [https://doi.org/10.1007/978-1-4612-5090-6\\_1](https://doi.org/10.1007/978-1-4612-5090-6_1).

WILSON, J.P.; GALLANT, J.C. Digital terrain analysis. In: \_\_\_\_\_ (Ed.). *Terrain analysis: principles and applications*. New York: **Wiley & Sons**, p.1-27, 2000.

XIE, X.L.; LI, A.B. Identification of soil profile classes using depth-weighted visible–near-infrared spectral reflectance. **Geoderma**, v.325, p.90–101, 2018. DOI: <https://doi.org/10.1016/j.geoderma.2018.03.029>.

ZEILHOFER, P. Soil Mapping in the Pantanal of Mato Grosso, Brazil, using Multitemporal Landsat TM data. **Wetlands Ecology and Management**, v.14, n.5, p.445-461, 2006. DOI: <https://doi.org/10.1007/s11273-006-0007-2>.