

Métodos para mapeamento digital de solos com utilização de sistema de informação geográfica

Methods for digital soil mapping using a geographical information system

Fabício Fernandes Coelho^I Elvio Giasson^{II}

RESUMO

Mapas pedológicos são fontes de informações primordiais para planejamento e manejo do uso do solo, porém apresentam altos custos de produção. A fim de produzir mapas de solos a partir de mapas existentes, neste trabalho, foram comparados métodos de classificação em estágio único (Regressões Logísticas Múltiplas Multinomiais e Bayes) e em estágios múltiplos (Classification and Regression Trees (CART), J48 e Logistic Model Trees (LMT)) com a utilização de sistemas de informações geográficas e de variáveis geomorfométricas para produção de mapas pedológicos com legenda original e simplificada. A base de dados foi gerenciada em aplicativo computacional ArcGis, em que as variáveis e o mapa original foram relacionados por meio de amostras de treinamento para os algoritmos. Os resultados dos algoritmos obtidos no software Weka foram implementados no ArcGis, para a confecção dos mapas. Foram geradas matrizes de erros para análise de acurácias dos mapas. As variáveis geomorfométricas de declividade, perfil e plano de curvatura, elevação e índice de umidade topográfica são aquelas que melhor explicam a distribuição espacial das classes de solo. Os métodos de classificação em estágio múltiplo apresentaram sensíveis melhoras nas acurácias globais, porém significativas melhoras nos índices Kappa. A utilização de legenda simplificada aumentou significativamente as acurácias do produtor e do usuário.

Palavras-chave: modelo digital de elevação, variáveis geomorfométricas, classificação em estágio único, classificação hierárquica.

ABSTRACT

Soil maps are sources of important information for land planning and management, but are expensive to

produce. This paper proposes testing and comparing single stage classification methods (Multiple Multinomial Logistic Regression and Bayes) and multiple stage classification methods (Classification and Regression Trees (CART), J48 and Logistic Model Trees (LMT)) using geographic information system and terrain parameters for producing soil maps with both original and simplified legend. The database was managed in ArcGis computer application in which the variables and the original map were related through training of the algorithms. The results from statistical software Weka were implemented in ArcGis environment to generate digital soil maps. The terrain parameters that best explained soil distribution were slope, profile and planar curvature, elevation, and topographic wetness index. The multiple stage classification methods showed small improvements in overall accuracies and large improvements in the Kappa index. Simplification of the original legend significantly increased the producer and user accuracies, however produced small improvements in overall accuracies and Kappa index.

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

INTRODUÇÃO

Mapas pedológicos são fontes de informações primordiais para planejamento e manejo de uso do solo. O Brasil possui grande parte de seu território coberto por mapas de solos pouco detalhados com escala de 1:1.000.000, produzidos a partir do ano de 1986 pelo projeto Radambrasil. Embora estejam disponíveis, esses dados têm escala muito pequena,

^IPrograma de Pós-graduação em Sensoriamento Remoto, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

^{II}Departamento de Solos, Faculdade de Agronomia, UFRGS, 90001-970, Porto Alegre, RS, Brasil. E-mail: giasson@ufrgs.br. Autor para correspondência.

não adequada para o gerenciamento municipal de propriedades rurais ou regiões de bacias hidrográficas. Mapas de solos mais detalhados, em escala maior, são escassos.

Como os custos necessários para elaboração de mapas em escalas maiores são altos, métodos de mapeamento supervisionados estão sendo testados e analisados com a finalidade de produzir mapas pedológicos com baixo custo-benefício. Para LAGACHERIE & MCBRATNEY (2007), o mapeamento digital de solos se baseia na geração de sistemas de informações que permitem estabelecer relações matemáticas entre variáveis ambientais e classes de solos e, assim, prever a distribuição espacial das classes de solos. Alguns métodos de mapeamento digital de solos já estão sendo utilizados, como regressões logísticas múltiplas multinomiais (GIASSON et al., 2006; CAMPLING et al., 2002), modelos logísticos com aplicação de componentes principais (TEN CATEN et al., 2009), redes neurais artificiais (SIRTOLI, 2008), classificação hierárquica e *support vector machine* (SVM) (BEHRENS & SCHOLTEN, 2007).

Considerando-se que classificadores hierárquicos têm a capacidade de converter processos de decisão complexos em uma série de decisões simples (BREIMAN et al., 1984), aumentando, em geral, a acurácia dos resultados, o presente trabalho tem por objetivo comparar métodos de classificação em estágio único (CEU), como regressão logística múltipla multinomial (RLMM) e classificador de Bayes, e métodos de classificação hierárquica, como LMT (*Logistic Model Trees*), CART (*Classification and Regression Trees*) e o algoritmo J48 (também conhecido como C4.5), para produção de mapas detalhados de solos a partir de variáveis geomorfométricas de uma área de referência com utilização de legenda original e simplificada.

MATERIAL E MÉTODOS

Integram a área de estudo os municípios de Ijuí, Bozano e parte de Coronel Barros, localizados na unidade geomorfológica do Planalto Riograndense, formação Serra Geral, do Estado do Rio Grande do Sul, Brasil, compreendendo uma área de aproximadamente 1018km²; a região é drenada pelo Rio Ijuí e seus afluentes. O mapeamento dos solos por métodos convencionais na escala 1:50.000 (ABRÃO & AZOLIN, 1970) identificou a ocorrência de sete classes de solo: a) Chernossolo Argilúvico Órtico (MTo) (80,97km² ou 7,96% da área total), b) Chernossolo Ebânico Órtico (MEo) (0,44km² ou 0,04% da área total), c) Latossolo Vermelho Distroférico Típico (LVdf) (622,84km² ou

61,19% da área total), d) Latossolo Vermelho Distrófico Húmico (LVdh) (2,76km² ou 0,27% da área total), e) Latossolo Vermelho Distrófico Típico (LVdt) (0,82km² ou 0,08% da área total), f) Neossolo Flúvico (RY) (111,93km² ou 11% da área total) e g) Neossolo Regolítico Eutrófico (RRe) (198,04km² ou 19,46% da área total).

O modelo digital de elevação (MDE) utilizado como base cartográfica para geração das variáveis geomorfométricas foi o SRTM (*Shuttle Radar Topography Mission*) (RABUS et al., 2003). No aplicativo computacional ArcGis 9.2 (ESRI, 2006), foram geradas sete variáveis geomorfométricas: a) declividade, b) curvatura planar, c) perfil de curvatura, d) curvatura (combinação entre curvatura planar e perfil de curvatura), e) direção do fluxo, f) acúmulo do fluxo e g) índice de umidade topográfica (IUT) (BEVEN & KIRKBY, 1979). A fim de trazê-las para mesma ordem de grandeza, todas variáveis foram padronizadas estatisticamente. Para selecionar as variáveis que possuem maior poder de discriminação entre as classes de solo, foi utilizado o método de seleção sequencial progressiva (*Sequential Forward Selection - SFS*) (JOHN et al., 1994) no *software* WEKA 3.5.8 (WITTEN & FRANK, 2005).

As amostras de treinamento foram geradas de maneira aleatória e corresponderam a 11.000 pontos (um ponto a cada 10,8km²). Para a utilização das RLMM e dos classificadores hierárquicos, as informações amostradas foram tabeladas e analisadas estatisticamente no *software* Weka 3.5.8. Já a classificação de Bayes foi realizada no aplicativo computacional ArcGis 9.2. Os CEU utilizados foram RLMM e Bayes. O modelo de regressão logística permite avaliar a probabilidade de ocorrência de uma variável resposta, considerando os valores de uma série de variáveis independentes que podem ser qualitativas ou quantitativas. Por ser utilizada mais que uma variável independente, o modelo é considerado múltiplo e, pelo fato de prever mais que duas classes, é considerado multinomial (ou politômico); logo, o modelo é denominado de regressão logística múltipla multinomial (RLMM), que é uma generalização do modelo logístico simples. O modelo é definido por J-1 equações, de forma que a ocorrência de uma classe j≠J é calculada em relação à classe referência J. Sua modelagem é dada por:

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

sendo $x_i' \beta_j = \sum_{k=0}^K \beta_{jk} x_{ik}$, com $x_{i0} = 1$ e x_i e β_j são os vetores (coluna) de dimensão (K+1), β_j são os coeficientes calculados pela função logística e x_i são as variáveis

geomorfométricas. As probabilidades das classes (p_1, \dots, p_j) podem ser definidas por:

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

$$p_j = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(x_i' \beta_j)} \quad (2)$$

A classificação de Bayes parte do pressuposto de que as funções de probabilidade dentro de cada classe, denotadas por $p(x/j_i)$, são conhecidas e que há dados sobre a probabilidade *a priori* - $P(j_i)$ - de cada classe. Assim, é possível estabelecer a probabilidade *a posteriori* - $P(j_i/x)$ - de um determinado pixel pertencer à classe j_i com a utilização do teorema de Bayes:

$$P(j_i/x) = \frac{p(x/j_i) \cdot P(j_i)}{p(x)} \quad (3)$$

em que $p(x)$ é a função mistura de densidade. Para estabelecer a probabilidade *a priori*, foram utilizadas as proporções das próprias amostras de treinamento em relação às classes, dado que foram amostras aleatórias e cobriam toda a área de estudo e todas as classes possíveis no mapa 1:50.000.

Os algoritmos de classificação hierárquica foram LMT, CART e J48, utilizados no *software* Weka 3.5.8, e seus resultados foram implementados no aplicativo computacional ArcGis 9.2 para produção dos mapas de solos. Os classificadores hierárquicos CART e J48 são formados por n -ésimas ramificações de uma classificação hierárquica, que representam regras de classificação com base na composição de informações das distribuições isoladas de cada variável.

O algoritmo LMT tem estrutura parecida com a do CART e J48, porém o método LMT possui, em seu nós terminais, ao invés de classes únicas, vetores de variáveis (x_i) e de coeficientes (β_j) de todas as classes. Por meio desses valores, pode ser calculada a função do LMT:

$$F(x) = x_i' \beta_j \quad j = 1, \dots, J \quad (4)$$

sendo $x_i' \beta_j = \sum_{k=0}^K \beta_{jk} x_{ik}$ com $x_{i0} = 1$ e x_i e β_j vetores (coluna)

de dimensão $(K+1)$, similar ao modelo logístico em estágio único. As probabilidades das classes, no LMT, são calculadas por meio da seguinte formulação:

$$p_j = \frac{\exp(F_j(x))}{\exp(F_1(x)) + \dots + \exp(F_J(x))} \quad e$$

$$p_j = \frac{\exp(F_j(x))}{\exp(F_1(x)) + \dots + \exp(F_J(x))} \quad j = 1, \dots, J-1 \quad (5)$$

As probabilidades resultantes são para cada nó terminal; dessa maneira, as probabilidades serão

válidas somente para aqueles pixels que respeitarem a regra de decisão da classificação hierárquica até o nó em questão.

Os mapas produzidos foram comparados com o mapa original a fim de realizar análise dos resultados por meio de matrizes de erros ou matrizes de confusão (CONGALTON, 1991). A matriz de erro expressa a relação entre os pixels classificados e originais, possibilitando a análise de quatro indicadores de acurácia: a) acurácia global, obtida pela divisão do número de pixels corretamente classificados (pc) com o total de pixels (n):

$$AG = \frac{pc}{n} \quad (6)$$

b) acurácia do produtor, obtida pela divisão dos pixels classificados corretamente de cada classe (pc_j) com o total de pixels da classe do mapa original (n_{j0}), indica em que proporção certa unidade de mapeamento foi classificada corretamente:

$$AP = \frac{pc_j}{n_{j0}} \quad (7)$$

c) acurácia do usuário, obtida pela divisão dos pixels classificados corretamente de cada classe (pc_j) com o total de pixels da classe do mapa produzido (n_{jp}), indica a certeza que o usuário tem, ao analisar o mapa, que um ponto específico, em uma determinada unidade de mapeamento, está classificado corretamente:

$$AU = \frac{pc_j}{n_{jp}} \quad (8)$$

d) coeficiente Kappa de Cohen (COHEN, 1960), o qual mede a concordância, sendo excluído o efeito da casualidade, dado por:

$$K = \frac{n \sum_{i=1}^j x_{ii} - \sum_{i=1}^j x_{i\oplus} x_{\oplus i}}{n^2 - \sum_{i=1}^j x_{i\oplus} x_{\oplus i}} \quad (9)$$

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

RESULTADOS E DISCUSSÃO

Do conjunto de variáveis utilizadas, as variáveis que melhor explicaram a relação entre a forma do terreno e a distribuição espacial das classes de solos, na presente área de estudo, foram: a) declividade, b) perfil de curvatura, c) elevação, d) curvatura planar e e) índice de umidade topográfica (IUT).

A tabela 1 mostra os resultados dos mapas produzidos com a utilização de CEU e classificação hierárquica a partir do mapa original de solos. O método RLMM apresentou acurácia global (AG) de 67,01% e índice Kappa de 32,81%, já o método de Bayes apresentou AG de 66,56% e índice Kappa de 35,85%. Observa-se que, em ambos os classificadores, não foram mapeadas as classes MTo, MEo, LVdh e LVdt, que são as classes com menor extensão de ocorrência.

Observa-se acurácia avaliada pelo índice Kappa 3% maior no classificador de Bayes e AG menos que 0,5% inferior em relação à RLMM. Como o índice Kappa considera a classificação ao acaso, o classificador de Bayes teve melhores resultados que RLMM para essa área de estudo.

O método CART apresentou AG de 68,02% e Kappa de 38,45%. O algoritmo J48, que tem a mesma estrutura de resposta do CART, apresentou resultados próximos a este, com 67,31% de AG e 38,60% de índice Kappa. Em ambos classificadores, as classes MTo, MEo, LVdh e LVdt não foram mapeadas. O algoritmo LMT apresentou AG de 68,38% e índice Kappa de 37,66% e foi o único algoritmo LMT entre os cinco métodos testados que mapeou a classe Mto, a qual apresentou 15,73% de AU e apenas 0,19% de AP. As classes MEo, LVdh e LVdt não foram mapeadas por nenhum método de classificação hierárquica.

Verifica-se que o total da área da classe MTo representa somente 7,96% da área total, já as classes MEo, LVdh e LVdt representam, cada uma, menos que 0,5% da área total, com respectivamente 0,04, 0,27 e 0,08%. Contrastando esses dados, a classe LVdf apresenta 61,16%; a classe RRE, 19,49%; e a classe RY, 11% da área total. Resultados de mapeamento digital de solos com a mesma magnitude de valores são apresentados em trabalhos de FIGUEIREDO (2006) e TEN CATEN et al. (2009).

Com a finalidade de verificar se a utilização de uma legenda simplificada aumentaria a acurácia dos mapas produzidos, mesmo com a perda de detalhamento das unidades de mapeamento de solos na escala original de 1:50.000, as classes foram agrupadas em quatro classes mais abrangentes: a) Neossolo Flúvico (RY), b) Neossolo Regolítico Eutrófico (RRe), c) Chernossolo (M, agrupamento de MTo e MT) e d) Latossolo Vermelho (LV, agrupamento de LVdf, LVdh e LVdt). A tabela 2 mostra os resultados dos mapas produzidos pelos CEU e os classificadores hierárquicos a partir do mapa de solos com legenda simplificada.

A figura 1 apresenta os mapas produzidos por meio dos métodos CEU e classificação hierárquica com legenda original e simplificada.

O método RLMM, com utilização de legenda simplificada, apresentou AG de 67,36% e Kappa de 33,14%. A classe RY apresentou 38,03 e 44,82% de AP e AU, respectivamente. A classe RRE apresentou 38,59% de AP e 56,30% de AU, e a classe LV apresentou 90,49% de AP e 71,99% de AU. O método de Bayes resultou em AG de 66,89% e Kappa de 36,17%. A classe RY apresentou 56,43% de AP e 41,22% de AU. A classe RRE apresentou 39,22% de AP e 55,57% de AU, e a classe LV apresentou 86,23% e 74,50% de AP e AU, respectivamente. Mesmo com a generalização das classes, a classe M não foi mapeada por nenhum CEU.

O mapa produzido pelo método CART resultou em AG de 68,37% e Kappa de 38,82%. A classe RY apresentou 52,06% de AP e 48,17% de AU. A classe RRE apresentou 47,90 e 54,38% de AP e AU, respectivamente, e a classe LV apresentou 86,67% de AP e 75,13% de AU. O método J48 resultou em 67,66% de AG e o maior índice Kappa com 38,97%. A classe RY teve 56,49% de AP e 46,29% de AU. A classe RRE apresentou 49,99% e 58,56% de AP e AU, respectivamente, e a classe LV apresentou 84,06% de AP e 75,81% de AU. Em ambos os métodos, não ocorreu o mapeamento da classe M. O método LMT teve a maior AG, com 68,72%, e Kappa de 38,02%. A classe RY teve 50,01 e 48,77% de AP e AU. A classe RRE apresentou 44,51% de AP e 58,56% de AU, e a classe LV teve 88,66% de AP e 73,88% de AU. O método LMT foi o único que mapeou a classe M e resultou em 0,19% de AP e 15,56% de AU.

CONCLUSÕES

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

Os métodos de classificação hierárquica apresentaram sensíveis melhoras nas AG e significativas melhoras nos índices Kappa, na AU e na AP. Entre os classificadores hierárquicos, o método J48 foi o que apresentou os maiores índices Kappa, porém as menores AG, ao passo que o método LMT apresentou as melhores AG e os menores índices Kappa, já o método CART apresentou resultados intermediários de AG e Kappa. Entre todos os CEU, o método RLMM foi aquele que teve os menores índices Kappa, porém obteve melhores AG. O método de Bayes, que é considerado um classificador ótimo, apresentou índice Kappa com valores próximos aos classificadores hierárquicos, porém apresentou as menores AG.

Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica.

Tabela 1 - Matriz de erro para avaliação dos mapas de solos com legenda original produzidos com a utilização de classificação hierárquica e classificação em estágio único (CEU).

		-----Mapa original-----									
	Classe	MTo	MEo	LVdf	LVdh	LVdt	RY	RRE	Total	AU (%)	
Mapa produzido por RLMM	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdf	6,03	0,04	55,31	0,27	0,08	5,79	9,80	77,31	71,54	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	AG = 67,01%	RY	0,63	0,00	2,35	0,00	0,00	4,18	2,17	9,33	44,83
	Kappa=32,81%	RRE	1,31	0,01	3,50	0,00	0,00	1,02	7,52	13,36	56,30
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
	AP (%)	*	*	90,45	*	*	38,03	38,59			
Mapa produzido por Bayes	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdf	5,67	0,03	52,71	0,25	0,07	3,96	8,49	71,19	74,04	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	AG = 66,56%	RY	1,15	0,01	4,34	0,01	0,00	6,21	3,35	15,06	41,22
	Kappa=35,85%	RRE	1,14	0,01	4,11	0,01	0,00	0,83	7,64	13,75	55,58
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
	AP (%)	*	*	86,19	*	*	56,43	39,22			
Mapa produzido por CART	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdf	5,55	0,03	52,96	0,27	0,08	3,66	8,41	70,95	74,64	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	AG = 68,02%	RY	0,82	0,00	3,59	0,00	0,00	5,73	1,74	11,89	48,18
	Kappa=38,45%	RRE	1,60	0,01	4,60	0,00	0,00	1,62	9,33	17,16	54,38
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
	AP (%)	*	*	86,60	*	*	52,06	47,90			
Mapa produzido por J48	MTo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdf	5,32	0,02	51,36	0,27	0,08	3,32	7,84	68,20	75,30	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	AG = 67,31%	RY	1,00	0,01	4,29	0,00	0,00	6,21	1,90	13,42	46,30
	Kappa=38,6%	RRE	1,64	0,02	5,50	0,00	0,00	1,47	9,74	18,38	53,01
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
	AP (%)	*	*	83,98	*	*	56,49	49,99			
Mapa produzido por LMT	MTo	0,01	0,00	0,04	0,00	0,00	0,00	0,04	0,10	15,73	
	MEo	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdf	5,88	0,02	54,19	0,26	0,08	4,33	9,04	73,82	73,42	
	LVdh	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	LVdt	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	*	
	AG = 68,38%	RY	0,78	0,00	3,25	0,00	0,00	5,50	1,74	11,28	48,79
	Kappa=37,66%	RRE	1,28	0,01	3,67	0,00	0,00	1,16	8,67	14,81	58,56
		Total	7,96	0,04	61,16	0,27	0,08	11,00	19,49	100,00	
	AP (%)	0,19	*	88,61	*	*	50,01	44,51			

* classe que não ocorreu no mapa produzido.

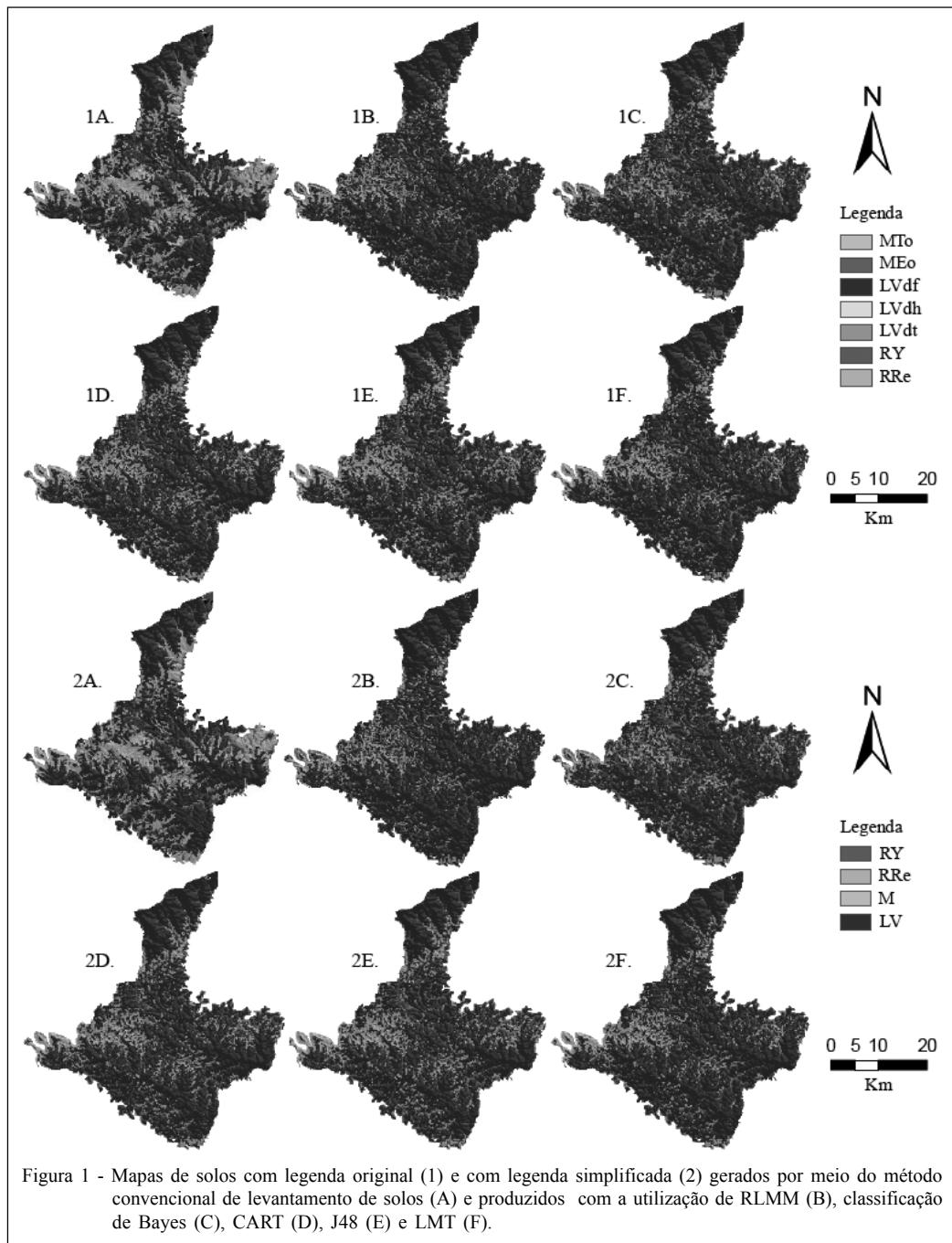
Tabela 2 - Matriz de erro para avaliação dos mapas de solos com legenda simplificada produzidos com a utilização de classificação hierárquica e classificação em estágio único (CEU).

-----Mapa original com legenda simplificada-----							
	Classe	RY	RRE	M	LV	Total	AU (%)
Mapa produzido por RLMM	RY	4,18	2,17	0,63	2,35	9,33	44,82
	RRE	1,02	7,52	1,31	3,50	13,36	56,30
	M	0,00	0,00	0,00	0,00	0,00	*
	LV	5,79	9,80	6,07	55,65	77,31	71,99
	Total	11,00	19,49	8,01	61,50	100,00	
AG = 67,36%							
Kappa=33,14%	AP (%)	38,03	38,59	*	90,49		
	Classe	RY	RRE	M	LV	Total	AU (%)
Mapa produzido por Bayes	RY	6,21	3,35	1,16	4,34	15,06	41,22
	RRE	0,83	7,64	1,16	4,13	13,75	55,57
	M	0,00	0,00	0,00	0,00	0,00	*
	LV	3,96	8,49	5,69	53,04	71,19	74,50
	Total	11,00	19,49	8,01	61,50	100,00	
AG = 66,89							
Kappa=36,17%	AP (%)	56,43	39,22	*	86,23		
	Classe	RY	RRE	M	LV	Total	AU (%)
Mapa produzido por CART	RY	5,73	1,74	0,83	3,59	11,89	48,17
	RRE	1,62	9,33	1,61	4,60	17,16	54,38
	M	0,00	0,00	0,00	0,00	0,00	*
	LV	3,66	8,41	5,57	53,31	70,95	75,13
	Total	11,00	19,49	8,01	61,50	100,00	
AG = 68,37%							
Kappa=38,82%	AP (%)	52,06	47,90	*	86,67		
	Classe	RY	RRE	M	LV	Total	AU (%)
Mapa produzido por J48	RY	6,21	1,90	1,01	4,30	13,42	46,29
	RRE	1,47	9,74	1,66	5,50	18,38	53,01
	M	0,00	0,00	0,00	0,00	0,00	*
	LV	3,32	7,84	5,34	51,70	68,20	75,81
	Total	11,00	19,49	8,01	61,50	100,00	
AG = 67,66%							
Kappa=38,97%	AP(%)	56,49	49,99	*	84,06		
	Classe	RY	RRE	M	LV	Total	AU (%)
Mapa produzido por LMT	RY	5,50	1,74	0,79	3,25	11,28	48,77
	RRE	1,16	8,67	1,30	3,68	14,81	58,56
	M	0,00	0,04	0,01	0,04	0,10	15,56
	LV	4,33	9,04	5,91	54,53	73,82	73,88
	Total	11,00	19,49	8,01	61,50	100,00	
AG = 68,72%							
Kappa=38,02%	AP(%)	50,01	44,51	0,19	88,66		

* classe que não ocorreu no mapa produzido.

A utilização de legenda simplificada não provocou grandes aumentos de AG e Kappa, porém proporcionou melhora na AU e na AP. Nos mapeamentos com legenda original, não se observou valor superior a 28,10% para AU e AP, ao passo que, com a utilização de legenda simplificada, o valor não baixou de 41,70% e chegou até aos 49,19%.

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



REFERÊNCIAS

ABRÃO, P.U.R.; AZOLIN, M.A.D. **Levantamento e utilização agrícola dos solos no Município de Ijuí**. Porto Alegre. INCRA. 1970. 102p. (Relatório Técnico).

BEHRENS, T.; SCHOLTEN, T. A comparison of data-mining techniques in predictive soil mapping. In: LAGACHERIE, P. et al. (Eds.). **Digital soil mapping, an introductory perspective**. Developments in soil science. Amsterdam: Elsevier, 2007. V.31, p.353-364.

BEVEN, K.; KIRKBY, N. A physically based variable contributing area model of basin hydrology. *Hydrological sciences. Bulletin des Sciences Hydrologiques*, v.24, p.43-69, 1979. Disponível em: <<http://dx.doi.org/10.1080/02626667909491834>>. Acesso em: 03 ago. 2010. doi: 10.1080/02626667909491834.

BREIMAN, L. et al. **Classification and regression trees (CART)**. Belmont, CA: Wadsworth International, 1984. 358p.

CAMPLING, P. et al. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Science Society*

of America Journal, v.66, p.1390-1401, 2002. Disponível em: <<https://www.agronomy.org/publications/sssaj/pdfs/66/4/1390>>. Acesso em: 03 ago. 2010. doi: 10.2136/sssaj2002.1390.

COHEN, J. A coefficient of agreement for nominal scales. **Journal of Educational and Measurement**, v.20, n.1, p.37-46, 1960. Disponível em: <<http://dx.doi.org/10.1177/001316446002000104>>. Acesso em: 03 ago. 2010. doi: 10.1177/001316446002000104.

CONGALTON, R.G. A review of assessing the accuracy of classification of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991. Disponível em: <[http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B)>. Acesso em: 03 ago. 2010. doi: 10.1016/0034-4257(91)90048-B.

ESRI. **ArcGis 9.2**. Redland, California. Software. 2006.

FIGUEIREDO, S.R. **Mapeamento supervisionado de solos através do uso de regressões** logísticas múltiplas e sistema de informações geográficas. 2006. 92f. Dissertação (Mestrado em Agronomia) – Programa de Pós-graduação em Ciência do Solo, Universidade Federal do Rio Grande do Sul, RS.

GIASSON, E. et al. Digital soil mapping using multiple logistic regressions on terrain parameters in Southern Brazil. **Scientia Agricola**, v.63, n.3, p.262-268, 2006. Disponível em: <<http://dx.doi.org/10.1590/S0103-90162006000300008>>. Acesso em: 03 ago. 2010. doi: 10.1590/S0103-90162006000300008.

JOHN, G.H. et al. Irrelevant features and the subset selection problem. In COHEN W.W.; HIRSH, H. **Machine learning:**

proceedings of the eleventh international conference. San Francisco: Morgan Kaufmann, 1994. p.121-129.

LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P. et al. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007. p.3-24.

RABUS, B. et al. The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar. **Journal of Photogrammetry & Remote Sensing**, v.57, p.241-262, 2003. Disponível em: <[http://dx.doi.org/10.1016/S0924-2716\(02\)00124-7](http://dx.doi.org/10.1016/S0924-2716(02)00124-7)>. Acesso em: 03 ago. 2010. doi:10.1016/S0924-2716(02)00124-7.

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

TEN CATEN, A. et al. Mapeamento digital de solos através da aplicação de componentes principais em modelos logísticos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2009, Natal, RN. **Anais...** São José dos Campos: INPE, 2009. p.7677-7684.

WITTEN, A.H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2.ed. San Francisco: Morgan Kaufmann, 2005. 560p.