

**COMPARAÇÃO DE MÉTODOS DE  
MAPEAMENTO DIGITAL DE SOLOS ATRAVÉS  
DE VARIÁVEIS GEOMORFOMÉTRICAS E  
SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS**

FABRÍCIO FERNANDES COELHO

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E  
METEOROLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO**

**COMPARAÇÃO DE MÉTODOS DE MAPEAMENTO DIGITAL DE  
SOLOS ATRAVÉS DE VARIÁVEIS GEOMORFOMÉTRICAS E  
SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS**

**FABRÍCIO FERNANDES COELHO**

**ORIENTADOR: PROF. DR. ELVIO GIASSON**

**BANCA EXAMINADORA: Profa. Dra. Eliana Lima da Fonseca  
Prof. Dr. Carlos Gustavo Tornquist  
Prof. Dr. Laurindo Antônio Guasselli**

**Dissertação de Mestrado apresentada como  
requisito para obtenção do Título de Mestre  
em Sensoriamento Remoto.**

Porto Alegre - 2010

## AGRADECIMENTO

Em primeiro lugar a Deus, fonte da ciência, por toda força nessa caminhada.

À minha família, especialmente aos meus pais Samuel José Coelho e Odila Fernandes Coelho pelo amor e dedicação e ao meu irmão Cássio Fernandes Coelho por estar sempre presente e pronto a ajudar.

À minha namorada Elisane da Costa, meu refúgio, pelo companheirismo, amor e paciência.

Aos meus colegas do curso de Pós-Graduação em Sensoriamento Remoto, em especial aos colegas Guilherme Oliveira, Vagner Conceição, Alan Bisso, Celso Rodrigues, Juliana Silveira, Daniel Zanotta e José Caresani pelo coleguismo construído.

Aos professores do curso de Pós-Graduação em Sensoriamento Remoto, em especial ao professor Dr. Elvio Giasson, exemplo de profissional, pela orientação e competência para nortear essa dissertação.

À Universidade Federal do Rio Grande do Sul (UFRGS) e ao Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPSRM) pela oportunidade que me foi dada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento do curso e à todos que de alguma maneira contribuíram para o desenvolvimento dessa dissertação.

*“A utilização do  
GIS está limitada exclusivamente  
pela imaginação humana”*

Jack Dangermond

## RESUMO

Mapas pedológicos são fontes de informações primordiais para planejamento e manejo de uso do solo, porém apresentam altos custos de produção. A fim de produzir mapas de solos a partir de mapas existentes, o presente trabalho objetiva testar e comparar métodos de classificação em estágio único (regressões logísticas múltiplas multinomiais e Bayes) e em estágios múltiplos (CART, J48 e LMT) com utilização de sistemas de informações geográficas e de variáveis geomorfométricas para produção de mapas pedológicos com legenda original e simplificada. A base de dados foi gerenciada em ambiente ArcGis onde as variáveis e o mapa original foram relacionados através de amostras de treinamento para os algoritmos. O resultado dos algoritmos obtidos no software Weka foram implementados no ArcGis para a confecção dos mapas. Foram gerados matrizes de erros para análise de acurácias dos mapas. As variáveis geomorfométricas de declividade, perfil e plano de curvatura, elevação e índice de umidade topográfica são aquelas que melhor explicam a distribuição espacial das classes de solo. Os métodos de classificação em estágio múltiplo apresentaram sensíveis melhoras nas acurácias globais, porém significativas melhoras nos índices Kappa. A utilização de legenda simplificada aumentou significativamente as acurácias do produtor e do usuário, porém sensível melhora na acurácia global e índice Kappa.

**Palavras-chave:** modelo digital de elevação, variáveis geomorfométricas, classificação em estágio único, classificador de decisão em árvore.

## ABSTRACT

Soil maps are sources of important information for land planning and management, but are expensive to produce. This study proposes testing and comparing single stage classification methods (multiple multinomial logistic regression and Bayes) and multiple stage classification methods (CART, J48 and LMT) using geographic information system and terrain parameters for producing soil maps with both original and simplified legend. In ArcGis environment terrain parameters and original soil map were sampled for training algorithms. The results from statistical software Weka were implemented in ArcGis environment to generate digital soil maps. Error matrices were generated for analysis accuracies of the maps. The terrain parameters that best explained soil distribution were slope, profile and planar curvature, elevation, and topographic wetness index. The multiple stage classification methods showed small improvements in overall accuracies and large improvements in the Kappa index. Simplification of the original legend significantly increased the producer and user accuracies, however produced small improvements in overall accuracies and Kappa index.

**Key words:** digital elevation model, terrain parameters, single stage classification, classification trees.

## LISTA DE FIGURAS

<b>Figura 01</b> - Localização da área de estudo.....	<b>5</b>
<b>Figura 02</b> - Mapa hipsométrico da área de estudo.....	<b>6</b>
<b>Figura 03</b> - Mapa de solos da área de estudo.....	<b>7</b>
<b>Figura 04</b> - Exemplo de Topossequência na Depressão Central do Rio Grande do Sul.....	<b>12</b>
<b>Figura 05</b> - Modelo dos processos geomorfológicos em uma vertente.....	<b>13</b>
<b>Figura 06</b> - Pedometria em relação a outros ramos da ciência.....	<b>15</b>
<b>Figura 07</b> - Submatriz de elevação utilizada para calcular as variáveis geomorfométricas.....	<b>18</b>
<b>Figura 08</b> - Determinação da direção do fluxo a partir de seus oito vizinhos.....	<b>22</b>
<b>Figura 09</b> - Limiar bayesiano $t$ e não bayesiano $t'$ .....	<b>24</b>
<b>Figura 10</b> - Estrutura geral de um CDA.....	<b>26</b>
<b>Figura 11</b> - Estrutura dos CDA CART e J48 (A) e LMT (B).....	<b>28</b>
<b>Figura 12</b> - Fluxo representativo dos procedimentos metodológicos.....	<b>29</b>
<b>Figura 13</b> - Estrutura dos CDA CART e J48.....	<b>35</b>
<b>Figura 14</b> - Estrutura do CDA LMT.....	<b>36</b>
<b>Figura 15</b> - Subconjunto das variáveis selecionadas a partir do método SFS no Weka .....	<b>39</b>
<b>Figura 16</b> - Histogramas das classes para cada variável geomorfométrica.....	<b>41</b>
<b>Figura 17</b> - Mapa de solos original (a) e mapa de solo produzido pela classificação de Bayes (b).....	<b>42</b>

<b>Figura 18</b> - Histogramas das variáveis para as classes agrupadas.....	<b>44</b>
<b>Figura 19</b> - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pela classificação de Bayes com legenda simplificada (b).....	<b>46</b>
<b>Figura 20</b> - Mapas de probabilidade de ocorrência de cada classe de solo a partir dos coeficientes do modelo logístico.....	<b>47</b>
<b>Figura 21</b> - Mapa de solo original (a) e mapa de solo produzido por RLMM (b).....	<b>49</b>
<b>Figura 22</b> - Mapa de solos com legenda simplificada (a) e mapa de solo produzido por RLMM com legenda simplificada (b).....	<b>51</b>
<b>Figura 23</b> - Árvore de decisão do CDA CART.....	<b>53</b>
<b>Figura 24</b> - Mapa de solo original (a) e mapa produzido pelo CDA CART (b).....	<b>54</b>
<b>Figura 25</b> - Mapa de solos com legenda simplificada (a) e mapa de solos produzido pelo CDA CART com legenda simplificada (b).....	<b>56</b>
<b>Figura 26</b> - Árvore de decisão do CDA J48.....	<b>57</b>
<b>Figura 27</b> - Mapa de solos original (a) e mapa produzido pelo CDA J48 (b).....	<b>59</b>
<b>Figura 28</b> - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pelo CDA J48 com legenda simplificada (b).....	<b>61</b>
<b>Figura 29</b> - Árvore de decisão resultante do método LMT.....	<b>62</b>
<b>Figura 30</b> - Mapa de solos original (a) e mapa produzido pelo CDA LMT (b).....	<b>65</b>
<b>Figura 31</b> - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pelo CDA LMT com legenda simplificada (b).....	<b>67</b>



## LISTA DE TABELAS

<b>Tabela 01</b> - Área ocupada pelas classes de solo, segundo mapeamento de ABRÃO & AZOLIN (1970) atualizado e digitalizado por BORTOLON (2008).....	<b>8</b>
<b>Tabela 02</b> - Matriz de correlação entre as variáveis geomorfométricas.....	<b>40</b>
<b>Tabela 03</b> - Matriz de erro do classificador de Bayes a partir do mapa original de solos.....	<b>43</b>
<b>Tabela 04</b> - Matriz de erro do classificador de Bayes a partir do mapa de solos, ambos com legenda simplificada.....	<b>45</b>
<b>Tabela 05</b> - Coeficientes do modelo logístico.....	<b>47</b>
<b>Tabela 06</b> - Matriz de erro do mapa produzido por RLMM a partir do mapa original de solos.....	<b>50</b>
<b>Tabela 07</b> - Matriz de erro do mapa produzido por RLMM a partir do mapa de solos, ambos com legenda simplificada.....	<b>52</b>
<b>Tabela 08</b> - Matriz de erro do mapa produzido pelo CDA CART e o mapa original de solos.....	<b>55</b>
<b>Tabela 09</b> - Matriz de erro do mapa produzido pelo CDA CART e o mapa de solos, ambos com legenda simplificada.....	<b>57</b>
<b>Tabela 10</b> - Matriz de erro do mapa produzido pelo CDA J48 e o mapa original de solos.....	<b>58</b>
<b>Tabela 11</b> - Matriz de erro do mapa produzido por J48 e o mapa de solos, com legenda simplificada.....	<b>60</b>
<b>Tabela 12</b> - Modelos logísticos resultandos nos nós terminais da árvore de decisão	

do LMT.....	63
<b>Tabela 13</b> - Matriz de erro do mapa produzido pelo classificador LMT e o mapa original de solos.....	64
<b>Tabela 14</b> - Matriz de erro do mapa produzido por LMT e o mapa de solos, ambos com legenda simplificada.....	66
<b>Tabela 15</b> - Síntese das AG e índice Kappa dos mapas produzidos com legenda original e simplificada.....	68
<b>Tabela 16</b> - Síntese das AU e AP dos mapas produzidos por legenda original e por legenda simplificada.....	70

## LISTA DE ABREVIATURAS

<b>AG</b>	Acurácia Global
<b>AP</b>	Acurácia do Produtor
<b>APm</b>	Acurácia do Produtor média
<b>AU</b>	Acurácia do Usuário
<b>AUm</b>	Acurácia do Usuário média
<b>CART</b>	<i>Classification And Regression Trees</i>
<b>CDA</b>	Classificação de Decisão em Árvore
<b>CEU</b>	Classificação em Estágio Único
<b>CSV</b>	<i>Comma Separated Values</i>
<b>D8</b>	<i>Deterministic Eight Neighbours</i>
<b>EMBRAPA</b>	Empresa Brasileira de Pesquisa Agropecuária
<b>ESRI</b>	<i>Environmental Systems Research Institute</i>
<b>FAO</b>	<i>Food and Agricultural Organization</i>
<b>GPS</b>	<i>Global Positioning System</i>
<b>InSAR</b>	<i>Interferometric Synthetic Aperture Radar</i>
<b>IUT</b>	Índice de Úmidade Topográfica
<b>LM</b>	<i>Logistic Model</i>
<b>LMT</b>	<i>Logistic Model Trees</i>
<b>LV</b>	Latossolo Vermelho
<b>LVdf</b>	Latossolo Vermelho Distroférico Típico
<b>LVdh</b>	Latossolo Vermelho Distrófico Húmico
<b>LVdt</b>	Latossolo Vermelho Distrófico Típico

<b>M</b>	Chernossolo
<b>MDE</b>	Modelo Digital de Elevação
<b>MEo</b>	Chernossolo Ebânico Órtico
<b>MNT</b>	Modelo Numérico do Terreno
<b>MTo</b>	Chernossolo Argilúvico Órtico
<b>NASA</b>	<i>National Aeronautics and Space Administration</i>
<b>RLMM</b>	Regressões Logísticas Múltiplas Multinomiais
<b>RNA</b>	Redes Neurais Artificiais
<b>RRe</b>	Neossolo Regolítico Eutrófico
<b>RY</b>	Neossolo Flúvico
<b>SiBCS</b>	Sistema Brasileiro de Classificação de Solos
<b>SFS</b>	<i>Sequential Forward Selection</i>
<b>SIG</b>	Sistemas de Informações Geográficas
<b>SRTM</b>	<i>Shuttle Radar Topography Mission</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>TauDEM</b>	<i>Terrain Analysis Using Digital Elevation Models</i>
<b>USDA</b>	<i>United States Department of Agriculture</i>
<b>WEKA</b>	<i>Waikato Environment for Knowledge Analysis</i>

## SUMÁRIO

Resumo.....	iv
Abstract.....	v
Lista de Figuras.....	vi
Lista de Tabelas.....	viii
Lista de Abreviaturas.....	x
<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>2. HIPÓTESE E OBJETIVOS.....</b>	<b>3</b>
2.1. HIPÓTESE.....	3
2.2. OBJETIVOS.....	3
2.2.1. Objetivo Geral.....	3
2.2.2. Objetivos Específicos.....	4
<b>3. LOCALIZAÇÃO E CARACTERIZAÇÃO DA ÁREA DE ESTUDO.....</b>	<b>5</b>
<b>4. REVISÃO BIBLIOGRÁFICA.....</b>	<b>9</b>
4.1. SOLOS.....	9
4.1.1. Definições de Solos.....	9
4.1.2. Classificação de Solos.....	10
4.2. RELAÇÃO SOLOS-GEOMORFOLOGIA.....	11
4.3. MAPEAMENTO DE SOLOS.....	14
4.3.1. Mapeamento Tradicional de Solos.....	14
4.3.2. Mapeamento Digital de Solos.....	14
4.4. VARIÁVEIS DO TERRENO.....	16

4.4.1. Modelo Digital de Elevação – SRTM.....	16
4.4.2. Variáveis Geomorfométricas.....	17
4.4.3. Variáveis Hidrográficas Derivadas do MDE.....	21
4.5. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU).....	23
4.5.1. Classificador de Bayes.....	23
4.5.2. Regressões Logísticas Múltiplas Multinomiais (RLMM).....	24
4.6. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA).....	26
<b>5. METODOLOGIA.....</b>	<b>29</b>
5.1. GERAÇÃO DA BASE DE DADOS.....	30
5.2. PROCESSAMENTO DAS VARIÁVEIS GEOMORFOMÉTRICAS.....	30
5.3. AMOSTRAGEM E SELEÇÃO DAS VARIÁVEIS.....	31
5.4. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU).....	32
5.4.1. Classificação de Bayes.....	32
5.4.2. Regressões Logísticas Múltiplas Multinomiais (RLMM).....	33
5.5. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA).....	34
5.6. ANÁLISE DAS ACURÁCIAS.....	37
<b>6. RESULTADOS.....</b>	<b>39</b>
6.1. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU).....	40
6.1.1. Classificação de Bayes.....	40
6.1.2. Regressões Logísticas Múltiplas Multinomiais (RLMM).....	47
6.2. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA).....	52
6.2.1. <i>Classification And Regression Trees</i> (CART).....	52
6.2.2. Classificador J48 (C4.5).....	57
6.2.3. <i>Logistic Model Trees</i> (LMT).....	62

6.3. ANÁLISE DOS RESULTADOS.....	68
<b>7. CONCLUSÕES.....</b>	<b>73</b>
<b>8. REFERÊNCIAS.....</b>	<b>75</b>

# 1. INTRODUÇÃO

A disponibilidade de informações sobre solos, sua característica e distribuição, são fundamentais para planejamento de atividades relacionadas ao manejo, uso e potencialidades de uso dos solos. O Brasil possui a totalidade de seu território coberto por mapas de solos, na escala 1:1.000.000, gerados através de levantamentos realizados em 1986 publicado pelo projeto RADAMBRASIL. O Rio Grande do Sul possui seu território coberto por mapa de solos, na escala 1:750.000, resultante do Levantamento de Reconhecimento de Solos do Estado do Rio Grande do Sul (BRASIL, 1973).

Embora esses dados estejam disponíveis para o público em geral, são dados com escala pequena, não adequadas para planejamento de uso e manejo dos solos em município, propriedade rural ou bacia hidrográfica (DALMOLIN et al, 2004), por exemplo. Mapas com escalas pequenas podem ser utilizados para planejamento em grandes regiões, zoneamento, construção de ferrovias e rodovias entre outros (EMBRAPA, 1995). Para estudos em escala municipal, o ideal é utilizar mapas com escalas maiores ou iguais a 1:50.000. Tanto no Brasil quanto no Rio Grande do Sul, mapas com escalas maiores ou iguais a 1:50.000 existem apenas em poucas regiões (FIGUEIREDO, 2006).

Como os custos e o tempo necessários para elaboração de mapas com escalas de maior detalhamento, utilizando métodos convencionais de mapeamento, são muitos altos, métodos de mapeamento supervisionados, por meios computacionais, estão sendo trabalhados e refinados, com a finalidade de subsidiar a falta desses dados. Deste modo, o uso de métodos organizados em sistema de informação geográfica (SIG), baseados na elaboração de variáveis derivadas do modelo digital de elevação (MDE) e associados à distribuição de ocorrência das classes de solos, assumem grande importância pelo alto custo-benefício para mapeamento de solos em escalas detalhadas e semi-detalhadas.

O mapeamento de solos realizado por métodos convencionais, é executado com base no modelo solo-paisagem, o que significa dizer que mudanças em um ou mais dos fatores de formação do solo permitem localizar com precisão os limites entre classes de solos diferentes. Considerando a relação dos solos com as variáveis



geomorfométricas, para LAGACHERIE & MCBRATNEY (2007), o mapeamento digital de solos se baseia na geração de sistemas de informações que permitem estabelecer relações matemáticas entre variáveis ambientais (atributos do terreno) e classes de solos e, assim, predizer a distribuição espacial das classes de solos com significativa semelhança aos mapas gerados por meios convencionais.

A grande vantagem do mapeamento digital de solos é a possibilidade de utilizar a modelagem para extrapolar os limites para áreas adjacentes ao mapa utilizado como referência, partindo da premissa que são áreas de mesma fisiografia, resultando em ganho de informações para áreas que não tinham informações alguma das classes de solos.

Alguns métodos de mapeamento digital de solos já estão sendo utilizados, como regressões logísticas múltiplas multinomiais (GIASSON et al., 2006; BAILEY, et al., 2003; CAMPLING et al., 2002), modelos logísticos com aplicação de componentes principais (TEN CATEN et al., 2009), redes neurais artificiais (SIRTOLI, 2008; BEHRENS et al., 2005), árvores de decisão (MORAN & BUI, 2002), *support vector machine* (SVM) (BEHRENS & SCHOLTEN, 2007), entre outros. Todos esses métodos preditivos estão condicionados a qualidade de entrada dos dados, ou seja, a resolução do MDE, as variáveis a serem utilizadas, a quantidade e a qualidade das amostras de treinamento.

## 2. HIPÓTESE E OBJETIVOS

### 2.1. HIPÓTESE

Pressupondo a relação existente entre classes de solo e as compartimentações do relevo, através de sistema de informações geográficas e métodos estatísticos, é possível estabelecer relações entre variáveis geomorfométricas (derivadas de um MDE) e classes de solos, que permitam a produção de mapas pedológicos com elevados custo-benefício e grau de semelhança a mapas realizados por métodos tradicionais.

### 2.2. OBJETIVOS

#### 2.2.1. OBJETIVO GERAL

Testar e comparar métodos de classificação em estágio único (regressões logísticas múltiplas multinomiais e Bayes) e classificação de decisão em árvore (*Classification and Regression Trees*, J48 e *Logistic Model Trees*) para produção de mapas pedológicos com legenda original e simplificada, com utilização de sistemas de informações geográficas e de variáveis geomorfométricas.

### 2.2.2. OBJETIVOS ESPECÍFICOS

- a) Obter variáveis geomorfométricas como declividade, curvatura, perfil de curvatura, curvatura planar, direção do escoamento, acúmulo do escoamento da água e índice de umidade topográfica (IUT) através do MDE oriundo do SRTM (*Shuttle Radar Topography Mission*);
- b) Estabelecer relações matemáticas entre o mapa de solo existente e as variáveis geomorfométricas utilizando de classificação em estágio único (CEU) e classificação de decisão em árvore (CDA).
- c) Produzir novos mapas de solos através de CEU e CDA a partir de variáveis geomorfométricas e mapa de solos existente, calcular e comparar as acurácias dos novos mapas digitais produzidos pelos diferentes métodos através de matrizes de confusão.

### 3. LOCALIZAÇÃO E CARACTERIZAÇÃO DA ÁREA DE ESTUDO

A área de estudo situa-se entre as latitudes  $28^{\circ}02'S$  e  $28^{\circ}33'S$  e as longitudes  $53^{\circ}39'O$  e  $54^{\circ}08'O$ . Integram a área de estudo os municípios de Ijuí, Bozano e parte de Coronel Barros, localizados na unidade geomorfológica do Planalto Meridional, formação Serra Geral, do Estado do Rio Grande do Sul, Brasil, com variação altimétrica de 210 a 570 metros, em relação ao nível do mar; a região é drenada pelo rio Ijuí e seus afluentes e compreende uma área de aproximadamente  $1.018\text{km}^2$ . A Figura 1 apresenta a localização da área de estudo.

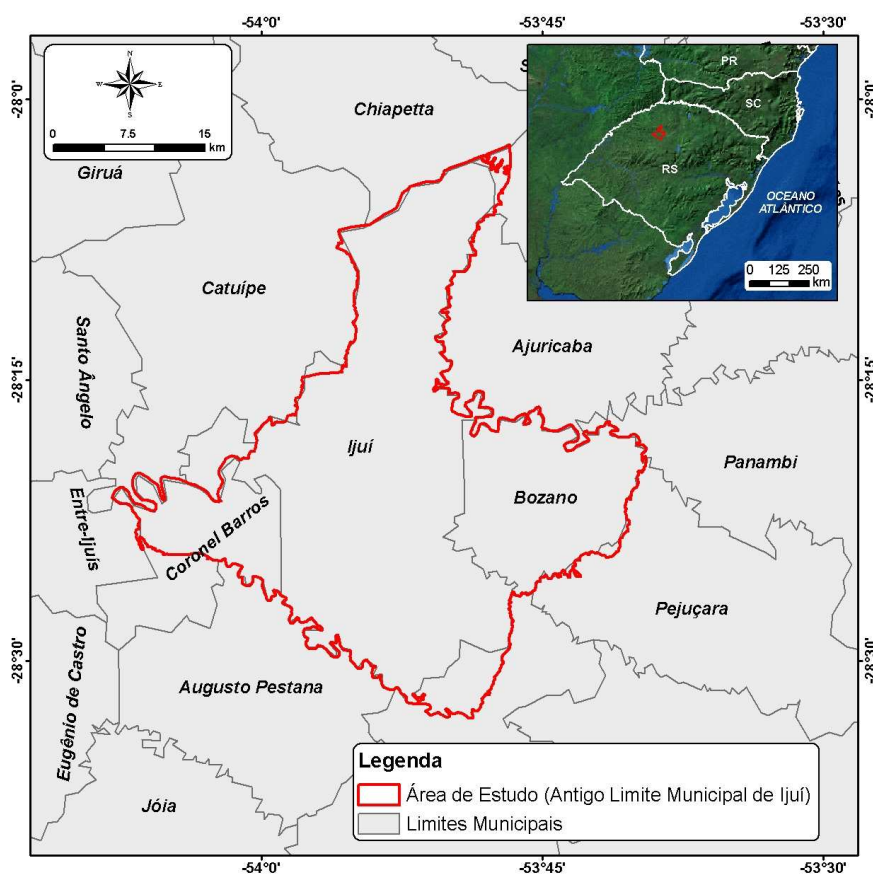


Figura 01 - Localização da área de estudo

Conforme MORENO (1961), o clima do Rio Grande do Sul, segundo o sistema internacional de classificação climática de Köppen, enquadra-se na zona

temperado úmido, ou “Cf”. No Estado, o tipo “Cf” se subdivide em “Cfa” e “Cfb”. O primeiro se caracteriza por apresentar chuvas durante todos os meses do ano e possuir a temperatura do mês mais quente superior a 22°C, e a do mês mais frio superior a 3°; o segundo apresenta chuvas durante todos os meses do ano, tendo a temperatura do mês mais quente inferior a 22°C e a do mês mais frio superior a 3°C. Assim, o Estado fica dividido em duas áreas, segundo a classificação de Köppen (“Cfa” e “Cfb”), sendo que a zona “Cfb” se restringe ao Planalto Meridional e ao escudo Sul-Riograndense, enquanto que as demais áreas pertencem à classe “Cfa”.

De acordo com MÜLLER FILHO (1970), a unidade do Planalto Meridional é caracterizada por uma superfície ondulada, formada por rochas vulcânicas básicas que variam para rochas vulcânicas ácidas no topo. As maiores altitudes estão localizadas a leste, ao passo que a oeste a escarpa erosiva se torna, gradativamente, mais rebaixada, gerando uma feição monoclinal. A Figura 2 apresenta a elevação da área de estudo.

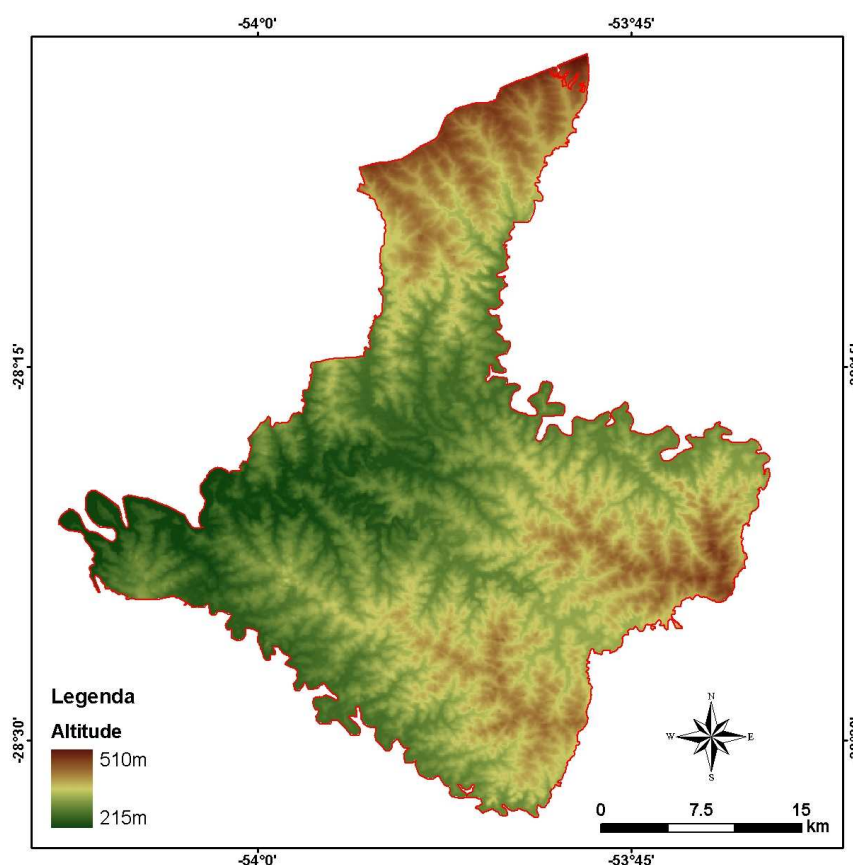


Figura 02 – Mapa hipsométrico da área de estudo

O mapeamento dos solos da área de estudo na escala 1:50.000 (ABRÃO & AZOLIN, 1970), identificou a ocorrência de sete classes de solo: a) Chernossolo Argilúvico Órtico (MTo), b) Chernossolo Ebânico Órtico (MEo), c) Latossolo Vermelho Distroférico Típico (LVdf), d) Latossolo Vermelho Distrófico Húmico (LVdh), e) Latossolo Vermelho Distrófico Típico (LVdt), f) Neossolo Flúvico (RY), g) Neossolo Regolítico Eutrófico (RRe). A Figura 3 apresenta a distribuição das classes de solos da área de estudo.

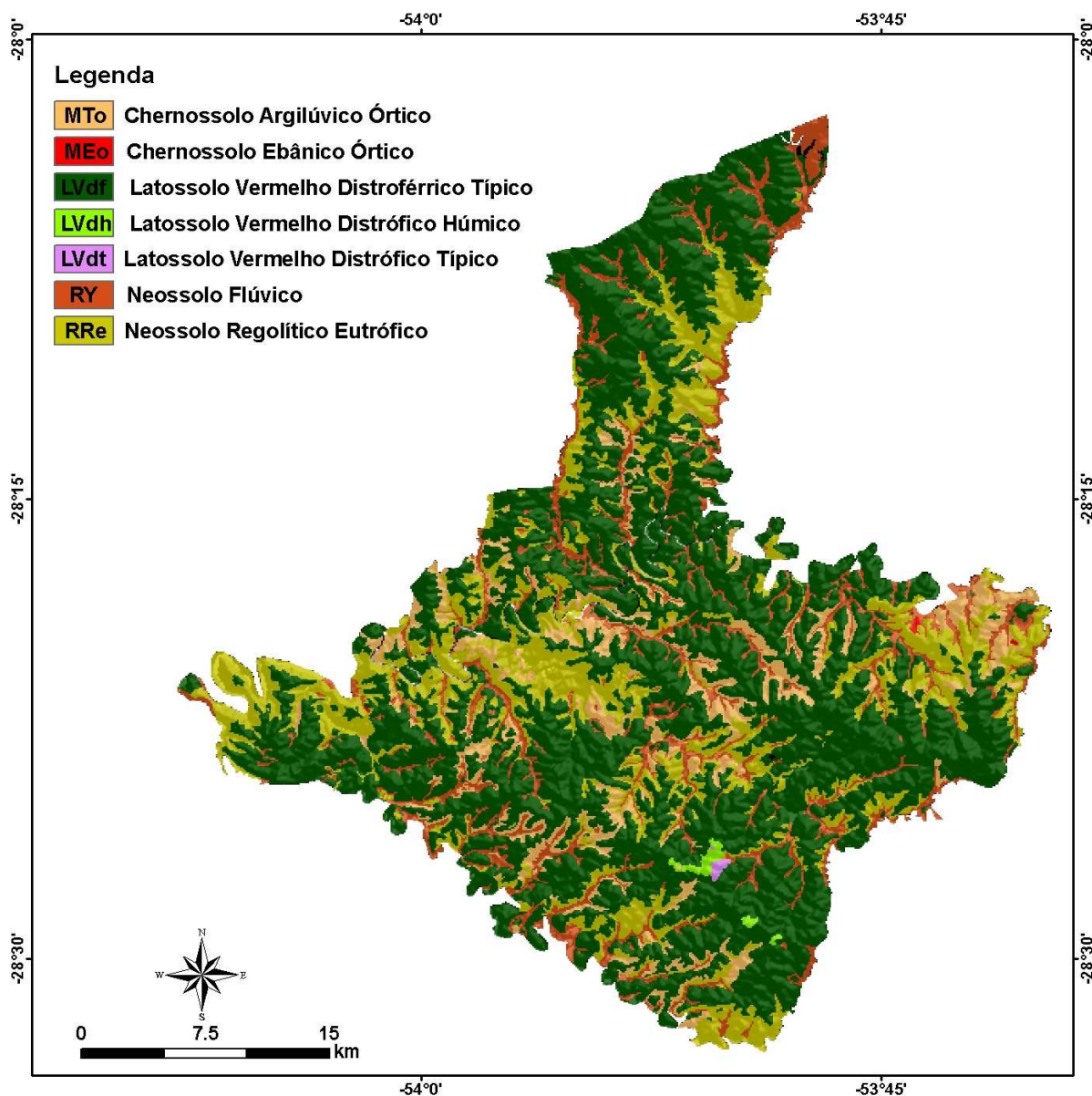


Figura 03 - Mapa de solos da área de estudo

Fonte: ABRÃO & AZOLIN (1970) atualizado e digitalizado por BORTOLON (2008)

Estão inseridos na Tabela 1 os dados correspondentes aos valores absolutos e relativos das áreas de cada classe de solo, mapeadas na escala 1:50.000, existente no local de estudo. Conforme a tabela, a classe de solo predominante é a LVdf com 61,19% da área total, ao passo que classes como MEO, LVdh e LVdt correspondem a menos de 1% da área total juntas.

Tabela 01 - Área ocupada pelas classes de solo, segundo mapeamento de ABRÃO & AZOLIN (1970) atualizado e digitalizado por BORTOLON (2008)

Classe	Nome da Unidade	Área (km <sup>2</sup> )	Área (%)
MTo	Chernossolo Argilúvico Órtico	80,97	7,96
MEo	Chernossolo Ebânico Órtico	0,44	0,04
LVdf	Latossolo Vermelho Distroférico Típico	622,84	61,19
LVdh	Latossolo Vermelho Distrófico Húmico	2,76	0,27
LVdt	Latossolo Vermelho Distrófico Típico	0,82	0,08
RY	Neossolo Flúvico	111,93	11,00
RRe	Neossolo Regolítico Eutrófico	198,04	19,46
	Total	1017,80	100,00

## 4. REVISÃO BIBLIOGRÁFICA

### 4.1. SOLOS

#### 4.1.1. DEFINIÇÕES DE SOLOS

O termo solo origina-se do Latim *solum* (suporte, superfície, base); sua definição tem diferentes significados segundo o conhecimento adquirido a seu respeito e o propósito para o qual se estuda. Para um agrônomo, ou para um agricultor, o solo é o meio necessário para o desenvolvimento dos vegetais, ao passo que para o engenheiro é o material que serve a base ou fundação de obras civis; para um geólogo o solo é visto como o produto do intemperismo das rochas na superfície enquanto para um arqueólogo é o material fundamental para suas pesquisas, por servir de registro de civilizações passadas; para o hidrólogo, o solo é o meio poroso que abriga reservatórios de águas subterrâneas (MILLER, 1993).

Entre diversos conceitos de solos, destacam-se (i) o conceito de solo como meio para desenvolvimento das plantas (HILGARD, 1914); (ii) o solo como regolito (RAMANN, 1919) e (iii) o solo como um corpo natural organizado (DOKOUCHEV, 1883).

No primeiro conceito, o objetivo final do solo é a produção de alimentos e fibras, conceito desenvolvido nas instituições de ensino e pesquisa agrícola. O segundo compreende o solo como regolito constituído por minerais originados do intemperismo da rocha, matéria orgânica originada por processos biológicos, minerais e matéria orgânica em estado coloidal ou em solução d'água e ar.

O terceiro conceito, o de solo como um corpo natural organizado, surgiu no final do século XIX, através do cientista Dokouchaev, quando o solo deixou de ser considerado simplesmente como um corpo inerte que reflete unicamente a composição da rocha que lhe deu origem (rocha matriz) para ser identificado como um material que evolui no tempo, sob ação das interações entre a litosfera, atmosfera, hidrosfera e a biosfera. JENNY (1941) apresenta o seguinte modelo:



$$S = f(cl, p, o, r, t) \quad (01)$$

O modelo expressa o solo ( $S$ ) como uma função das interações entre os fatores ambientais de formação clima ( $cl$ ), material parental ( $p$ ), organismos vivos ( $o$ ) e relevo ( $r$ ) atuando ao longo do tempo ( $t$ ).

Conforme TEIXEIRA et. al. (2000), existe uma definição simples, a qual se adapta aos propósitos da Ciência da Terra, que considera solo como produto do intemperismo, do remanejo e da organização das camadas superiores da crosta terrestre, sob ação da atmosfera, da hidrosfera, da biosfera e das trocas de energia envolvidas.

#### 4.1.2. CLASSIFICAÇÃO DE SOLOS

O mapeamento dos solos, pressupõe que exista uma classificação dos solos. A classificação pode ser feita segundo diferentes critérios variáveis de país para país, devido a diferentes ênfases dadas aos fatores morfogenéticos envolvidos na formação do solo.

As classificações mais conhecidas são a francesa, a classificação adotada pela FAO (*Food and Agricultural Organization*) na sistematização da carta mundial de solos e a classificação mais difundida a “*Soil Taxonomy*” (Taxonomia dos Solos). Publicada em 1938, a “*Soil Taxonomy*” foi desenvolvida nos Estados Unidos da América e na edição de 1999 (USDA, 1999) diferencia doze ordens de solos, subdividindo em sub-ordens, grandes grupos, grupos, famílias e séries.

No Brasil, desde a década de 1960, o serviço pedológico da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) realiza levantamentos de solos sistemáticos do território nacional. A partir desses levantamentos, em 1999, a EMBRAPA consolida o Sistema Brasileiro de Classificação de Solos (SiBCS) (EMBRAPA, 1999). Atualizado em 2006 (EMBRAPA, 2006) o Sistema Brasileiro de Classificação de Solos é resultado de projeto nacional que envolveu a comunidade científica de Ciência do Solo. O desenvolvimento do sistema passou por diversas fases,

desde a definição de classes até um sistema mais elaborado segundo princípios lógicos de taxonomia.

Do modo como está estruturado, similar à *Soil Taxonomy*, o sistema permite a classificação de todos os tipos de solos do território nacional em seis níveis categóricos diferentes: Ordem, Subordem, Grande Grupo, Subgrupo, Família e Série, correspondendo cada nível a um grau de detalhamento, em ordem crescente.

## 4.2. RELAÇÃO SOLOS – GEOMORFOLOGIA

CHRISTOFOLETTI (1980) define a Geomorfologia como a ciência que estuda a gênese e a evolução das formas do relevo sobre da Terra, onde essas formas são resultantes dos processos atuais ou pretéritos ocorridos nos litotipos existentes. Segundo o mesmo, os processos ou fatores determinantes que definem essa evolução podem ser exógenos, como o clima e vegetação, bem como podem ser endógenos, como a tectônica e a geologia. Por ser um atributo geométrico da superfície terrestre, o relevo oferece indicações sobre diversas variáveis ambientais importantes na delimitação do terreno e no mapeamento de solo em diferentes escalas.

De acordo com BRIGGS & SHISHIRA (1985), a principal aplicação do relevo, para o mapeamento de outras variáveis ambientais, é a abordagem quantitativa e a divisão do terreno em unidades de características qualitativamente homogêneas.

Considerando a relação solos-relevo, LEPSH & BUOL (1974) estabeleceram o conceito de Topossequência, que ocorre quando a diferenciação de um conjunto de solos se deve, principalmente, à sua posição no relevo, como apresenta a Figura 4.

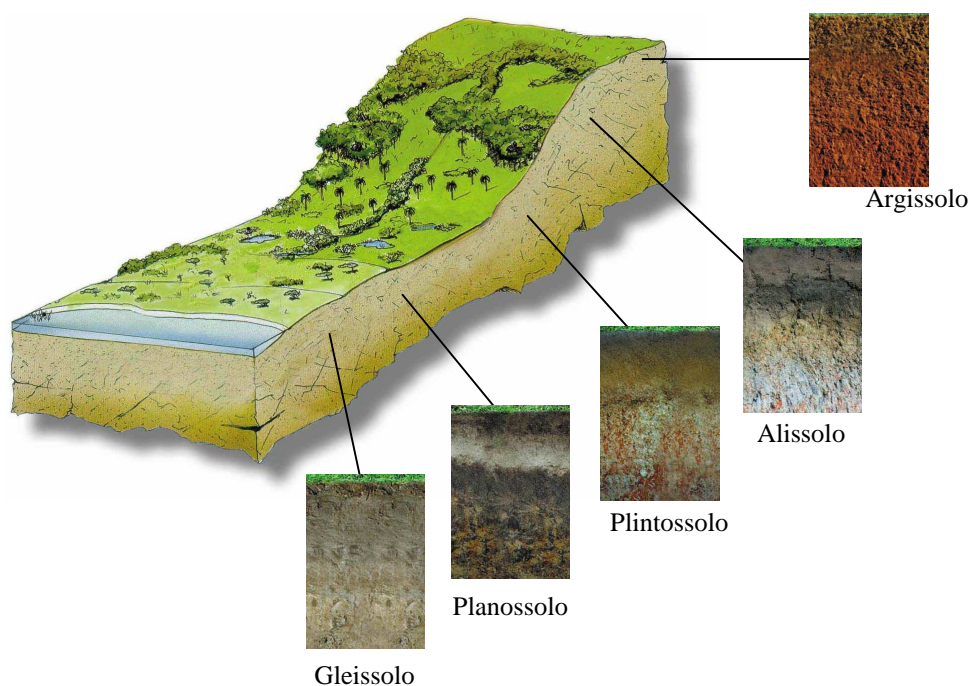


Figura 04 – Exemplo de Topossequência na Depressão Central do Rio Grande do Sul  
 Fonte: Adaptado de STRECK (2002)

Diferentes tipos de solos em diferentes posições do relevo remontam à idéia de que os processos de formação também são diferentes para cada segmento de uma vertente. Frequentemente a porção superior de uma encosta se caracteriza pela erosão, a porção intermediária se caracteriza pelo transporte de sedimentos e o sopé pela deposição do material sedimentar; dessa maneira, as partes mais altas são mais drenadas, em relação as partes mais baixas, e esta diferença de drenagem é responsável pelas mudanças graduais nas características dos solos (GERRARD, 1981).

DALRYMPLE et al. (1968), apresentou um modelo hipotético dos processos geomorfológicos dominantes em nove unidades de superfície em uma vertente (Figura 5). Segundo o mesmo, na primeira unidade (Interflúvio), é onde ocorrem os processos pedogenéticos associados com o movimento vertical subsuperficial da água; na segunda unidade (declive de infiltração), ocorre o processo de Eluviação química e mecânica pelo movimento lateral subsuperficial da água; na terceira unidade (declive convexo de rastejamento), há o rastejamento do solo com a formação de terraços; na quarta unidade (frente de queda), há o movimento de queda, escorregamento e o desgaste físico e químico do solo; na quinta unidade (meia encosta de transporte), ocorre, através do movimento de massa, transporte de material e

formação de terraços com ação superficial e subsuperficial da água; na sexta unidade (sopé coluvial), ocorre redeposição de matéria por movimentos de massa e lavagem superficial, formação de leques, transporte de material, rastejamento e ação superficial e subsuperficial da água; na unidade sete (saliente coluvial), há deposição aluvial e processos resultantes da ação subsuperficial da água; na oitava unidade (canal barreira), ocorrem os processos de erosão, afundamento e queda dos solos; e na nona unidade (canal leito), ocorre o transporte do material para o vale pela ação da superfície da água, gradação e erosão periódicas.

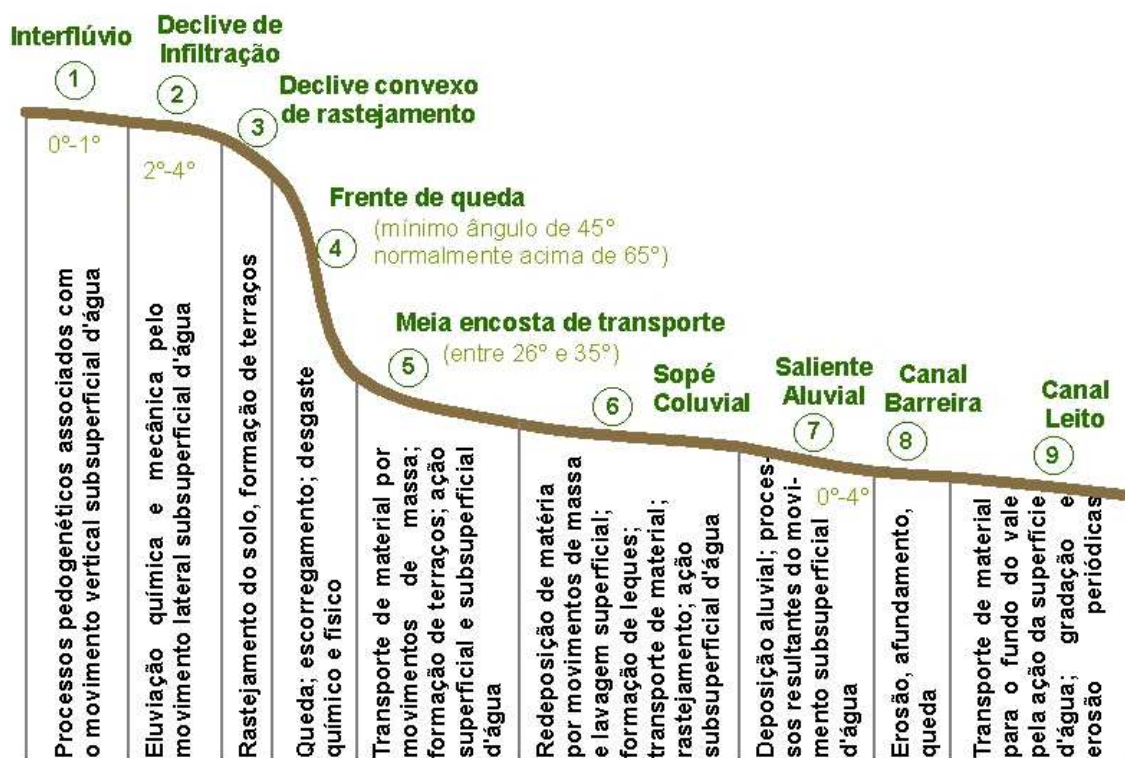


Figura 05 - Modelo dos processos geomorfológicos em uma vertente  
 Fonte: Adaptado de DALRYMPLE et al. (1968).

### 4.3. MAPEAMENTO DE SOLOS

#### 4.3.1. MAPEAMENTO TRADICIONAL DE SOLOS

O mapeamento de solos envolve a determinação da localização e distribuição dos diferentes solos que ocorrem numa região, coleta de informações sobre sua localização, natureza, propriedades, potencial e usos. Os métodos tradicionais baseiam-se nas relações entre solos e fisiografias (IPPOLITI-RAMIRO et. al., 2005). Tais métodos utilizam observação estereoscópica de pares de aerofotos, e a partir de então, é realizada interpretação visual das feições do relevo e, após, é feito levantamento de campo onde os solos são identificados e mapeados conforme a escala pretendida. Os solos são classificados segundo o Sistema Brasileiro de Classificação de Solos, considerando diversas características diagnósticas, tais como abundância e qualidade de material orgânico nos horizontes superficiais, espessura e número de horizontes, cor, material de origem, entre outros (PINTO et al., 1999).

#### 4.3.2. MAPEAMENTO DIGITAL DE SOLOS

O mapeamento digital de solos se desenvolveu através do advento de uma série de tecnologias que possibilitam sua execução, como SIGs, Sensoriamento Remoto, GPS (*Global Positioning System*), entre outros (MCBRATHNEY et.al., 2003). O avanço e popularização das geotecnologias tornaram os processos computacionais complexos mais rápidos, como, por exemplo, as operações algébricas em um SIG. Juntamente e através do avanço das geotecnologias, o campo da cartografia, em geral, também teve grande desenvolvimento que permitiu o advento e crescimento da pedometria como Ciência do Solo (COOK et al., 1996).

Segundo MCBRATNEY (1996), a Pedometria estabelece relações matemáticas entre classes e propriedades de solos e as variáveis preditoras (variáveis

geomorfométricas) para o estudo da distribuição e gênese de solos. A Figura 6 representa esquematicamente a construção da pedometria em relação a outros ramos da ciência..

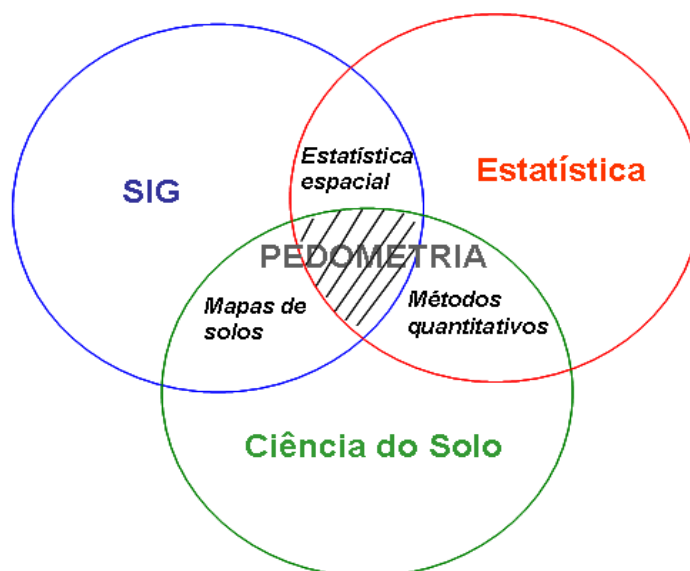


Figura 06 – Pedometria em relação a outros ramos da ciência  
Fonte: Adaptado de HENGL (2003)

Uma definição formal de mapeamento digital de solos foi apresentada para a comunidade científica em 2004, durante o “*I Global Workshop on Digital Soil Mapping*” realizado em Montpellier na França. Os artigos defendidos no evento foram publicados no livro “*Digital Soil Mapping – An introductory perspective*” na série *Developments in Soil Science*, volume 31 da Elsevier (MCBRATNEY et al., 2007). Conforme MCBRATNEY et al., (2007), mapeamento digital de solos é definido como a criação e manipulação de sistemas espaciais de informação de solos, através do uso de modelos numéricos para a inferência das variações espaciais e temporais dos tipos de solos e de suas propriedades, a partir de observações e conhecimento dos solos e de variáveis ambientais correlacionadas.

Como exemplo de aplicação do mapeamento digital de solos, GIASSON et al. (2006) realizaram estudo na região sul do Brasil; os resultados chegaram a apresentar até 71% de coincidência (acurácia global) nos mapas produzidos com o mapa original, utilizando-se do método de regressões logísticas múltiplas multinomiais. Também utilizando-se do modelo logístico TEN CATEN et al. (2009)

realizaram mapeamento digital de solos com aplicação de componentes principais. Os resultados mostraram a aplicabilidade do uso de análise de componentes principais para redução das variáveis e análise da correlação entre as mesmas e ainda consideraram satisfatórios os resultados do mapeamento.

Utilizando-se de métodos hierárquicos de classificação, BEHRENS & SCHOLTEN (2007), testaram os métodos de árvores de decisão e *support vector machine* (SVM) para o mapeamento digital de solos. Segundo os autores, os melhores resultados ocorreram nos métodos de árvores de decisão e concluíram que o método SVM não é adequado para mapeamento digital de solos pois retorna baixas acurácias e o processamento dos dados é lento. O estudo realizado por SARMENTO et al., (2009), comparando os métodos de decisão em árvore e redes neurais artificiais (RNA), apresentou melhores acurácias com o método de decisão em árvore e de mais fácil implementação em SIG do que o método de redes neurais artificiais.

A tese de SIRTOLI (2008), além de utilizar apenas as variáveis derivadas do MDE, utilizou variáveis espectrais de imagens de satélite e mapas de geologia. O método de classificação utilizado em seu estudo foi o de Redes Neurais Artificiais (RNAs) para discriminação das classes de solo e seus resultados indicaram que a metodologia é útil para a melhoria da qualidade dos levantamentos preliminares de solos.

#### 4.4. VARIÁVEIS DO TERRENO

##### 4.4.1. MODELO DIGITAL DE ELEVAÇÃO - SRTM

Liderada pela NASA (*National Aeronautics and Space Administration*), durante 11 dias, em Fevereiro de 2000, a missão espacial SRTM (*Shuttle Radar Topography Mission*) obteve dados sobre mais de 80% da superfície terrestre, entre as latitudes 60°N e 57°S, através de radar de abertura sintética interferométrico (InSAR – *Interferometric Synthetic Aperture Radar*) que adquiriu

dados nas bandas C e X a bordo do ônibus espacial Endeavour (RABUS et al., 2003). O sistema contava com duas antenas de recepção separadas por um mastro de 60m, o que possibilitou a aquisição dos dados em uma mesma órbita, garantindo a melhor qualidade dos mesmos (RABUS et al., 2003).

Os MDEs do SRTM, disponibilizados para a América do Sul, possuem resolução espacial de 3 arco-segundos (cerca de 90m) e resolução vertical nominal de 6m (RABUS et al., 2003). Segundo VALERIANO (2008), toda aquisição dos dados SRTM foi planejada a suprir mapeamentos em escalas entre 1:100.000 e 1:250.000.

#### 4.4.2. VARIÁVEIS GEOMORFOMÉTRICAS

Conforme MOORE et. al., (1993), a obtenção de variáveis geomorfométricas pode ser diferenciada entre variáveis primárias e secundárias; as variáveis primárias seriam aquelas derivadas diretamente do MDE; as secundárias são as variáveis que envolvem a combinação de variáveis primárias (geralmente índices geomorfométricos). Entre as principais variáveis primárias estão a declividade, orientação das vertentes, plano e perfil de curvatura.

Para GALLANT & WILSON (2000), a declividade é considerada um dos atributos topográficos mais importantes que controlam os processos pedológicos, pois influencia diretamente na velocidade do fluxo superficial e sub-superficial de água, logo, influencia no teor da água do solo e no potencial dos processos erosivos e deposicionais. BURROUGH (1986) define declividade como sendo um plano tangente a superfície, expresso como a mudança de elevação sobre certa distância, representada no formato digital na resolução de uma grade regular.

Segundo CHAGAS (2006), a orientação das vertentes, é um atributo pouco explorado no estudo das relações solo paisagem no Brasil, talvez devido à dificuldade de sua obtenção pelo método tradicional da fotointerpretação. Para MOORE et.al. (1993), a orientação das vertentes tem relação direta com a evapotranspiração, insolação, teor de água no solo e, conseqüentemente, com os atributos do solo.



Para SIRTOLI (2008), o plano e o perfil de curvatura representam as formas do relevo, sendo importantes atributos na distinção de unidades geomorfológicas. O primeiro é referente à forma da vertente, em plano, podendo ser convergente, divergente ou planar; o segundo se refere à forma da vertente, analisada em perfil, podendo ser convexa, côncava ou retilínea.

O perfil de curvatura tem importância na aceleração ou desaceleração do fluxo de água em uma encosta, logo, tem influência direta na erosão do solo. Quando a declividade diminui, na direção da orientação da vertente, o terreno é côncavo; a convexidade se dá quando a declividade aumenta na direção da orientação da vertente; quando a declividade não se altera no perfil, o terreno é caracterizado como reto.

O plano de curvatura, ou curvatura planar, pode ser convergente, divergente ou reto e tem influência na acumulação da umidade e do fluxo da água superficial e sub-superficial do terreno. Caracterizam-se como terrenos convergentes aqueles que as direções de maior declividade, em diferentes pontos do terreno, tendem a se encontrar; os terrenos, em que as direções de maior declividade, em diferentes pontos, tendem a se separar, são considerados divergentes.

Para o cálculo das variáveis geomorfológicas, ZEVENBERGEN & THORNE (1987) apresentaram uma submatriz genérica, representada pela Figura 7.

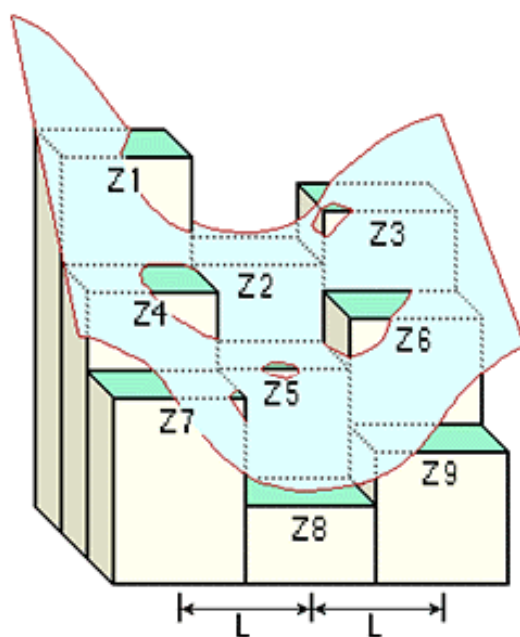


Figura 07 - Submatriz de elevação utilizada para calcular as variáveis geomorfológicas  
Fonte: Adaptado de ZEVENBERGEN & THORNE (1987)

A submatriz representa uma superfície côncava onde cada célula possui uma unidade de medida de elevação (componente Z), de um MDE com determinada resolução espacial. Os valores L, representam a distância entre duas cotas de elevação. Trata-se de uma janela flutuante (móvel), de dimensão 3x3, que utiliza 9 pixels para calcular o valor do pixel central (Z5). Assim que é calculado o valor de um pixel, a janela é centralizada no pixel seguinte até que seja calculado todo o MDE.

A partir da submatriz, são calculados vários parâmetros intermediários (de A a H) que serão utilizados para o cálculo das variáveis geomorfológicas:

$$A = [(Z1 + Z3 + Z7 + Z9)/4 - (Z2 + Z4 + Z6 + Z8)/2 + Z5]/L4 \quad (02)$$

$$B = [(Z1 + Z3 - Z7 - Z9)/4 - (Z2 - Z8)/2]/L3 \quad (03)$$

$$C = [(-Z1 + Z3 - Z7 + Z9)/4 + (Z4 - Z6)]/2L3 \quad (04)$$

$$D = [(Z4 + Z6)/2 - Z5]/L2 \quad (05)$$

$$E = [(Z2 + Z8)/2 - Z5]/L2 \quad (06)$$

$$F = (-Z1 + Z3 + Z7 - Z9)/4L2 \quad (07)$$

$$G = (-Z4 + Z6)/2L \quad (08)$$

$$H = (Z2 - Z8)/2L \quad (09)$$

O píxel central da submatriz é calculado a partir de seus oito vizinhos, através dos parâmetros intermediários listados acima e da janela móvel, originando as variáveis geomorfológicas através das seguintes equações:

$$\text{Declividade (graus)} = (\sqrt{([dz/dx]^2 + [dz/dy]^2)}) * 57,29578 \quad (10)$$

Sendo:

$$[dz/dx] = ((Z3 + 2Z6 + Z9) - (Z1 + 2Z4 + Z7)) / 8 * (\text{tamanho\_da\_célula}) \quad (11)$$

$$[dz/dy] = ((Z7 + 2Z8 + Z9) - (Z1 + 2Z2 + Z3)) / 8 * (\text{tamanho\_da\_célula}) \quad (12)$$

(BURROUGH, 1998)

$$\text{Orientação das Vertentes} = 57,29578 * \arctan 2([dz/dy], -[dz/dx]) \quad (13)$$

Sendo:

$$[dz/dx] = ((Z3 + 2Z6 + Z9) - (Z1 + 2Z4 + Z7)) / 8 \quad (14)$$

$$[dz/dy] = ((Z7 + 2Z8 + Z9) - (Z1 + 2Z2 + Z3)) / 8 \quad (15)$$

(BURROUGH, 1998)

$$\text{Curvatura Planar} = 2 * (DH^2 + EG^2 - FGH) / (G^2 + H^2) \quad (16)$$

(ZEVENBERGEN & THORNE, 1987)

$$\text{Perfil de Curvatura} = -2 * (DG^2 + EH^2 + FGH) / (G^2 + H^2) \quad (17)$$

(ZEVENBERGEN & THORNE, 1987)

O Índice de Umidade Topográfica (IUT) (KIRKBY, 1975) é uma variável secundária utilizada para quantificar o efeito da topografia em processos hidrológicos e modelar a distribuição espacial das zonas de saturação superficial e conteúdo de água nas paisagens. O IUT é definido como uma função de declividade e

da área de contribuição por unidade de largura ortogonal à direção do fluxo (KIRKBY, 1975). Sua formulação é dada por:

$$IUT = \ln(\mathbf{A_s} / \tan \mathbf{b}) \quad (18)$$

Onde:

$A_s$  = a área de contribuição multiplicada pelo tamanho do pixel (m<sup>2</sup>);

$\mathbf{b}$  = declividade da célula.

#### 4.4.3. VARIÁVEIS HIDROGRÁFICAS DERIVADAS DO MDE

Conforme VALERIANO & MORAES (2001), entre as características do relevo mais importantes a indicar diferenças entre solos e rochas está a rede de drenagem. Esta advém da interação dos fatores associados ao relevo, geologia, clima, vegetação, textura e permeabilidade do solo.

Modelos hidrológicos tipicamente empregam equações matemáticas para representar os processos hidrológicos como a transformação da chuva em vazão e o escoamento da água pela superfície do terreno. Apesar dos modelos matemáticos gerarem dificuldade para a representação de processos físicos e discretização de processos contínuos, TUCCI (1998) considera que os modelos matemáticos têm a vantagem de permitir a geração de resultados para diferentes situações com alta velocidade de resposta.

Dado a grande influência da hidrografia na pedogênese e na escultura do relevo, na identificação de áreas de acúmulo sujeitas a inundação bem como de áreas de intenso escoamento superficial, as variáveis hidrográficas, derivadas do MDE, são utilizadas, juntamente, com as variáveis geomorfológicas, em modelos estatísticos para o mapeamento digital de solos. As variáveis direção do fluxo de escoamento, acumulação do fluxo e comprimento do fluxo ganham destaque devido suas importância na gênese dos solos.

A direção do fluxo de escoamento consiste em considerar uma única direção do fluxo para cada pixel do MDE, sendo essa atribuída para um de seus oito vizinhos, considerando uma janela móvel de dimensão 3x3. Um dos métodos mais comumente utilizados é o D8 (*deterministic eight neighbours* ou determinístico oito vizinhos) proposto por JENSON & DOMINGUE (1988). A determinação da direção do fluxo é realizada a partir da escolha da direção que proporcione a maior declividade, calculada como sendo a diferença de elevação entre o pixel vizinho e o pixel central, dividida pela distância entre eles, como apresenta a Figura 8.

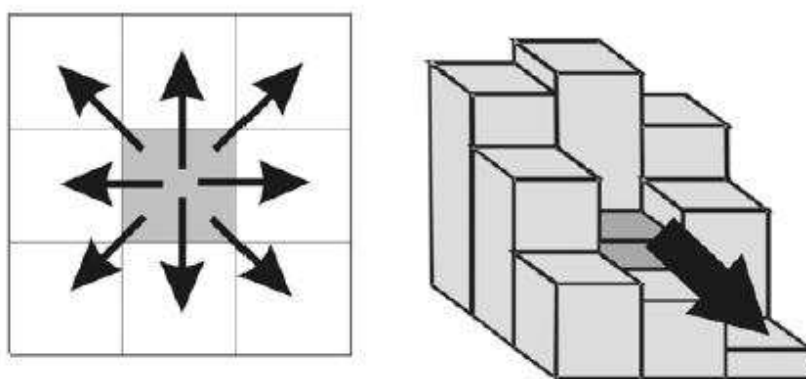


Figura 08 - Determinação da direção do fluxo a partir de seus oito vizinhos

A acumulação do fluxo representa as áreas de drenagem acumulada sobre uma superfície e pode ser determinada exclusivamente com base nas direções do fluxo. Conforme JENSON & DOMINGUE (1988), trata-se de um arquivo matricial onde cada pixel tem como atributo o valor correspondente ao somatório das áreas superficiais de todos os pixels cujo escoamento contribui para o pixel em questão.

Do mesmo modo que a acumulação do fluxo, a variável comprimento do fluxo pode ser gerada exclusivamente com base nas direções do fluxo. O comprimento do fluxo representa os comprimentos acumulados no percurso da drenagem, desde seu pixel de origem até um ponto de estagnação da água onde o fluxo se encerra. Logo, o valor atribuído a cada pixel indica a distância do mesmo ao exutório da bacia, ou outro ponto referencial, ao longo da drenagem.

## 4.5. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU)

### 4.5.1. CLASSIFICADOR DE BAYES

É um classificador amplamente utilizado para classificação de imagens de Sensoriamento Remoto e que será testado para estimação de mapas de solos. A classificação de Bayes parte do pressuposto de que as funções de probabilidade dentro de cada classe, denotadas por  $P(x/\omega_i)$ , são conhecidas e que há dados sobre a probabilidade *a priori*, denotada por  $P(\omega_i)$ , de cada classe. Assim, é possível estabelecer a probabilidade *a posteriori*, denotada por  $P(\omega_i/x)$ , de um determinado pixel pertencer a classe  $\omega_i$ . Conforme FUKUNAGA (1990), a regra de classificação de Bayes se baseia simplesmente nas probabilidades de  $x$  pertencer a cada uma das classes  $\omega_i$  e pode ser descrita da seguinte maneira:

$$X \in \omega_i, \text{ se } P(\omega_i/x) > P(\omega_j/x) \quad (19)$$

A probabilidade *a posteriori* pode ser obtida utilizando o teorema de Bayes:

$$P(\omega_i/x) = P(x/\omega_i) \cdot P(\omega_i) / P(x) \quad (20)$$

Sendo  $P(x)$  a função mistura de densidade, ou função densidade ponderada entre as duas classes.

A Figura 9 exemplifica o caso unidimensional da regra de Bayes, onde o erro bayesiano é dado pelo menor limiar possível  $t$  (onde o erro é dado pela região  $A+B+C$ ); caso fosse tomado qualquer outro limiar,  $t'$  por exemplo, a região do erro sob as duas curvas é aumentada pela área  $D$ , sendo assim, maior que o erro de bayesiano. Devido a escolha do melhor limiar, conseqüentemente a minimização da probabilidade do erro, Bayes é considerado como classificador ótimo (GONZALEZ & WOODS, 1992).

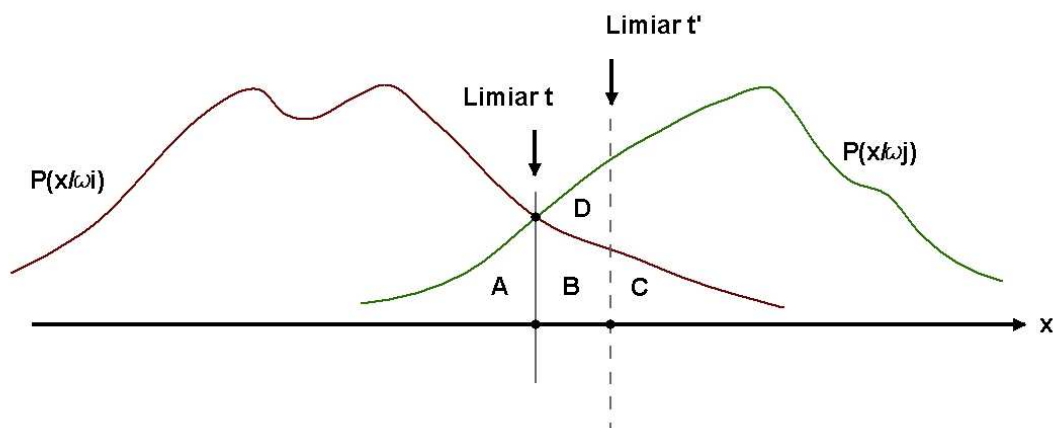


Figura 09 - Limiar bayesiano  $t$  e não bayesiano  $t'$   
 Fonte: Adaptado de THERRIEN (1989)

#### 4.5.2. REGRESSÕES LOGÍSTICAS MÚLTIPLAS MULTINOMIAIS (RLMM)

O modelo de regressão logística nos permite avaliar a probabilidade de ocorrência de uma variável resposta, considerando os valores de uma série de variáveis independentes que podem ser qualitativas ou quantitativas. Por ser utilizado mais que uma variável independente, o modelo é considerado múltiplo e, pelo fato de prever mais que duas classes, é considerado multinomial (ou politômico); logo, o modelo é denominado de regressão logística múltipla multinomial (RLMM) que é uma generalização do modelo logístico simples (AGRESTI, 1990).

O modelo é definido por  $J-1$  equações, ou seja, o número de equações é definido pela quantidade de classes menos um, de forma que a ocorrência de uma classe  $j \neq J$  é calculada em relação à classe referência  $J$ . Temos sua modelagem dada por:

$$\log\left(\frac{p_j}{p_J}\right) = x_i' \beta_j \quad j = 1, \dots, J-1 \quad (21)$$

sendo:

$$x_i' \beta_j = \sum_{k=0}^K \beta_{jk} x_{ik} \quad (22)$$

com  $x_{i0} = 1$  e  $x_i$  e  $\beta_j$  vetores (coluna) de dimensão  $(K+1)$ . Ou seja, é a multiplicação entre o vetor de coeficientes calculados pela função logística ( $\beta_j$ ) e as variáveis geomorfométricas ( $x_i$ ). As probabilidades das classes ( $p_j, \dots, p_J$ ) podem ser definidas por:

$$p_j = \frac{\exp(x_i' \beta_j)}{1 + \sum_{j=1}^{J-1} \exp(x_i' \beta_j)} \quad \text{e} \quad p_J = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(x_i' \beta_j)} \quad (23)$$

As RLMM têm sido utilizadas, em várias áreas das Ciências da Terra, para diferentes fins como obter probabilidade de classes de drenagem (CAMPLING et al. 2002), prospecção geológica (AGTERBERG, 1992), probabilidade de deslizamento de massa (OHLMACHER & DAVIS, 2003), como ferramenta para gerenciamento ambiental (ARBESÚ & FELICÍSIMO, 2002) e também para predição de classes de solo através da referência de campo.

Alguns exemplos do uso das RLMM para predição de classes de solos são os estudos de BAILEY, et al (2003), os quais utilizaram as RLMM para produção de mapas de solo na Inglaterra e observaram que o resultado, em relação com a referência de campo, ficaram entre 47% a 65%. No Brasil, os estudos de GIASSON et al. (2006), utilizaram as RLMM para produção de mapas de solo na região sul, os resultados chegaram a apresentar até 71% de coincidência com o mapa referência. FIGUEIREDO (2006) comparou o resultado das RLMM para predição de classes de solo em três localidades com fisiografia contrastante no Rio Grande do Sul, os resultados variaram de 31% a 68% de coincidência com os mapas originais. O autor concluiu que as RLMM têm potencial para produção de mapas pedológicos e que a eficiência do modelo depende da qualidade dos dados utilizados.

TEN CATEN et al. (2009), fizeram uso das RLMM com aplicação de componentes principais para estimativa de solo no Sul do Brasil, afim de diminuir a



dimensionalidade das variáveis, com perda irrelevante de informação. O resultado obtido ficou em 79,40% de coincidência com o mapa referência.

#### 4.6. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA)

Os CDA representam um conjunto de regras sobre uma sequência hierárquica estruturada com a finalidade de particionar os dados. Segundo BREIMAN et al. (1984), a mais importante função do CDA é sua capacidade de converter processos de decisão complexos em uma série de decisões simples.

Os CDA CART (*classification and regression trees*) (BREIMAN et al., 1984) e J48, também conhecido como C4.5 (QUINLAN, 1993) são formados por n-ésimas ramificações de uma árvore, que representam regras de classificação com base na composição de informações das distribuições isoladas de cada variável; observando os valores dessa variável, divide-se a região de valores em duas subregiões, através de uma pergunta que é respondida em termos de “sim” e “não”; no segundo passo, uma outra variável será analisada, em cada uma das subregiões, e dividida em mais duas partes através do mesmo critério de resposta “sim” e “não”. A árvore termina quando os nós (terminais) possuem apenas uma classe. SAFAVIAN & LANDGREBE (1991) definem a estrutura geral de um CDA de acordo com a Figura 10.

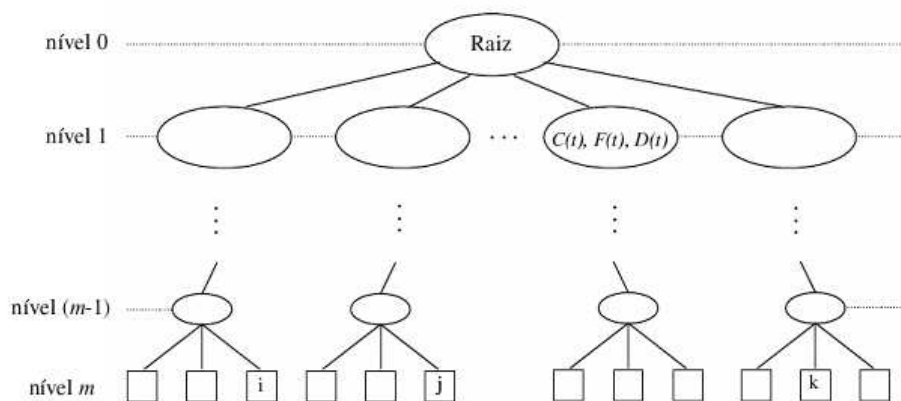


Figura 10 - Estrutura geral de um CDA.

Fonte: Adaptado de SAFAVIAN & LANDGREBE (1991)

No nível 0 está localizado o nó raiz, o qual contém todas as variáveis e todas as classes. Nos nós seguintes, estão representadas as derivações do nó raiz, definido em função de três componentes:

- 1) Uma regra de classificação (classificador);
- 2) Conjunto de variáveis;
- 3) Conjunto de classes.

Cada nó está representado por uma terna (C(t), F(t) e D(t), onde C(t) representa as classes do nó t, F(t) as variáveis do nó t, e D(t) a regra de classificação do nó t). O nível m representa os nós terminais, os quais devem conter apenas uma classe.

O CDA LMT (*Logistic Model Trees*) (LANDWEHR, et al., 2005) tem estrutura parecida com a do CART e J48, porém o método LMT possui, em seu nós terminais, ao invés de classes únicas, vetores de variáveis ( $x_i$ ) e de coeficientes ( $\beta_j$ ) de todas as classes. Através desses valores, podemos calcular a função do LMT:

$$F(x) = x_i \cdot \beta_j \quad j = 1, \dots, J \quad (24)$$

onde:

$$x_i \cdot \beta_j = \sum_{k=0}^K \beta_{jk} x_{ik} \quad (25)$$

com  $x_{i0} = 1$  e  $x_i$  e  $\beta_j$  vetores (coluna) de dimensão (K+1), similar ao modelo logístico em estágio único. As probabilidades das classes, no LMT, são calculadas através da seguinte formulação:

$$p_j = \frac{\exp F_j(x)}{\exp F_1(x) + \dots + \exp F_J(x)} \quad \text{e} \quad p_J = \frac{\exp F_J(x)}{\exp F_1(x) + \dots + \exp F_J(x)} \quad j = 1, \dots, J-1 \quad (26)$$

As probabilidades resultantes são para cada nó terminal, ou seja, as probabilidades serão válidas somente para aqueles pixels que respeitarem a regra de

decisão da árvore até o nó em questão. A Figura 11 ilustra a estrutura dos três métodos de CDA's.

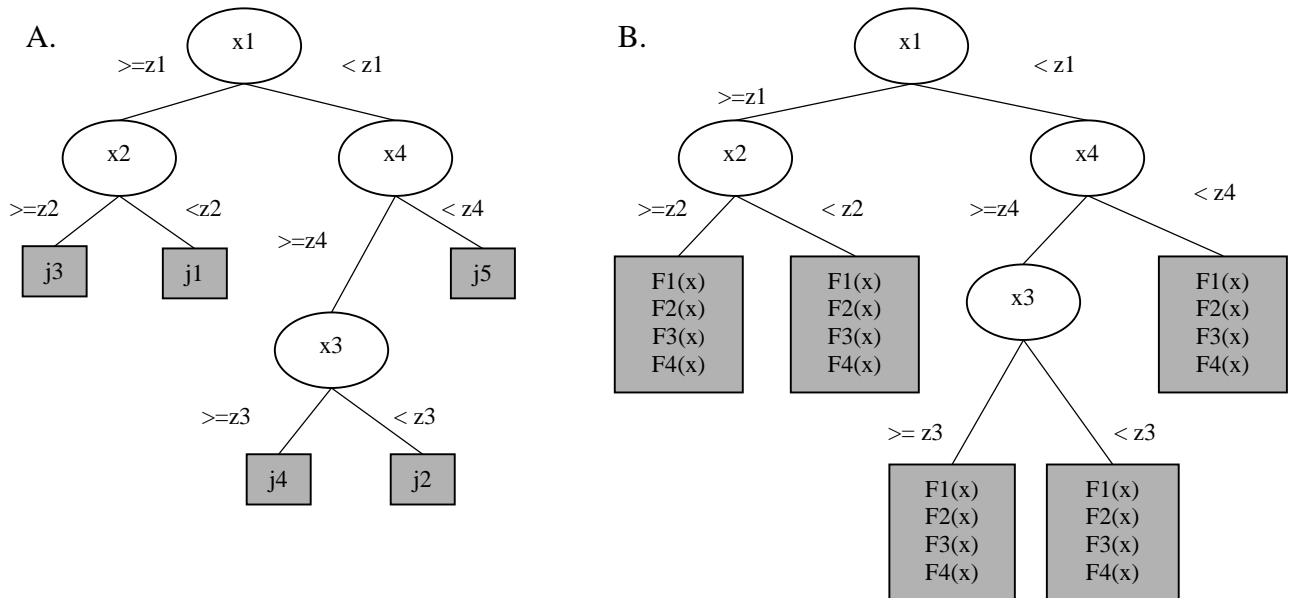


Figura 11 - Estrutura dos CDA CART e J48 (A) e LMT (B)

## 5. METODOLOGIA

A partir da criação de uma base de dados, variáveis geomorfométricas foram derivadas do MDE do SRTM e foram estatisticamente padronizadas. Com as novas variáveis e o mapa original de solos, foram geradas amostras aleatórias para o processo de classificação. Foram testados cinco métodos de classificação, sendo que o método de Bayes foi executado em ambiente ArcGis 9.2 (ESRI - *Environmental Systems Research Institute*, 2006) e os demais (RLMM, CART, J48 e LMT) foram executados no *software* estatístico WEKA 3.5.8 (*Waikato Environment for Knowledge Analysis*) (HALL, et. al., 2009); após a obtenção dos dados no WEKA 3.5.8, os mapas foram implementados em ambiente ArcGis 9.2 através de ferramentas de álgebra de mapas. Os procedimentos metodológicos podem ser esquematizados, para um melhor entendimento, conforme a Figura 12.

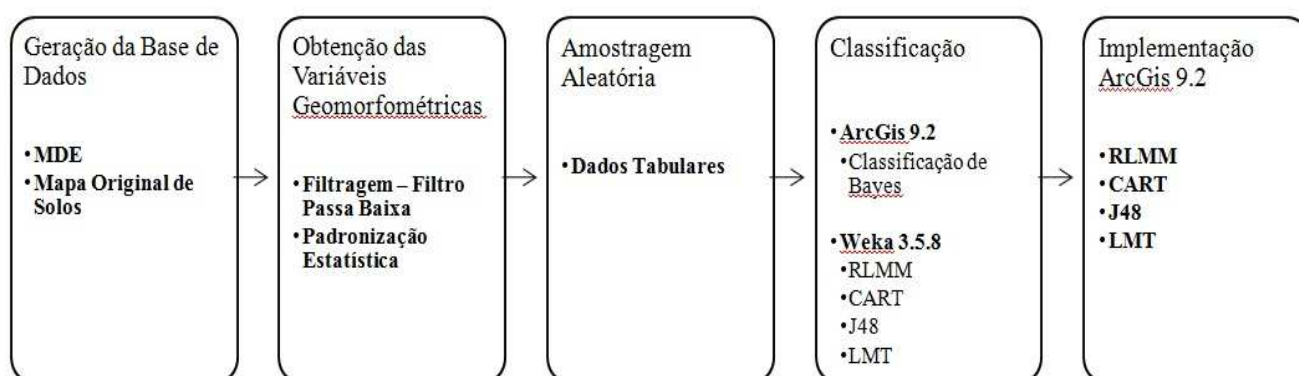


Figura 12 - Fluxo representativo dos procedimentos metodológicos

## 5.1. GERAÇÃO DA BASE DE DADOS

O primeiro passo para o mapeamento digital de solos foi a construção de uma base de dados espacial em um Sistema de Informações Geográficas (SIG). A construção da base de dados consistiu na seleção e relacionamento dos dados que irão alimentar o SIG como o MDE, as variáveis geomorfométricas, o mapa de solos confeccionado por ABRÃO & AZOLIN (1970), atualizado e digitalizado por BORTOLON (2008), compatível com a escala 1:50.000 e os pontos amostrais.

O SIG utilizado foi o ArcGis 9.2 com extensões *Spatial Analyst* e *TauDEM (Terrain Analysis Using Digital Elevation Models)* (TARBOTON, 2008). Também foi utilizado, para classificação supervisionada, o *software* estatístico WEKA 3.5.8; o *software*, que foi desenvolvido na universidade de Waikato na Nova Zelândia, é distribuído gratuitamente e pode ser obtido através do site <http://www.cs.waikato.ac.nz/ml/weka/>.

## 5.2. PROCESSAMENTO DAS VARIÁVEIS GEOMORFOMÉTRICAS

Em ambiente ArcGis 9.2, a partir do MDE, utilizando-se da extensão *Spatial Analyst* foram gerados seis variáveis geomorfométricas, além da própria variável elevação, sendo elas: a) declividade, b) curvatura planar, c) perfil de curvatura, d) curvatura (combinação entre curvatura planar e perfil de curvatura), e) direção do fluxo e f) acúmulo do fluxo; ainda em ambiente ArcGis 9.2, porém utilizando-se da extensão *TauDEM* foi gerada a variável IUT (índice de umidade topográfica) (KIRKBY, 1975).

Para realçar os grandes padrões das variáveis originadas, com a finalidade de eliminar alguma possível imperfeição no MDE, partindo da premissa que a maioria das mudanças de feições espaciais na natureza se dão de forma gradual, como a elevação, curvatura, mudança de paisagem, entre outros, foi utilizado o filtro passa baixa, uma técnica bastante difundida em Sensoriamento Remoto para suavizar os detalhes em uma imagem de satélite. O filtro passa baixa atenua ou elimina os

elementos de alta frequência. As altas frequências espaciais significam mudanças rápidas em poucos pixels, representando mudanças bruscas no modelo numérico do terreno (MNT) (ROSENFELD, 1978).

Cada uma das oito variáveis são MNT representando temas, ou informações diferentes uma das outras, logo suas escalas de valores e unidades de medidas também são diferentes. Por exemplo, a variável elevação não tem pixels negativos e representa a medida em metros, ao passo que a variável perfil de curvatura têm valores negativos e representa o quão côncavo ou convexo é o terreno, já a variável declividade representa a medida em graus e tem um outro tipo de escala de valores. Afim de trazê-las para uma mesma ordem de grandeza, todas as variáveis passaram pelo processo de padronização estatística, conforme a equação:

$$V_p = \frac{(V_o - \mu)}{\sigma} \quad (27)$$

onde  $V_p$  é a variável padronizada,  $V_o$  a variável original,  $\mu$  a média da variável e  $\sigma$  o desvio-padrão da variável. Dessa maneira, todas as variáveis terão médias 0 (zero) e desvio padrão igual a 1 (um). Para tanto, foi utilizada, em ambiente ArcGis 9.2, a extensão *Spatial Analyst* e a ferramenta *Raster Calculator*.

### 5.3. AMOSTRAGEM E SELEÇÃO DAS VARIÁVEIS

Para realizar o processo de amostragem, foi utilizada a ferramenta *Create Random Points* encontrada no ArcGis 9.2. Essa ferramenta permite serem criados pontos amostrais, dentro de uma área de interesse, de forma aleatória, para não tornar a classificação supervisionada subjetiva e tendenciosa. Foram criados 11.000 pontos amostrais aleatórios, o que corresponde a um ponto a cada 10,80km<sup>2</sup>.

Após criados os pontos, as informações de cada uma das variáveis geomorfométricas, referentes a cada ponto amostral, foram inseridas na tabela de atributos dos pontos. Ou seja, automaticamente foram criadas oito colunas, na tabela de atributos dos pontos amostrais, representando as oito variáveis utilizadas no estudo;

após, a informação do pixel de uma variável, sobreposto por um ponto amostral, foi adicionada na tabela de atributos dos pontos na respectiva coluna da variável, resultando em uma tabela com 11.000 linhas (representando os 11.000 pontos amostrais) e oito colunas (representando as oito variáveis). Para tanto, foi utilizada a ferramenta *Sample*, encontrada na extensão *Spatial Analyst* em ambiente ArcGis 9.2.

Afim de compatibilizar a tabela de atributos para ser utilizada no *software* estatístico WEKA 3.5.8, a tabela foi exportada para arquivo do tipo CSV (*Comma Separated Values*, ou valores separados por vírgula).

Com a tabela exportada do ambiente ArcGis 9.2, no formato CSV, e importada no *software* WEKA, foi realizado o processo de seleção de variáveis pelo método de seleção sequencial progressiva, conhecido como SFS (*Sequential Forward Selection*) (JOHN et al., 1994). O algoritmo SFS está estruturado para identificar iterativamente o melhor subconjunto de variáveis que possa ser obtido pela adição de uma variável por iteração, ao invés de selecionar um conjunto completo de variáveis numa única iteração.

## 5.4. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU)

### 5.4.1. CLASSIFICAÇÃO DE BAYES

Em ambiente ArcGis 9.2 e com as variáveis que mais influenciaram nas ocorrências das classes de solo, selecionadas pelo método SFS, foi realizado o processo de classificação de Bayes.

Primeiramente foi gerado uma assinatura dos pontos amostrais com a finalidade de verificar o comportamento multivariado das amostras em relação as variáveis selecionadas. Para tanto, foi utilizada a ferramenta *Create Signatures*, encontrada na extensão *Spatial Analyst*; nessa ferramenta são inseridas as variáveis desejadas e os pontos amostrais em formato *shapefile*, ou outro formato padrão do ArcGis, resultando em um arquivo com extensão \*.gsg que contém as informações multivariadas sobre os pontos.

O próximo passo foi realizar o processo de classificação, utilizando-se da ferramenta *Maximum Likelihood Classification* (ou classificação por máxima verossimilhança) também encontrada na extensão *Spatial Analyst* do ArcGis, onde o arquivo de assinatura multivariada, de extensão \*.gsg, serviu como parâmetro de entrada de informação multivariada. A classificação por máxima verossimilhança é uma variação da classificação de Bayes pois considera todas as probabilidades *a priori* iguais, porém nessa ferramenta podemos inferir a probabilidade *a priori* para todas as classes, o que resulta, então, na classificação de Bayes. Como as amostras de treinamento foram extraídas de forma aleatória e de uma área onde se conhece toda a verdade terrestre, o mapa inteiro, a probabilidade *a priori* foi estabelecida pela proporção de amostras de cada classe, logo, uma classe de solo com maior área de ocorrência teve sua probabilidade *a priori* maior que uma classe de solo com menor área de ocorrência no mapa original.

#### 5.4.2. REGRESSÕES LOGÍSTICAS MÚLTIPLAS MULTINOMIAIS (RLMM)

Em ambiente WEKA 3.5.8, foram realizados os métodos das RLMM, CART, J48 e LMT. Para cada método, foi resultante um arquivo de texto contendo os coeficientes dos classificadores, logo foi possível a implementação, em ambiente ArcGis 9.2, de mapas através da ferramenta *Raster Calculator*, encontrada na extensão *Spatial Analyst*.

Para o método das RLMM, foi resultante uma tabela contendo o valor  $\beta_j$  (coeficientes calculados) do modelo em relação as variáveis ( $x_i$ ). Utilizando-se da ferramenta *Raster Calculator*, foram gerados mapas das probabilidades de ocorrência das classes ( $p_1, \dots, p_J$ ), calculadas da seguinte maneira:

$$p_j = \frac{\exp(x_i \cdot \beta_j)}{1 + \sum_{j=1}^{J-1} \exp(x_i \cdot \beta_j)} \quad \text{e} \quad p_J = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(x_i \cdot \beta_j)} \quad (28)$$



sendo  $p_j$  as probabilidades da primeira à penúltima classe e  $p_j$  a probabilidade da última classe. Os mapas de probabilidade produzidos têm, como valor do pixel, a probabilidade, entre 0 e 1, daquele pixel pertencer àquela classe, dessa maneira foi calculado, pixel a pixel, aquele que tinha a maior probabilidade para a classe em questão, sendo que o somatório do vetor de pixels deve ser, obrigatoriamente, igual a 1, ou 100%. Em um exemplo genérico de três classes temos:

$$[MC1] = [pC1] > [pC2] \text{ OR } [pC1] > [pC3] \quad (29)$$

$$[MC2] = [pC2] > [pC1] \text{ OR } [pC2] > [pC3] \quad (30)$$

$$[MC3] = [pC3] > [pC1] \text{ OR } [pC3] > [pC2] \quad (31)$$

onde:

[MC1] é o mapa da classe 1

[MC2] é o mapa da classe 2

[MC3] é o mapa da classe 3

[pC1] é a probabilidade de ocorrência da classe 1

[pC2] é a probabilidade de ocorrência da classe 2

[pC3] é a probabilidade de ocorrência da classe 3

após calculados os mapas de cada classe, é confeccionado o mapa final [MF]:

$$[MF] = [MC1] \text{ OR } [MC2] \text{ OR } [MC3] \quad (32)$$

## 5.5. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA)

Os mapas gerados pelos métodos CDA tiveram maior custo de tempo e trabalho para serem produzidos. Para os algoritmos CART e J48, foi utilizado a ferramenta *Raster Calculator* e seus operadores lógicos nível a nível da árvore até

chegar aos nós finais, como mostra o exemplo genérico abaixo (Figura 13) com três classes (A, B e C) e quatro variáveis (x1, x2, x3 e x4):

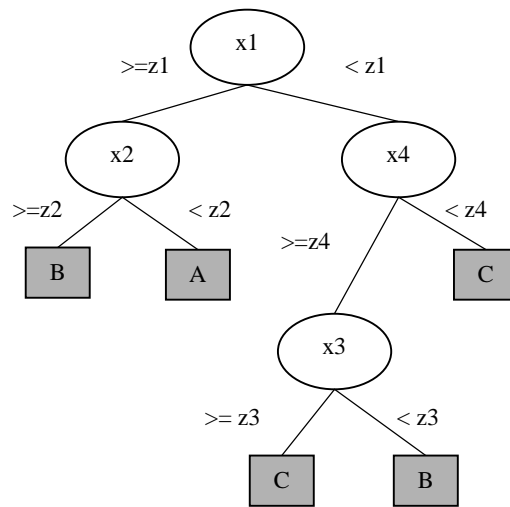


Figura 13 – Estrutura dos CDA CART e J48

Para a implementação da árvore em ambiente ArcGis, são necessárias várias fórmulas no *Raster Calculator*, como as demonstradas a seguir:

$$[MCA] = x1 \geq z1 \text{ OR } x2 < z2 \quad (33)$$

$$[MCB\_1] = x1 \geq z1 \text{ OR } x2 \geq z2 \quad (34)$$

$$[MCB\_2] = x1 < z1 \text{ OR } x4 \geq z4 \text{ OR } x3 < z3 \quad (35)$$

$$[MCC\_1] = x1 < z1 \text{ OR } x4 < z4 \quad (36)$$

$$[MCC\_2] = x1 < z1 \text{ OR } x4 \geq z4 \text{ OR } x3 \geq z3 \quad (37)$$

onde:

[MCA] é o mapa da classe A

[MCB] é o mapa da classe B sendo subdivido em [MCB\_1] e [MCB\_2] por estar em mais de um nó terminal.

[MCC] é o mapa da classe C, sendo subdivido em [MCC\_1] e [MCC\_2] por estar em mais de um nó terminal.

Após calculados todos os nós terminais, os mapas correspondendo as mesmas classes são agrupados:

$$[MCB] = [MCB_1] \text{ OR } [MCB_2] \quad (38)$$

$$[MCC] = [MCC_1] \text{ OR } [MCC_2] \quad (39)$$

e, para confeccionar o mapa final [MF] tem-se:

$$[MF] = [MCA] \text{ OR } [MCB] \text{ OR } [MCC] \quad (40)$$

Ainda utilizando-se do exemplo com quatro variáveis ( $x_1$ ,  $x_2$ ,  $x_3$  e  $x_4$ ) e três classes (A, B e C), para o algoritmo LMT (Figura 14), os passos, para implementação em ambiente ArcGis 9.2, foram os mesmos até os nós terminais da árvore, onde foram implementadas fórmulas para se chegar à probabilidade de ocorrência das classes em cada nó terminal.

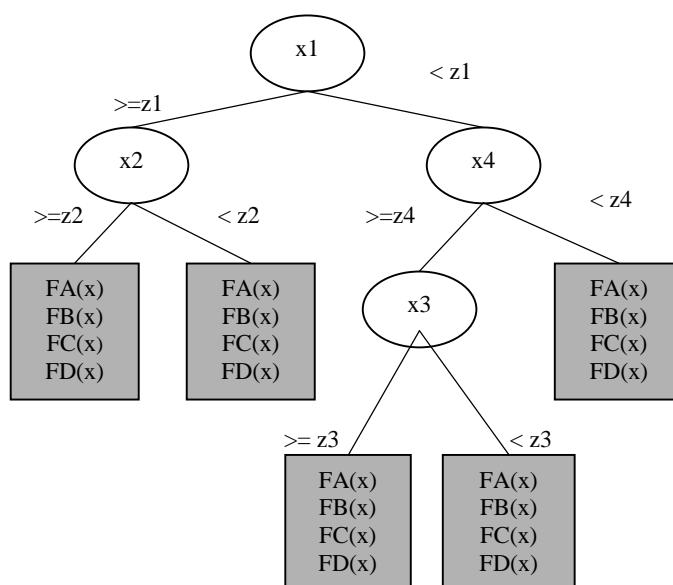


Figura 14 – Estrutura do CDA LMT

Através das funções resultantes, para cada classe, nos nós terminais, é possível estabelecer as probabilidades de ocorrência das classes para cada nó terminal:

$$p_j = \frac{\exp F_j(x)}{\exp F_1(x) + \dots + \exp F_j(x)} \quad \text{e} \quad p_J = \frac{\exp F_J(x)}{\exp F_1(x) + \dots + \exp F_J(x)} \quad j = 1, \dots, J-1 \quad (41)$$

Os mapas de probabilidade produzidos têm, como valor do pixel, a probabilidade, entre 0 e 1, daquele pixel pertencer àquela classe; ao final, a classe cujo pixel tinha maior probabilidade de ocorrência foi a escolhida para rotular aquele pixel, e assim por diante, até que todos os pixels fossem rotulados por alguma classe. Calculados todos os nós terminais, os mapas correspondentes a mesma classe foram agrupados e após confeccionado o mapa final.

## 5.6. ANÁLISE DAS ACURÁCIAS

Os mapas produzidos foram comparados, pixel a pixel, com o mapa original a fim de realizar análise dos resultados por meio de matrizes de erros, ou matriz de confusão (CONGALTON, 1991). A matriz de erro, neste caso, expressa a relação entre os pixels classificados e originais, possibilitando a análise de quatro indicadores de acurácia, sendo eles:

- a) acurácia global, obtida pela divisão do número de pixels corretamente classificados com o total de pixels, indica a proporção de pixels corretamente classificados em relação ao número total de pixels;
- b) acurácia do produtor, calculado, classe a classe, dividindo os pixels classificados corretamente com o total de pixels do mapa original, indica a proporção que uma certa unidade de mapeamento foi classificada corretamente;
- c) acurácia do usuário, calculada, classe a classe, dividindo os pixels classificados corretamente com o total de pixels do mapa produzido, indica a certeza que o

usuário tem, ao analisar o mapa, que um ponto específico, em uma certa unidade de mapeamento, está classificado corretamente; e

- d) coeficiente Kappa de Cohen (COEHN, 1960), o qual mede o quanto da concordância observada se aproxima da concordância perfeita, retirando-se o efeito da concordância ao acaso. O índice kappa pode ser calculado da seguinte forma:

$$K = \frac{n \sum_{i=1}^j x_{ii} - \sum_{i=1}^j x_{i\oplus} x_{\oplus i}}{n^2 - \sum_{i=1}^j x_{i\oplus} x_{\oplus i}} \quad (42)$$

sendo  $x_{ii}$  o valor da linha  $i$  e coluna  $i$  da matriz de erros;  $x_{i\oplus}$  o total de pixels da linha  $i$ ;  $x_{\oplus i}$  o total de pixels da coluna  $i$ ;  $n$  o número de pixels; e  $j$  o número total de classes.

## 6. RESULTADOS E DISCUSSÃO

A Figura 15 apresenta o subconjunto de variáveis selecionadas.

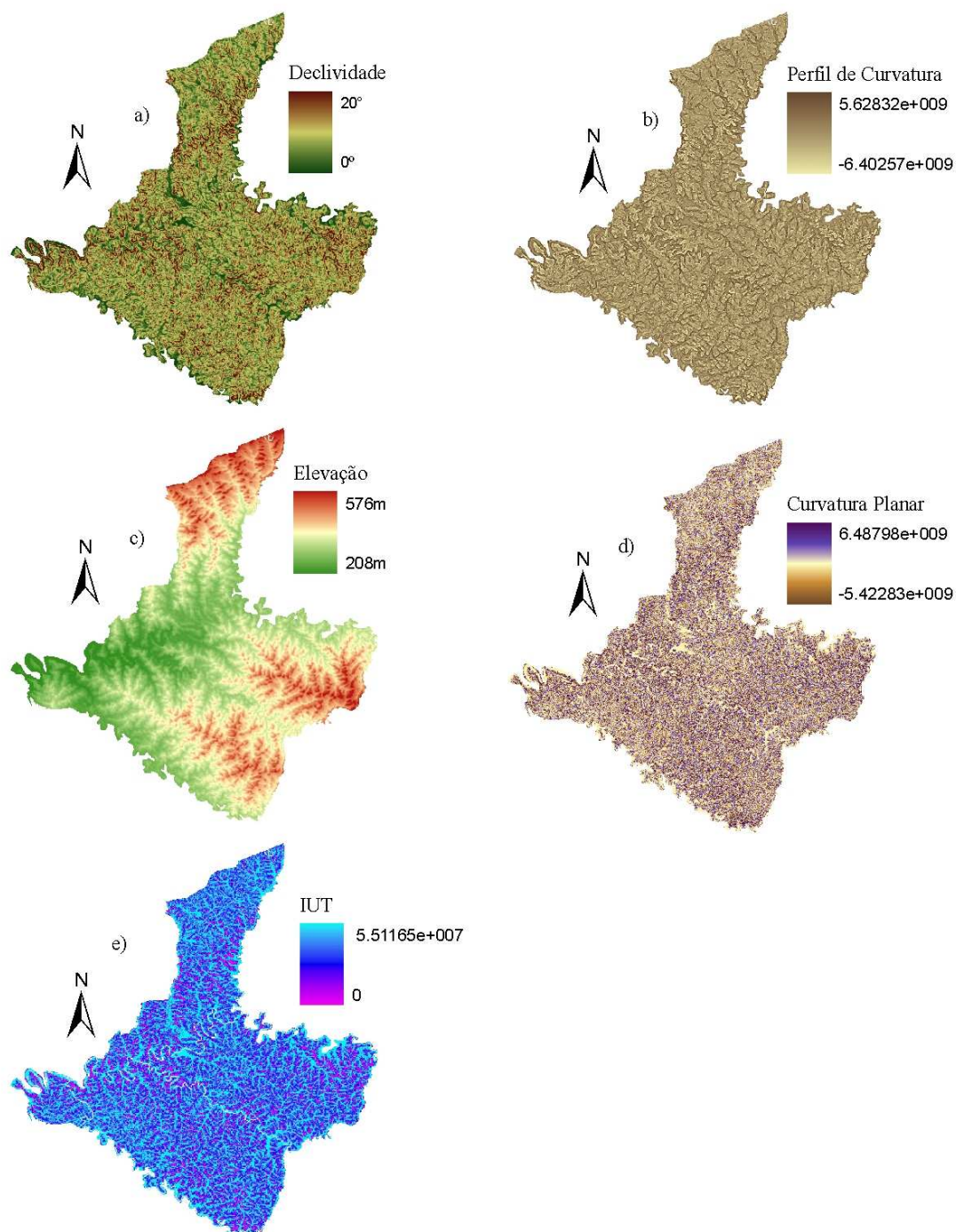


Figura 15 – Subconjunto das variáveis selecionadas a partir do método SFS no Weka 3.5.8. a)Declividade; b)Perfil de Curvatura; c)Elevação; d)Curvatura Planar e e)IUT

Com 11.000 pontos amostrais aleatórios, o que corresponde a cerca de um ponto a cada 10,80km<sup>2</sup>, as variáveis geomorfométricas foram selecionadas pelo método SFS (*Sequential Forward Selection*) com a finalidade de apresentar o subconjunto de variáveis que mais influencia na ocorrência das classes de solo na área de estudo. Em ordem de importância, as variáveis selecionadas foram: a) declividade, b) perfil de curvatura, c) elevação, d) curvatura planar, e e) índice de umidade topográfica (IUT).

Para verificar, numericamente, a correlação das variáveis, foi gerada a Tabela 2.

Tabela 02 - Matriz de correlação entre as variáveis geomorfométricas.

Matriz de Correlação					
Variável	Declividade	Peril de Curvatura	Elevação	Curvatura Planar	IUT
Declividade	1,00000	-0,49353	0,14676	0,27336	0,78943
Perfil de Curvatura		1,00000	-0,18465	-0,43769	-0,61950
Elevação			1,00000	0,01202	0,08446
Curvatura Planar				1,00000	0,60873
IUT					1,00000

As maiores correlações se dão entre as variáveis declividade e IUT com 0,78943 e as variáveis curvatura planar e IUT, com 0,60873 de correlação; entre as variáveis perfil de curvatura e IUT, observa-se correlação negativa de -0,61950. A variável elevação é destaque por apresentar correlação baixa com as demais variáveis, ou seja, os valores mais próximos a zero.

## 6.1. CLASSIFICAÇÃO EM ESTÁGIO ÚNICO (CEU)

### 6.1.1. CLASSIFICAÇÃO DE BAYES

Para análise do comportamento univariado das amostras nas variáveis padronizadas, foram gerados histogramas, para melhor visualização, na Figura 16.

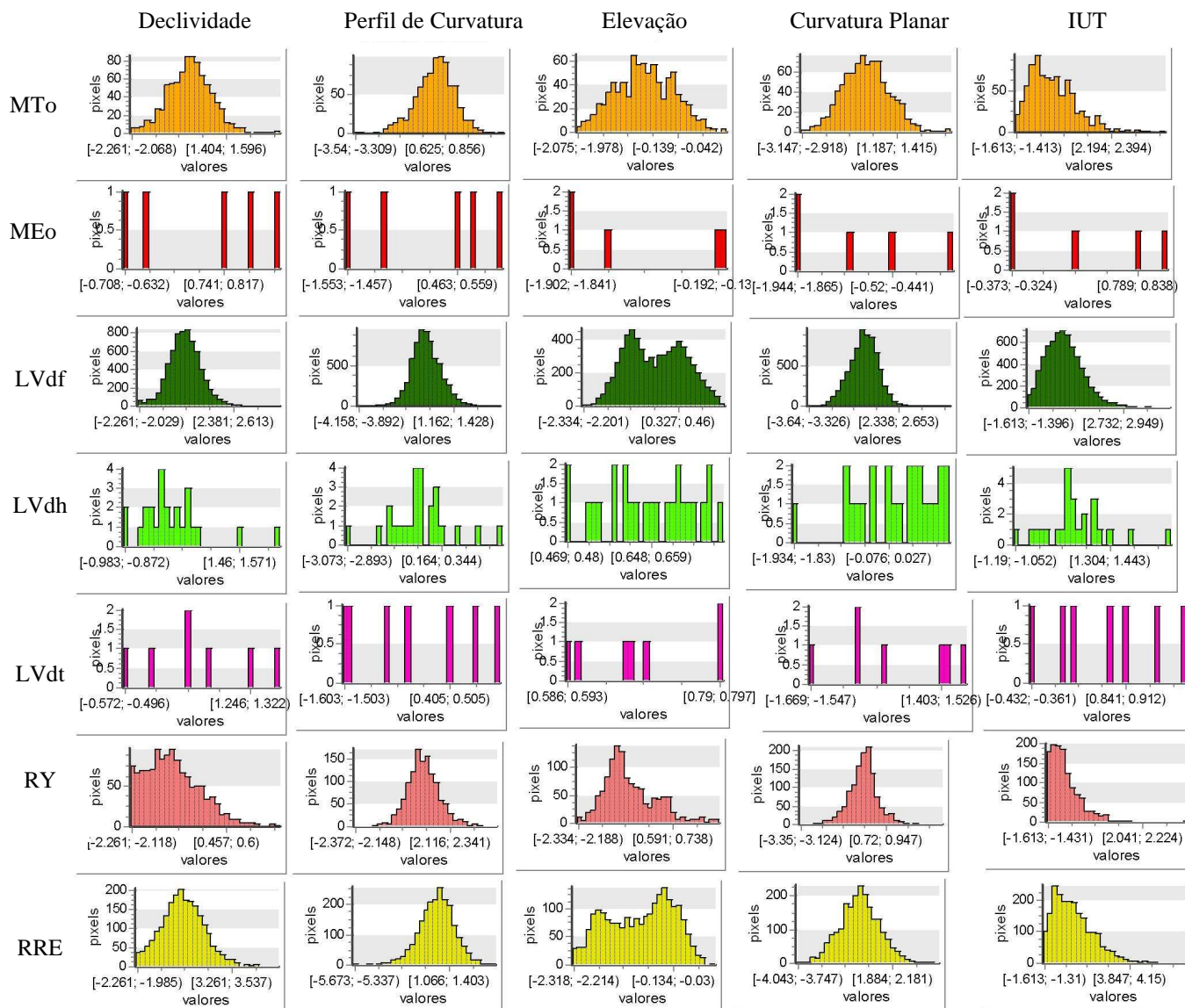


Figura 16 - Histogramas das classes para cada variável geomorfológica

Destaca-se, nas classes MEo, LVdh e LVdt, a distribuição irregular das amostras, o que pode ser explicado pelo pequeno número de amostras obtidos para essas classes. O histograma da classe LVdf e a variável elevação apresenta comportamento bimodal, o que pode ocasionar confusão para o classificador, bem como ocorre na classe RRE e a variável elevação; verifica-se que os demais resultados se aproximam de uma distribuição normal univariada.

Ao estabelecer uma relação visual entre o mapa original de solos e o mapa produzido pelo método de Bayes (Figura 17), observa-se certa semelhança, que não pode ser mensurada somente pelo critério visual, porém, por essa semelhança, podemos afirmar que o método se adequou na predição das classes de solo.



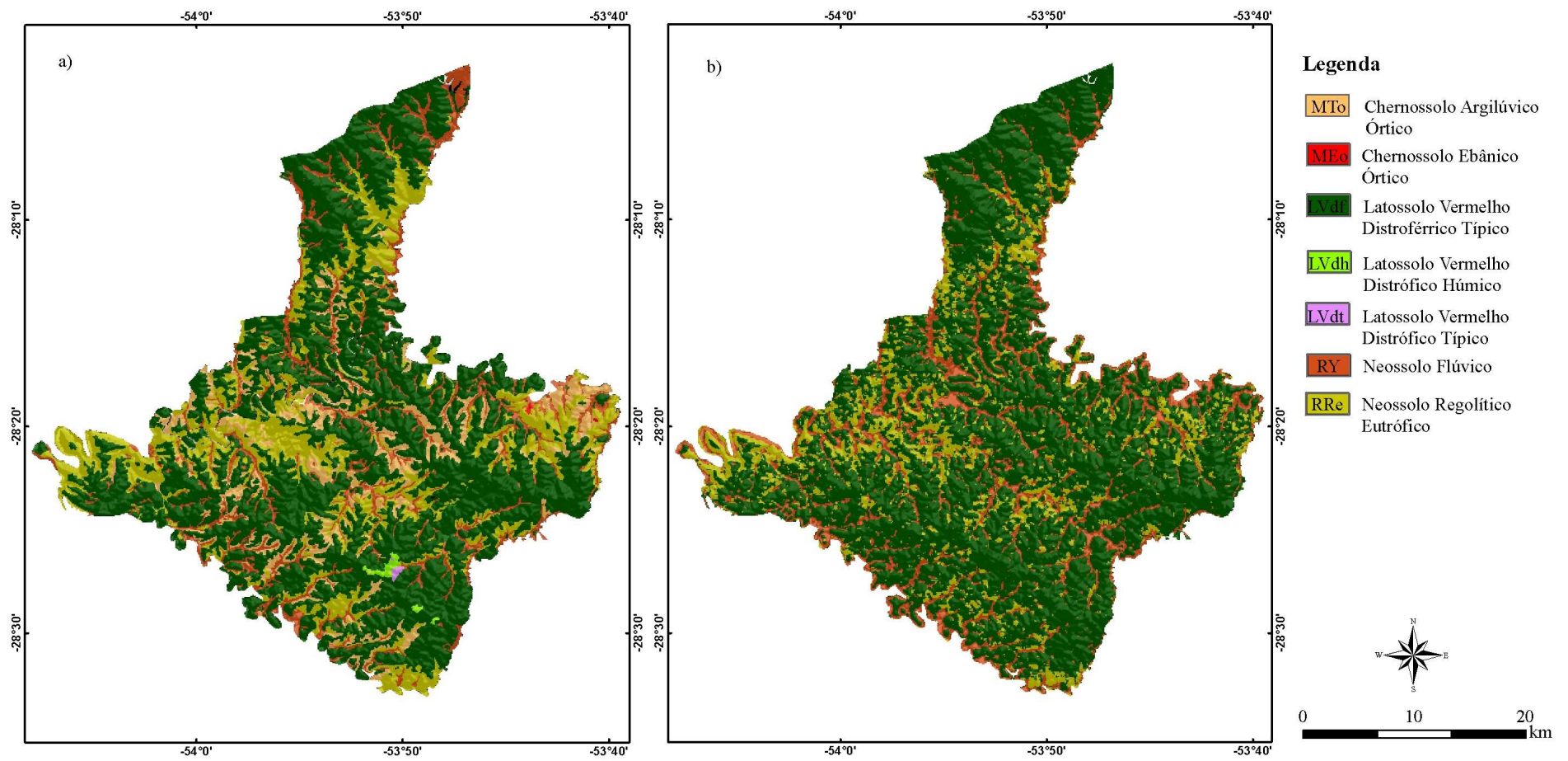


Figura 17 - Mapa de solos original (a) e mapa de solo produzido pela classificação de Bayes (b)

O grau de semelhança pode ser obtido através de matriz de erro (Tabela 3). Nesta, observa-se que o método de Bayes apresentou acurácia global (AG) de 66,56% e índice Kappa de 35,85%; a classe com maior acurácia foi a LVdf com 74,04 % de acurácia do usuário (AU) e 86,19% de acurácia do produtor (AP), a classe RRE apresentou 55,58% de AU porém AP de 39,22% e a classe RY apresentou a AU com 41,22% e 56,43% de AP; as classes MTo, MEo, LVdh e LVdt não foram mapeadas pelo algoritmo.

Tabela 03 - Matriz de erro do classificador de Bayes a partir do mapa original de solos

		MAPA ORIGINAL									
		Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)
MAPA PRODUZIDO	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdf	5,67	0,03	52,71	0,25	0,07	3,96	8,49	71,19	74,04	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	RY	1,15	0,01	4,34	0,01	0,00	6,21	3,35	15,06	41,22	
BAYES	RRE	1,14	0,01	4,11	0,01	0,00	0,83	7,64	13,75	55,58	
	Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00		
	AP (%)	*	*	86,19	*	*	56,43	39,22			
		AG = 66,56%				Kappa = 35,85%					

\* classe não ocorrente no mapa produzido.

Os resultados apresentam a dificuldade do método preditivo identificar as classes MTo, MEo, LVdh e LVdt. Isso pode ser explicado pelo fato das área dessas classes serem, exceto a MTo, praticamente irrelevantes em relação ao total da área. O total de pixels da classe MTo representa 7,96% do total, ou seja, sua área é relativa a 7,96% da área total, já as classes MEo, LVdh e LVdt representam, cada uma, menos que 0,5% da área total, com respectivamente 0,04%, 0,27% e 0,08%. Somando-se a área dessas três últimas classes temos apenas 0,39% da área total.

Com a finalidade de verificar se a utilização de uma legenda simplificada aumentaria a acurácia dos mapas produzidos, mesmo com a perda de detalhamento de solos na escala original de 1:50.000, as classes foram agrupadas em quatro classes maiores: Neossolo Flúvico (RY), Neossolo Regolítico Eutrófico (RRe), Chernossolo (M, agrupamento de MTo e MT) e Latossolo Vermelho (LV, agrupamento de LVdf, LVdh e LVdt), respeitando as características de cada tipo de solo. Os novos histogramas foram gerados e agrupados conforme apresenta a Figura 18.

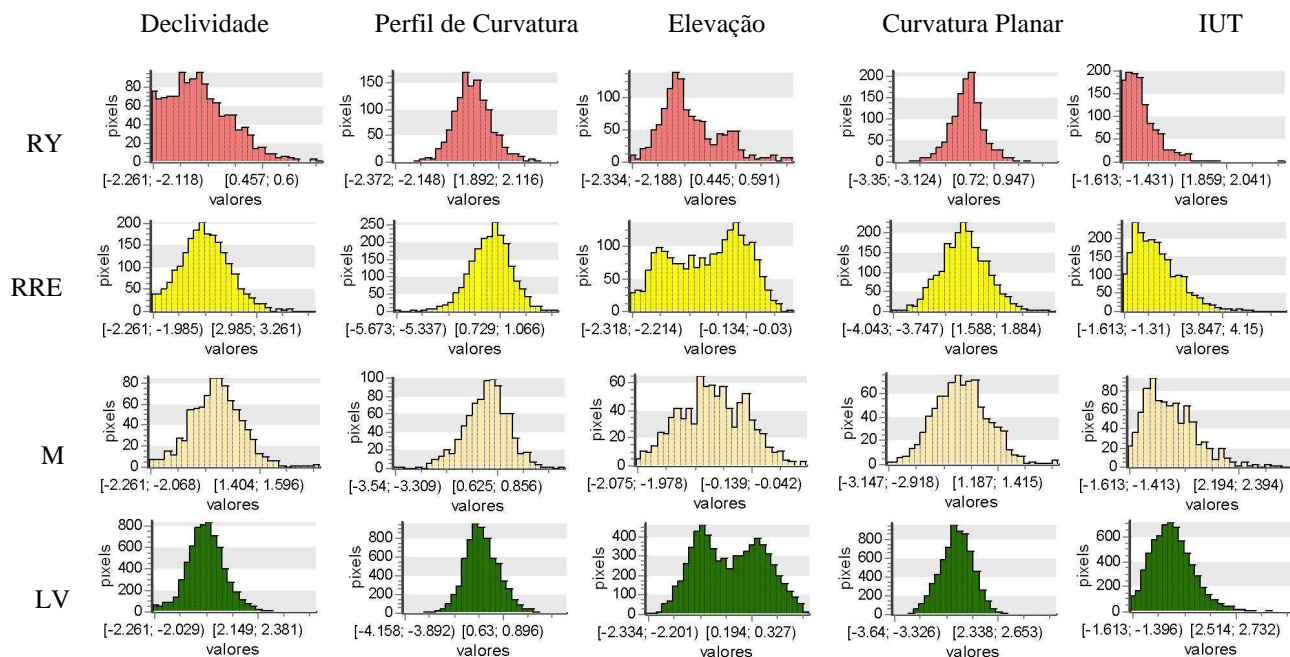


Figura 18 - Histogramas das variáveis para as classes agrupadas

Verifica-se melhor distribuição, em relação aos histogramas das classes não agrupadas, onde os histogramas aproximam-se das distribuição normal, com exceção daqueles relacionados com a variável elevação das classes RRE e LV, que apresentam-se de forma bimodal.

Com a simplificação da legenda, notou-se que ocorreu um pequeno aumento na AG que passou de 66,56% para 66,89%, o que pode ser considerado como ínfimo devido ao número de pixels classificados. Em relação ao índice Kappa, também houve um ligeiro ganho, passando de 35,85% pra 36,17%.

Em relação às AU e AP de cada classe, obteve-se os seguintes resultados (Tabela 4): classe RY com 56,43% de AP e 41,22% de AU, a classe RRE com 39,22% de AP e 55,57% de AU, e a classe LV 86,23% e 74,50% de AP e AU respectivamente. Mesmo com a generalização das classes, a classe M não foi mapeada pelo método bayesiano. Em relação à esses indicadores podemos afirmar que não houve diferença entre a classificação com a legenda original e a simplificada.

Tabela 04 - Matriz de erro do classificador de Bayes a partir do mapa de solos, ambos com legenda simplificada

		MAPA COM LEGENDA SIMPLIFICADA						
		Classe	RY	RRE	M	LV	Total	AU (%)
MAPA PRODUZIDO		RY	6,21	3,35	1,16	4,34	15,06	41,22
		RRE	0,83	7,64	1,16	4,13	13,75	55,57
		M	0,00	0,00	0,00	0,00	0,00	*
		LV	3,96	8,49	5,69	53,04	71,19	74,50
		Total	11,00	19,49	8,01	61,50	100,00	
BAYES		AP (%)	56,43	39,22	*	86,23		
			AG = 66,89%			Kappa = 36,17%		

\* classe não ocorrente no mapa produzido.

A Figura 19 estabelece a relação visual entre o mapa original e o produzido pelo classificador de Bayes, ambos com legenda simplificada.

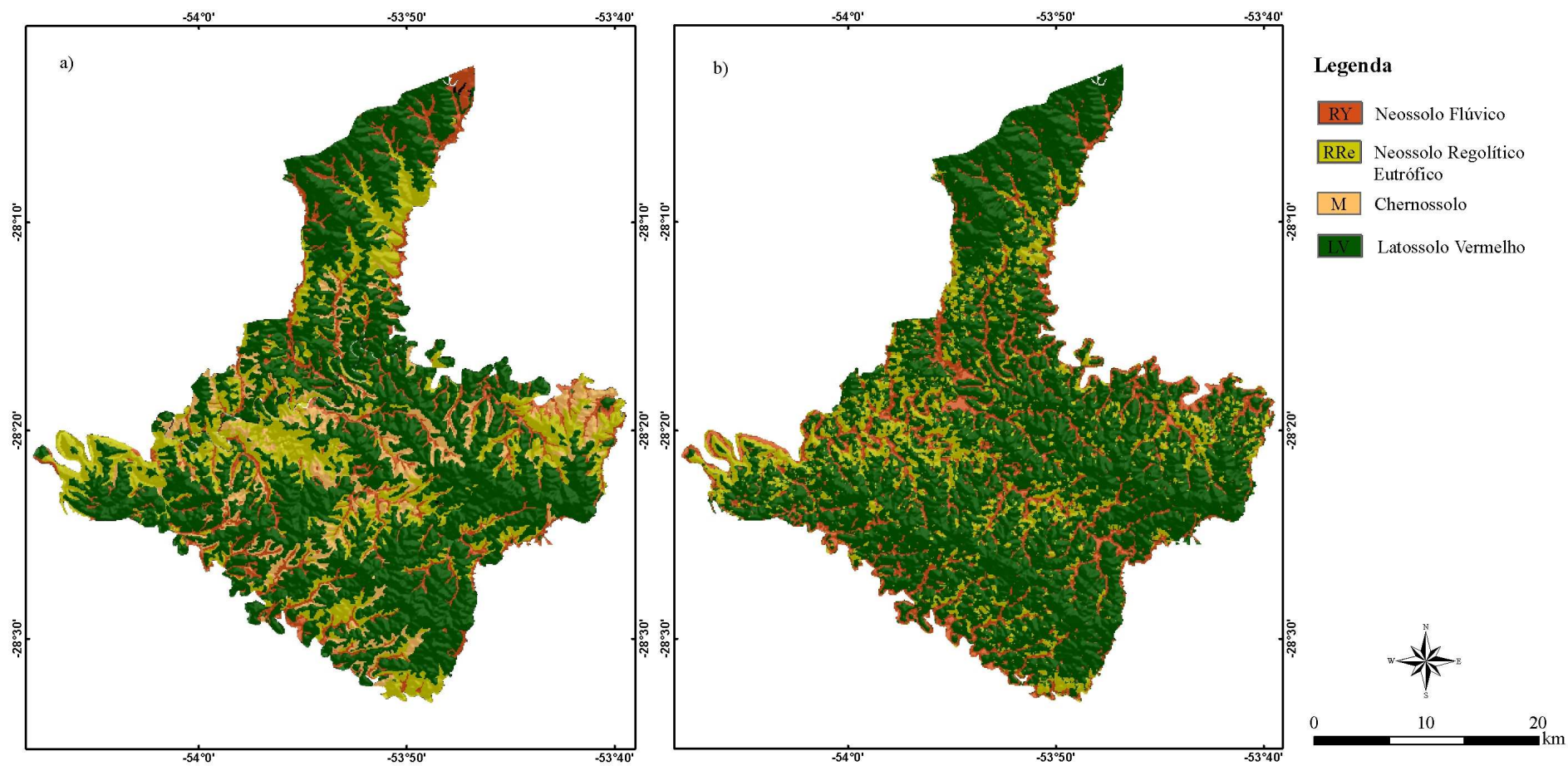


Figura 19 - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pela classificação de Bayes com legenda simplificada (b)

### 6.1.2. REGRESSÕES LÓGICAS MÚLTIPLAS MULTINOMIAIS (RLMM)

Os coeficientes das RLMM, obtidos no *software* estatístico WEKA e após implementados no SIG, podem ser visualizados na Tabela 5.

Tabela 05 - Coeficientes do modelo logístico

Variável	MTo / LVdt	LVdf / LVdt	RY / LVdt	RRe / LVdt	MEo / LVdt	LVdh / LVdt
Elevação	-3,1333	-2,2986	-2,3424	-3,3953	-4,1441	-0,6427
IUT	-0,7421	-0,7221	-1,4432	-0,3663	-0,0591	0,5596
Declividade	-0,0112	-0,5960	-1,2439	0,4467	-0,1023	-1,5396
Curvatura Planar	-0,1915	-0,1052	0,2036	-0,2837	-1,3031	-0,4572
Perfil de Curvatura	-0,3469	-1,1273	-0,0177	0,0637	-0,5699	-1,0542
Constante	5,2279	7,7603	5,0045	5,6458	-1,4078	1,5715

As probabilidades de ocorrências de cada classe, elaboradas a partir dos parâmetros e coeficientes do modelo logístico descrito acima, foram espacializadas (Figura 20), resultando em sete mapas de probabilidade, correspondendo às sete classes de solo presentes na área de estudo.

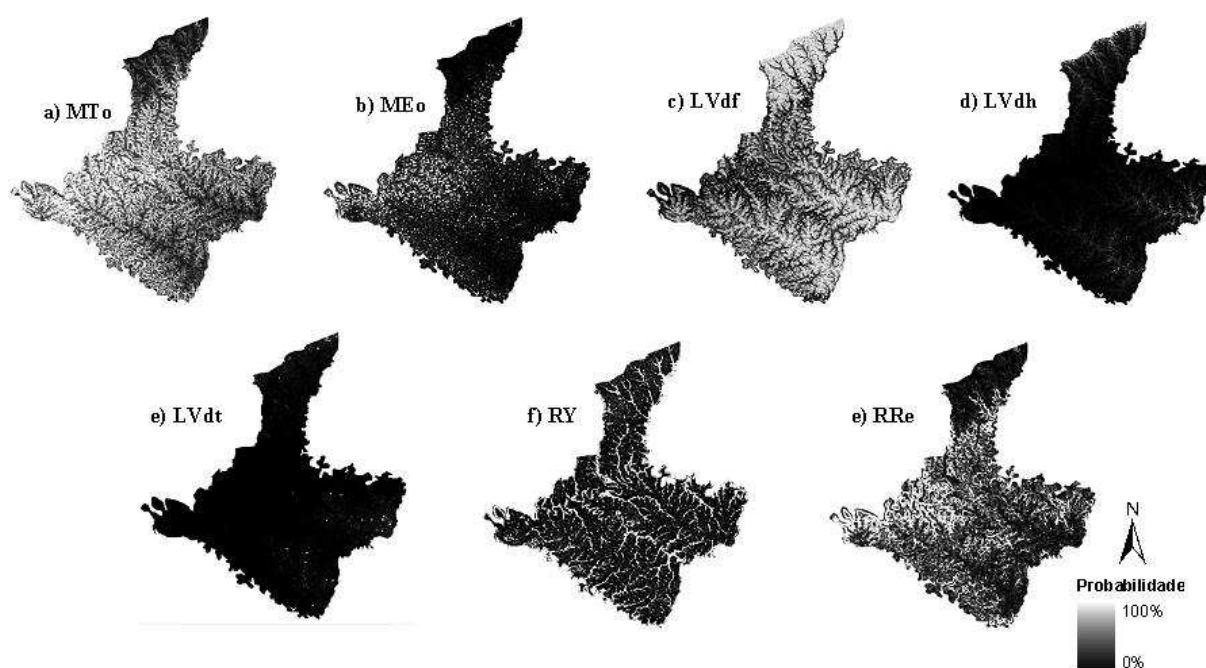


Figura 20 - Mapas de probabilidade de ocorrência de cada classe de solo a partir dos coeficientes do modelo logístico

Cada mapa de probabilidade possui como valores limites 0% (quando a probabilidade de ocorrência da classe em questão no pixel, é zero) e 100% (quando a probabilidade de ocorrência da classe em questão no pixel é 100%), ou seja, na Figura 22, quanto mais claros os pixels, maior a probabilidade de ocorrência da classe em questão.

É possível observar que os mapas das classes LVdh e LVdt são praticamente todos escuros, ou seja, as probabilidade de ocorrência dessas classes são baixas em praticamente toda região de estudo. Também é possível notar, visualmente, a forte relação entre a classe RY com os cursos d'água; as maiores probabilidades de ocorrência, dessa classe, se dão ao longo dos cursos d'água ao passo que o mapa LVdf apresenta relação inversa aos cursos d'água, pois suas maiores probabilidades de ocorrência se dão nas cabeceiras e divisores de água.

A partir do cruzamento dos mapas de probabilidade de cada classe, onde as classes estabelecidas, para cada pixel, foram aquelas que tiveram maior probabilidade de ocorrência entre todas as classes para o pixel em questão, foi possível a produção do mapa estabelecido por RLMM (Figura 21).



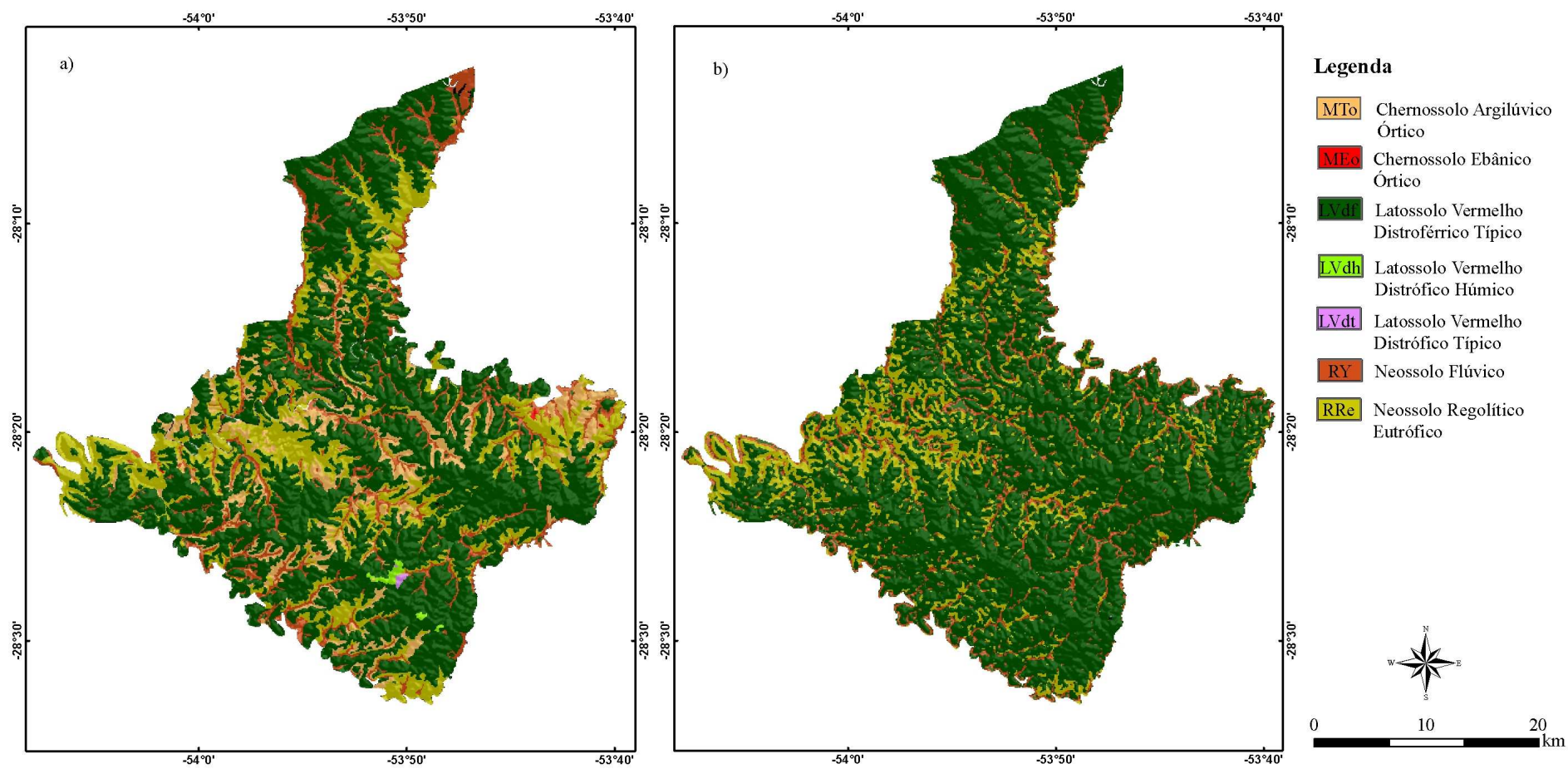


Figura 21 - Mapa de solo original (a) e mapa de solo produzido por RLMM (b)



O método RLMM apresentou acurácia global (AG) de 67,01% e índice Kappa de 32,81%; a classe com maior acurácia foi a LVdf, com 71,54% de acurácia do usuário (AU) e 90,45% de acurácia do produtor (AP), seguida pelas classes RRE, com 56,30% de AU e 38,59% de AP, e a classe RY com 44,83% de AU e 38,03% de AP; as classes MTo, MEo, LVdh e LVdt não foram mapeadas (Tabela 6).

Tabela 06 - Matriz de erro do mapa produzido por RLMM a partir do mapa original de solos

		MAPA ORIGINAL									
MAPA PRODUZIDO		Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)
RLMM		MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		Meo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		LVdf	6,03	0,04	55,31	0,27	0,08	5,79	9,80	77,31	71,54
		LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		RY	0,63	0,00	2,35	0,00	0,00	4,18	2,17	9,33	44,83
		RRE	1,31	0,01	3,50	0,00	0,00	1,02	7,52	13,36	56,30
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
		AP (%)	*	*	90,45	*	*	38,03	38,59		
			AG = 67,01%				Kappa = 32,81%				

\* classe não ocorrente no mapa produzido.

A partir da simplificação da legenda, com o agrupamento das classes em quatro classes (RY, RRe, M e LV), respeitando as características de cada tipo de solo, foi realizado a análise sobre um possível aumento nas acurácias do modelo logístico (Figura 22).

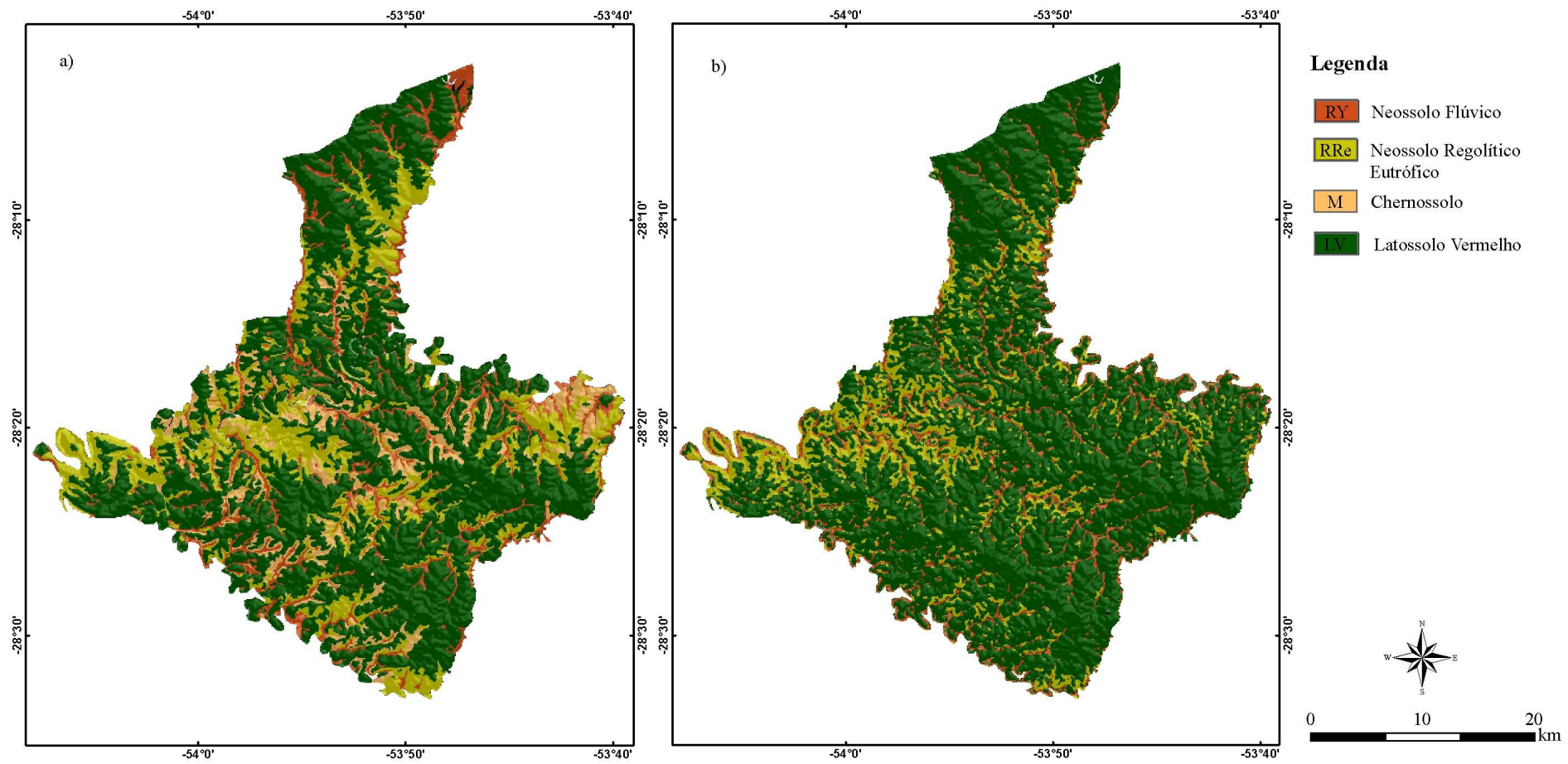


Figura 22 - Mapa de solos com legenda simplificada (a) e mapa de solo produzido por RLMM com legenda simplificada (b)

Com a simplificação da legenda, notou-se que ocorreu um pequeno aumento na AG, a qual passou de 67,01% para 67,36%, o que pode ser considerado, do mesmo modo que no método bayesiano, como ínfimo devido ao número de pixels classificados. Em relação ao índice Kappa, também houve um ligeiro ganho, passando de 32,01% pra 33,14%.

Conforme a Tabela 7, temos os seguintes AU e AP para cada classe: a classe RY apresentou 38,03% e 44,82% de AP e AU respectivamente, a classe RRE 38,59% de AP e 56,30% de AU, a classe LV 90,49% de AP e 71,99% de AU; e a classe M não foi mapeada.

Tabela 07 - Matriz de erro do mapa produzido por RLMM a partir do mapa de solos, ambos com legenda simplificada

		MAPA COM LEGENDA SIMPLIFICADA					
Classe		RY	RRE	M	LV	Total	AU (%)
MAPA PRODUZIDO	RY	4,18	2,17	0,63	2,35	9,33	44,82
	RRE	1,02	7,52	1,31	3,50	13,36	56,30
	M	0,00	0,00	0,00	0,00	0,00	*
	LV	5,79	9,80	6,07	55,65	77,31	71,99
	Total	11,00	19,49	8,01	61,50	100,00	
RLMM	AP (%)	38,03	38,59	*	90,49		
		AG = 67,36%			Kappa = 33,14%		

\* classe não ocorrente no mapa estimado.

## 6.2. CLASSIFICAÇÃO DE DECISÃO EM ÁRVORE (CDA)

### 6.2.1. CLASSIFICATION AND REGRESSION TREES (CART)

Os resultados, obtidos no *software* WEKA 3.5.8, apresentaram uma árvore com 15 nós terminais, 7 níveis hierárquicos e de tamanho 28, que representam os 28 passos para o término da classificação (Figura 23).

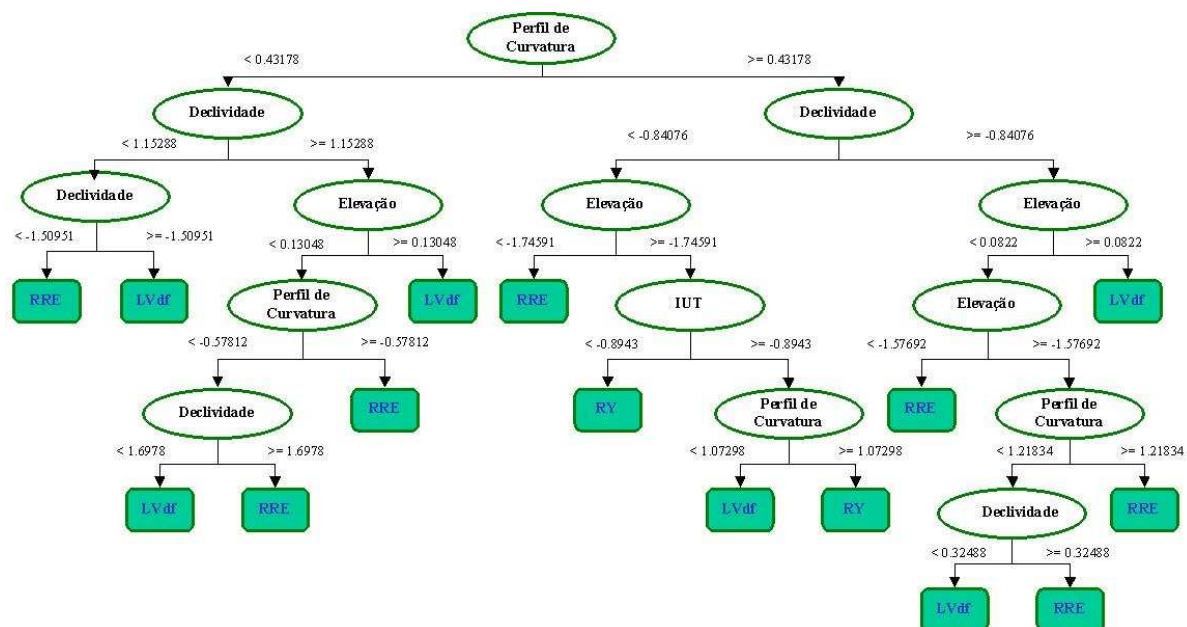


Figura 23 - Árvore de decisão do CDA CART

Em ambiente ArcGis 9.2, a árvore gerada no WEKA 3.5.8 foi implementada, através de álgebra de mapas, o que resultou no mapa apresentado na Figura 24. Visualmente, é possível observar que a semelhança entre o mapa original e o produzido por CART é maior que os mapas produzidos por CEU, e que mostra a eficácia do método hierárquico para a classificação.

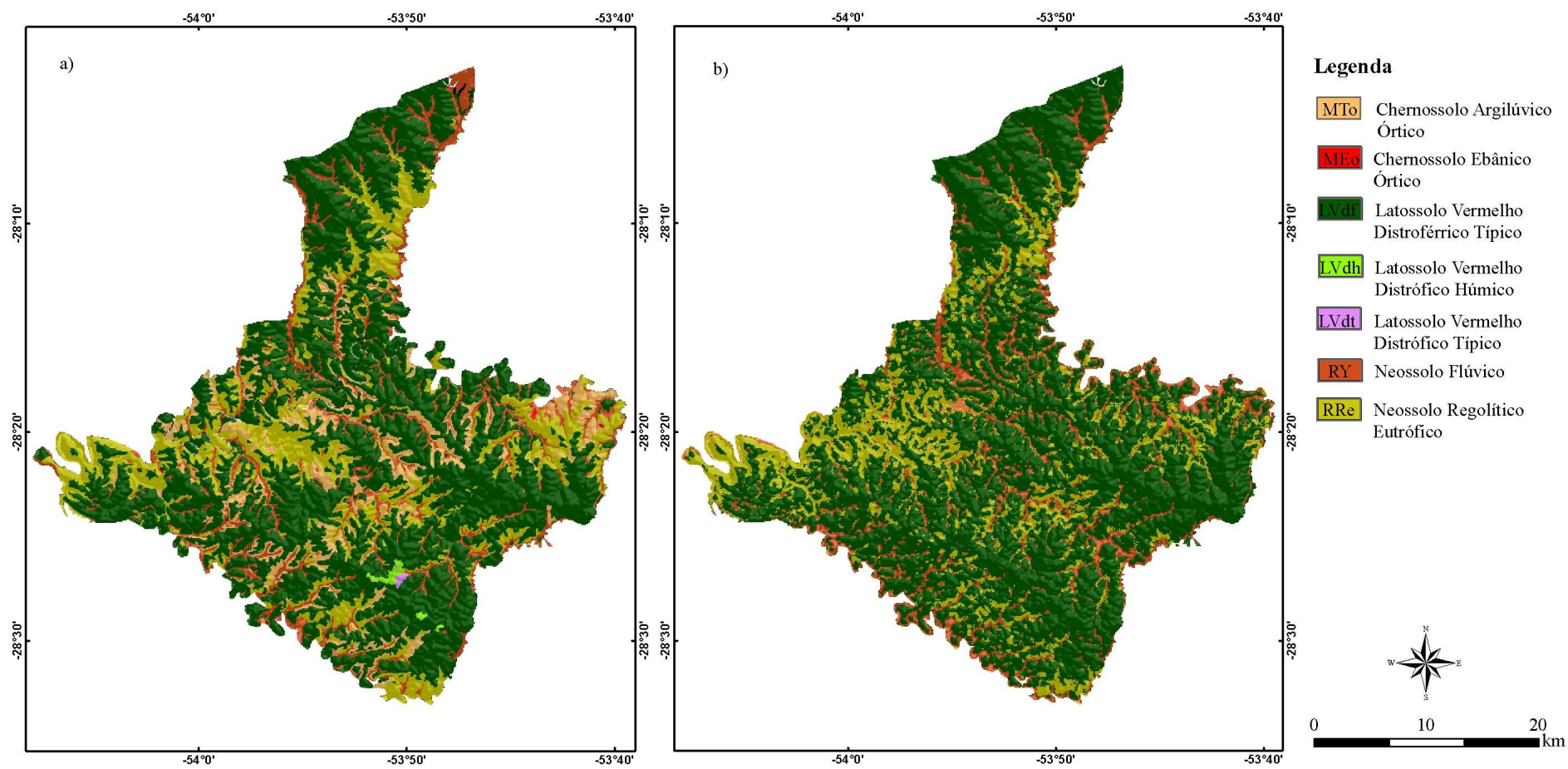


Figura 24 - Mapa de solo original (a) e mapa produzido pelo CDA CART (b)

Observa-se, na Tabela 8, que o método CART apresentou AG de 68,02% e Kappa de 38,45%; a classe LVdf apresentou as melhores AU e AP com, respectivamente, 74,64% e 86,60%; a classe RY teve 48,18% de AU e 52,06% de AP e a classe RRE 54,38% e 47,90% de AU e AP, respectivamente.

Tabela 08 - Matriz de erro do mapa produzido pelo CDA CART e o mapa original de solos

		MAPA ORIGINAL									
		Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)
MAPA PRODUZIDO  CART	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdf	5,55	0,03	52,96	0,27	0,08	3,66	8,41	70,95	74,64	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	RY	0,82	0,00	3,59	0,00	0,00	5,73	1,74	11,89	48,18	
	RRE	1,60	0,01	4,60	0,00	0,00	1,62	9,33	17,16	54,38	
	Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00		
	AP (%)	*	*	86,60	*	*	52,06	47,90			
					AG = 68,02%		Kappa = 38,45%				

\* classe não ocorrente no mapa estimado.

Comparando com os métodos CEU, teve-se sensível aumento nas AG (Bayes com 66,56% e RLMM com 67,01%) e aumento significativo no índice Kappa (Bayes com 35,85% e RLMM com 32,81%).

Com a simplificação da legenda, observou-se sensível melhora no mapa, como apresenta a Figura 25.



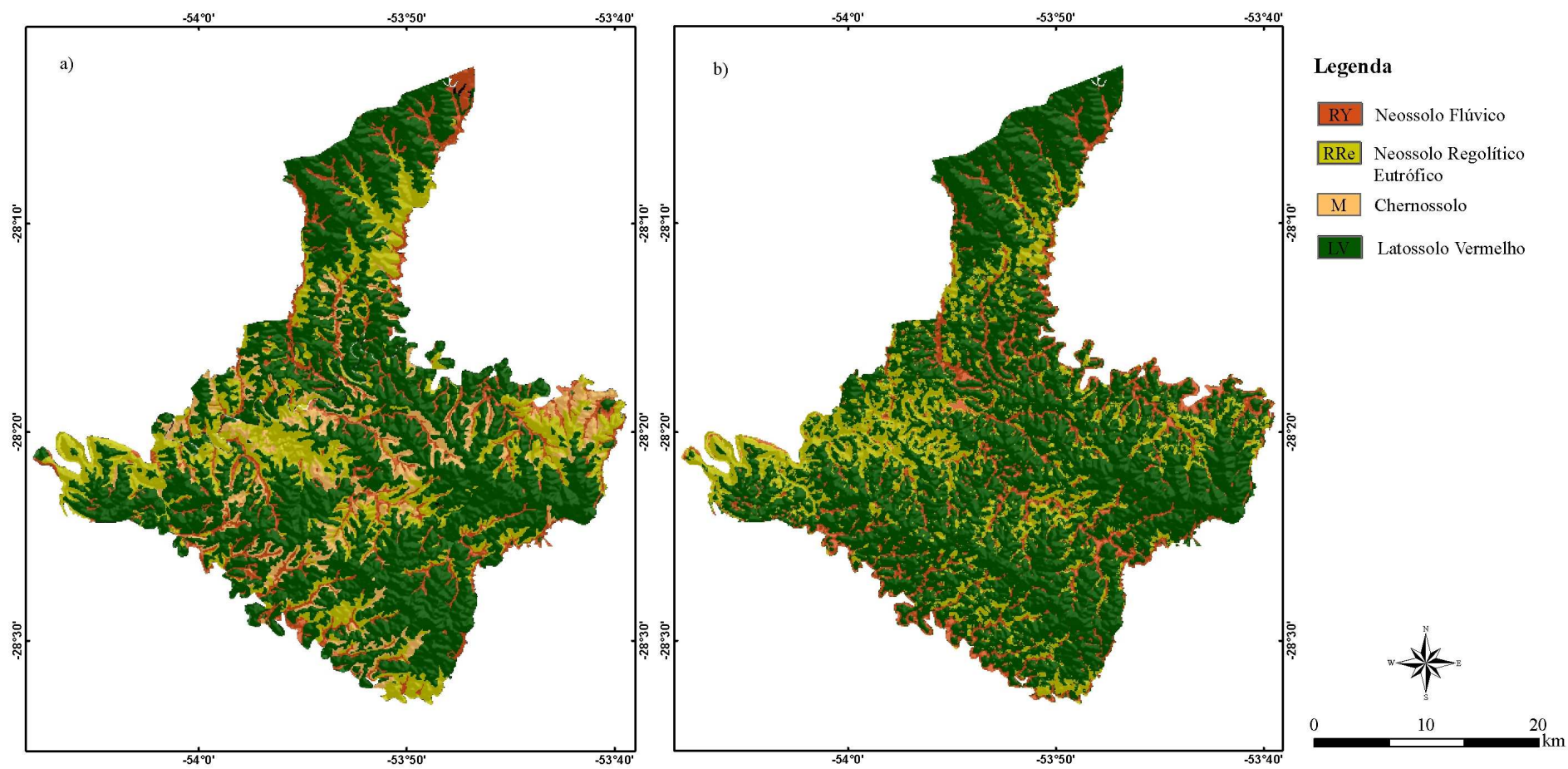


Figura 25 - Mapa de solos com legenda simplificada (a) e mapa de solos produzido pelo CDA CART com legenda simplificada (b)

O mapa produzido pelo método CART, com legenda simplificada, apresentou os seguintes resultados (Tabela 9): AG de 68,37% e Kappa de 38,82%, a classe RY teve 52,06% de AP e 48,17% de AU, a classe RRE apresentou 47,90% e 54,38% de AP e AU respectivamente, a classe LV teve 86,67% de AP e 75,13% de AU; a classe M não foi mapeada.

Tabela 09 - Matriz de erro do mapa produzido pelo CDA CART e o mapa de solos, ambos com legenda simplificada

		MAPA COM LEGENDA SIMPLIFICADA						
		Classe	RY	RRE	M	LV	Total	AU (%)
MAPA PRODUZIDO	RY		5,73	1,74	0,83	3,59	11,89	48,17
	RRE		1,62	9,33	1,61	4,60	17,16	54,38
	M		0,00	0,00	0,00	0,00	0,00	*
	LV		3,66	8,41	5,57	53,31	70,95	75,13
CART	Total		11,00	19,49	8,01	61,50	100,00	
	AP (%)		52,06	47,90	*	86,67		
			AG = 68,37%		Kappa = 38,82%			

\* classe não ocorrente no mapa estimado.

## 6.2.2. CLASSIFICADOR J48 (C4.5)

A árvore de decisão gerada pelo algoritmo J48 (Figura 26) apresentou 14 nós terminais, 7 níveis hierárquicos e teve tamanho 26, sendo ligeiramente menor que a árvore gerada pelo método CART.

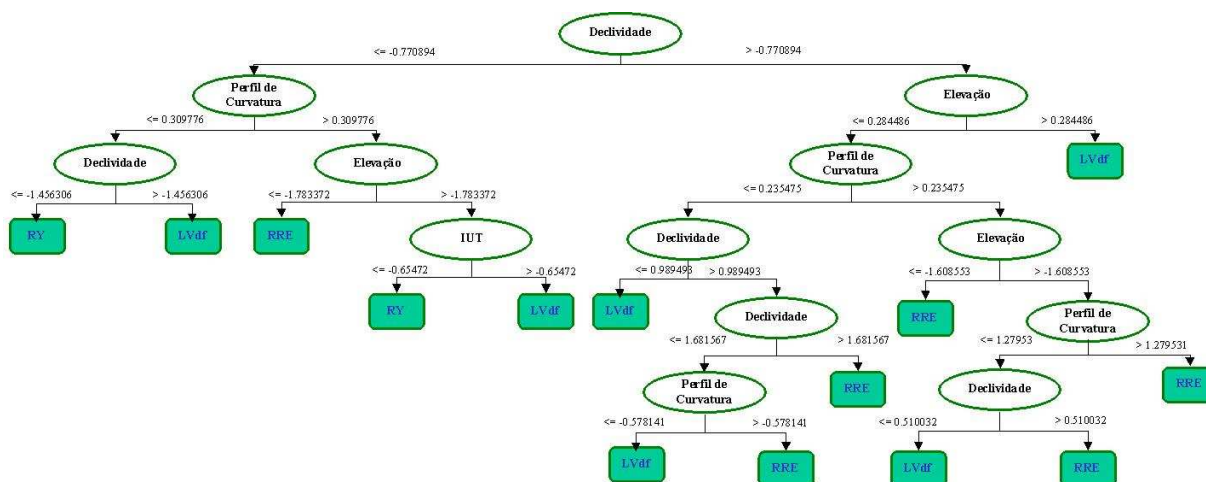


Figura 26 - Árvore de decisão do CDA J48



O algoritmo J48, que tem a mesma estrutura de resposta do CART (árvore de decisão cujos nós finais são as classes), apresentou resultados próximos a este com 67,31% de AG e 38,60% de índice Kappa; a classe LVdf resultou nas melhores AU (75,30%) e AP (83,98%); a classe RY teve 46,30% de AU e 56,49% de AP e a classe RRE apresentou 53,01% e 49,99% de AU e AP respectivamente. Como nos demais classificadores testados, até então, as classes MTo, MEo, LVdh e LVdt não foram mapeadas. Os resultados comparativos entre o mapa produzido e original podem ser visualizados na Tabela 10.

Tabela 10 - Matriz de erro do mapa produzido pelo CDA J48 e o mapa original de solos

		MAPA ORIGINAL									
		Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)
MAPA PRODUZIDO  J48	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdf	5,32	0,02	51,36	0,27	0,08	3,32	7,84	68,20	75,30	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
	RY	1,00	0,01	4,29	0,00	0,00	6,21	1,90	13,42	46,30	
	RRE	1,64	0,02	5,50	0,00	0,00	1,47	9,74	18,38	53,01	
	Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00		
	AP (%)	*	*	83,98	*	*	56,49	49,99			
	AG = 67,31%					Kappa = 38,60%					

\* classe não ocorrente no mapa estimado.

A Figura 27 apresenta o resultado da implementação da árvore em ambiente ArcGis 9.2. A relação entre o mapa produzido e o estimado apresenta, visualmente, semelhança principalmente nas classes de solo com maior área de ocorrência.

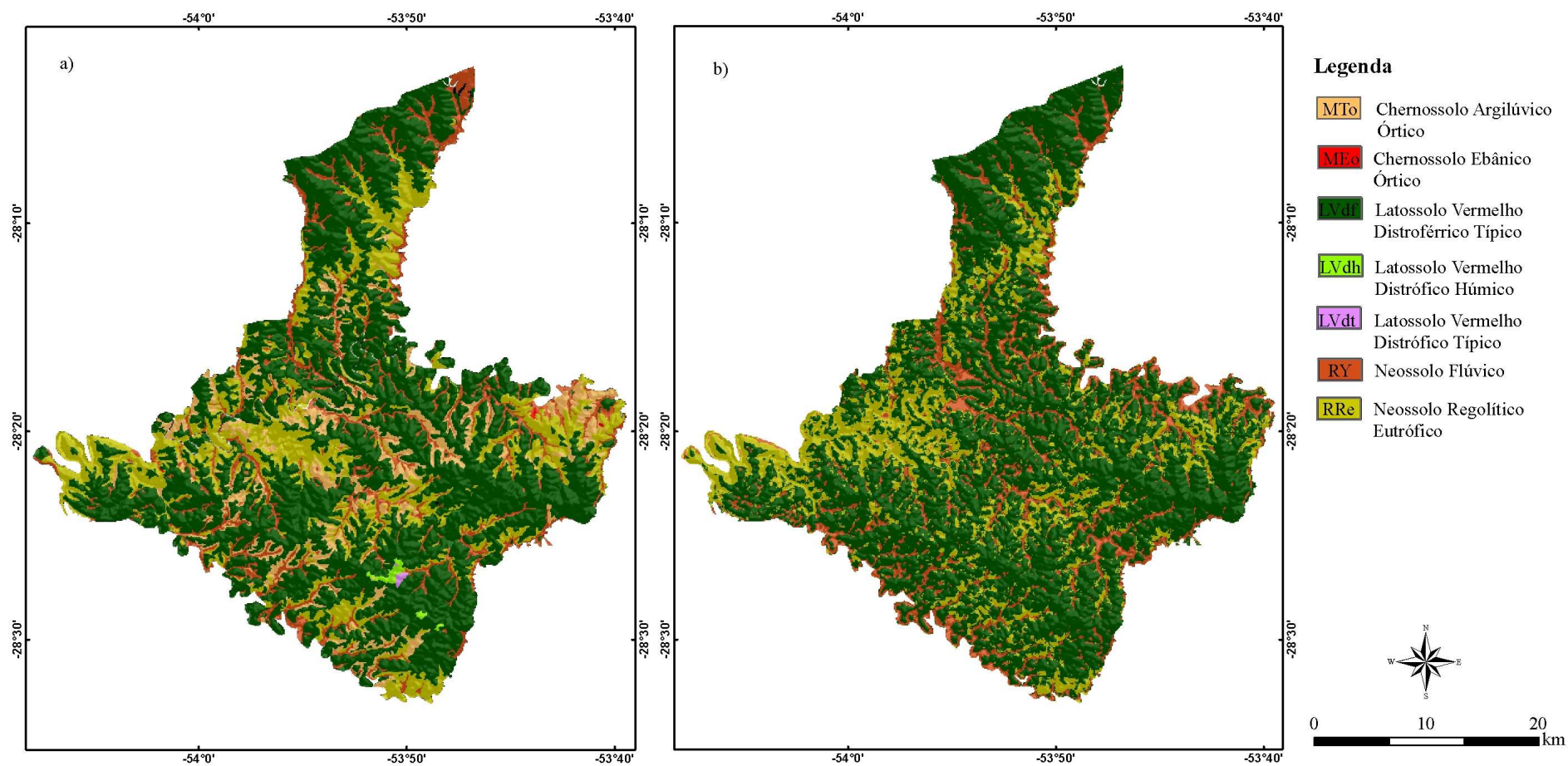


Figura 27 - Mapa de solos original (a) e mapa produzido pelo CDA J48 (b)

A simplificação da legenda gerou sensível aumento na semelhança entre os mapas, conforme apresenta a Tabela 11 e a Figura 28. O método J48, com simplificação da legenda, resultou em 67,66% de AG e o maior índice Kappa com 38,97%. A classe RY teve 56,49% de AP e 46,29% de AU, a classe RRE apresentou 49,99% e 58,56% de AP e AU respectivamente, a classe LV apresentou 84,06% de AP e 75,81% de AU. A classe M também não foi mapeada pelo J48.

Tabela 11 - Matriz de erro do mapa produzido por J48 e o mapa de solos, com legenda simplificada

		MAPA COM LEGENDA SIMPLIFICADA						
		Classe	RY	RRE	M	LV	Total	AU (%)
MAPA PRODUZIDO  J48	RY		6,21	1,90	1,01	4,30	13,42	46,29
	RRE		1,47	9,74	1,66	5,50	18,38	53,01
	M		0,00	0,00	0,00	0,00	0,00	*
	LV		3,32	7,84	5,34	51,70	68,20	75,81
	Total		11,00	19,49	8,01	61,50	100,00	
	AP(%)		56,49	49,99	*	84,06		
			AG = 67,66%			Kappa = 38,97%		

\* classe não ocorrente no mapa estimado.

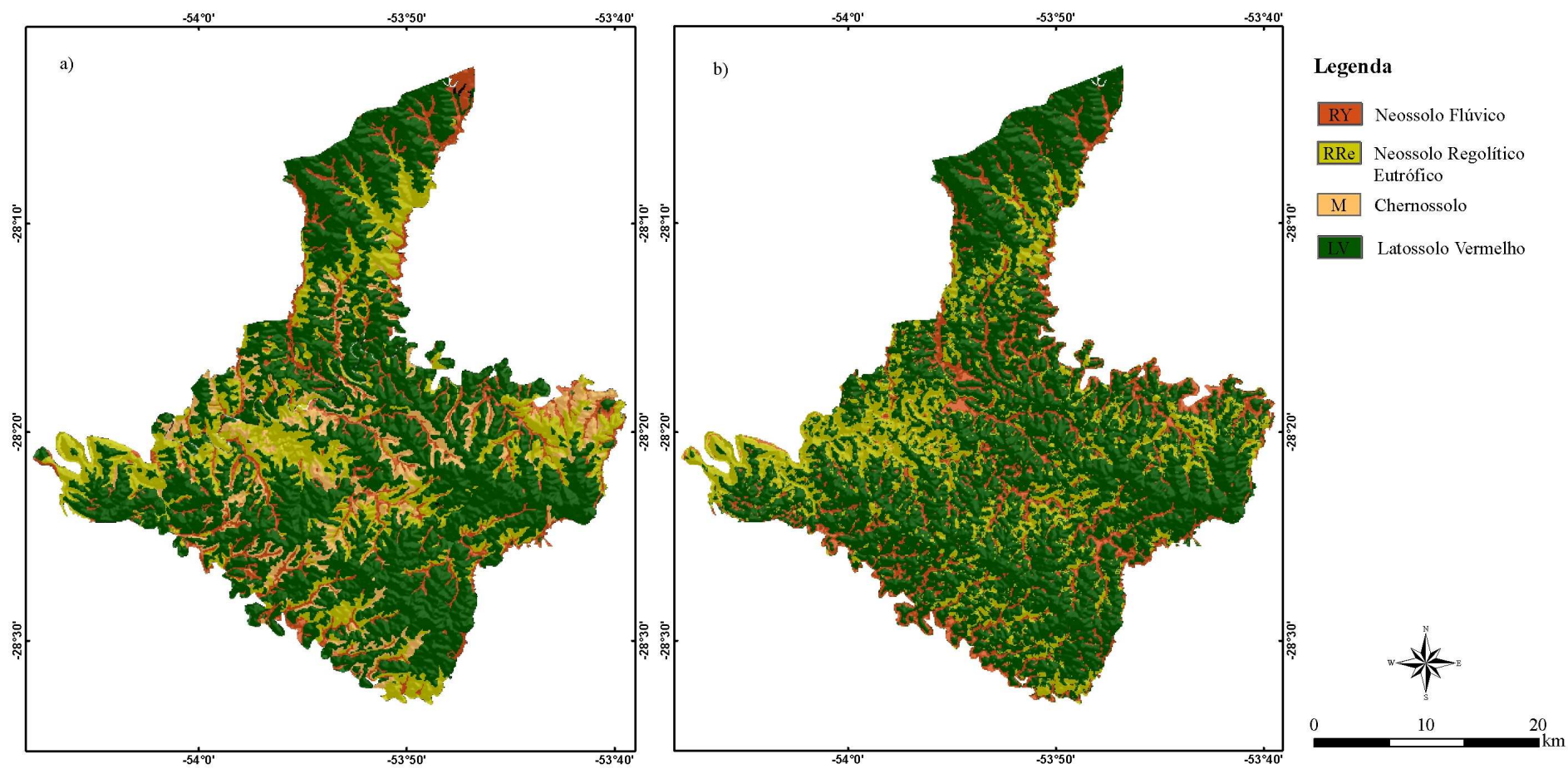


Figura 28 - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pelo CDA J48 com legenda simplificada (b)

### 6.2.3. LOGISTIC MODEL TREES (LMT)

No modelo logístico de decisão em árvore, diferentemente dos métodos CART e J48, os nós finais não representam classe única, mas sim modelos de probabilidade para as classes (LM – *logistic model*). Com 7 nós terminais, os quais resultam nos modelos logísticos de probabilidade (LM<sub>1</sub>,...LM<sub>7</sub>), 5 níveis hierárquicos e tamanho 12 (sendo os 12 passos para a finalização do processo) a árvore resultante do método LMT (Figura 29) foi significativamente menor que as árvores oriundas dos métodos CART e J48.

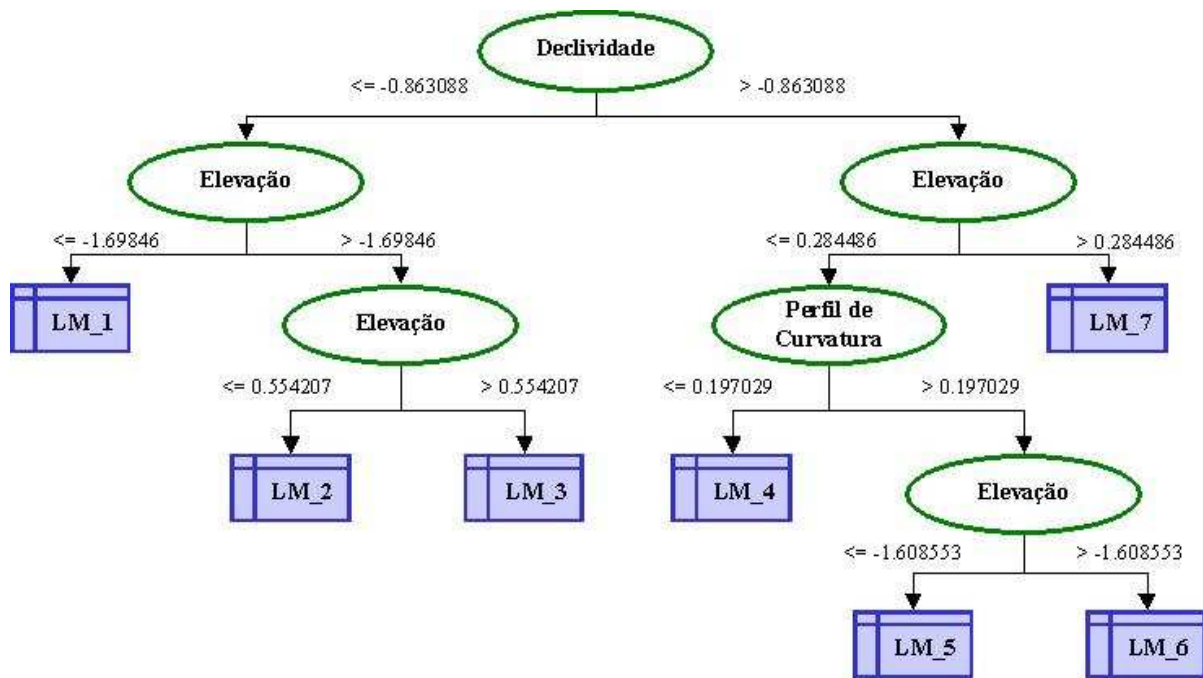


Figura 29 - Árvore de decisão resultante do método LMT

Os parâmetros e coeficientes dos modelos logísticos resultantes nos nós terminais da árvore (LM) foram agrupados na Tabela 12.

Tabela 12 - Modelos logísticos resultandos nos nós terminais da árvore de decisão do LMT

<i>LOGISTIC MODEL 1 (LM_1)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	0.83	2.09	1.75	-3.94	-30.45	-8.10	-29.42
Elevação	0.49	0.34	1.71	-2.46	-1.09	2.03	2.50
IUT	0.03	-0.08	-0.45	-0.14	0.40	-3.93	0.67
Declividade	-0.07	-0.28	-1.21	-0.13	0.34	9.69	0.12
Curvatura Planar	-0.84	0.05	0.05	0.10	-1.00	1.62	0.00
Perfil de Curvatura	-0.38	-0.96	0.05	0.57	-0.15	-2.31	0.09
<i>LOGISTIC MODEL 2 (LM_2)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	-0.35	2.16	0.11	-0.64	-42.45	-5.69	-41.42
Elevação	-0.04	0.11	0.28	0.14	-1.09	2.18	2.50
IUT	-0.22	0.34	-0.81	0.40	0.40	-5.62	0.67
Declividade	0.06	-0.35	-0.84	-0.59	0.34	14.02	0.12
Curvatura Planar	0.07	-0.22	0.08	0.02	-1.00	2.13	0.00
Perfil de Curvatura	-0.08	-0.89	-0.16	0.55	-0.15	-3.21	0.09
<i>LOGISTIC MODEL 3 (LM_3)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	-12.53	-6.61	-6.21	-12.82	-42.45	6.70	-41.42
Elevação	-0.14	-3.49	5.61	0.06	-1.09	-8.07	2.50
IUT	-0.22	0.35	-0.82	0.38	0.40	-4.42	0.67
Declividade	0.04	-14.40	1.53	-0.60	0.34	3.74	0.12
Curvatura Planar	0.06	-0.20	-3.46	0.02	-1.00	2.13	0.00
Perfil de Curvatura	-0.10	-1.62	0.52	0.45	-0.15	-3.23	0.09
<i>LOGISTIC MODEL 4 (LM_4)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	0.07	2.35	-1.28	0.10	-6.43	-27.79	-29.38
Elevação	-0.21	0.22	-0.26	-0.17	-0.59	2.03	2.62
IUT	-0.12	-0.15	-0.16	0.19	0.07	1.10	0.64
Declividade	0.15	-0.49	-0.57	0.94	0.61	-1.17	0.12
Curvatura Planar	0.00	0.04	-0.25	-0.29	-0.61	-0.28	0.10
Perfil de Curvatura	0.07	-0.80	1.42	-0.06	-0.80	-0.60	0.23
<i>LOGISTIC MODEL 5 (LM_5)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	4.42	-1.09	-1.22	-0.76	-11.84	-39.79	-41.38
Elevação	1.48	-1.48	-0.18	-1.39	-4.58	2.03	2.62
IUT	-0.15	-0.01	-0.18	0.12	1.40	1.10	0.64
Declividade	-1.02	-0.28	0.38	0.87	-3.37	-1.17	0.12
Curvatura Planar	-0.55	0.06	0.52	-0.08	-2.80	-0.28	0.10
Perfil de Curvatura	-2.95	-1.27	0.73	0.41	-2.49	-0.60	0.23
<i>LOGISTIC MODEL 6 (LM_6)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	0.49	2.06	-0.94	0.45	-6.42	-39.79	-41.38
Elevação	-0.30	0.15	-0.44	-0.09	2.61	2.03	2.62
IUT	-0.15	0.04	-0.62	0.10	0.64	1.10	0.64
Declividade	-0.06	-0.35	-0.43	0.79	0.94	-1.17	0.12
Curvatura Planar	-0.09	-0.09	0.32	-0.15	-1.66	-0.28	0.10

Perfil de Curvatura	-0.58	-0.92	0.42	0.27	-0.64	-0.60	0.23
<i>LOGISTIC MODEL 7 (LM_7)</i>							
Variável	MTo	LVdf	RY	RRE	MEo	LVdh	LVdt
Constante	0.78	2.82	-2.00	1.56	-18.77	-1.26	-3.13
Elevação	-4.36	0.07	2.47	-5.55	-1.50	-0.73	0.34
IUT	-0.11	-0.22	-0.85	0.42	0.48	1.18	0.54
Declividade	0.55	-0.54	-0.63	1.05	0.20	-1.53	0.01
Curvatura Planar	-0.18	0.02	0.47	-0.30	-1.14	-0.41	0.07
Perfil de Curvatura	-0.23	-0.65	0.63	0.31	-0.29	-0.66	0.27

Após a implementação da árvore, conforme a Tabela 13, o algoritmo LMT apresentou AG de 68,38% e índice Kappa de 37,66%; a classe que apresentou as melhores acurácias foi a LVdf com 73,42% de AU e 88,61% de AP; a classe RY resultou em 48,79% de AU e 50,01% de AP e a classe RRE 58,56% de AU e 44,51% de AP. Mesmo com baixos valores, o algoritmo LMT foi o único dentre os cinco métodos testados que mapeou a classe MTo a qual apresentou 15,73% de AU e apenas 0,19% de AP; as classes MEo, LVdh e LVdt também não foram mapeadas pelo método LMT. O mapa produzido é apresentado na Figura 30.

Tabela 13 - Matriz de erro do mapa produzido pelo classificador LMT e o mapa original de solos

		MAPA ORIGINAL									
		Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)
MAPA PRODUZIDO  LMT		MTo	0,01	0,00	0,04	0,00	0,00	0,00	0,04	0,10	15,73
		MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		LVdf	5,88	0,02	54,19	0,26	0,08	4,33	9,04	73,82	73,42
		LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*
		RY	0,78	0,00	3,25	0,00	0,00	5,50	1,74	11,28	48,79
		RRE	1,28	0,01	3,67	0,00	0,00	1,16	8,67	14,81	58,56
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
		AP (%)	0,19	*	88,61	*	*	50,01	44,51		
			AG = 68,38%				Kappa = 37,66%				

\* classe não ocorrente no mapa estimado.



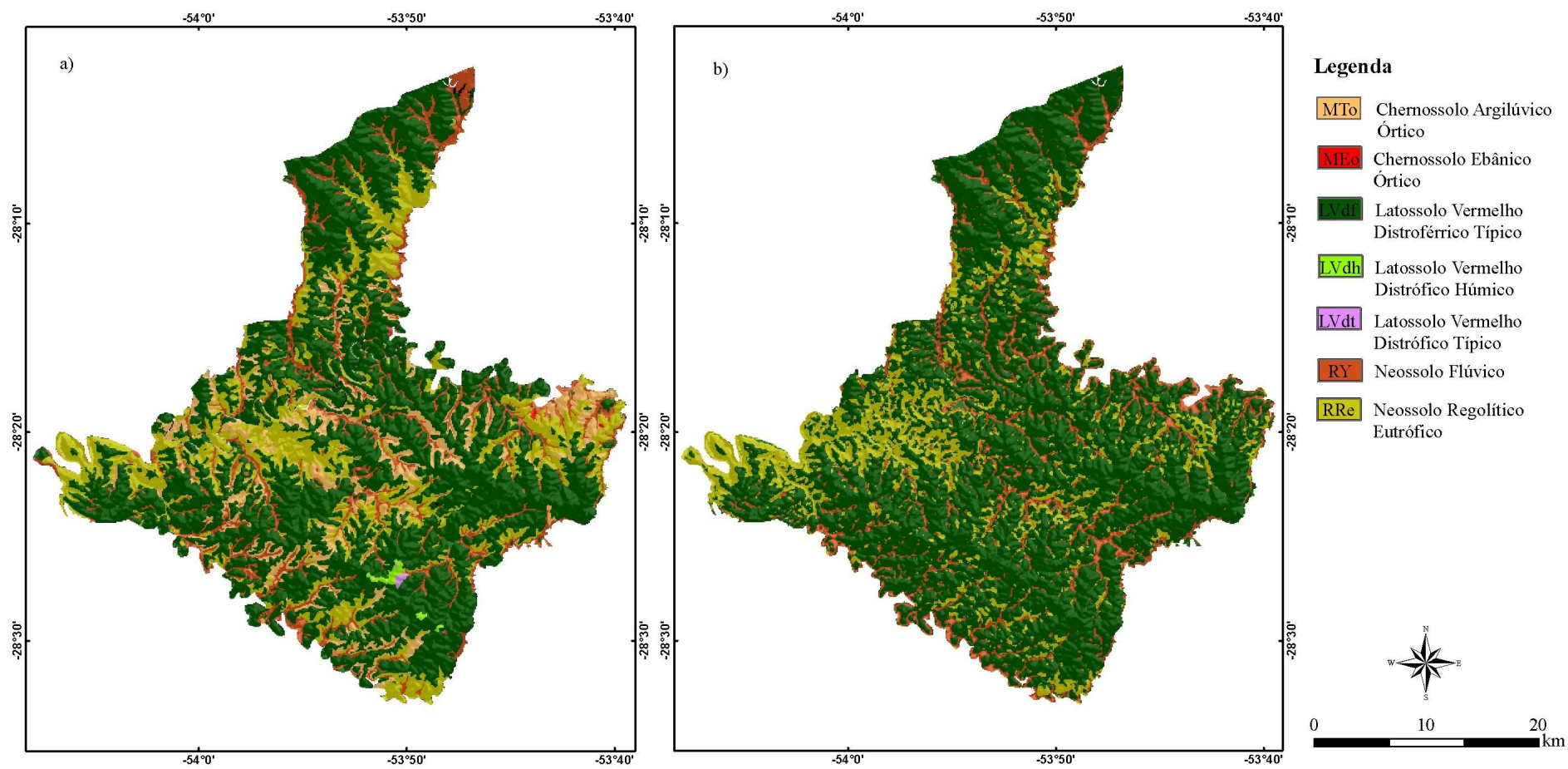


Figura 30 - Mapa de solos original (a) e mapa produzido pelo CDA LMT (b)



A simplificação da legenda, conforme a matriz de erros (Tabela 14) resultou em AG de 68,72% ( a maior obtida entre todos os métodos) e Kappa de 38,02%. A classe RY teve 50,01% e 48,77% de AP e AU, a classe RRE apresentou 44,51% de AP e 58,56% de AU, a classe LV teve 88,66% de AP e 73,88% de AU; o método LMT foi o único que mapeou a classe M e resultou em 0,19% de AP e 15,56% de AU.

Tabela 14 - Matriz de erro do mapa produzido por LMT e o mapa de solos, ambos com legenda simplificada

		MAPA COM LEGENDA SIMPLIFICADA						
		Classe	RY	RRE	M	LV	Total	AU (%)
MAPA PRODUZIDO	RY		5,50	1,74	0,79	3,25	11,28	48,77
	RRE		1,16	8,67	1,30	3,68	14,81	58,56
	M		0,00	0,04	0,01	0,04	0,10	15,56
	LV		4,33	9,04	5,91	54,53	73,82	73,88
	Total		11,00	19,49	8,01	61,50	100,00	
LMT	AP(%)		50,01	44,51	0,19	88,66		
			AG = 68,72%			Kappa = 38,02%		

\* classe não ocorrente no mapa estimado.

A Figura 31 apresenta a comparação entre o mapa produzido pelo modelo logístico de decisão em árvore e o mapa de solos com legenda simplificada.

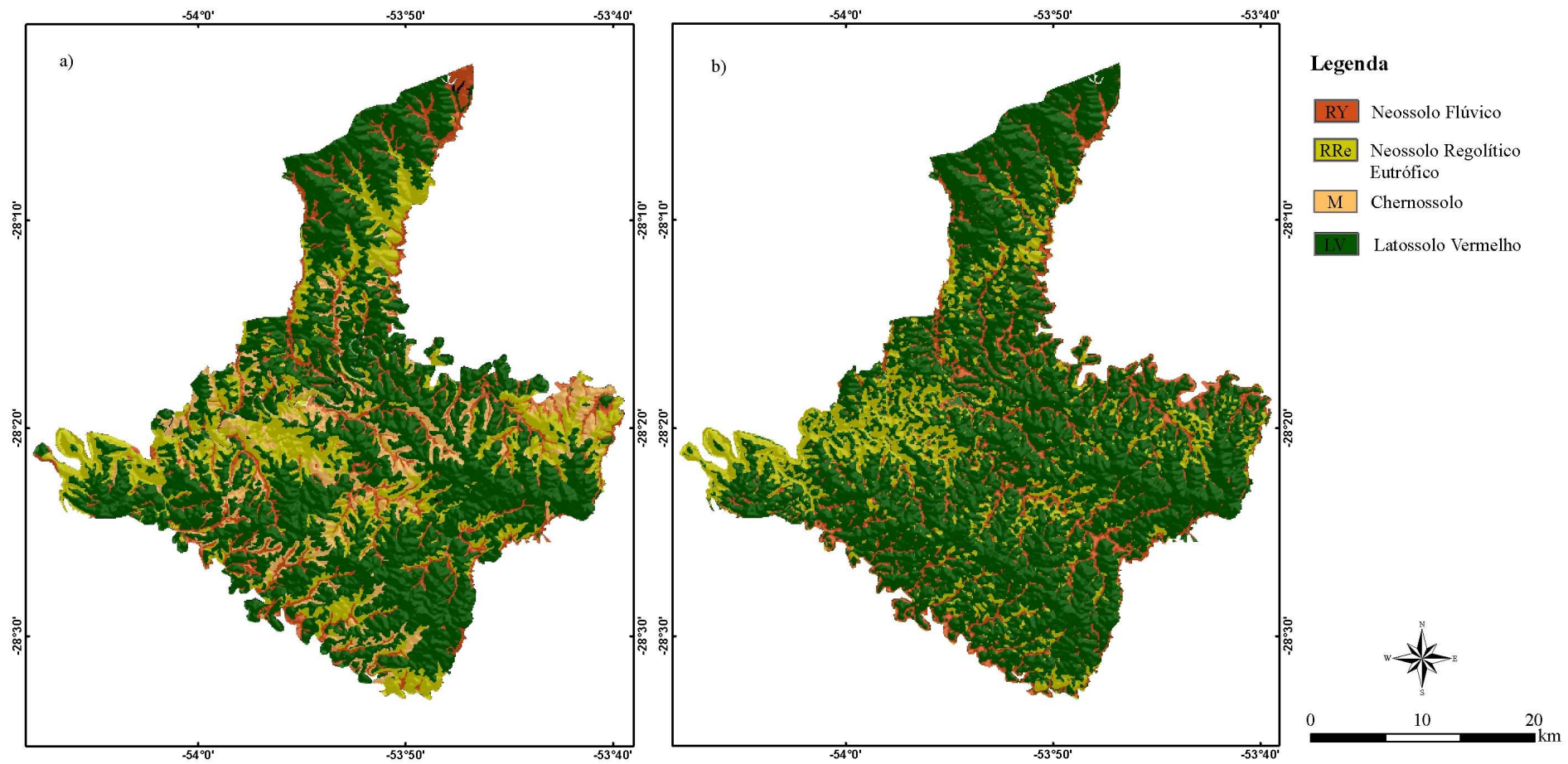


Figura 31 - Mapa de solos com legenda simplificada (a) e mapa de solo produzido pelo CDA LMT com legenda simplificada (b)

### 6.3. ANÁLISE DOS RESULTADOS

Embora os resultados obtidos não tenham apresentados altas acurácias e maior semelhança dos mapas produzidos com o mapa base, resultados de mapeamento digital de solos com a mesma grandeza de valores são apresentados em trabalhos de HENGL & ROSSITER (2003), FIGUEIREDO (2006), TEN CATEN et al. (2009) e SARMENTO et al (2009).

O motivo do não mapeamento das classes MEO, LVdh e LVdt por nenhum algoritmo e a classe MTO ser mapeada somente pelo método LMT, pode ser explicado pela proporção de pixels ou a menor área de ocorrência das mesmas. A ocorrência de pixels da classe MTO representa somente 7,96% do total, ou seja, sua área é relativa a 7,96% da área total, já as classes MEO, LVdh e LVdt representam, cada uma, menos que 0,5% da área total, com respectivamente 0,04%, 0,27% e 0,08%. Somando-se a área dessas três últimas classes temos apenas 0,39% da área total. Contrastando esses dados, temos a classe LVdf com 61,16%, a classe RRE 19,49% e RY com 11% da área total.

Afim de comparar quantitativamente os métodos, foi elaborada a Tabela 15 que apresenta, de forma agrupada, os valores de AG e índice Kappa dos mapas produzidos com legenda original e simplificada.

Tabela 15 - Síntese das AG e índice Kappa dos mapas produzidos com legenda original e simplificada

MÉTODOS		ACURÁCIAS					Média Kappa
		AG (%)			Kappa (%)		
		Original	Simplificada	Média AG	Original	Simplificada	
CEU	Logística	67,01	67,36	66,96	32,81	33,14	34,49
	Bayes	66,56	66,89		35,85	36,17	
CDA	CART	68,02	68,37	68,08	38,45	38,82	38,42
	J48	67,31	67,66		38,60	38,97	
	LMT	68,38	68,72		37,66	38,02	

Observa-se que as maiores acurácias ocorreram nos mapas com legenda simplificada com utilização dos CDA; nota-se também que há sensível melhora

nas AG dos CDA em relação aos CEU, com aumento de pouco mais de 1%, ao passo que se observou significativa melhora nos índices Kappa dos CDA, que teve uma média geral de 38,42% em relação aos CEU, com média geral de 34,49%.

Destacam-se as maiores AG obtidas pelo método LMT com 68,38% no mapa estimado com legenda original e 68,72% no mapa com legenda simplificada, porém esse método apresentou os menores índices Kappa entre os CDA, com 37,66% no mapa com legenda original e 38,02% no mapa com legenda simplificada. Outro destaque é o método J48 que apresentou os melhores Kappa com 38,60% no mapa com legenda original e 38,97% no mapa com legenda simplificada, porém apresentou as menores AG entre os CDA, com 67,31% no mapa com legenda original e 67,66% no mapa com legenda simplificada; já o método CART foi o que apresentou, entre os CDA, os valores intermediários de AG, com 68,02% no mapa com legenda original e 68,37% no mapa com legenda simplificada, e também os valores intermediários entre os índices Kappa com 38,45% no mapa com legenda original e 38,82% no mapa com legenda simplificada.

Os métodos de CEU apresentaram as acurácias mais baixas. O destaque é o método logístico que apresentou índices Kappa muito abaixo dos outros métodos com apenas 32,81% no mapa com legenda original e 33,14% no mapa com legenda simplificada, porém teve as AG sensivelmente melhores que as obtidas pelo método de Bayes com 67,01% no mapa com legenda original e 67,36 no mapa com legenda simplificada ao passo que Bayes teve suas AG de 66,56% e 66,89% nos mapas com legenda original e simplificada respectivamente; o método de Bayes teve Kappa com valores mais próximos aos dos CDA com 35,85% no mapa com legenda original e 36,17% no mapa com legenda simplificada.

A Tabela 16 apresenta a comparação entre as AU e AP dos diferentes classificadores para cada classe dos mapas produzidos com legenda original e os mapas produzidos com legenda simplificada.

Tabela 16 - Síntese das AU e AP dos mapas produzidos por legenda original e por legenda simplificada

MAPA COM LEGENDA ORIGINAL							
Classe	Área (Km <sup>2</sup> )	MÉTODOS / ACURÁCIAS	RLMM	Bayes	CART	J48	LMT
MTo	80,97	AU (%)	*	*	*	*	15,73
		AP (%)	*	*	*	*	0,19
MEo	0,44	AU (%)	*	*	*	*	*
		AP (%)	*	*	*	*	*
LVdf	622,84	AU (%)	71,54	74,04	74,64	75,30	73,42
		AP (%)	90,45	86,19	86,60	83,98	88,61
LVdh	2,76	AU (%)	*	*	*	*	*
		AP (%)	*	*	*	*	*
LVdt	0,82	AU (%)	*	*	*	*	*
		AP (%)	*	*	*	*	*
RY	111,93	AU (%)	44,83	41,22	48,18	46,30	48,79
		AP (%)	38,03	56,43	52,06	56,49	50,01
RRe	198,04	AU (%)	56,30	55,58	54,38	53,01	58,56
		AP (%)	38,59	39,22	47,90	49,99	44,51
Médias		AUm (%)	24,67	24,41	25,31	24,94	28,07
		APm (%)	23,87	25,98	26,65	27,21	26,19
MAPA COM LEGENDA SIMPLIFICADA							
Classe	Área (Km <sup>2</sup> )	MÉTODOS / ACURÁCIAS	RLMM	Bayes	CART	J48	LMT
RY	111,93	AU (%)	44,82	41,22	48,17	46,29	48,77
		AP (%)	38,03	56,43	52,06	56,49	50,01
RRe	198,04	AU (%)	56,30	55,57	54,38	53,01	58,56
		AP (%)	38,59	39,22	47,90	49,99	44,51
M	81,41	AU (%)	*	*	*	*	15,56
		AP (%)	*	*	*	*	0,19
LV	626,42	AU (%)	71,99	74,50	75,13	75,81	73,88
		AP (%)	90,49	86,23	86,67	84,06	88,66
Médias		AUm (%)	43,28	42,82	44,42	43,78	49,19
		APm (%)	41,78	45,47	46,66	47,64	45,84

\* classe não ocorrente no mapa estimado

Destacam-se os significativos aumentos de valores nas médias das AU (AUm) e nas médias das AP (APm) dos mapas produzidos com legenda simplificada, com valores na ordem de 40% em relação aos mapas produzidos com legenda original com valores na ordem de 20%. Observa-se que o método LMT apresentou, em ambos mapeamentos, as maiores AU. O método LMT também se diferenciou por ser o único que mapeou as classes MTo, no mapeamento com legenda original, e a classe M, no mapeamento com legenda simplificada. O método RLMM se destaca por apresentar as AP mais altas de 91,45% para a classe LVdf, no mapeamento com legenda original, e

90,49% para a classe LV, no mapeamento com legenda simplificada. As classes LVdf e LV tiveram as maiores AP e AU em todos os métodos.

O estudo de HENGL & ROSSITER (2003) apresentou resultado de 45,3% de acurácia global e 42,6% no índice Kappa para mapeamento digital de solos no Nordeste da Croácia com a área de estudo compreendendo 1.062km<sup>2</sup>. O estudo de FIGUEIREDO (2006), comparou o resultado do mapeamento digital de solos em três diferentes regiões do Rio Grande do Sul (Brasil). A primeira área, no município de Ibirubá, o autor obteve 61,4% de acurácia global e 38,3% de índice Kappa, porém a área de estudo correspondia a 720km<sup>2</sup>. Para a segunda área, correspondente à bacia hidrográfica do Arroio Portão, os resultados obtidos ficaram em 67,7% na acurácia global e 50,3% no índice Kappa, porém a área de estudo correspondia a apenas 224 km<sup>2</sup>. A terceira área, correspondendo parte do Município de Viamão, o autor obteve 40,8% de acurácia global e 31,8% no índice Kappa. A área foi correspondente a 1.000km<sup>2</sup>.

O estudo de TEN CATEN et al (2009) chegou a apresentar 74,30% de acurácia global e 58,20% de índice Kappa no mapa predito, com utilização de análise de componentes principais. A área de estudo compreendeu 874km<sup>2</sup>. Já o estudo de SARMENTO et al (2009) chegou a apresentar 71,29% de acurácia global e 58,13% de índice Kappa com utilização de CDA, porém a área compreendida foi de apenas 6,735 km<sup>2</sup>.

É possível observar, na literatura citada e no presente trabalho, a relação do valor das acurácias com a área mapeada. Em geral, observa-se que quanto maior a área, menor o poder de discriminação do modelo, resultando em acurácias mais baixas. Isso pode ser explicado pela confusão e mistura dos valores das variáveis geomorfométricas para cada classe. Quanto maior a área, a tendência é ter maior mistura de valores para uma mesma classe. O mesmo ocorre quando existe um grande número de variáveis. Em relação ao número de variáveis, o estudo de TEN CATEN (2009), o qual realizou mapeamento digital de solos a partir de componentes principais, cujo objetivo é reduzir as variáveis sem grande perda de informação, apresentou maiores acurácias.

Um dos fatores que podem ter contribuído para a redução das acurácias foi a utilização do MDE do SRTM pois toda aquisição dos dados SRTM foi planejada a suprir mapeamentos em escalas entre 1:100.000 e 1:250.000 (VALERIANO, 2008) e o mapa de solos, utilizado como referência, possui escala de

1:50.000. A utilização de um MDE grosseiro pode ocasionar perda de detalhamento e maior mistura dos valores das variáveis para cada classe de solo, o que pode aumentar a probabilidade de erro do classificador.

## 7. CONCLUSÕES

Os resultados mostram que os CDA, como já esperado e citado em literatura, têm maior poder de discriminação que os CEU também para mapeamento supervisionado de solos com variáveis geomorfométricas.

Os métodos de CDA apresentaram sensíveis melhoras nas AG e significativas melhoras nos índices Kappa e nas AU e AP. O método CDA J48 foi aquele que apresentou os maiores índices Kappa, porém as menores AG dentre os CDA, ao passo que o método LMT apresentou as melhores AG e os menores índices Kappa dentre os CDA; o método CART foi o CDA que apresentou resultados intermediários de AG e Kappa.

Dentre todos os métodos CEU, o método RLMM foi aquele que teve os menores índices Kappa, porém, obteve melhores AG que o método de Bayes, o qual é um método considerado ótimo, que teve índice Kappa mais próximo aos CDA, porém as menores AG entre todos os métodos.

A utilização de legenda simplificada não provocou grandes ganhos de AG e Kappa, porém proporcionou significativa melhora nas AU e AP, como foi apresentado nos mapeamentos com legenda original onde não se observou valor superior a 28,10%, para esses índices, ao passo que com a utilização de legenda simplificada o valor não baixou de 41,70% e chegou até aos 49,19%.

Os melhores métodos para esta área de estudo não foram os tradicionais métodos de CEU RLMM e nem o classificador de Bayes, porém foram os CDA, os quais obtiveram valores de acurácias próximos entre si, com a utilização de legenda simplificada, o que proporciona maior confiança com o significativo aumento nas AU e AP.

A contribuição do trabalho para o futuro do mapeamento digital de solos foi a de apresentar diferentes métodos matemáticos e que, para uma mesma área, diferentes métodos de CEU apresentam praticamente a mesma resposta, bem como os métodos CDA, e que o último apresenta melhores resultados. Os valores obtidos foram satisfatórios levando em consideração a área mapeada, a qual passou dos 1.000 km<sup>2</sup>. Cabe ainda futuros estudos para testar os mesmos métodos e/ ou outros métodos, nessa mesma área, subdividindo-a em áreas menores, bem como utilizar MDEs mais refinados



e analisar os resultados com os obtidos na totalidade da área. Ainda é possível almejar estudos que utilizem os modelos e extrapolem para áreas adjacentes da utilizada no presente estudo e que tenham alguma informação referencial, na escala 1:50.000, para testar o potencial do modelo para predição em áreas sem amostras de treinamento, somente amostras de teste.

## 8. REFERÊNCIAS

ABRÃO, P. U. R. & AZOLIN, M. A. D. **Levantamento e utilização agrícola dos solos no Município de Ijuí**. Relatório Técnico. INCRA. 1970.

AGRESTI, A. **Categorical data analysis**. New York: John Wiley & Sons, 1990. 558p.

AGTERBERG, F. P. **Combining indicator patterns in weights of evidence modeling for resource evaluation**. *Nonrenewable Resources* 1 (1), 39-50. 1992.

ARBESÚ, R. A. & FELICISSIMO, A. M. **GIS and logistic regression as tools for environmental management: a coastal cliff vegetation model in Northern Spain**. In Brebia C (ed). *Management Information Systems 2002: GIS and Remote Sensing*. Wessex Institute of Technology: 215-224. 2002.

BAILEY, N.; CLEMENTS, T.; LEE, J. T.; THOMPSON, S. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. **Journal of Environmental Management**, v. 67, n. 4, p. 395-407, 2003.

BEHRENS, T.; FÖRSTER, H.; SCHOLTEN, T.; STEINRÜCKEN, U.; SPIES, E. D.; GOLDSCHMIDT, M. Digital soil mapping using artificial neural networks. **Journal of Plant Nutrition and Soil Science**, v. 168, n.1, p. 21-33, 2005.

BEHRENS, T. & SCHOLTEN, T. A comparison of data-mining techniques in predictive soil mapping. In: LAGACHERIE, P; MCBRATNEY, A.B.; VOLTZ, M (Eds), **Digital Soil Mapping, an introductory perspective**. *Developments in soil science*, vol. 31. Amsterdam: Elsevier, 2007 p. 353-364.

BORTOLON, E. S. O. **Simulação do impacto da agricultura sobre os estoques de carbono orgânico do solo e o balanço de CO<sub>2</sub> regionais integrando o modelo Century e o Sistema de Informações geográficas**. 2008. 161p. Tese (Doutorado em Ciência do Solo) - Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

BRASIL, Ministério da Agricultura. **Levantamento de reconhecimento dos solos do estado do Rio Grande do Sul**. Recife, 1973. 431p. (Boletim Técnico, 30).

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, R. A. **Classification and regression trees (CART)**. Belmont, CA: Wadsworth International, 1984. 358p.

BRIGGS, D. J. & SHISHIRA, E. K. **Soil variability in geomorphologically defined survey units in the Albudeite area of Murcia Province, Spain**. In: Jungerius, P.D.

(ed.). Soils and geomorphology (Catena supplement no 6). Braunschweig: Catena Verlag, 1985. cap. 2, p.69-84.

BURROUGH, P. A. **Principles of geographical information systems of land resources assesment**. Francis & Taylor, 1986. 185p.

BURROUGH, P. A. & MCDONELL, R. A. **Principles of Geographical Information Systems** (Oxford University Press, New York), p. 190. 1998.

CAMPLING, P.; GOBIN, A.; FEYEN, J. Logistic modeling to spatially predict the probability of soil drainage classes. **Soil Science Society of America Journal**, v.66, p. 1390-1401, 2002.

CHAGAS, C. da S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. 223 f. Tese (Doutorado em solos e nutrição de plantas) – Universidade Federal de Viçosa, Viçosa, 2006.

CHRISTOFOLETTI, A. **Geomorfologia**. 2. ed. São Paulo: Edgar Blüchler, 1980.

COHEN, J. A coeficient of agreement for nominals scales. **Journal of Educational and Measurement**, Washington, v.20, n.1, p.37-46, 1960.

CONGALTON, R. G. A review of assessing the accuracy of classification of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991.

COOK, S. E.; CORNER, R.; GREALISH, G. J.; GESSLER, P. E.; CHARTRES, C. J. A rule-based system to map soil properties. **Soil Science Society of America Journal**, Madison, v.60, p.1893–1900, 1996.

DALMOLIN, R. S. D.; KLAMT, E.; PEDRON, F. A.; AZEVEDO, A. C. Relação entre as características e o uso das informações de levantamentos de solos de diferentes escalas. **Ciência Rural**, Santa Maria, v.34, n.5, p. 1479-1486, 2004.

DALRYMPLE, J. B.; BLONG, R. J.; CONACHER, A. J. A hypothetical nine unit landsurface model. **Zeitschrift Für Geomorphologie**, v. 12, p. 60-76, 1968.

DOKOUCHEV V.V. **Le Chernozem Russe**. St. Pétersburg, 1883.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Procedimentos Normativos de Levantamentos Pedológicos**. Brasília, 1995. 116p.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. Rio de Janeiro, 1999. 412p.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 2.ed. Rio de Janeiro, 2006. 306p.

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - ESRI. **ArcGis 9.2**. Redland, California, 2006.

FIGUEIREDO, S. R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. Dissertação (Mestrado em Ciência do Solo) – Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2006.

FUKUNAGA, K. **Introduction to Statical Pattern Recognition**. Ed. Boston: Academic Press, 2<sup>nd</sup> edition, 1990.

GALLANT, J. C. & WILSON, J. P. **Primary topographic attributes**. In: WILSON, J. P. & GALLANT, J. C. (Eds.). *Terrain Analysis: Principles and applications*. New York: John Wiley & Sons, 2000. p.51-85.

GERRARD, A. J. **Soils and landforms**. 1.ed. London: George Allen & Unwin, 1981. 219p.

GIASSON, E.; CLARKE, R. T.; JUNIOR, A. V. I.; MERTEN, G. H.; TORNQUIST, C.G. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, v.63, n.3, p.262-268, 2006.

GONZALEZ, R. C. & WOODS, R. **Digital Image Processing**. Addison-Wesley Publishing Press, 2<sup>nd</sup> edition, 1992.

HALL, M.; FRANK, E.; HOLMES, G. PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. **The WEKA Data Mining Software: An Update**; SIGKDD Explorations, Volume 11, Issue 1. 2009.

HENGL, T. & ROSSITER, D.G. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. **Soil Science Society of America Journal**, Madison, v. 67, p. 1810–1822, 2003.

HILGARD, E. W.: **"Soils"** The Macmillan Company, New York, 1914.

IPPOLITI-RAMIRO, G. A.; COSTA, L. M.; SCHAEFER, C. E. G. R.; FERNANDES FILHO, E. I.; GAGGERO, M. R. Análise digital do terreno: ferramenta na identificação de pedoformas em microbacia na região de "Mar de Morros" (MG) **Revista Brasileira de Ciência do Solo**, Viçosa, v.29, n.2, p. 269-276, 2005.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. New York, London: McGraw-Hill, 1941, 324 p.

JENSON, S. K. & DOMINGUE, J. O. **Extracting topographic structure from digital elevation data for geographic information system analysis.** Photogramm. Eng. Remote Sens., 54(11), 1593-1600. 1988.

JOHN, G. H.; KOHAVI, R.; PFLEGER, K. Irrelevant features and the subset selection problem. In COHEN W. W. & HIRSH, H. **Machine Learning: Proceedings of the Eleventh International Conference.** San Francisco: Morgan Kaufmann, 1994. p.121-129.

KIRKBY, M. J. **Hydrography modeling strategies.** In: PEEL et al. (Ed). Progress in physical and human geography. London: Heinemann, 1975. p. 69-90.

LAGACHERIE, P. & MCBRATNEY, A. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie, P.; Mcbratney, A.; Voltz, M. (Ed.). **Digital soil mapping: an introductory perspective.** Amsterdam: Elsevier, 2007, p. 3-22.

LANDWEHR, N.; HALL, M.; FRANK, E. **Logistic Model Trees.** Machine Learning. 95 (1-2): 161-205. 2005.

LEPSCH, I.F. & BUOL, S.W. Investigations in an Oxisol-Ultisol toposequence in S.Paulo State, Brazil. **Soil Sci. Soc. Am. Proc.**, Madison, 38:491-496, 1974.

LEPSH, I. F. & BUOL, S. W. **Investigations in an Oxisol-Ultisol toposequence in Sao Paulo State, Brazil.** Soil Sci. Soc. Am. Proc., Madison, v.38, p.491-493, 1974

MCBRATNEY, A. B. A thoeretical framework for land evaluation. **Geoderma**, Amsterdam, v. 72, p. 198-199, 1996.

MCBRATNEY, A. B. & LAGACHERIE P.; VOLTZ, M. **Digital Soil Maping, An Introductory Perspective.** Developments in Soil Science. Volume 31. Amsterdam: Elsevier, 2007. 600p.

MCBRATNEY, A.B.; MENDONÇA SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, p. 3-52, 2003.

MILLER, F. P. Soil Science: **A scope broader than its identity.** Soil Sci. Soc. Am. J., 57:299 e 564, 1993

MOORE, I. D.; GEISSLER, P. E.; NIELSEN, G. A; PETERSON, G. A. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, v. 57, p. 443-452, 1993.

MORAN, C. J. & BUI, E. Spatial data mining for enhanced soil map modelling. **International Journal of Geographical Information Science**, London, v. 16, p. 533-549, 2002.

MORENO, J. A. **Clima do Rio Grande do Sul**. Porto Alegre: Governo do Estado do Rio Grande do Sul, Secretaria da Agricultura, 1961, 41p.

MÜLLER FILHO, I. L. **Notas para o Estudo da Geomorfologia do Rio Grande do Sul, Brasil**. Publicação Especial nº1. Santa Maria, Imprensa Universitária, Universidade Federal do Rio Grande do Sul, 1970.

OHLMACHER, G. C.; DAVIS, J. C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. **Engineering Geology**, v.69, p.331-343, 2003.

PINTO, L. F. S.; PAULETTO E. A.; GOMES A. S.; SOUZA R. O. **Caracterização de solos de várzea**. In: A. da S. Gomes & E.A. Pauletto, eds. Manejo do solo e da água em áreas de várzea. Embrapa, CPACT, Pelotas, 1999. p.11-36.

QUINLAN, R. **C4.5: Programs for Machine Learning**. Morgan Kauffman Publishers, San Mateo, CA. 1993.

RABUS, B.; EINEDER, M.; ROTH, A.; BAMLER, R. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborn radar. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 57, p. 241-262, 2003. doi: 10.1016/S0924-2716(02)00124-7

RAMANN, E.: Der Boden und sein geographischen Wert, **Mitteilungen der Geographischen Gesellschaft Milnchen**, 13 : 1-14, 1919.

ROSENFELD, A. **Image Processing and Recognition**. Technical Report 664. University of Maryland Computer Vision Laboratory, 1978..

SAFAVIAN, S. R. & LANDGREBE, D. A. “A survey of decision tree methodology”. IEEE Trans. Systems, Man and Cybernetics, vol 21, no. 3, 1991.

SARMENTO, E. C.; WEBER, E.; GIASSON, E.; FLORES, C. A.; HASENACK, H. **Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital do Vale dos Vinhedos, RS, Brasil**. In: Congresso Brasileiro de Ciência do Solo, 32. Anais. Fortaleza: SBCS, 2009. CD-ROM.

SIRTOLI, A. E. **Mapeamento de solos com auxílio da geologia, atributos do terreno e índices espectrais integrados por redes neurais artificiais**. 2008. 96f. Tese Doutorado em Geologia – Curso de Pós-Graduação em Geologia, Universidade Federal do Paraná.

STRECK, E. V.; KAMPV, N. DALMOLIN, R. S. D.; KLAMT, E.; NASCIMENTO, P. C.; SCHNEIDER, P. **Solos do Rio Grande do Sul**. Porto Alegre: EMATER/RS. UFRGS, 2002.

- TARBOTON, D. G. **Terrain Analysis Using Digital Elevation Models (TAUDEM)**. Utah State University, USA. 2008.
- TEIXEIRA, W.; TOLEDO, M. C. M.; FAIRCHILD, T. M.; TAIOLI, F. **Decifrando a Terra**. Universidade de São Paulo, São Paulo, Brasil, 558pp. 2000.
- TEN CATEN, A.; DALMOLIN, R. S. D.; RUIZ, L. F. C.; SEBEM, E.; PEREIRA, R. S. P. Mapeamento digital de solos através da aplicação de componentes principais em modelos logísticos. In: Simpósio Brasileiro de Sensoriamento Remoto, Natal, RN, 2009. **Anais XIV**. INPE:, 2009. p. 7677-7684.
- TERRIEN, C. W., **Decision Estimation and Classification, An Introduction to Pattern Recognition and Related Topics**. John Wiley & Sons, 1989.
- TUCCI, C. E. M. **Modelos Hidrológicos**. 2ª edição, Editora da UFRGS, Porto Alegre, 678 p. 1998.
- UNITED STATES DEPARTMENT OF AGRICULTURE - USDA, **Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys**. United States Department of Agriculture. Natural Resources Conservation Service, Agriculture Handbook, 1975.
- UNITED STATES DEPARTMENT OF AGRICULTURE - USDA, **Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys**. United States Department of Agriculture. Natural Resources Conservation Service, Agriculture Handbook, n 436 Second Edition, 1999.
- VALERIANO, M. M. **Topodata: guia para utilização de dados geomorfométricos locais**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2008.
- VALERIANO, M. M. & MORAES, J. F. L. **Extração de rede de drenagem e divisores por processamento digital de dados topográficos**. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10., 2001, Foz do Iguaçu, CD-ROM... Foz do Iguaçu, Arquivo 084.pdf. 8p.
- ZEVEBERGEN, L. W. & THORNE C. R. **Quantitative Analysis of Land Surface Topography**. Earth Surface Processes and Landforms 12: 47-56. 1987.