

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPUS DE JABOTICABAL**

**MAPEAMENTO DIGITAL DE SOLOS DA QUADRÍCULA DE
RIBEIRÃO PRETO - SP PELO MÉTODO RANDOM FOREST**

Matheus Felipe Oliveira

Geógrafo

JABOTICABAL – SÃO PAULO – BRASIL

2016

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**MAPEAMENTO DIGITAL DE SOLOS DA QUADRÍCULA DE
RIBEIRÃO PRETO - SP PELO MÉTODO RANDOM FOREST**

Matheus Felipe Oliveira

Orientador: Prof. Dr. José Eduardo Corá

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Agronomia (Ciência do Solo).

JABOTICABAL – SÃO PAULO – BRASIL

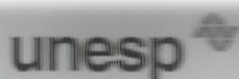
2016

Oliveira, Matheus Felipe
O48m Mapeamento digital de solos da quadrícula de Ribeirão Preto - SP
pelo método Random Forest / Matheus Felipe Oliveira. --
Jaboticabal, 2015
x, 112 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientador: José Eduardo Corá
Banca examinadora: Célia Regina Paes Bueno, Antonio Sérgio
Ferraudo, Waldir de Carvalho Junior
Bibliografia

1. Variáveis geomorfológicas. 2. Geoprocessamento. 3. Random
Forest. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e
Veterinárias.

CDU 528.7



UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

TÍTULO: MAPEAMENTO DIGITAL DE SOLOS DA QUADRÍCULA DE RIBEIRÃO PRETO - SP
PELO MÉTODO RANDOM FOREST

AUTOR: MATHEUS FELIPE OLIVEIRA

ORIENTADOR: Prof. Dr. JOSE EDUARDO CORA

CO-ORIENTADORA: Profa. Dra. CELIA REGINA PAES BUENO

Aprovado como parte das exigências para obtenção do Título de MESTRE EM AGRONOMIA
(CIÊNCIA DO SOLO), pela Comissão Examinadora:

Profa. Dra. CELIA REGINA PAES BUENO

Departamento de Solos e Adubos / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. WALDIR DE CARVALHO JUNIOR

Instituição Embrapa Solos, Rio de Janeiro/RJ

Participação por videoconferência

Prof. Dr. ANTONIO SÉRGIO FERRAUDO

Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Data da realização: 08 de dezembro de 2015.

DADOS CURRICULARES DO AUTOR

MATHEUS FELIPE OLIVEIRA – é natural de Ribeirão Preto, São Paulo. Possui graduação em Geografia (Licenciatura e Bacharelado) pela Unesp Campus de Ourinhos. No primeiro semestre de 2010 foi bolsista Prograd pelo grupo Cenpea (Centro de Estudos de Percepção e Educação Ambiental) realizando trabalhos de extensão no Parque Ecológico Municipal Bióloga Tânia Mara Netto Silva. Em 2011, passou a ser bolsista do grupo PET (Programa de Educação Tutorial) onde desenvolveu projetos de extensão na área de educação e começou sua iniciação científica na área de Bacias Hidrográficas, Uso e Ocupação do Solo, Geoprocessamento e Sensoriamento Remoto. Participa do grupo de pesquisa Centro de Estudo e Divulgação de Informações sobre Áreas Protegidas, Bacias Hidrográficas e Geoprocessamento (CEDIAP- GEO). No primeiro semestre de 2013, foi estagiário na empresa Engemap Geoinformação. No segundo semestre, já concluída sua graduação, ingressou no Mestrado no programa de Pós-Graduação em Agronomia (Ciência do Solo) da Faculdade de Ciências Agrárias e Veterinárias, Unesp Campus de Jaboticabal, sendo bolsista CAPES por 24 meses. Participa do projeto de extensão ECOS voltado para o ensino da formação e conservação de solos, além de ser autor do projeto de extensão Mapeamento das características físicas e químicas do solo em núcleos do assentamento Mário Lago, em Ribeirão Preto.

"Se eu vi mais longe, foi por estar sobre ombros de gigantes."

Isaac Newton

A minha irmã Ana Paula e a minha companheira Tatiana por todas as formas de incentivo, além do simples fato de existirem na minha vida.

Ofereço

Aos meus pais Paulo e Beatriz, pela luta e pela dedicação que me fizeram alcançar meus objetivos.

Dedico

AGRADECIMENTOS

À Professora Dr.^a Célia Regina Paes Bueno pela paciência, confiança e pelo apoio dado em toda minha trajetória no mestrado.

Ao amigo Gabriel Damasceno por ter me encorajado nessa jornada além do apoio dado em todas as etapas, especialmente as mais difíceis.

Ao amigo Rafael Crivelenti por me apresentar ao Mapeamento Digital de Solos.

Aos amigos Leandro, Ronaldo e Bernardo seja pelo tempo de convívio ou pelas contribuições dadas.

Aos amigos da República Boi Banguelo que me acolheram como um membro de família no momento em que mais precisei.

Aos orientados da professora Célia: James, Marcelino, Raquel e Jasmine, pela amizade e pela ajuda nos trabalhos de campo.

A bibliotecária Sueli Cristina dos Santos Pacheco, Instituto Agrônomo de Campinas, pela atenção e pela vontade de ajudar pouco vista hoje em dia.

Ao pesquisador Waldir de Carvalho Junior, da Embrapa Solos do Rio de Janeiro, pela disponibilidade e profissionalismo, que contribuiu de maneira muito importante neste trabalho.

Aos funcionários do departamento de Solos e Adubos da Unesp Jaboticabal, em especial o técnico Dejair, por me acompanhar pelas longas viagens nos trabalhos de campo.

Ao meu grande amor Tatiana por me apoiar e me incentivar, além de compreender os momentos em que estive ausente.

Aos meus pais e minha irmã pelos conselhos que me deram baseados em suas incríveis experiências pessoais.

SUMÁRIO

RESUMO.....	xii
ABSTRACT	xiii
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	2
2.1. Os levantamentos de solos.....	2
2.2. Histórico do levantamento de solos no Brasil.....	4
2.3. Os métodos de levantamento de solos	6
2.3.1. O mapeamento convencional de solos	6
2.3.2. O mapeamento digital de solos.....	8
2.4. Variáveis preditoras no Mapeamento Digital de Solos	10
2.5. O método Random Forest.....	12
2.6. Avaliação da acurácia	17
2.7. Escala efetiva	19
3. MATERIAIS E MÉTODOS	21
3.1. Caracterização da área de estudo	21
3.1.1. Localização	21
3.1.2. Geologia.....	22
3.1.3. Relevo.....	25
3.1.4. Clima, vegetação e hidrografia	28
3.2. Materiais utilizados.....	31
3.2.1. Base cartográfica digital.....	31
3.2.2. Obtenção das variáveis.....	33
3.2.3. Processamento dos dados.....	40
4. RESULTADOS E DISCUSSÃO	42
4.1. Descrição das variáveis dos fatores <i>s.c.o.r.p.a.n.</i>	42
4.1.1. Fator <i>s</i>	42
4.1.2. Fator <i>r</i>	45

4.1.3.	Fator p	54
4.2.	Análise dos dados das amostragens	55
4.3.	Avaliação dos modelos Random Forest ajustados	60
4.4.	Avaliação da predição das classes de solo	62
4.5.	Importância das variáveis	68
4.6.	Mapas das predições das classes de solos	70
5.	CONCLUSÕES	84
6.	REFERÊNCIAS.....	85

MAPEAMENTO DIGITAL DE SOLOS DA QUADRÍCULA DE RIBEIRÃO PRETO - SP PELO MÉTODO RANDOM FOREST

RESUMO – O presente estudo buscou desenvolver um modelo capaz de compreender as relações solo-paisagem para a predição de classes de solo das folhas do IBGE de Ribeirão Preto, Serrana, Cravinhos e Bonfim Paulista, que constituem a quadrícula de Ribeirão Preto. Para isto, foram utilizadas informações contidas em um mapa pedológico convencional semidetalhado na escala 1:100.000, um Modelo Digital de Elevação (MDE) com resolução espacial de 30 metros, além do mapa geológico na escala 1:50.000. Do mapa geológico foi obtida a litologia e do MDE, foram obtidas as variáveis geomorfológicas por meio de técnicas de geoprocessamento. Todas essas informações foram relacionadas em uma matriz, de onde foram selecionadas três amostragens estratificadas de acordo com a área das classes, extraído-se dados para treino e teste, que foram utilizados para aplicação em modelos do método Random Forest e avaliação da acurácia. Foram testados diferentes ajustes, com aplicação dos modelos nas classes no segundo e terceiro nível categórico. Com uma amostragem que compreende apenas 0,43% do total da área, o modelo para o segundo nível categórico apresentou uma exatidão global de 62,5%, com o mapa digital de solos apresentando uma persistência de 70,63% das classes do mapa original, valores maiores do que os apresentados para o terceiro nível categórico, com exatidão global de 57,1% e persistência de 44,24%. As variáveis mais importantes na compreensão das relações solo-paisagem foram Litologia, Elevação, Declividade e Distância da rede de drenagem. O estudo mostrou que a metodologia empregada é capaz de contribuir para criação de mapas de solo, com a possibilidade de ser empregado em áreas onde não há informações de solos pré-existent, de maneira rápida e menos onerosa, auxiliando o trabalho dos pedólogos.

Palavras-chave: variáveis geomorfológicas, geoprocessamento, Random Forest.

THE DIGITAL SOIL MAPPING OF RIBEIRÃO PRETO - SP GRID BY RANDOM FOREST METHOD

ABSTRACT – This study aimed to develop a model to understand the soil-landscape relationships to predict soil classes of topographic sheets of IBGE from Ribeirão Preto, Serrana, Cravinhos and Bonfim Paulista, constituting the grid Ribeirão Preto. For this, we used information included in a conventional semi-detailed soil map at 1:100,000 scale, a Digital Elevation Model (DEM) with a spatial resolution of 30 meters, in addition to the geological map at 1: 50,000 scale. From geological map was obtained lithology and from MDE were obtained the geomorphometric variables through geoprocessing techniques. All this information was linked in a matrix, from which they were selected three stratified sampling according to the area of classes, extracting data for training and testing, which were used for use in models of Random Forest method and evaluation of accuracy. Adjustments were tested with application of models in classes on the second and third categorical level. With a sample comprising only 0.43% of the total area, the model for the second categorical level had an overall accuracy of 62.5%, with the digital soil map showing a persistence of 70.63% of classes from original map, higher values than those presented for the third categorical level, with an overall accuracy of 57.1% and persistence of 44.24%. The most important variables in understanding the soil-landscape relationships were Lithology, Elevation, Slope Distance and drainage network. The study showed that the method is able to contribute to the creation of soil maps, with the possibility of being employed in areas where there is no pre-existing soil information quickly and less costly way, assisting the work of soil scientists.

Keywords: geomorphometric variables, geoprocessing, Random Forest.

ÍNDICE DE TABELAS

Tabela 1 - Níveis de detalhamento por escalas, área mínima mapeável e densidade de observações. Adaptado de IBGE (2007).....	3
Tabela 2 - Matriz de confusão com proporções de observações dentro de c classes preditas i e classes observadas j	17
Tabela 3 - Escala de classificação do valor K	19
Tabela 4 – Unidades geológicas, litologia, idade e área em hectares e porcentagem das respectivas classes no total da quadrícula.	25
Tabela 5 - Trecho da matriz de dados grid.....	39
Tabela 6 - Tabela de conversão da classificação de solos antiga para atual do mapa semidetalhado da quadrícula de Ribeirão Preto, com a área de cada classe em hectares e porcentagem.....	43
Tabela 7 - Indicadores e cálculo da escala efetiva para o mapa pedológico da quadrícula de Ribeirão Preto.....	44
Tabela 8 - Distribuição das classes de elevação por área na quadrícula de Ribeirão Preto.....	46
Tabela 9 - Distribuição das classes de declividade por área na quadrícula de Ribeirão Preto de acordo com as classes definidas por Lepsch et al. (1991).	47
Tabela 10 - Distribuição das classes de Aspecto por área na quadrícula de Ribeirão Preto.....	49
Tabela 11 - Distribuição das classes de curvatura por área na quadrícula de Ribeirão Preto.....	50
Tabela 12 - Distribuição das classes de distância da rede de drenagem por área na quadrícula de Ribeirão Preto.....	51
Tabela 13 - Distribuição da intervalo de classes de CTI por área na quadrícula de Ribeirão Preto.	52
Tabela 14 - Distribuição da intervalo de classes de SPI por área na quadrícula de Ribeirão Preto.	54
Tabela 15 - Distribuição da intervalo de classes de Litologia por área na quadrícula de Ribeirão Preto	55

Tabela 16 - Quantidade de pontos para treino, teste e o total de cada amostragem aleatória.....	55
Tabela 17 - Estatística descritiva dos conjuntos de amostragem e do grid da quadrícula.....	58
Tabela 18 - Matriz de confusão do modelo Random Forest com amostragem A2, <i>ntree</i> = 300 e <i>mtry</i> = 4 no segundo nível categórico.	63
Tabela 19 - Matriz de confusão do modelo Random Forest com amostragem A2, <i>ntree</i> = 300 e <i>mtry</i> = 4 no terceiro nível categórico.	67
Tabela 20 - Caracterização do ponto 3.	78
Tabela 21- Caracterização do ponto 5.	79
Tabela 22 - Caracterização do perfil 1350.....	80
Tabela 23 - Caracterização do perfil 1346.....	81
Tabela 24- Caracterização do ponto 33.	82

ÍNDICE DE FIGURAS

Figura 1 - Representação esquemática de árvore de decisão onde Xi são as variáveis; a, b, c, d, e são os limites da regra de decisão; A, B e C correspondem as classes.	13
Figura 2 - Funcionamento do método Random Forest. Adaptado de OSHIRO (2013).	15
Figura 3 - Mapa de localização da quadrícula de Ribeirão Preto.	22
Figura 4 - Mapa geológico da quadrícula de Ribeirão Preto (adaptado de Sinelli et al., 1973). Sedimentos Aluvionares (Qa) – areias e argilas; Sedimentos Correlatos a Formação Itaqueri (KTii) – arenitos conglomeráticos; Formação Botucatu (JKb) – arenitos eólicos; Formação Pirambóia (TrJp), – arenitos fluviais; Formação Serra Geral (JKsg) – Basaltos e intercalações de arenitos intertrapeanos; Formações Superficiais (TQir) – areias inconsolidadas e Intrusivas Básicas (sill) (JKβ) – diabásios.	24
Figura 5 – Vista do relevo testemunho tabuliforme.	26
Figura 6 - Modelo tridimensional da quadrícula de Ribeirão Preto: a) visada NO-SE; b) visada NE-SO.	27
Figura 7 - Balanço hídrico do município de Ribeirão Preto. Fonte: Sentelhas et al. (1999).	28
Figura 8 - Balanço hídrico do município de São Simão. Fonte: Sentelhas et al. (1999).	29
Figura 9 - Mapa da hidrografia da quadrícula de Ribeirão Preto.	31
Figura 10 - Funcionamento da ferramenta <i>Fill</i>	34
Figura 11 - Combinações de curvatura (ESRI, 2014).	35
Figura 12 – Exemplo de funcionamento do modelo D8 para o cálculo da direção de fluxo.	36
Figura 13 – Acumulação de fluxo.	37
Figura 14 - Fluxograma da metodologia utilizada.	41
Figura 15 - Mapa Pedológico da Quadrícula de Ribeirão Preto (adaptado de Oliveira & Prado, 1987).	44
Figura 16 - Mapa da elevação da quadrícula de Ribeirão Preto.	45

Figura 17 - Mapa de declividade da quadrícula de Ribeirão Preto.....	47
Figura 18 – Mapa do Aspecto da quadrícula de Ribeirão Preto.....	48
Figura 19 - Mapa de curvatura da quadrícula de Ribeirão Preto.....	50
Figura 20 - Mapa da distância da rede de drenagem.....	51
Figura 21 - Mapa de CTI da quadrícula de Ribeirão Preto.....	52
Figura 22 - Mapa de SPI da quadrícula de Ribeirão Preto.....	53
Figura 23 - Mapa da Litologia na quadrícula de Ribeirão Preto.	54
Figura 24 – Distribuição percentual das amostras nas diferentes amostragens de treino para cada classe em relação ao grid no segundo nível categórico.	56
Figura 25 – Distribuição percentual das amostras nas diferentes amostragens de treino para cada classe em relação ao grid no terceiro nível categórico.	56
Figura 26 - Histogramas das variáveis para o grid e para as amostragens A1, A2 e A3.....	59
Figura 27 - Taxa de erro OOB pelo número de árvores no segundo nível categórico.	61
Figura 28 - Taxa de erro OOB pelo número de árvores no terceiro nível categórico.	61
Figura 29 - Distribuição da média das variáveis dos dados de treino entre as classes MT e RL (valores normalizados entre 0 e 1).	64
Figura 30 - Distribuição da média das variáveis dos dados de treino entre as classes LV, LVA, NV e RQ (valores normalizados entre 0 e 1).....	65
Figura 31 - Distribuição da média das variáveis dos dados de treino entre as classes CX, GM, GX e OX (valores normalizados entre 0 e 1).	66
Figura 32 - Distribuição da média das variáveis dos dados de treino entre as classes LVd, LVdf, LVef, LVwf, NVef e RQo (valores normalizados entre 0 e 1).....	68
Figura 33 - Gráfico de importância das variáveis pelas estimativas MDA e MDG. a) segundo nível categórico; b) terceiro nível categórico.	69
Figura 34 - Mapa de solos original (a) e mapa de solos digital (b) da quadrícula de Ribeirão Preto no segundo nível categórico.....	70
Figura 35 - Comparação e variação da área das classes de solo entre o mapa original e o mapa digital no segundo nível categórico.....	71

Figura 36 - Mapa da mudança ou persistência de classes no mapa digital em relação ao mapa original no segundo nível categórico.	72
Figura 37 - Porcentagem das classes dentro das associações fragmentadas no mapa digital para o segundo nível categórico.	73
Figura 38 - Mapa de solos original (a) e mapa de solos digital (b) da quadrícula de Ribeirão Preto no terceiro nível categórico.....	74
Figura 39 - Comparação e variação da área das classes de solo entre o mapa original e o mapa digital no terceiro nível categórico.....	74
Figura 40 - Mapa da mudança ou persistência de classes no mapa digital em relação ao mapa original no terceiro nível categórico.....	76
Figura 41 - Porcentagem das classes dentro das associações fragmentadas no mapa digital para o terceiro nível categórico.....	77
Figura 42 - Perfil do Nitossolo Vermelho eutrófico no município de Dumont - SP.	78
Figura 43 - Localização do ponto 3 obtido em campo e sua comparação no mapa original e no mapa digital do terceiro nível categórico.....	79
Figura 44- Paisagem de ocorrência do Neossolo Quartzarênico órtico.	79
Figura 45 - Localização do ponto 5 obtido em campo e sua comparação no mapa original e no mapa digital do terceiro nível categórico.....	80
Figura 46 - Perfil do Latossolo Vermelho acrífico no município de Jardinópolis - SP. Fonte: Oliveira & Prado (1987).	80
Figura 47 - Localização do perfil 1350 e sua comparação no mapa original e no mapa digital do terceiro nível categórico.	81
Figura 48 - Perfil do Neossolo Litólico distrófico no município de Luís Antônio - SP. Fonte: Oliveira & Prado (1987).	81
Figura 49 - Localização do perfil 1346 e sua comparação no mapa original e no mapa digital do terceiro nível categórico.	82
Figura 50 - Localização do perfil 33 e sua comparação no mapa original e no mapa digital do segundo nível categórico.	83

1. INTRODUÇÃO

O Brasil possui uma grande diversidade de solos por conta de uma variedade biológica, climática, geológica e geomorfológica, proporcionadas pela sua extensão territorial de dimensões continentais. Esta diversidade de solos dá suporte a biomas de grande importância, centros urbanos e áreas rurais, sendo seu conhecimento imprescindível para sua conservação e preservação.

Os mapas pedológicos que representam a distribuição espacial dos tipos de solos se tornam essenciais para o uso e manejo de maneira racional e sustentável. As informações contidas no mapa pedológico possibilitam ao poder público a criação de políticas, construção de instrumentos e leis que permitam a formulação de diretrizes de preservação, recuperação e desenvolvimento (CARVALHO, et al., 2013).

De acordo com Oliveira (2005), os mapas pedológicos são também documentos muito úteis no planejamento inicial de cidades, ou em planos diretores, indicando locais mais apropriados para escavações, na seleção de áreas para estabelecimento de cinturão verde, cemitérios, aterros sanitários, entre outros usos, além de auxiliar na formulação de políticas de uso do solo.

No Brasil existem extensas áreas que carecem de informação completa e em escala adequada para solucionar problemas de uso, manejo, conservação, prevenção e recuperação de áreas degradadas, agrícolas e não agrícolas (SANTOS, 2007).

A cobertura de mapas de solos do Brasil é pouco detalhada (escala 1:1.000.000), realizadas pelo projeto RADAM (Radar na Amazônia) e pela Embrapa Solos (DEMATTE et. al, 2004). O estado de São Paulo foi pioneiro no levantamento sistemático de solos por quadrículas na escala 1:100.000, porém os mapeamentos realizados recobrem apenas 15% da área total do estado (CRIVELENTI, 2009).

Os métodos de levantamento tradicionais para escalas detalhadas e semidetalhadas apresentam tempo de execução elevados. Métodos de mapeamento supervisionados por meios computacionais estão cada vez mais sendo trabalhados

e refinados para subsidiar esta falta de dados com maior detalhamento (COELHO, 2010).

Parte-se da hipótese de que é possível criar um modelo capaz de compreender a relação solo-paisagem da área de estudo por meio do mapa pedológico semidetalhado, do mapa geológico e de variáveis geomorfométricas, utilizando uma baixa densidade de amostras em relação a área total, capaz de reproduzir o mapa original com certo grau de semelhança. O objetivo do presente trabalho é executar o mapeamento digital das folhas do IBGE de Ribeirão Preto, Serrana, Bonfim Paulista e Cravinhos, que constituem a quadrícula de Ribeirão Preto, avaliando o desempenho dos modelos Random Forest para as classes de solo segundo e terceiro nível categórico.

2. REVISÃO DE LITERATURA

2.1. Os levantamentos de solos

Os levantamentos de solos de acordo com a Embrapa (1995), tem como objetivo identificar e separar unidades de mapeamento, que são grupos de áreas de solos criados para mostrar sua distribuição espacial, segundo uma unidade taxonômica. Um levantamento é constituído em sua forma final, por um mapa e um texto explicativo, que define, descreve e interpreta as classes de solos componentes de unidades de mapeamento.

Uma unidade de mapeamento é o agrupamento de área de solos, estabelecido para possibilitar a representação cartográfica, distribuição espacial, extensão e limites dos solos (IBGE, 2007). Uma unidade de mapeamento pode ser constituída de uma única unidade taxonômica, conhecida como unidade simples, ou por várias unidades taxonômicas, conhecidas como unidades combinadas (EMBRAPA, 1995).

Entre as unidades combinadas, encontram-se as associações, que são constituídas pelo agrupamento de duas ou mais unidades taxonômicas distintas, apresentando limites nítidos ou pouco nítidos entre si, podendo ser separadas por

levantamentos de maior detalhe. As associações são estabelecidas pela necessidade de generalizações cartográficas, em função da escala e do padrão de ocorrência dos solos em uma área. É designado no mapa pela junção dos nomes de duas ou mais classes ligados pelo sinal “+” (IBGE, 2007; EMBRAPA, 1995).

Os levantamentos podem ser identificados como esquemático, exploratório, reconhecimento, semidetalhado, detalhado e ultradetalhado. Estes níveis se diferenciam em função da precisão, escala cartográfica e objetivo do trabalho. A densidade de observações que corresponde aos exames visuais e coletas de campo varia em função dos diferentes níveis (Tabela 1).

Tabela 1 - Níveis de detalhamento por escalas, área mínima mapeável e densidade de observações. Adaptado de IBGE (2007).

Níveis de detalhamento	Escalas usuais	Área mínima mapeável (ha)	Densidade de observações (obs./ha)
Ultradetalhado	1:500 - 1:5000	0,001 - 10	0,20 - 4
Detalhado	1:7000 - 1:20000	0,19 - 1,60	0,02 - 0,2
Semidetalhado	1:25000 - 1:50000	2,50 - 10	0,04 - 2
Reconhecimento	1:50000 - 1:750000	10 - 2250	< 0,04
Exploratório	1:750000 - 1:2500000	2250 - 25000	sem especificação

Os levantamentos exploratórios são apropriados para grandes extensões territoriais, onde necessita-se de informações de natureza qualitativa do solo, identificando áreas de maior ou menor potencial, prioritárias ao desenvolvimento regional. Os levantamentos de reconhecimento são executados para avaliação qualitativa e semiquantitativa, visando a estimativa do potencial de uso agrícola e não agrícola (EMBRAPA, 1995).

Os levantamentos semidetalhados são utilizados desde estudos ambientais para empreendimentos localizados, manejos e conservação do solo, microbacias, projetos agrícolas, além de outras propriedades importantes para fins de engenharia civil, considerando que os levantamentos semidetalhados podem ser necessários em áreas de diversas extensões (áreas especiais, microrregiões, bacias hidrográficas, municípios e estados).

Os levantamentos detalhados e ultradetalhados acumulam um grande número de características para individualização das classes de solos em níveis categóricos mais baixos. São utilizados em escalas grandes, para projetos de uso intensivo dos solos, em áreas relativamente pequenas no caso do levantamento detalhado e muito pequenas para o ultradetalhado. São utilizadas para decisões localizadas e solução de problemas específicos (EMBRAPA, 1995).

2.2. Histórico do levantamento de solos no Brasil

No ano de 1934, os estudos de solos até então eram denominados estudos agrológicos e se iniciaram na região Nordeste, com o intuito de conter a seca de 1930 por meio da construção de açudes (CARVALHO, et al., 2013). O Instituto José Augusto Trindade, em Sousa na Paraíba, é considerado pioneiro em realizar estudos detalhados de solos. O mapa agrológico da bacia de irrigação do açude de São Gonçalo é considerado o primeiro mapa de solos do Brasil (IBGE, 2007).

O conhecimento sobre o solo até então não estava organizado em um sistema taxonômico, mas o fato da agricultura constituir uma das principais bases da economia no Brasil criou a necessidade em se ter um maior conhecimento a respeito dos solos e seus potenciais agrícolas. Dessa forma, o governo passou a fomentar pesquisas para atender os anseios da economia nacional (CARVALHO, et al., 2013).

Em 1950, o Instituto Agrônomo em parceria com o Departamento de Água e Energia e o governo do estado de São Paulo, realizaram levantamentos detalhados na bacia de Taubaté, no vale do Rio Paraíba, para suprir a carência de dados sobre características físicas e químicas dos solos da região, sendo necessária para o planejamento geral de irrigação e drenagem. Estes esforços resultaram em três relatórios e três mapas na escala 1:10.000, onde são descritos os solos e suas principais características químicas e físicas (VERDADE et al., 1961).

No ano de 1960 foi publicado o mapa de solos do Estado de São Paulo na escala de 1:500.000, trabalho realizado pelo Centro Nacional de Pesquisa de Solo, atual Embrapa Solos. Este mapa foi atualizado em uma parceria do Instituto Agrônomo com a Embrapa Solos, projeto iniciado no ano de 1996, que contribuiu

para resolver antigos problemas e incorporar novos conhecimentos (ROSSI & OLIVEIRA, 2000).

O esforço pioneiro do governo na década de 1970 para realizar pesquisas sobre os recursos naturais do território brasileiro resultou no projeto RADAM, organizado pelo Ministério de Minas e Energia. A princípio, priorizou-se a coleta de dados sobre recursos minerais, solos, vegetação, uso da terra e cartografia na faixa de influência da rodovia Transamazônica. Foi utilizado para as pesquisas, o radar de visada lateral (SLAR – sidelooking airborne radar), que superou a dificuldade na obtenção de imagens em condições de alta nebulosidade (OLIVEIRA, 2011).

O êxito no trabalho resultou na expansão da série “Levantamento de Recursos Naturais” que incluía o mapa temático de solos na escala 1:1.000.000, recobrindo a partir de 1975 todo o território nacional, no qual passou denominar-se Projeto RADAMBRASIL (CARVALHO, et al., 2013).

Devido ao reduzido número de profissionais e a escassez de recursos financeiros, os levantamentos realizados pelo projeto RADAMBRASIL eram de caráter exploratório, enquanto os levantamentos realizados pelo Centro Nacional de Pesquisa de Solos, atual Embrapa Solos, eram de reconhecimento (CHAGAS, 2006).

Levantamentos semidetalhados de solos do estado de São Paulo na escala 1:100.000 foram realizados a partir de 1975, pela Seção de Pedologia (antiga seção de Agroecologia) do Instituto Agrônomo (IAC), recobrindo uma área de apenas 15% do estado (SILVA, 2012).

A realização de levantamentos pedológicos no Brasil contribuiu para diversos órgãos de assistência técnica, planejamento e execução de projetos na esfera federal ou estadual, para fins de seleção de áreas para colonização, estudos de viabilidade técnica de projetos de irrigação e drenagem, introdução de novas culturas agrícolas, planejamento regional e local, zoneamentos pedoclimáticos, assistência a áreas inundadas por represas hidrelétricas, seleção de áreas experimentais e atendimento de instituições de ensino (EMBRAPA, 1995).

Apesar dessas contribuições, o levantamento de solos foi considerado como atividade não ligada diretamente a produção, sendo sua relação custo/benefício pouco compreendida e difícil de ser estimada. Houve uma série de restrições

orçamentárias que desacelerou a atualização progressiva do conhecimento de solos a partir de 1974, onde observa-se que vários estados da Federação são carentes de informações de solos (EMBRAPA, 1995).

Recentemente, técnicas de sensoriamento remoto, estatística, modelagem, sistema de posicionamento global, o uso generalizado da informática e o acesso instantâneo a informação, transformaram radicalmente os procedimentos para levantamento e atualização das bases dos mapas e seus derivados (CARVALHO, et al., 2013). Novas oportunidades surgem para a ciência do solo, como o Mapeamento Digital de Solos, uma abordagem surgida em 1970 e que foi introduzida no Brasil na década de 2000, com estudos cada vez mais frequentes no país (TEN CATEN, 2012).

2.3. Os métodos de levantamento de solos

2.3.1. O mapeamento convencional de solos

O método de Levantamento Pedológico convencional, ou também conhecido como método clássico, é baseado no conceito de solo como um corpo natural organizado, na qual é resultado da interação complexa de fatores pedogenéticos, que foi proposto por Dokouchaev. A partir deste conceito, o solo deixou de ser considerado simplesmente como um corpo inerte, resultado de uma rocha matriz, passando a ser identificado como um material que evolui no tempo, sob ação das interações entre a litosfera, atmosfera, hidrosfera e a biosfera (COELHO, 2010).

O Levantamento convencional, é baseado no conceito de que o solo é resultado da interação de seu ambiente de formação (QI & ZHU, 2003). O processo de formação do solo proposto por Jenny (1941) relaciona os principais fatores identificados por Dokuchaev, exposto na equação 1 abaixo:

$$S = f(cl.p.o.r.t) \quad (1)$$

onde o solo (S) é resultado da interação dos fatores clima (*Climate* – cl), organismos (*Organisms* – o), relevo (*Relief* – r), material de origem (*Parental*

material – p) e tempo (*Time* – t), sendo esta última a única variável independente. Este método é conhecido como CLORPT, devido as iniciais dos fatores pedogenéticos levados em consideração.

O modelo utilizado pelo pedólogo para determinar as classes de solo e delimitar seus limites, consiste na correlação entre os tipos de solos e o ambiente onde se encontram, baseados na abordagem CLORPT. Esta abordagem é considerada qualitativa, de natureza empírico-determinística, realizada por meio da análise fisiográfica da paisagem e pela interpretação de fotografias aéreas (MEDONÇA-SANTOS & MANZATTO, 2007).

O Levantamento Pedológico convencional é baseado no modelo de representação conhecido como discreto, na qual os solos no campo são representados por meio de polígonos e mostram a distribuição de uma classe de solo particular, que é classificando de acordo com o sistema taxonômico vigente e estabelece limites entre as classes definidas no mapa, tendo como base o pressuposto que os fatores de formação (CLORPT) controlam a distribuição dos diferentes solos na paisagem (MEDONÇA-SANTOS & SANTOS, 2003; ZHU, 1997).

O modelo de representação discreto limita o tamanho da unidade de mapeamento, sendo assim, unidades que apresentam tamanho menor do que o estabelecido acabam sendo ignoradas ou anexadas em unidades maiores, sendo criadas associações de solos que expressam a inclusão de diferentes solos no polígono (ZHU, 2001).

O método convencional não considera a dependência espacial entre as unidades de mapeamento, principalmente em levantamentos detalhados ou áreas onde os limites entre os solos não são óbvios. Dessa forma, o modelo discreto, utilizado no mapeamento convencional não é realístico, por que os limites naturais na paisagem tendem a ser mais graduais que abruptos (MEDONÇA-SANTOS & SANTOS, 2003).

As características e propriedades dos solos variam espacialmente, sendo esta variabilidade resultante da atuação do clima e organismos sobre o material de origem, sendo condicionadas pelo relevo e pelo tempo (SOIL SURVEY DIVISION STAFF, 1993). A desvantagem da abordagem CLORPT é a de não tratar

adequadamente as autocorrelações espaciais das propriedades dos solos, especialmente a nível local (MCBRATNEY et al., 2000).

Os pedólogos interpretam as características locais de topografia e vegetação como indícios de combinações dos fatores de formação dos solos, podendo inferir os limites entre classes e propriedades dos solos. Porém, seu paradigma é inadequado, pois ignora a variabilidade espacial dos fatores de formação e do próprio solo resultante (BURROUGH & MCDONNELL, 1998).

A produção manual dos mapas de solos pelo levantamento convencional dificulta sua atualização rápida e acurada. A interpretação visual de fotografias aéreas consome muito tempo e é um processo propenso a erros, podendo ter como resultado, interpretações erradas durante o delineamento dos limites dos solos (ZHU et al., 2001).

2.3.2. O mapeamento digital de solos

Devido ao caráter qualitativo do mapeamento convencional, este método passou a ser alvo de críticas. Novas abordagens utilizando modelagem quantitativa surgem com o objetivo de classificar e estudar os padrões de variação espacial dos solos na paisagem, objetivando melhorar seu conhecimento, com mais precisão e qualidade. Os métodos quantitativos foram coletivamente categorizados em um campo emergente da ciência do solo chamado de pedometria (MENDONÇA-SANTOS & SANTOS, 2003; MCBRATNEY, 2000).

O termo pedometria foi cunhado por Alex B. McBratney, um neologismo das palavras pedo (solo) e metria (medida). Consiste na aplicação de métodos matemáticos e estatísticos para modelagem quantitativa de solos, com o objetivo de analisar sua distribuição, propriedades e comportamentos (HENGL, 2003).

A pedometria lida com a incerteza em modelos de solos de variação determinística ou estocástica, imprecisão ou falta de conhecimento dos processos e propriedades do solo (MCBRATNEY, 2000).

Para tratar o problema da variabilidade do solo sob uma ótica diferente, adota-se o modelo contínuo, na qual se assume que as características dos solos variam gradualmente no espaço e que assim devem ser representadas. Dessa

forma, modelos geoestatísticos são aplicados a Ciência do Solo, na tentativa de representar a variabilidade espacial dos atributos do solo e sua representação contínua (MEDONÇA-SANTOS & MANZATTO, 2007).

O mapeamento pedométrico é geralmente caracterizado como uma produção geoestatística quantitativa da geoinformação do solo, sendo referido como mapeamento preditivo de solos ou mapeamento digital de solos, pois depende muito do uso de tecnologias da informação (HENGL, 2003).

De acordo com Lagacherie & McBratney (2007), o mapeamento digital de solos consiste na criação e alimentação de sistemas de informação espacial de solos por meio de modelos numéricos, de modo a prever a variabilidade espacial e temporal de classes e propriedades do solo, a partir de observações, conhecimento e dados de covariáveis ambientais relacionados.

Uma abordagem proposta por McBratney et al. (2003) para o mapeamento digital de solos é a adaptação do modelo teórico de Jenny (1941), onde é incluída a posição espacial e a possibilidade de prever o solo a partir de informações sobre ele mesmo (mapas existentes de classes e propriedades). Este modelo é conhecido como modelo *s.c.o.r.p.a.n*, representado na equação 2 abaixo:

$$S_{c,a} = f(s.c.o.r.p.a.n) \quad (2)$$

onde:

$S_{c,a}$ = Solo, classe ou atributo.

s = *soil* (solo ou propriedade do solo em um dado ponto);

c = *climate* (clima, propriedades climáticas do clima em um dado ponto);

o = *organisms* (organismos, cobertura vegetal e uso do solo);

r = *topography* (relevo, atributos da paisagem);

p = *parent material* (material de origem, litologia);

a = *age* (idade, tempo);

n = *space* (espaço, posição espacial).

Este modelo consiste na descrição empírica de relações entre o solo e fatores referenciados espacialmente com o intuito de usá-los como funções de predição espacial do solo (MCBRATNEY, 2003).

A possibilidade de prever o solo a partir de informações já existentes com a criação de regras de classificação pode ser empregada para o treinamento de novos pedólogos ou para produção de novos mapas de solos em áreas onde não há informações de solos, desde que as mesmas relações solo-paisagem sejam válidas (QI & ZHU, 2003).

Para a modelagem da equação *s.c.o.r.p.a.n*, ou seja, para prever de forma quantitativa as relações entre classes ou propriedades de solos com seu ambiente, são utilizados diversos métodos como: modelos lineares generalizados, árvores de regressão e de classificação, redes neurais, sistemas fuzzy e a geoestatística (MEDONÇA-SANTOS & MANZATTO, 2007).

Os dados e informações sobre as variáveis ambientais utilizadas na construção dos modelos mais comumente utilizados são a altitude, declividade, o índice de umidade que podem ser derivadas do Modelo Digital de Elevação (MDE), entre outros. Informações como Índice de Vegetação por Diferença Normalizada (NDVI) e composição mineralógica do solo podem ser extraídas de bandas espectrais (CARVALHO, et al., 2013; MEDONÇA-SANTOS & MANZATTO, 2007).

O modelo *s.c.o.r.p.a.n* juntamente com as funções espaciais de predição de solos, tem se mostrado um método importante, especialmente em áreas onde os recursos em dados e informações de solos são escassos, como é o caso do território brasileiro (MEDONÇA-SANTOS & MANZATTO, 2007).

2.4. Variáveis preditoras no Mapeamento Digital de Solos

O modelo *s.c.o.r.p.a.n*. possui sete fatores ou conjunto de variáveis independentes que o torna diferente do modelo proposto por Jenny, necessitando assim de mais informações sobre cada uma delas. Os dados que irão representar os fatores propostos no modelo scorpan são escolhidos de acordo com as fontes disponíveis e segundo alegações científicas. As ciências da terra e pesquisas ecológicas dos últimos 20 anos tem contribuído para a criação de variáveis

ambientais de entrada que representam os fatores do modelo s.c.o.r.p.a.n. (MCBRATNEY, 2003).

A evolução do mapeamento digital de solos é contemporânea, e ocorre de acordo com o desenvolvimento das infraestruturas de dados espaciais, que fornecem cada vez mais embasamento para se representar os fatores de formação do solo, por exemplo, o relevo por meio do Modelo Digital de Elevação (MDE), e os organismos do solo por meio de índices de vegetação. A forma clássica de observação e caracterização de solos no campo está cada vez mais integrada aos Sistemas de Informação Geográfica (SIG), graças às novas ferramentas, como o GPS (LAGACHERIE & MCBRATNEY, 2007).

A maioria dos estudos utiliza o MDE como fonte principal de dados auxiliares, seguido de imagens de sensoriamento remoto e dados preexistentes de cobertura do solo. Diferentes atributos do relevo podem ser parametrizados a partir de um MDE, como altitude, inclinação, Aspecto, diferentes curvaturas, área curva ascendente, índice topográfico composto, entre outros (MCBRATNEY, 2003).

Para informações sobre o material de origem pode utilizar-se mapas geológicos, porém estes mapas devem concentrar-se mais sobre a litologia do que sobre a estratigrafia para ser mais útil para a previsão do solo (MCBRATNEY, 2003). No Brasil, a pequena escala de alguns mapas geológicos faz com que informações sobre o material de origem muitas vezes deixem de ser utilizadas no Mapeamento Digital de Solos (TEN CATEN, 2012).

A posição topográfica influencia a formação dos horizontes do solo e de suas respectivas características. A organização da superfície do terreno de acordo com modelos geomorfológicos formais de paisagem e as relações entre forma e paisagem, mostram potencial para melhorar a predição de solos (JENNY, 1941; MOORE et al., 1993).

A geomorfometria, conhecida como geomorfologia quantitativa ou modelagem digital do terreno, tenta descrever quantitativamente as formas da superfície da terra pelo uso conjunto de matemáticas, engenharia, ciências da terra e da computação. As variáveis geomorfométricas podem ser divididas em primárias e secundárias (PIKE, 2000).

As variáveis primárias são derivadas do MDE, tendo como principais variáveis a elevação, declividade, plano e perfil de curvatura, caminho de fluxo e área específica de contribuição. As variáveis secundárias ou compostas são resultados de combinações entre os atributos primários, podendo ser utilizados para caracterizar a variabilidade espacial de processos que ocorrem na paisagem (MOORE et al., 1993).

A inclinação local determina não só a intensidade de processos como erosão e redistribuição de sedimentos, mas também a capacidade de drenagem local, exercendo influência sobre os padrões de fluxo de água ao nível da paisagem. Características topográficas como curvatura, declividade e Aspecto influenciam as condições hidrológicas de um local e as diferentes condições de umidade do solo e padrões de fluxo (CARRÉ; MCBRATNEY, 2005).

A representação do fator *o* no modelo *s.c.o.r.p.a.n*, que compreende a ação dos organismos, cobertura vegetal e uso do solo, representa um grande desafio. A modificação da cobertura vegetal original, a presença de nuvens, a prática do plantio direto, entre outros, são obstáculos para a obtenção do fator *o* e comprometem a utilização de mapas de uso da terra e índices de vegetação como o NDVI.

2.5. O método Random Forest

O método Random Forest é uma extensão da técnica de modelos baseados em árvores. Os modelos de árvores consistem na criação de uma hierarquia de nós internos e externos conectados por ramos. O nó interno ou raiz é a unidade de tomada de decisão, que avalia qual será o próximo nó descendente ou filho realizando um teste lógico. O nó externo não possui descendente, e é conhecido também como folha ou nó terminal, sendo associado a um rótulo ou um valor (BREIMAN et al., 1984).

De acordo com McBratney et al. (2003), ao invés de um modelo ser ajustado para os dados, é gerada uma estrutura em árvore que divide os dados de forma recursiva para um número de grupos, onde cada divisão é escolhida de modo a maximizar alguma medida de diferença na variável de resposta. Modelos baseados em árvores determinam automaticamente quais as variáveis e pontos de divisão, e

também a forma (topologia) da árvore. É possível utilizar dados categóricos e contínuos para predição de classes de solos.

A representação esquemática de uma árvore de decisão é apresentada na Figura 1.

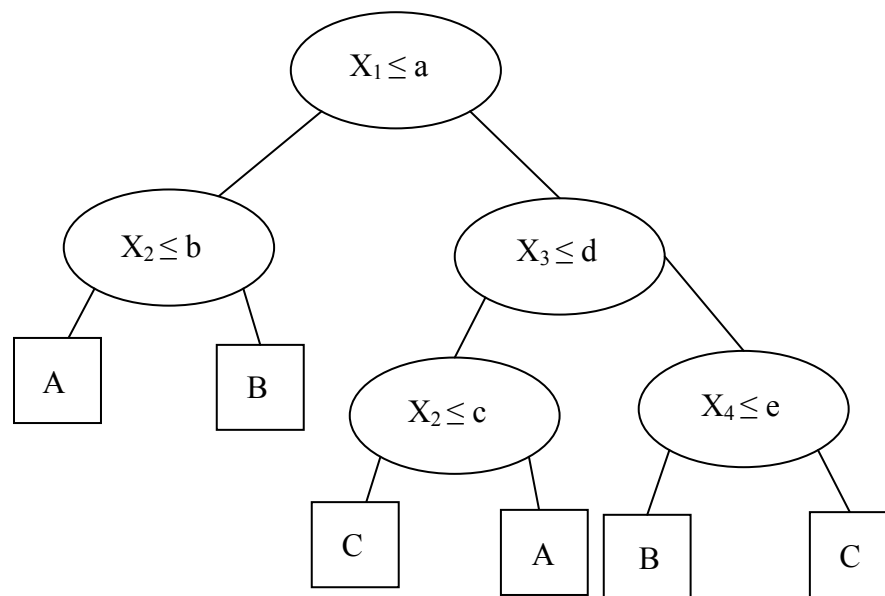


Figura 1 - Representação esquemática de árvore de decisão onde X_i são as variáveis; a, b, c, d, e são os limites da regra de decisão; A, B e C correspondem as classes.

A técnica de árvores de decisão para a predição de classes de solos tem sido utilizada por diversos autores. Entre as vantagens na utilização das árvores de decisão na predição de classes de solo, destacam-se a maior facilidade de implementação e a maior acurácia em relação a outros métodos de classificação (COELHO & GIASSON, 2010; CRIVELANTI, 2009; SARMENTO, 2009).

As predições de atributos e classes de solos pela técnica de árvores de decisão tem sido cada vez mais utilizadas devido sua robustez no tratamento de um grande volume de dados, porém, são poucos os trabalhos que dão ênfase a análise das regras de decisão geradas na modelagem. (TEN CATEN, 2011).

A extensão da técnica de árvores de decisão, o método Random Forest, é uma combinação de preditores baseados em árvore, onde cada árvore depende de

um vetor aleatório, que são amostrados com mesma distribuição e de forma independente para todas as árvores da floresta. A combinação de grandes conjuntos de árvores de decisão tende a uma maior precisão. A acurácia da predição tem melhores resultados quando múltiplos modelos são gerados e agregados produzindo uma estimativa (BREIMAN, 2001; MCBRATNEY et al., 2003).

Este método trabalha com um número m de variáveis retirado do total de variáveis M do conjunto de dados treino, de modo que em cada nó m variáveis são selecionadas de forma aleatória para fora do conjunto M , utilizado para dividir o nó. O valor de m especificado é mantido constante.

Entre as abordagens aplicadas a métodos de aprendizagem estatística, o Random Forest utiliza o método *bagging*. O método *bagging* consiste na geração de amostragens aleatórias dos dados para cada árvore de decisão, que são chamadas de amostras *bootstrap*. O *bootstrap* é uma técnica de amostragem com reposição, a partir do conjunto de treinamento inicial, no qual são selecionadas amostras aleatoriamente para um novo subconjunto de treinamento.

Na amostra *bootstrap*, 63% das observações originais podem ocorrer pelo menos uma vez. Os 37% dos dados que não ocorrem no *bootstrap* são chamados de out-of-bag (OOB) (CUTLER et al., 2007). A árvore de decisão é ajustada para cada amostra *bootstrap*, porém cada nó possui um valor m fixo de variáveis. A árvore é cultivada em sua maior extensão possível, sem poda. O desenho esquemático do funcionamento do Random Forest é apresentado na Figura 2.

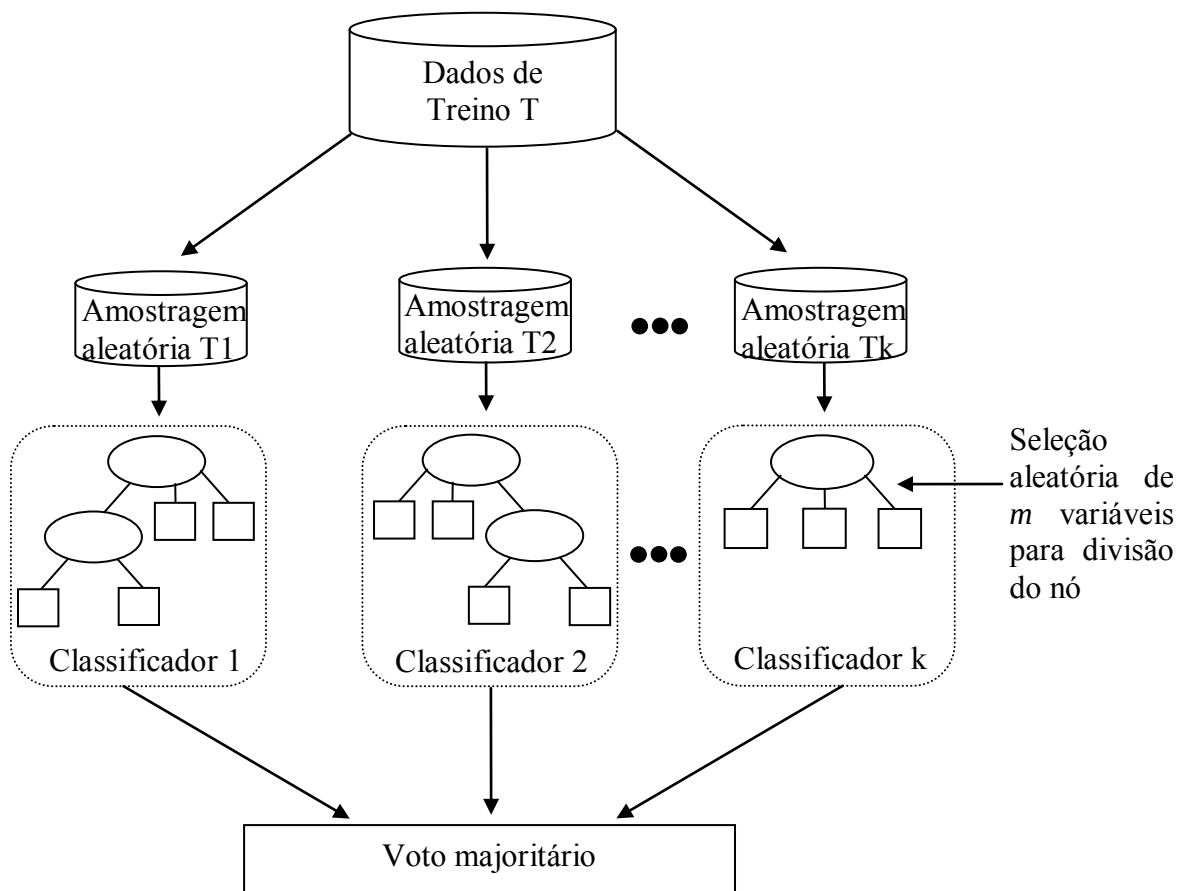


Figura 2 - Funcionamento do método Random Forest. Adaptado de OSHIRO (2013).

Os dados OOB são usados para obter uma estimativa imparcial para o erro de generalização do conjunto combinado de árvores, além de estimativas de importância das variáveis (BREIMAN, 2001). As estimativas de importância das variáveis são feitas de duas formas conhecidas como *Mean decrease accuracy* (MDA) e *Mean decrease Gini* (MDG).

O MDA é uma medida baseada na permutação da variável, onde os verdadeiros valores da variável são substituídos por valores gerados aleatoriamente para cada árvore no conjunto, avaliando o impacto desta troca na classificação. Se não houver nenhum impacto sobre o erro da árvore a importância da variável diminui (BREIMAN, 2001).

O MDG baseia-se na qualidade de cada nó em uma variável de uma árvore de decisão. Quando uma determinada variável é utilizada para dividir um nó, é

calculado o coeficiente de Gini para os nós descendentes comparando-se com o nó original. As alterações na Gini são somadas para cada variável e normalizado no final do cálculo. Uma variável que produz uma alta homogeneidade nos nós descendentes resulta em um alto valor de Gini (BREIMAN, 2001).

O método Random Forest ainda é pouco explorado em pesquisas nacionais de mapeamento digital de solos se comparado a outros métodos, com destaque para trabalhos recentes como o dos autores Bhering et al. (2015) e Chagas et al. (2015) no mapeamento de atributos do solo, e Pinheiro et al. (2015) no mapeamento de classes de solo.

Na literatura internacional, são encontrados poucos trabalhos que utilizam esta técnica para a predição de classes de solo, entre eles, destacam-se os trabalhos de Stum (2010), Häring et al. (2012) e Heung et al. (2014).

Stum (2010) realizou seu trabalho em uma bacia hidrográfica árida e semi-árida do oeste de Utah, utilizando amostras coletadas em campo, imagens Landsat 7 Enhanced Thematic Mapper Plus (ETM+) e variáveis derivadas de MDE (10 metros de resolução espacial), sendo estas, as que apresentaram uma maior força de predição. Para o autor, o Random Forest funcionou como um poderoso preditor de classes de solo e seus resultados facilitaram ainda mais a compreensão das relações solo-paisagem.

Häring et al. (2012) utilizaram o Random Forest para a desagregação espacial das unidades de um mapa de solos do estado da Baviera na Alemanha, com o intuito de refinar as unidades de mapa de solo em que dois ou mais tipos de solo são associados. Os autores chegaram à conclusão de que as variáveis geomorfométricas tiveram uma influência significativa nas relações solo-paisagem, e que podem ser utilizadas para desagregação de associações.

Na pesquisa de Heung et al. (2014), o método Random Forest foi utilizado na compreensão da relação entre o material de origem e a topografia, utilizando variáveis derivadas do MDE (100 metros de resolução espacial), pontos de treinamento e validação aleatórios obtidos de um mapa convencional de solos, além de amostras de campo também obtidas para validação. Os resultados dos autores mostraram o método Random Forest como eficaz em abordagens de aprendizado de máquina e mineração de dados.

2.6. Avaliação da acurácia

A acurácia pode ser definida como o grau de conformidade entre o mapa de solos e a realidade. A precisão pode ser definida como a medida da incerteza associada ao conjunto de procedimentos usados para mapear uma determinada realidade. A medida da acurácia serve para validar ou invalidar o mapa de solos, enquanto a medida de precisão é um subproduto do método do mapeamento de solos oferecendo o erro de predição (FINKE, 2007).

A qualidade de um mapa de solos pode ser avaliada pela comparação entre o observado e o predito. Esta comparação serve como resultado para validação de uma determinada classe de solo ao nível de observação, ou seja, se é verdadeiro ou falso. O ideal é que uma amostra seja retirada onde haja a ocorrência de todas as classes de solos. Os resultados são armazenados em uma matriz de confusão, que é a base multinominal para testes (FINKE, 2007).

Tabela 2 - Matriz de confusão com proporções de observações dentro de c classes preditas i e classes observadas j .

		Classes observadas			Total
		$j = 1$	...	$j = c$	
Classes preditas	$i = 1$	$P_{1,1}$	...	$P_{1,c}$	$P_{1+} = \sum_{j=1}^c P_{1j}$
	...	...	P_{ij}	...	$P_{i+} = \sum_{j=1}^c P_{ij}$
	$i = c$	$P_{c,1}$	...	$P_{c,c}$	$P_{c+} = \sum_{j=1}^c P_{cj}$
	Total	$P_{+1} = \sum_{i=1}^c P_{i1}$	$P_{+j} = \sum_{i=1}^c P_{ij}$	$P_{+c} = \sum_{i=1}^c P_{ic}$	$\sum_{i=1}^c \sum_{j=1}^c P_{ij} = 1$

De acordo com Congalton & Green (2009), uma matriz de confusão compara informações de pontos de referência com informações do mapa. Compreende a uma matriz quadrada com números estabelecidos de linhas e colunas. As colunas são assumidas como corretas e as linhas são usadas para exibir as classes ou dados

classificados. Com a matriz de confusão gerada adequadamente, é possível aplicar técnicas básicas para realizar uma avaliação de precisão, tais como o índice Kappa e exatidão global.

A exatidão global é a razão entre a soma das células da diagonal p_{ii} (número de acertos) e o total de amostras n , sendo expressa pela fórmula

$$\theta_1 = \frac{\sum_{i=1}^c p_{ii}}{n} \quad (3)$$

A matriz de confusão apresenta as precisões individuais de cada categoria no mapa, que são descritas pelos erros de inclusão (comissão) e erros de exclusão (omissão), representando de maneira eficaz a precisão do mapa. O erro de comissão ocorre quando uma área está incluída em uma categoria incorreta. Já o erro de omissão ocorre quando uma área é excluída da categoria que pertence. Cada erro no mapa é uma omissão da categoria correta e uma comissão da categoria incorreta (CONGALTON & GREEN, 2009).

A exatidão do usuário refere-se ao erro de comissão e a exatidão do produtor refere-se ao erro de omissão. Essas formas de exatidão são maneiras de representar a precisão individual das categorias ao invés de apresentar apenas a exatidão global da classificação (CONGALTON & GREEN, 2009). A exatidão do usuário é obtida pela fórmula

$$\frac{P_{ii}}{P_{i+}} \quad (4)$$

onde P_{ii} são as diagonais e P_{i+} representa a soma da linha i . A exatidão do produtor é obtida pela fórmula

$$\frac{P_{jj}}{P_{+j}} \quad (5)$$

onde P_{jj} é a são as diagonais e P_{+j} é a soma da coluna j .

O índice Kappa é uma técnica multivariada discreta utilizada na avaliação de precisão para determinar estatisticamente se uma matriz de erro é diferente de

outra. Esta medida se baseia na diferença entre o acordo real e o acordo de chance que é indicado pelos totais das linhas e colunas (CONGALTON & GREEN, 2009).

A fórmula para o índice Kappa pode ser expressa da seguinte forma:

$$\theta_2 = \sum_{i=1}^c P_i + P + i \quad (6)$$

$$K = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (7)$$

onde K é o índice Kappa, θ_1 representa a exatidão geral, c é o número de linhas da matriz e $P+i$ é a soma da coluna i .

Landis & Koch (1977) propuseram uma classificação de escala do índice Kappa, que varia de mapas com ausência de concordância (< 0) ou com total concordância ($= 1$) (Tabela 3).

Tabela 3 - Escala de classificação do valor K.

Valor K	Grau de acordo
< 0	Pobre
$0 - 0,20$	Fraco
$0,21 - 0,40$	Bom
$0,41 - 0,60$	Moderado
$0,61 - 0,81$	Forte
$0,81 - 1,00$	Quase perfeito

A diferença entre a exatidão global e o índice Kappa é que a primeira incorpora apenas a diagonal, excluindo os erros de omissão e comissão, enquanto a segunda inclui diretamente os elementos de fora da diagonal. Dependendo do erro incluído na matriz, essas medidas podem ser discrepantes (CONGALTON & GREEN, 2009).

2.7. Escala efetiva

A escala efetiva é uma medida de precisão que demonstra se a complexidade dos padrões representados no mapa corresponde a escala dita nominal, ou seja, a escala de apresentação. Quanto maior é a escala, mais detalhado é o mapa e maior a possibilidade de seus delineamentos apresentarem maior exatidão e precisão (FINKE, 2007; OLIVEIRA, 2005).

A escala efetiva (*effective scale number* – ESN) é calculada de acordo com a equação proposta por Forbes et al. (1987):

$$ESN = NSN \times \frac{IMR}{2} \quad (8)$$

onde *IMR* é o índice de máxima redução e *NSN* é a escala nominal (*nominal scale number* - NSN). O *IMR* é calculado por meio da seguinte equação:

$$IMR = \sqrt{\frac{ASD}{MLD}} \quad (9)$$

onde *ASD* é o tamanho médio de delimitação (*average size delineation* – ASD) e *MLD* é a área mínima mapeável (*minimum legible delineation* - MLD), usualmente sendo utilizado o valor de 0,4 cm² (FINKE, 2007; IBGE, 2007).

O *ASD* e o *MLD* (em hectares) são obtidos pelas seguintes equações:

$$ASD = \sqrt{\frac{\sum_{j=1}^m A_j}{m}} \quad (10)$$

$$MLD = \frac{NSN^2 \times 0,4}{10^8} \quad (11)$$

onde *A_j* é a área do j-ésimo polígono, *m* é o total de polígonos contidos no mapa.

3. MATERIAIS E MÉTODOS

3.1. Caracterização da área de estudo

3.1.1. Localização

A área de estudo compreende a quadrícula de Ribeirão Preto do mapa pedológico semidetalhado na escala 1:100.000, elaborado por Oliveira & Prado (1987). Situada no nordeste do estado de São Paulo, a quadrícula abrange 4 folhas das cartas topográficas do IBGE na escala 1:50.000, conforme a Figura 3, sendo elas as folhas de Ribeirão Preto (SF-23-V-C-I-1), Serrana (SF-23-V-C-I-2), Bonfim Paulista (SF-23-V-C-I-3) e Cravinhos (SF-23-V-C-I-4).

A quadrícula de Ribeirão Preto possui uma área de aproximadamente 287.557 hectares, sendo circunscrita pelas coordenadas 7675764 de latitude norte e 7619488 de latitude sul e 240941 de longitude oeste e 188060 de longitude leste, pela projeção Universal Transversa de Mercator (UTM), datum SIRGAS 2000. Os municípios total ou parcialmente inclusos na quadrícula são: Altinópolis, Barrinha, Batatais, Brodowski, Cravinhos, Dumont, Guatapar, Jardinpolis, Lus Antnio, Pontal, Pradpolis, Ribeir Preto, Santa Cruz da Esperana, So Simo, Serra Azul, Serrana e Sertozinho.

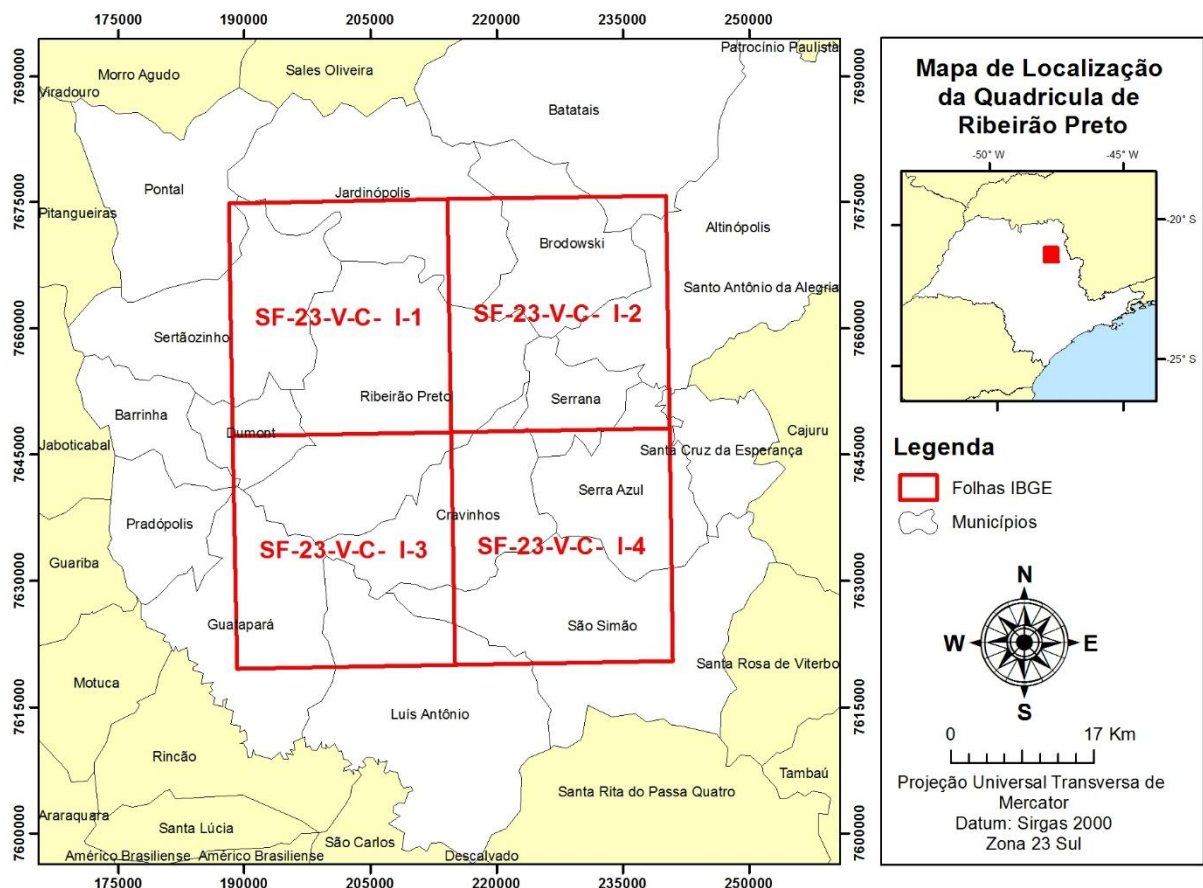


Figura 3 - Mapa de localização da quadricula de Ribeirão Preto.

3.1.2. Geologia

Na quadricula de Ribeirão Preto encontramos quatro unidades litoestratigráficas segundo Oliveira & Prado (1987):

1. Grupo São Bento: compostas pela Formação Serra Geral, Formação Botucatu e Formação Pirambóia.
2. Sedimentos correlatos à Formação Itaqueri
3. Sedimentos aluvionais
4. Sedimentos continentais indiferenciados

A Formação Serra Geral é constituída de rochas basálticas originadas a partir de atividades magmáticas fissurais. A espessura dessa formação decresce de oeste para leste, com espessura média de 65m na região de Ribeirão Preto, podendo também desaparecer. Entre Ribeirão Preto e Cássia dos Coqueiros se encontra o

afloramento do Aquífero Guarani, que devido as altas cotas altimétricas (800m), constituem áreas de recarga do aquífero no estado de São Paulo (VILLAR, 2008).

A Formação Botucatu tem origem na deposição de arenitos através de processos eólicos em ambiente desértico, sendo estes de granulação fina e média, não ultrapassando a espessura de 80 m na região de Ribeirão Preto. O teor de silte e argila é inferior a 10% (FERREIRA, 1992; SILVA et al., 2008).

A Formação Pirambóia é de origem fluvial, sendo constituída por arenitos intercalados com camadas de siltito e folhelho, com arenitos de granulação média a muito fina com 20% de teor de silte e argila. Na rodovia SP-333, que liga Ribeirão Preto a Cajuru, as formações apresentam boas exposições, com espessura máxima na região de 140 m (FERREIRA, 1992; SILVA et al., 2008). Esta formação possui posição litoestratigráfica inferior a formação Botucatu, por isso ocorre em cotas mais baixas, ficando em posição intermediária entre esta formação e a Serra Geral (ALMEIDA et al., 1981; OLIVEIRA & PRADO, 1987).

Os Sedimentos correlatos a formação Itaqueri ocorrem na região de Franca, compondo uma sequência de arenitos, siltitos e conglomerados oligomíticos. Localiza-se em uma pequena mancha ao norte da quadrícula, próximo à rodovia que vai de Ribeirão Preto a Brodowski (OLIVEIRA & PRADO, 1987).

Os Sedimentos aluvionais ou aluviões quaternários são depósitos colúvio-eluviais recentes, correspondentes a sedimentos areno-argilosos incoesos, não apresentando estruturas sedimentares. São desenvolvidos sob substrato arenoso. Ocorrem nas planícies aluviais dos principais cursos d'água da quadrícula (KOTCHETKOFF-HENRIQUES, 2003; OLIVEIRA & PRADO, 1987; VILLAR, 2008).

Os Sedimentos continentais indiferenciados ou Formações superficiais são compostas por areias inconsolidadas, que são sobrepostas à Formação Serra Geral e à Formação Botucatu, caracterizando-se por sedimentos Cenozóicos arenosos passando a areno-argilosos, com uma espessura da ordem de 20 metros (SINELLI, 1971).

O mapa da Figura 4 mostra a mapa geológico da quadrícula de Ribeirão Preto de acordo com as unidades geológicas. A Tabela 4 apresenta as unidades, sua litologia, idade geológica a área de cada classe em hectares e porcentagem.

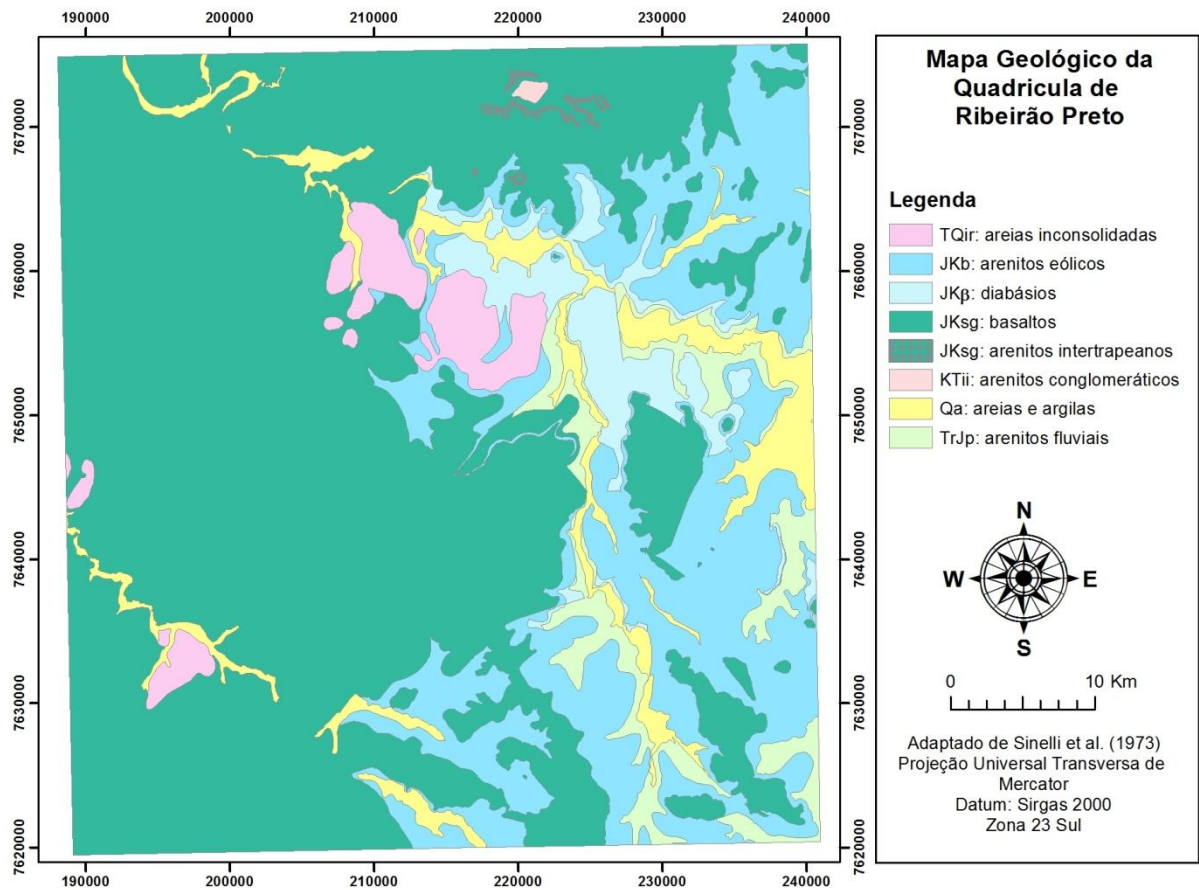


Figura 4 - Mapa geológico da quadrícula de Ribeirão Preto (adaptado de Sinelli et al., 1973). Sedimentos Aluvionares (Qa) – areias e argilas; Sedimentos Correlatos a Formação Itaqueri (KTii) – arenitos conglomeráticos; Formação Botucatu (JKb) – arenitos eólicos; Formação Pirambóia (TrJp), – arenitos fluviais; Formação Serra Geral (JKsg) – Basaltos e intercalações de arenitos intertrapeanos; Formações Superficiais (TQir) – areias consolidadas e Intrusivas Básicas (sill) (JK β) – diabásios.

Tabela 4 – Unidades geológicas, litologia, idade e área em hectares e porcentagem das respectivas classes no total da quadrícula.

Unidade	Litologia	Idade	Área (hectares)	Área (%)
Formação Serra Geral	Basaltos	Cretáceo inferior - Jurássico superior	179323.3734	62.36%
Formação Serra Geral	Arenitos interderrames	Cretáceo inferior - Jurássico superior	406.4210451	0.14%
Aluviões Quaternários	Areias e argilas Arenitos fluviais, imaturos	Quaternário	15904.47668	5.53%
Formação Piramboia		Jurássico - Triássico	11836.39629	4.12%
Intrusivas Básicas (sill)	Diabásios	Mesozóico	11707.84749	4.07%
Formações Superficiais	Areias inconsolidadas	Cenozóico	9529.548933	3.31%
Formacao Botucatu	Arenitos eólicos	Cretáceo inferior - Jurássico	58589.08757	20.37%
Sedimentos Correlatos a Formação Itaqueri	Arenitos conglomeráticos	Cretácio superior	259.8632668	0.09%
Total			287557.0147	100.00%

3.1.3. Relevô

Predomina na quadrícula de Ribeirão Preto um relevo ondulado e suave ondulado, constituído por uma sucessão de morros de topo subaplainado ou arredondado, com alguns casos ultrapassando 3000 metros de comprimento, com declividade entre 3 e 10%. Ocorrem na paisagem em meio a superfície suave ondulada, formas isoladas ou plataformas interflúvias amplas e contínuas de pequenos testemunhos tabuliformes (Figura 5). Nessa forma de relevo, há a ocorrência de solos rasos em suas bordas (OLIVEIRA & PRADO, 1987; SINELLI et al., 1980).



Figura 5 – Vista do relevo testemunho tabuliforme.

Os pequenos testemunhos tabuliformes, também evidenciados na área de estudo de Ten Caten (2008), apresentam uma maior resistência aos processos de intemperismo, destacando-se na paisagem na medida em que sua encosta retrocede, apresentando assim, valores elevados de declividade.

O limite entre as bacias do rio Pardo ao norte e a do Moji-Guaçu ao sul, ocorre na quadrícula no sentido NO-SE, formando uma plataforma fluvial indo de Dumont a Cravinhos, como pode ser observado na Figura 6.

Ao longo deste divisor de águas ocorrem solos com alto teor de areia grossa quartzosa e teores de Fe_2O_3 baixo. Na vertente sul desta plataforma fluvial e na porção nordeste da quadrícula o relevo é bastante movimentado predominando solos rasos. A sudeste na abrangência do município de Bonfim Paulista e em direção ao município de Cravinhos, a altitude se eleva rapidamente, atingindo um máximo em torno de 900 metros. (OLIVEIRA & PRADO, 1987).

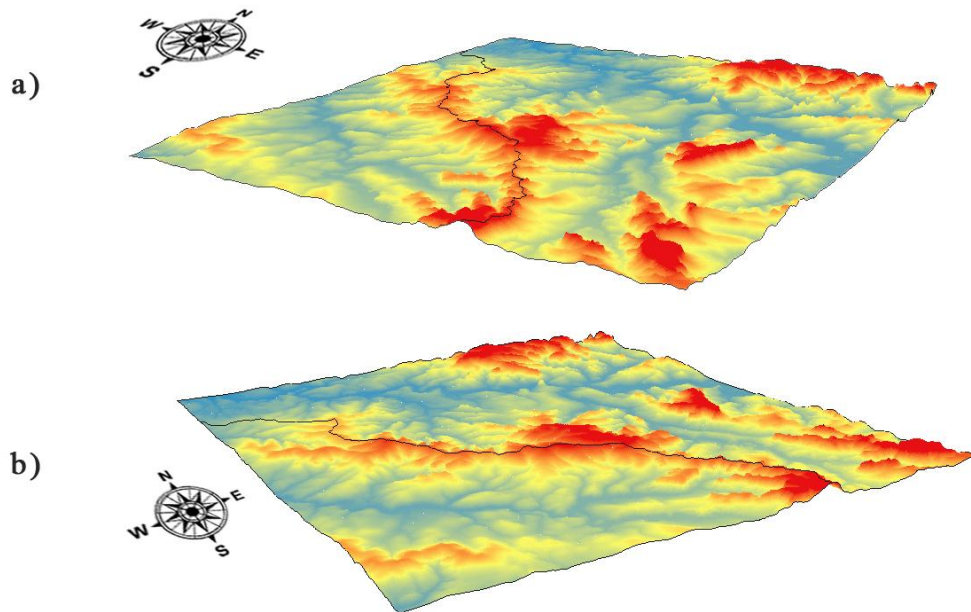


Figura 6 - Modelo tridimensional da quadrícula de Ribeirão Preto: a) visada NO-SE; b) visada NE-SO.

De acordo com Oliveira & Prado (1987), o relevo da quadrícula de Ribeirão Preto tem grande importância nos fenômenos de abrasão, transporte e sedimentação. Ainda de acordo com os autores, são identificados 3 ambientes particulares de pedogênese:

- 1- **Planícies aluvionais:** zonas de deposição alúvio-coluvial, na qual os processos pedogenéticos são condicionados em meio pouco oxigenado, desenvolvendo uma morfologia típica de ambiente redutor.
- 2- **Áreas dissecadas de relevo acentuado:** caracteriza-se pelas ações de rejuvenescimento e possui pedoclima seco. Predominam processos de transformação e perda.
- 3- **Áreas de relevo suave ondulado:** possui pedoclima úmido, com maior quantidade de água percolando no solo e portanto, dando origem a solos mais profundos. Predominam processos de transformação e adição.

3.1.4. Clima, vegetação e hidrografia

O clima da quadrícula segundo a classificação climática de Koppen enquadra-se nos tipos Aw e Cwb. O clima Aw, é do tipo tropical apresentando verão chuvoso e inverno seco, com média do mês mais frio superior a 18°C. Este clima é predominante na quadrícula, e situa-se nas partes mais baixas (500m a 700m). O clima Cwb é do tipo temperado, apresentando verão chuvoso e inverno seco, com temperatura do mês mais quente inferior a 22°C. Situa-se nas partes mais elevadas da quadrícula, na região serrana (OLIVEIRA & PRADO, 1987).

De acordo com Oliveira & Prado (1987), o balanço hídrico de Ribeirão Preto é válido para maior parte da quadrícula, onde predominam as partes mais baixas do relevo, enquanto o balanço hídrico de São Simão representa as áreas mais elevadas da paisagem regional.

A Figura 7 e Figura 8 que mostram o balanço hídrico dos municípios de Ribeirão Preto e São Simão, evidenciando que nos dois municípios não há uma grande variação, o que caracteriza o clima da quadrícula como homogêneo.

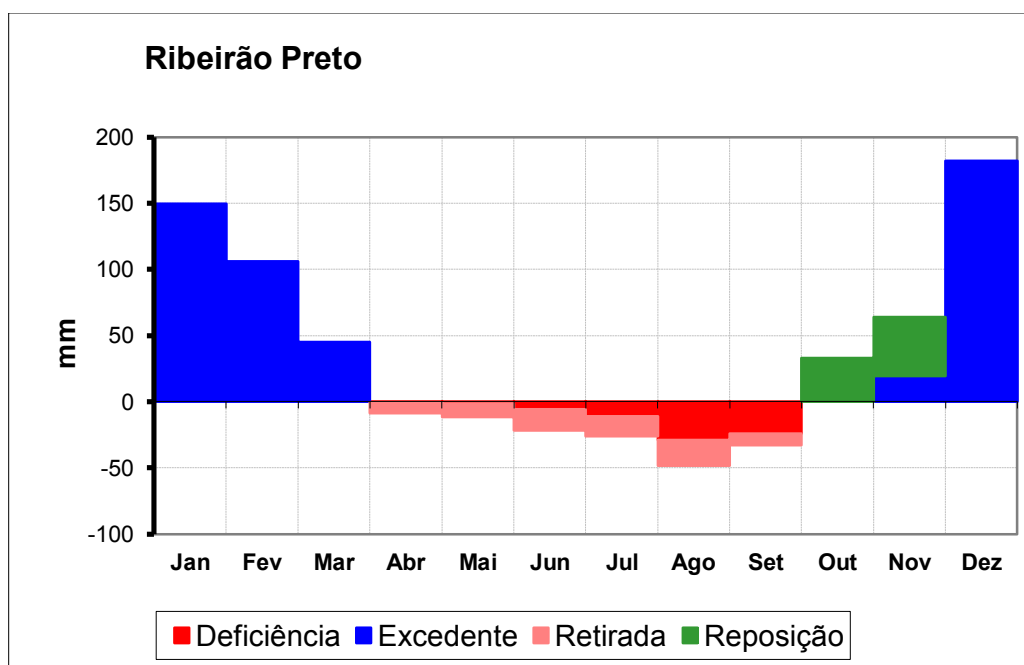


Figura 7 - Balanço hídrico do município de Ribeirão Preto. Fonte: Sentelhas et al. (1999).

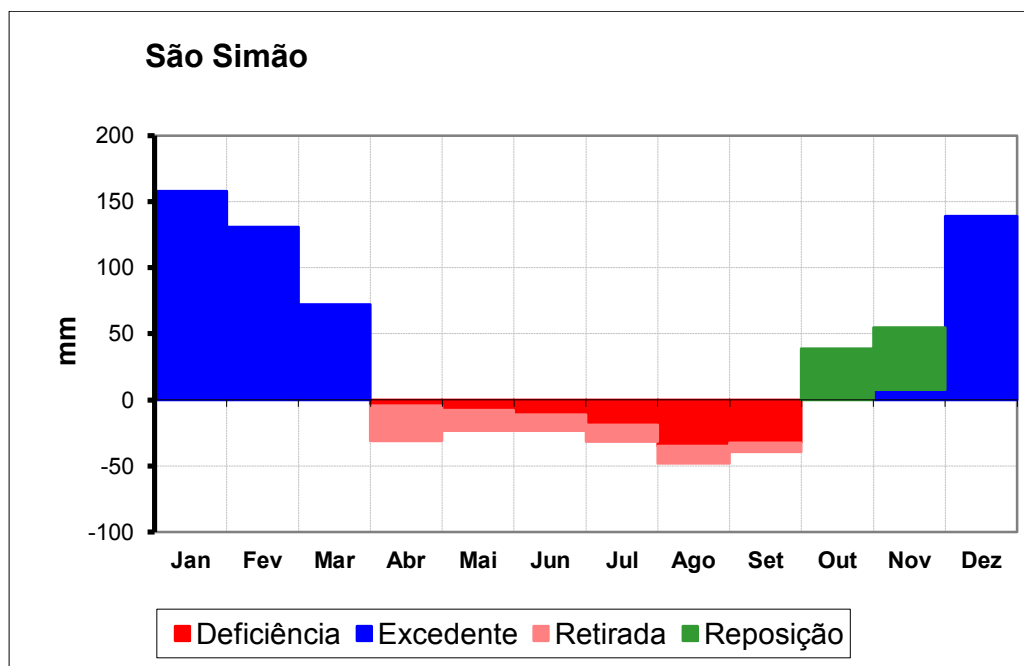


Figura 8 - Balanço hídrico do município de São Simão. Fonte: Sentelhas et al. (1999).

A vegetação atual da quadrícula de Ribeirão Preto é composta basicamente por extensas plantações de cana-de-açúcar, culturas de café, citrus, cultivos anuais (milho, feijão e soja), pastagens e reflorestamento. Encontram-se também fragmentos remanescentes de vegetação primitiva, como cerrados, cerradões e campos higrófilos (KOTCHETKOFF-HENRIQUES, 2003; OLIVEIRA & PRADO, 1987).

Desde 1975 houve uma expansão das áreas cultivadas com cana de açúcar na região de Ribeirão Preto, município onde encontram-se as maiores fazendas de cana-de-açúcar e usinas que dispõe de solos de boa qualidade. O município localizado no centro das plantações de cana de açúcar tornou-se referência no setor sucroalcooleiro, considerada a “capital do agrobusiness” (KOHLHEPP, 2010).

Segundo dados do Canasat-Área de Rudorff et al. (2010), 53% da área dos 17 municípios que compõem a quadrícula de Ribeirão Preto são ocupadas por cana-de-açúcar. Os dados do Canasat-Colheita de Aguiar et al. (2011) mostram que sobre 52% dessa área cultivada com cana é realizada a colheita crua, ou seja, com práticas de manejo de plantio direto.

Observando os dados descarta-se a hipótese, no presente trabalho, sobre a utilização de índices de vegetação, pois de acordo com Ten Caten (2011), o plantio direto compromete o sinal refletido do solo para os sensores orbitais, dificultando a obtenção de dados confiáveis.

A hidrografia da quadrícula de Ribeirão Preto é composta por uma densa rede de rios, córregos, açudes e represas ou lagos. Os destaques são os rios Pardo, Tamanduá, Onça e Pântano, além de córregos como Guataparazinho, Serra Azul e Espraiado (OLIVEIRA & PRADO, 1987).

O mapa da Figura 9 mostra os principais cursos hídricos e o limite do divisor de águas entre as bacias hidrográficas do rio Pardo e do rio Mogi-Guaçu.

Fica evidente no mapa a ocorrência de um paralelismo entre os rios principais das bacias hidrográficas do rio Mogi-Guaçu e do rio Pardo, que apresentam uma retilinidade exibida por longos segmentos em seus percursos. Esta rede hidrográfica que engloba o médio rio Pardo ao norte, e o Ribeirão da Onça e Mogi-Guaçu ao sul condiciona o relevo ao entalhamento realizado por estes três cursos principais, onde correm de maneira quase perpendicular, seus afluentes (SINELLI et al., 1980).

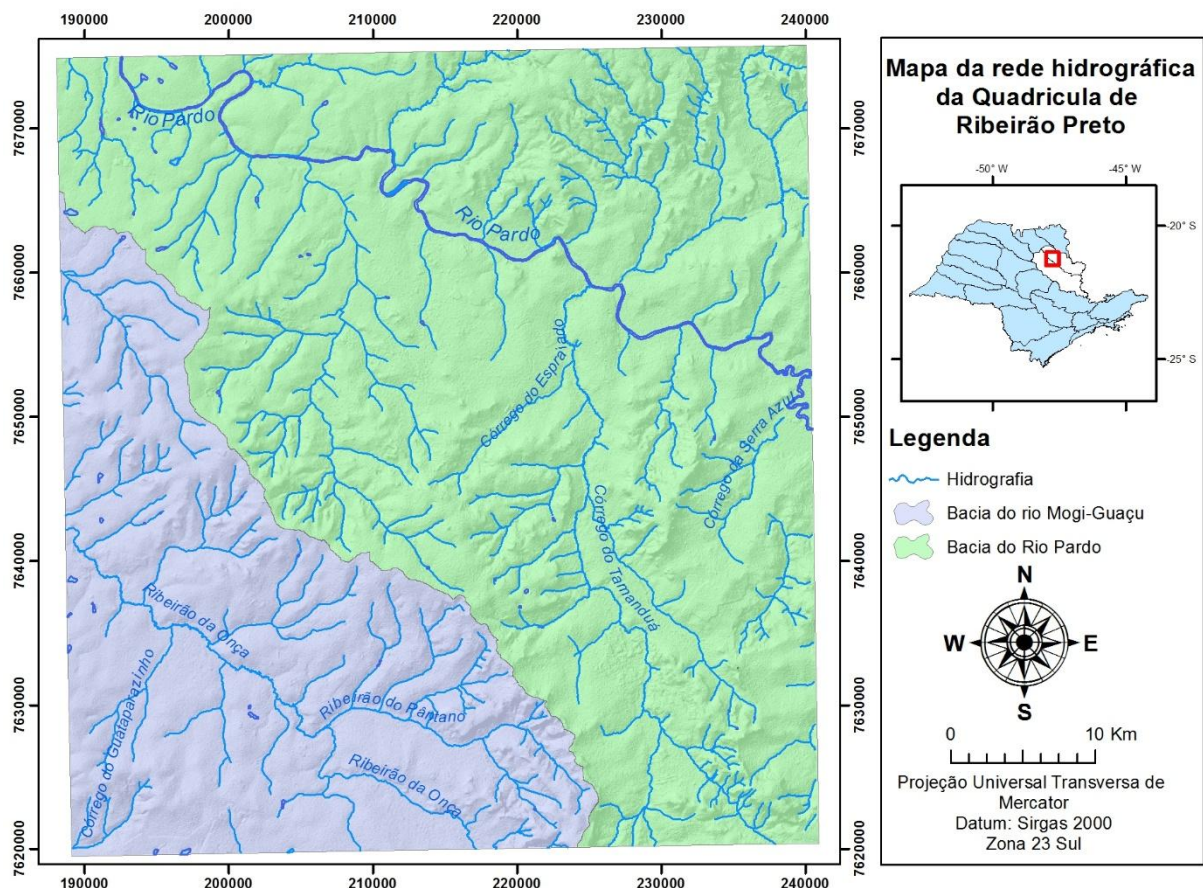


Figura 9 - Mapa da hidrografia da quadrícula de Ribeirão Preto.

3.2. Materiais utilizados

3.2.1. Base cartográfica digital

O presente estudo baseia-se no modelo *s.c.o.r.p.a.n* proposto por McBratney et al. (2003), onde os fatores são obtidos por diferentes fontes de dados.

O fator *s* (informações sobre solos) foi obtido do Boletim Científico e do mapa pedológico semidetalhado (escala 1:100.000) da quadrícula de Ribeirão Preto de Oliveira & Prado (1987). O levantamento realizado por estes autores contou com a observação de 1570 locais, com coleta de material de uma ou mais camadas para análise em 495 pontos. Desconsiderando a área representada pela mancha urbana, rios e açudes, obtém-se uma densidade de um ponto observado a cada 1,7 km².

Foram identificadas 24 unidades no levantamento semidetalhado. As unidades de mapeamento que apresentam associações de duas classes, conforme definido por Oliveira & Prado (1987), contém 60% da primeira e 40% da segunda. As unidades que apresentam associações de três classes contém 50% da primeira, 30% da segunda e 20% da terceira.

As informações dos solos da quadrícula de Ribeirão Preto foram adequadas ao novo Sistema Brasileiro de Classificação de Solos (SiBCS) da Embrapa (2013). A adequação das classes do mapa pedológico semidetalhado da quadrícula de Ribeirão Preto foi realizada com base nas informações dos perfis de solos contidos no Boletim Científico. Em unidades com associações de classes, foi escolhida a primeira classe da legenda, ou seja, a classe com maior porcentagem. Em casos de associações entre tipos de terreno e classes de solos, foram desconsiderados os tipos de terreno.

De acordo com Zhu et al. (2001), a resolução é ditada pela qualidade da base de dados digital, sendo assim, sendo necessário uma avaliação da qualidade do mapa de solos em questão. A medida utilizada foi o cálculo da escala efetiva, proposta por Forbes et al. (1987), baseando-se nos polígonos encontrados no mapa.

Para o fator *r* (relevo), utilizou-se uma imagem de Modelo Digital de Elevação SRTM (Shuttle Radar Topography Mission), com resolução espacial de 30 metros. A imagem foi obtida no sistema Earth Explorer do *United States of Geological Survey* (U.S.G.S). Para o fator *p* (material de origem), utilizou-se o mapa geológico de Sinelli et al. (1973) na escala 1:50.000. Foram também utilizadas no trabalho cartas topográficas do IBGE, para extração de rodovias, hidrografia e área urbana.

Optou-se por utilizar apenas atributos do relevo juntamente com a litologia, desconsiderando o fator *c* (clima), por conta da homogeneidade climática da quadrícula. Também foi desconsiderado o fator *o* (organismos) devido à grande alteração antrópica em decorrência de intensas atividades agropecuárias.

O mapa pedológico, geológico e as cartas topográficas foram digitalizados. Para organização de toda base cartográfica digital, foi utilizado o sistema de informação geográfica (SIG) ArcMap 10.2.2 (ESRI, 2014), no qual foram criados dois bancos de dados geográficos (*geodatabase*) contendo arquivos vetoriais e matriciais necessários para a execução do trabalho. O primeiro banco reuniu arquivos brutos,

ainda não processados, e o segundo reuniu os arquivos prontos para serem utilizados.

Em um primeiro momento, para que os mapas pedológico, geológico e as cartas topográficas apresentassem sua posição espacial real de acordo com o sistema de referência indicado, obedecendo ao fator n (posição espacial), executou-se o processo de georreferenciamento. De acordo com Piroli (2010), o georreferenciamento consiste em registrar espacialmente na imagem a sua posição conhecida no terreno, em coordenadas de referência reconhecidas tais como a Universal Transversa de Mercator (UTM).

Após georreferenciadas, as feições espaciais dos mapas em formato raster foram vetorizadas e as informações contidas nas legendas foram inseridas na tabela de atributos das feições vetoriais.

Os arquivos vetoriais do mapa geológico, mapa pedológico e das cartas topográficas passaram por uma conversão de seus respectivos sistemas de referência de coordenadas para o datum Sirgas 2000, que se tornou o sistema geodésico de referência oficial no Brasil a partir de fevereiro de 2015.

Das cartas topográficas do IBGE foram vetorizadas as rodovias em formato linha, a hidrografia, extraíndo polígonos para representação de rios com largura considerável, represas e lagos, além de linhas para representação de rios com menor largura. A área urbana dos municípios também foi vetorizada em forma de polígono.

3.2.2. Obtenção das variáveis

O MDE da imagem SRTM é um arquivo matricial (raster), onde o valor de cada pixel da grade corresponde a altitude. Para ser utilizado, o arquivo raster foi submetido a uma remoção de depressões espúrias, utilizando-se a ferramenta “*Fill*”, do módulo Spatial Analyst. Esta ferramenta remove picos anômalos e corrige áreas com ausência de dados (Figura 10).

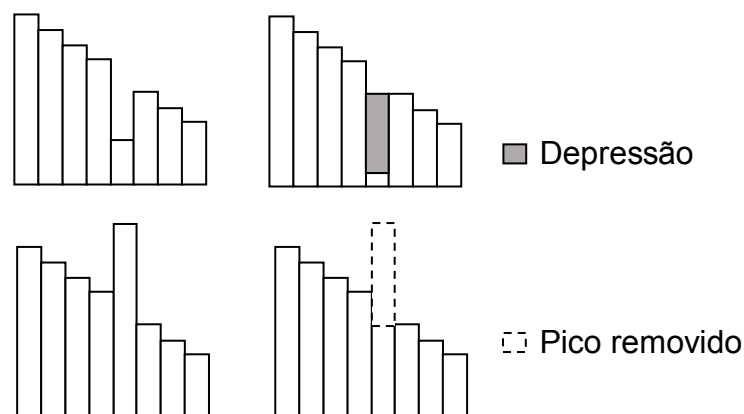


Figura 10 - Funcionamento da ferramenta *Fill*.

A partir da imagem MDE corrigida, foram extraídas as variáveis primárias Aspecto, Curvatura e Declividade, criados respectivamente com as ferramentas *Aspect*, *Curvature* e *Slope*, do módulo *Surface*.

A variável Aspecto, também chamado de direção da inclinação, exerce influência na quantidade e intensidade da radiação solar em que uma localização é exposta. O regime de temperatura afeta os processos biológicos e químicos do solo, bem como a evaporação. A variável Aspecto geralmente é incorporado em modelos preditivos onde a extensão espacial da predição abrange grandes áreas com inúmeras encostas e inclinações (CARRÉ; MCBRATNEY, 2005; THOMPSON et al., 2012).

A Declividade corresponde à inclinação da superfície, sendo a relação entre o ângulo em um plano horizontal e o plano tangente, sendo expressa em graus ou porcentagem. A declividade tem importância no regime hídrico do solo, o fluxo de água superficial e subsuperficial, o impacto e a velocidade de escoamento e na infiltração de água da chuva. É considerada uma das mais importantes variáveis primárias que controlam os processos pedogenéticos (WILSON & GALLANT, 2000).

A Curvatura é uma variável que tem importância na medida do fluxo de água local, em formas côncavas e convexas do relevo. É útil para interpretação de processos que ocorrem em vertentes, como a redistribuição de água no solo por meio de fluxo e infiltração vertical, conforme pode ser observado na Figura 11 (THOMPSON et al., 2012).

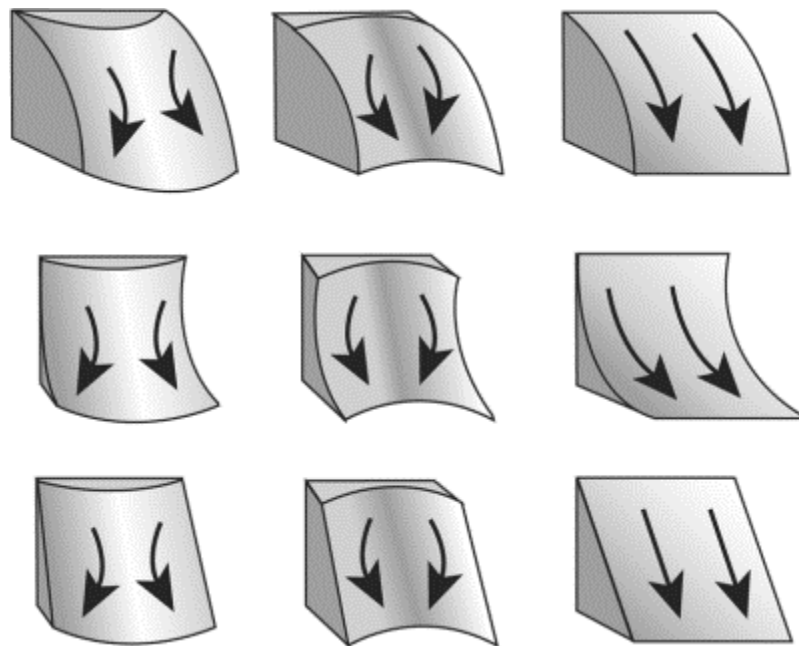


Figura 11 - Combinações de curvatura (ESRI, 2014).

As variáveis geomorfométricas secundárias são derivadas de combinações das variáveis primárias. As variáveis secundárias utilizadas no trabalho foram o Stream Power Index (SPI) e o Compound Topographic Index (CTI).

O SPI é a medida de poder erosivo da água baseado na suposição que o fluxo de descarga é proporcional a uma área específica de influência. Tem sido utilizado em estudos de geomorfologia, transporte de sedimentos e erosão (WILSON & GALLANT, 2000).

O CTI é calculado como o logaritmo natural da relação entre área específica de influência sobre a inclinação da vertente. Este índice foi desenvolvido para ser utilizado em estudo de topossequências de áreas declivosas (THOMPSON et al., 2012).

Para gerar o CTI e o SPI foi necessário antes extrair do MDE variáveis hidrológicas. Utilizando o módulo *Hidrology*, obteve-se a direção de fluxo por meio da ferramenta *Flow Direction*, que posteriormente serviu como arquivo de entrada na ferramenta *Flow Accumulation*, para a obtenção da acumulação de fluxo.

A direção de fluxo consiste em identificar o caminho mais íngreme para o fluxo da água, considerando que a água escoará para uma das 8 células adjacentes

de acordo com a ação da gravidade. O modelo utilizado pelo ArcMap é conhecido como D8. Para calcular a direção de fluxo utiliza-se o MDE como entrada, resultando em um raster que codifica a direção de fluxo de cada pixel (Figura 12).

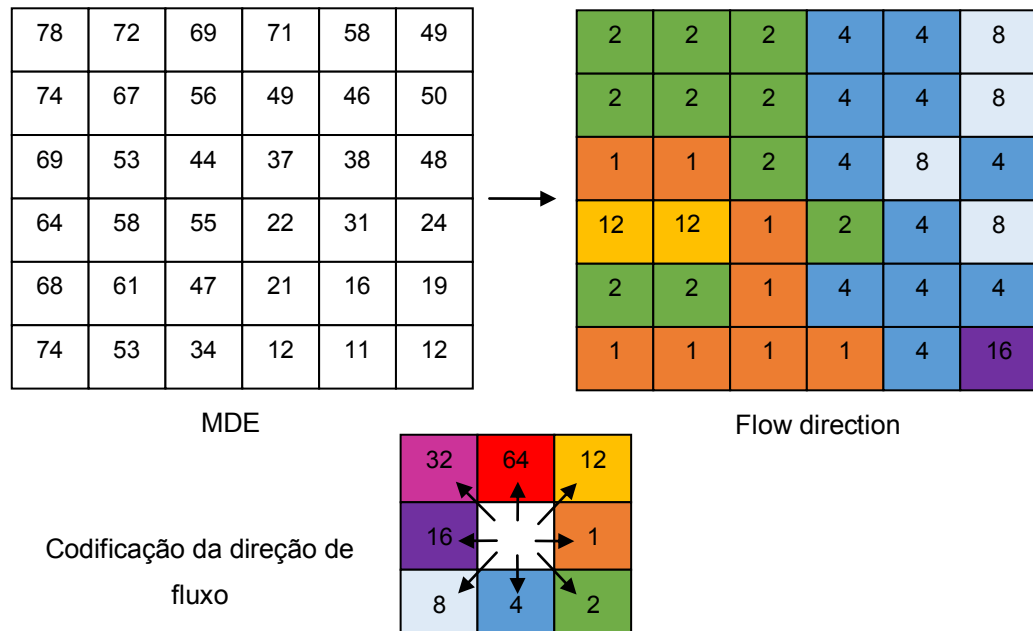


Figura 12 – Exemplo de funcionamento do modelo D8 para o cálculo da direção de fluxo.

A acumulação de fluxo calcula o peso acumulado de cada célula que flui para dentro de outra célula encosta abaixo. Células com elevados valores possuem fluxo concentrado, enquanto células que apresentam o valor 0 são elevações topográficas, podendo ser utilizadas para indicar cumes e divisores de águas. A acumulação de fluxo é calculada pela ferramenta *Flow Accumulation* e utiliza o raster *Flow Direction* como entrada (Figura 13).

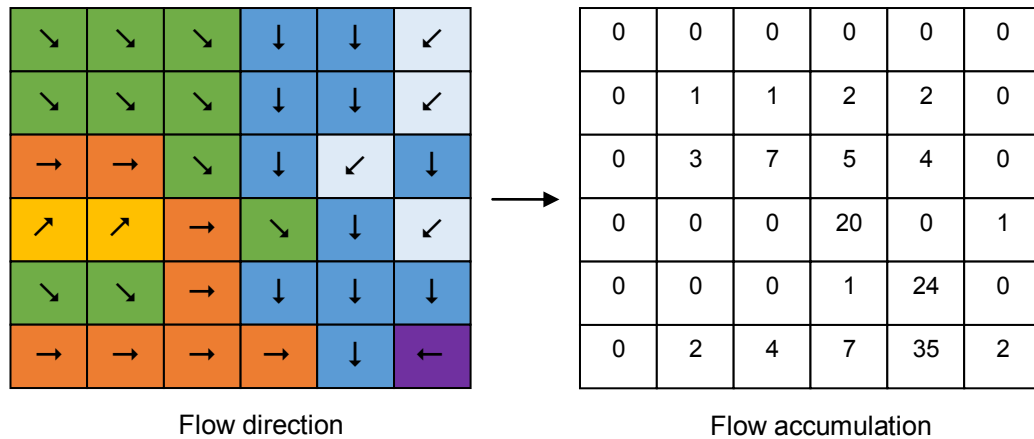


Figura 13 – Acumulação de fluxo.

Para o cálculos das variáveis CTI e SPI foi necessário converter a acumulação de fluxo em área de influência. Para isso, foi calculada a seguinte equação:

$$As = (\text{Flow Accumulation} + 1) * \text{Tamanho do Pixel} \quad (12)$$

onde o tamanho do pixel corresponde a resolução espacial do MDE SRTM (em m²). A declividade obtida em graus precisou ser convertida para radianos, com a seguinte fórmula:

$$(\beta * 1.570796)/90 \quad (13)$$

onde β é a declividade em graus e o valor 1,570796 corresponde a $\pi/2$.

Para obter o CTI utilizou-se a seguinte equação:

$$\text{CTI} = \ln \left(\frac{As}{\tan \beta} \right) \quad (14)$$

onde As é a área de contribuição (em m² m⁻¹) e β é a declividade (em graus) (WILSON & GALLANT, 2000).

O SPI foi obtido com a seguinte equação:

$$\text{SPI} = \ln (As \tan \beta) \quad (15)$$

onde A_s é a área de contribuição (em m^2 por m^{-1}) e β é a declividade (em graus) (WILSON & GALLANT, 2000). As equações foram calculadas com a ferramenta *Raster Calculator*.

A variável distância da rede de drenagem foi obtida pelo cálculo da distância Euclidiana da rede hidrográfica da quadrícula, com a ferramenta *Euclidean Distance*, tendo como arquivo de entrada a hidrografia em formato vetorial. O cálculo da distância Euclidiana é feito a partir do centro da célula de origem para o centro de cada uma das células circundantes, onde o arquivo de saída contém a distância medida a partir de cada célula para a fonte mais próxima

Os valores das variáveis Aspecto, Curvatura, Declividade, Distância da rede de drenagem, Elevação, CTI e SPI foram incorporados com a ferramenta *Extract values to points* à tabela de atributos de um vetor de pontos, contendo um ponto para cada pixel da quadricula (aproximadamente um ponto a cada $900 m^2$). Os mapas geológico e pedológico em formato vetorial foram incorporados a tabela de atributos com ferramenta *Spatial join*.

Dessa forma, obteve-se uma matriz de dados, denominada grid, contendo oito colunas representando as variáveis, duas colunas representando as classes de solo no segundo e terceiro nível categórico, duas colunas com as coordenadas UTM x e y de cada pixel, além de uma coluna de identificação.

A matriz de dados grid apresenta um total de 3.031.857 linhas, que correspondem a um ponto centralizado em cada pixel da quadrícula, desconsiderando-se os pixels dentro de polígonos de áreas urbanas e cursos hídricos, conforme o exemplo da Tabela 5.

Tabela 5 - Trecho da matriz de dados grid.

Aspec	CTI	Curv	Dist	Elev	Lito	Decl	SPI	Classe 2	Classe 3	X	Y	ID
135,00	7,45	0,11	1451,51	772,00	Basaltos e diabásios	1,00	-0,63	LV	LVdf	240059,41	7675735,67	58
315,00	10,99	-0,44	302,40	617,00	Arenitos	2,68	4,87	RQ	RQo	237398,30	7675705,43	89
296,57	10,06	-0,66	332,64	619,00	Arenitos	6,33	5,66	RQ	RQo	237428,54	7675705,43	90
283,67	9,70	-0,22	362,88	625,00	Arenitos	8,95	6,00	RQ	RQo	237458,78	7675705,43	91
287,70	9,43	-0,66	393,12	631,00	Arenitos	11,53	6,25	RQ	RQo	237489,02	7675705,43	92
287,02	9,38	-0,11	423,36	640,00	Arenitos	11,96	6,28	RQ	RQo	237519,26	7675705,43	93
303,31	9,27	-0,66	453,60	646,00	Arenitos	9,82	5,77	RQ	RQo	237549,50	7675705,43	94
293,55	8,88	-0,66	544,32	664,00	Basaltos e diabásios	9,97	5,41	RL	RLe	237640,22	7675705,43	97

A partir do total de amostras da matriz de dados grid, foram extraídas 3 amostragens estratificadas com um maior adensamento de amostras em classes com menor representatividade em área. As amostragens foram criadas para o segundo nível categórico (subordem) e para o terceiro nível categórico (grandes grupos) das classes de solo de acordo com o SiBCS da Embrapa (2013).

A representação das classes no segundo nível categórico é resultado de uma generalização taxonômica das classes do terceiro nível. A única mudança entre os dois níveis representados nas amostragens foi a generalização das classes Latossolo Vermelho distrófico (LVd), Latossolo Vermelho distroférrico (LVdf), Latossolo Vermelho eutroférrico (LVef) e Latossolo Vermelho ácrico (LVwf) em uma única classe, o Latossolo Vermelho (LV). Dessa forma, o segundo nível categórico contém 10 classes e o terceiro nível categórico contém 13 classes. As demais classes apenas apresentaram uma mudança em sua legenda, e não uma generalização.

O segundo nível categórico (subordens) apresenta as características da gênese do solo, ou seja, pode indicar tanto a intensidade do processo de formação do solo quanto a atuação de outros processos que agiram de maneira conjunta ou afetaram processos dominantes no primeiro nível. Em outras palavras, o segundo nível categórico expressa o grau de pedogênese (EMBRAPA, 2013).

Em níveis elevados, que possuem maior generalização como as subordens, as diferenças nas características ocorrem do resultado direto dos processos de formação do solo ou que afetam estes processos (EMBRAPA, 2013).

O terceiro nível categórico (grandes grupos) apresentam os Aspectos morfológicos do solo, tais como o tipo e arranjo dos horizontes, atividade de argila, condições de saturação do complexo sortivo por bases ou por alumínio ou por sódio e/ou por sais solúveis. É a partir deste nível que podem ser identificadas características qualitativas, onde as evidências de horizontes ou propriedades irá indicar diferentes tipos de uso e manejo, relacionados ao desenvolvimento de plantas e de água no solo (EMBRAPA, 2013).

3.2.3. Processamento dos dados

Os modelos de predição de classes foram criados no software livre R, utilizando o pacote randomForest (R DEVELOPMENT CORE TEAM, 2015; LIAW & WIENER 2002). Foram selecionados aleatoriamente para cada amostragem estratificada aproximadamente 80% dos dados para treino. O restante dos dados das amostragens, aproximadamente 20%, foram selecionados para teste, sendo utilizada para a avaliação da acurácia. Feita as seleções aleatórias de treino e teste em cada amostragem, foram calculados os valores mínimos, máximos, a média e o desvio padrão das variáveis.

De acordo com Ten Caten (2011), é necessário avaliar o ganho preditivo dos modelos na medida em que eles se tornam mais complexos. Dessa forma, foram testadas as 3 diferentes amostragens A1, A2 e A3, para as classes de solo no 2º e 3º nível categórico. Além disso os modelos Random Forest foram ajustados com diferentes valores de *ntree* (número de árvores para o crescimento da floresta) e *mtry* (número de variáveis selecionadas aleatoriamente para divisão em cada nó da árvore). Os valores de *ntree* testados foram de 200, 300 e 400 árvores por modelo. Os valores de *mtry* testados foram de 2, 4 e 6 variáveis.

Para comprovar se houveram diferenças significativas entre todos os modelos gerados, utilizou-se os valores das médias das taxas de erro OOB de cada modelo para análises estatísticas ANOVA e Tukey ($\alpha = 0,05$).

A avaliação da acurácia foi feita aplicando os modelos para predição nos dados selecionados para teste, onde calculou-se a matriz de confusão destes dados, obtendo assim os valores de exatidão global e do índice Kappa, que avaliam

respectivamente o total de pixels classificados corretamente e a concordância da classificação com os dados de referência.

Os modelos que apresentaram os melhores valores de acurácia, para o segundo e para o terceiro nível categórico, de acordo com valores de *ntree* e *mtry* ajustados e com a amostragem utilizada, serviram para o ajuste do modelo de predição das classes de solo. As classes de solos foram preditas aplicando-se o modelo ajustado a matriz de dados grid.

A matriz de dados grid foi utilizada como arquivo entrada no ArcMap, sendo transformada em um vetor de pontos de acordo com suas coordenadas e convertida para o formato raster, onde obteve-se o arquivo matricial com a grade de linhas e colunas contendo os pixels que representam as classes de solos preditas.

A Figura 14 apresenta de maneira esquemática o fluxograma da metodologia empregada no trabalho.

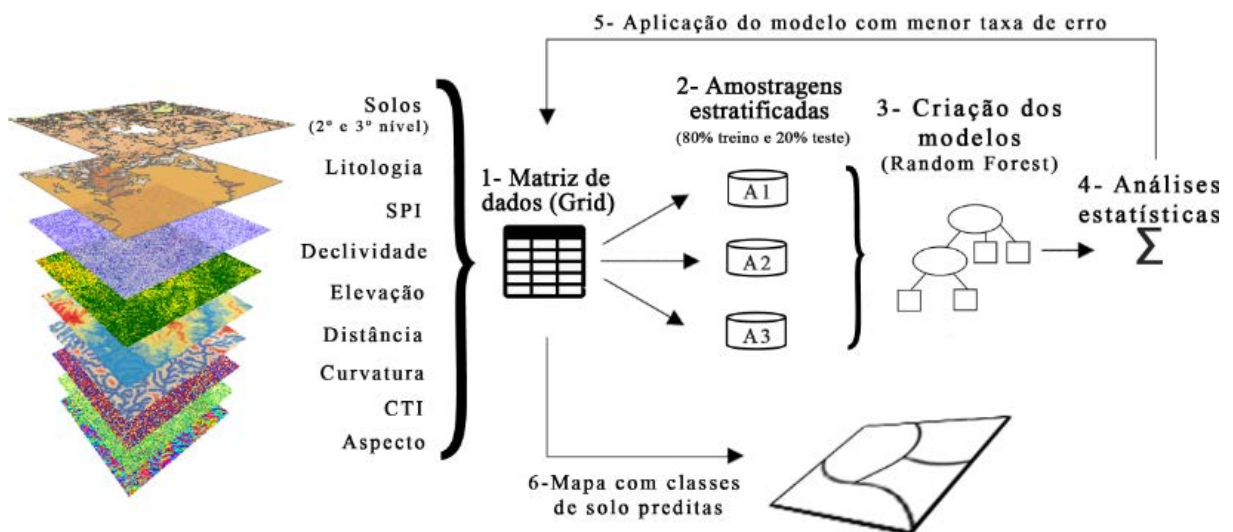


Figura 14 - Fluxograma da metodologia utilizada.

Posteriormente a elaboração dos mapas digitais para o segundo e terceiro nível categórico, foi verificado o grau de concordância bem como a comparação com o mapa original.

Sendo o mapa digital de solos uma representação matricial, pode ocorrer uma redução da generalização dos solos no domínio espacial, onde os detalhes

espaciais da variação dos solos podem ser representados com grande resolução (ZHU et al., 2001). Dessa forma, foram analisadas áreas contendo associações de classes, com o intuito de verificar se houve uma desagregação entre as classes associadas. Diferentemente do trabalho de Häring et al. (2012), que verificou a desagregação das associações com dados de campo, aqui apenas realizou-se a tabulação da área das classes obtidas nos mapas digitais com as associações de classes, com o intuito de verificar se o erro associado ao modelo compreende as classes associadas.

Avaliou-se a correspondência entre o mapa digital e alguns perfis de solo coletados em campo. Foram destacados apenas locais onde houve equivalência da verdade de campo com as mudanças observadas no mapa digital em relação ao mapa original. Foram obtidos dois perfis levantados em campo, dois perfis do Boletim Científico do Levantamento Semidetalhado da quadrícula de Ribeirão Preto de Oliveira & Prado (1987) e um perfil do projeto RADAMBRASIL, extraído do trabalho de Benedetti et al. (2008). Não foram obtidos perfis suficientes para ser feita uma validação, apenas para medida de avaliação pontual do que foi distinguido do mapa original.

4. RESULTADOS E DISCUSSÃO

4.1. Descrição das variáveis dos fatores *s.c.o.r.p.a.n.*

4.1.1. Fator *s*

A partir da adequação ao SiBCS, foram encontradas 13 classes de solo no terceiro nível categórico. A mostra as classes da quadrícula de Ribeirão Preto segundo a classificação antiga do mapa pedológico semidetalhado de Oliveira & Prado (1987) juntamente com as classes adequadas para a classificação atual e suas respectivas áreas em hectares e porcentagem. Foram desconsiderados na somatória de área das classes de solos os cursos hídricos, represas, lagos e a mancha urbana dos municípios.

Tabela 6 - Tabela de conversão da classificação de solos antiga para atual do mapa semidetalhado da quadrícula de Ribeirão Preto, com a área de cada classe em hectares e porcentagem.

Classificação antiga	Sigla antiga e Unidades	Classificação atual	Sigla atual	Área (hectares)	Área (%)
Cambissolo distrófico	Cb	Cambissolo Háplico Tb distrófico típico	CXbd	3394,53	1,22%
Brunizem Avermelhado	BV	Chernossolo Argilúvico órtico	MTo	317,68	0,11%
Associação de Gley pouco húmico e Gley húmico	Hi-2	Gleissolo Háplico Tb distrófico típico	GXbd	6904,74	2,49%
Associação de Gley húmico e Gley pouco húmico	Hi-1	Gleissolo Melânico Tb distrófico	GMbd	324,36	0,12%
Latossolo Roxo ácrico	LRa, LRv, LRe	Latossolo Vermelho ácriférico	LVwf	109591,44	39,50%
Latossolo Vermelho Amarelo álico	LV-1, LV-2, LV-3, LV-4, LV-5	Latossolo Vermelho Amarelo distrófico	LVAd	10850,50	3,91%
Latossolo Roxo distrófico	LRd-1 e LRd-2	Latossolo Vermelho distroférico	LVdf	31422,70	11,33%
Latossolo Vermelho Escuro álico	LE-1, LE-2, LE-3 e LE-4	Latossolo Vermelho distrófico	LVd	31201,31	11,25%
Latossolo Roxo eutrófico	LRr e LRe	Latossolo Vermelho eutroférico	LVef	41641,31	15,01%
Solos Litólicos eutróficos	Li-1	Neossolo Litólico eutrófico	RLe	13769,06	4,96%
Areia Quartzosa álica	AQ	Neossolo Quartzarênico órtico	RQo	17923,22	6,46%
Terra Roxa Estruturada eutrófico	TE	Nitossolo Vermelho eutroférico	NVef	6051,19	2,18%
Solo Orgânico	Hi-3	Organossolo háplico sáprico	OXs	4062,35	1,46%

O mapa adaptado é apresentado na Figura 15, o qual mostra a distribuição geográfica das classes de solo já adequadas ao novo sistema de classificação.

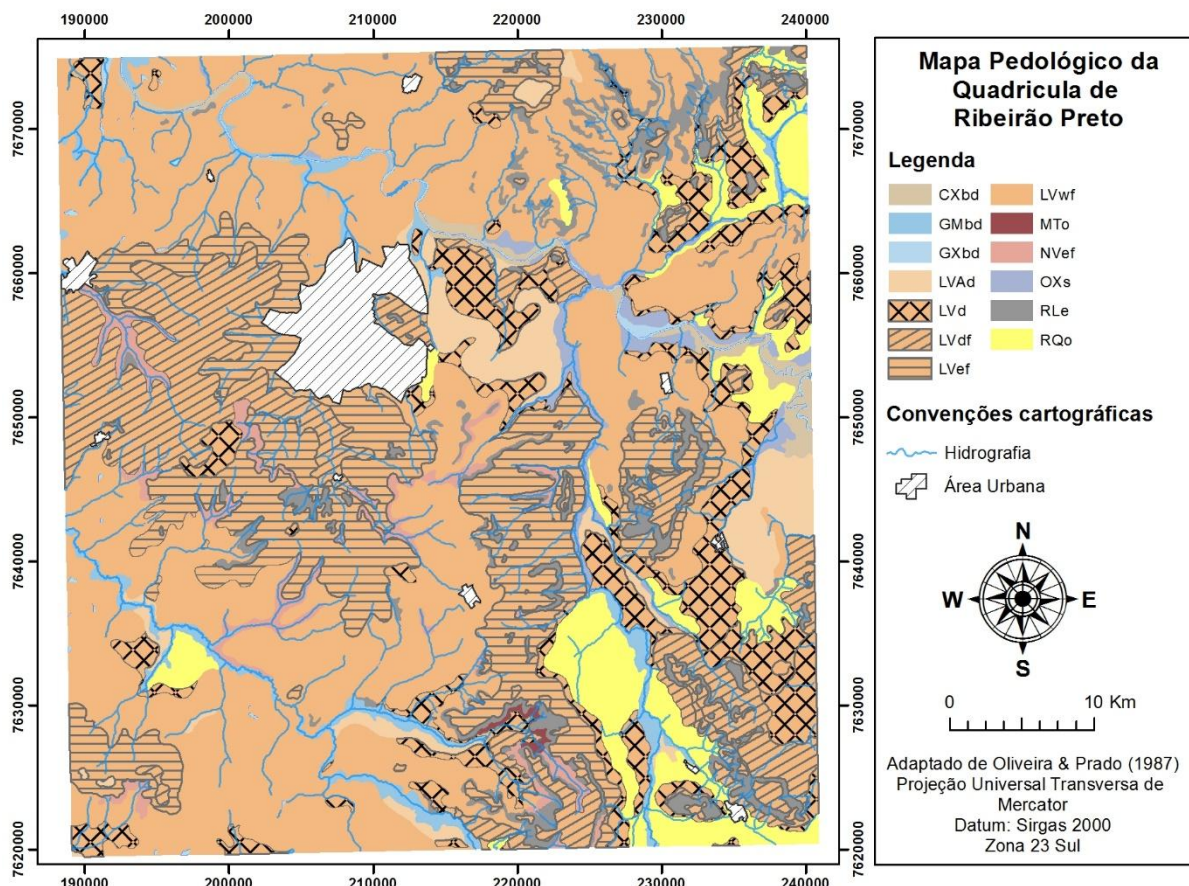


Figura 15 - Mapa Pedológico da Quadrícula de Ribeirão Preto (adaptado de Oliveira & Prado, 1987).

A medida da escala efetiva do mapa é apresentada na Tabela 7.

Tabela 7 - Indicadores e cálculo da escala efetiva para o mapa pedológico da quadrícula de Ribeirão Preto.

	Área (ha)	Polígonos	Escala nominal	ASD	MLD	IMR	Escala efetiva
Mapa original	277452,18	328	1:100,000	845,8908	40	4,598616	1:229.930

Como pode ser observada a escala efetiva é cerca de quatro vezes inferior se comparado com sua escala nominal. Sendo assim, o mapa utilizado se enquadraria em um nível de detalhamento de reconhecimento, segundo o que recomenda o SiBCS Embrapa. Este resultado pode estar relacionado as condições locais, já que há um predomínio do Latossolo Vermelho.

4.1.2. Fator r

O fator r compreende as variáveis primárias Elevação, Aspecto, Curvatura e Declividade e as variáveis secundárias CTI e SPI, além da variável Distância da rede de drenagem.

A quadrícula de Ribeirão Preto possui uma diferença de altitude em toda sua área de aproximadamente 470 metros. Há um predomínio de baixas altitudes, com cerca de 21% da quadrícula variando entre 489 e 556 metros. As altitudes elevadas, acima de 770 metros, representam apenas 4,81% da área.

O mapa da Figura 16 mostra o modelo digital de elevação representando a variável altimetria. A Tabela 8 mostra a distribuição das classes de elevação por área.

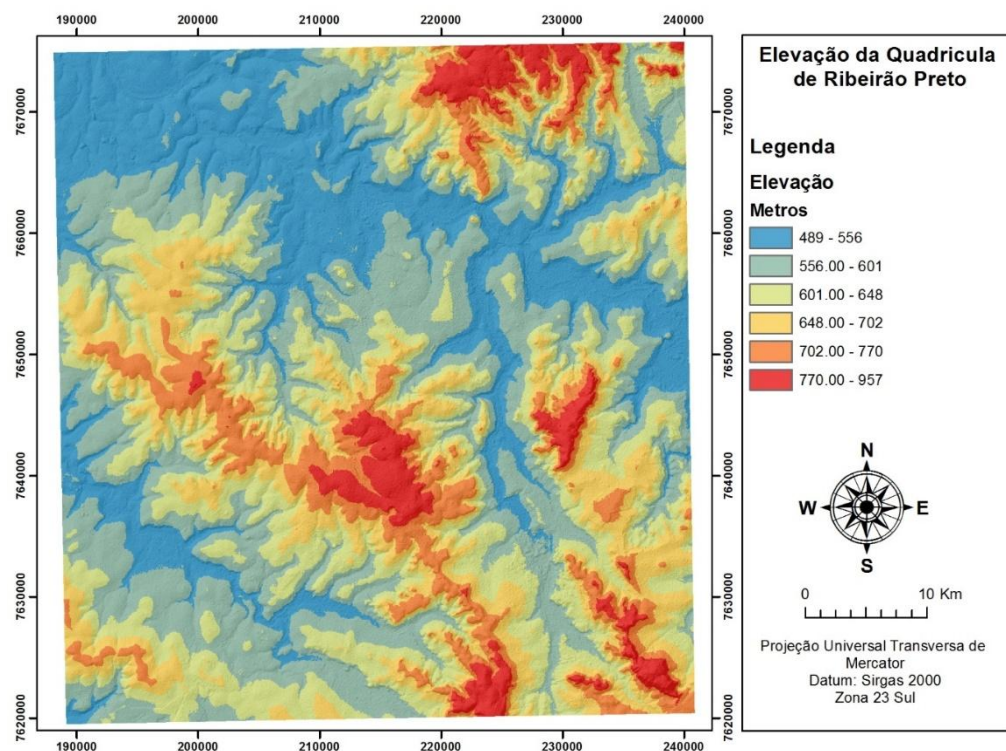


Figura 16 - Mapa da elevação da quadrícula de Ribeirão Preto.

Tabela 8 - Distribuição das classes de elevação por área na quadrícula de Ribeirão Preto.

Classe	Área (%)
489-556	21,24
557-601	27,77
602-648	22,00
649-702	15,29
703-770	8,89
771-957	4,81

O mapa das classes de declividade é apresentado na Figura 17. A Tabela 9 mostra a distribuição das classes de declividade por área na quadrícula.

De acordo com a classificação de declividade proposta por Lepsch et al. (1991), as classes foram divididas em plano (0-3%), suave ondulado (3-6%), moderadamente ondulado (6-12%), ondulado (12-20%), forte ondulado (20-40%) e montanhoso (>40%).

O relevo predominante na quadrícula é o moderadamente ondulado, com 34,20%, seguido do relevo suave ondulado, que corresponde a 28,63% da área. O relevo plano abrange 22% da quadrícula, se concentrando em sua maior parte ao Norte, às margens do Rio Pardo e seus afluentes, e ao Sul, nas margens do Ribeirão da Onça e seus afluentes. O relevo ondulado apresenta 10,55% de área. O relevo fortemente ondulado apresenta 3,70% e o montanhoso 0,58%, se concentrando à nordeste e centro-sul, sendo encontrado nestes locais solos mais rasos.

O relevo de maneira geral é pouco movimentado, pois as classes ondulado, fortemente ondulado e montanhoso, correspondem apenas 14,83% da área.

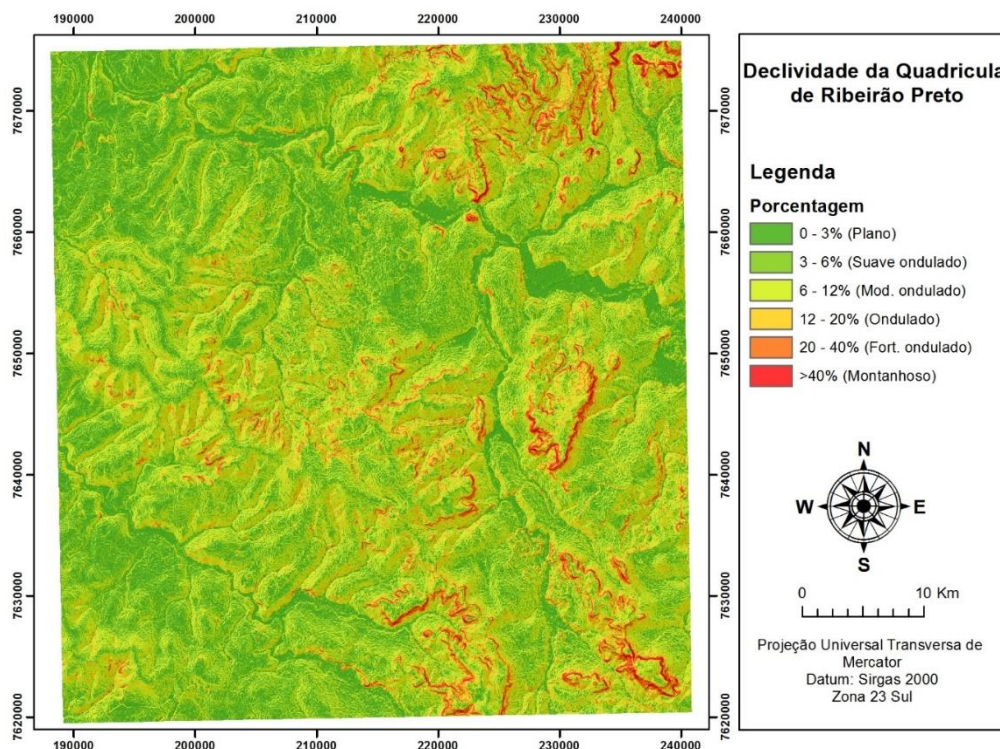


Figura 17 - Mapa de declividade da quadrícula de Ribeirão Preto.

Tabela 9 - Distribuição das classes de declividade por área na quadrícula de Ribeirão Preto de acordo com as classes definidas por Lepsch et al. (1991).

Classe de Declividade	Relevo	Área (%)
0 - 3%	Plano	22,34
3 - 6%	Suave Ondulado	28,63
6 - 12%	Moderadamente Ondulado	34,20
12 - 20%	Ondulado	10,55
20 - 40%	Fortemente Ondulado	3,70
>40%	Montanhoso	0,58

A variável Aspectogra representa a direção da declividade, sendo importante para determinar a exposição das vertentes a disponibilidade de energia na forma de calor, o que influencia nos processos de intemperismo e pedogênese. Por apresentar uma grande extensão, a quadrícula de Ribeirão Preto possui uma grande variedade de orientações do relevo que são bem distribuídas. O mapa da Figura 18 apresenta as classes de Aspecto e a Tabela 10 mostra a distribuição destas classes por área na quadrícula.

Não há a predominância de uma determinada orientação de vertente, que fica evidenciado pela média de área de 12% para cada sentido geográfico. Áreas de relevo plano, que possuem declividade zero, não apresentam uma orientação de vertente definida, sendo assim, o valor atribuído para diferenciá-la das demais áreas é -1 . Estas áreas podem ser bem identificadas no mapa, nas várzeas do rio Pardo, ao norte da quadrícula sentido SE-NO, representando apenas 3% da área.

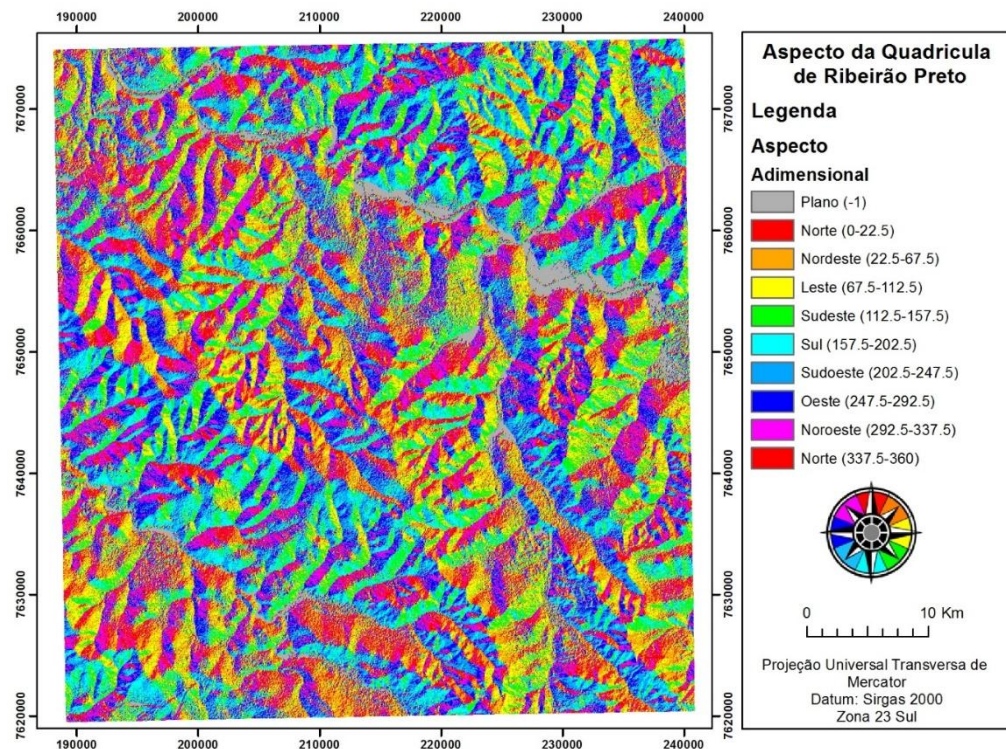


Figura 18 – Mapa do Aspecto da quadrícula de Ribeirão Preto.

Tabela 10 - Distribuição das classes de Aspecto por área na quadrícula de Ribeirão Preto.

Classe	Área (%)
Plano	3,21
Nordeste	11,95
Leste	12,27
Sudeste	11,86
Sul	10,77
Sudoeste	11,49
Oeste	12,90
Noroeste	13,19
Norte	12,31

A curvatura apresenta formas côncavas, convexas e planares, que exercem influência na distribuição de água e sedimentos das partes mais elevadas às partes mais baixas. O mapa da Figura 19 exhibe de forma simplificada a reclassificação dos valores de curvatura em classes: superfícies planas (valores entre -0,05 e 0,05); superfícies côncavas (valores inferiores a 0,05); superfícies convexas (valores superiores a 0,05). A Tabela 11 mostra as classes de curvatura por área na quadrícula.

A forma convexa de curvatura (valores positivos) representa 42,3% da área da quadrícula, indicando locais de transporte de sedimentos, que dão origem a solos mais profundos e evoluídos.

A forma côncava de curvatura (valores negativos) possui uma área de 40,4%, representando locais de acúmulo de sedimentos e de retenção de umidade prolongada, ocorrendo solos mais rasos, pois há uma menor variação da declividade. De acordo com Chagas (2006), a curvatura côncava apresenta maior fluxo convergente de água, o que favorece a morfogênese, tornando esses solos rejuvenescidos pela erosão.

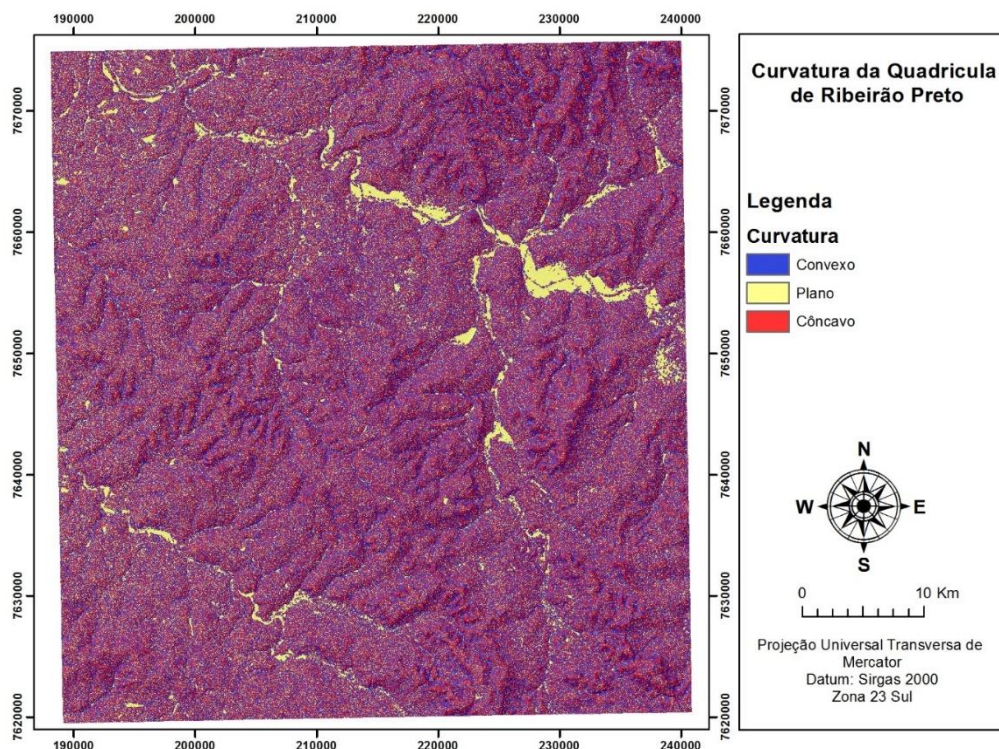


Figura 19 - Mapa de curvatura da quadrícula de Ribeirão Preto.

Tabela 11 - Distribuição das classes de curvatura por área na quadrícula de Ribeirão Preto.

Classe	Área (%)
Convexo	42,28
Plano	17,35
Côncavo	40,37

A distância da rede de drenagem é apresentada no mapa da Figura 20, de acordo com suas classes definidas. A distribuição das classes de distância por área é apresentada na Tabela 12.

A densa rede hidrográfica da quadrícula de Ribeirão Preto faz com que a distância da drenagem não seja elevada, apresentando uma área de 30,9% com distâncias que variam de 0 a 313 metros e 26,9% com distâncias que variam de 314 a 659 metros.

Contribui para esta alta densidade a grande quantidade de nascentes que ocorrem no divisor de águas das bacias dos rios Pardo e Mogi-Guaçu, que corta a quadrícula no sentido NO-SE.

Distâncias da rede drenagem que variam de 1500 a 4200 metros correspondem a apenas 7% da quadrícula, restringindo-se a locais com elevada altitude.

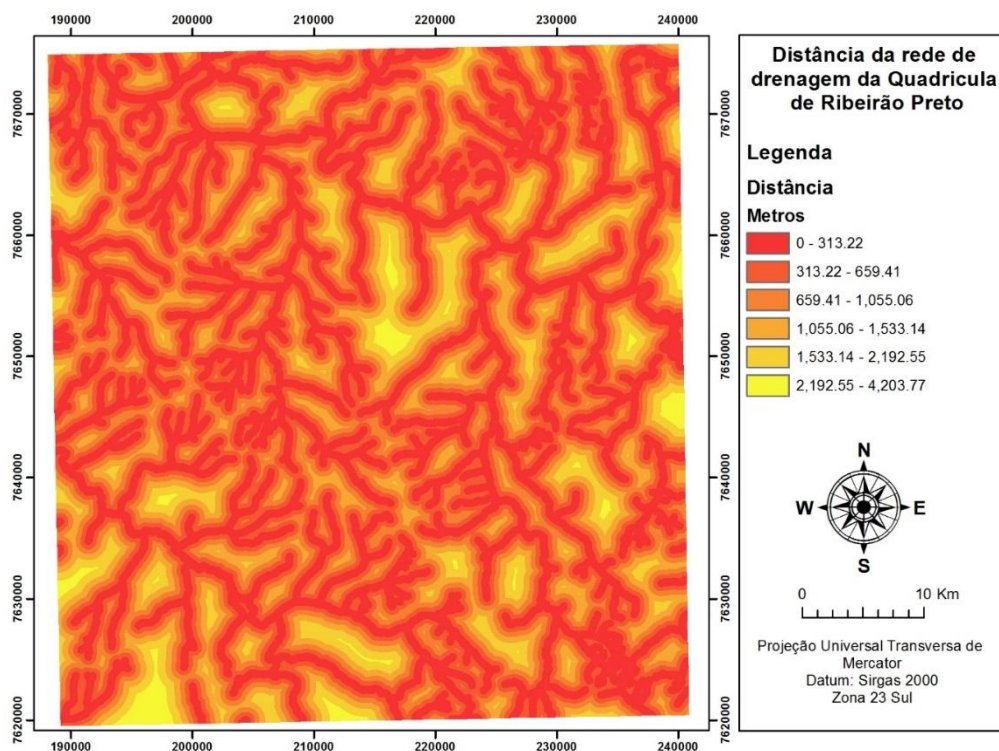


Figura 20 - Mapa da distância da rede de drenagem.

Tabela 12 - Distribuição das classes de distância da rede de drenagem por área na quadrícula de Ribeirão Preto.

Classe	Área (%)
0-313	30,91
314-659	26,93
660-1055	21,87
1056-1533	12,87
1534-2192	6,13
2193-4203	1,26

A Figura 21 apresenta o mapa do CTI e a Tabela 13 mostra a distribuição da área para os intervalos definidos.

Valores elevados de CTI coincidem com áreas de acumulação, ou seja, áreas planas, de baixadas, onde é menos provável o escoamento. Estas são

representadas no mapa pelas classes de valores maiores que 9,45 que compreendem juntas apenas 16,47% da área.

Áreas de topos de morro localizadas ao centro-sul e nordeste da quadrícula apresentam baixos valores de CTI, sendo representas pelas classes com valores de 3,33 a 9,45. São locais de drenagem do relevo, que compreendem 83,53% da área.

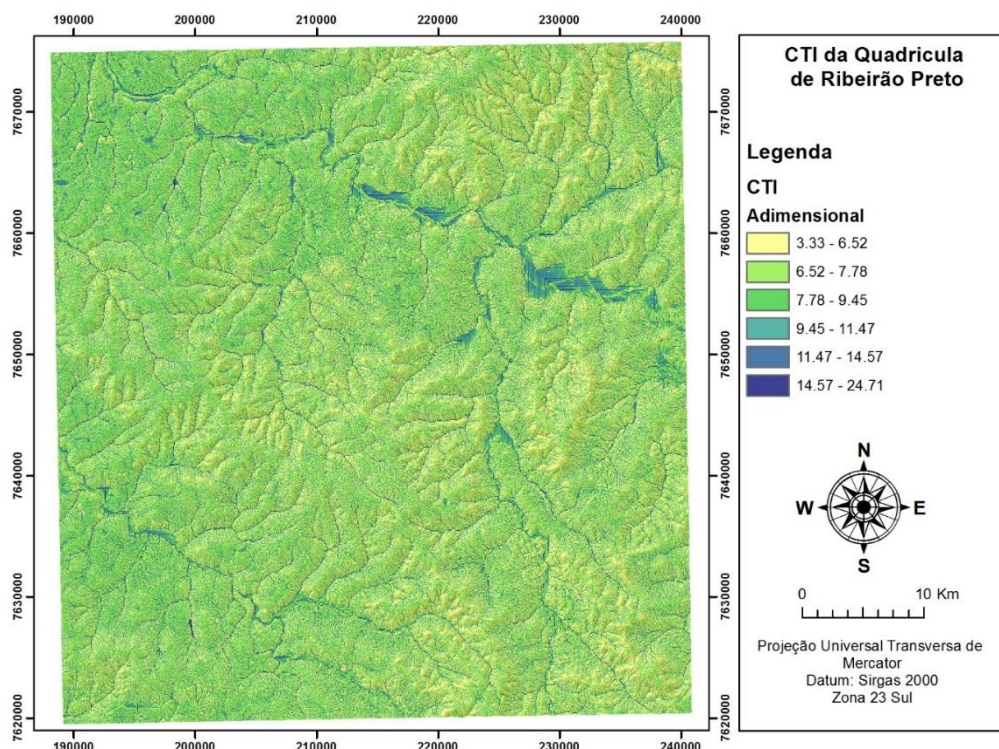


Figura 21 - Mapa de CTI da quadrícula de Ribeirão Preto.

Tabela 13 - Distribuição do intervalo de classes de CTI por área na quadrícula de Ribeirão Preto.

Classe	Área (%)
3,33 - 6,52	30,79
6,52 - 7,78	34,28
7,78 - 9,45	18,43
9,45 - 11,57	10,83
11,47 - 14,57	4,40
14,57 - 24,71	1,24

A Figura 22 e a Tabela 14 mostram respectivamente o mapa de SPI e área dos intervalos de classes definidos.

Os valores elevados de SPI indicam um maior fluxo e consequentemente maior transporte de sedimentos, havendo assim predomínio da morfogênese em relação a pedogênese. São locais onde a água flui com maior energia, onde predominam processos erosivos, sendo responsáveis pela limitação continuada no desenvolvimento de solos (SIRTOLI et al., 2008; THOMPSON et al., 2012).

Os valores do intervalo de 7,85 a 15,87 indicam um maior poder de escoamento e se encontram nas vertentes do relevo, que de acordo com Thompson et al. (2012), coincidem com locais de elevada área contribuição específica e declividade, ou seja, é onde se espera que esta água continue a fluir por causa de gradientes de declividade mais acentuados.

Valores baixos indicam solos mais desenvolvidos com predomínio da pedogênese, ou de caráter hidromórfico quanto localizados em áreas de baixadas ou várzeas (SIRTOLI et al., 2008).

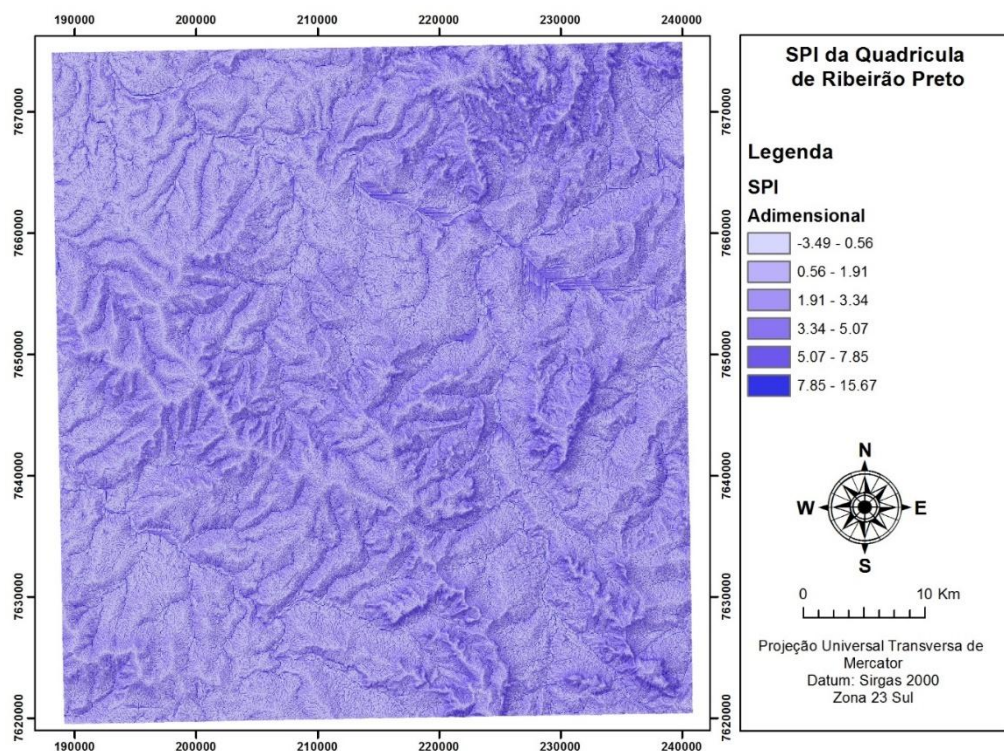


Figura 22 - Mapa de SPI da quadricula de Ribeirão Preto.

Tabela 14 - Distribuição do intervalo de classes de SPI por área na quadrícula de Ribeirão Preto.

Intervalo de classes	Área (%)
-3,49 – 0,56	17,83
0,56 – 1,91	31,41
1,91 – 3,34	26,18
3,34 – 5,07	15,22
5,07 – 7,85	7,73
7,85 – 15,67	1,59

4.1.3. Fator *p*

A litologia da quadrícula de Ribeirão Preto é predominantemente formada por basaltos e diabásios da formação Serra Geral, que compreendem 66,44% da área. Os arenitos das Formações Botucatu, Pirambóia e do grupo Bauru compreendem 24,72% da área. Sedimentos inconsolidados correspondem a uma área de 8,84%. A Figura 23 e a Tabela 15 apresentam respectivamente o mapa e a área das classes de Litologia da quadrícula de Ribeirão Preto.

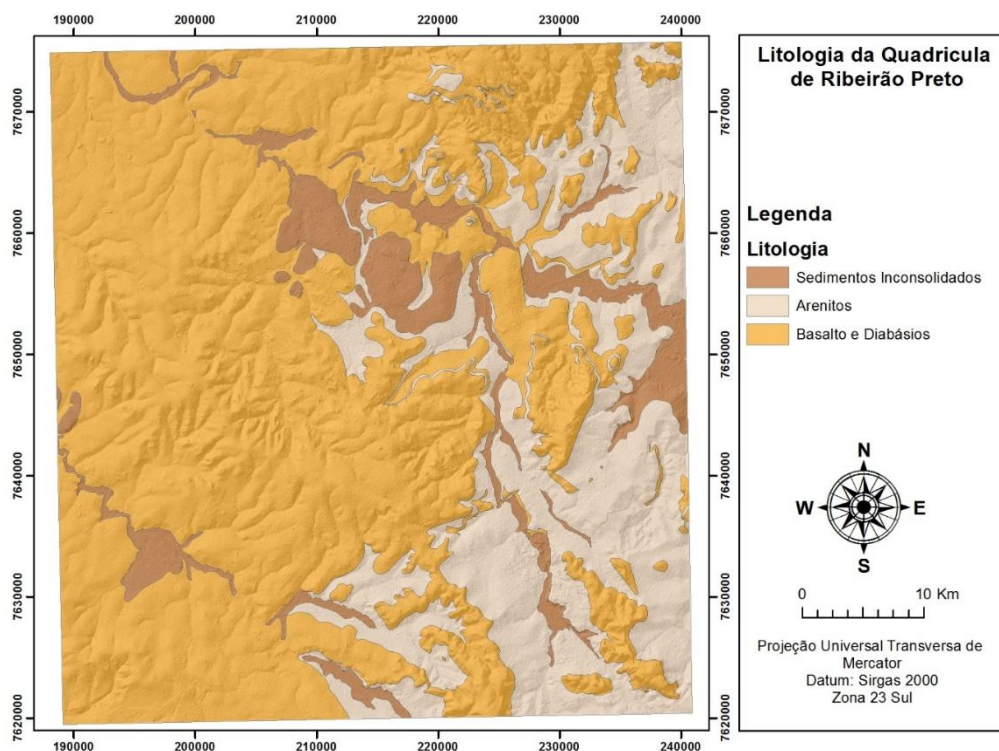


Figura 23 - Mapa da Litologia na quadrícula de Ribeirão Preto.

Tabela 15 - Distribuição do intervalo de classes da Litologia por área na quadrícula de Ribeirão Preto

Litologia	Área (%)
Arenitos	24,72%
Sedimentos Inconsolidados	8,84%
Basaltos e Diabásios	66,44%

4.2. Análise dos dados das amostragens

As amostragens, a quantidade de amostras aleatórias selecionados para treino e teste dos modelos e o total amostrado em relação a quadricula é apresentado na Tabela 16.

Tabela 16 - Quantidade de pontos para treino e teste, total de cada amostragem aleatória e densidade de amostragem da quadricula.

Amostragem	Treino	Teste	Total de amostras	Densidade de amostragem da quadrícula (%)
A1	5060	1264	6324	0.21%
A2	7973	1993	9966	0.33%
A3	10481	2459	12940	0.43%

Observa-se que a porcentagem do total amostrado situa-se em um valor abaixo de 0,5% do total de pontos contidos em cada pixel da quadricula. Buscou-se obter uma baixa densidade de amostras em relação a área da quadrícula, porém com uma boa representatividade das variáveis entre as classes.

O gráfico das Figura 24 e Figura 25 apresentam respectivamente a densidade das amostras em cada amostragem para o segundo e terceiro nível categórico dos dados de treino, comparando-se com o total de pixels que cada classe possui no grid.

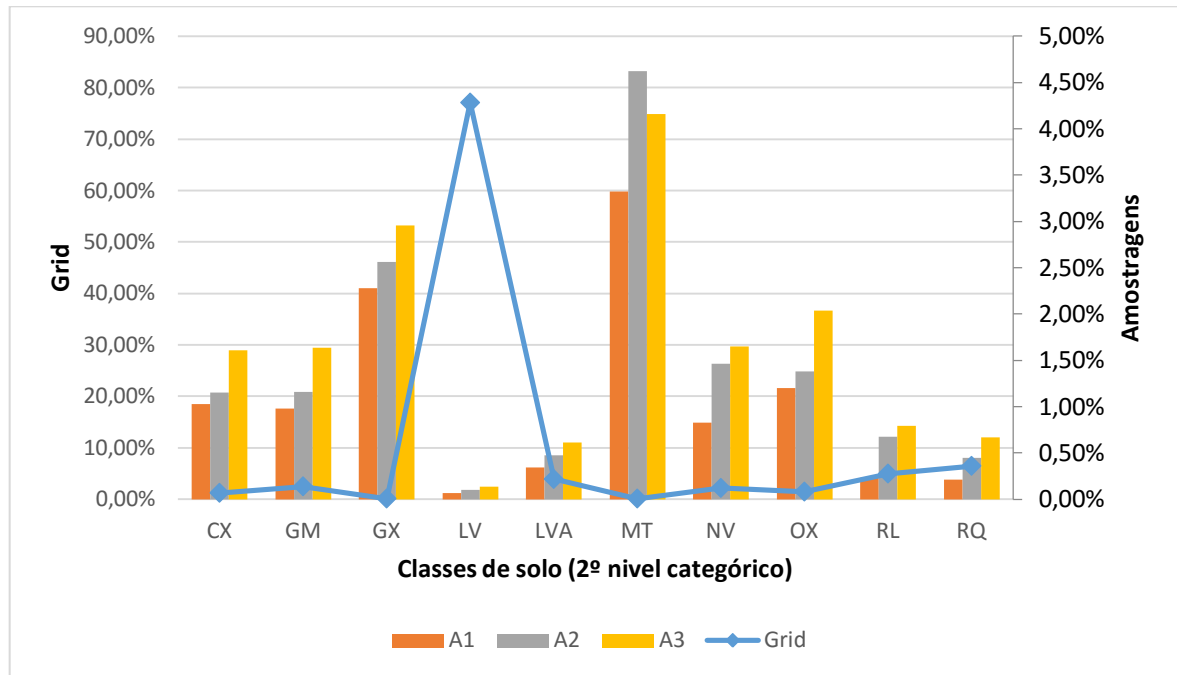


Figura 24 – Distribuição percentual das amostras nas diferentes amostragens de treino para cada classe em relação ao grid no segundo nível categórico.

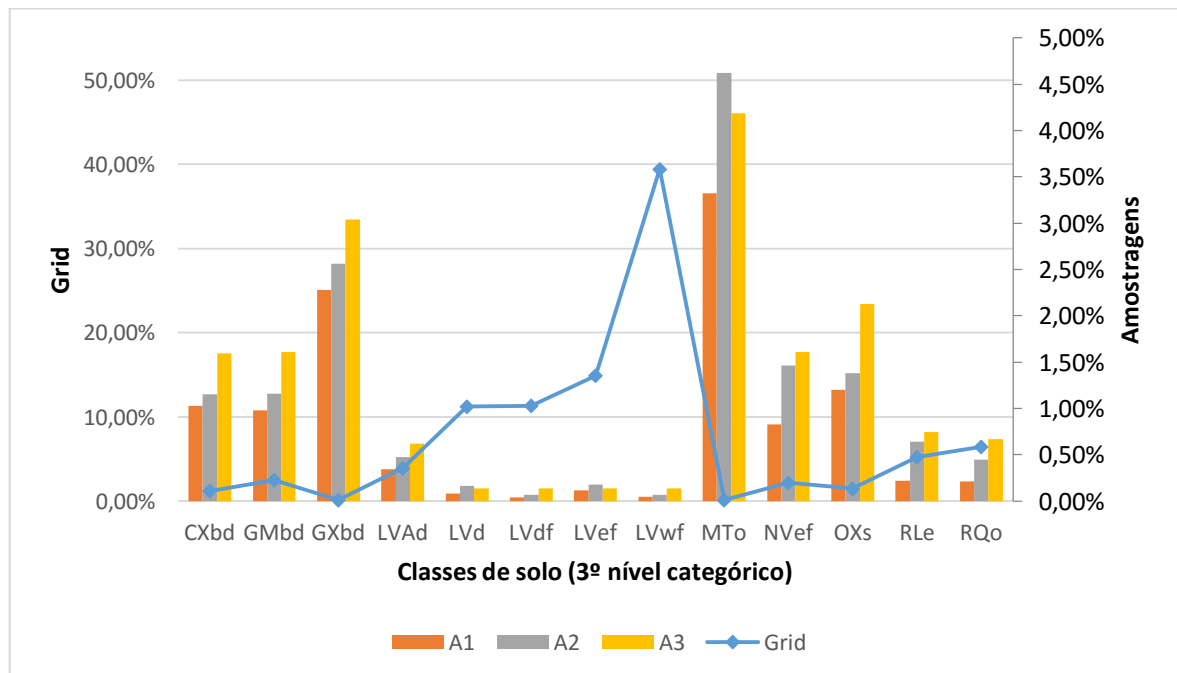


Figura 25 – Distribuição percentual das amostras nas diferentes amostragens de treino para cada classe em relação ao grid no terceiro nível categórico.

Há um grande desequilíbrio na área das classes de solo que ocorrem na quadrícula, conforme pode ser observado no gráfico pela porcentagem de cada

classe no grid. As maiores densidades de amostras ocorrem nas classes que tem a menor área ocupada na quadrícula.

Com a amostragem estratificada valorizando as classes menos representativas em área, o desequilíbrio entre as classes é suavizado. As classes com menor representatividade como CX (CXbd), GM (GMbd), GX (GXbd) e MT (MTo) foram as que apresentaram as maiores densidades de amostras, enquanto as classes LV (correspondente a generalização das classes LVd, LVdf, LVef e LVwf) com maior porcentagem de área no grid, apresentaram uma menor densidade de amostras em relação ao total do grid.

A estatística descritiva das variáveis em cada amostragem e no grid é apresentado na Tabela 17.

Tabela 17 - Estatística descritiva dos conjuntos de amostragem e do grid da quadrícula.

Variáveis	Pontos	Mínimo	Máximo	Média	Desvio Padrão
Aspecto	Grid	-1,00	359,76	173,97	108,44
	A1	-1,00	358,40	154,30	116,08
	A2	-1,00	359,09	158,40	113,65
	A3	-1,00	359,29	158,91	113,77
CTI	Grid	3,34	24,71	7,71	2,04
	A1	4,13	22,34	8,13	2,55
	A2	3,87	22,16	7,96	2,51
	A3	3,82	22,87	7,93	2,42
Curvatura	Grid	-6,45	8,97	0,00	0,36
	A1	-2,84	3,50	-0,01	0,37
	A2	-3,61	4,37	0,00	0,40
	A3	-3,39	3,50	-0,01	0,39
Distância	Grid	0,00	4203,80	669,00	533,13
	A1	0,00	3356,60	500,40	498,15
	A2	0,00	3664,00	521,20	504,26
	A3	0,00	3507,80	532,00	510,39
Elevação	Grid	489,00	957,00	620,20	74,71
	A1	493,00	901,00	595,70	75,41
	A2	493,00	952,00	606,90	76,67
	A3	493,00	944,00	605,80	77,93
Declividade	Grid	0,00	56,90	4,39	3,58
	A1	0,00	38,00	4,43	3,97
	A2	0,00	35,52	5,01	4,55
	A3	0,00	47,32	4,82	4,48
SPI	Grid	-3,50	15,68	2,02	1,94
	A1	-3,50	15,04	2,34	2,31
	A2	-3,50	15,26	2,38	2,24
	A3	-3,50	15,68	2,29	2,16

Comparando os conjuntos de amostragem com o grid, observa-se que de maneira geral os valores mínimos e máximos, a média e o desvio padrão das variáveis não apresentam grandes diferenças, o que mostra a boa representatividade das variáveis em cada amostragem, contemplando um intervalo de valores que representa a totalidade do grid.

A Figura 26 apresenta os histogramas das variáveis para cada amostragem e para o grid.

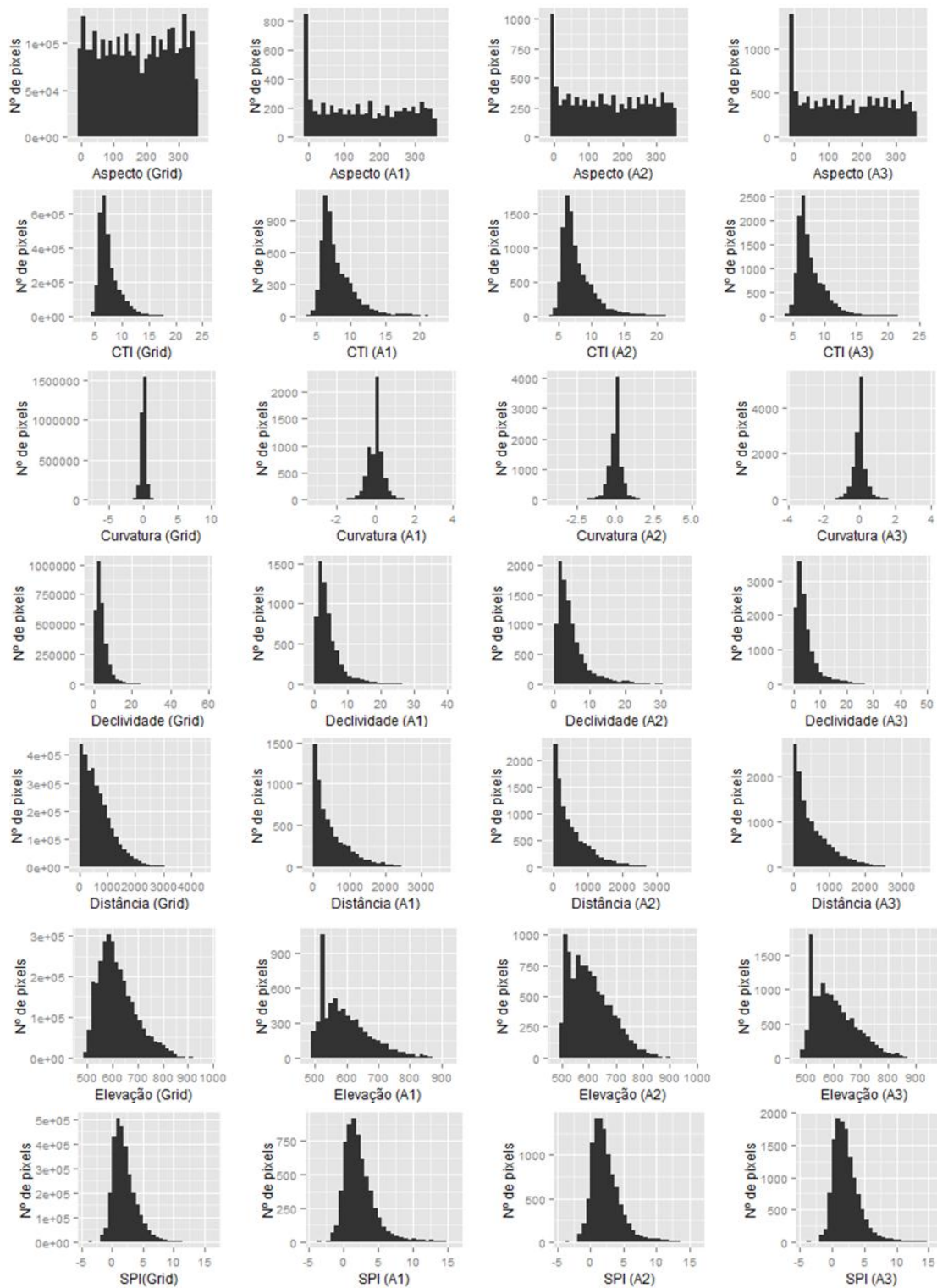


Figura 26 - Histogramas das variáveis para o grid e para as amostragens A1, A2 e A3.

Em uma comparação do comportamento e distribuição dos valores dentro de cada variável, observa-se que os histogramas são similares, no qual grande parte não possui distribuição normal, com exceção da variável Curvatura. A variável Aspecto apresenta uma distribuição irregular.

Destaca-se que as amostragens não superestimam determinados intervalos de valores em detrimento de outros, ou seja, comparando-se com o grid, nota-se que o comportamento é semelhante. Os histogramas revelam que as amostras obtidas do grid não alteram a distribuição dos valores das variáveis, onde cada parte (amostragem) corresponde de maneira significativa ao todo (grid).

4.3. Avaliação dos modelos Random Forest ajustados

Um total de 54 modelos com diferentes ajustes foram criados, avaliados e analisados estatisticamente em conjunto, para cada nível.

O teste ANOVA realizado para o segundo e terceiro nível categórico, mostrou que as amostragens apresentam diferenças significativas nas médias da taxa de erro OOB ($p < 0,05$), no qual a amostragem A3 foi a que apresentou o menor valor.

Os ajustes dos diferentes valores de variáveis em cada árvore também apresentaram diferenças significativas nas médias das taxas de erro OOB, porém o teste Tukey mostrou que os ajustes em 4 ou 6 não diferem entre si. Sendo assim, o ajuste do mtry com valores acima de 4 variáveis não representaria uma redução na taxa de erro, apenas um maior gasto de tempo no processamento dos dados, pois mais da metade das variáveis passariam a ser utilizadas para dividir o nó de cada uma das árvores, o que também prejudicaria a aleatoriedade do processo de seleção das variáveis.

Os valores de ntree não apresentaram diferenças significativas entre os valores 200, 300 e 400. Utilizando a amostragem A3 e mtry igual a 4, testou-se o número de árvores no modelo. Os gráficos das Figura 27 e Figura 28 mostram a variação da taxa de erro OOB para as classes no segundo e terceiro nível categórico respectivamente, em função do aumento do número de árvores.

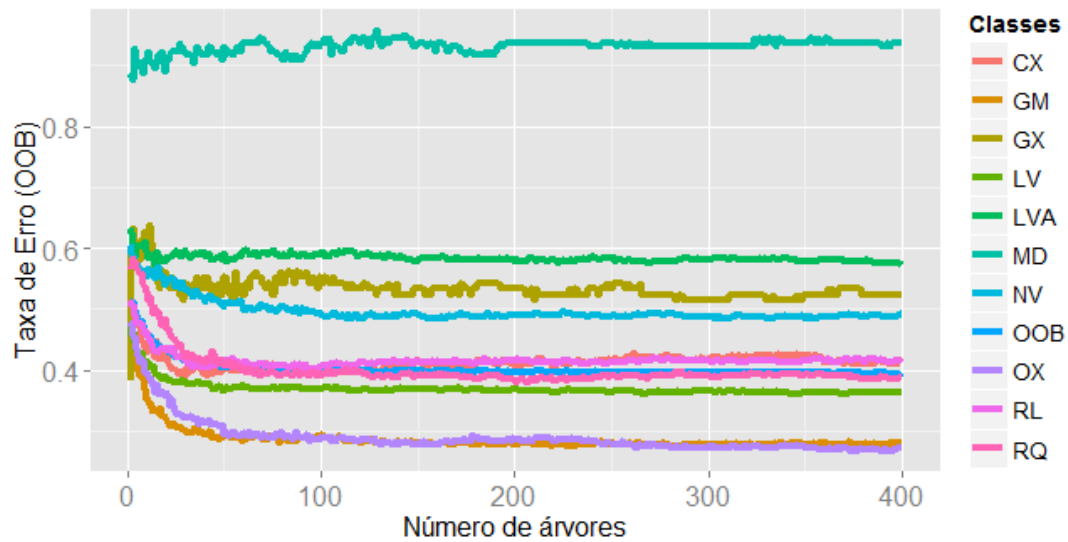


Figura 27 - Taxa de erro OOB pelo número de árvores no segundo nível categórico.

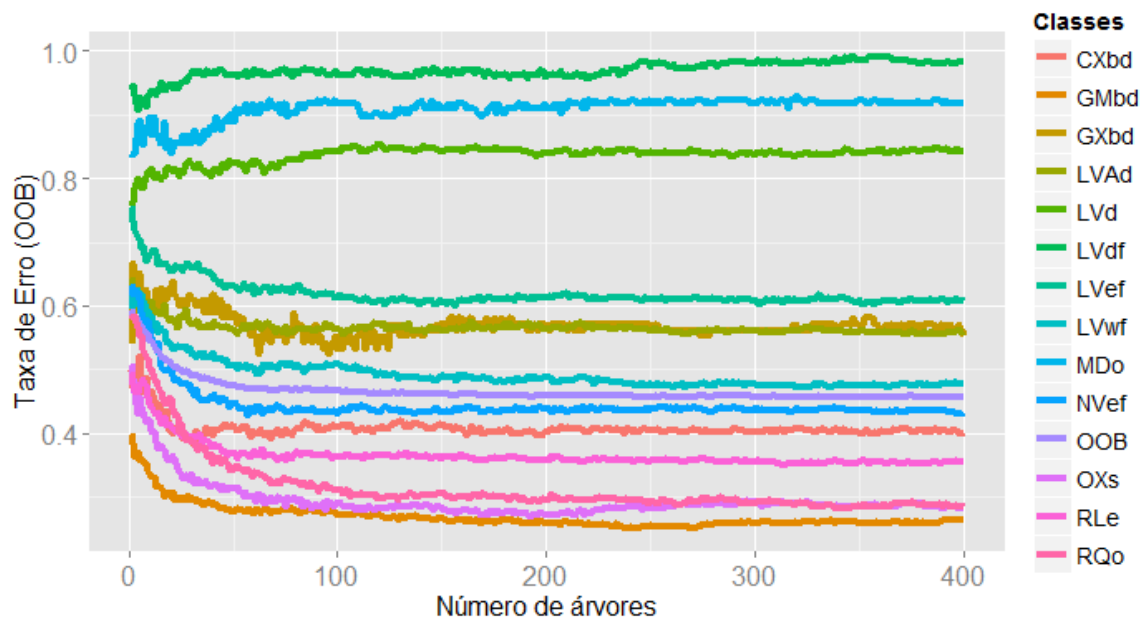


Figura 28 - Taxa de erro OOB pelo número de árvores no terceiro nível categórico.

Observa-se que a taxa de erro OOB geral e as taxas de erro para cada classe se estabilizam a partir do crescimento da floresta com 200 árvores, apresentando um erro geral em torno de 40% para o segundo nível categórico e 45% para o terceiro nível categórico. O aumento do número de árvores para um valor acima de 200 representaria apenas um maior gasto de tempo para o processamento dos

dados, já que não houve diferenças significativas na diminuição da taxa de erro OOB. Estes resultados indicam que o ajuste do valor *ntree* para 200 árvores no modelo não causaria prejuízo na predição das classes.

As médias das taxas de erro OOB na comparação entre os níveis categóricos apresentaram diferenças significativas ($p < 0,05$). O segundo nível categórico apresentou taxas de erro mais baixas do que o terceiro nível categórico. Isto indica que a generalização das classes LVd, LVdf, LVeF e LVwf resultou em um maior ganho preditivo.

4.4. Avaliação da predição das classes de solo

Aplicando o modelo Random Forest ajustado de acordo com a avaliação preliminar nas classes do segundo nível categórico, obteve-se na avaliação da acurácia o valor de exatidão global de 62,5% e índice Kappa de 0,62. Este resultado apresenta um grau de concordância forte, de acordo com a classificação dos valores do índice Kappa propostos por Landis & Koch (1977), pois o resultado obtido situa-se entre o intervalo 0,61 e 0,80. A matriz de confusão é apresentada na Tabela 18, que mostra os resultados de Exatidão do usuário (EU), Exatidão do produtor (EP) e Exatidão global (EG).

Tabela 18 - Matriz de confusão do modelo Random Forest com amostragem A3, $n_{tree} = 300$ e $m_{try} = 4$ no segundo nível categórico.

Classes	CX	GM	GX	LV	LVA	MT	NV	OX	RL	RQ	Total	EU
CX	82	13	2	2	1	0	0	22	0	0	122	67.2
GM	5	222	0	21	9	2	1	21	1	29	311	71.4
GX	1	1	13	0	0	0	0	1	0	0	16	81.3
LV	4	18	0	488	34	3	77	3	65	45	737	66.2
LVA	1	6	0	20	83	0	1	0	0	7	118	70.3
MT	0	0	0	2	0	4	0	0	0	0	6	66.7
NV	0	3	0	69	1	2	119	0	29	0	223	53.4
OX	44	20	7	3	5	0	2	134	0	7	222	60.4
RL	0	0	0	54	2	8	48	0	177	4	293	60.4
RQ	1	19	0	95	45	16	4	3	14	214	411	52.1
Total	138	302	22	754	180	35	252	184	286	306	2459	
EP	59.4	73.5	59.1	64.7	46.1	11.4	47.2	72.8	61.9	69.9		
EG	62.5											
Kappa	0.62											

A exatidão do produtor, que considera a exclusão do pixel de uma classe no qual ele deveria ser designado, apresentou valores abaixo de 50% apenas para as classes NV, MT e LVA. Estes baixos resultados ocorrem pela inclusão de pixels na classe incorreta devido à dificuldade de discriminação das classes por meio de determinadas variáveis, diminuindo assim os resultados de exatidão global e do índice Kappa.

O pior resultado na exatidão do produtor foi o da classe MT, que apresentou um valor de 11,4%. Do total de pixels correspondentes a esta classe, 8 foram designados para a classe RL e 16 para o RQ. Este baixo valor pode estar relacionado a baixa quantidade de amostras de teste selecionadas para esta classe. Além disso, este baixo valor pode ser explicado pela confusão entre as classes MT e RL que acontece pela semelhança entre as variáveis discriminantes Aspecto, CTI, Elevação e SPI, como pode ser visto na Figura 29, devido a ocorrência destas classes em relevo ondulado ou fortemente ondulado.

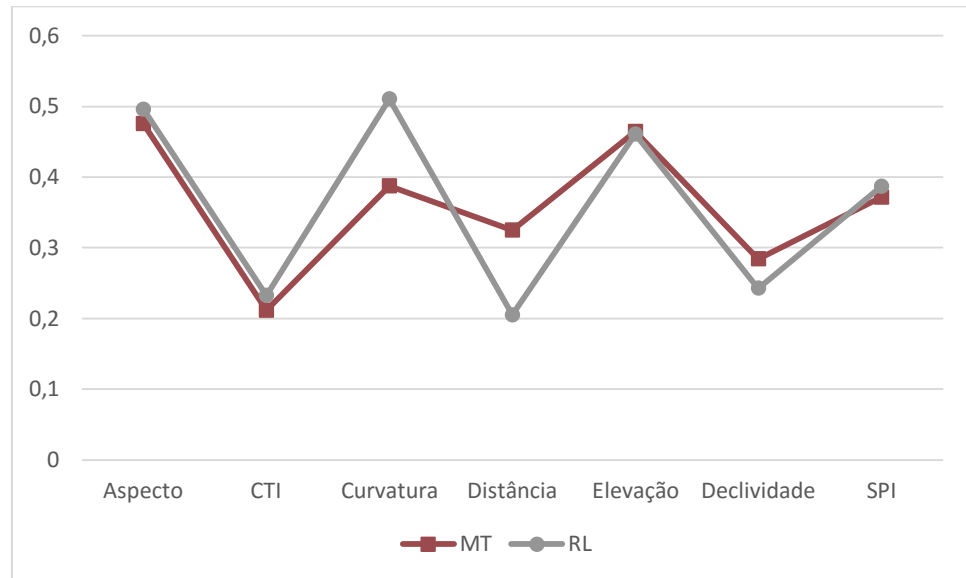


Figura 29 - Distribuição da média das variáveis dos dados de treino entre as classes MT e RL (valores normalizados entre 0 e 1).

Nas colunas LV e RQ, ambas as classes apresentam elevados valores de pixels designados incorretamente entre si. Isso ocorre pois estas classes possuem um ambiente de ocorrência semelhante, sendo originados em áreas de relevo plano e suave ondulado, o que explica a proximidade das variáveis Aspecto e Declividade, como apresentado na Figura 30. A variável Litologia tem grande importância na diferenciação destas classes, pois são oriundas de material de origem diferentes.

As classes NV e LV são formadas por materiais de origem semelhantes, como rochas básicas (basaltos e diabásios), e portanto, possuem Litologia semelhante, o que explica as inclusões incorretas destas classes entre si. As variáveis Curvatura, Distância, Elevação e Declividade são capazes de discriminar o ambiente de ocorrência entre estas classes (Figura 30), pois o NV ocorre em encostas onduladas, fortemente onduladas ou em áreas bem drenadas, próximas a cursos d'água, diferentemente do LV.

A classe LVA apresentou uma exatidão do produtor de 46,1%, onde o maior número de pixels classificados incorretamente foram para as classes LV, com 34 pixels, e RQ, com 45 pixels. As inclusões incorretas das classes LV e RQ acontecem pela semelhança do relevo em que ocorrem, plano ou suave ondulado,

apresentando valores próximos entre as variáveis Aspecto, CTI e Distância para o RQ, e Aspecto, CTI e SPI para o LV (Figura 30).

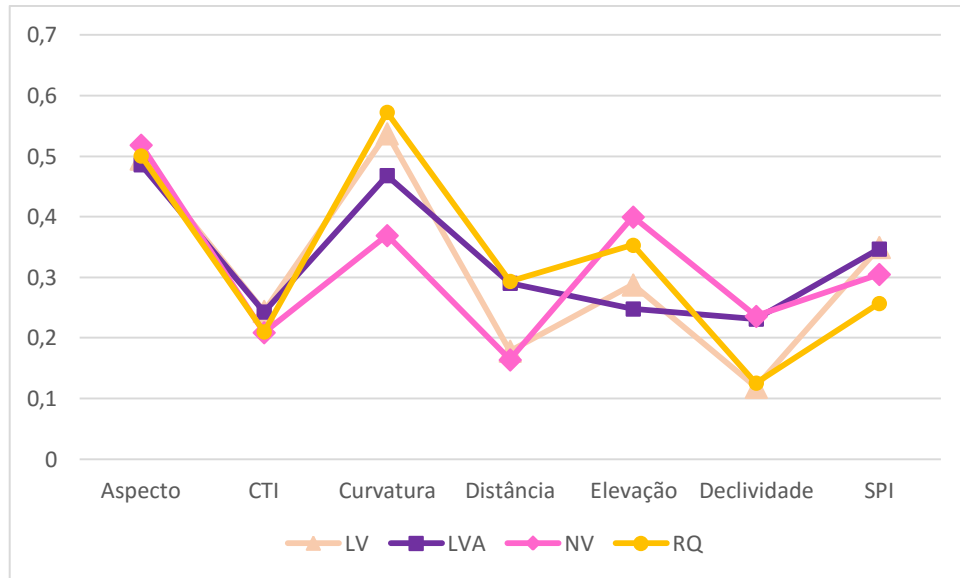


Figura 30 - Distribuição da média das variáveis dos dados de treino entre as classes LV, LVA, NV e RQ (valores normalizados entre 0 e 1).

Apesar da pequena quantidade de amostras de teste, a classe GX apresentou uma exatidão do produtor de 59,1%. Isto se deve ao fato desta classe possuir uma grande discriminação entre as classes, como pode ser observado na Figura 31. Do total de pixels desta classe, 2 foram classificados na classe CX e 7 na classe OX. Estas classes incluídas incorretamente na coluna GX correspondem a solos que tem ambiente de ocorrência semelhante.

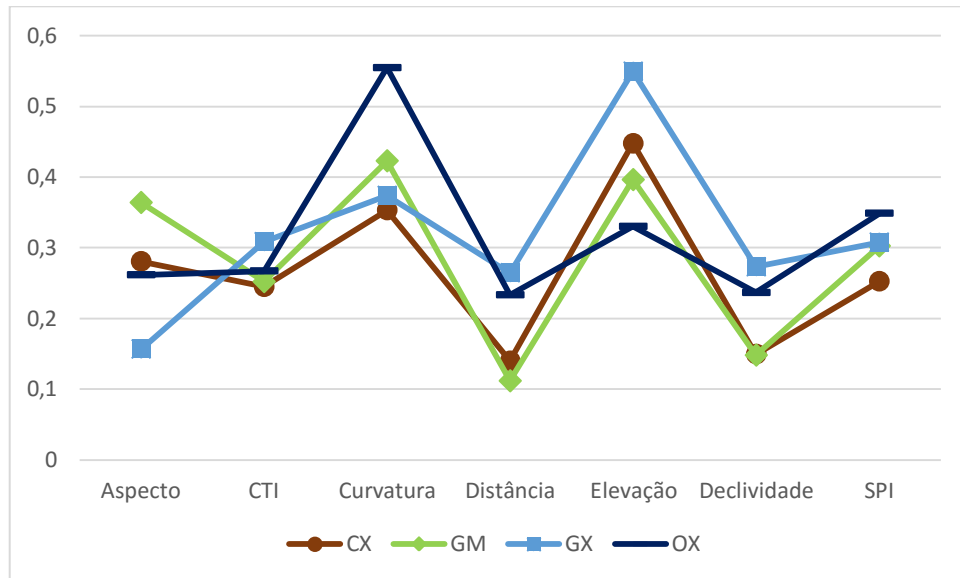


Figura 31 - Distribuição da média das variáveis dos dados de treino entre as classes CX, GM, GX e OX (valores normalizados entre 0 e 1).

Para o modelo no terceiro nível categórico, a avaliação da acurácia resultou em uma exatidão global de 57,1% e um índice Kappa de 0,57. Estes resultados demonstram uma perda de acurácia em relação ao modelo aplicado para as classes no segundo nível categórico, onde há a generalização de classes. De acordo com a classificação proposta por Landis & Koch (1977), o grau de concordância é moderado, pois o índice Kappa situa-se entre 0,41 e 0,60.

Giasson et al. (2008) chegaram a um resultado semelhante utilizando regressões logísticas, onde obtiveram um índice Kappa de 0,36 para a legenda original e de 0,54 para uma legenda simplificada, ou seja, para a generalização das classes.

A matriz de confusão para o modelo no terceiro nível categórico é apresentada na Tabela 19.

Tabela 19 - Matriz de confusão do modelo Random Forest com amostragem A3, $n_{tree} = 300$ e $m_{try} = 4$ no terceiro nível categórico.

Classes	CXbd	GMbd	GXbd	LVAd	LVd	LVdf	LVef	LVwf	MTo	NVef	OXs	RLe	RQo	Total	EU
CXbd	84	13	1	2	0	1	0	2	0	0	20	0	1	124	67.7
GMbd	7	222	0	10	5	0	2	17	2	1	21	2	30	319	69.6
GXbd	0	1	14	0	0	0	0	0	0	0	2	0	0	17	82.4
LVAd	1	9	0	85	9	2	1	12	0	0	0	0	9	128	66.4
LVd	0	1	0	3	7	0	0	3	1	1	2	0	6	24	29.2
LVdf	0	0	0	3	1	21	7	3	0	3	0	4	0	42	50.0
LVef	0	0	0	1	3	21	28	9	0	15	0	10	3	90	31.1
LVwf	3	13	0	13	26	24	40	232	1	21	1	29	4	407	57.0
MTo	0	0	0	0	0	0	1	1	4	0	0	1	0	7	57.1
NVef	0	2	0	1	3	20	33	39	2	145	0	35	2	282	51.4
OXs	42	22	7	4	1	0	1	0	0	2	135	0	7	221	61.1
RLe	0	0	0	2	4	17	24	25	8	58	0	188	6	332	56.6
RQo	1	19	0	56	54	18	6	31	17	6	3	17	238	466	51.1
Total	138	302	22	180	113	124	143	374	35	252	184	286	306	2459	
EP	60.9	73.5	63.6	47.2	6.2	16.9	19.6	62.0	11.4	57.5	73.4	65.7	77.8		
EG	57.1														
Kappa	0.57														

O terceiro nível categórico em relação ao segundo nível categórico, apresentou um sensível aumento nos valores de exatidão do produtor nas classes CXbd, GXbd, LVAd, OXs e RLe, enquanto as classes NVef e RQo apresentaram um elevado aumento. Conforme visto anteriormente, no segundo nível categórico houve uma confusão da classe LV com as classes RQ e NV. Com a divisão da classe LV em LVd, LVdf, LVef e LVwf, diminuiu-se o número de inclusões incorretas em RQo e NVef, que correspondem as mesmas classes do segundo nível categórico, apenas com legendas diferentes.

A classe LV na matriz do segundo nível categórico apresentou uma exatidão do produtor de 64,7%, com 488 pixels classificados corretamente. Já no terceiro nível categórico, o melhor resultado foi para a classe LVwf, com 62%, enquanto as classes LVd, LVdf e LVef apresentaram valores abaixo de 20%. Além disso, a soma dos pixels classificados corretamente entre estas classes foi de 288, ou seja, 200

pixels a mais foram classificados incorretamente, o que explica a queda na exatidão global e no índice Kappa.

Observa-se na Figura 32 que grande parte das variáveis são incapazes de discriminar as classes LVd, LVdf, LVef e LVwf umas das outras, o que fez com que estas classes tivessem um número maior de pixels designados incorretamente entre si, e dessa forma, diminuiu-se as inclusões incorretas nas classes NV e RQ.

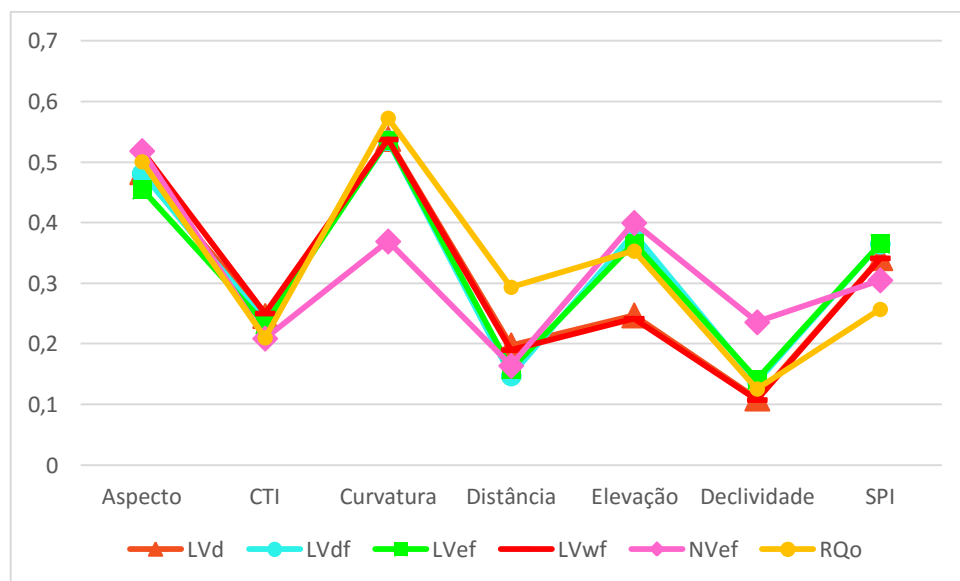


Figura 32 - Distribuição da média das variáveis dos dados de treino entre as classes LVd, LVdf, LVef, LVwf, NVef e RQo (valores normalizados entre 0 e 1).

Alguns autores utilizando o método de árvores de decisão obtiveram os seguintes resultados: Giasson et al. (2011) com um total amostrado de 0,54% obtiveram um Kappa de 0,52; Coelho & Giasson (2010) com um total amostrado de 8,75% obtiveram um Kappa de 0,38; Crivelenti (2009) com um total amostrado de 10% obteve um Kappa de 0,43. Sendo assim, os valores obtidos pelo método Random Forest com um total amostrado de 0,43% (amostragem A3) para os dois níveis categóricos se mostraram satisfatórios, se comparados a estes trabalhos.

4.5. Importância das variáveis

A importância das variáveis utilizadas na criação dos modelos para o segundo e terceiro nível categórico são apresentadas por ordem de importância na Figura 33.

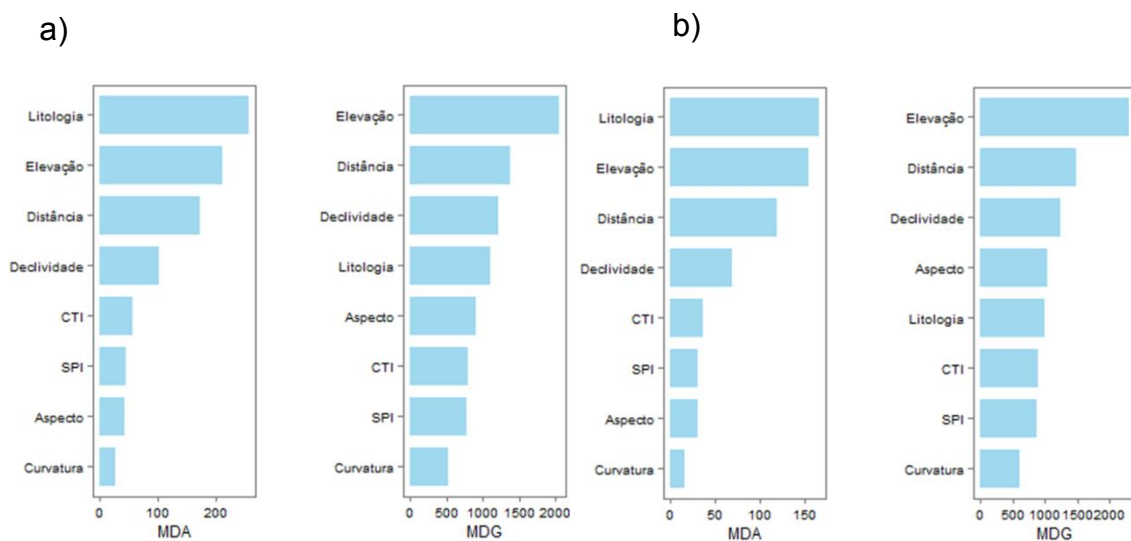


Figura 33 - Gráfico de importância das variáveis pelas estimativas MDA e MDG. a) segundo nível categórico; b) terceiro nível categórico.

Em uma comparação da importância das variáveis do segundo nível categórico com o terceiro, nota-se que as estimativas MDA mantiveram-se semelhantes, enquanto na estimativa MDG houve o aumento da importância para a variável Aspecto no terceiro nível categórico.

A variável Litologia possui a maior importância no MDA, ou seja, sua exclusão ou permutação representaria um aumento na taxa de erro. Na estimativa de importância MDG, a Litologia aparece em quarto lugar. Isso ocorre devido a variável possuir apenas três unidades litológicas, o que não contribui para uma maior homogeneidade dos nós e das folhas no modelo Random Forest. Sendo assim, é uma variável importante para caracterizar as classes de solos, porém não possui categorias suficientes para uma pureza que induza a uma classe de solo específica.

No segundo nível categórico, a variável Aspecto que aparece em penúltimo lugar na estimativa MDA, está em quinto lugar na MDG, o que se explica pela sua capacidade de diferenciação de algumas classes, principalmente a de solos que ocorrem em áreas planas, como CX, GM, GX e OX. Em ambas as estimativas a Curvatura apresentou os piores resultados, enquanto as variáveis CTI e SPI apresentaram valores semelhantes e baixos.

As variáveis Litologia, Elevação, Distância da rede de drenagem e Declividade estão nesta ordem entre as primeiras nas estimativas MDA e MDG, com uma grande importância na discriminação das classes de solo. De acordo com Chagas (2006), a elevação está entre os atributos mais efetivos para auxiliar na realização de levantamento de solos em média escala. Giasson et al. (2006) e Crivelenti (2009) utilizando metodologias diferentes de predição de classes solos, também obtiveram entre as variáveis mais importantes a Litologia, Elevação, Distância da rede de drenagem e Declividade.

4.6. Mapas das predições das classes de solos

O mapa digital de solos da quadrícula de Ribeirão Preto em comparação visual ao mapa original, no segundo nível categórico, é apresentado na Figura 34. A variação da área das classes do mapa digital em relação ao mapa original é apresentado na Figura 35.

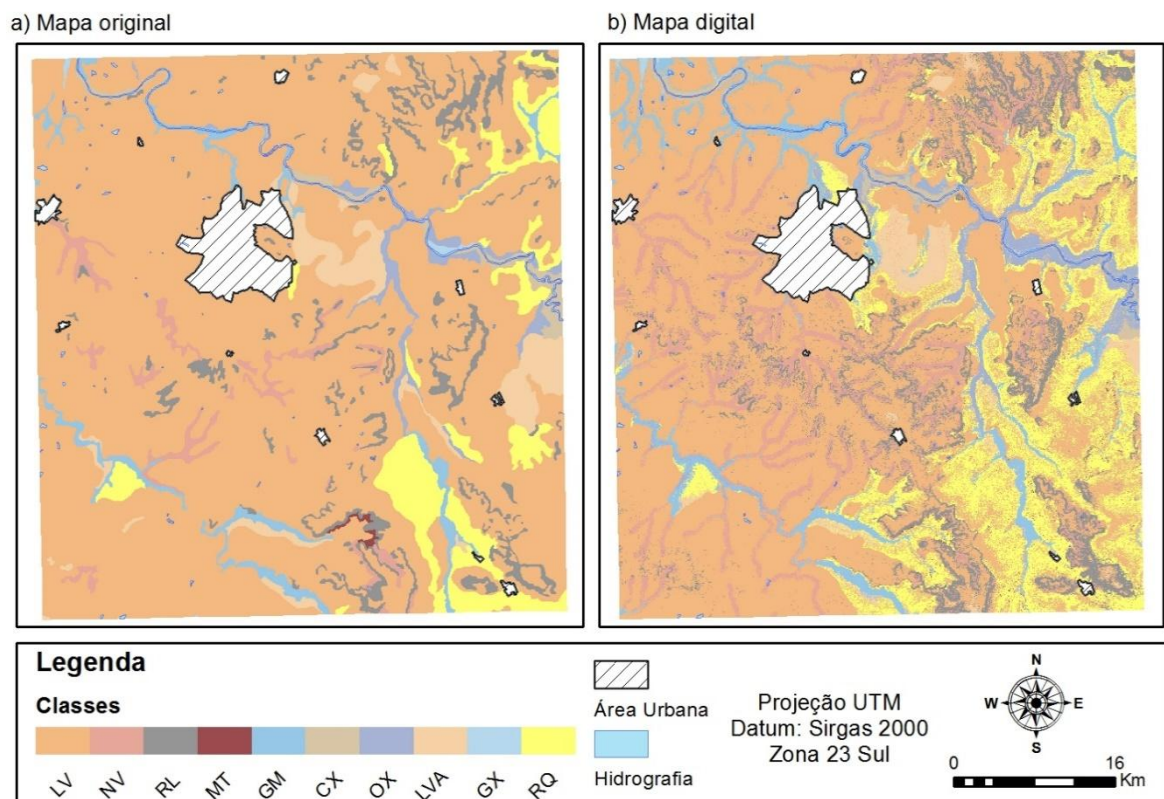


Figura 34 - Mapa de solos original (a) e mapa de solos digital (b) da quadrícula de Ribeirão Preto no segundo nível categórico.

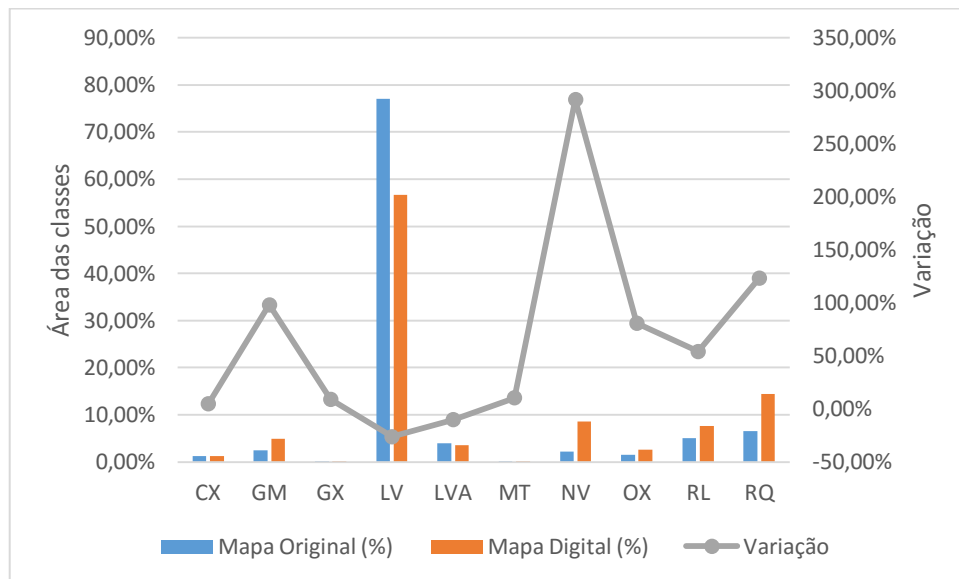


Figura 35 - Comparação e variação da área das classes de solo entre o mapa original e o mapa digital no segundo nível categórico.

A classe LV apresentou uma redução 56.349 ha (26,37%) em relação ao total de sua área, porém aparece ainda como a classe com maior representatividade na quadrícula, mesmo com a menor densidade de amostragem utilizada na criação do modelo. A classe LVA apresentou uma redução de 1105 ha (10,19%).

As classes RQ e NV foram as que apresentaram a maior variação positiva em relação a área de suas classes no mapa original, com um aumento de 22.098 ha (123,29%) e 17.646 ha (291,77%) respectivamente, podendo facilmente ser identificado visualmente na comparação dos mapas na Figura 34. As classes GM, OX e RL apresentaram um aumento de 6772 ha (98,02%), 3277 ha (80,72%) e 7438 ha (54,02%) respectivamente.

Uma menor variação foi apresentada pelas classes MT, GX e CX, que apesar de apresentarem as maiores densidades de amostras, tiveram sensível aumento de 33 ha (10,42%), 29 ha (8,93%) e 161 ha (4,76%) respectivamente.

O mapa digital apresentou 70,63%, de persistência, ou seja, classes que foram mantidas igualmente ao mapa original. Este valor é superior ao da exatidão global encontrado para o modelo aplicado nos dados de teste, que foi de 62,5%. O restante, 29,37% da área, apresentou mudanças nas classes de solo em relação ao

mapa original, que estão associadas aos erros do modelo, conforme pode ser observado na Figura 36.

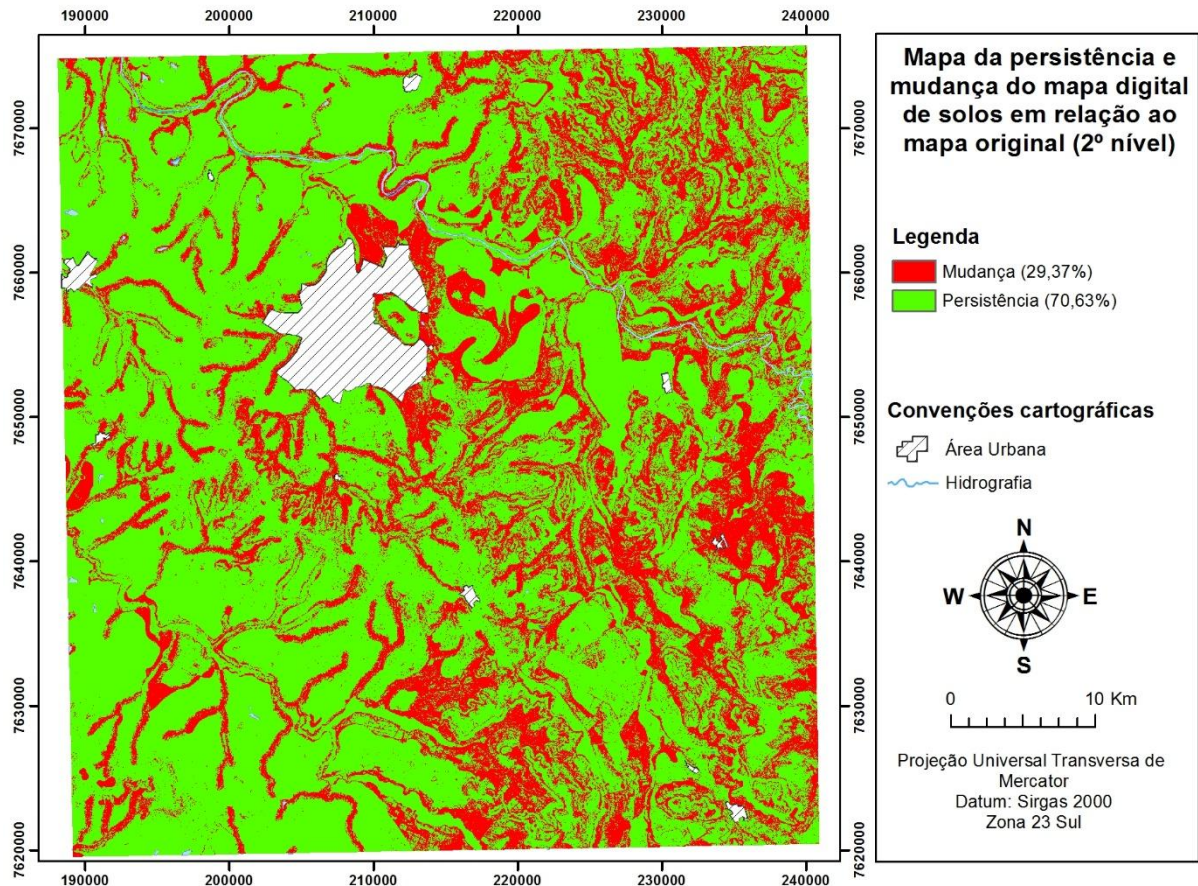


Figura 36 - Mapa da mudança ou persistência de classes no mapa digital em relação ao mapa original no segundo nível categórico.

Foi feita a tabulação da área entre as associações de classes e a área das mudanças em relação ao mapa original, ou seja, sob o erro do modelo, constatando-se a desagregação de alguns tipos de associações entre duas classes.

Na Figura 37 são apresentadas apenas as desagregações que obedeceram a proporção entre as classes com maior e menor área dentro da associação. Essas desagregações não respeitam fielmente as porcentagens indicadas no mapa original, de 60% para a primeira classe da legenda e 40% para a segunda (OLIVEIRA & PRADO, 1987). Além disso, outras classes foram incluídas em menor proporção além das indicadas nas legendas das associações de classes.

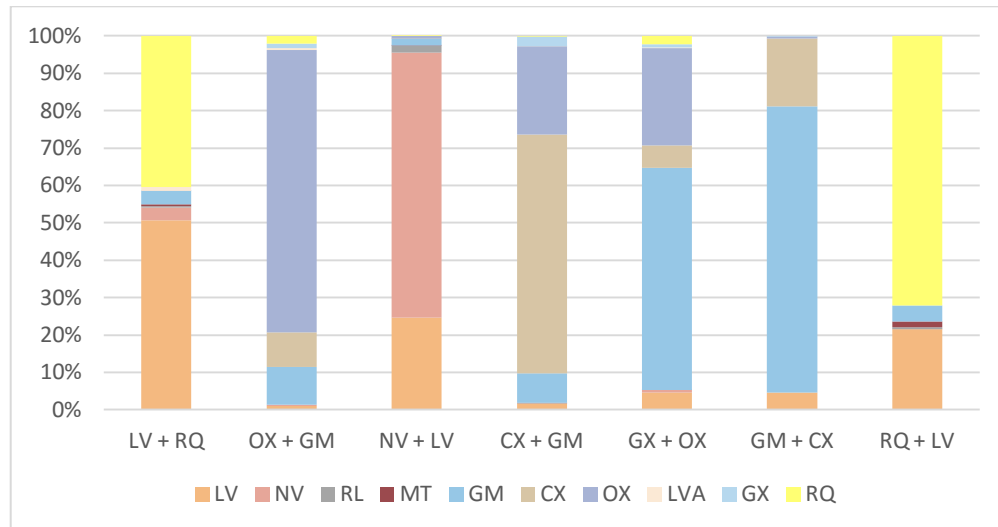


Figura 37 - Porcentagem das classes dentro das associações fragmentadas no mapa digital para o segundo nível categórico.

A maior parte da desagregação de classes dentro das associações ocorreu entre solos hidromórficos (GX, GM e OX) ou solos com ambiente de ocorrência semelhante, como a classe CX. As duas associações entre LV e RQ também foram desagregadas, em grande parte pela variável Litologia, já que as variáveis geomorfométricas apresentaram uma dificuldade na discriminação destas classes devido à semelhança do ambiente de ocorrência. Entre a associação NV + LV, a Litologia é semelhante, sendo as variáveis geomorfométricas responsáveis por essa desagregação.

Para o terceiro nível categórico, a comparação entre o mapa original e o mapa digital de solos é apresentado na Figura 38. A variação da área das classes do mapa digital em relação ao mapa original é apresentada no gráfico da Figura 39.

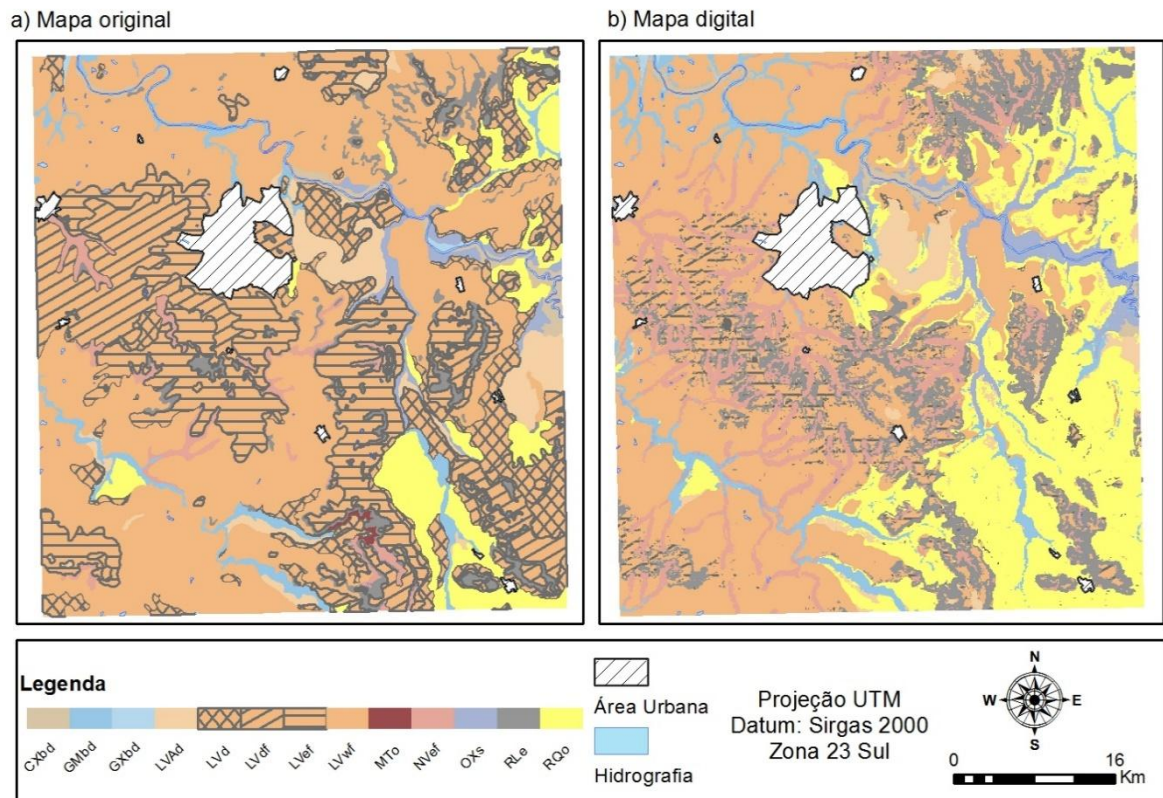


Figura 38 - Mapa de solos original (a) e mapa de solos digital (b) da quadrícula de Ribeirão Preto no terceiro nível categórico.

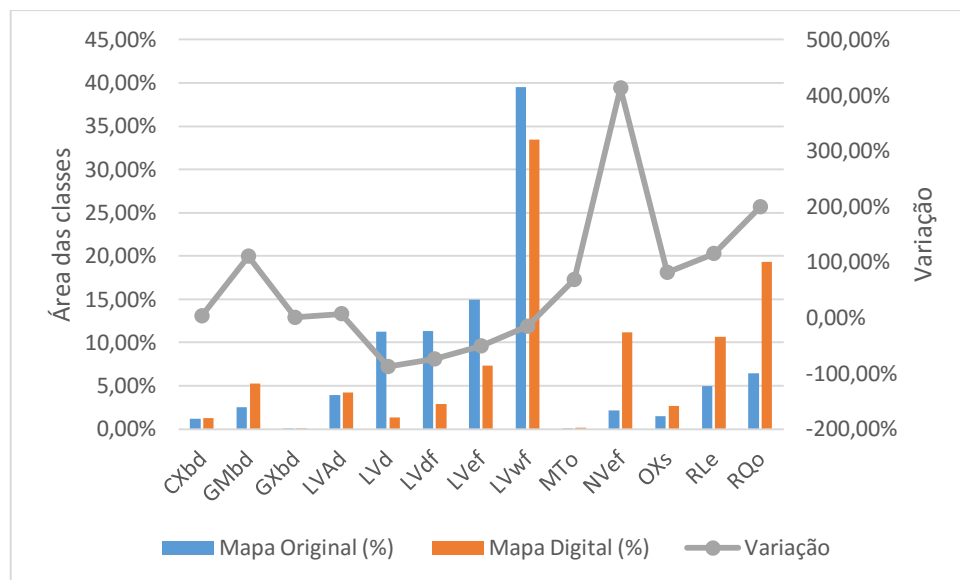


Figura 39 - Comparação e variação da área das classes de solo entre o mapa original e o mapa digital no terceiro nível categórico.

As classes LVd, LVdf, LVef e LVwf foram as únicas a apresentarem uma redução em relação ao total de sua área, totalizando uma redução de 88.643 há (228,17%), maior do que a apresentada na predição do segundo nível categórico para a classe LV. A classe LVwf permaneceu como a de maior representatividade na quadrícula, enquanto as classes LVd, LVdf e LVef perderam sua representatividade, ficando bem evidente na identificação visual da comparação entre os mapas da Figura 38.

As maiores variações foram para as classes NVef com 24.971 ha (412,9%), RQo com 35.723 ha (199,30%), RLe com 15.894 ha (115,43%) e GMbd com 7642 ha (110,41%). A classe OXs, apresentou um resultado semelhante ao segundo nível categórico, com aumento de 3306 ha (80,72%). A classe MTo, diferentemente do segundo nível categórico, apresentou um aumento de 217 ha (68,53%) em relação a sua área do mapa original. As classes CXbd, GXbd, e LVAd apresentaram um sensível aumento, inferior a 10%.

A persistência do mapa digital para o terceiro nível categórico foi de 55,76%, valor próximo ao obtido da exatidão global, de 57,1%, enquanto as mudanças em relação ao mapa original foram de 44,24%, maior do que a apresentada para o segundo nível categórico, resultado esperado, já que a exatidão global do modelo para a predição do terceiro nível categórico também foi menor.

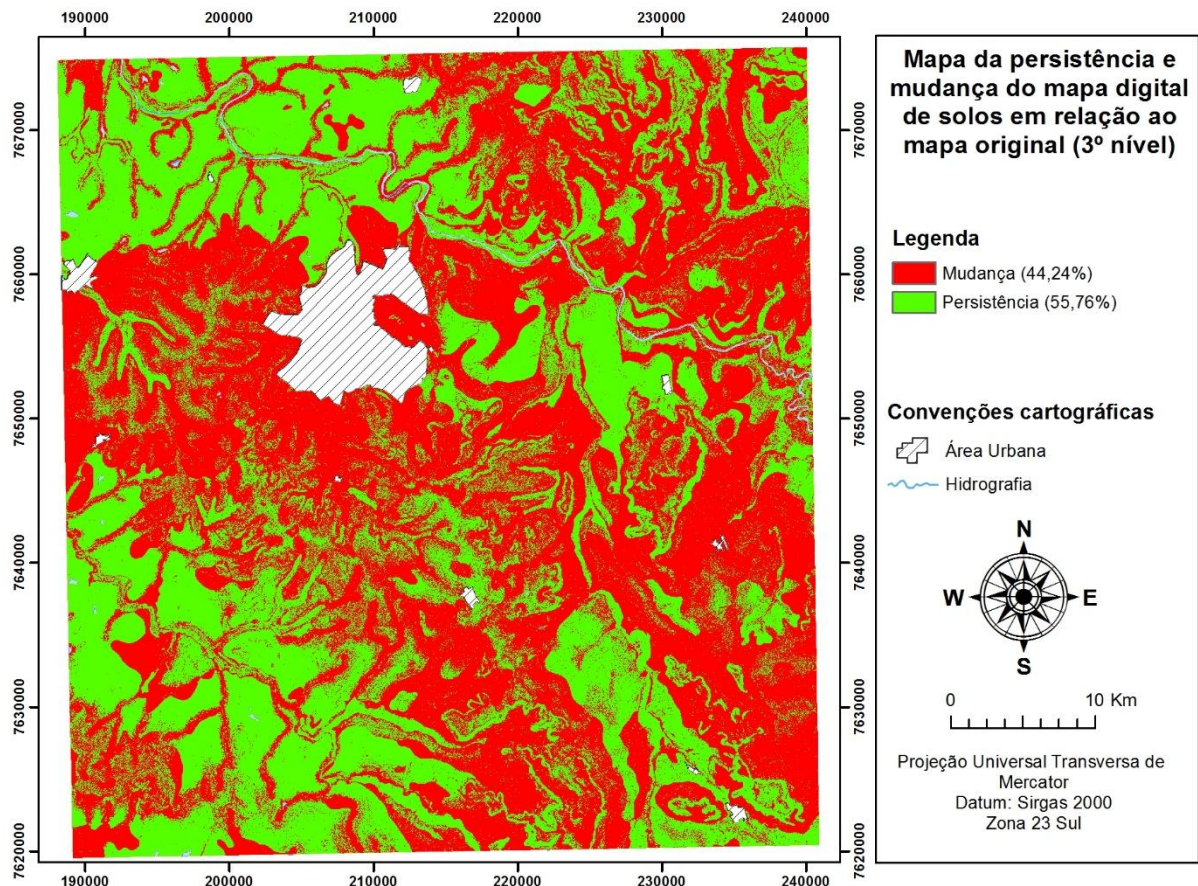


Figura 40 - Mapa da mudança ou persistência de classes no mapa digital em relação ao mapa original no terceiro nível categórico.

Apesar de uma elevada mudança em relação ao segundo nível categórico, houve um menor número de desagregações entre as classes associadas, que respeessem a condição de manter a proporção das classes com maior e menor área dentro da associação.

Diferentemente do segundo nível categórico, não houve um predomínio da desagregação de classes de solos hidromórficos ou mesmo de ambientes de ocorrência semelhantes. A desagregação das associações para o terceiro nível categórico é apresentado no gráfico da Figura 41.

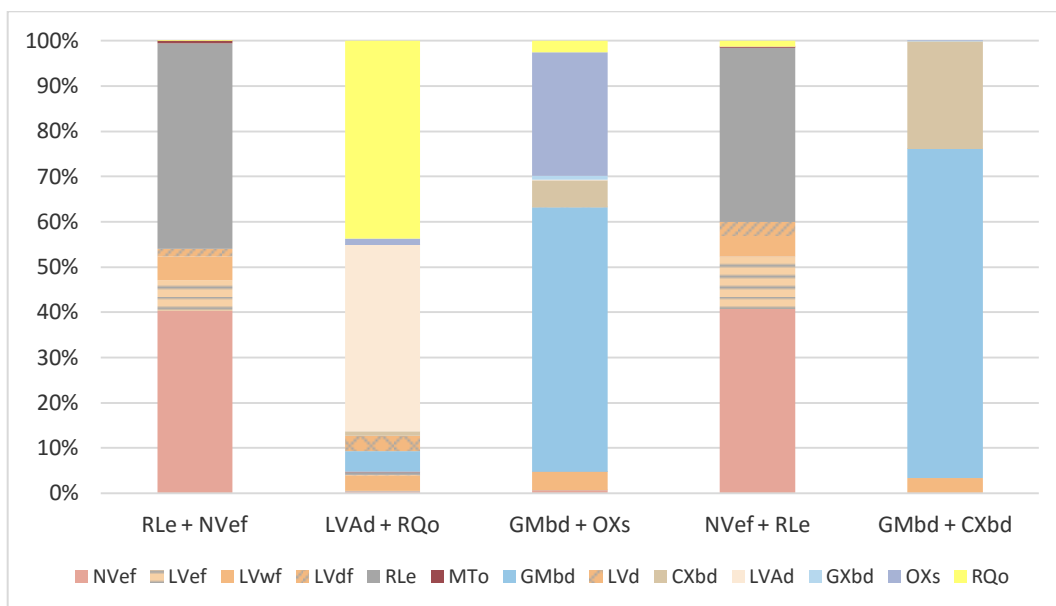


Figura 41 - Porcentagem das classes dentro das associações fragmentadas no mapa digital para o terceiro nível categórico.

As três classes que obtiveram o maior aumento de sua área em relação ao mapa original nos dois níveis categóricos analisados foram as classes NV (NVef), RQ (RQo) e GM (GMbd) nesta ordem.

O elevado aumento da área da classe NVef e de sua legenda simplificada NV, ocorre predominantemente a oeste da quadrícula, conforme pode ser visto pela interpretação visual dos mapas digitais, formando-se nas porções médias das vertentes das bacias do rio Mogi-Guaçu e do Pardo, onde predominam relevo suave a forte ondulado. É nesta região da quadrícula que predominam as rochas básicas (basalto e diabásio), o mesmo material de origem dos Latossolos. Com a aplicação das variáveis geomorfométricas, possivelmente houve um aumento na capacidade de distinção entre estas classes.

O aumento da área da classe RQo e de sua legenda simplificada RQ, se deve ao fato de que o mapa original foi elaborado a partir do mapa geológico do estado de São Paulo do IPT (1981), na escala 1:500.000, de acordo com o apresentado por Oliveira & Prado (1987). Este mapa apresenta uma escala cerca de 20 vezes inferior ao mapa elaborado por Sinelli et al. (1973) na escala 1:50.000, utilizado para a criação dos modelos de predição. Como no leste da quadrícula há um predomínio dos arenitos da formação Botucatu e Pirambóia, além de sedimentos

inconsolidados, sendo estes materiais de origem importantes na distinção da classe RQo das demais, nota-se um aumento predominante desta classe nesta região.

A classe GMbd e sua legenda simplificada GM, apesar de possuir uma semelhança no ambiente de ocorrência com solos hidromórficos e com a classe CXbd (CX), apresentou uma distinção destas classes por meio das variáveis geomorfométricas, e com isto apresentou um elevado aumento, mesmo possuindo uma densidade de amostras menor em relação a estas classes.

A partir das mudanças evidenciadas, foram contrastados entre mapa original e digital alguns pontos obtidos em campo, principalmente das classes que obtiveram os maiores ganhos de área. São mostrados apenas os perfis onde ocorre equivalência com as mudanças ocorridas no mapa digital. São apresentados cinco perfis distintos, contendo a verdade de campo das classes NVef, RQo, RLe e LVwf.

Tabela 20 - Caracterização do ponto 3.



Ponto 3

Classificação atual: Nitossolo Vermelho eutroférico

Localização: Folha de Ribeirão Preto.

Coordenadas 47,97 W e 21,24 S.

Município de Dumont - SP.

Altitude: 620 metros

Situação e declive: ondulado e 10%.

Relevo: ondulado.

Material de origem: intemperismo do basalto da formação Serra Geral.

Vegetação atual: Milho.

Responsável: Matheus Felipe Oliveira

Figura 42 - Perfil do Nitossolo Vermelho eutroférico no município de Dumont - SP.

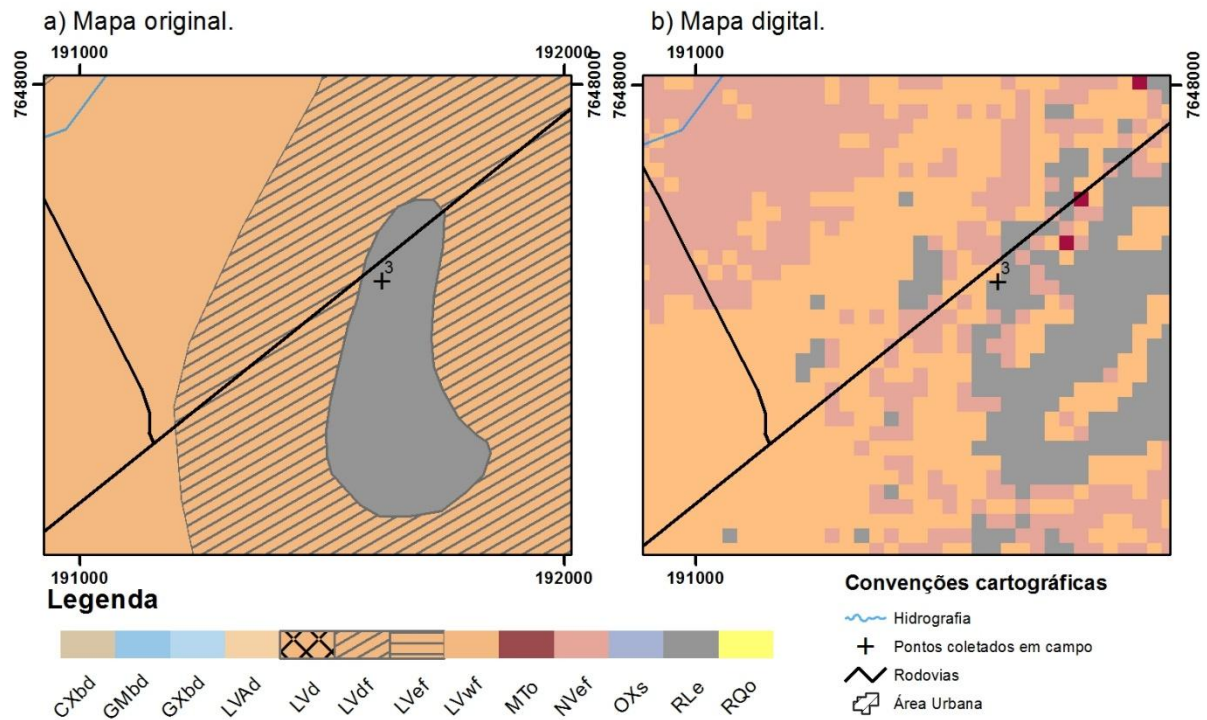


Figura 43 - Localização do ponto 3 obtido em campo e sua comparação no mapa original e no mapa digital do terceiro nível categórico.

Tabela 21- Caracterização do ponto 5.



Figura 44- Paisagem de ocorrência do Neossolo Quartzarênico órtico.

Ponto 5

Classificação atual: Neossolo Quartzarênico órtico

Localização: Folha de Ribeirão Preto.

Coordenadas 47,75 W e 21,17 S. Município de Ribeirão Preto - SP.

Altitude: 550 metros

Situação e declive: plano inferior a 3%.

Relevo: aplainado.

Material de origem: arenitos eólicos da formação Botucatu.

Vegetação primária: Cerradão.

Vegetação atual: Cana de açúcar.

Responsável: Matheus Felipe Oliveira

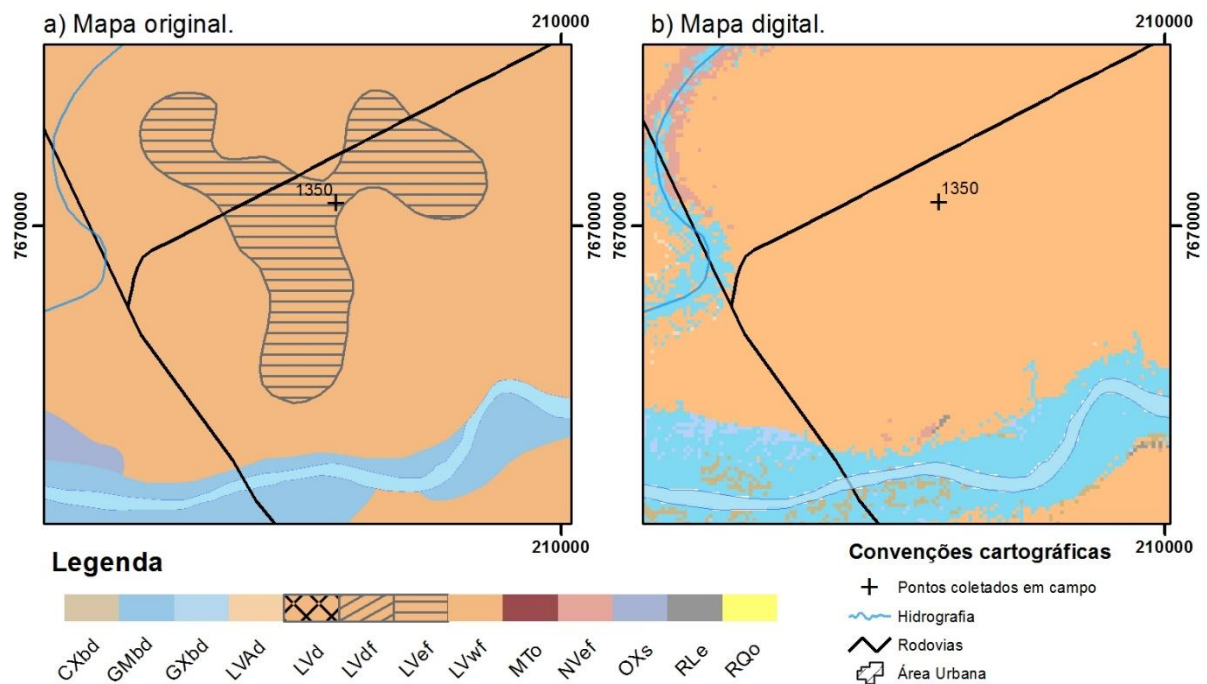


Figura 47 - Localização do perfil 1350 e sua comparação no mapa original e no mapa digital do terceiro nível categórico.

Tabela 23 - Caracterização do perfil 1346.



Figura 48 - Perfil do Neossolo Litólico distrófico no município de Luís Antônio - SP. Fonte: Oliveira & Prado (1987).

Perfil 1346

Classificação antiga: Latossolo Roxo concrecionário

Classificação atual: Neossolo Litólico eutrófico

Localização: Folha de Cravinhos.

Coordenadas UTM 218-220km E. e 7622-7624N. Município de Luís Antônio - SP.

Altitude: 750 metros

Situação e declive: topo aplainado de morrote inferior a 2%.

Relevo: aplainado.

Material de origem: produtos de intemperização de diabásio, afetados por retrabalhamento do local.

Vegetação primária: Mata subtropical caducifólia.

Responsável: Oliveira & Prado (1987)

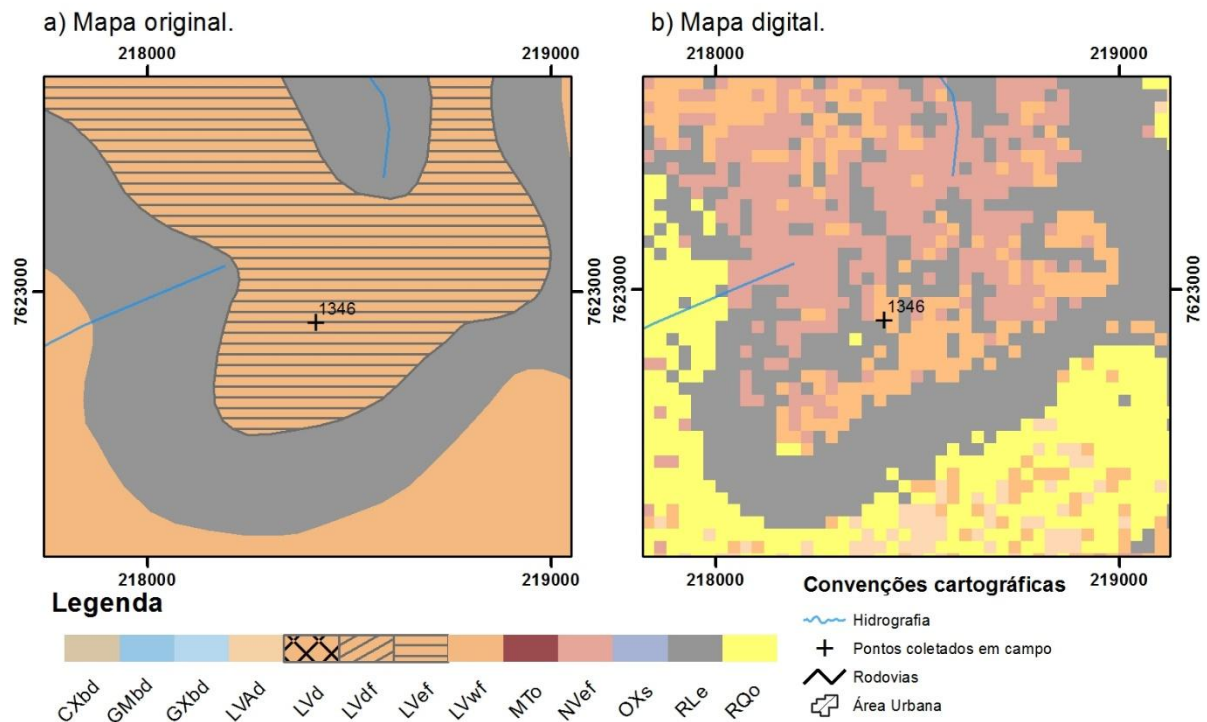


Tabela 24- Caracterização do ponto 33.

Ponto 33

Classificação antiga: Terra Roxa

Classificação atual: Nitossolo Vermelho

Localização: Folha de Ribeirão Preto. Coordenadas 47,88 W e 21.2 S.

Altitude: 600 metros

Situação e declive: plano inferior a 6%.

Relevo: ondulado.

Material de origem: intemperismo do basalto da formação Serra Geral.

Responsável: RADAM Brasil

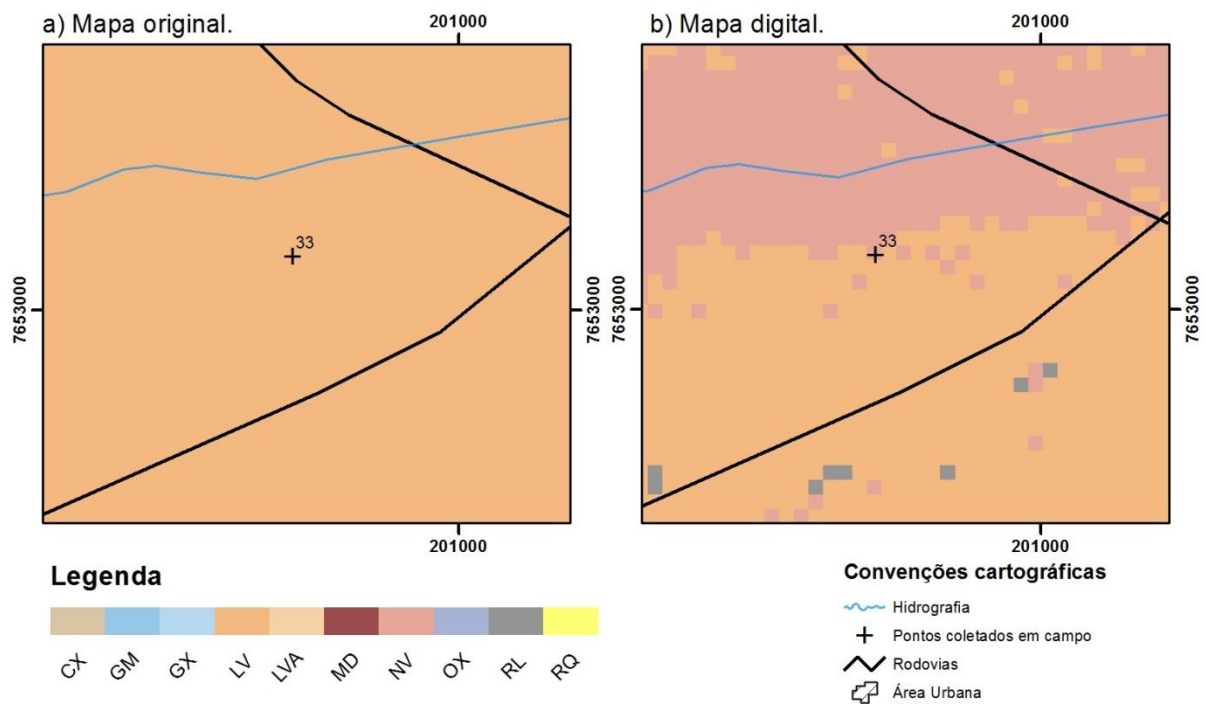


Figura 50 - Localização do perfil 33 e sua comparação no mapa original e no mapa digital do segundo nível categórico.

Observa-se que a maioria das equivalências entre a verdade de campo e as mudanças ocorrem no mapa digital do terceiro nível categórico, isso devido a maior diferença em relação ao mapa original (44,24%) se comparado ao segundo nível categórico (29,37%). O ponto 33 obtido do RADAMBRASIL teve equivalência apenas no segundo nível categórico. Já o ponto 5 obtido em campo e o perfil 1350 do Boletim Científico, tem equivalência tanto no terceiro nível categórico quanto no segundo.

Esta verdade de campo mostra a evidência das incertezas quanto ao nível de detalhe do mapa tomado como referência. Estas incertezas já haviam sido apresentadas anteriormente, pela diferença entre a escala nominal apresentada e sua escala efetiva, que questionam o nível de detalhe presumido pelo mapa. Estas incertezas podem ser inerentes ao processo de delimitação de classes por meio do método convencional ou podem estar relacionadas a homogeneidade da área em questão, que possui um predomínio da classe Latossolo Vermelho.

Ainda que estas informações mostrem que o mapa digital correspondeu em alguns casos a verdade de campo, o número de pontos avaliados é pouco

significativo para uma validação das classes preditas nas áreas que não correspondem ao mapa original, porém, de acordo com Zhou et al. (2014), as restrições financeiras e de pessoal, tornam difíceis os trabalhos de campo para coleta de amostras suficientes para uma avaliação do erro estaticamente significativa, principalmente em uma área com a extensão da quadrícula de Ribeirão Preto.

5. CONCLUSÕES

Os resultados obtidos permitem concluir que:

- a) O método Random Forest juntamente com as variáveis utilizadas, foram capazes de compreender as relações solo-paisagem na representação do mapa original.
- b) A generalização das classes contribuiu para um maior ganho preditivo do modelo.
- c) Com uma pequena densidade de amostragem em relação ao total da área, conseguiu-se obter resultados satisfatórios na representação do mapa original tomado como referência, se comparado a trabalhos semelhantes
- d) As variáveis geomorfológicas Elevação, Declividade e Distância da rede de drenagem e a variável Litologia, foram as mais importantes para explicar as correlações ambientais da área.
- e) A avaliação preliminar dos modelos Random Forest e a estratificação das amostragens por classes tiveram grande importância para a criação de um modelo que apresentasse uma capacidade relevante de processamento dos dados com uma baixa taxa de incerteza.
- f) Após criado a rotina do modelo em ambiente computacional, é possível seu fácil manuseio para ajuste e inserção de novas variáveis, utilizando dados e sistemas disponíveis gratuitamente.
- g) Esta metodologia pode ser empregada para subsidiar levantamentos de solos, podendo ser extrapolados para áreas com características semelhantes, auxiliando nos delineamentos de maneira menos subjetiva com um menor gasto tempo.

- h) É possível desagregar associações de classes, porém é necessário em estudos futuros, uma comparação com dados obtidos em campo para validação dos resultados.

6. REFERÊNCIAS

- AGUIAR, D. A. et al. Remote Sensing Images in Support of Environmental Protocol: Monitoring the Sugarcane Harvest in São Paulo State, Brazil. **Remote Sensing**, 2011. Disponível em: <<http://www.mdpi.com/2072-4292/3/12/2682>>. Acesso em: 03 Mai. 2014.
- ALMEIDA, F. F. M. et al. **Mapa Geológico do Estado de São Paulo: Nota Explicativa**. São Paulo: Instituto de Pesquisas Tecnológicas - IPT, v. 1, 1981. 126 p. Escala 1:500.000.
- BENEDETTI, M. M. et al. Representatividade e potencial de utilização de um banco de dados de solos do Brasil. **Revista Brasileira de Ciência do Solo [Online]**, 32, n. 6, 2008. 2591-2600. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832008000600036&lng=en&nrm=iso>. Acesso em: 22 Nov. 2014.
- BHERING, S. B. et al. Influência do modelo digital de elevação no mapeamento digital de atributos do solo por random forest (RF). **Anais do XXXV Congresso Brasileiro de Ciência do Solo**, Natal - RN, 2015. Disponível em: <http://www.cbcs2015.com.br/anais/index_int.php?id_trabalho=1443&ano=&ev=#menuanais>. Acesso em: 10 Nov. 2015.
- BREIMAN, L. Random Forests. **Machine Learnings**, v. 45, p. 5-32, 2001. Disponível em: <<http://link.springer.com/article/10.1023%2FA%3A1010933404324>>. Acesso em: 05 Mai. 2015.
- BREIMAN, L.; FRIEDMAN, J. H.; OSHLEN, R. A. **Classification and Regression Trees**. Wadsworth: [s.n.], 1984.
- BURROUGH, P. A.; MCDONNELL, R. A. Principle of Geographical Information Systems. **Oxford University Press**, p. 17-34, 1988. Disponível em: <<http://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo->

IV/Burrough%20Peter%20A%20y%20McDonnell%20Rachael%20A%20(1998)%20Principles%20of%20geographical%20information%20systems.pdf>. Acesso em: 03 Dez. 2014.

CARVALHO, C. C. N. et al. Histórico do levantamento de solos no Brasil. **Revista Brasileira de Cartografia**, n. 65/5, 2013. 997-1013. Disponível em: <<http://www.lsie.unb.br/rbc/index.php/rbc/article/download/722/630>>. Acesso em: 02 Dez. 2014.

CARRÉ, F.; MCBRATNEY, A. B. Digital terrain mapping. **Geoderma**, v. 128, p. 340-353, 2005.

CHAGAS, C. D. S. et al. MAPEAMENTO DIGITAL DA CTC EM SOLOS DO SEMIÁRIDO BRASILEIRO. **Anais do XXXV Congresso Brasileiro de Ciência do Solo**, Natal - RN, 2015. Disponível em: <http://www.cbcs2015.com.br/anais/index_int.php?id_trabalho=365&ano=&ev=#menuanais>. Acesso em: 10 Nov. 2015.

CHAGAS, C. S. **Mapeamento Digital de Solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. Viçosa: Tese (Doutorado), 2006. 223 p. Disponível em: <http://www.tede.ufv.br/tedesimplificado/tde_arquivos/25/TDE-2006-11-07T142021Z-64/Publico/texto%20completo.pdf>. Acesso em: 03 Jun. 2014.

COELHO, F. F. **Comparação de métodos no mapeamento digital de solos através de variáveis geomorfométricas e sistemas de informações geográficas**. Porto Alegre: Dissertação de Mestrado, 2010. Disponível em: <http://www.ufrgs.br/srm/ppgsr/publicacoes/Dissert.fabricio_coelho.pdf>. Acesso em: 12 Jul. 2014.

COELHO, F. F.; GIASSON, E. Métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, Santa Maria, v. 40, n. 10, p. 2099-2106, 13 Out. 2010.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2ª. ed. Boca Raton: CRC Press, 2009.

CRIVELANTI, R. C. **Mineração de dados para inferência da relação solo-paisagem em mapeamentos digitais de solos**. Campinas: Dissertação (Mestrado em Concentração em Gestão de Recursos Agroambientais) - Instituto Agrônomo,

2009. Disponível em:
<http://www.iac.sp.gov.br/areadoinstituto/posgraduacao/dissertacoes/pb1219307.pdf>
 >. Acesso em: 17 Mar. 2014.
- CUTLER, R. D. et al. Random Forests for classification in Ecology. **Ecology**, v. 88, p. 2783-2792, 2007. Disponível em:
[http://www.sciencedirect.com/science/refhub/S0016-7061\(13\)00344-3/rf0060](http://www.sciencedirect.com/science/refhub/S0016-7061(13)00344-3/rf0060).
 Acesso em: 03 Mai. 2015.
- DEMATTE, J. A. M. et al. Comparação entre mapas de solos obtidos por sensoramento remoto espectral e pelo método convencional. **Pesquisa agropecuária brasileira**, Brasília, 39, 2004. 1219-1229. Disponível em:
<http://www.scielo.br/pdf/pab/v39n12/22864.pdf>>. Acesso em: 03 Ago. 2014.
- DIAS, C. Solo brasileiro agora tem mapeamento digital. **Embrapa Solos**, 2014. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/2062813/solo-brasileiro-agora-tem-mapeamento-digital>>. Acesso em: 22 Nov. 2014.
- EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 3. ed. Rio de Janeiro: Embrapa Solos, 2013.
- EMBRAPA. CENTRO NACIONAL DE PESQUISAS DE SOLOS. **Procedimentos normativos para levantamentos pedológicos**. Brasília: Embrapa-SPI, 1995. 166 p. Disponível em:
<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/81913/1/PROCEDIMENTOS-NORMATIVOS-MANUAL.pdf>>. Acesso em: 16 Out. 2014.
- FERREIRA, S. T. **Estudo da vulnerabilidade à contaminação da águas subterrâneas na Região de Ribeirão Preto-SP**. Rio Claro: Dissertação (Mestrado) - Universidade Estadual Paulista (UNESP) Campus de Rio Claro, 1992. 80 p.
- FORBES, T. R.; ROSSITER, D. & V. W. A. **Guidelines for evaluating the adequacy of soil resource**. Ithaca: Cornell University Department of Agronomy, Soil Management Support Services (SMSS), Soil Conservation Service, USDA, 1987. 52 p. Disponível em: <http://eprints.icrisat.ac.in/8766/1/RP%2D02147.pdf>>. Acesso em: 20 Mar. 2015.
- GIASSON, E. et al. Decision trees for digital soil mapping on subtropical basaltic steep slopes. **Scientia Agrícola**, v. 68, p. 167-174, 2011. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162011000200006&lng=en&nrm=iso>. Acesso em: 14 Out. 2014.

HÄRING, T. et al. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. **Geoderma**, 2012. Disponível em: <http://www.researchgate.net/publication/255721682_Spatial_disaggregation_of_complex_soil_map_units_A_decision_tree_based_approach_in_Bavarian_forest_soils>. Acesso em: 20 Out. 2015.

HENGL, T. **Pedometric mapping**: bridging the gaps between conventional and pedometric approaches. Enschede: Wageningen University - PhD Thesis, 2003. 200 p.

HEUNG, B.; BULMER, C.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, p. 141-154, 2014. Disponível em: <http://www.researchgate.net/publication/259089742_Predictive_soil_parent_material_mapping_at_a_regional-scale_A_Random_Forest_approach>. Acesso em: 20 Mar. 2015.

IBGE. **Manual Técnico de Pedologia**. 2. ed. Rio de Janeiro: Manuais Técnicos em Geociências, v. 4, 2007. Disponível em: <ftp://geoftp.ibge.gov.br/documentos/recursos_naturais/manuais_tecnicos/manual_tecnico_pedologia.pdf>. Acesso em: 01 Mai. 2014.

IPT. INSTITUTO DE PESQUISAS TECNOLÓGICAS. **Mapa geológico do Estado de São Paulo**. São Paulo: (Publicação 1184. Memorial descritivo. Escala 1:500.000, Color), 1981.

JENNY, H. **Factors of soil formation, a system of quantitative pedology**. New York: McGraw-Hill, 1941. 281 p.

KOENOW-PINHEIRO, H. S. et al. MODELOS BASEADOS EM ÁRVORES PARA PREDIÇÃO DE CLASSES DE SOLO. **Anais do XXXV Congresso Brasileiro de Ciência do Solo**, Natal - RN, 2015. Disponível em: <http://www.cbcs2015.com.br/anais/index_int.php?id_trabalho=351&ano=&ev=#menuanais>. Acesso em: 10 Nov. 2015.

KOHLHEPP, G. Análise da situação da produção de etanol e biodiesel no Brasil. **Estudos Avançados**, São Paulo, v. 24, n. 68, 2010. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010340142010000100017&lng=en&nrm=iso>. Acesso em: 15 Jan. 2014.

KOTCHETKOFF-HENRIQUES, O. **Caracterização da vegetação natural de Ribeirão Preto, SP - Bases para conservação**. Ribeirão Preto: Tese de Doutorado - Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2003.

LAGACHERIE, P.; MCBRATNEY, A. B. A spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. B.; VOLTZ, M. **Digital Soil Mapping: An introductory perspective**. [S.l.]: [s.n.], v. 31, 2006. p. 3-22.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, n. 33, p. 159-174, 1977.

LEPSCH, I. F. et al. **Manual para levantamento utilitário do Meio Físico e Classificação de Terras no Sistema de Capacidade de Uso**. Campinas: Sociedade Brasileira de Ciência do Solo, 1991.

LIAW, A.; WIENER, W. Classification and regression by random forest. **R News** 2, p. 18-22.

MCBRATNEY, A. B. et al. An overview of pedometric techniques for use in soil. **Geoderma**, v. 97, p. 293-327, 2000.

MCBRATNEY, A. B. et al. On a Digital Soil Mapping. **Geoderma**, 17, 2003. 3-52. Disponível em:

<<https://djfextranet.agrsci.dk/sites/soilsensors/public/Documents/MHG/On%20digital%20soil%20mapping.pdf>>. Acesso em: 03 Dez. 2014.

MENDONÇA-SANTOS, M. D. L.; MANZATTO, C. V. Pedologia ou Pedometria? Modelo discreto ou modelo contínuo? Mapeamento tradicional ou digital? **Boletim Informativo - Sociedade Brasileira de Ciência do Solo**, v. 32, n. 1, p. 21-26, 2007.

MENDONÇA-SANTOS, M. L.; SANTOS, H. G. Mapeamento digital de classes e atributos de solos: métodos, paradigmas e novas técnicas. **Documentos**, Rio de Janeiro, v. 55, 2003.

MOORE, I. D. et al. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, v. 57, p. 443-452, 1993.

- NEUMANN, M. R. B. **Mapeamento Digital de Solos, no Distrito Federal**. Brasília: Tese de Doutorado - Insituto de Geociências da Universiade de Brasília, 2012.
- OLIVEIRA, J. B. **Pedologia aplicada**. 2. ed. Piracicaba: FEELQ, 2005.
- OLIVEIRA, J. B.; PRADO, H. **Levantamento pedológico semidetalhado do Estado de São**: quadrícula de Ribeirão Preto. Campinas: Instituto Agrônômico de Campinas, 1987.
- OLIVEIRA, S. A. M. et al. Digitalização e disponibilização dos filmes originais dos Projetos RADAM/RADAMBRASIL - Exemplos para aplicação. **Anais XV Simpósio Brasileiro de Sensoriamento Remoto**, Curitiba, p. 8303-8309, 2011. Disponível em: <<http://www.dsr.inpe.br/sbsr2011/files/p0440.pdf>>. Acesso em: 20 jan. 2015.
- OLIVEIRA, S. R. D. M.; ZURMELY, H. R. Sistema de Informação de Solos Brasileiro. **SISOLOS**, 2014. Disponível em: <<http://www.sisolos.cnptia.embrapa.br/>>. Acesso em: 28 Abr. 2015.
- OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. Ribeirão Preto: Dissertação (Mestrado em Bioinformática) - Universidade de São Paulo, 2013.
- PEREIRA, I. Z. C.; FERNANDES, M. N.; MENEZES, P. M. L. Modelo conceitual e procedimentos metodológicos para a análise e interpretação de imagens no Projeto RADAM. **Anais**.
- PIROLI, E. L. **Introdução ao geoprocessamento**. Ourinhos: Unesp - Campus Experimental de Ourinhos, 2010.
- QI, F.; ZHU, A. X. Knowledge discovery from soil maps using inductive learning. **International Journal of Geographical Information Science**, v. 17, n. 8, p. 771-795, 2003. Disponível em: <http://solimserver.geography.wisc.edu/pdfs/QiFeng_IJGIS2003.pdf>. Acesso em: 27 Out. 2014.
- R DEVELOPMENT CORE TEAM. R: a Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, Viena, austria, 2015. Disponível em: <www.R-project.org>.
- ROSSI, M.; OLIVEIRA, J. B. O mapa pedológico do Estado de São Paulo. **O Agrônômico**, Campinas, v. 52, n. 1, p. 21-23, 2000. Disponível em:

<<http://www.iac.sp.gov.br/publicacoes/agronomico/pdf/mapa.pdf>>. Acesso em: 01 Mai. 2014.

RUDORFF, B. F. T. et al. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. **Remote Sensing**, p. 1057-1076, 2010. Disponível em: <<http://www.mdpi.com/2072-4292/2/4/1057>>. Acesso em: 03 Mai. 2014.

SANTOS, H. G. Importância e evolução dos levantamentos de solos no Brasil. **Boletim Informativo - Sociedade Brasileira de Ciência do Solo**, v. 32, n. 1, p. 18-20, 2007.

SARMENTO, E. C. **Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos do Vale dos Vinhedos, RS, Brasil**. Porto Alegre: Dissertação (Mestrado) Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, 2010. Disponível em: <<http://hdl.handle.net/10183/25977>>. Acesso em: 11 Dez. 2014.

SENTELHAS, P. C. et al. Balanços Hídricos Climatológicos do Brasil - 500 balanços hídricos de localidades brasileiras. **Esalq**, 1999. Disponível em: <<http://www.lce.esalq.usp.br/bhbrasil/Saopaulo>>. Acesso em: 10 Mar. 2015.

SILVA, C. C. **Mapeamento Digital de Classes de Solo: Aplicação da metodologia na folha de Botucatu (SF-22-Z-B-IV-3) e validação de campo**. Campinas: Dissertação (Mestrado) em Agricultura Tropical e Subtropical - Instituto Agronômico (IAC), 2012. Disponível em: <<http://www.iac.sp.gov.br/areadoinstitutoposgraduacao/dissertacoes/pb12123.10%20Cristiano%20Cassiano%20da%20Silva.pdf>>. Acesso em: 07 Ago. 2014.

SILVA, F. P. et al. Arcabouço geológico e hidrofacies do Sistema Aquífero Guarani, no município de Ribeirão Preto (SP). **Revista Brasileira de Geociências**, v. 38, n. 1, p. 56-67, 2008.

SINELLI, O. et al. **Mapa geológico do nordeste do Estado de São Paulo**. Ribeirão Preto: Convênio CNEC/FFCL USP, 1973.

SINELLI, O. et al. A Hidrogeologia de Ribeirão Preto. **Congresso Brasileiro de Águas Subterrâneas**, Recife, 1980. 319-335. Disponível em: <<http://aguassubterraneas.abas.org/asubterraneas/article/viewFile/23871/15944>>. Acesso em: 04 Jun. 2013.

SIRTOLI, A. E. et al. Atributos topográficos secundários no mapeamento de pedoformas. **Geociências**, São Paulo, n. 1, p. 63-77, 2008.

SOARES, P. C. et al. Geologia do nordeste do Estado de São Paulo. **Congresso Brasileiro de Geologia**, Aracaju, 1, 1973. 209-236.

SOIL SURVEY DIVISION STAFF. **Soil Survey Manual**: revised. Washington: Soil Conservation Service. U.S. Department of Agriculture Handbook 18., 1993. 437 p. Disponível em:

<http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_050993.pdf>.

Acesso em: 22 Out. 2014.

STUM, A. K. **Random Forests Applied as a Soil Spatial Predictive Model in Arid Utah**. [S.l.]: [s.n.], 2010. Disponível em:

<<http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1732&context=etd>>.

Acesso em: 04 Nov. 2014.

TEN CATEN, A. **Aplicação de componentes principais e regressões logísticas múltiplas em sistema de informações geográficas para o mapeamento digital de solos**. Santa Maria, RS: Dissertação (mestrado) – Universidade Federal de Santa Maria - Centro de Ciências Rurais, Programa de Pós-Graduação Ciência do Solo, 2008. Disponível em:

<http://w3.ufsm.br/ppgcs/disserta%E7%F5es%20e%20teses/Alexandre_ten_Caten%20PPGCS%20UFSM.pdf>. Acesso em: 17 Abr. 2015.

TEN CATEN, A. **Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos**. Santa Maria: Tese (Doutorado) - Universidade Federal de Santa Maria - RS, 2011. Disponível em: <<http://w3.ufsm.br/ppgcs/disserta%E7%F5es%20e%20teses/teses/Alexandre%20ten%20Caten%20TESE.pdf>>. Acesso em: 17 Abr. 2015.

TEN CATEN, A. et al. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, Santa Maria, v. 42, n. 11, p. 1989-1997, Novembro 2012. Disponível em: <<http://www.scielo.br/pdf/cr/v42n11/a31912cr2012-0022r1.pdf>>. Acesso em: 20 Mai. 2014.

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

Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782012001100013&lng=en&nrm=iso>. Acesso em: 3 Mar. 2015.

THOMPSON, J. A. et al. Digital Soil Mapping: Interactions with and Applications for Hydropedology. **Hydropedology**, n. 1, p. 665-709, 2012.

VERDADE, F. C. et al. Solos da Bacia de Taubaté (vale do Paraíba). Levantamento de reconhecimento. Séries monotípicas, suas propriedades genético-morfológicas, físicas e químicas. **Bragantia**, Campinas, v. 20, n. 4, p. 43-322, 1961. Disponível em: <<http://www.scielo.br/pdf/brag/v20nunico/04.pdf>>. Acesso em: 24 Jan. 2015.

VILLAR, P. C. **Gestão das Áreas de Recarga do Aquífero Guarani**: o caso do município de Ribeirão Preto, São Paulo. São Paulo: Dissertação (Mestrado em Ciência Ambiental), 2008.

WILSON, J. P.; GALLANT, J. C. **Terrain Analysis**: Principles and Applications. New York: John Wiley & Sons, Inc., 2000.

ZAMBON, M. et al. Effect of Alternative Splitting Rules on Image Processing Using Classification Tree Analysis. **Photogrammetric Engineering & Remote Sensing**, v. 72, n. 1, p. 25-30, Janeiro 2006.

ZHOU, B.; ZHANG, X.; WHANG, R. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. **Journal of Zhejiang University Science**, Zhejiang, v. 5, n. 7, p. 782-795, 2004.

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

ZHU, A. X. et al. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. **Soil Science Society of America Journal**, 2001. Disponível em: <http://www.colorado.edu/geography/leyk/geog_5113/readings/zhu_etal_2001_SSSAJ.pdf>. Acesso em: 24 Out. 2014.