

ALEXSON DE MELLO CUNHA

**SELEÇÃO DE VARIÁVEIS AMBIENTAIS E DE ALGORITMOS DE
CLASSIFICAÇÃO PARA MAPEAMENTO DIGITAL DE SOLOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2013

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

Cunha, Alexson de Mello, 1973-
C972s Seleção de variáveis ambientais e de algoritmos de
2013 classificação para mapeamento digital de solos / Alexson de
Mello Cunha. – Viçosa, MG, 2013.
x, 121f : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: João Luiz Lani.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Mapeamento do solo. 2. Redes neurais (Computação).
3. Solos - Classificação. 4. Mapeamento digital. 5. Estatística.
I. Universidade Federal de Viçosa. Departamento de Solos.
Programa de Pós-Graduação em Solos e Nutrição de Plantas.
II. Título.

CDD 22. ed. 631.47

ALEXSON DE MELLO CUNHA

**SELEÇÃO DE VARIÁVEIS AMBIENTAIS E DE ALGORITMOS DE
CLASSIFICAÇÃO PARA MAPEAMENTO DIGITAL DE SOLOS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de Doctor *Scientiae*.

APROVADA: 16 de agosto de 2013.



Diego Lang Burak

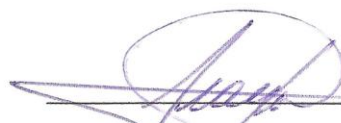


Vicente Paulo Soares

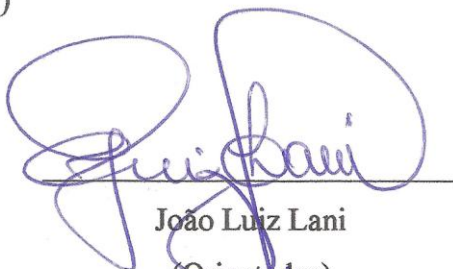


Elpídio Inácio Fernandes Filho

(Co-orientador)



Márcio Rocha Francelino



João Luiz Lani

(Orientador)

AGRADECIMENTOS

À Deus.

À minha amada esposa Ingrid, pelo carinho, dedicação, paciência e incentivo.

À Universidade Federal de Viçosa, em particular ao Departamento de Solos.

Ao Instituto Nacional de Colonização e Reforma Agrária – INCRA, por conceder o afastamento e poder me dedicar em tempo integral ao curso de doutorado.

Aos meus pais, pela vida, educação e incentivo ao estudo.

Aos meus irmãos, Gláucio, Rogério, Andressa e Gláucia, cunhados, cunhadas e sobrinhos pelo apoio e torcida.

Ao Professor João Luiz Lani, pela dedicação na orientação, amizade e por acreditar em meu potencial.

Ao professor Elpídio Inácio Fernandes Filho, pelo grande empenho na co-orientação e pela amizade.

Aos professores Gerson Rodrigues dos Santos, Márcio Francelino e Vicente Soares pelas sugestões.

Ao professor Diego Lang Burak, da Universidade Federal do Espírito Santo, pelo apoio nos trabalhos de campo e pela amizade.

Ao professor João Ker, pelas importantes sugestões e materiais que contribuíram para a realização deste trabalho.

Ao Instituto Capixaba de Pesquisa e Extensão Rural – Incaper, pelo apoio logístico nos trabalhos de campo.

Ao professor José Ambrósio, pela valiosa co-orientação no projeto de pesquisa.

A todos os amigos do NEPUT, pelo apoio.

A todos os professores do Departamento de solos, pelo conhecimento transmitido nas disciplinas e seminários e a todos os amigos do labgeo, em especial Eliana Souza, Filipe Trindade, Flávia Xavier e Renato Taques, os quais contribuíram de forma direta e significativa para que pudesse concluir o trabalho de Tese.

Aos amigos Rafaela Schaedler, Caio Lins, Thalita Avelar Monteiroe Samuel Valadares.

A todos os servidores do Departamento de Solos, em especial à Luciana e Claudia.

BIOGRAFIA

ALEXSON DE MELLO CUNHA, filho de Vicente Bodevan da Cunha e Ilzeny Machado de Mello Cunha, nasceu em Guaçuí, ES, em 09 de dezembro de 1973.

Em 1989, iniciou o curso técnico em agropecuária na Escola Agrotécnica Federal de Alegre, formando-se em 1991.

Em 1995, iniciou o curso de Agronomia no Centro de Ciências Agrárias da Universidade Federal do Espírito Santo, concluindo-o em 2001. No mesmo ano, iniciou o curso de Mestrado em Solos e Nutrição de Plantas, no Departamento de Solos da Universidade Federal de Viçosa.

Em 2005, por meio de aprovação em concurso público, entra em exercício no Instituto Nacional de Colonização e Reforma Agrária – INCRA, onde atua no cargo de Engenheiro Agrônomo, Carreira de Perito Federal Agrário.

Em 2010, iniciou o curso de Doutorado em Solos e Nutrição de Plantas pelo Departamento de Solos - DPS da Universidade Federal de Viçosa - Minas Gerais, submetendo-se à defesa de Tese em agosto de 2013.

ÍNDICE

1. INTRODUÇÃO GERAL	1
1.1 REFERÊNCIAS	2
2. CAPÍTULO 1 - ESPACIALIZAÇÃO DA PRECIPITAÇÃO PLUVIOMÉTRICA NO ESTADO DO ESPÍRITO SANTO POR MEIO DO USO DE KRIGAGEM E COKRIGAGEM	4
2.1. RESUMO.....	4
2.2. ABSTRACT	4
2.3. INTRODUÇÃO.....	5
2.4. MATERIAL E MÉTODOS.....	7
2.5. RESULTADOS E DISCUSSÃO.....	11
2.6. CONCLUSÕES	23
3. CAPÍTULO 2 - SELEÇÃO DE PREDITORES AMBIENTAIS NO MAPEAMENTO DIGITAL DE SOLOS DA BACIA DO RIO MUQUI DO NORTE, ESTADO DO ESPÍRITO SANTO, BRASIL	27
3.1. RESUMO.....	27
3.2. ABSTRACT	28
3.3. INTRODUÇÃO.....	28
3.4. MATERIAL E MÉTODOS.....	30
3.4.1. Caracterização da área.....	30
3.4.2. Variáveis discriminantes	32
3.4.3. Seleção das variáveis discriminantes	36
3.4.4. Trabalhos de campo	38
3.5. RESULTADOS E DISCUSSÃO.....	38
3.5.1. Seleção de atributos.....	38
3.6. CONCLUSÕES	49
3.7. REFERÊNCIAS BIBLIOGRÁFICAS	50
4. CAPÍTULO 3 - PREDIÇÃO DE CLASSES DE SOLO POR DIFERENTES ALGORITMOS E VARIÁVEIS PREDITIVAS AMBIENTAIS ..	55
4.1. RESUMO.....	55
4.2. ABSTRACT	55
4.3. INTRODUÇÃO.....	56
4.4. MATERIAL E MÉTODOS.....	57

4.5. RESULTADOS E DISCUSSÃO.....	65
4.6. CONCLUSÕES	76
4.7. REFERÊNCIAS BIBLIOGRÁFICAS	76
5. CAPÍTULO 4 - ATRIBUTOS DO TERRENO, CLIMÁTICOS E BALANÇO HÍDRICO NO MAPEAMENTO DIGITAL DE SOLOS DO ASSENTAMENTO RURAL SEZÍNIO FERNANDES, LINHARES, ESPÍRITO SANTO.....	80
5.1. RESUMO.....	80
5.2. ABSTRACT	81
5.3. INTRODUÇÃO.....	81
5.4. MATERIAL E MÉTODOS.....	82
5.5. RESULTADOS E DISCUSSÃO.....	88
5.6. CONCLUSÕES	97
5.7. REFERÊNCIAS BIBLIOGRÁFICAS	98
6. CONCLUSÕES GERAIS	102
7. APÊNDICES	103

RESUMO

CUNHA, Alexson de Mello, D.Sc., Universidade Federal de Viçosa, julho de 2013. **Seleção de variáveis ambientais e de algoritmos de classificação paramapeamento digital de solos.** Orientador: João Luiz Lani. Co-orientadores: Elpídio Inácio Fernandes Filho e José Ambrósio Ferreira Neto.

No mapeamento digital de solos têm sido utilizados os atributos do terreno como as principais variáveis preditivas ambientais. Outras variáveis, relacionadas aos processos pedogenéticos, como as climáticas, não tem sido geralmente utilizadas. Nesse contexto, os objetivos deste estudo foram: (1) comparar o método da cokrigagem ordinária em com o da krigagem na espacialização da precipitação pluvial no Estado do Espírito Santo; (2) identificar no mapeamento de solos, em escalas diferentes, quais as variáveis mais relevantes para um melhor desempenho de predição das classes de solo, tendo como estudo duas áreas distintas; (3) avaliar o desempenho dos algoritmos classificadores NaiveBayes, rede neural multilayerperceptron - MLP, SimpleCart e J48. Na avaliação da krigagem e cokrigagem utilizaram-se dados de 108 postos pluviométricos e variáveis secundárias como altitude e distância do mar. Para a avaliação das variáveis e algoritmos classificadores no mapeamento de média escala (1:100.000) o estudo foi realizado na bacia do rio Muqui do Norte, sul do Estado do Espírito Santo. Utilizaram-se 598 amostras de treinamento (10 classes de solos) e 45 variáveis relacionadas a diferentes fatores de formação dos solos, tais como: atributos do terreno, geologia, geomorfologia, clima, balanço hídrico e índices derivados das bandas 1, 3, 4 e 5 do sensor TM Landsat 5. Essas variáveis foram submetidas a diferentes métodos de seleção de atributos baseadas em correlação – CFS; em consistência – CSE; ganho de informação - IA e “ReliefF”, disponíveis no software Weka 3.6.8. Nesse software foi aplicado o classificador SimpleCart para avaliar a efetividade da predição com os subconjuntos de variáveis selecionadas. A avaliação dos quatro classificadores foi realizada com o uso das 45 variáveis e as selecionadas pelo algoritmo “ReliefF”. Para o mapeamento detalhado realizado no Assentamento Rural Sezínio Fernandes, em Linhares, ES, utilizaram-se 259 amostras de treinamento (3 classes de solos) e 19 variáveis preditivas (atributos do terreno, climáticas e balanço hídrico) nas predições feitas pelos classificadores SimpleCart, J48, MLP e NaiveBayes. As predições foram avaliadas com base na validação cruzada e comparações dos mapas elaborados com o mapa convencional de referência. Os resultados da interpolação sugerem que se deve preferir a cokrigagem e o uso de grades regulares para amostragem

de variáveis secundárias. Os resultados da seleção de atributos para mapeamento da bacia do rio Muqui do Norte indicaram que os algoritmos “ReliefF” e o CSE, ambos limitados a 10 atributos, foram os que apresentaram árvores menos complexas e sem perda significativa na exatidão da predição em relação ao grupo de 45 variáveis. A exatidão da classificação, indicada pelo Kappa de 0,60, foi considerada muito boa. As variáveis selecionadas pelo “ReliefF” foram geologia, geomorfologia e principalmente atributos do terreno e elementos do balanço hídrico, como excedente hídrico, deficiência hídrica e evapotranspiração potencial. Os algoritmos NaiveBayes, MLP e SimpleCart apresentaram desempenhos de predição semelhantes (Kappa 0,60 a 0,66), superiores ao J48. A maior concordância com o mapa de referência obtida pelo algoritmo MLP, seguido do SimpleCart, J48 e NaiveBayes foi de 55, 52, 51 e 48%, respectivamente. Nas predições de classes de solos do Assentamento Sezinio Fernandes as variáveis excedente hídrico, deficiência hídrica e temperatura do ar foram relevantes. No entanto, as pequenas amplitudes de valores apresentadas pelas variáveis climáticas e balanço hídrico não são provavelmente suficientes para propiciar condições pedogenéticas diferenciadas na área de estudo. Não houve diferenças significativas entre os valores de Kappa (0,77 a 0,82) dos três algoritmos de classificação. As maiores concordâncias com o mapa convencional foram obtidas para o algoritmo J48, seguido do NaiveBayes e do SimpleCart, utilizando somente atributos de terreno como variáveis preditivas. As árvores de decisão por produzirem resultados de mais fácil entendimento e apresentarem em geral exatidões semelhantes aos classificadores NaiveBayes e rede neural MLP, podem ser consideradas como de grande potencial para se consolidarem no mapeamento digital de solos.

ABSTRACT

CUNHA, Alexson de Mello, D.Sc., Universidade Federal de Viçosa, July, 2013. **Selection of environmental variables and classification algorithms for digital soil mapping.** Advisers: João Luiz Lani. Co-adviser: Elpídio Inácio Fernandes Filho and José Ambrósio Ferreira Neto.

In digital soil mapping the terrain attributes have been used as the main environmental predictor variables. Other variables related to pedogenic processes, such as climate changes, has not been used often. In this context, the objectives of this study were: (1) to evaluate the method of cokriging ordinary kriging compared to the spatial distribution of rainfall in the State of Espírito Santo, (2) identify soil mapping at different scales, which variables most relevant for better performance prediction of soil classes, whose study two distinct areas, (3) evaluate the performance of the NaiveBayes classifier algorithm, neural network multilayerperceptron - MLP SimpleCart and J48. In the evaluation of kriging and cokriging were used data from 108 rain gauges and secondary variables such as altitude and distance from the sea. For the evaluation of the variables and the mapping algorithms classifiers medium scale (1:100,000) the study was conducted in the river basin Muqui North, south of Espírito Santo. We used 598 training instances (10 soil classes) and 45 variáveis related to different factors of soil formation, such as terrain attributes, geology, geomorphology, climate, water balance and indices derived from bands 1, 3, and 4 5 Landsat 5 TM. These variables were subjected to different methods of feature selection based on correlation - CFS, in consistency - CSE, information gain - and IA "ReliefF", available in the software Weka 3.6.8. This software was applied SimpleCart classifier to evaluate the effectiveness of the prediction with the subsets of selected variables. The evaluation of the four binders was performed using the 45 variables and the algorithm selected by "ReliefF." For detailed mapping conducted in Rural Settlement Sezínio Fernandes, Linhares, ES, we used 259 training instances (three soil classes) and 19 predictor variables (terrain attributes, climate and water balance) in the predictions made by the classifiers SimpleCart, J48, MLP and NaiveBayes. The predictions were evaluated based on cross-validation and comparisons of maps made with the conventional map reference. The interpolation results suggest that cokriging to be preferred to the use of regular grids for sampling secondary variables. The results of the selection of attributes for mapping river basin Muqui North indicated that the algorithms "ReliefF" and CSE, both limited

to 10 attributes were those with less complex trees and without significant loss in accuracy prediction compared to group 45 variables. The classification accuracy, indicated by Kappa of 0.60 was considered very good. The variables selected by "ReliefF" were geology, geomorphology and especially terrain attributes and elements of the water balance, as water surplus, water deficit and potential evapotranspiration. Algorithms NaiveBayes, MLP and SimpleCart showed similar performance prediction (Kappa 0.60 to 0.66), higher than the J48. The highest agreement with the reference map obtained by the MLP algorithm, followed by SimpleCart, J48 and NaiveBayes was 55, 52, 51 and 48%, respectively. The predictions of soil from Settlement Sezinio Fernandes variables water surplus, water deficit and air temperature were relevant. However, the small amplitude values presented by climatic variables and water balance are probably not sufficient to provide different pedogenetic conditions in the study area. There were no significant differences between the Kappa values (0.77 to 0.82) of the three classification algorithms. The greatest agreement with the conventional map was obtained for the algorithm J48, followed by NaiveBayes and SimpleCart, using only the terrain attributes as predictor variables. Decision trees for producing results more easily understood and presented in general accuracies similar to NaiveBayes classifiers and neural network MLP, may be regarded as of great potential to consolidate the digital soil mapping.

1. INTRODUÇÃO GERAL

O levantamento de solos é fundamental em estudos que visam o planejamento de uso da terra. São considerados elementos básicos para definição da aptidão das terras com vistas ao melhor uso, em especial, nas atividades agropecuárias e florestais.

A diversidade de solos, relevo, regime pluvial e uso da terra resulta em grande diversidade de ambientes, que é o caso do Espírito Santo (Lani et al., 2001). Isto implica a necessidade de informações mais detalhadas de solos para uma escolha mais cuidadosa das terras a serem utilizadas, seja pela agricultura familiar ou pelo agronegócio. Idealmente, a tomada de decisão quanto à destinação das terras e ao seu manejo racional deve ser baseada nas características intrínsecas dos solos, clima e relevo. Muitas vezes são necessárias práticas de uso e manejo diferenciadas para uma produção sustentada. Nesse contexto, não basta o conhecimento das classes de solos ou suas características em determinados locais de uma área, elevando-se de importância a espacialização das mesmas, a qual é obtida por meio do levantamento e mapeamento de solos. Mapas de solos atualizados, contemplando não só escala maiores, como nomenclaturas das unidades de mapeamento atualizadas, permitem conseqüentemente a elaboração de cartas de aptidão agrícola que possibilitam o melhor planejamento de uso do solo, seja em nível local ou regional.

Novos métodos de mapeamento de solos que fazem uso de geotecnologias referentes a sensoriamento remoto e sistemas de informação geográfica, têm permitido a predição das classes de solos (McBratney et al., 2003) e a elaboração de mapas digitais de solos. O mapeamento digital de solos tem sido uma ferramenta que caminha para sua consolidação como técnica de auxílio ao mapeamento convencional de solos (Ten Caten, 2011). Estas técnicas integram o método convencional aos métodos quantitativos, de modo a incorporar o conhecimento qualitativo de pedólogos na predição de propriedades ou classes de solos (Mackenzie & Ryan, 1999).

O uso de geotecnologias, modelos digitais de elevação (MDE) e imagens de sensoriamento remoto orbital, tem desempenhado importante papel no mapeamento e vem crescendo de forma significativa em levantamentos de solos (Nolasco-Carvalho et al., 2009). O MDE auxilia de maneira eficaz no mapeamento de solos, pois são baseados em fatores pedogenéticos e na relação solo-paisagem (Souza Junior & Demattê, 2008). Os padrões de relevo são relevantes, principalmente se usados no prognóstico de classes de solos em pequenas áreas (Mendonça-Santos e Santos, 2003).

Apesar do relevo (r) ser uma importante variável ambiental correlacionada com os tipos de solos, a utilização de outros fatores de formação do solo como, material de origem (p), clima (c), organismos (o) e tempo (t) são desejáveis na modelagem. No entanto, grande parte destes estudos incorpora um único ou dois fatores (McBratney et al., 2003). Neste aspecto, os poucos trabalhos que utilizam variáveis climáticas estão voltados para o mapeamento de atributos dos solos (Böhner & Selige, 2006). Não há relato na literatura sobre o desempenho dessas variáveis na predição de classes de solo.

Regressões logísticas múltiplas multinomiais (Giasson et al., 2006), modelos logísticos com aplicação de componentes principais (Ten Caten et al., 2011), redes neurais artificiais (Chagas et al., 2010), árvores de decisão (Moran & Bui, 2002), entre outros, são alguns métodos utilizados em estudos sobre mapeamento digital de solos. No entanto, trabalhos que mostram comparações de desempenho entre diferentes algoritmos classificadores têm sido ainda escassos.

Dessa forma, este trabalho teve como objetivos avaliar dentre uma série de variáveis ambientais preditivas, relacionadas a diferentes fatores de formação dos solos, quais as que propiciavam melhor desempenho de predição no mapeamento digital de classes de solos. Avaliar diferentes algoritmos classificadores e variáveis ambientais no mapeamento digital em diferentes escalas em duas áreas distintas. Para isso, o trabalho foi estruturado em quatro capítulos. No primeiro capítulo, o objetivo foi avaliar a krigagem e a cokrigagem como métodos de interpolação de dados de precipitação pluvial com vistas a produzir mapas de precipitação, uma vez que esta foi uma das variáveis ambientais preditivas utilizadas posteriormente no mapeamento digital de classes de solos. No segundo capítulo, objetivou-se avaliar diferentes algoritmos de seleção de atributos a fim de se obter os melhores subconjuntos de variáveis ambientais preditivas no mapeamento digital com a utilização de árvore de decisão como algoritmo classificador. O terceiro capítulo teve como objetivo avaliar diferentes algoritmos classificadores e variáveis preditivas em duas escalas de mapeamento. No quarto capítulo, avaliaram-se diferentes algoritmos classificadores e variáveis preditivas, porém em uma área distinta da primeira e em escala mais detalhada.

1.1 REFERÊNCIAS

BÖHNER, J. & SELIGE, T. Spatial Prediction of Soil Attributes Using Terrain Analysis and Climate Regionalization. In: BÖHNER, J.; MCCLOY, K.R. & STROBL, J. SAGA - **Analysis and Modeling Applications**. Göttinger

- Geographische Abhandlungen, 115:13-27, 2006. Disponível em: http://ufpr.dl.sourceforge.net/project/saga-gis/SAGA%20-%20Documentation/GGA115/gga115_02.pdf. Acesso em 18/02/2011.
- CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.R. & CARVALHO JUNIOR, W. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. **Pesquisa Agropecuária Brasileira**, 45:497-507, 2010.
- GIASSON, E.; CLARKE, R. T.; JUNIOR, A. V. I.; MERTEN, G. H. & TORNQUIST, C.G. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, 63:262-268, 2006.
- LANI, J.L.; REZENDE, S.B. & RESENDE, M. Estratificação de ambientes com base nas classes de solos e outros atributos na bacia do rio Itapemirim, Espírito Santo. **Revista Ceres**, 48:239-261, 2001.
- McBRATNEY, A. B., MENDONÇA-SANTOS, M. L., MINASNY, B. On digital soil mapping. **Geoderma**, 117:3-52, 2003.
- McKENZIE, N.J. & RYAN, P.J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, 89:67-94, 1999.
- MENDONÇA-SANTOS, M.L. & SANTOS, H.G. **Mapeamento digital de classes e atributos de solos - métodos, paradigmas e novas técnicas**. Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA, Rio de Janeiro, Documentos 55, 19p. 2003.
- MORAN, C. J. & BUI, E. Spatial data mining for enhanced soil map modeling. **International Journal of Geographical Information Science**, London, 16:533-549, 2002.
- NOLASCO-CARVALHO, C.C; FRANCA-ROCHA, W. & UCHA, J.M. Mapa digital de solos: uma proposta metodológica usando inferência fuzzy. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 13:46-55, 2009.
- SOUZA JUNIOR, J.G.A. & DEMATTÊ, J.A.M. Modelo digital de elevação na caracterização de solos desenvolvidos de basalto e material arenítico. **Revista Brasileira de Ciência do Solo**, 32:449-456, 2008.
- TEN CATEN, A.; DALMOLIN, R.S.D.; PEDRON, F.A.; MENDONÇA-SANTOS, M.L. Componentes principais como preditores no mapeamento digital de classes de solos. **Ciência Rural**, 41:1170-1176, 2011.

2. CAPÍTULO 1- ESPACIALIZAÇÃO DA PRECIPITAÇÃO PLUVIOMÉTRICA NO ESTADO DO ESPÍRITO SANTO POR MEIO DO USO DE KRIGAGEM E COKRIGAGEM

Submetido à Revista Pesquisa Agropecuária Brasileira

2.1. RESUMO

O objetivo deste trabalho foi avaliar a cokrigagem ordinária com dados de altitude e distância do mar em comparação à krigagem ordinária na espacialização da precipitação pluvial dos períodos anual, seco e úmido no Estado do Espírito Santo. Os dados de altitude e distância do mar foram obtidos em pontos de amostragens em grades regulares e irregulares. Foram utilizados dados de 108 postos pluviométricos distribuídos na área de estudo. A avaliação dos métodos e das variáveis foi realizada com base na validação cruzada, considerando os erros dos valores preditos, o ajuste dos modelos de regressão linear para valores observados e preditos. Em geral a grade regular para amostragem das covariáveis apresentou melhor acurácia de predição em relação à grade irregular. A cokrigagem produziu resultados mais acurados que a krigagem, verificados por pequenas diferenças nos erros médios absolutos, mas capaz de produzir mapas estatisticamente diferentes. Dessa maneira, a interpolação por cokrigagem e uso de grades regulares para amostragem são preferíveis, principalmente se as covariáveis são de fácil obtenção e de baixo custo.

Termos para indexação: linha de costa, geoestatística, predição espacial.

RAINFALL SPATIALIZATION IN THE STATE OF ESPÍRITO SANTO BY USING KRIGING AND COKRIGING

2.2. ABSTRACT

The objective of this work was to evaluate the ordinary co-kriging algorithm with utilization of elevation and distance from the sea data when comparing to ordinary kriging in the rainfall spatialization associated to wet, dry and annual periods for the State of Espírito Santo, Brazil. Altitude data and distance from the sea were obtained in sampling sites on regular and irregular grids. Data from 108 rain gauge stations

distributed in the study area were used. The evaluation of methods and variables was based on cross-validation, considering the errors of the predicted values, the fit of linear regression models to observed and predicted values. In general, the regular grid sampling of the covariates showed better prediction accuracy compared to the irregular grid. The cokriging produced more accurate results than kriging, checked by small differences in mean absolute errors, but able to produce statistically different maps. Thus, the cokriging interpolation and use of regular grids for sampling are preferable, especially if the covariates are easy to obtain and inexpensive.

Index terms: coastline, geostatistic, spatial prediction

2.3. INTRODUÇÃO

A estimativa da variabilidade espacial de chuvas é importante para a determinação do potencial agrícola em uma região e identificação de áreas de risco, onde elevados índices de precipitação pluvial podem levar às inundações (Goovaerts, 2000) e ou movimentos de massa ao longo de encostas (Porto Junior et al., 2012).

Frequentemente, os dados de precipitação pluvial são obtidos em postos de coleta pontuais, o que gera a necessidade de utilização de métodos de interpolação espacial para o seu mapeamento, dentre os quais se destaca a krigagem (Carvalho & Assad, 2005; Viola et al., 2010). A krigagem ordinária, um método geoestatístico univariado, tem sido muito utilizada devido a sua eficiência na interpolação de dados pluviométricos (Carvalho & Assad, 2005; Viola et al., 2010; Carvalho et al., 2012). A extensão multivariada da krigagem, conhecida como cokrigagem, é utilizada quando existe correlação espacial cruzada entre a variável primária e uma variável secundária (Isaaks & Srivastava, 1989). A utilidade da cokrigagem é muitas vezes reforçada pelo fato de que a variável primária de interesse pode ser subamostrada em relação a outras covariáveis correlacionadas espacialmente (Isaaks & Srivastava, 1989). Nesse aspecto, a cokrigagem pode ser mais interessante que a krigagem, devido à baixa densidade de postos pluviométricos existentes nas áreas de interesse, especialmente em áreas montanhosas de difícil instalação e manutenção de equipamentos.

O método da cokrigagem, para fins de espacialização da precipitação pluvial, envolve geralmente variáveis topográficas (Goovaerts, 2000). Tais variáveis têm baixo custo como fonte de informação secundária, se obtidas de modelos de elevação digital

(Goovaerts, 2000). Dentre essas variáveis, a altitude tem sido mais utilizada e estudos têm demonstrado melhorias na acurácia da predição de chuvas com a utilização desta covariável na cokrigagem (Diodato & Ceccarelli, 2005; Moral, 2010; Viola et al., 2010).

Conforme Goovaerts (2000), ao se comparar a cokrigagem com dados de elevação como variável secundária com a krigagem, pode se verificar a superioridade da cokrigagem. Se a correlação entre a precipitação e a altitude aumenta, aumenta-se também o ganho da cokrigagem em relação à krigagem. Uma correlação fraca pode fornecer resultados muito similares entre krigagem e cokrigagem, conforme verificado por Rocha et al. (2012) ao analisar a influência de uma variável secundária como uma função da correlação com a variável primária para a cokrigagem colocalizada. A fim de se obter resultados mais acurados, Moral (2010) sugere selecionar o mais apropriado algoritmo geostatístico para cada área de estudo. Nesse contexto, deve-se verificar também qual a melhor variável secundária, bem como o seu melhor plano de amostragem.

Conforme sugere Goovaerts (2000), outros descritores ambientais devem ser investigados, os quais podem explicar a grande proporção da variabilidade espacial apresentada pela precipitação pluvial. Alguns autores têm levado em conta a influência marítima na precipitação em áreas costeiras e sugerido a distância da costa como uma importante variável no ajustamento de modelos de regressão (Johansson & Chen, 2003). No entanto, diferentemente da altitude, em que são conhecidos os bons resultados do uso desta como variável auxiliar na estimativa da precipitação pluvial (Diodato & Ceccarelli, 2005; Viola et al., 2010), para a variável distância da costa há ainda necessidade de avaliar de sua utilização como auxiliar na cokrigagem.

A configuração da amostragem dos dados é outro fator que interfere no desempenho preditivo da krigagem (Li & Heap, 2011). A amostragem regular e a aleatória tem sido duas estratégias comumente utilizadas em geostatística. A regular cobre sistematicamente a área de interesse, mas não representa distâncias menores que o tamanho da grade. A irregular representa as variações a distâncias mais curtas, no entanto, o espalhamento dos pontos no espaço geográfico é menor que o da grade regular, resultando muitas vezes em menor exatidão global (Hengl, 2007). No entanto, Lin & Heap (2011) ao estudar diferentes métodos de interpolação verificaram maior exatidão da predição com o uso de grades irregulares. Para o caso da krigagem

ordinária, Englund et al. (1992) verificaram que o padrão da amostragem não foi significativo em determinar o desempenho do interpolador.

Diante do exposto, os objetivos deste trabalho foram avaliar a cokrigagem ordinária com dados de altitude e distância do mar em comparação à krigagem ordinária na espacialização da precipitação pluvial dos períodos anual, seco e úmido no Estado do Espírito Santo.

2.4. MATERIAL E MÉTODOS

As séries de dados de precipitação pluvial utilizadas neste estudo foram obtidas na base de dados da Agência Nacional de Águas – ANA, a partir do sistema de informações hidrológicas Hidroweb (ANA, 2011) e da base de dados do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural - INCAPER. No total utilizaram-se 108 postos pluviométricos localizados no Estado do Espírito Santo, sendo 96 da ANA e 12 do INCAPER. Considerou-se 25 anos de dados, obtidos dos registros do período de 1976 a 2006.

Os pontos de medição da precipitação estão geograficamente distribuídos de forma irregular na área de estudo (Figura 1), sendo 92 postos inseridos no Estado do Espírito Santo e, os outros 16 postos nos Estados de Minas Gerais, Rio de Janeiro e Bahia. A inclusão destes postos fora dos limites da região objeto de estudo foi utilizada com o objetivo de minimizar o efeito de borda no processo de interpolação (Viviani & Manzato, 2005). Foram utilizados dados das séries de precipitação consistidos, disponíveis no sítio da ANA (<http://hidroweb.ana.gov.br/>). No entanto, algumas séries mensais apresentavam falhas, as quais foram preenchidas pelo método da ponderação regional (Bertoni & Tucci, 2007). Foram selecionados pelo menos três postos vizinhos, considerando séries de 10 anos de dados. As consistências das séries de dados do INCAPER e dos preenchimentos das falhas foram obtidas pelo método da dupla massa (Bertoni & Tucci, 2007). As médias mensais de precipitação foram obtidas e agrupadas em três períodos para análise dos interpoladores: anual; período seco (maio a setembro), e; período úmido (outubro a abril).

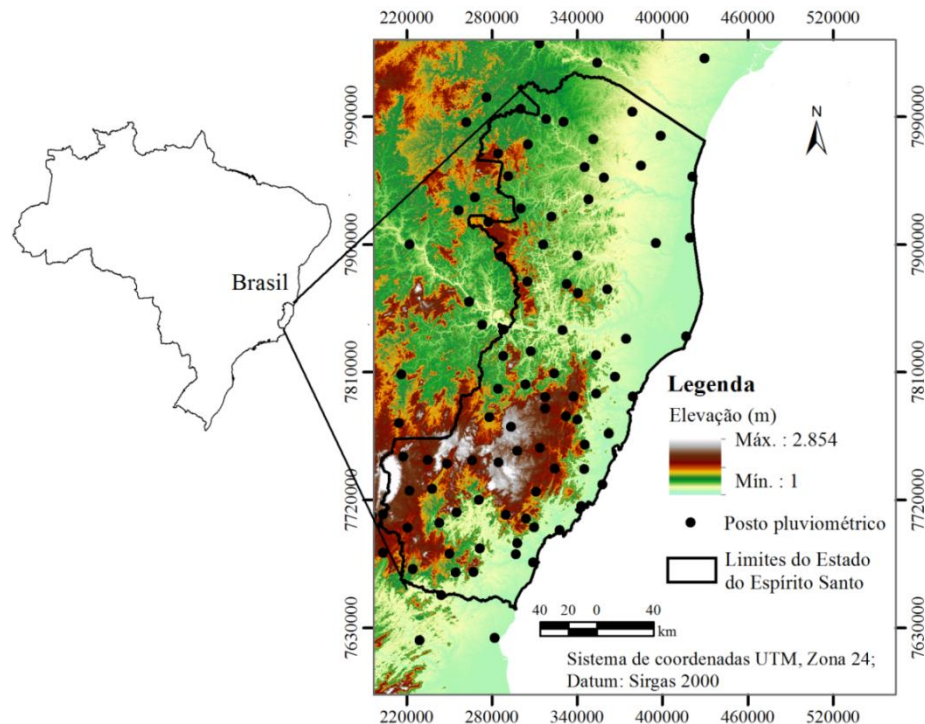


Figura 1. Localização dos postos pluviométricos na área de estudo.

Foram utilizadas a altitude e a distância do mar como variáveis secundárias na cokrigagem. Estas covariáveis foram obtidas em grades regulares, com espaçamento de 5.000 x 5.000 metros, o que gerou 3.014 pontos amostrais para a área. O espaçamento da grade foi escolhido de forma arbitrária e não se avaliou qual a melhor densidade de pontos. A mesma quantidade de pontos foi utilizada para se obter as covariáveis em grades irregulares. Obedeceu-se uma distância mínima de 1.000 metros entre pontos para assegurar um melhor espalhamento dos mesmos. Para cada um desses pontos foram obtidos os dados de altitude pela ferramenta “extract values to points” do ArcGis 10.0 (Esri, 2011), a partir de uma imagem SRTM (Shuttle Radar Topography Mission) com resolução de 90 metros. Os valores de altitude foram calculados das células adjacentes com valores válidos, utilizando-se interpolação bilinear. As distâncias dos pontos amostrais em relação ao mar foram obtidas com a ferramenta “Near” do Arcgis 10.0 (Esri, 2011), com a qual se determinou a distância de cada ponto amostral em relação à linha de costa. Considerou-se como costa a face do polígono do mapa estadual que limita com o oceano Atlântico (Figura 1).

Os dados de precipitação pluvial, altitude e distância do mar foram submetidos à análise de correlação de Spearman. Os dados de precipitação pluvial referentes aos três

períodos considerados foram submetidos à análise estatística descritiva. Essas análises foram realizadas com o uso do programa R (R DEVELOPMENT CORE TEAM, 2011).

Foi utilizado o pacote *geoR* (Ribeiro Junior & Diggle, 2001) do programa R (R DEVELOPMENT CORE TEAM, 2011), e por meio da análise dos semivariogramas experimentais nas direções 0°, 45°, 90° e 135° foi realizada a verificação de anisotropia dos dados de precipitação pluviométrica. Para os demais procedimentos utilizou-se o pacote “Geostatistical Analyst” do *software* Arcgis 10.0. (ESRI, 2011), no qual foram realizadas as análises de tendência, o ajuste do semivariograma e a validação cruzada. O mesmo programa também foi utilizado para realizar a análise geoestatística em razão de este software permitir a construção de covariogramas cruzados com mais de uma variável secundária.

A krigagem e a cokrigagem ordinária foram realizadas com prévia remoção de tendência dos dados e avaliados, que conforme Diggle e Ribeiro Júnior (2007) consiste na krigagem universal. Foi aplicado o modelo que melhor se ajustou aos semivariogramas. Realizaram-se duas cokrigagens: com a utilização de uma covariável apenas, ou seja, cokrigagem precipitação x altitude e cokrigagem precipitação x distância do mar. Avaliou-se também a cokrigagem com três conjuntos de dados, sendo um primário (precipitação) e dois secundários (altitude e distância do mar), para cada período de precipitação pluvial avaliado.

Os parâmetros dos modelos foram estimados automaticamente com foco nos resultados da validação cruzada. Os raios de busca foram limitados às distâncias correspondentes ao comprimento do maior semieixo da elipse de busca. O número de setor utilizado foi igual a um, o qual foi escolhido conforme o melhor resultado obtido na validação cruzada. O número de setores indica em quantas partes a elipse é dividida para busca de pontos vizinhos.

Mapas dos totais de precipitação para os três períodos avaliados e mapas de erro-padrão da predição foram gerados. O mapa de erro-padrão da krigagem, referido como o desvio-padrão da predição para qualquer ponto individual, foi obtido com a finalidade de informar sobre a confiança dos valores interpolados na área de interesse.

As interpolações foram avaliadas por meio do procedimento de validação cruzada (Isaaks & Srivastava, 1989). Foi utilizado o ajuste do modelo de regressão linear dos valores observados em função dos valores estimados. Quanto mais próximas de 1 e 0 forem as estimativas obtidas para β_1 e β_0 , respectivamente, melhor o método de

interpolação ou modelo de dependência espacial (Isaaks & Srivastava, 1989). Utilizaram-se também os valores obtidos para o erro médio absoluto (EMA), em percentagem, e raiz do erro quadrático médio (REQM), os quais são definidos pelas seguintes equações:

$$EMA(\%) = \left(\sum_{i=1}^n \frac{|\hat{z}(s_i) - z(s_i)|}{z(s_i)} \cdot 100 \right) \cdot \frac{1}{n}$$

$$REQM = \sqrt{\frac{\sum_{i=1}^n (\hat{z}(s_i) - z(s_i))^2}{n}}$$

Em que $z(s_i)$ representa a precipitação observada, $\hat{z}(s_i)$ os valores preditos e “n” o número de postos pluviométricos.

A semelhança entre os mapas produzidos foi avaliada pelos valores do índice Kappa (Congalton & Green, 2009), considerando-se como referência o mapa gerado com a utilização do interpolador “inverso de uma potência da distância”, dado pela equação:

$$xp = \frac{\sum_{i=1}^n \left(\frac{1}{d_i^m} * x_i \right)}{\sum_{i=1}^n \left(\frac{1}{d_i^m} \right)}$$

Em que: x_p = atributo interpolado; x_i = valor do atributo do i-ésimo ponto de amostragem; d_i = distância euclidiana entre o i-ésimo ponto de vizinhança e o ponto amostrado; n = número de amostras, e m = expoente da distância euclidiana.

Foram utilizados valores de expoentes das distâncias otimizados pelo software ArcGis 10.0 no qual é avaliado vários valores de expoente para identificar aquele que produz o menor valor da raiz do erro quadrático médio (REQM). Foram utilizados um mínimo de 10 e um máximo de 15 vizinhos mais próximos. Os valores dos expoentes foram: 2,82 para o período úmido, 3,84 para o período seco e 3,15 para o período anual. Os valores de precipitação foram agrupados em dez classes com base em intervalos geométricos. Verificou-se a significância valores dos índices Kappa pelo teste estatístico Z (Congalton & Green, 2009).

Foi realizada uma análise visual dos mapas de precipitação anual gerados pelos métodos utilizados e uma comparação dos mapas de erros padrão da predição da precipitação média anual, considerando as diferentes Macrorregiões do Estado do

Espírito Santo (IJSN, 2012). Para tanto, foram realizados recortes dos mapas gerados para todo o Estado, de maneira que estes foram segmentados nas seguintes Macrorregiões: Metropolitana, Sul, Noroeste e Norte.

2.5. RESULTADOS E DISCUSSÃO

A maior variabilidade do regime pluvial entre os postos pluviométricos ocorre na estação seca, conforme o coeficiente de variação apresentado (Tabela 1), o que pode evidenciar conforme Mello & Silva (2009) diferentes regiões climáticas. Além desse fato, a alta variabilidade nessa época pode estar relacionada a frentes frias. Segundo Viola et al. (2010), maiores coeficientes de variação nos meses secos foram encontrados por, para o estado de Minas Gerais. Esses autores associaram a variabilidade dos dados à presença de frentes frias mais intensas na região sul e que se enfraquecem à medida que avançam para a direção nordeste, indicando a existência de correlação entre a precipitação e a altitude.

Tabela 1. Estatística descritiva dos dados de precipitação pluvial média dos períodos úmido, seco e anual e dos dados de altitude e de distância do mar e coeficientes de correlação de Spearman, para o Estado do Espírito Santo.

Variável	un.	Média	V (%)	s ²	Min.	Max.	S	K	Coeficiente de correlação	
									Alt.	Dist.
Úmido	mm	1.016,0	15,0	23,289	644,40	1.451	0,30	-0,10	0,404**	-0,003
Seco	mm	216,7	38,9	7,098	83,29	448	0,70	-0,38	-,398**	-0,780**
Anual	mm	1.232,6	16,1	39,455	812,90	1.852	0,46	0,08	0,174	-0,347**
Alt.	m	261,1	104,7	74,73	4,00	1.041	1,13	0,08		0,599**
Dist.	km	68,8	65,7	2,044	0,08	191	0,34	-0,64		

Nível de significância: ** p<0,01. CV = coeficiente de variação; s² = variância; S = assimetria; K = curtose; Min. = mínimo; Max. = máximo.

Períodos: se referem a períodos para análise dos interpoladores: período anual; período seco (maio a setembro), e; período úmido (outubro a abril).

São vários os fatores que influenciam a distribuição espacial da precipitação, como passagens de frentes frias, fenômenos globais de circulação como o El Niño e La Niña, o gradiente térmico entre o mar e o continente e o relevo (Mello & Silva, 2009). Este último pode considerado de relevância no Estado do Espírito Santo, em razão dos

coeficientes de correlações apresentados principalmente entre a precipitação do período úmido com a altitude (Tabela 1). Segundo Carvalho & Assad (2005), índices pluviométricos mais elevados no litoral são relacionados ao relevo concordante (serras que se dispõem paralelamente à linha da costa) e as vertentes a barlavento (face em que o vento incide), as quais são mais expostas aos ventos úmidos. Nesse aspecto, conforme Carvalho et al. (2012) o relevo configura-se como um importante fator na formação das chuvas orográficas.

O maior coeficiente de correlação linear foi verificado entre a precipitação da estação seca e a distância do mar, o que indica uma forte associação negativa entre as variáveis. A utilização da variável altitude resultou em correlações moderadas e significativas com as precipitações das estações úmida e seca. No entanto, ao contrário do comportamento da precipitação na estação úmida, a correlação negativa entre as duas variáveis na estação seca indicou que nesta época do ano há uma redução da precipitação com o aumento da altitude. Nesse sentido, as correlações verificadas entre altitude e distância do mar e a precipitação em pelo menos uma estação do ano indica o potencial destas nas interpolações por meio de métodos bivariados. Silva & Lima encontraram baixa correlação entre altitude e precipitação para o Estado do Espírito Santo e, no entanto, obtiveram resultados satisfatórios de predição da precipitação pluvial média mensal.

Os patamares dos semivariogramas direcionais (Figura 2) mostraram-se aproximadamente iguais aos valores das variâncias dos dados (Tabela 1). Uma expressiva similaridade na região de influência da dependência espacial dos semivariogramas experimentais direcionais foi verificada apenas para a precipitação na estação úmida (Figura 2A). Nesse caso, conforme Carvalho & Assad (2005), os padrões de dependência espacial podem ser considerados idênticos nas direções 0°, 45°, 90° e 135°. Com base nestes comportamentos apresentados, é possível avaliar que o modelo isotrópico é suficiente para descrever a estrutura da continuidade espacial da precipitação da estação úmida. Os semivariogramas direcionais dos períodos seco e anual apresentaram comportamento diferenciado da continuidade espacial da precipitação, especialmente nas regiões mais próximas de atingir o patamar (Figura 2B e 2C), o que um indicativo da presença de anisotropia, na qual a variabilidade é dependente da direção (Goovaerts, 2000).

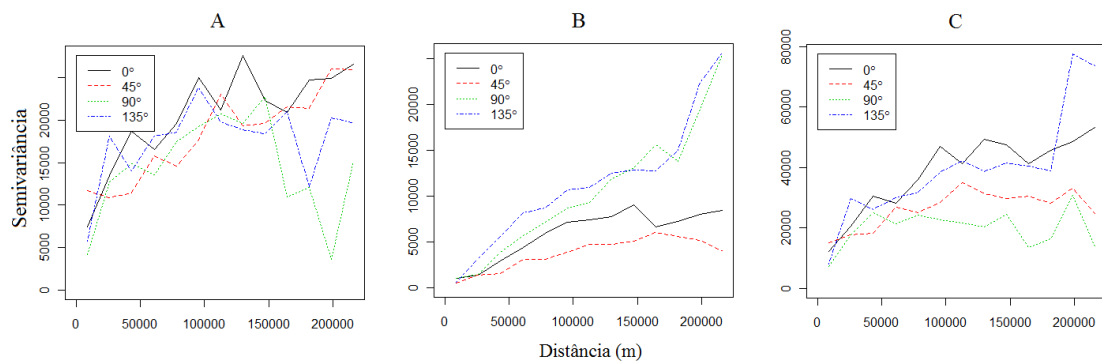


Figura 2. Semivariogramas experimentais para os dados de precipitação pluvial dos períodos úmido (A), seco (B) e anual (C), em quatro direções (0° , 45° , 90° e 135°).

No período seco e anual, a anisotropia encontrada foi a do tipo “combinada”, que é a mais comumente presente em variáveis ambientais (Camargo et al., 2004). Este tipo de anisotropia combina a anisotropia geométrica (refletida pelo mesmo patamar e diferentes alcances) com a anisotropia zonal (em que os semivariogramas direcionais apresentam os mesmos alcances e diferentes patamares) (Camargo et al., 2004). Nesses dois períodos, os semivariogramas direcionais das precipitações apresentaram comportamentos mais semelhantes entre si em distâncias mais curtas entre os pares de pontos (Figura 2). Como, na prática, a estimação dos semivariogramas experimentais é realizada com mais precisão para distâncias próximas da origem (Carvalho et al. 2009) e estes não indicaram acentuada anisotropia para a variável primária, a análise geostatística foi realizada admitindo-se uma distribuição isotrópica. A condição de isotropia para os dados de precipitação foram então assumidas para simplificação de procedimentos tanto para a construção dos semivariogramas quanto para a dos covariogramas cruzados, conforme análises direcionais realizadas. Segundo Carvalho et al. (2009), diferenças de altitudes podem influenciar a variografia dos dados de precipitação. Nesse aspecto, o relevo do Estado do Espírito Santo, com grandes mudanças de altitudes a curtas distâncias do litoral, apresenta-se como um provável fator a influenciar a dependência espacial da precipitação nas diferentes direções. No entanto, a causa da anisotropia usualmente não é conhecida (Krishna Murthy and Abbaisah, 2007).

Os modelos de semivariograma e covariograma que melhor se ajustaram aos dados de precipitação e das covariáveis foram o exponencial e o esférico (Tabela 2). No caso da krigagem ordinária, o modelo esférico se mostrou mais ajustado aos dados de

precipitação no período seco e anual. Na krigagem ordinária dos dados do período úmido e para os covariogramas cruzados, o modelo exponencial foi o que melhor se ajustou. Em um estudo semelhante, porém com objetivo de espacializar a precipitação pluvial mensal média do Estado do Espírito Santo por cokrigagem, Silva & Lima (2011) obtiveram melhor ajuste com o modelo esférico. Na mesma área, Castro (2008) obteve para a precipitação média anual um melhor ajuste com a utilização do modelo linear. Nesse sentido, os resultados deste e de outros trabalhos (Castro, 2008; Silva & Lima, 2011) para a área de estudo são indicativos de que o modelo mais adequado para interpolação de dados de precipitação por meio de método geoestatístico no Estado pode depender de vários fatores relacionados aos conjuntos de dados, como quantidade de postos utilizados, tempo considerado na série histórica, períodos do ano entre outros.

Tabela 2. Modelos e parâmetros semivariogramas e covariogramas cruzados para espacialização dos dados de precipitação pluvial média dos períodos úmido, seco e anual, para o Estado do Espírito Santo.

Período	Semivariograma / covariograma	Modelo	Efeito pepita - C_0	Variância estrutural - C_1	Alcance - a (km)
Úmido	P	Exponencial	928	19.677	75
	P x Alt.	Exponencial		19.044	94
	P x Dist.	Exponencial		-282	75
Seco	P	Esférico	600	3.055	75
	P x Alt.	Exponencial		610	68
	P x Dist.	Exponencial		-598	74
Anual	P	Exponencial	0	27.022	30
	P x Alt.	Exponencial		25.248	75
	P x Dist.	Exponencial		1.173	75

Os alcances apresentados nos semivariogramas para a precipitação anual nesse estudo (Tabela 2) foram menores que aqueles encontrados por Castro (2008) e Silva & Lima (2011) para as precipitações médias mensais. No entanto, os resultados encontrados por aqueles autores foram mais próximos aos dos períodos seco e úmido desse trabalho. O alcance significa, conforme Carvalho et al., (2012), que os pontos localizados numa área com raio menor ou igual ao valor deste parâmetro apresentam precipitações semelhantes e estão correlacionados entre si, podendo estimar com maior precisão valores não medidos nessa área. Em geral, independente do período considerado, os alcances obtidos nos covariogramas apresentaram média de 76,8 km, ou seja, mais de duas vezes o alcance verificado no semivariograma gerado com os dados

de precipitação anual. Apesar do menor alcance, a semivariograma da precipitação anual revelou a mais elevada dependência espacial, indicada pelos valores de efeito pepita e variância estrutural.

Na validação cruzada para interpolação dos dados de precipitação anual, os métodos krigagem ordinária e cokrigagem ordinária produziram resultados comparáveis de erros dos valores preditos, parâmetros de regressão e coeficiente de correlação dos valores observados versus valores preditos (Tabela 3). Entre os dois métodos estocásticos utilizados, a cokrigagem com a altitude produziu o valor mais baixo de REQM, um dos menores erros médios absolutos (EMA) e o mais elevado R^2 . Os erros absolutos médios dos valores preditos, abaixo de 10%, foram em geral semelhantes aos encontrados por Castro (2008) ao interpolar a precipitação média anual para a mesma área de estudo. Os resultados encontrados por esse autor corroboram os resultados verificados para a krigagem e as cokrigagens com dados de altitude e de distância do mar.

Na avaliação das grades regulares e irregulares para obtenção das covariáveis com vistas à utilização da cokrigagem, verificou-se que as grades regulares, para os dados de altitude, apresentaram em geral maior eficácia na predição espacial da precipitação média anual (Tabela 3).

No período úmido, os resultados da validação cruzada da cokrigagem e da krigagem mostraram que os menores valores de REQM, maiores coeficientes de determinação ($R^2 = 0,58$) e melhores ajustes dos parâmetros das regressões foram obtidos para cokrigagem com altitude e cokrigagem com altitude mais distância do mar. Estes resultados evidenciam, em muitos casos, uma melhor estimativa da precipitação da estação úmida por meio da cokrigagem com presença da covariável altitude, quando comparada com as estimativas apresentadas pela krigagem ordinária.

Na comparação entre grades regulares e irregulares para o período úmido, os valores de EMA e REQM apresentaram comportamentos semelhantes aos do período anual. As cokrigagens que envolveram a utilização da covariável altitude apresentaram menores erros dos valores preditos. Dessa maneira, ao considerarem-se os períodos anual e úmido, a utilização de grade regular tendeu a permitir desempenho superior com relação aos parâmetros da regressão, obtendo-se os valores mais baixos de β_0 e mais elevados de β_1 . Estes dois coeficientes obtidos da regressão linear entre os valores observados e os preditos pelas krigagens e cokrigagens fornecem a qualidade do ajuste

dos modelos matemáticos aos semivariogramas (Carvalho et al., 2009). Indicativos de melhores desempenhos na interpolação geoestatística por meio dos coeficientes β_0 e β_1 foram encontrados também por Soares et al. (2008) e Carvalho et al. (2009).

Tabela 3. Estimativas dos parâmetros da validação cruzada para os semivariogramas ajustados aos dados de precipitação dos períodos úmido, seco e anual, utilizando-se krigagem ordinária e cokrigagem.

Método/ covariável	Modelo	Grade	Erro de predição		R ²	Regressão	
			EMA (%)	REQM (mm)		$\hat{\beta}_0$	$\hat{\beta}_1$
----- <i>Período úmido</i> -----							
KO	Exponencial		8,16	104,5	0,53	519,3	0,497
CKA	Exponencial	Regular	7,68	98,4	0,58	445,1	0,581
	Exponencial	Irregular	7,99	103,2	0,54	497,2	0,513
CKD	Exponencial	Regular	8,13	104,3	0,53	518,0	0,498
	Exponencial	Irregular	8,18	104,9	0,53	522,9	0,494
CKAD	Exponencial	Regular	7,67	98,53	0,58	451,2	0,554
	Exponencial	Irregular	8,39	107,5	0,50	548,8	0,460
----- <i>Período seco</i> -----							
KO	Esférico		12,33	38,76	0,77	19,26	0,897
CKA	Exponencial	Regular	12,05	38,33	0,78	20,74	0,891
	Exponencial	Irregular	12,07	38,23	0,78	21,09	0,889
CKD	Exponencial	Regular	12,44	38,68	0,78	23,15	0,882
	Exponencial	Irregular	12,37	38,88	0,77	20,45	0,892
CKAD	Exponencial	Regular	11,92	37,70	0,79	20,27	0,893
	Exponencial	Irregular	12,22	38,36	0,78	16,37	0,910
----- <i>Período Anual</i> -----							
KO	Esférico		8,25	135,5	0,53	527,0	0,580
CKA	Exponencial	Regular	7,62	124,1	0,61	480,1	0,606
	Exponencial	Irregular	8,11	132,9	0,54	447,2	0,636
CKD	Exponencial	Regular	8,32	135,4	0,51	452,7	0,636
	Exponencial	Irregular	8,28	134,9	0,51	448,0	0,640
CKAD	Exponencial	Regular	7,60	123,6	0,61	446,6	0,633
	Exponencial	Irregular	8,06	132,5	0,54	444,2	0,638

KO – krigagem ordinária; CKA – cokrigagem com altitude; CKD – cokrigagem com distância do mar; CKAD – cokrigagem com altitude e distância do mar; EMA – erro médio absoluto; REQM – raiz do erro quadrático médio; R² – coeficiente de determinação entre valores preditos e observados.

Períodos: se referem a períodos para análise dos interpoladores: período anual; período seco (maio a setembro), e; período úmido (outubro a abril).

Por meio da análise dos resultados de validação para a estação seca, nota-se que, no conjunto, a cokrigagem com altitude e distância do mar apresentou melhor desempenho na predição espacial em relação à krigagem e cokrigagem com altitude. Diferentemente do observado para os períodos anual e úmido, a covariável distância do mar mostrou-se eficaz para predição da precipitação nesse período, se associada à altitude. Este resultado pode estar relacionado às precipitações orográficas, que no caso

período seco para o Espírito Santo, parece ser mais eficientemente predita considerando simultaneamente a altitude e distancia do mar. Conforme os coeficientes de correlação verificados, quanto menor a altitude e maior a proximidade do mar maior a precipitação nesse período. Essa influência da distância do mar na melhoria da interpolação está relacionada com a boa correlação entre esta variável e a precipitação nesse período. No entanto, esperava-se melhor desempenho da cokrigagem com distância do mar se considerado que o maior coeficiente de correlação entre as covariáveis e a precipitação pluvial foi verificado para a distância do mar e precipitação do período seco.

Os valores de EMA para o período seco foram superiores aos observados para os períodos anual e úmido (Tabela 3). Comportamento semelhante e valores compatíveis de EMA foram encontrados por Silva et al. (2011) com a utilização da krigagem ordinária para interpolação de dados de precipitação mensal no Estado do Espírito Santo. Neste caso, maiores erros de predição podem evidenciar a maior variabilidade espacial da precipitação da estação seca, que apresenta maior coeficiente de variação (38,9%). Nesse sentido, Viola et al. (2010) concluíram que a maior variabilidade dos índices pluviométricos na época seca para o Estado de Minas Gerais dificultou a reconstituição da continuidade espacial pelos interpoladores e produziu maiores erros de predição. A alta variabilidade espacial da época seca origina-se da grande amplitude de precipitação nos diferentes postos pluviométricos e está relacionada a vários fatores (Mello & Silva, 2009), sendo um deles a própria magnitude dos valores de precipitação nesse período, pois como é baixa, uma pequena alteração nos mesmos pode representar variabilidade significativa, impactando nos valores do coeficiente de variação. Conforme verificado por Viola et al. (2010) para o estado de Minas Gerais, as frentes frias é um dos fatores que conduz a uma maior variabilidade das chuvas, as quais são mais reduzidas ao norte daquele Estado. No caso do Espírito Santo, além desses fatores, a alta variabilidade na época seca pode ser influenciada por um provável efeito das precipitações orográficas.

Embora a avaliação das grades regulares e irregulares tenha sido para as covariáveis, os resultados observados são em geral consistentes com a teoria geoestatística. Na krigagem, para um mesmo tamanho amostral, o desempenho preditivo é afetado pelo plano de amostragem (Li & Heap, 2011) e, de modo geral, as grades regulares resultam em variância da predição minimizada, pois fornecem melhores estimativas em relação às amostragens aleatórias em grades irregulares

(Hengl, 2007). Entretanto, o padrão de amostragem regular ou irregular pode, em alguns casos, não resultar em diferenças significativas (Li & Heap, 2011). Nesse sentido, diferente das interpolações nos períodos úmido e anual, com base nos erros de predição, o uso de grades regulares para as covariáveis na estação seca produziu na maioria dos casos estimativas equiparáveis ao de grades irregulares (Tabela 3).

As pequenas diferenças nos valores de EM e REQM nas comparações entre krigagem e cokrigagem e entre grade regular e irregular são reforçadas pela análise estatística dos valores dos índices Kappa (Tabela 4). No período úmido, a quase totalidade das comparações apresentou diferença significativa, mas em três situações não houve diferença significativa entre os mapas gerados, quais sejam: na comparação entre krigagem versus cokrigagem com distância do mar em grade regular; cokrigagem com altitude versus cokrigagem com altitude e distância do mar, ambas em grade regular; cokrigagem com distância do mar em grade regular versus cokrigagem com distância do mar e grade irregular. Estas situações correspondem aos casos em que as diferenças nos erros médios são mais baixas (Tabela 3), o que indica que apenas nesses casos os mapas não são estatisticamente diferentes pelo teste Z, ao nível de 5% de probabilidade. No período seco, o mesmo ocorre também em três situações: na comparação entre krigagem versus cokrigagem com distância do mar em grade regular; cokrigagem com altitude em grade regular versus cokrigagem com altitude em grade irregular; cokrigagem com altitude e distância do mar em grade regular versus cokrigagem com distância do mar em grade irregular. No caso da precipitação anual, todos os mapas comparados apresentaram diferenças estatísticas entre si pelo teste Z, o que constitui um indicativo de que as pequenas diferenças de EMA e REQM apresentadas na Tabela 3 foram suficientes para revelar em muitos casos a superioridade da cokrigagem em relação à krigagem e das grades regulares em relação às irregulares.

Nesse sentido, a cokrigagem proporcionou uma ligeira superioridade na predição da precipitação para os três períodos avaliados. Segundo Isaaks & Srivastava (1989), é possível uma redução na variância do erro de estimação com a cokrigagem quando existe correlação estatística entre elas e o semivariograma cruzado demonstra dependência entre as variáveis. No entanto, os baixos coeficientes de correlação verificados entre as variáveis e as covariáveis utilizadas podem ter limitado um maior ganho no desempenho da cokrigagem. De acordo com Rocha et al. (2012), uma maior influência da variável secundária ocorre quando a correlação entre esta e a variável

primária é mais alta. Os resultados do estudo desses autores mostram claramente que quando a correlação é baixa, a variável primária é retida como informação mais confiável no processo de interpolação por cokrigagem.

Tabela 4. Matriz de significância entre índices Kappa, variâncias do Kappa (σ^2) dos diferentes mapas de precipitação nos períodos úmido, seco e anual, obtidos por cokrigagem, ao se considerar como referência os mapas gerados por meio do interpolador inverso de uma potência da distância.

Tipo de grade / Método	----- Grade regular -----				----- Grade irregular -----			
	KO	CKA	CKD	CKAD	CKA	CKD	CKAD	
----- Período úmido -----								
Kappa	0,66	0,54	0,66	0,54	0,59	0,66	0,50	
σ^2 (x 10 ⁻⁶)	4,8608	5,5745	4,8628	5,5934	5,3508	4,8709	5,7380	
Grade regular	KO	0,00	-36,2*	-0,04	-37,22*	-23,10*	-0,39	-50,33*
	CKA		0,00	36,08*	-1,10	12,97*	35,72*	-14,03*
	CKD			0,00	-37,19*	-23,06*	-0,36	-50,29*
Grade irregular	CKA				0,00	14,07*	36,83*	-12,92*
	CKA					0,00	22,70*	-27,05*
	CKD						0,00	-49,93*
CKA							0,00	
----- Período seco -----								
Kappa	0,62	0,66	0,62	0,61	0,65	0,61	0,59	
σ^2 (x 10 ⁻⁶)	5,1042	4,8371	5,0636	5,1699	4,8724	5,1501	5,2715	
Grade regular	KO	0,00	12,16*	1,81	-3,76*	10,50*	-2,32*	-9,67*
	CKA		0,00	-10,35*	-15,93*	-1,66	-14,48*	-21,86*
	CKD			0,00	-5,57*	8,70*	4,13*	-11,49*
Grade irregular	CKA				0,00	148*	1,44	-5,91*
	CKA					0,00	-12,83*	-20,20*
	CKD						0,00	-7,35*
CKA							0,00	
----- Período anual -----								
Kappa	0,60	0,52	0,57	0,51	0,55	0,57	0,54	
σ^2 (x 10 ⁻⁶)	5,1273	5,4680	5,3139	5,4987	5,4204	5,3173	5,4384	
Grade regular	KO	0,00	-25,56*	-11,74*	-29,27*	-17,70*	-11,56*	-20,15*
	CKA		0,00	13,79*	-3,68*	7,80*	13,96*	5,37*
	CKD			0,00	-17,48*	-5,97*	0,18	-8,40*
Grade irregular	CKA				0,00	11,48*	17,65*	9,05*
	CKA					0,00	6,14*	-2,43*
	CKD						0,00	-8,58*
CKA							0,00	

KO – krigagem ordinária; CKA – cokrigagem com altitude; CKD – cokrigagem com distância do mar; CKAD – cokrigagem com altitude e distância do mar; * significativo a 5% de probabilidade pelo teste Z.

Os valores de precipitação média anual variam predominantemente de forma crescente no sentido norte-sul do Estado, com os maiores valores na região serrana sul (Figura 3C). Esta distribuição espacial da precipitação observada na Figura 3 está de acordo com Silva & Lima (2011), que verificaram que os menores índices de precipitação ocorrem na região norte, oeste e litoral sul do estado do Espírito Santo; e com os maiores índices, em parte, no litoral sul e região serrana sul.

Nos mapas de erro-padrão da predição, que medem a confiança dos valores interpolados, constata-se que as localizações próximas aos postos pluviométricos geralmente possuem erros-padrão menores que as mais distantes (Figura 4). Conforme Mello Junior et al. (2006), esse comportamento é característico da variância de krigagem. Uma maior variância representa, também, a maior incerteza para a predição, o que indica a necessidade de um número maior de postos pluviométricos na região norte do Estado, caso o objetivo seja a redução de áreas com elevada incerteza. Além deste fato, de acordo com Webster & Oliver (2007), para se obter uma estimativa confiável de variogramas isotrópicos são necessários dados de pelo menos 144 pontos de amostragem e que em torno de 400 pontos permite a estimação com grande precisão.

Os maiores erros-padrão das predições referentes à precipitação média anual ocorreram geralmente em áreas situadas próximas às regiões nordeste e norte do estado. Na cokrigagem com altitude, que obteve melhor desempenho na interpolação, conforme resultados da validação cruzada, a classe de erro-padrão com valores menores que 89,2 mm ocupam 10,5% da área do Estado (Figura 4). Na krigagem ordinária, essa mesma classe ocupa um total de 4,3% da área. Numa comparação entre regiões, no norte do Estado a ocorrência dessa classe foi de 0,5%, enquanto na metropolitana alcançou 4,6%. Nas áreas ao norte do Estado, os valores preditos de precipitação estão supostamente associados com um maior nível de incerteza, a qual se relaciona com a baixa densidade de postos pluviométricos. Além disso, as áreas com maior incerteza da estimação foram reduzidas nos mapas de precipitação gerados por cokrigagem.

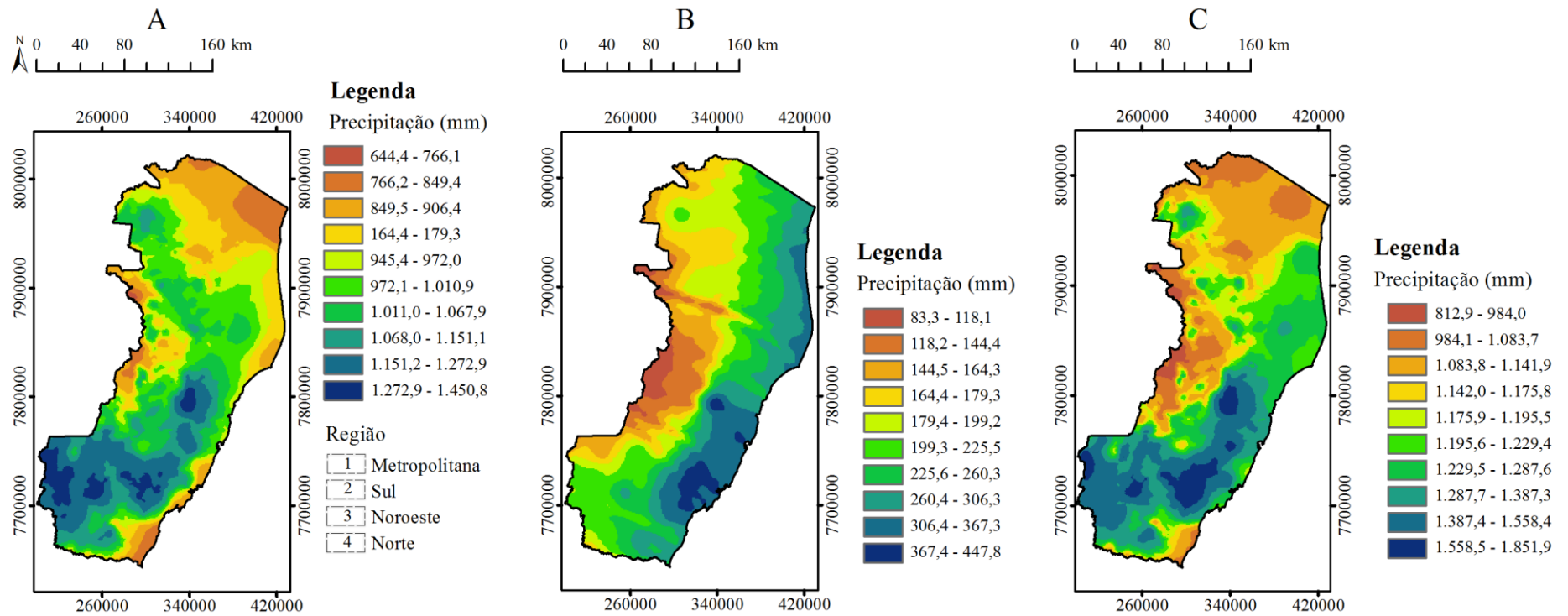


Figura 3. Mapas de precipitação pluvial dos períodos úmido, seco e anual estimados por cokrigagem com altitude (A), cokrigagem com altitude mais distância do mar (B) e cokrigagem com altitude (C), respectivamente.

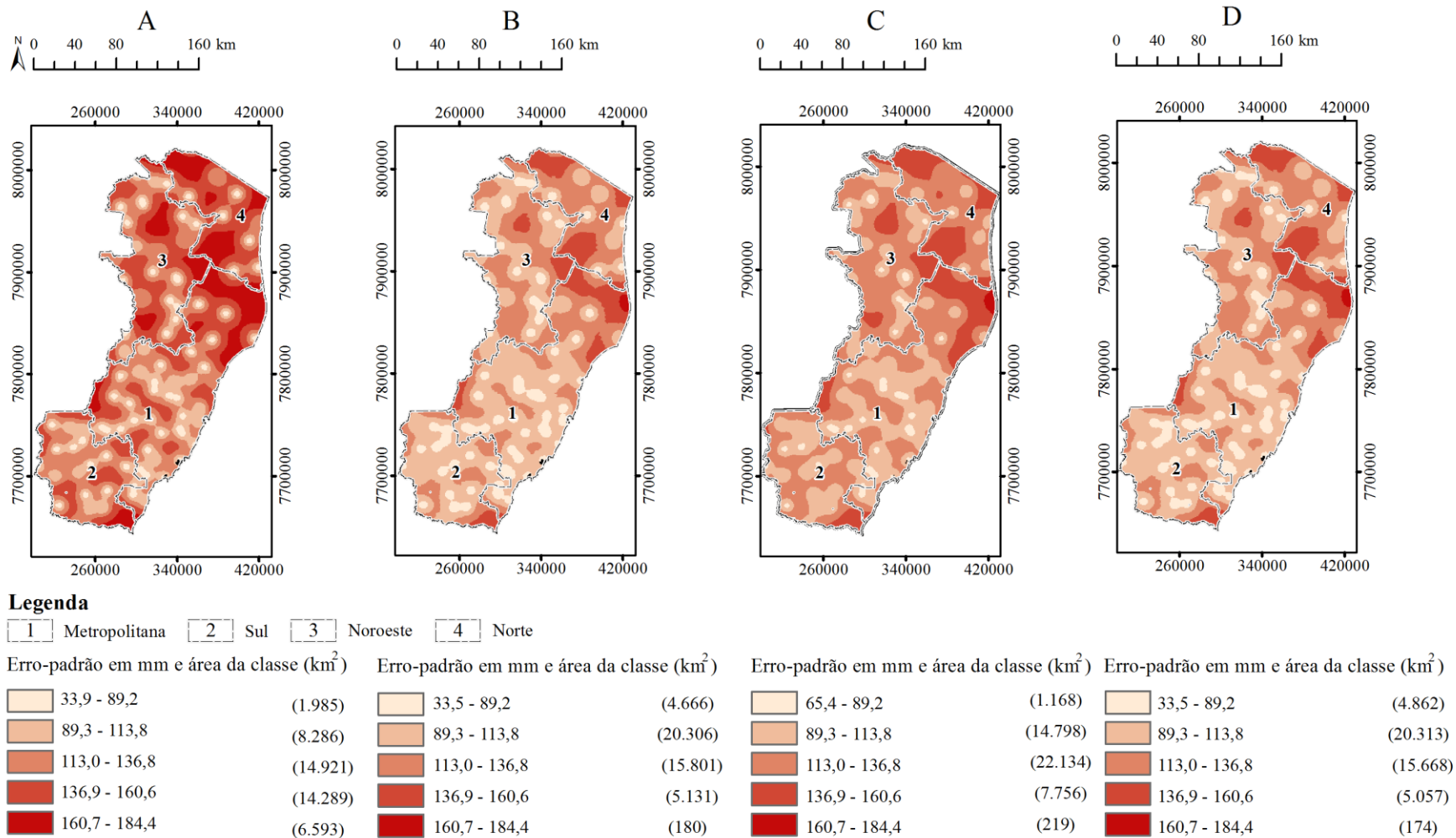


Figura 4. Mapas de erro-padrão da predição para precipitação pluvial anual (mm) obtidos por krigagem (A), cokrigagem usando altitude (B), distância do mar (C) e altitude e distância do mar (D).

2.6. CONCLUSÕES

1. A cokrigagem fornece menores erros médios absolutos e erros quadráticos médios na validação cruzada, além de mapas estatisticamente diferentes da krigagem ordinária e com menores erros-padrão da estimativa, na interpolação da precipitação pluvial no estado do Espírito Santo.

2. A utilização da altitude como covariável possibilita melhor desempenho na interpolação por cokrigagem em relação à distância do mar.

2.7. REFERÊNCIAS

- AGÊNCIA NACIONAL DE ÁGUA - ANA. HIDROWEB: **Dados Hidrológicos**. Disponível em: <<http://hidroweb.ana.gov.br/>>. Acesso em: 18 out. 2011.
- BERTONI J.C.; TUCCI, C.E.M. Precipitação. In: Tucci, C.E.M. **Hidrologia: Ciência e aplicação**. Porto Alegre: UFRGS, 2007. p.177-241.
- CAMARGO, E.C.G.; FUCKS, S.D.; CÂMARA, G. Análise espacial de superfícies. In: DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.M.V. (Ed.). **Análise espacial de dados geográficos**. Planaltina: Embrapa Cerrados, 2004. p.79-122.
- CARVALHO, J.R.P.; ASSAD, E.D. Análise espacial da precipitação pluviométrica no Estado de São Paulo. *Engenharia Agrícola*, v.25, p.377-384, 2005.
- CARVALHO, J.R.P.; ASSAD, E.D.; PINTO, H.S. Interpoladores geoestatísticos na análise da distribuição espacial da precipitação anual e de sua relação com altitude. **Pesquisa Agropecuária Brasileira**, v.47, p.1235-1242, 2012.
- CARVALHO, J.R.P.; VIEIRA, S.R.; GREGO, C.R. Comparação de métodos para ajuste de modelos de semivariograma da precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.4, p.443-448, 2009.
- CASTRO, F.S. **Zoneamento agroclimático para a cultura do pinus no estado do Espírito Santo**. 2008. 101p. Tese (Mestrado) - Universidade Federal do Espírito Santo, Alegre.
- CONGALTON, R.G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2.ed. Boca Raton: Taylor and Francis, 2009.

- DIGGLE, P.J.; RIBEIRO JUNIOR, P.J. **Model-based geostatistics**. New York: Springer, 2007. 228 p.
- DIODATO, N.; CECCARELLI, M. Interpolation processes using multivariate geostatistics for mapping of climatological precipitation mean in the Sannio Mountains (Southern Italy). **Earth Surface Processes and Landforms**, v.30, p.259-268, 2005.
- ESRI. ArcGIS 10.0. Environmental Systems Research Institute, Redlands, CA, USA. <http://www.esri.com>, 2011.
- GOOVAERTS, P. Geostatistical Approaches for Incorporating Elevation Into the Spatial Interpolation of Rainfall. **Journal of Hydrology**, v.228, n.1/2, p.113-129, 2000.
- HENGL, T. **A Practical Guide to Geostatistical Mapping of Environmental Variables**. Luxembourg: Office for Official Publication of the European Communities, 2007. 146p.
- INSTITUTO JONES DOS SANTOS NEVES – IJSN. **Limite entre Macrorregiões do ES**. 2012. Disponível em: <http://www.ijsn.es.gov.br/Mapas/Consulta/?module=Search&action=Results&lookfor=&type=AllFields&filter%5B%5D=format%3A%22Arquivo+Shapefile%22&view=list>. Acesso em 13/02/2013.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University, 1989. 561p.
- JOHANSSON, B.; CHEN, D. The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modeling. **International Journal of Climatology**, v.23, p.1523-1535, 2003.
- KRISHNA MURTHY B.R.; ABBAIAH, G. Geostatistical analysis for estimation of mean rainfalls in Andhra Pradesh India. **International Journal of Geology**, v.3, p.35–51, 2007.
- LI, J.; HEAP, A.D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. **Ecological Informatics**, v.6, p.228–241, 2011.
- MELLO, C.R.; SILVA, A.M. Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.13, n.1, p.68-74, 2009.

- MORAL, F.J. Comparison of different geostatistical approaches to map climate variables: application to precipitation. **International Journal of Climatology**, v.30, p.620–631, 2010.
- PORTO JUNIOR, R.; PIRES, B.P.; GOUVEIA, G.R.; BRANDÃO, V.S.; COUTINHO, N.M. Caracterização geológico-geotécnica do movimento de massa ocorrido entre as praias da Prainha e Grumari, no município do Rio de Janeiro, RJ. **Anuário do Instituto de Geociências**, v.35, p.5-13, 2012.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2011. Disponível em: <<http://www.R-project.org>>
- RIBEIRO JÚNIOR, P.J.; DIGGLE, P.P. GeoR: A package for geostatistical analysis. **R – News**, v.21, p.15-18, 2001.
- ROCHA, M.M.; YAMAMOTO, J.K.; WATANABE, J.; FONSECA, P.P. Studying the influence of a secondary variable in Collocated Cokriging estimates. **Anais da Academia Brasileira de Ciências**, v.84, p.335-346, 2012.
- SILVA, J.W.; GUIMARÃES, E.C. & TAVARES, M. Variabilidade temporal da precipitação mensal e anual na estação climatológica de Uberaba-MG. **Ciência e Agrotecnologia**, v.27, p.665-674, 2003.
- SILVA, K.R.; CECÍLIO, R.A.; XAVIER, A.C.; PEZZOPANE, J.R.M.; GARCIA, G.O. Interpolação Espacial da Precipitação no Estado do Espírito Santo. **Floresta e Ambiente**, v.18, p.417-427, 2011.
- SILVA, S.A.; LIMA, J.S.S. Número de postos pluviométricos necessários para a estimativa da precipitação mensal no estado do Espírito Santo, Brasil. **Revista Brasileira de Meteorologia**, v.26, p.555-560, 2011.
- SOARES, V.P.; ZANETI, L.Z.; SANTOS, N.T.; LEITE, H.G. Análise espacial da distribuição de cigarras (*Quesada gigas* Oliver) em povoamentos de paricá (*Schizolobiumamazonicum* Huber exDucke) na região de Dom Eliseu, PA. **Revista Árvore**, v.32, n.2, p.251-258, 2008.
- VIOLA, M.R.; MELLO, C.R.; PINTO, D.B.F.; MELLO, J.M.; ÁVILA, L.F. Métodos de interpolação espacial para o mapeamento da precipitação pluvial. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.14, p.970-978, 2010.
- VIVIANE, E.; MANZATO, G.G. Geração de modelos digitais de superfície por meio de plataformas computacionais com estrutura vetorial e raster. **Ciência & Engenharia**, v.15, p.27-34, 2005.

WEBSTER, R.; OLIVER, M. **Geostatistics for environmental scientists**. Chichester:
John Wiley & Sons, 2007. 315p.

3. CAPÍTULO 2 - SELEÇÃO DE PREDITORES AMBIENTAIS NO MAPEAMENTO DIGITAL DE SOLOS DA BACIA DO RIO MUQUI DO NORTE, ESTADO DO ESPÍRITO SANTO, BRASIL

Submetido à Revista Geoderma

3.1. RESUMO

Estudos sobre métodos de mapeamento digital de solos têm sido realizados em vários países e busca-se um aperfeiçoamento na quantidade de variáveis preditoras ambientais e na seleção das mesmas. Na busca desse propósito, avaliaram-se diferentes variáveis ambientais no mapeamento digital de solos em uma bacia hidrográfica no sul do Estado do Espírito Santo. Para isso, utilizou-se um algoritmo de classificação por árvore de decisão. As variáveis (45) relacionaram-se a diferentes atributos do terreno, geologia, geomorfologia, clima e índices derivados das bandas 1, 3, 4 e 5 do sensor Landsat 5. Essas variáveis foram submetidas a diferentes métodos de seleção de atributos (“CFSSubsetEval”, “ConsistencySubsetEval”, “InfogainAttributeEval” e “ReliefAttributeEval”), disponíveis no software Weka 3.6.8. Subsequentemente, com a utilização do mesmo software, foi aplicado o classificador SimpleCart para avaliar a efetividade da predição com os subconjuntos de variáveis selecionadas. Os resultados indicaram que somente nove variáveis selecionadas pelo algoritmo “ReliefAttributeEval” foram suficientes para fornecer uma classificação considerada muito boa (Kappa de 0,60) combinada com um menor tamanho da árvore de decisão. Este subconjunto de nove variáveis envolveu a combinação de variáveis ambientais relacionadas aos atributos do terreno, geologia, geomorfologia e variáveis climáticas. Estes resultados evidenciaram que a utilização de atributos climáticos pode ser muito útil no mapeamento digital de solos.

Termos de indexação: pedometria, seleção de atributos, árvore de decisão.

**SELECTION OF ENVIRONMENTAL PREDICTORS FOR DIGITAL SOIL
MAPPING IN THE MUQUI DO NORTE RIVER BASIN, ESPÍRITO SANTO
STATE, BRAZIL**

3.2. ABSTRACT

Studies on digital soil mapping have been conducted in several countries in search of improvements in the amount of environmental predictor variables and their selection. For this purpose, the different environmental variables in digital soil mapping were evaluated in a watershed in the south of the state of Espírito Santo, using one classification algorithm per decision tree. The 45 study variables were related to the different properties of terrain, geology, geomorphology, climate, water balance and indices derived from bands 1, 3, 4 and 5 of the Thematic Mapper (TM) sensor of the Landsat 5 satellite. These variables were subjected to different methods of attribute selection ("CFSSubsetEval", "ConsistencySubsetEval", "InfogainAttributeEval" and "ReliefFAttributeEval"), a feature of software Weka 3.6.8. Subsequently, using the same software, SimpleCart classifier was applied to evaluate the effectiveness of the prediction with the subsets of the selected variables. Results indicated that only nine variables selected by the algorithm "ReliefAttributeEval" were sufficient to provide a very good classification (Kappa 0.60), combined with a smaller decision tree size. This subset of nine variables involved the combination of environmental variables related to attributes of terrain, geology, geomorphology, and water balance variables. These results show that climatic and water balance variables are useful for digital soil mapping at small and medium scales.

3.3. INTRODUÇÃO

As predições de classes ou de atributos dos solos baseiam-se usualmente no modelo "scorpan" formulado por McBratney et al. (2003), fundamentado nos cinco fatores de formação do solo (Jenny, 1941): material de origem, relevo, clima, organismos e tempo. Os atributos do terreno ou topográficos são os preditores mais utilizados no mapeamento digital de solos (McBratney et al., 2003). Isto se deve ao fato de que o relevo é um dos mais importantes fatores de formação dos solos (Jenny, 1941) e, também, à disponibilidade de modelos digitais de elevação (Behrens et al., 2010), como os oriundos do SRTM (Shuttle Radar Topography Mission) e de cartas topográficas. No entanto, a necessidade de se incluir nos modelos outras variáveis ambientais relacionadas à formação dos solos é de grande importância para aumentar a eficácia da predição de classes de solos pelo mapeamento digital.

O uso de variáveis climáticas na espacialização de atributos do solo não tem sido amplamente relatadas na literatura (McBratney et al., 2003; Grunwald, 2009), possivelmente pela falta de dados ou mapas climáticos detalhados. Esses trabalhos que utilizam variáveis climáticas estão especificamente voltados para o mapeamento de propriedades dos solos (Böhner & Selige, 2006). Este fato induz à necessidade de se avaliar a utilização dessas variáveis como auxiliares aos atributos do terreno na predição espacial de classes de solo, especialmente em mapeamentos de médias a pequenas escalas (menores que 1:100.000, por exemplo).

McBratney et al. (2003), sugere que deveria-se tentar uma modelagem que representasse todos os fatores de formação do solo. Uma modelagem com maior número possível desses fatores geraria elevado número de covariáveis, pois apenas com o fator relevo, por meio de um modelo digital de elevação – MDE, dezenas de covariáveis preditoras podem ser obtidas (Moore & Wilson, 1992; Moore et al., 1993; Riley et al., 1999; Shary et al., 2002). Mesmo em um ambiente com relativamente poucos dados disponíveis, pode ser definido um número quase infinito de variáveis preditoras, as quais caracterizam os processos fundamentais de formação do solo, (Mora-Vallejo et al., 2008). No entanto, o uso de preditores ambientais adequados pode ser mais importante que o método de predição (Mackenzie & Ryan, 1999).

A utilização de grande quantidade de variáveis, devido ao fato de se desconhecer em muitos casos os atributos mais relevantes, pode dificultar a obtenção de modelos consistentes. Atributos redundantes e irrelevantes têm efeitos adversos sobre a mineração de dados, resultando em reduzida acurácia da classificação, e são, portanto, indesejáveis (Yang et al., 2011). Há benefícios potenciais na seleção de atributos, tais como: facilita a visualização e compreensão dos dados; reduz os requisitos de armazenamento e o tempo de treinamento, a dimensionalidade dos dados, e; melhora o desempenho de predição (Guyon & Elisseeff, 2003).

Conforme Yang et al. (2011), técnicas de seleção de atributos têm sido utilizadas em vários domínios da ciência. No entanto, poucos trabalhos sobre mapeamento digital de solos os têm utilizado (Coelho & Giasson, 2010). Isto pode ser devido ao reduzido número de variáveis utilizadas, o que torna menos relevante dos algoritmos de seleção. Nesse sentido, na tentativa de incorporar nos modelos um maior número de variáveis relacionadas aos diferentes fatores de formação do solo, torna-se necessário decidir

dentro da grande gama de preditores possíveis de serem gerados, por aqueles de maior e melhor potencial preditivo (Hengl et al., 2007).

Ao se considerar esses aspectos, os principais objetivos deste estudo foram: (1) identificar, entre diversas variáveis, com a utilização de diferentes métodos de seleção de atributos, qual o conjunto de variáveis é mais relevante para uma predição mais exata das classes de solo da bacia do rio Muqui do Norte, ES; (2) verificar se as variáveis climáticas e do balanço hídrico são úteis para tornar o processo de predição de classes de solo mais efetivo e; (3) avaliar o desempenho de predição com a utilização de atributos climáticos e do balanço hídrico em comparação aos atributos do terreno.

3.4. MATERIAL E MÉTODOS

3.4.1. Caracterização da área

O estudo foi desenvolvido na bacia hidrográfica do rio Muqui do Norte, afluente do rio Itapemirim, no sul do Estado do Espírito Santo. A área da bacia é de 683,90km² e situa-se entre as coordenadas UTM 7.666.276 e 7.693.910 m N e 248.742 e 309.050 m E, zona 24S (Figura 1). A área é representativa da região montanhosa da costa atlântica brasileira e apresenta uma grande diversidade geomorfológica (Quadro 1), geológica (Quadro 2) e de solos, além de altitudes que variam desde próximo ao Oceano Atlântico a mais de 1.100 m. O clima é do tipo Aw, com estação seca no inverno e verão quente e chuvoso (Köppen & Geiger, 1928). A fim de englobar uma área mais representativa da classe dos Chernossolos, foi acrescentado um “buffer” de 4 km à área da bacia, que corresponde a 690 km² em torno da bacia.

Quadro 1. Unidades geomorfológicas da área de estudo da bacia do rio Muqui do Norte, no Estado do Espírito Santo

FID*	Unidades geomorfológicas
0	Acumulação Fluvial
1	Maçãos do Caparaó
2	Tabuleiros Costeiros
3	Planícies Costeiras, Complexos Deltaicos, Estuarinos e Praias
4	Patamares Escalonados do Sul Capixaba
5	Colinas e Maçãos Costeiros

*FID – Identificador na tabela de atributos das imagens no formato raster.

Fonte: IJSN (2012).

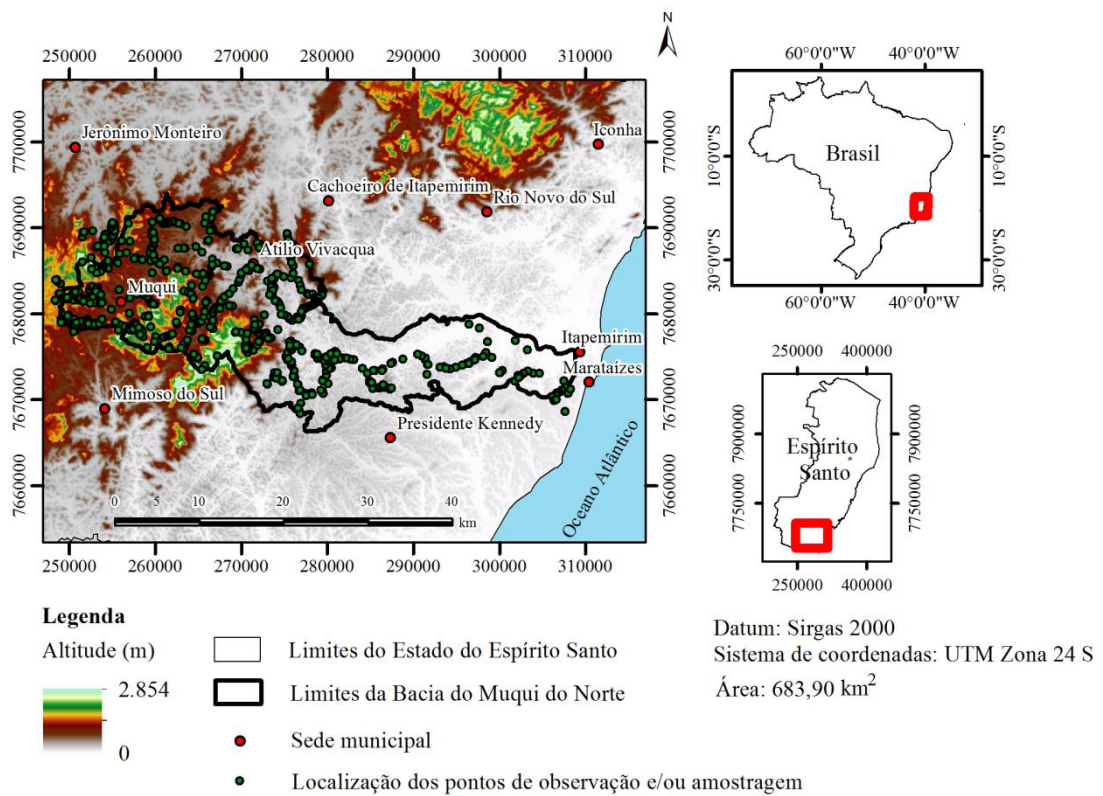


Figura 1. Localização e altimetria da bacia do rio Muqui do Norte no Estado do Espírito Santo.

Quadro 2. Litologias da área de estudo da bacia do rio Muqui do Norte, no Estado do Espírito Santo

FID*	Litologia
0	Dunas, restingas e cordões litorâneos
1	Arenitos e sedimentos areno-argilosos com laterização
2	Gnaisses granatíferos e kingizitos
3	Areias, cascalhos e argilas inconsolidadas
4	Hornblenda gnáisses, anfíbolitos, silimanita-granada-biotita gnáisses
5	Gnaisses bandados, milonitos e migmatitos
6	Granitos pórfiros grosseiros, sienitos, rochas básicas e intermediárias
7	Granitos sintectônicos com domínios restritos migmáticos
8	Gnaisses granitóides de composição granítica a tonalítica
9	Anfíbolitos
10	Granitos pós-tectônicos em geral cinza e de granulação média a fina
11	Rochas enderbíticas a charnockíticas
12	Rochas metabásicas

*FID – Identificador na tabela de atributos das imagens no formato raster.

Fonte: Brasil (1983).

3.4.2. Variáveis discriminantes

Utilizaram-se quarenta e cinco variáveis discriminantes no mapeamento digital de solos. Dentre elas, 28 atributos do terreno, 1 geomorfológica, 2 geológicas, 7 climáticas, 4 do balanço hídrico e 3 derivadas de dados do Landsat 5 TM (Quadro 3). Os atributos do terreno foram derivados de modelo digital de elevação (MDE), com 30 m de resolução espacial, obtido com a opção “topo to raster” do programa ArcGIS 10.0 (ESRI, 2011), gerado utilizando as curvas de nível equidistantes de 20 metros e a hidrografia, obtidas das cartas topográficas do IBGE, na escala de 1:50.000, disponível em www.geobases.es.gov.es.

Quadro 3. Preditores ambientais utilizados no mapeamento digital de solos da bacia do rio Muqui do Norte, ES

Variável e unidade	Fator ambiental*	Referência	Tipo	Média/ intervalo
1. Compacidade da base da elevação	Relevo	a	Q ¹	1,03 - 2,24
2. Fator forma da base da elevação	Relevo	a	Q ¹	0,25 – 1,00
3. Área basal das elevações (km ²)	Relevo		Q ¹	0,00 - 8,60
4. Amplitude altimétrica(m)	Relevo		Q ¹	0 – 1,046
5. Altitude (m)	Relevo		Q ¹	0 – 1,167
6. Distância à drenagem (m)	Posição espacial		Q ¹	0 – 1.754,18
7. Segunda derivada da declividade	Relevo		Q ¹	0 - 38,98
8. Razão de elevação do relevo	Relevo	b	Q ¹	0 - 0,89
9. Dissecação de Martonne	Relevo	c	Q ¹	0 – 1
10. Índice de Rugosidade do terreno (m)	Relevo	d	Q ¹	0 - 6,68
11. Índice de corrente máximo de fluxo	Relevo	e	Q ¹	0,58 - 11,91
12. Índice topográfico combinado	Relevo	e	Q ¹	2,91 - 30,76
13. Índice capacidade de transporte de sedimentos	Relevo	f	Q ¹	0 - 190
14. Temperatura média anual (°C)	Clima	g	Q ⁴	17,4 - 25,6
15. Temperatura média das mínimas anual (°C)	Clima	g	Q ⁴	12,1 - 20,3
16. Temperatura média das máximas anual (°C)	Clima	g	Q ⁴	22,7 - 30,8
17. Precipitação pluvial média anual (mm)	Clima		Q ⁴	1.029 – 1.428
18. Precipitação pluvial do período úmido (mm)	Clima		Q ⁴	765 – 1,223
19. Precipitação pluvial do período seco (mm)	Clima		Q ⁴	192,7 - 295,0
20. NDVI – Normalized difference vegetation index	Organismo	h	Q ⁵	0,92 - 0,81
21. CMI – “Clay mineral index”	Propriedade do solo	i	Q ⁵	0 - 72
22. IOI – “Iron oxide index”	Propriedade do solo	i	Q ⁵	0,06 - 1,26
23. Geologia	Material origem	j	C ²	13

Continuação do Quadro 3.

Variável e unidade	Fator ambiental*	Referência	Tipo	Média/ intervalo
24. Radiação solar (MWh m ⁻²)	Clima	k	Q ⁴	0,57 - 1,95
25. Aspecto (°)	Relevo	l	C ¹	8
26. Declividade (°)	Relevo	l	Q ¹	0,00 - 42,60
27. Densidade de drenagem (km km ⁻²)	Posição espacial	a	Q ¹	0,47 - 5,10
28. Distância mais próxima do mar (m)	Posição espacial		Q ¹	0,01 – 61,726
29. Distância às linhas de falha (m)	Posição espacial		Q ²	0,002 – 18,252
30. Geomorfologia	Relevo	m	C ³	6
31. “Northness”	Relevo	n	Q ¹	-1 - 1
32. “Eastness”	Relevo	n	Q ¹	-1 - 1
33. Área com curvatura retilínea (%)	Relevo		Q ¹	0 - 100
34. Área com curvatura côncava (%)	Relevo		Q ¹	0 -100
35. Área de topo (%)	Relevo		Q ¹	0 - 100
36. Área de topo + área convexa (%)	Relevo		Q ¹	0 -100
37. Topo, convexa, côncava e terraço	Relevo		C ¹	4
38. Curvatura (° m ⁻¹)	Relevo	o	C ¹	3
39. Perfil de curvatura (° m ⁻¹)	Relevo	o	C ¹	3
40. Plano de curvatura (° m ⁻¹)	Relevo	o	C ¹	3
41. Deficiência hídrica (mm)	Clima		Q ⁶	84 - 380
42. Evapotranspiração potencial (mm)	Clima		Q ⁶	1.225 – 1.461
43. Excedente hídrico (mm)	Clima		Q ⁶	56 - 191
44. Evapotranspiração real (mm)	Clima		Q ⁶	1.092 – 1.248
45. Altitude relativa à drenagem (m)	Clima		Q ¹	0 – 1.028

C – Classe; Q – quantitativo. ¹ atributo do terreno; ² variável geológica; ³ variável geomorfológica; ⁴ variável climática; ⁵ variável derivada de dados do Landsat-5; ⁶ variável do balanço hídrico. a - Villela & Mattos (1975); b - Pike & Wilson (1971); c - Evans (1972); d - Riley et al. (1999); e - Moore et al. (1993); f - Moore & Wilson (1992); g - Castro (2010); h - Rouse et al. (1974); i - Sabins (1987); j - Brasil (1983); k - Fu & Rich (2002); l - Burrough & McDonell (1998); m - IJNS (2012); n - Alexander et al. (2006); o - Moore et al. (1991). * Fator ambiental que melhor caracteriza as variáveis, considerando-se o modelo scorpan de McBratney et al. (2003), em que s = propriedade do solo em um ponto; c = clima; o = organismos; r = topografia; p = material originário; a = idade; n = posição espacial.

Os atributos geomorfológicos constaram do Mapa de Unidades Geomorfológicas do Estado do Espírito Santo na escala 1:250.000 (IJSN, 2012) e de mapa pedogeomorfológico, gerado a partir do MDE, com a delimitação das áreas de topo, encostas côncavas, encostas convexas e terraços.

As unidades litológicas e as linhas de falhas foram obtidas, respectivamente, dos mapas de Geologia do Projeto RADAMBRASIL (Brasil, 1983) e do geológico da CPRM (2012), ambos na escala 1:1.000.000.

As áreas basais das elevações foram delimitadas com base no delineamento das bacias hidrográficas com ferramenta “basin” do Arcgis 10, a partir do raster de direção de fluxo gerado do MDE com valores invertidos (multiplicados por -1). O raster com os limites das bases das elevações (bacias invertidas) foi convertido para o formato de arquivo tipo vetor (shapefile) e realizada a eliminação dos polígonos com áreas muito pequenas, inferiores à 0,0576 ha. Em seguida, com a utilização da ferramenta “Minimum Bounding Geometry”, foram criados retângulos de menores larguras envolvendo os polígonos de entrada, ou seja, os polígonos que representam as bases das elevações. Com essa ferramenta, um novo arquivo shape do tipo polígono (retângulos) foi criado, contendo informações de suas medidas como comprimento e largura. A medida comprimento e área da base da elevação foram utilizadas para o cálculo do fator de forma - Kf e o coeficiente de compacidade - Kc. Para o cálculo da densidade de drenagem, utilizou-se o comprimento dos canais de drenagem do IBGE e as áreas das bases das elevações.

A amplitude altimétrica - Δa foi calculada da seguinte forma: $\Delta a = h_{\max} - h_{\min}$, onde h_{\max} = altitude máxima relativa a cada polígono referente à base de elevação e; h_{\min} = altitude mínima relativa a cada polígono referente à base de elevação. A altitude relativa à drenagem de cada base de elevação foi obtida pela subtração da altitude mínima de cada base pelos valores de altitude do MDE.

Outras variáveis relacionadas aos polígonos das bases das elevações foram o mapa pedogeomorfológico, as áreas de topo, de topo + encostas convexas, de encostas côncavas e de curvatura retilínea.

Para elaborar o mapa pedogeomorfológico, a área de estudo foi dividida em seis subáreas com diferentes classes de amplitudes altimétricas: 0 a 50 m (muito baixa); 50,1 a 100 m (baixa); 100,1 a 250 m (média); 250,1 a 500 m (alta); 500,1 a 750 m (muito alta); 750,1 a 1.028,2 m (extremamente alta). Este procedimento foi necessário para melhor definir os diferentes limites, para cada subárea, dos segmentos planos e suaves ondulados (0 a 10% de declividade) da paisagem que separam topos e terraços. Os limites de separação entre topos e terraços foram estabelecidos com base em diferentes altitudes relativas que melhor se ajustaram visualmente às curvas de nível. Para as subáreas de muito baixa, baixa, média, alta, muito alta e extremamente alta, os valores considerados como limiares entre topos e terraços foram respectivamente de 25 m, 40 m, 85 m, 150 m, 250 m e 150 m. Nas áreas com declividade superior a 10% constaram

os valores de curvatura. Dessa forma, foi possível estratificar pedogeomorfologicamente a bacia em topos e terraços com declividades entre 0 e 10% e em encostas côncavas e convexas com declividades superiores a 10%.

As superfícies com informações de distância à drenagem, distância mais próxima do mar e distância às linhas de falha foram obtidas no Arcgis 10.0 pelo cálculo da distância euclidiana. As feições de interesse mais próximas contidas em arquivos no formato “shapefile” foram: drenagem da carta 1:50.000 do IBGE, linha de costa obtida das malhas municipais do IBGE e linhas de falhas digitalizadas do mapa geológico 1:1.000.000 do Radambrasil (Brasil, 1983).

Dados do sensor TM do Landsat-5 (maio de 2010) foram utilizados para derivar os índices minerais de argila (“clay mineral index”, CMI), óxido de ferro (“iron oxide index”, IOI) e vegetação por diferença normalizada, definido como NDVI (“normalized difference vegetation index”) = $(IVP - Vis)/(IVP + Vis)$, onde IVP é a intensidade de reflectância no infravermelho próximo e Vis a intensidade de reflectância do vermelho no visível. O CMI e o IOI foram obtidos de acordo com Sabins (1987). O primeiro por meio da divisão da banda 5 (1,55–1,75 μm) pela banda 7 (2,08–2,35 μm) e o segundo por meio da divisão da banda 3 (0,63–0,69 μm) pela banda 1 (0,45–0,52 μm). Os três índices foram obtidos com a utilização do ArcGIS 10.0 (ESRI, 2011).

Os dados climáticos utilizados como variáveis discriminantes foram: precipitação pluvial, temperaturas médias das mínimas, médias e máximas do ar e radiação solar global. A precipitação para a bacia foi obtida por meio de cokrigagem dos dados dos postos pluviométricos localizados no Estado e áreas vizinhas, obtidos no sítio <http://hidroweb.ana.gov.br/> e junto ao Instituto Capixaba de Pesquisa e Extensão Rural - INCAPER. Para o período úmido (outubro a abril) e anual a variável auxiliar foi a altitude e para o período seco (maio a setembro) a altitude mais a distância do mar.

As temperaturas médias do ar foram obtidas por meio de modelagem a partir de dados obtidos em estações meteorológicas distribuídas no Estado. As equações de regressão foram desenvolvidas por Castro et al. (2010) para permitir estimativas de temperatura em função da altitude, latitude e longitude.

A partir dos dados de precipitação e da estimativa da temperatura média do ar para todos os locais dos postos, calculou-se os elementos do balanço hídrico: a deficiência hídrica, evapotranspiração potencial, evapotranspiração real e excedente hídrico, pelo método proposto por Thornthwaite & Mather (1955). Os cálculos foram

realizados com o auxílio do programa "BHnorm", elaborado em planilha Excel por Rolim et al. (1998), assumindo-se uma Capacidade Máxima de Armazenamento de água no solo igual a 100 mm.

A quantificação da variável radiação solar global foi realizada por meio da ferramenta de análise "área solar radiation" disponível no software ArcGis 10, a qual considera como dados de entrada o MDE.

Os valores das variáveis intervalares foram reescalados entre o intervalo 0 e 1. Este procedimento foi realizado com a ferramenta "raster calculator" do ArcGis 10.0, utilizando a expressão $x' = (x - x_{min}) / (x_{max} - x_{min})$, onde x é o valor da variável, x_{min} e x_{max} são respectivamente o menor e o maior valor que esta variável assume, e x' o novo valor de x , normalizado. Nas variáveis com distribuição simétrica que apresentaram valores excessivamente elevados em um número muito reduzido de células (ruídos e outliers), os mesmos foram substituídos pelos valores correspondentes à média aritmética \pm três vezes o desvio-padrão. Para as variáveis com distribuição assimétrica os valores foram substituídos por aqueles correspondentes a 0,5 % ou 1% dos maiores valores, no caso de assimetria positiva.

3.4.3. Seleção das variáveis discriminantes

No ArcGis 10.0, o arquivo de ponto no formato "shapefile", referente aos pontos de observação, foi utilizado para obter os valores das variáveis discriminantes dos arquivos no formato raster. A tabela de atributos foi exportada para o formato "dbf" e convertida para ".arff", cujo formato é requerido para análises no software de mineração de dados denominado "WEKA", Versão 3.6.8 (Hall et al., 2009).

No software WEKA foi realizado o processo de seleção de variáveis de modo a se obter subconjuntos de variáveis de maior poder discriminante das classes de solo. Foram avaliados quatro algoritmos de seleção de atributos. Os dois primeiros avaliam o valor de subconjuntos de variáveis, quais sejam: "CFSSubsetEval" – CFS (Hall, 1998) e "ConsistencySubsetEval" – CSE (Liu & Setiono, 1996). Os outros métodos utilizados foram o "InfogainAttributeEval" – IA (Hall e Holmes, 2003) e o "ReliefAttributeEval" – RA (Kononenko, 1994), os quais avaliam variáveis individualmente e gera um lista em ranking. Os métodos de busca testados para os dois primeiros algoritmos de seleção foram dois: Bestfirst D1-N5 e SubsetSizeForwardSelection – I – K50 – T0 – F5 – S1 –

Z false – E. Para os algoritmos avaliadores de atributos individualmente, utilizou-se o método de busca denominado “Ranker”. Para estes algoritmos “Infogain” e “ReliefF”, os números de atributos mantidos nos subconjuntos foram de 12, 22 e 25, a fim de facilitar a comparação com os subconjuntos gerados pelos métodos “CFSSubsetEval” e “ConsistencySubsetEval”. Os métodos de busca foram utilizados com parâmetros padrão do software e a avaliação foi realizada sobre os conjuntos de dados de treinamento.

Avaliaram-se também os desempenhos de grupos compostos exclusivamente de variáveis climáticas em comparação a grupos de variáveis referentes aos atributos do terreno e combinações destas duas variáveis com outras, como geologia, geomorfologia e dados do Landsat-5.

A efetividade da classificação com as variáveis discriminantes selecionadas foi avaliada com o algoritmo de aprendizagem “SimpleCart”, que é um método de classificação por árvore de decisão baseado em Breiman et al. (1984), disponível no software WEKA. Os parâmetros para realizar a classificação foram aqueles fornecidos como padrão do software. A avaliação foi realizada por meio da validação cruzada, considerando o número de “folds” (subconjuntos de dados) igual a 10, onde cada um destes “folds” é utilizado para teste e o restante para treinamento.

A validação cruzada foi utilizada para avaliação das previsões em razão dos levantamentos de solos convencionais existentes na área estarem em escalas menores que 1:400.000, escalas incompatíveis com o nível de detalhe exigido (entre 1:100.000 e 1:50.000), para comparação com os mapas digitais produzidos pelos diferentes conjuntos de variáveis.

A avaliação dos resultados da classificação foi realizada com 598 instâncias de treinamento, distribuídas entre nove classes de solo e uma de afloramento de rocha (Quadro 4). Para tanto, foram comparados os percentuais de instâncias classificadas corretamente e os valores da estatística Kappa, que verifica, inicialmente, se a classificação difere de uma classificação casual. A significância dos valores dos índices Kappa foi verificada pelo teste estatístico Z (Congalton & Green, 2009).

Quadro 4. Classes de solos identificadas na bacia do rio Muqui do Norte, Espírito Santo, e número de instâncias para treinamento dos modelos por árvore de decisão

Unidade de mapeamento	Classe de solo observada	Quantidade de amostras
LVA	LATOSSOLO VERMELHO-AMARELO	181
PV	ARGISSOLO VERMELHO	78
PVA	ARGISSOLO VERMELHO-AMARELO	45
PA	ARGISSOLO AMARELO	38
MX	CHERNOSSOLO ARGILÚVICO	34
CX	CAMBISSOLO HÁPLICO	46
GX	GLEISSOLO HÁPLICO	41
RY	NEOSSOLO FLÚVICO	18
RL	NEOSSOLO LITÓLICO	59
AR*		58
Total		598

*Afloramento rochoso.

3.4.4. Trabalhos de campo

A área da bacia foi percorrida para identificação das classes de solos. Em 598 locais foram identificadas as classes de solos a partir de observações de perfis em barrancos de cortes de estrada e por meio da utilização de ortofotomosaico com resolução espacial de 1 metro (IEMA, 2011), em locais de difícil acesso. As coordenadas dos locais foram obtidas com utilização de GPS de navegação (modelo Garmin GPSMAP 76). Dois perfis foram descritos e amostrados conforme Santos et al. (2005) (um NEOSSOLO FLÚVICO e um GLEISSOLO HÁPLICO) e em outros três locais foram coletadas amostras nos horizontes A e B (um LATOSSOLO VERMELHO-AMARELO e dois ARGISSOLOS VERMELHOS) para análises físicas e químicas (dados não apresentados). Dados de outros seis perfis representativos da área foram obtidos da literatura (Embrapa, 1978; Ferreira Neto et al., 2009). Ao todo foram identificadas nove classes de solos (Quadro 4), as quais foram classificadas até o segundo nível categórico de acordo com o Sistema Brasileiro de Classificação (Embrapa, 2006).

3.5. RESULTADOS E DISCUSSÃO

3.5.1. Seleção de atributos

Os resultados (Quadro 5) mostram que, entre os algoritmos de seleção de subconjuntos (CFS e CSE), a maior redução de atributos foi obtida com a utilização do algoritmo CSE. Esse resultado está de acordo com a característica do método CSE. Segundo Hall & Holmes (2003), o CSE desenvolve uma avaliação do valor de um subconjunto de atributos por meio do nível de consistência nos valores da classe quando as instâncias de treinamento são projetadas sobre o subconjunto de atributos. Neste método, a seleção é realizada com o objetivo de encontrar o menor subconjunto de atributos com mesma capacidade de distinguir classes do conjunto completo de atributos. No caso do CFS (Hall, 1999), selecionam-se subconjuntos de atributos altamente correlacionados com a classe, mas os atributos têm baixa correlação entre eles.

Quanto aos métodos de busca, tanto para CFS quanto para CSE, foram encontradas diferenças no número de atributos selecionados. A redução do número de atributos foi mais eficaz, principalmente no caso da utilização do CFS, com “SubsetSizeForwardSelection”. Este método de busca reduziu em mais de 50% os atributos em relação ao método de busca “BestFirst”. Conforme Gutlein et al. (2009), em geral este método de busca produz menores subconjuntos de atributos sem reduzir a acurácia.

Os atributos selecionados pelo algoritmo CSE apresentaram coincidência menor que 50% em relação aos atributos selecionados pelo CFS. No caso dos algoritmos IAE e RAE, que elaboram ranking de atributos, verificou-se que os mesmos apresentaram maior similaridade quanto aos atributos selecionados. O IAE (Hall & Holmes, 2003) baseia-se no ganho de informação em relação à classe pelo conhecimento da presença ou ausência de um atributo. O RAE amostra aleatoriamente uma instância e então localiza seus vizinhos mais próximos da mesma e de diferentes classes. Os valores dos atributos dos vizinhos são comparados ao da instância amostrada e é calculada a relevância de cada atributo (Kononenko, 1994). Embora os algoritmos sejam bem diferentes, principalmente em termos de complexidade, a coincidência na seleção de atributos por meio do IAE e RAE, com rankings limitados a 24 atributos, foi em torno de 80%. As similaridades mais baixas nos atributos selecionados foram encontradas na comparação entre os algoritmos de seleção de subconjuntos (CSF e CSB) e os algoritmos de seleção de atributos individuais (IAE e RAE). Estes resultados indicam considerável divergência na seleção de atributos pelos diferentes métodos avaliados.

Quadro 5. Atributos selecionados pelos diferentes algoritmos de seleção e métodos de busca

Algoritmo de seleção	Método de busca	Atributos selecionados (un)	Atributos selecionados
CFSSubsetEval (CFS)	Bestfirst D1-N5	24	4, 5, 6, 10, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23, 24, 26, 27, 28, 33, 36, 41, 43, 44, 45
	SubsetSizeForwardSelection – I – K45 – T0 – F5 – S1 – Z false – E	10	4, 6, 10, 14, 20, 23, 24, 27, 41, 45
ConsistencySubsetEval (CSE)	Bestfirst D1-N5	10	4, 8, 10, 21, 22, 25, 27, 28, 41, 45
	SubsetSizeForwardSelection – I – K45 – T0 – F5 – S1 – Z false – E	8	4, 16, 22, 25, 27, 40, 41, 45
InfogainAttributeEval (IAE)	Ranker-T -1.7976931348623157E308 -N 24	24	28, 41, 5, 15, 14, 42, 16, 23, 27, 4, 43, 10, 26, 45, 18, 44, 13, 36, 29, 30, 19, 17, 37, 35
	Ranker-T -1.7976931348623157E308 -N 10	10	28, 41, 5, 15, 14, 42, 16, 23, 27, 4
	Ranker-T -1.7976931348623157E308 -N 8	8	28, 41, 5, 15, 14, 42, 16, 23
ReliefFAttributeEval (RAE)	Ranker-T -1.7976931348623157E308 -N 24	24	23, 30, 37, 26, 43, 39, 28, 42, 41, 4, 15, 5, 14, 10, 18, 16, 13, 19, 25, 40, 17, 45, 38, 44
	Ranker-T -1.7976931348623157E308 -N 10	10	23, 30, 37, 26, 43, 39, 28, 42, 41, 4
	Ranker-T -1.7976931348623157E308 -N 8	8	23, 30, 37, 26, 43, 39, 28, 42

4 – amplitude altimétrica; 5 – altitude; 6 – distância à drenagem; 8 – razão de elevação do relevo; 10 – índice de rugosidade do terreno; 12 – índice topográfico combinado; 13 – índice de capacidade de transporte de sedimentos; 14 – temperatura média anual; 15 – temperatura média das mínimas anual; 16 – temperatura média das máximas anual; 17 – precipitação pluvial média anual; 18 – precipitação pluvial do período úmido; 19 – precipitação pluvial do período seco; 20 – NDVI; 21 – CMI; 22 – IOI; 23 – geologia; 24 – radiação solar; 25 – aspecto; 26 – declividade; 27 – densidade de drenagem; 28 – distância do mar; 29 – distância às linhas de falha; 30 – geomorfologia; 33 – área com curvatura retilínea; 35 – área de topo; 36 - área de topo + área convexa; 37 – topo, convexo, côncavo e terraço; 38 – curvatura; 39 – perfil de curvatura; 40 – plano de curvatura; 41 – deficiência hídrica; 42 – evapotranspiração potencial; 43 – excedente hídrico; 44 – evapotranspiração real; 45 – altitude relativa à drenagem.

Entre os cinco atributos de maior importância há ocorrência de pelo menos uma variável climática ou do balanço hídrico. Isto é observado com os resultados da aplicação dos algoritmos RA e, especialmente o IAE (Quadro 5), o qual aponta as variáveis “distância do mar” (28), “deficiência hídrica” (41), “altitude” (5), “temperatura mínima” (15) e “temperatura média do ar” (14) como as cinco mais importantes. No entanto, conforme Bherens et al. (2010), o resultado mais interessante na avaliação sobre a seleção de atributos é a relação entre o número de atributos selecionados e o resultado da predição.

A importância dos atributos selecionados foi verificada em termos de exatidão da classificação (Quadro 6). Os resultados mostram não haver grande discrepância entre os valores do índice Kappa referentes às predições com os diferentes subconjuntos de variáveis. Mesmo com 50% de divergência nos atributos selecionados (Quadro 5), como no caso da comparação entre CSF10 e CSE10, não houve diferença significativa pelo teste Z a 5% de probabilidade. Isto decorre, provavelmente, em razão de que atributos considerados menos relevantes por um método podem melhorar o desempenho da classificação quando combinados com outros atributos para compor subconjuntos formados por outros métodos de seleção.

Quadro 6. Índice Kappa, Variância do Kappa (vâr), número de folhas (NF) e de nós (NN) obtidos com 44 atributos e com os diferentes subconjuntos de atributos, utilizando o algoritmo SimpleCart de árvore de decisão e avaliação por meio da validação cruzada

Grupo de atrib.	45 atrib.	CFS 24	IAE 24	RAE 24	CFS 10	IAE 10	RAE 10	CSE 10	CSE 8	IAE 8	RAE 8
NN	59	50	103	59	77	105	65	61	93	55	45
NF	30	99	52	30	39	53	33	31	47	28	23
Kappa	0,609	0,575	0,556	0,596	0,610	0,601	0,598	0,632	0,613	0,544	0,536
Var (x 10 ⁻³)	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
45 atrib.	0,00	-1,06	-1,67	-0,41	0,02	-0,24	-0,35	0,72	0,14	-2,02*	-2,25*
CFS24		0,00	-0,61	0,65	1,08	0,82	0,72	1,78	1,20	-0,96	-1,19
IAE24			0,00	1,26	1,68	1,42	1,32	2,39*	1,80	-0,36	-0,59
RAE24				0,00	0,43	0,17	0,07	1,13	0,55	-1,61	-1,84
CFS10					0,00	-0,26	-0,36	0,70	0,12	-2,03*	-2,26*
IAE10						0,00	-0,10	0,97	0,38	-1,78	-2,01*
RAE10							0,00	1,07	0,49	-1,68	-1,91
CSB8								0,00	-0,58	-2,74*	-2,97*
CSL8									0,00	-2,16*	-2,39*
IAE8										0,00	-0,23
RAE8											0,00

45 atrib. – todos os 45 atributos considerados no estudo; CFS24 – 24 atributos selecionados pelo CFSSubsetEval, Bestfirst; IAE24 – 24 atributos selecionados pelo InfogainAttributeEval, Ranker; RAE24 – 24 atributos selecionados pelo ReliefFAttributeEval, Ranker; CFS10 – 10 atributos selecionados pelo CFSSubsetEval, GreedyStepwise e SubsetSizeForwardSelection; IAE10 – 10 atributos selecionados pelo InfogainAttributeEval, Ranker; CSE10 – 10 atributos selecionados pelo ConsistencySubsetEval, Bestfirst; CSE8 – 8 atributos selecionados pelo ConsistencySubsetEval, SubsetSizeForwardSelection; IAE8 – 8 atributos selecionados pelo InfogainAttributeEval, Ranker; RAE8 – 8 atributos selecionados pelo ReliefFAttributeEval, Ranker. * significativo a 5% de probabilidade pelo teste Z.

Um dos objetivos da seleção de atributos está relacionado a uma redução no número de atributos, e visa aumentar ou manter os níveis de acurácia da predição em relação ao conjunto original de variáveis (Hall & Holmes, 2003). IAE8 e RAE8 foram dois dos três subconjuntos de variáveis preditoras ambientais que apresentaram menor

número de atributos, porém apresentaram os índices Kappa inferiores aos demais subconjuntos. O menor número de atributos selecionados sem perda de desempenho na classificação em relação ao conjunto com 45 atributos foi obtido com os subconjuntos CSE8, CFS10, CSE10, IAE10 e RAE10. Os índices Kappa desses subconjuntos apresentaram valores em torno de 0,600 indicando boa a muito boa qualidade dos modelos produzidos, de acordo com a classificação do valores do índice adotada por Landis & Koch (1977).

Não houve diferenças estatísticas entre CSE8, CFS10, CSE10, IAE10 e RAE10. No entanto, CSE10 e RAE10 foram os subconjuntos que apresentaram as árvores de decisão com os menores números de nós (tamanho da árvore) e de folhas (nó que indica uma classe). Estes são dois parâmetros indispensáveis para se avaliar quando são utilizados algoritmos de classificação por árvore de decisão. Conforme Bui & Moran (2003) árvores menores são desejáveis, pois o tamanho reflete a complexidade do modelo.

As árvores de decisão construídas pelo algoritmo SimpleCart, as quais geraram a matriz de significância do Kappa (Quadro 6) não utilizaram todas as variáveis contidas no grupo com 45 atributos e nem todas daqueles subgrupos que continham variáveis selecionadas pelos diferentes algoritmos de seleção. No Quadro 7 é apresentada uma melhor caracterização quanto ao número de variáveis, separadas por tipo. Os números totais de variáveis preditoras ambientais efetivamente utilizadas variaram de 6 (RAE8) a 21 (45atrib.) em razão da utilização da poda (minimal cost-complexity pruning) para o desenvolvimento das árvores de decisão.

Quadro 7. Atributos efetivamente utilizados na construção da árvore de decisão

Atributos	45 atrib.	CFS24	IAE24	RAE24	CFS10	IAE10	RAE10	CSE10	CSE8	IAE8	RAE8
NAUA	21	19	19	17	10	9	9	8	8	7	6
Terreno	9	7	7	6	5	4	4	5	5	2	2
Climático e do balanço hídrico	8	8	10	9	3	5	3	1	2	4	2
Geológico	1	0	1	1	0	0	1	0	0	0	1
Geomorfológico	1	1	1	1	1	0	1	0	0	1	1
Landsat 5	2	3	0	0	1	0	0	2	1	0	0

45 atrib. – todos os 45 atributos considerados no estudo; CFS24 – 24 atributos selecionados pelo CFSSubsetEval, Bestfirst; IAE24 – 24 atributos selecionados pelo InfogainAttributeEval, Ranker; RAE24 - 24 atributos selecionados pelo ReliefFAttributeEval, Ranker; CFS10 – 10 atributos selecionados pelo CFSSubsetEval, GreedyStepwise e SubsetSizeForwardSelection; IAE10 – 10 atributos selecionados pelo InfogainAttributeEval, Ranker; CSE10 - 10 atributos selecionados pelo ConsistencySubsetEval, Bestfirst; CSE8 – 8 atributos selecionados pelo ConsistencySubsetEval, SubsetSizeForwardSelection; IAE8 – 8 atributos selecionados pelo InfogainAttributeEval, Ranker; RAE8 - 8 atributos selecionados pelo ReliefFAttributeEval, Ranker. NAUA – número de atributos efetivamente utilizados para construir a árvore.

Verifica-se que variáveis climáticas ou do balanço hídrico foram utilizadas em todas as árvores de decisão. No RA10 as variáveis do balanço hídrico envolvidas foram “deficiência hídrica”, “evapotranspiração potencial” e “excedente hídrico”. Essas quatro variáveis estão entre as seis mais importantes entre as climáticas e do balanço hídrico, conforme apontado pelo método de seleção de atributos “InfogainAttributeEval” (Quadro 8). Essas variáveis climáticas ou do balanço hídrico ocorreram simultaneamente em outros subconjuntos de variáveis, mas o subconjunto que apresentou a melhor combinação de variáveis para gerar o mapa de classes de solos (Figura 2) foi o subconjunto RA10. O mapa elaborado a partir do CSE10 e a matriz de confusão (dados não mostrados) mostraram visível superestimação da classe dos Chernossolos, a qual é de fato mais restrita ao noroeste da bacia.

Quadro 8. Ranking de variáveis climáticas e do balanço hídrico obtidas pelos métodos de seleção de atributos Infogain e Relief

Método de seleção	Atributos selecionados
Infogain	Deficiência hídrica, temperatura mínima, temperatura média, evapotranspiração potencial, temperatura máxima, excedente hídrico, precipitação período úmido, evapotranspiração real, precipitação período seco, precipitação anual e radiação solar
ReliefF	Excedente hídrico, temperatura mínima, deficiência hídrica, temperatura média, precipitação período úmido, precipitação período seco, temperatura máxima, evapotranspiração potencial, precipitação anual, evapotranspiração real, radiação solar

Nota: ranking da esquerda para direita.

Outras variáveis do RA10 utilizadas na árvore de decisão (Figura 3), além daquelas do balanço hídrico, foram: “geologia”; “geomorfologia”; “toposseqüencia”; “declividade”; “distância do mar” e; “amplitude altimétrica. A amplitude altimétrica e a deficiência hídrica foram as únicas variáveis que ocorreram em todas as árvores de decisão que não apresentaram perda de desempenho em relação ao grupo inicial com 45 atributos. Isto evidencia uma maior relevância dessas duas variáveis no processo de mineração de dados com vistas à predição de classes de solo na área estudada. A bacia do rio Itapemirim, na qual se insere a bacia do rio Muqui do Norte, apresenta ambientes com ênfase em solos que estão mais ou menos dispostos paralelamente à linha de costa. Altitude, relevo, geologia e clima foram considerados por Lani et al. (2001) como elementos importantes para a sua estratificação com base em unidades de mapeamento de solos. Este fato sugere que a seleção das variáveis discriminantes ora apresentada neste trabalho está em linha com esses autores.

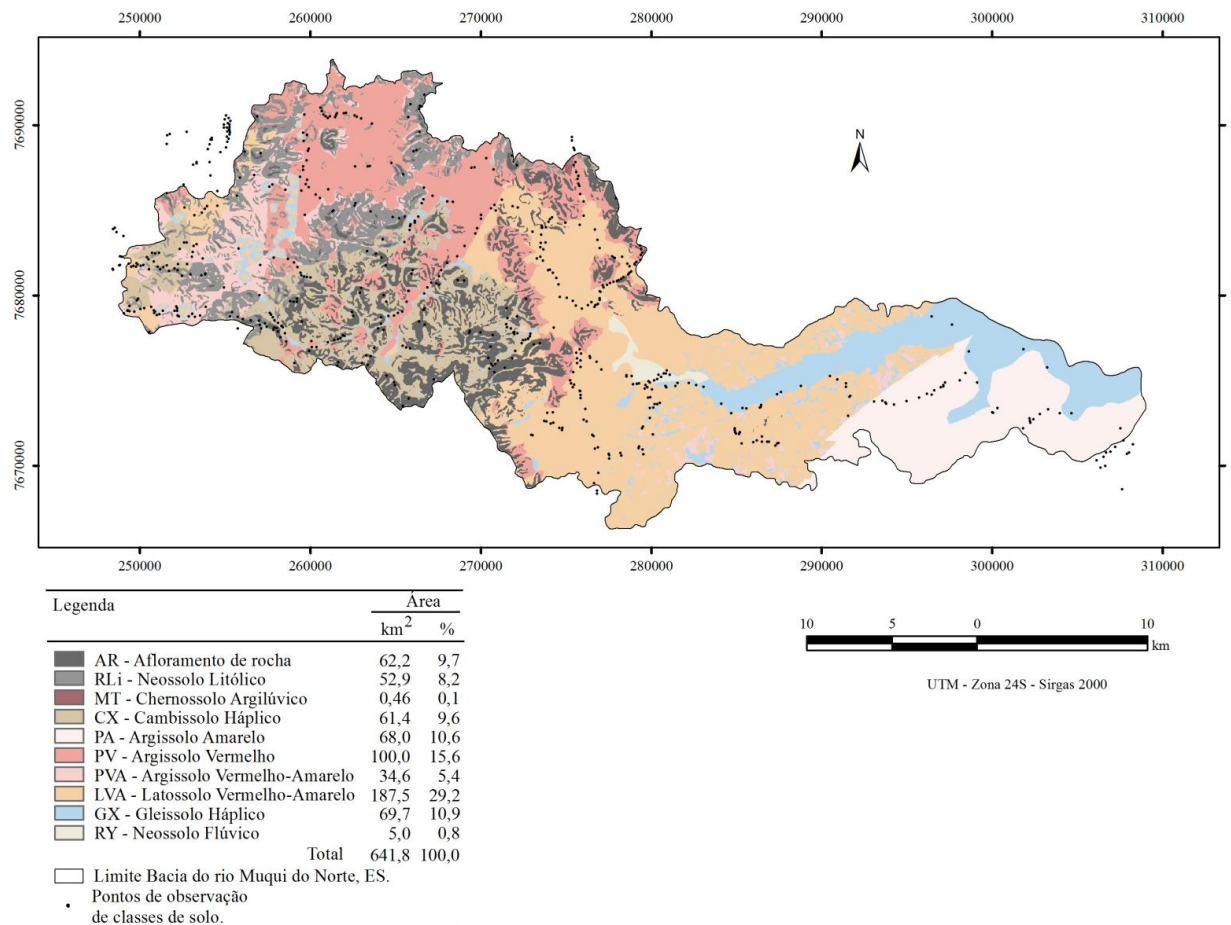


Figura 2. Mapa de solos elaborado por árvore de decisão a partir de dez atributos selecionados pelo algoritmo de seleção “ReliefAttributeEval” (geologia, geomorfologia, topossequencia, declividade, excedente hídrico, distância do mar, evapotranspiração potencial, deficiência hídrica e amplitude altimétrica).

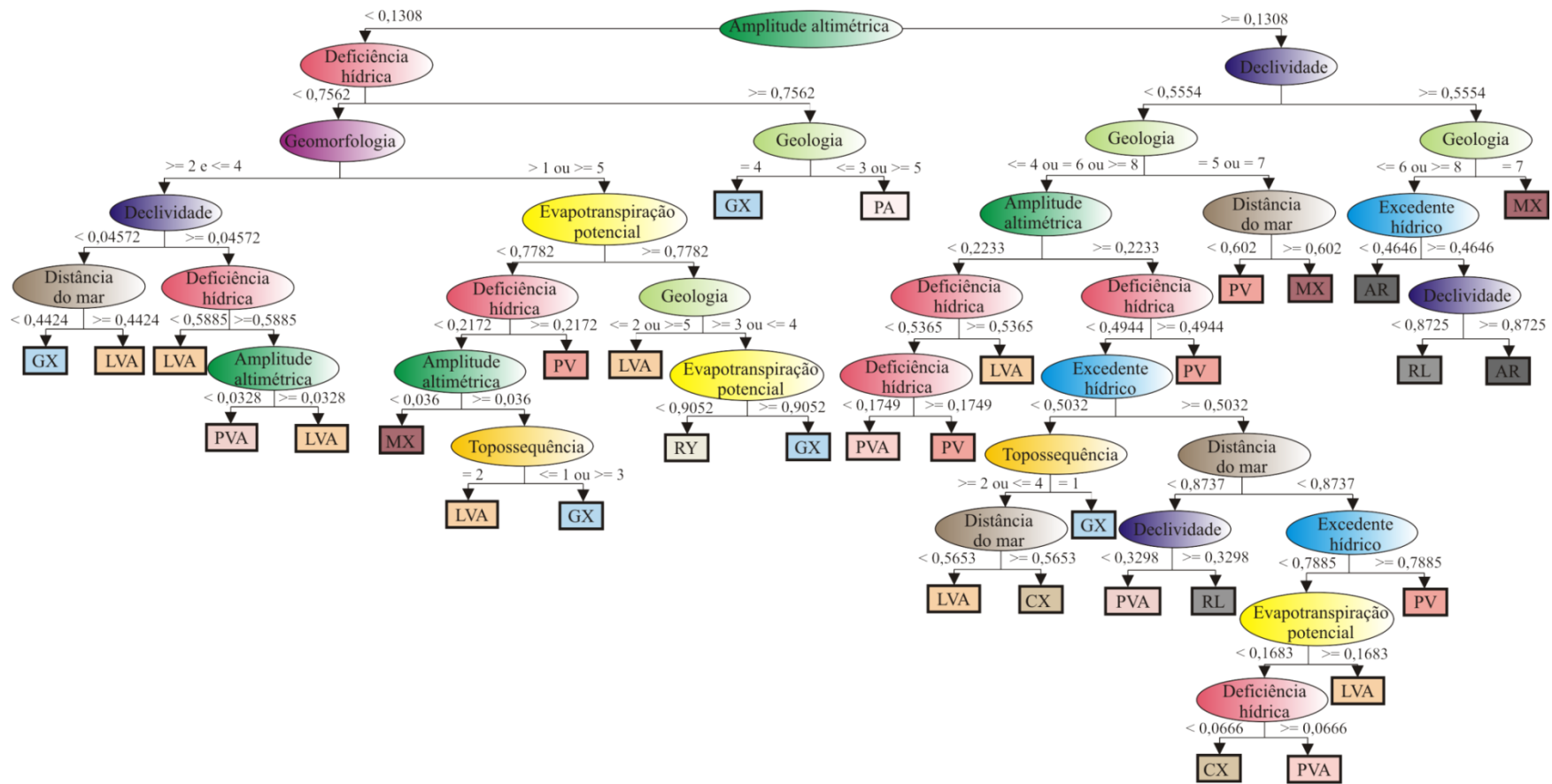


Figura 3. Árvore de decisão obtida pelo algoritmo SimpleCart. AR - Afloramento de rocha; RL – NEOSSOLO LITÓLICO; MT – CHERNOSSOLO ARGILÚVICO; CX – CAMBISSOLO HÁPLICO; PA – ARGISSOLO AMARELO; PV – ARGISSOLO VERMELHO; PVA – ARGISSOLO VERMELHO-AMARELO; LVA – LATOSSOLO VERMELHO-AMARELO; GX – GLEISSOLO HÁPLICO; RY – NESSOLO FLÚVICO.

3.5.2. Comparação da predição com variáveis climáticas, do balanço hídrico e do terreno

Embora haja influência de tempos pretéritos na gênese dos solos, o que poderia levar a uma baixa correlação entre as classes e as condições climáticas atuais (Ten Caten et al., 2012), nesta seção são apresentados e discutidos os resultados da comparação da exatidão da predição de classes de solos por diferentes subconjuntos de variáveis preditoras. A ênfase é mostrar, principalmente, o potencial de uso de variáveis climáticas no mapeamento digital de classes de solos.

A predição de classes de solo com a utilização exclusiva das variáveis preditoras do terreno (VT) apresentou o valor mais baixo do índice Kappa, que foi de 0,506 (Quadro 9). O uso das variáveis climáticas e do balanço hídrico isoladamente (VC) produziu Kappa com valor superior (0,554) ao proporcionado pelas variáveis do terreno, embora não tenha sido verificado diferença significativa a 5% de probabilidade entre estes dois conjuntos de variáveis, pelo teste Z. Os valores de Kappa com a utilização das variáveis climáticas em combinação com outras variáveis (VCL, VCGG e VCLGG) apresentaram-se também superiores aos dos subconjuntos de variáveis do terreno com as mesmas combinações (VTL, VTGG e VTLGG).

Quadro 9. Índice Kappa, Variância do Kappa (vâr), número de folhas e tamanho da árvore (nº de nós) obtidos com diferentes subconjuntos de atributos, utilizando o algoritmo SimpleCart de árvore de decisão e avaliação por meio da validação cruzada

Grupo de atributos	VT	VTL	VTGG	VTLGG	VC	VCL	VCGG	VCLGG	VTCLGG
Nós (nº)	47	69	101	65	43	39	67	35	65
Folhas (nº)	24	35	51	33	22	20	34	18	33
Kappa	0,506	0,568	0,537	0,575	0,554	0,632	0,615	0,619	0,625
vâr	0,00054	0,00054	0,00054	0,00053	0,00052	0,00049	0,00051	0,00050	0,00049
VT	0,00	1,89	0,93	2,10*	1,47	3,93*	3,35*	3,50*	3,71*
VTL		0,00	-0,95	0,21	-0,43	2,01*	1,44	1,58	1,79
VTGG			0,00	1,16	0,53	2,97*	2,40*	2,54*	2,75*
VTLGG				0,00	-0,64	1,80	1,23	1,37	1,58
VC					0,00	2,46*	1,89	2,03*	2,24*
VCL						0,00	-0,56	-0,43	-0,22
VCGG							0,00	0,13	0,34
VCLGG								0,00	0,21
VTCLGG									0,00

VT – variáveis do terreno; VTL – atributos do terreno + dados do Landsat5; VTGG – atributos do terreno + geologia + geomorfologia; VTLGG – atributos do terreno + dados do Landsat 5 + geologia + geomorfologia; VC – Atributos climáticos + balanço hídrico; VCL – atributos climáticos + balanço hídrico + dados do Landsat 5; VCGG – atributos climáticos + balanço hídrico + geologia + geomorfologia; VCLGG – atributos climáticos + balanço hídrico + dados do Landsat 5 + geologia + geomorfologia; VTCLGG – atributos do terreno + atributos climáticos + balanço hídrico + dados do Landsat 5 + geologia + geomorfologia.

Considerando-se uma avaliação visual dos mapas gerados a partir das variáveis do terreno e das variáveis climáticas e do balanço hídrico (Figura 4), verifica-se, no geral, uma significativa concordância entre os mesmos. Há, visualmente, uma maior concordância para as unidades de mapeamento G, PA e LVA, situadas respectivamente em áreas de relevo predominantemente plano, suave ondulado e ondulado a forte ondulado. Essa concordância (60,1%) pode ser facilmente observada no mapa de concordância entre o mapa de solos elaborado a partir de variáveis preditoras climáticas e o mapa de solos produzido a partir de variáveis do terreno (Figura 5).

Nas áreas mais dissecadas, da porção central até a porção oeste da bacia, os mapas são mais discordantes. Essa discordância pode ser explicada pela maior complexidade da distribuição espacial dos solos nessa porção da bacia, em razão da geologia e relevo mais complexos (Brasil, 1983).

Segundo Gonzalez et al. (2008), o relevo exerce maior influência sobre as características dos solos quando se considera pequenas áreas, ao passo que as variações em razão do clima são muito pequenas. Por outro lado, quando a área de estudo é mais extensa, o clima exerce influência mais discernível sobre a formação do solo, e tende a dominar o poder de influência dos atributos do terreno e geologia sobre os tipos de solo. Dessa maneira, nessas áreas, onde há maior diversidade de relevo, o mapa gerado com a utilização de variáveis do terreno mostrou uma maior capacidade de detalhamento espacial no delineamento das unidades de mapeamento. Isto sugere que apesar dos valores de Kappa terem sido em geral mais elevados para as predições realizadas com as variáveis climáticas, as variáveis do terreno são fundamentais para o mapeamento digital de áreas montanhosas, especialmente em áreas com alta variabilidade de relevo a curtas distâncias.

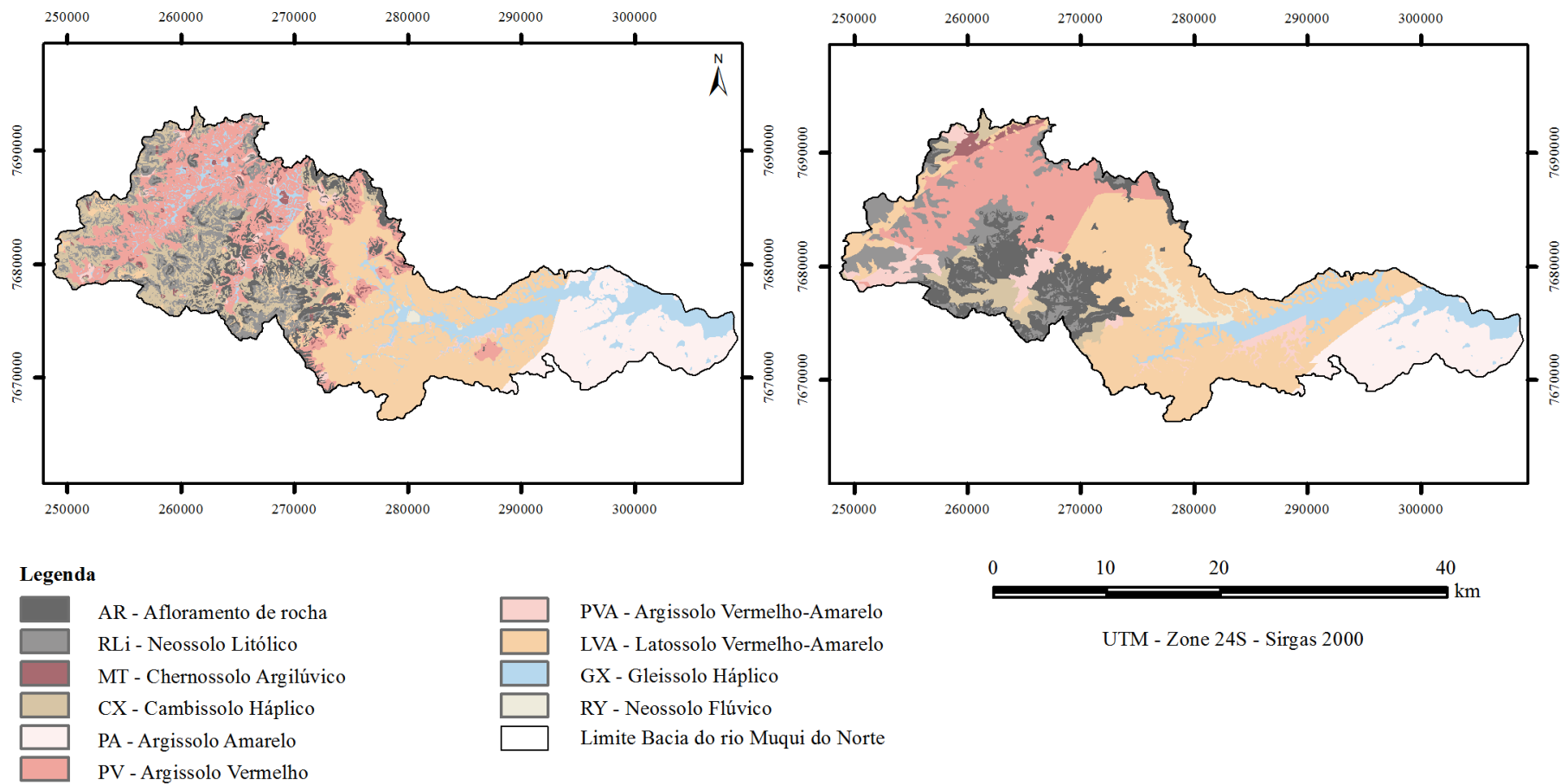


Figura 4. Mapas de solos elaborados por árvore de decisão a partir de variáveis do terreno (A) e de variáveis climáticas e do balanço hídrico (B) da bacia do rio Muqui do Norte, Espírito Santo.

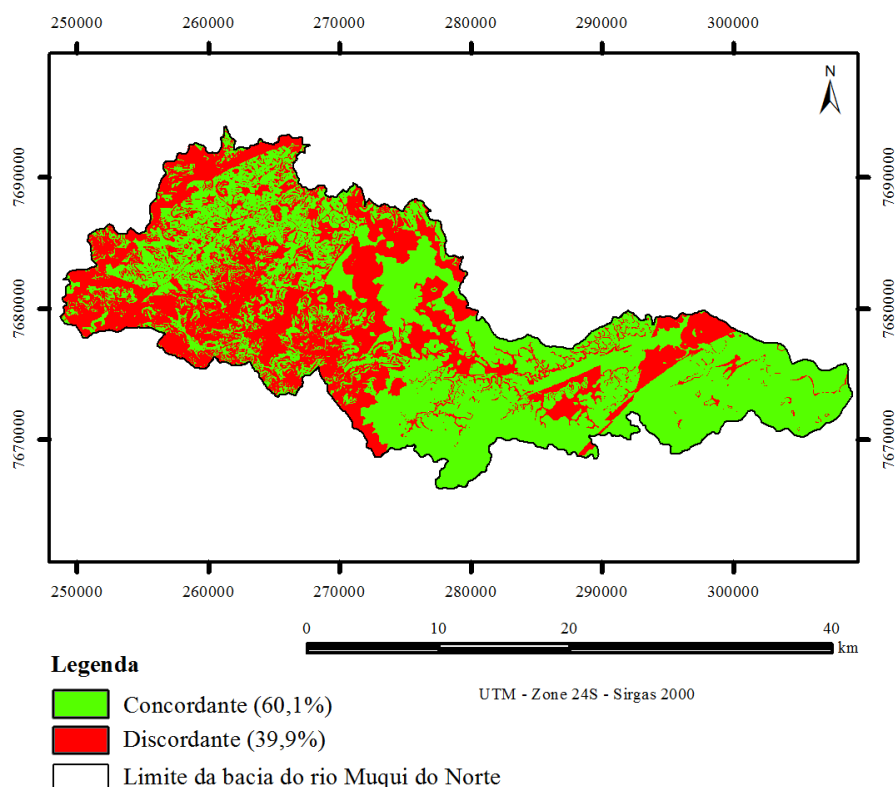


Figura 5. Comparação entre o mapa gerado a partir de variáveis climáticas com o gerado a partir de variáveis do terreno.

3.6. CONCLUSÕES

O método de seleção de atributos “ReliefAttributeEval” gerou o melhor conjunto de variáveis com base nos resultados da validação cruzada, matriz de confusão e tamanho da árvore de decisão. Apresentou, desta maneira, maior potencial de utilização em estudos de mapeamento digital de solos que envolvam elevado número de variáveis preditoras.

O subconjunto que se apresentou mais efetivo e eficiente envolveu a combinação de variáveis do terreno, geologia, geomorfologia e variáveis do balanço hídrico. Das nove variáveis, três foram do balanço hídrico, indicando a necessidade da utilização dessas variáveis em mapeamento digital de classes de solo com vista a alcançar maior exatidão da predição.

As variáveis climáticas e do balanço hídrico apresentaram desempenho preditivo equiparável às do terreno e com significante percentual de concordância entre os mapas produzidos, o que reforça a necessidade de se incluir esta variável em estudos de mapeamento digital de solos.

3.7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDER, S.M.; LOGAN, T.B. & PAQUET, P.C. Spatiotemporal co-occurrence of cougars (*Felisconcolor*), wolves (*Canis lupus*) and their prey during winter: a comparison of two analytical methods. *Journal of Biogeography*, 33: 2001–2012, 2006.
- ALKIMIM, A.F. Geoambientes, morfometria e solos da bacia do rio Benevente, ES. Viçosa, Universidade Federal de Viçosa, 2009. 109p. (Tese de Mestrado)
- BEHRENS, T.; ZHU, A.; SCHMIDT, K. & SCHOLTEN, T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155:175-185, 2010.
- BÖHNER, J. & SELIGE, T. Spatial Prediction of Soil Attributes Using Terrain Analysis and Climate Regionalisation. In: BÖHNER, J.; MCCLOY, K.R. & STROBL, J. **SAGA - Analysis and Modelling Applications**. Göttinger Geographische Abhandlungen, 115:13-27, 2006. Disponível em: http://ufpr.dl.sourceforge.net/project/saga-gis/SAGA%20-%20Documentation/GGA115 /gga115_02.pdf. Acesso em 18/02/2011.
- BRASIL, Ministério das Minas e Energia, Secretaria Geral, Folhas SF 23/24 Rio de Janeiro/Vitória: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra / Projeto RADAMBRASIL, Rio de Janeiro, 1983, 775p.
- BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A. & STONE, C.J. **Classification and regression trees**. Wadsworth, Monterey, CA, USA, 1984.
- BUI, E. N. & MORAN, C. J. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. *Geoderma*, 111:21-44, 2003.
- BURROUGH, P. A. & MCDONELL, R. A. **Principles of Geographical Information Systems**. New York, Oxford University Press, 1998, 190p.

- CASTRO, F. S.; PEZZOPANE, J.R.M.; CECÍLIO, R.A.; PEZZOPANE, J.E.M. Uso de imagens de radar na espacialização da temperatura do ar. **Idesia**, 28:69-79, 2010.
- COELHO, F. F. & GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, 40:2099-2106, 2010.
- CONGALTON, R.G. & GREEN, K. Assessing the accuracy of remotely sensed data: principles and practices. 2.ed. Boca Raton, Taylor and Francis, 2009.
- COMPANHIA DE PESQUISA DE RECURSOS MINERAIS - CPRM. **Mapa Geológico do Estado do Espírito Santo**. Disponível em: <http://www.cprm.gov.br/>, Acesso em 06 de fevereiro de 2012.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema brasileiro de classificação de solos**. 2. ed. Rio de Janeiro, 2006. 306p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA. Serviço Nacional de Levantamento e Conservação do Solo. **Levantamento de reconhecimento dos solos do Estado do Espírito Santo**. Rio de Janeiro, 1978, 461p. (Boletim Técnico)
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - ESRI, Software ArcGis Desktop, version 10, 2011.
- EVANS, I.S. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley, R.J. **Spatial Analysis in Geomorphology**. New York, Harper & Row, 1972.
- EVANS, J. Evans TNC Modeling Tools. ArcGIS - Geomorphometry and Gradient Metrics Toolbox. 2012. Disponível em: <http://conserveonline.org/workspaces/emt/documents/all.html>.
- FERREIRA NETO, J.A. & LANI, J.L. **Estudos visando o licenciamento ambiental de assentamentos rurais do Estado do Espírito Santo**. Projeto de Assentamento Monte Alegre. Viçosa, MG. NEPUT/UFV. 214p. 2009.
- FU, P. & RICH, P.M. A geometric solar radiation model with applications in agriculture and forestry. **Computers and Electronics in Agriculture**, 37:25-35, 2002.
- GONZALEZ, J.P.; JARVIS, A.; COOK, S.E.; OBERTHÜR, T.; RINCON-ROMERO, M. BAGNELL, J.A. & M. DIAS, B. Digital Soil Mapping of Soil Properties in Honduras Using Readily Available Biophysical Datasets and Gaussian Processes.

- In: HARTEMINK A.E., MC BRATNEY A. & MENDONÇA-SANTOS M.L. (Org.). **Digital soil mapping with limited data**. Amsterdam, Elsevier, 2008. p.367-380.
- GUTLEIN, M.; FRANK, E; HALL, M.; KARWATH, A. Large-scale attribute selection using wrappers, in: Proceedings of the IEEE Symposium Series on Computational Intelligence and Data Mining (CIDM-2009), 2009, pp. 332–339.
- GUYON, I. & ELISSEEFF, A. An introduction to variable and feature selection. **Journal of Machine Learning Research**, 3:1157–1182, 2003.
- GRUNWALD, S. Multi-criteria characterization of recent digital soil mapping and modeling approaches. **Geoderma**, 152:195-207, 2009.
- HALL, M. A. Correlation-based feature selection for machine learning. Hamilton, University of Waikato, 1998. (Tese de doutorado)
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P. & WITTEN, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, Volume 11, Issue 1, 2009.
- HALL, M. & HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. **IEEE Transactions on Knowledge and Data Engineering**, 15:1437–1447, 2003.
- HENGL, T.; TOOMANIAN, N.; REUTER, H.I. & MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, 140:417-427, 2007.
- INSTITUTO ESTADUAL DO MEIO AMBIENTE E RECURSOS HIDRICOS DO ESPÍRITO SANTO - IEAMA. Subgerência de geomática. Disponível em: <<http://www.meioambiente.es.gov.br/default.asp>>. Acesso em 06 jun. 2011.
- INSTITUTO JONES DOS SANTOS NEVES – IJSN. **Mapeamento geomorfológico do Estado do Espírito Santo**, Vitória, ES, 2012, 19p. (Nota técnica, 28)
- JENNY, H. Factors of Soil Formation. McGraw-Hill, New York. 1941.
- KONONENKO, I. Estimating attributes: Analysis and extensions of relief. In: Proceedings of the seventh European Conference on Machine Learning, pages 171-182. Springer-Verlag, 1994.
- KÖPPEN, W. & GEIGER, R. *Klimate der Erde*, Gotha: Verlag Justus Perthes, 1928,
- LANDIS J. & KOCH G. The measurement of observer agreement for categorical data. **Biometrics**, 33:159-174, 1977.

- LANI, J.L.; REZENDE, S.B. & RESENDE, M. Estratificação de ambientes com base nas classes de solos e outros atributos na bacia do rio Itapemirim, Espírito Santo. **Revista Ceres**, 48:239-261, 2001.
- LIU, H. & SETIONO, R., A probabilistic approach to feature selection: A filter solution," in Proceedings of the 13th International Conference on Machine Learning. 1996, 319-327, Morgan Kaufmann.
- MCBRATNEY, A. B.; MENDONÇA-SANTOS, M. L. & MINASNY, B. On digital soil mapping. **Geoderma**, 117:3–52, 2003.
- MCKENZIE, N.J. & RYAN, P.J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, 89:67–94, 1999.
- MORA-VALLEJO, A.; CLASSENS, L.; STOOORVOGEL, J. & HEUVELINK, G.B.M. Small scale digital soil mapping in southeastern Kenya. **Catena**, 76:44–53, 2008.
- MOORE, I.D.R.; GRAYSON, B. & LANDSON, A.R. Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. **Hydrological Processes**, 5:3–30, 1991.
- MOORE, I.D.; GESSLER, G.A. & PETERSON, G.A. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, 57:443-452, 1993.
- MOORE, I.D. & WILSON, J.P. Length–slope factors for the Revised Universal Soil Loss Equation: simplified method for estimation. **Journal of Soil and Water Conservation**, 47:423–428, 1992.
- PIKE, R.J. & WILSON S.E. Elevation relief ratio, hypsometric integral, and geomorphic area altitude analysis. **Geological Society of America Bulletin**. 82:1079-1084, 1971.
- RILEY, S.J.; DEGLORIA, S.D. & ELLIOT, R., A terrain ruggedness index that quantifies topographic heterogeneity. **Intermountain Journal of Sciences**, 5:23-27, 1999.
- ROLIM, G.S.; SENTELHAS, P.C. & BARBIERI, V. Planilhas no ambiente EXCEL para os cálculos de balanços hídricos: normal, sequencial, de cultura e de produtividade real e potencial. **Revista Brasileira de Agrometeorologia**, 6:133-137, 1998.
- ROUSE, J.W.J.; HAAS, R.H.; DEERING, D.W.; SCHELL, J.A. & HARLAN, J.C. Monitoring the vernal advancement and retrogradation (Green Wave Effect) of

- natural vegetation, NASA/GSFC, Type III Final Report, Greenbelt, MD, 1974. 371p.
- SABINS F. F. **Remote Sensing Principles and Interpretation**. 2ed., New York, W.H. Freeman & Co, 1987.
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H.C. **Manual de descrição e coleta de solo no campo**. 5 ed. revisada e ampliada, Viçosa, Sociedade Brasileira de Ciência do Solo, 2005. 100p.
- SHARY, P.A.; SHARAYAB, L.S. & MITUSOV, A.V. Fundamental quantitative methods of land surface analysis. **Geoderma**, 107:1–32, 2002.
- TEN CATEN, A.; DALMOLIN, R.S.D.; MENDONÇA-SANTOS, M.L.; GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, 42:1990-1997, 2012.
- THORNTHWAITE, C.W. An approach toward a rational classification of climate. **Geographical Review**, 38:55-94, 1948.
- VILLELA, S.M. & MATTOS, A. **Hidrologia aplicada**. São Paulo, McGraw-Hill do Brasil, 1975. 245p.
- YANG, Q.; SHAO, J.; SCHOLZ, M. & PLANT, C. Feature selection methods for characterizing and classifying adaptive Sustainable Flood Retention Basins. **Water Research**, 45:993–1004, 2011.

4. CAPÍTULO 3 - PREDIÇÃO DE CLASSES DE SOLO POR DIFERENTES ALGORITMOS E VARIÁVEIS PREDITIVAS AMBIENTAIS

4.1. RESUMO

Diferentes algoritmos de classificação têm apresentado desempenhos preditivos satisfatórios, mas ainda há carência de estudos que contribuam para a consolidação do mapeamento digital de solos. Dessa maneira, objetivou-se avaliar o desempenho dos algoritmos NaiveBayes, rede neural multilayer perceptron, SimpleCart e J48 na predição de classes de solos em uma sub-bacia hidrográfica no sul do Estado do Espírito Santo. Quarenta e cinco variáveis preditoras foram utilizadas, entre elas diferentes atributos do terreno, geologia, geomorfologia, clima e índices derivados das bandas 1, 3, 4 e 5 do sensor Landsat 5. Outro grupo com dez variáveis, selecionadas do total (45) pelo algoritmo ReliefF, foi avaliado. As predições foram avaliadas com base na validação cruzada e comparações dos mapas preditos com um mapa convencional de referência. A seleção de atributos foi eficaz apenas para as árvores de decisão. Os resultados da validação cruzada mostraram que os quatro algoritmos classificadores apresentaram desempenhos de predição semelhantes, com valores de Kappa que variaram de 0,66 a 0,75. Na comparação entre os mapas preditos e o mapa de referência, considerando-se os primeiros componentes das unidades de mapeamento, as concordâncias foram de 55%, 52%, 51% e 48% para os algoritmos multilayer perceptron, SimpleCart, J48 e NaiveBayes, respectivamente.

Termos para indexação: mapeamento digital de solos, redes neurais, árvore de decisão.

PREDICTION OF SOIL CLASSES BY DIFFERENT ALGORITHMS AND ENVIRONMENTAL PREDICTORS

4.2. ABSTRACT

Different classification algorithms have shown satisfactory predictive performances, but there is still a lack of studies that contribute to the consolidation of digital soil mapping. Thus, this study aimed to evaluate the performance of algorithms NaiveBayes, multilayer perceptron neural network, SimpleCart and J48 in predicting soil classes in a sub-basin in southern region of Espírito Santo State. Forty-five predictor variables were

used, including different attributes of terrain, geology, geomorphology, climate and indices derived from bands 1, 3, 4 and 5 of Landsat 5. Another group with ten variables, selected of the total (45) by ReliefF algorithm was evaluated. The predictions were evaluated based on cross-validation and comparisons of predicted maps with a conventional map reference. The feature selection was effective only for decision trees. The cross-validation results show that the four classifiers presented similar performance of prediction, with Kappa values ranging from 0.66 to 0.75. Comparing the predicted maps and map reference, considered the first components of the map units, the concordances were 55%, 52%, 51% and 48% for the multilayer perceptron, SimpleCart, J48 and NaiveBayes algorithms, respectively.

Index terms: digital soil mapping, neural networks, decision tree.

4.3. INTRODUÇÃO

Estudos sobre métodos de mapeamento digital de solos têm sido realizados em vários países. No Brasil, pesquisas nesse tema têm gerado crescente número de publicações científicas devido à importância desse tema na obtenção de informação mais detalhadas de solos (Ten Caten et al., 2012). As técnicas mais utilizadas nesses estudos fazem uso da mineração de dados que é considerada uma etapa no processo de obtenção do conhecimento de banco de dados (Han & Kamber, 2006). Classificadores bayesianos, árvores de decisão e redes neurais artificiais têm sido alguns dos modelos preditores frequentemente utilizados (McBratney et al., 2003; Ten Caten et al., 2012).

O classificador Bayesiano, provavelmente o mais utilizado em aplicações de mineração de dados, e o de árvores de decisão são algoritmos relativamente rápidos (Hall e Holmes, 2003). No mapeamento digital de solos estes dois algoritmos em geral têm alcançado boa acurácia (Coelho & Giasson, 2010). A vantagem da utilização de modelos de árvores de decisão está na facilidade de interpretar os resultados (Shatar & McBratney, 1999). O uso das redes neurais também tem possibilitado a produção de mapas de solos com elevada acurácia (Chagas et al, 2010). Embora forneçam melhor acurácia preditiva em muitas aplicações, as redes neurais não apresentam um modelo de fácil compreensão devido à complexidade da rede (Qi & Zhu, 2003). Apesar de estudos mostrarem que os classificadores Bayesianos, árvores de decisão e redes neurais sejam bons modelos preditores, raros trabalhos têm mostrado comparações entre os diferentes

algoritmos classificadores para o mapeamento de classes de solos (Coelho & Giasson, 2010; Sarmiento et al., 2012).

Estudos com predições de classes de solos com base em pontos de observação de campo (Chagas et al., 2010; Chagas et al., 2011) têm apresentado bons desempenhos dos modelos preditivos utilizados. Esses resultados têm ocorrido tanto na fase de treinamento dos modelos preditivos quanto na comparação com mapas de referências convencionais. Nos estudos de mapeamento digital de solos um maior número de algoritmos e de variáveis preditoras precisam ser comparados e validados com dados de observações de campo, de maneira a aproximar os mapas preditos com a realidade de campo (Sarmiento, 2010).

Nesse contexto, o objetivo deste trabalho foi avaliar o desempenho dos algoritmos classificadores NaiveBayes, rede neural multilayer perceptron e árvores de decisão, SimpleCart e J48, e comparar dois conjuntos de variáveis preditoras ambientais na predição de classes de solos na área da bacia do rio Muqui do Norte, sul do Estado do Espírito Santo.

4.4. MATERIAL E MÉTODOS

O estudo foi desenvolvido em uma área com 683,9 km², na bacia hidrográfica do rio Muqui do Norte, afluente do rio Itapemirim, no sul do Estado do Espírito Santo (Figura 1). Encontra-se entre as coordenadas UTM 7666276 e 7693910 m N e 248742 e 309050 m E, zona 24S. O clima é do tipo Aw, com estação seca no inverno e verão quente e chuvoso. A fim de englobar uma área mais representativa da classe dos Chernossolos, foi acrescentado um “buffer” de 4 km à área da bacia, que corresponde a 690 km² em torno da bacia.

O relevo da área é diverso e seis unidades geomorfológicas são encontradas: Planície Costeira e Colinas e Maciços Costeiros, ambas de pouca representatividade; Acumulação Fluvial; Tabuleiros Costeiros; Patamares Escalonados do Sul Capixaba e Maciços do Caparaó, em geral com altitudes mais elevadas (IJSN, 2012). Os solos são formados de diferentes materiais de origem, como sedimentos fluviais, sedimentos arenos-argilosos da Formação Barreiras e rochas graníticas, gnáissicas, enderbiticas, charnockiticas, hornblenda-gnáisse, anfibólitos e outras do Complexo Paraíba do Sul (Brasil, 1983).

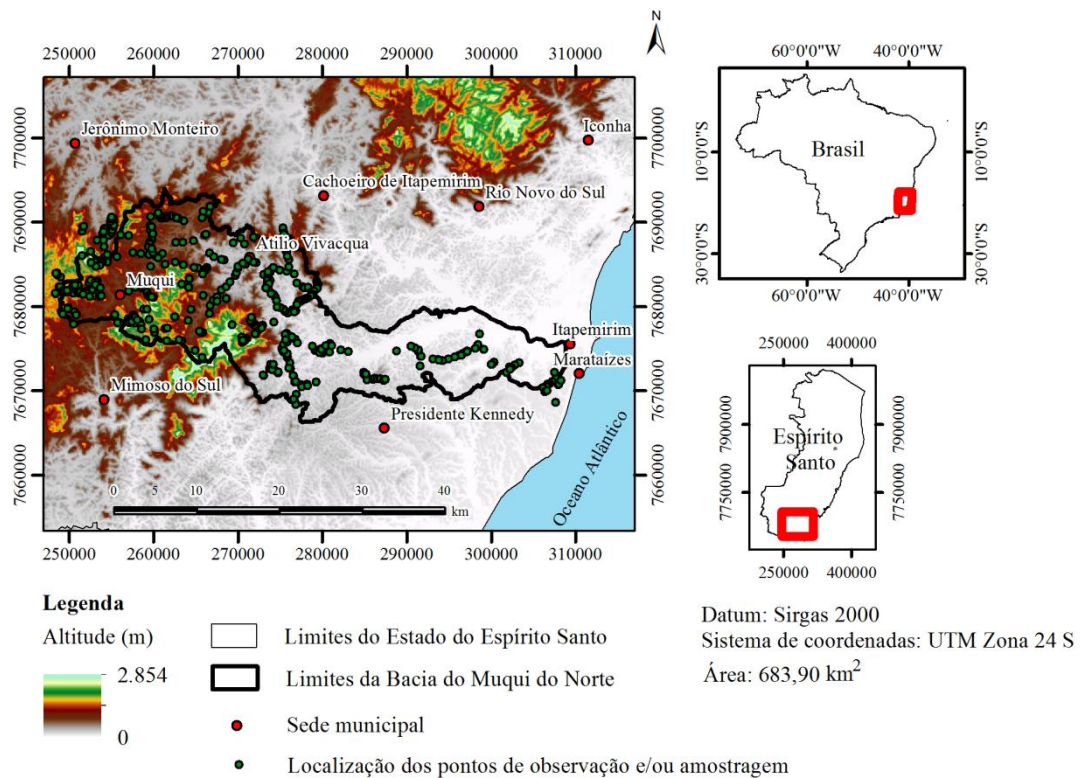


Figura 1. Localização e altimetria da bacia do rio Muqui do Norte no Estado do Espírito Santo.

Foram realizadas 598 observações, entre coletas de pontos de coordenadas com GPS de navegação modelo Garmin GPSMAP 76 e identificação por meio da utilização de ortofotomosaico com resolução espacial de 1 metro (IEMA, 2013), em locais de difícil acesso. Em alguns pontos foram coletadas amostras para análises físicas e químicas. Nove classes de solos e uma de afloramentos de rocha foram identificados na área (Quadro 1), das quais entre as principais, três foram descritas e amostradas conforme Santos et al. (2005) e classificadas de acordo com o Sistema Brasileiro de Classificação. Os demais perfis utilizados foram obtidos de estudos preexistentes.

Utilizaram-se quarenta e cinco variáveis discriminantes no mapeamento digital de solos da área. Dentre elas, 28 atributos do terreno, 1 geomorfológica, 2 geológicas, 11 climáticas e 3 derivadas de dados do Landsat 5 TM (Quadro 2). Os atributos do terreno foram derivados de modelo digital de elevação (MDE), com 30 m de resolução espacial, obtido com a opção “topo to raster” do programa ArcGIS 10.0 (ESRI, 2011). As curvas de nível equidistantes de 20 metros e a hidrografia foram obtidas das cartas topográficas do IBGE, na escala de 1:50.000, disponível em www.geobases.es.gov.es.

Quadro 1. Classes de solos identificadas na bacia do rio Muqui do Norte, ES

Classes de solo observadas	Perfis representativos	Fonte	Município
LATOSSOLO VERMELHO-AMARELO	LATOSSOLO VERMELHO-AMARELO distrófico*	Ferreira Neto et al. (2009)	Muqui
ARGISSOLO VERMELHO	ARGISSOLO VERMELHO eutrófico*	Embrapa (1978)	Cachoeiro de Itapemirim
ARGISSOLO VERMELHO-AMARELO	ARGISSOLO VERMELHO-AMARELO eutrófico*	Ferreira Neto et al. (2009)	Muqui
ARGISSOLO AMARELO	ARGISSOLO AMARELO Distrocoeso típico	Embrapa (1978)	Marataízes
CHERNOSSOLO ARGILÚVICO	CHERNOSSOLO ARGILÚVICO	Embrapa (1978)	Castelo
CAMBISSOLO HÁPLICO	CAMBISSOLO HÁPLICO Tb Distrófico léptico	Embrapa (1978)	Alfredo Chaves
GLEISSOLO HÁPLICO	GLEISSOLO HÁPLICO	Descrito para este estudo	Itapemirim
NEOSSOLO FLÚVICO	NEOSSOLO FLÚVICO Tb Eutrófico gleissólico	Descrito para este estudo	Atílio Vivacqua
NEOSSOLO LITÓLICO	NEOSSOLO LITÓLICO eutrófico típico	Alkimin (2009)	Anchieta
Afloramento de rocha			

*Descrito e coletado dois horizontes diagnósticos (A e B).

Os atributos geomorfológicos constaram do Mapa de Unidades Geomorfológicas do Estado do Espírito Santo (IJSN, 2012) e de um mapa pedogeomorfológico, gerado a partir do MDE, com a delimitação das áreas de topo, encostas côncavas, encostas convexas e terraços.

As unidades litológicas e as linhas de falhas foram obtidas, respectivamente, dos mapas de Geologia do Projeto RADAMBRASIL (Brasil, 1983) e do mapa geológico ao milionésimo da CPRM (2012).

As áreas basais das elevações foram delimitadas com base no delineamento das bacias hidrográficas com ferramenta “basin” do Arcgis 10, a partir do raster de direção de fluxo gerado do MDE com valores invertidos (multiplicados por -1). O raster com os limites das bases das elevações (bacias invertidas) foi convertido para o formato de arquivo tipo vetor (shapefile) e realizada a eliminação dos polígonos com áreas muito pequenas, inferiores à 0,0576 ha. Em seguida, com a utilização da ferramenta “Minimum Bounding Geometry”, foram criados retângulos de menores larguras envolvendo os polígonos de entrada, ou seja, os polígonos que representam as bases das elevações. Com essa ferramenta, um novo arquivo shape do tipo polígono (retângulos) foi criado, contendo informações de suas medidas como ângulos de direção do maior eixo, comprimento e largura. A medida comprimento e área da base da elevação foram utilizadas para o cálculo do fator de forma - Kf e o coeficiente de compacidade - Kc.

Para o cálculo da densidade de drenagem, utilizou-se o comprimento dos canais de drenagem do IBGE e as áreas das bases das elevações.

Quadro 2. Variáveis discriminantes utilizadas no mapeamento digital de solos da bacia do rio Muqui do Norte, ES

Variável
----- Derivadas do MDE (carta topográfica) -----
Rugosidade
Altitude
Compacidade
Fator forma
Área basal das elevações
Altitude relativa à menor cota das bases das elevações
Distância à drenagem
Segunda derivada da declividade
Curvatura
Razão de elevação do relevo
Dissecação
Índice de corrente máximo de fluxo
Índice topográfico combinado
Índice de capacidade de transporte de sedimentos
Aspecto
Declividade
Densidade de drenagem
Distância mais próxima do mar
Percentual de área com curvatura retilínea
Percentual de área com curvatura côncava
Percentual de área de topo
Percentual de área de topo + área convexa
“Northness”
“Eastness”
Toposequencia – topo, convexa, côncava e terraço
Perfil de curvatura
Plano de curvatura
----- Derivadas da imagem Landsat 5 TM -----
NDVI
CMI – “Clay mineral index”
IOI – “Iron oxide index”
----- Derivadas de mapas -----
Geologia
Distância às linhas de falha
Geomorfologia
----- Climáticas e balanço hídrico -----
Temperatura média anual
Temperatura média das mínimas anual
Temperatura média das máximas anual
Precipitação pluvial média anual
Precipitação pluvial do período seco
Precipitação pluvial do período úmido
Radiação solar
Deficiência hídrica
Evapotranspiração potencial
Evapotranspiração real
Excedente hídrico

A amplitude altimétrica - Δa foi calculada da seguinte forma: $\Delta a = h_{\max} - h_{\min}$, onde h_{\max} = altitude máxima relativa a cada polígono referente à base de elevação e; h_{\min} = altitude mínima relativa a cada polígono referente à base de elevação. A altitude relativa à drenagem de cada base de elevação foi obtida pela subtração da altitude mínima de cada base pelos valores de altitude do MDE.

Outras variáveis relacionadas aos polígonos das bases das elevações foram o mapa pedogeomorfológico, os percentuais das áreas de topo, de topo + encostas convexas, de encostas côncavas e de curvatura retilínea.

Para produzir o mapa pedogeomorfológico, primeiramente a área de estudo foi dividida em seis subáreas com diferentes classes de amplitudes altimétricas: 0 a 50 m (muito baixa); 50,1 a 100 m (baixa); 100,1 a 250 m (média); 250,1 a 500 m (alta); 500,1 a 750 m (muito alta); 750,1 a 1.028,2 m (extremamente alta). Este procedimento foi necessário para melhor definir os diferentes limites, para cada subárea, dos segmentos planos e suaves ondulados (0 a 10% de declividade) da paisagem que separam topos e terraços. Os limites de separação entre topos e terraços foram estabelecidos com base em diferentes altitudes relativas que melhor se ajustaram visualmente às curvas de nível. Para as subáreas de muito baixa, baixa, média, alta, muito alta e extremamente alta, os valores considerados como limiares entre topos e terraços foram respectivamente de 25 m, 40 m, 85 m, 150 m, 250 m e 150 m. Nas áreas com declividade superior a 10% constaram os valores de curvatura. Dessa forma, foi possível estratificar pedogeomorfologicamente a bacia em topos e terraços com declividades entre 0 e 10% e em encostas côncavas e convexas com declividades superiores a 10%.

As superfícies com informações de distância à drenagem, distância mais próxima do mar e distância às linhas de falha foram obtidas no Arcgis 10.0 pelo cálculo da distância euclidiana. As feições de interesse mais próximas contidas em arquivos no formato "shapefile" foram: drenagem da carta 1:50.000 do IBGE, linha de costa obtida das malhas municipais do IBGE e linhas de falhas digitalizadas do mapa geológico 1:1.000.000 do Radambrasil (Brasil, 1983).

Dados do sensor TM do Landsat-5 (maio de 2010) foram utilizados para derivar os índices minerais de argila ("clay mineral index", CMI), óxido de ferro ("iron oxide index", IOI) e vegetação por diferença normalizada, definido como NDVI ("normalized difference vegetation index") = $(IVP - Vis)/(IVP + Vis)$, onde IVP é a intensidade de reflectância no infravermelho próximo e Vis a intensidade de reflectância do vermelho

no visível. O CMI e o IOI foram obtidos de acordo com Sabins (1987). O primeiro por meio da divisão da banda 5 (1,55–1,75 μm) pela banda 7 (2,08–2,35 μm) e o segundo por meio da divisão da banda 3 (0,63–0,69 μm) pela banda 1 (0,45–0,52 μm). Os três índices foram obtidos com a utilização do ArcGIS 10.0 (ESRI, 2011).

Os dados climáticos utilizados como variáveis discriminantes foram: precipitação pluvial, temperaturas médias das mínimas, médias e máximas do ar e radiação solar global. A precipitação para a bacia foi obtida por meio de cokrigagem dos dados dos postos pluviométricos localizados no Estado e áreas vizinhas, obtidos no sítio <http://hidroweb.ana.gov.br/> e junto ao Instituto Capixaba de Pesquisa e Extensão Rural - INCAPER. Para o período úmido (outubro a abril) e anual a variável auxiliar foi a altitude e para o período seco (maio a setembro) a altitude mais a distância do mar.

As temperaturas médias do ar foram obtidas por meio de modelagem a partir de dados obtidos em estações meteorológicas distribuídas no Estado. As equações de regressão foram desenvolvidas por Castro et al. (2010) para permitir estimativas de temperatura em função da altitude, latitude e longitude.

A partir dos dados de precipitação e da estimativa da temperatura média do ar para todos os locais dos postos, calculou-se os elementos do balanço hídrico: a deficiência hídrica, evapotranspiração potencial, evapotranspiração real e excedente hídrico, pelo método proposto por Thornthwaite & Mather (1955). Os cálculos foram realizados com o auxílio do programa "BHnorm", elaborado em planilha Excel por Rolim et al. (1998), assumindo-se uma Capacidade Máxima de Armazenamento de água no solo igual a 100 mm.

A quantificação da variável radiação solar global foi realizada por meio da ferramenta de análise "área solar radiation" disponível no software ArcGis 10, a qual considera como dados de entrada o MDE.

Os valores das variáveis intervalares foram reescalados entre o intervalo 0 e 1. Este procedimento foi realizado com a ferramenta "raster calculator" do ArcGis 10.0, utilizando a expressão $x' = (x - x_{min}) / (x_{max} - x_{min})$, onde x é o valor da variável, x_{min} e x_{max} são respectivamente o menor e o maior valor que esta variável assume, e x' o novo valor de x , normalizado. Nas variáveis com distribuição simétrica que apresentaram valores excessivamente elevados em um número muito reduzido de células (ruídos e outliers), os mesmos foram substituídos pelos valores correspondentes à média aritmética \pm três vezes o desvio-padrão. Para as variáveis com distribuição assimétrica

os valores foram substituídos por aqueles correspondentes a 0,5 % ou 1% dos maiores valores.

Foram realizadas predições de classes de solos com a utilização de quatro algoritmos de aprendizagem disponíveis no software WEKA, Versão 3.6.8 (Witten & Frank, 2005): NaiveBayes (classificador probabilístico baseado em inferência bayesiana), multilayer perceptron - MLP (rede neural artificial), SimpleCart e J48 (os dois últimos classificadores em árvore de decisão). Cada um dos quatro classificadores foi aplicado a dois conjuntos de variáveis preditoras ambientais: um com 45 variáveis e outro com 10 variáveis selecionadas das 45 por meio da aplicação do algoritmo de seleção de atributos denominado “ReliefFAttributeEval” (Kononenko, 1994), também disponível no WEKA 3.6.8.

Os parâmetros para realizar as classificações pelo algoritmo NaiveBayes foram aqueles fornecidos como padrão do software. Para utilização da rede neural (MLP), diferentes arquiteturas foram testadas. Utilizou-se a interface do Weka denominada “Experimenter”, na qual se executam algoritmos com diferentes parâmetros e conjuntos de dados e permite coletar estatística de desempenho e realizar testes de significância dos resultados. Variou-se o número de neurônios na camada interna de 10 a 55, correspondentes à quantidade de classes e ao número de classes mais a quantidade de variáveis ambientais utilizadas, respectivamente. Avaliou-se também diferentes valores para a taxa de aprendizagem (0,01 a 0,3) para o momentum factor (0,1 a 1,0), todos contendo o mesmo número de neurônios na camada de saída, correspondente ao número de unidades de mapeamento (10). Após essa avaliação, utilizou-se a interface “explorer” do Weka 3.6.8. Foram realizadas as classificações usando-se 16 e 18 neurônios nas camadas internas, 0,12 e 0,08 de taxas de aprendizagem e 0,8 e 0,3 de “momentum factor” para os conjuntos de 10 e 45 atributos, respectivamente. Foram considerados 10 mil ciclos de aprendizagem. Os parâmetros foram avaliados com base no valor do Kappa apresentado com base nas classes preditas e observadas para o treinamento.

Para realizar as classificações pelo algoritmo SimpleCart foram previamente testadas variações nos parâmetros “minNumObject” (2 a 5), numFoldsPruning (1 a 10) e “seed” (1 a 10). Os valores desses parâmetros utilizados para predição das classes foram iguais a 2, 7 e 7 para o conjunto de 10 variáveis e iguais a 2, 5 e 1 para o conjunto de 45 variáveis. O método de poda da árvore utilizado foi o “minimal cost-complexity

pruning”. Realizou também uma classificação com o SimpleCart com 10 atributos com a utilização dos parâmetros fornecidos como padrão do software.

Para o algoritmo J48, foram testadas variações nos parâmetros “confidence factor” (0,1, 0,125, 0,25, 0,50 e 0,75), “minNumObject” (1 a 5) e unpruned (sem poda). Os valores dos parâmetros selecionados para desenvolver a predição de classes de solo foram 0,125 e 2 para confidence factor e “minNumObject”, respectivamente. A poda foi desenvolvida (unpruned = false). O método de poda reducedErrorPruning não foi selecionado.

A avaliação dos resultados foi realizada com base no índice Kappa, derivado de uma matriz de confusão (Congalton & Green, 2009). Essa medida estatística foi obtida por meio da validação cruzada, considerando o número de “folds” (subconjuntos de dados) igual a 10, onde cada um destes “folds” é utilizado para teste e o restante para treinamento.

A generalização, ou seja, a classificação das amostras que não foram utilizadas para treinamento foi realizada no software WEKA 3.6.8, com base nos modelos gerados. Essas amostras classificadas foram implementadas no ArcGis para gerar as imagens raster que se constituíram nos mapas de solos.

As classificações foram desenvolvidas também considerando apenas sete classes de solos (Quadro 2) e um total de 452 instâncias de treinamento. A redução da quantidade de classes de solos teve o objetivo de proporcionar uma comparação entre o mapeamento gerado pelos quatro classificadores e um mapa de solos de referência. Tendo em vista que a área estudada não possui mapas de solos detalhados ou semidetalhados, utilizou-se o mapa de solos do levantamento original do projeto RADAMBRASIL, em escala 1:250.000, correspondente às Folhas SF 23/24 Rio de Janeiro/Vitória (Brasil, 1983). O mapa foi digitalizado em mesa digitalizadora pelo Instituto Estadual de Meio Ambiente e Recursos Hídricos do Espírito Santo - IEMA, mantendo-se a escala de 1:250.000. Para este estudo, o arquivo obtido no formato shape foi recortado com base nos limites da bacia do rio Muqui do Norte, ES. Foi realizada uma comparação com os primeiros e outra com os segundos componentes das unidades de mapeamento do RADAMBRASIL, as quais foram atualizadas até o segundo nível categórico, de acordo com o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006). Dessa maneira, o arquivo shape foi convertido para o formato raster com tamanho das células de 250 m x 250 m, assumindo que a menor área discernível é de 1 mm x 1 mm e considerando que o tamanho do pixel pode ser calculado

multiplicando-se o denominador da escala por 0,001 m (McBratney et al., 2003). A exatidão desse mapa de referência foi verificada com base nos pontos de observação disponíveis. Foi necessário tornar os mapas preditos compatíveis em termos de escala com o mapa do Projeto RADAMBRASIL. Para tanto, foi realizada uma reamostragem dos tamanhos de pixel de 30 m para 250 m nos mapas de solos gerados pelos diferentes algoritmos de classificação.

Quadro 2. Unidades de mapeamento do mapa de solos do Projeto RADAMBRASIL e primeiros componentes atualizados até o segundo nível categórico do Sistema Brasileiro de Classificação do Solo (Embrapa, 2006)

Unidade de mapeamento original*	Unidade de mapeamento atualizada
LVa15 - Latossolo Vermelho Amarelo Álico + Podzólico Vermelho Amarelo Álico	Latossolo Vermelho Amarelo + Argissolo Vermelho Amarelo
LVa34 - Latossolo Vermelho Amarelo Álico + Latossolo Variação Una Álico + Cambissolo Álico	Latossolo Vermelho Amarelo + Latossolo Vermelho Amarelo + Cambissolo Háptico
LVa2 - Latossolo Vermelho Amarelo Álico	Latossolo Vermelho Amarelo
PEe6 - Podzólico Vermelho Escuro eutrófico Tb + Podzólico Vermelho Amarelo eutrófico Tb	Argissolo Vermelho + Argissolo Vermelho Amarelo
PEe9 - Podzólico Vermelho Escuro eutrófico Tb + Podzólico Vermelho Amarelo eutrófico Tb	Argissolo Vermelho + Argissolo Vermelho Amarelo
Ca32 - Cambissolo Álico + Latossolo Vermelho Amarelo Álico	Cambissolo Háptico + Latossolo Vermelho Amarelo
Re2 - Solos Litólicos eutróficos + Podzólico Vermelho Escuro eutrófico	Neossolo Litólico + Argissolo Vermelho
PAa3 - Podzólico Amarelo Álico + Latossolo Amarelo Álico	Argissolo Amarelo + Latossolo Amarelo
PAa4 - Podzólico Amarelo Álico + Latossolo Amarelo Álico	Argissolo Amarelo + Latossolo Amarelo
Ad2 - Aluvial distrófico + Glei Pouco Húmico distrófico	Neossolo Flúvico + Gleissolo Háptico
BV3 - Brunizém Avermelhado + Podzólico Vermelho Escuro eutrófico Tb	Chernossolo Argilúvico + Argissolo Vermelho

* Considerando somente as associações de classes de solos.

4.5. RESULTADOS E DISCUSSÃO

Os valores de Kappa dos algoritmos de árvore de decisão não apresentaram diferenças significativas ao se utilizar os conjuntos com 10 e 45 variáveis preditoras ambientais (Quadro 3). Verifica-se, no caso dos algoritmos MLP e NaiveBayes, que os mesmos apresentaram diferenças ao se comparar os valores de Kappa para os dois conjuntos de variáveis avaliados. Nesse caso, o conjunto com maior quantidade de variáveis proporcionou valores de Kappa mais elevados. Esses dois algoritmos (NaiveBayes e MLP) apresentaram perda de acurácia de predição em razão da seleção de atributos, diferentemente do observado para as árvores de decisão (SimpleCart e J48). Embora a falta de seleção de atributos possa resultar em redução da exatidão na

classificação (Yang et al., 2011), esses resultados indicam que, para o caso dessa base de dados, os algoritmos NaiveBayes e MLP apresentaram-se robustos mesmo com a existência de atributos redundantes.

Quadro 3. Índice Kappa, Variância do Kappa (vâr), número de folhas e tamanho da árvore (nº de nós), obtidos na validação cruzada para os diferentes algoritmos de classificação e conjuntos de atributos, considerando-se 10 unidades de mapeamento

Método/ Nº de atributos	CART/45	J48/45	MLP/45	Bayes/45	CART/10	J48/10	MLP/10	Bayes/10	CART/10*
Nós (und)	59	101			65	122			65
Folhas (und)	30	65			33	84			33
Kappa	0,64	0,56	0,63	0,66	0,60	0,51	0,50	0,56	0,60
Vâr (x 10 ⁻⁴)	4,8	5,4	5,0	4,6	5,0	5,3	5,4	5,2	5,0
CART/45	0	-2,53*	-0,44	0,53	-1,31	-4,03*	-4,22*	-2,54*	-1,31
J48/45		0	2,08*	3,08*	1,23	-1,46	-1,65	0,01	1,23
MLP/45			0,00	0,97	-0,86	-3,57*	-3,75*	-2,08*	-0,87
Bayes/45				0,00	-1,86	-4,60*	-4,79*	-3,09*	-1,86
CART/10					0,00	-2,72*	-2,91*	-1,23	0,00
J48/10						0,00	-0,19	1,49	2,72*
MLP/10							0,00	1,68	2,90*
Bayes/10								0,00	1,22
CART/10*									0,00

* significativo a 5% de probabilidade pelo teste Z. CART/45 – Algoritmo SimpleCart com 45 atributos; J48/45 – Algoritmo J48 com 45 atributos; MLP/45 – algoritmo MLP com 45 atributos; Bayes/45 – Algoritmo NaiveBayes com 45 atributos; CART/10 – Algoritmo SimpleCart com 10 atributos; J48/10 – Algoritmo J48 com 10 atributos; MLP/10 – algoritmo MLP com 10 atributos; Bayes/10 – Algoritmo NaiveBayes com 10 atributos.

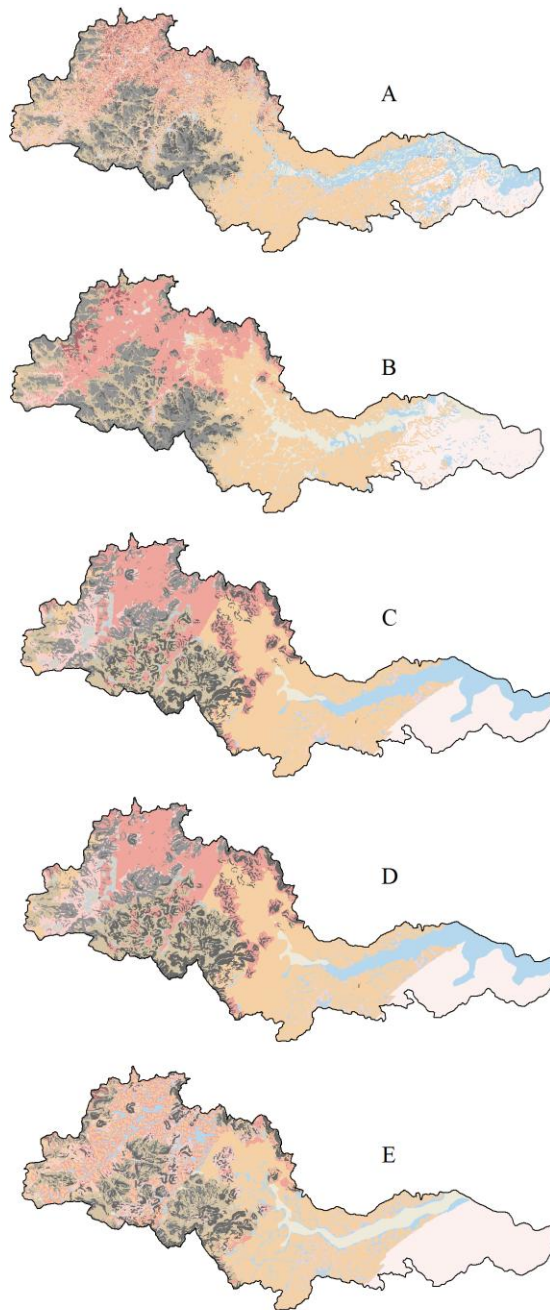
Ao comparar os dois algoritmos de árvore de decisão, verifica-se que o SimpleCart apresentou acurácia superior ao J48, obtendo-se diferença no valor de Kappa de 0,09 e 0,08, respectivamente, quando se utilizou 10 e 45 atributos. Além de valor de Kappa mais elevado, o SimpleCart produziu árvore de decisão mais simples, com menor tamanho, expresso pelos menores números de nós e de folhas. Quanto à comparação entre o SimpleCart configurado com parâmetros (*minNumObject*, *mumFoldsPruning* e *seed*) padrão do software e o SimpleCart com parâmetros otimizados, verificou-se não existir diferenças estatísticas na exatidão pelo teste Z (Quadro 3). Valores de Kappa mais baixos e próximos entre si foram encontrados por Coelho & Giasson (2010) para os classificadores SimpleCart e J48, ao compararem os mapas preditos com um mapa convencional. Em outro estudo, Giasson et al. (2011) compararam cinco algoritmos de árvore de decisão e obtiveram melhor desempenho na predição com a utilização do SimpleCart. Esses resultados sugerem que os diferentes algoritmos possivelmente apresentam desempenhos influenciados pelas características

das áreas de estudo, como quantidade de classes de solos a serem mapeadas, quantidade e combinações das variáveis preditoras utilizadas.

Na comparação geral, embora sem diferenças significativas entre eles, os maiores valores de Kappa foram encontrados para os algoritmos NaiveBayes, MLP e SimpleCart. Visualmente, os mapas preditos apresentam-se diferentes (Figura 2). As concordâncias entre os mapas preditos por MLP, SimpleCart e J48, em relação ao mapa gerado por NaiveBayes, foram de 56%, 55% e 52%, respectivamente.

No caso dos classificadores de NaiveBayes e MLP, os maiores valores de Kappa foram obtidos com a utilização de 45 variáveis preditoras. Para SimpleCart e J48, não se verificaram alterações significativas nos valores de Kappa em razão da mudança na quantidade de variáveis utilizadas. Este resultado sugere que as árvores de decisão apresentam a vantagem de se utilizar um menor número de variáveis preditoras sem perda da acurácia da predição. Uma menor quantidade de atributos permite simplificar o modelo e melhorar o seu entendimento, pois conforme Giasson et al. (2011), a análise das regras de decisão das árvores permite explicitar as relações entre os solos ou aos seus fatores de formação e as variáveis preditoras ambientais utilizadas na modelagem.

Métodos como redes neurais artificiais e Bayes podem apresentar maiores detalhes espaciais no delineamento de unidades de mapeamento em relação aos mapas convencionais (Chagas et al., 2010). Nesse sentido, o mapa predito por MLP mostrou uma maior complexidade na distribuição das classes de solo. Isso é facilmente visualizado na região em que predomina PV, cuja unidade de mapeamento apresenta-se delimitada entre a unidade LVA. No entanto, conforme observações da paisagem em campo, nessa região da bacia são comuns áreas contíguas mais extensas de PV, sem a presença da unidade LVA. Nessa região, o mapa predito apresenta também ocorrência de Chernossolos distribuídos de forma generalizada entre PV e LVA, enquanto em campo verifica-se esta unidade localizada principalmente no extremo oeste da bacia, associados às rochas máficas. Comportamento semelhante quanto à ocorrência de LVA em região de PV foi verificado para o algoritmo J48. Destaca-se ainda que J48 não mapeou as unidades GX e RY, as quais ocorrem no ambiente da Formação Barreiras, a leste da bacia. Dessa maneira, ao considerar o aspecto visual dos mapas preditos, NaiveBayes e SimpleCart aparentemente refletiram melhor a distribuição das classes observadas em campo.



Legenda	MLP / 45		Bayes / 45		SimpleCart / 10		J48 / 10		SimpleCart / 10*	
	km ²	%	km ²	%	Área km ²	%	km ²	%	km ²	%
Limites da bacia										
AR - Afloramento de rocha	30,5	4,5	27,9	4,1	69,5	10,2	60,9	8,9	62,2	9,7
MT - Chernossolo Argilúvico	11,9	1,7	8,9	1,3	0,5	0,1	1,7	0,2	0,46	0,1
CX - Cambissolo Háplico	63,2	9,2	80,0	11,7	63,8	9,3	73,7	10,8	61,4	9,6
GX - Gleissolo Háplico	77,9	11,4	24,2	3,5	78,5	11,5	54,6	8,0	69,7	10,9
RY - Neossolo Flúvico	24,6	3,6	46,5	6,8	5,4	0,8	31,5	4,6	5,0	0,8
LVA - Latossolo Vermelho-Amarelo	259,7	38,0	188,4	27,5	195,4	28,6	213,2	31,2	187,5	29,2
PA - Argissolo Amarelo	43,3	6,3	89,1	13,0	73,0	10,7	110,1	16,1	68,0	10,6
PVA - Argissolo Vermelho-Amarelo	31,5	4,6	11,2	1,6	36,6	5,4	49,0	7,2	34,6	5,4
PV - Argissolo Vermelho	62,2	9,1	125,3	18,3	105,7	15,4	59,8	8,7	100,0	15,6
RLi - Neossolo Litólico	79,1	11,6	82,4	12,0	55,5	8,1	29,5	4,3	52,9	8,2

MLP/45 – algoritmo MLP com 45 atributos; Bayes/45 – Algoritmo Bayes com 45 atributos; SimpleCART/10 – Algoritmo SimpleCart com 10 atributos; SimpleCART/10* – Algoritmo SimpleCart com 10 atributos em que os parâmetros para realizar a classificação foram aqueles fornecidos como padrão do software; J48/10 – Algoritmo J48 com 10 atributos.

Figura 2. Mapas de solos da bacia do rio Muqui do Norte, com dez unidades de mapeamento, gerados pelos diferentes algoritmos de classificação. A) MLP/45; B) Bayes/45, C) SimpleCart/10*, D) SimpleCart/10 e J48/10.

Ao considerar um menor número de classes de saída (7), os algoritmos MLP, SimpleCart e J48 apresentaram elevação nos respectivos valores de Kappa (Quadro 4). Para esses três algoritmos, a redução de dez para sete classes resultou em elevação de até 29% nos valores de Kappa, conforme observado para as predições feitas pelo algoritmo J48. Nessa condição, a maior exatidão da classificação foi observada para o algoritmo MLP, seguido por SimpleCart e J48, ao passo que NaiveBayes não respondeu em termos de melhoria da exatidão devido à redução do número de classes de solos. Verifica-se que esses resultados se devem principalmente à supressão da classe PVA, a qual resultava em confusão com a classe LVA e PV, principalmente (Quadro 5).

Quadro 4. Índice Kappa, Variância do Kappa (vâr), número de folhas e tamanho da árvore (nº de nós), obtidos na validação cruzada para os diferentes algoritmos de classificação e conjuntos de atributos, considerando-se 7 unidades de mapeamento

Método/Quantidade de atributo	MLP/45	Bayes/45	CART/10	J48/10
Nós (und)			43	103
Folhas (und)			22	71
Kappa	0,75	0,66	0,71	0,71
Vâr (x 10 ⁻⁴)	5,92	6,64	06,39	6,18
MLP/45	0,00	-2,50*	-1,02	-1,10
Bayes/45		0,00	1,46	1,40
CART/10			0,00	-0,08
J48/10				0,00

* significativo a 5% de probabilidade pelo teste Z. MLP/45 – algoritmo MLP com 45 atributos; Bayes/45 – Algoritmo NaiveBayes com 45 atributos; CART/10 – Algoritmo SimpleCart com 10 atributos; J48/10 – Algoritmo J48 com 10 atributos.

O mapa de solos do Projeto RADAMBRASIL apresentou valor de Kappa de 0,37, obtido da matriz de confusão com base nos pontos de observação utilizados para o treinamento e geração dos modelos de predição. Tal valor apresenta-se bem inferior àqueles encontrados para as predições realizadas pelos diferentes algoritmos, com base nos resultados da validação cruzada (Quadro 4). Estes resultados corroboram a necessidade de se preferir validação com dados de observação de campo (Sarmiento, 2010).

Embora comparações com mapas convencionais indiquem somente a capacidade de reprodução do mapa pelos modelos preditivos (Ten Caten et al., 2012), os mesmos servem como uma importante referência para a validação e consolidação dos métodos de mapeamento digital de solos. A comparação dos mapas gerados pelos diferentes algoritmos com o mapa de referência (Brasil, 1983), com base nos índices Kappa, mostrou que as predições pelo algoritmo MLP, seguido pelo SimpleCart, foram os que

geraram os mapas mais semelhantes ao convencional (Quadro 6). Valores de Kappa sem diferenças significativas foram obtidos para os algoritmos J48 e NaiveBayes, os quais apresentaram desempenhos inferiores aos do MLP e do SimpleCart. Desempenhos de predição superiores em relação ao algoritmo NaiveBayes foram apresentados pelo SimpleCart e por redes neurais artificiais conforme encontrados por Coelho & Giasson (2010) e Carvalho Junior et al. (2011), respectivamente, o que corrobora os resultados encontrados na predição de classes de solo da bacia do rio Muqui do Norte, ES.

Quadro 5. Matriz de confusão obtida na classificação das dez classes de solos pelo algoritmo MLP, considerando-se a utilização das quarenta e cinco variáveis ambientais preditoras

	AR	MT	CX	GX	RY	LVA	PA	PVA	PV	RL	Total	Exatidão do usuário (%)	
Classes no mapa predito	AR	46	0	2	0	0	2	0	1	1	6	58	79,3
	MT	0	26	0	0	1	0	0	1	6	0	34	76,5
	CX	4	2	28	0	1	4	0	4	1	2	46	60,9
	GX	0	1	2	18	3	5	4	4	4	0	41	43,9
	RY	0	0	0	4	10	2	0	1	1	0	18	55,6
	LVA	1	0	3	8	2	132	3	11	19	2	181	72,9
	PA	0	0	0	1	0	1	36	0	0	0	38	94,7
	PVA	1	0	3	3	0	12	3	15	8	0	45	33,3
	PV	0	4	2	1	2	18	0	4	47	0	77	60,2
	RL	8	0	0	0	0	1	0	0	0	50	59	84,7
	Total	60	33	39	35	19	177	46	41	87	60	598	
Exatidão do produtor (%)		76,7	78,8	70,0	51,4	52,6	74,6	78,3	36,6	54,0	83,3		

Quadro 6. Kappa dos mapas de solo generalizados para 1:250.000 gerados pelo MLP, NaiveBayes, Cart e J48 em comparação ao Mapa de referência (RADAMBRASIL), considerando-se os primeiros componentes das unidades de mapeamento

Método/Quantidade de atributo	MLP/45	Bayes/45	CART/10	J48/10
EG	0,55	0,48	0,52	0,51
Kappa	0,41	0,36	0,39	0,38
Var ($\times 10^{-5}$)	3,3	3,0	3,2	3,2
MLP/45	0,00	-6,03*	-2,13*	-4,18*
Bayes/ 45		0,00	3,87*	1,78
CART/10			0,00	-2,06*
J48/10				0,00

* significativo a 5% de probabilidade pelo teste Z. MLP/45 – algoritmo MLP com 45 atributos; Bayes/45 – Algoritmo NaiveBayes com 45 atributos; CART/10 – Algoritmo SimpleCart com 10 atributos; J48/10 – Algoritmo J48 com 10 atributos.

A unidade LVA foi apresentada como a de maior ocorrência nos mapas preditos pelos quatro algoritmos utilizados, em acordo com o mapa referência (Figura 3). Por outro lado, esses algoritmos mostraram que a unidade de mapeamento com menor

percentual em área foi a dos Chernossolos Argilúvicos - MT, o que difere do observado no mapa de referência. Este resultado pode indicar uma superestimação da área de Chernossolos no mapa convencional.

As concordâncias entre o mapa gerado por MLP e o mapa de referência dos primeiros componentes das unidades de mapeamento (Figura 4) estão localizados principalmente nas áreas de ocorrência das unidades LVA, PA e RY (Figura 3A e 3B). Estas três unidades estão relacionadas às áreas relativamente menos dissecadas, onde predominam relevos que variam de plano a ondulado. Nas áreas de relevo mais movimentado, principalmente da porção central a oeste da bacia, ocorreram os maiores percentuais de discordância. Em parte, esses resultados podem estar relacionados à maior complexidade de solos dessas áreas, não representadas completamente no mapa convencional, ou devido aos seus percentuais relacionados aos segundos componentes das associações. Entre as causas das discordâncias há de se considerar outro aspecto, o qual está relacionado à delimitação dos polígonos dos mapas convencionais. Discordâncias provavelmente relacionadas a essa questão podem ser verificadas próximas aos limites entre as unidades de mapeamento LVA e RY (Figuras 3 e 4), localizadas a oeste da bacia. Segundo Zhu et al. (2001), variações graduais e tênues nas condições ambientais são de difícil localização por meio do método convencional, o que gera uma incerteza quanto à real localização das transições entre classes na paisagem. Neste sentido, Ten Caten et al. (2012) afirmam que os mapas de solos convencionais têm nas bordas dos polígonos a região de maior variabilidade, devido à subjetividade, a qual é intrínseca ao próprio método.

A natureza generalizada do mapa de solos convencional, por conta da limitação de escala pode dificultar a análise de concordância com os mapas inferidos (Zhu et al., 2001; Chagas, 2006). Como se observa no Quadro 7, as unidades de mapeamento do mapa convencional utilizado como referência apresentam associações de até 3 classes de solos, além da existência de inclusões, enquanto nos mapas preditos as unidades correspondem a apenas uma classe de solo. Para contornar esse problema, foi realizada uma análise conforme Chagas (2006), na qual foram determinadas quais classes de solos identificadas nos mapas preditos ocorrem nos polígonos correspondentes a cada unidade de mapeamento do mapa de solos convencional.

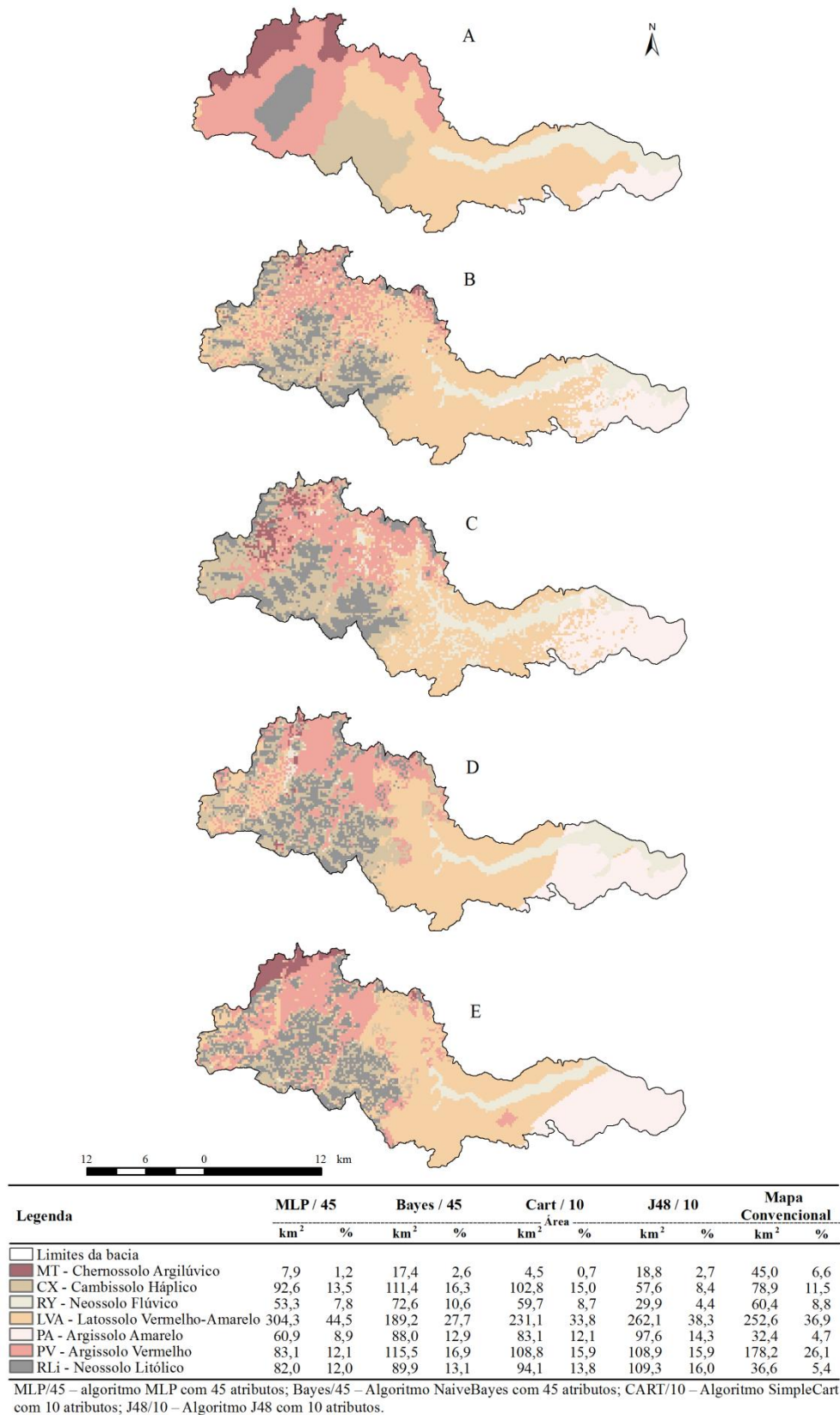


Figura 3. Mapade solos convencional considerando os primeiros componentes das sete unidades de mapeamento da bacia do rio Muqui do Norte (A) e mapas gerados pelos diferentes algoritmos MLP (B), NaiveBayes (C), SimpleCart (D) e J48 (E).

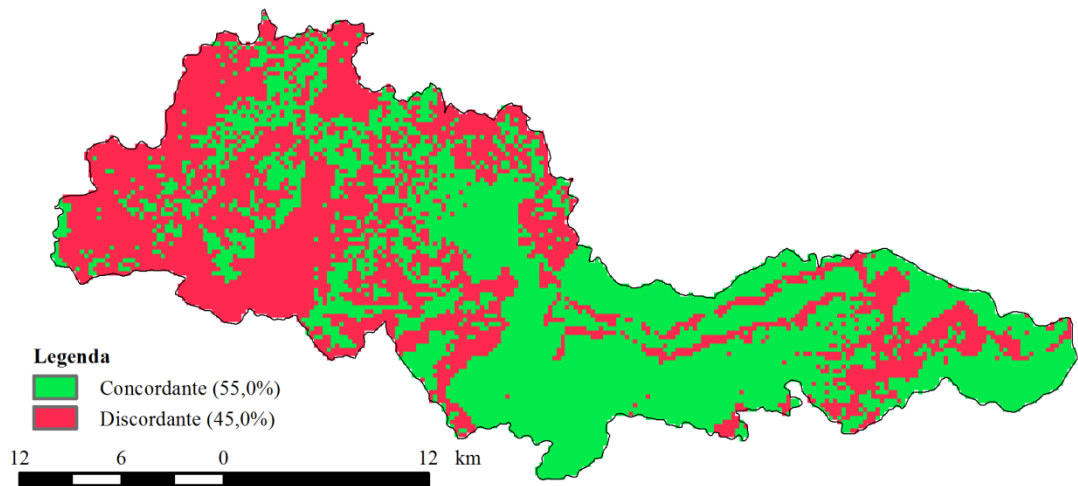


Figura 4. Mapa de concordância entre o mapa gerado rede neural MLP / 45 e o mapa de referência (B).

Quadro 7. Percentual de área das classes de solos identificadas nos mapas preditos correspondentes à área de cada unidade de mapeamento do mapa de solos de referência (RADAMBRASIL).

----- Mapa convencional -----		----- Mapa predito -----				
Unidades de mapeamento original	Associações atualizadas	Classes de solos	MLP/45 (%)	Bayes/45 (%)	CART/10 (%)	J48/10 (%)
LVa15 (PV+CX)	LVA + PVA	MT	0,0	0,0	0,0	0,0
		CX	0,0	0,0	0,0	0,2
		RY	4,6	15,3	5,0	3,9
		LVA	79,7	62,1	66,7	67,0
		PA	15,7	22,6	28,3	27,1
		PV	0,0	0,0	0,0	1,9
		RL	0,0	0,0	0,0	0,0
LVa34 e LVa2 (PVA+ PV + AR)	LVA + CX	MT	0,1	0,0	0,0	0,0
		CX	0,5	2,1	3,9	3,5
		RY	2,4	16,3	1,6	1,6
		LVA	82,9	57,2	67,2	76,2
		PA	0,1	0,0	0,4	0,0
		PV	13,5	24,4	24,7	17,7
		RL	0,6	0,1	2,2	1,0
PEe6 (RL+AR) e PEe9 (RL + PVA + AR)	PV + PVA	MT	2,1	5,4	1,0	1,2
		CX	20,9	30,4	25,8	12,8
		RY	0,4	1,8	2,0	0,3
		LVA	32,4	6,2	16,2	25,7
		PA	0,0	0,0	0,0	0,0
		PV	28,3	39,0	33,5	35,5
RL	15,9	17,4	21,4	24,5		

Continua...

Quadro 7. Continuação

----- Mapa convencional -----		----- Mapa predito -----				
Unidades de mapeamento original	Associações atualizadas	Classes de solos	MLP/45 (%)	Bayes/45 (%)	CART/10 (%)	J48/10 (%)
Ca32 (LVA + CX + RL)	CX + LVA	MT	0,3	0,0	0,0	0,0
		CX	40,3	38,9	42,3	26,9
		RY	1,0	2,3	0,6	0,2
		LVA	21,2	13,6	18,8	17,5
		PA	0,0	0,0	0,0	0,0
		PV	3,0	6,8	8,2	10,5
		RL	34,2	38,4	30,1	44,8
Re2 (AR)	RL + PV	MT	1,4	3,8	0,9	0,0
		CX	32,8	31,5	32,6	21,4
		RY	0,9	0,3	3,1	0,0
		LVA	11,5	2,7	6,3	8,0
		PA	0,0	0,0	0,0	0,0
		PV	17,4	21,2	14,4	22,4
		RL	36,1	40,5	42,7	48,2
PAa3 (RQ) e PAa4	PA + LA	MT	0,0	0,0	0,0	0,0
		CX	0,0	0,0	0,0	0,0
		RY	1,0	0,0	2,9	0,0
		LVA	11,6	12,9	4,3	5,8
		PA	87,5	87,1	92,9	94,2
		PV	0,0	0,0	0,0	0,0
		RL	0,0	0,0	0,0	0,0
Ad2 (OJ)	RY + GX	MT	0,0	0,0	0,0	0,0
		CX	0,0	0,0	0,0	0,0
		RY	68,9	46,3	71,6	35,5
		LVA	19,2	15,0	16,8	25,8
		PA	11,9	38,7	11,6	38,7
		PV	0,0	0,0	0,0	0,0
		RL	0,0	0,0	0,0	0,0
BV3 (PVA + RL + AR)	MT + PV	MT	7,2	14,4	5,4	36,8
		CX	24,6	29,3	17,4	5,3
		RY	0,1	0,0	1,7	0,0
		LVA	13,5	1,9	10,3	10,0
		PA	0,0	0,0	0,0	0,0
		PV	25,7	23,6	33,1	21,9
		RL	28,9	30,7	32,2	26,0

LVA - Latossolo Vermelho Amarelo; PVA - Argissolo Vermelho Amarelo; PV - Argissolo Vermelho; CX - Cambissolo Háplico; RL - Neossolo Litólico; PA - Argissolo Amarelo; RY - Neossolo Flúvico; GX - Gleissolo Háplico; MT - Chernossolo Argilúvico; AR - Afloramento de rocha; RQ - Neossolo Quartzarênico; OJ - Organossolo tiomórfico; MLP/45 - algoritmo MLP com 45 atributos; Bayes/45 - Algoritmo NaiveBayes com 45 atributos; CART/10 - Algoritmo SimpleCart com 10 atributos; J48/10 - Algoritmo J48 com 10 atributos.

Obs.: Entre parênteses, após os componentes das associações do mapa convencional, estão as inclusões, ou seja, as classes que ocupam extensão inferior a 15% do total da área de determinada unidade de mapeamento. Em negrito, os primeiros componentes para cada algoritmo classificador.

Em geral, os resultados mostram que foram identificadas nos mapas preditos mais classes de solos do que no mapa convencional de solos. Estes resultados evidenciam o maior detalhamento na distribuição espacial das classes de solos nos mapas preditos, o que está de acordo com os resultados encontrados por Chagas (2006) ao avaliar os

algoritmos Maxver e redes neurais artificiais no mapeamento digital de solos. Nesse sentido, uma das vantagens do mapeamento digital de solos esta na possibilidade de se obter com facilidade os percentuais de área das associações e inclusões, podendo-se fazer constar essas informações nas legendas dos mapas. Estas são informações relevantes com relação à confiabilidade dos mapas de solos convencionais, no entanto, tais informações muitas vezes não estão disponíveis nesses mapas.

Os quatro algoritmos classificadores (MLP, NaiveBayes, SimpleCart e J48), indicaram os mesmos primeiros componentes na maioria das unidades de mapeamento do mapa convencional. Este fato pode ser explicado pelas pequenas diferenças nos valores de Kappa quando se comparou a eficácia da predição dos quatro algoritmos, conforme se observou no Quadro 4. As maiores divergências na indicação dos primeiros componentes pelos quatro algoritmos ocorreram na unidade de mapeamento BV3 que contém associações de Chernossolo e Argissolo Vermelho. A maior divergência nesta unidade pode estar relacionada ao fato da pequena extensão de área desta unidade e também devido à provável superestimação desta unidade, especialmente com referência à classe Chernossolo no mapa convencional. Além desta divergência de resultados entre os algoritmos para esta unidade de mapeamento, verifica-se que nesta última a classe de maior ocorrência é a de Neossolo Litólico.

Em todas as outras unidades as classes de solos com maiores extensões de área corresponderam às mesmas classes dos primeiros componentes do mapa convencional, o que implica num indicativo de coerência deste último com os mapas preditos. Por outro lado, em quase todas as unidades, exceto na unidade MT+PV, não houve concordância quanto aos segundos componentes do mapa convencional com os mapas preditos. Essa ausência de concordância quanto aos segundos componentes se devem em parte ao fato de que as classes PVA, LA e GX não foram consideradas no mapeamento digital. Outro provável fator está relacionado à generalização do mapa convencional, demonstrando um mais baixo nível de confiabilidade dos segundos componentes e das informações sobre as inclusões. Um exemplo desta discordância em nível de segundo componente e inclusão pode ser vista para o caso da unidade Ca32 (Quadro 7), na qual consta LVA como segundo componente desta associação, enquanto nos mapas preditos a classe RL é aparece como o segundo componente. São duas classes de solo muito distintas pedogeneticamente e que a princípio pode-se explicar pela deficiência da modelagem ou pela baixa confiabilidade do mapa convencional em

relação aos segundos componentes das associações e das inclusões conforme já exposto. Considerando o último caso, a classe RL que é considerado como inclusão na unidade Ca32, tornaria um segundo componente, de acordo com os mapas preditos.

4.6. CONCLUSÕES

Os resultados da validação cruzada mostraram que os algoritmos classificadores MLP, NaiveBayes, SimpleCart e J48 apresentaram desempenhos de predição semelhantes. Isto indica que se deve dar preferência à utilização das árvores de decisão, considerando-se o interesse por modelos que possibilitem o mais fácil entendimento das relações entre os atributos preditores e as classes de solos.

A redução da quantidade de classes de saída diminuiu a confusão entre unidades de mapeamento, o que em geral melhorou o desempenho de predição dos classificadores MLP, SimpleCart e J48.

A seleção de variáveis preditoras pelo algoritmo ReliefF resultou em menor desempenho de predição pela rede neural multilayer perceptron e pela classificador de NaiveBayes. Diferentemente, a redução das variáveis preditoras ambientais não causou efeito negativo no desempenho de predição das árvores de decisão na predição de classes de solos, o que permite a simplificação dos modelos gerados por esses classificadores.

4.7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALKIMIM, A.F. **Geoambientes, morfometria e solos da bacia do rio Benevente, ES.** 2009. Dissertação (Mestrado em Agronomia (Solos e Nutrição de Plantas)) - Universidade Federal de Viçosa. 109p.
- BRASIL. Ministério das Minas e Energia. Secretaria Geral. **Folhas SF 23/24 Rio de Janeiro/Vitória: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra / Projeto RADAMBRASIL.** Rio de Janeiro,1983. 775p.
- CARVALHO JUNIOR, W.;CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.G.; BHERING, S.B. &FRANCELINO, M.R. Digital soilscape mapping of tropical hillslope areas by neural

- networks. **Scientia Agricola**, 68:691-696, 2011. Disponível em: <http://www.scielo.br/pdf/sa/v68n6/v68n6a14.pdf>. Acesso em: 28 abr. 2013. doi: <http://dx.doi.org/10.1590/S0103-90162011000600014>.
- CASTRO, F. da S.; PEZZOPANE, J.R.M.; CECÍLIO, R.A. & PEZZOPANE, J.E.M. Uso de imagens de radar na espacialização da temperatura do ar. *Idesia*, 28:69-79, 2010.
- CHAGAS, C. S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. Viçosa, Universidade Federal de Viçosa, 2006. 223p. (Tese de Doutorado)
- CHAGAS, C.S.; CARVALHO JUNIOR, W. & BHERING, S.B. Integração de dados do quickbird e atributos do terreno no mapeamento digital de solos por redes neurais artificiais. **Revista Brasileira de Ciência do Solo**, 35:693-704, 2011.
- CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.R. & CARVALHO JUNIOR, W. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. **Pesquisa Agropecuária Brasileira**, 45:497-507, 2010.
- COELHO, F.F. & GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, 40:2099-2106, 2010. Disponível em: <http://www.scielo.br/pdf/cr/v40n10/a728cr2677.pdf>. Acesso em: 04 abr. 2013. doi: 10.1590/S0103-84782010005000156.
- CONGALTON, R.G. & GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2.ed. Boca Raton, Taylor and Francis, 2009.
- CPRM. Mapa Geológico do Estado do Espírito Santo. Disponível em: <http://www.cprm.gov.br/>>. Acesso em 06 de fevereiro de 2012.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 2. ed. Rio de Janeiro, 2006. 306p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Serviço Nacional de Levantamento e Conservação do Solo. **Levantamento de reconhecimento dos solos do Estado do Espírito Santo**. Rio de Janeiro. 1978. 461p. (Boletim Técnico)
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - ESRI. Software ArcGis Desktop, version 10. 2011.

- FERREIRA NETO, J.A. & LANI, J.L. **Estudos visando o licenciamento ambiental de assentamentos rurais do Estado do Espírito Santo**. Projeto de Assentamento Monte Alegre. Viçosa, MG. NEPUT/UFV. 214p. 2009.
- GIASSON, E.; SARMENTO, E.C.; WEBER, E.; FLORES, C.A. & HASENACK, H. Decision trees analysis for digital soil mapping on subtropical basaltic steeplands. **Scientia Agricola**, 68:167-174, 2011.
- HALL, M. & HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. **IEEE Transactions on Knowledge and Data Engineering**, 15:1437–1447, 2003.
- HAN, J. & KAMBER, M. **Data mining: concepts and techniques**. 2.ed. San Francisco, Morgan Kaufmann, 2006. 770p.
- INSTITUTO ESTADUAL DE MEIO AMBIENTE E RECURSOS HÍDRICOS – IEMA. Ortofotomosaico IEMA 2007/2008. Disponível em: <http://www.meioambiente.es.gov.br/default.asp>. Acesso em 02/05/2013.
- INSTITUTO JONES DOS SANTOS NEVES - IJSN. Mapeamento geomorfológico do Estado do Espírito Santo. Vitória, ES, 2012, 19f.: il. (Nota técnica, 28)
- KONONENKO, I. Estimating attributes: Analysis and extensions of relief. In: Proceedings of the seventh European Conference on Machine Learning, pages 171-182. Springer-Verlag, 1994.
- MCBRATNEY, A. B.; MENDONÇA-SANTOS, M. L. & MINASNY, B. On digital soil mapping. **Geoderma**, 117:3–52, 2003.
- QI, F. & ZHU, A. X. Knowledge discovery from soil maps using inductive learning. **International Journal of Geographical Information Science**. 17.:771-795, 2003. Disponível em: <http://www.tandfonline.com/doi/pdf/10.1080/13658810310001596049>. Acesso em: 08 abr. 2013. doi: 10.1080/13658810310001596049.
- ROLIM, G. S.; SENTELHAS, P. C. & BARBIERI, V. Planilhas no ambiente EXCEL para os cálculos de balanços hídricos: normal, sequencial, de cultura e de produtividade real e potencial. **Revista Brasileira de Agrometeorologia**, 6:133-137, 1998.
- SABINS F. F. **Remote Sensing Principles and Interpretation**. 2ed., New York, W.H. Freeman & Co, 1987.

- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H.C. **Manual de descrição e coleta de solo no campo**. 5 ed. revisada e ampliada. Viçosa: Sociedade Brasileira de Ciência do Solo, 100 p. 2005.
- SARMENTO, E. C. Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil. Porto Alegre, RS, Universidade Federal do Rio Grande do Sul. 2010. (Tese de Doutorado).
- SARMENTO, E.C.; GIASSON, E.; WEBER, E.; FLORES, C.A. & HASENACK, H. Prediction of soil orders with high spatial resolution: response of different classifiers to sampling density. **Pesquisa Agropecuária Brasileira**, 47:1395-1403, 2012.
- SHATAR, T.M. & MCBRATNEY, A.B. Empirical modelling of relationships between sorghum yield and soil properties. *Precision Agriculture*, 1:249– 276, 1999.
- TEN CATEN, A.; DALMOLIN, R.S.D. & RUIZ, L.F.C. Digital soil mapping: strategy for data processing. **Revista Brasileira de Ciência do Solo**, 36:1.083-1.092, 2012.
- TEN CATEN, A.; DALMOLIN, R.S.D.; MENDONÇA-SANTOS, M.L. & GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, 42:1990-1997, 2012.
- THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical Review**, v.38, p.55-94, 1948.
- WITTEN, I.H. & FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed. San Francisco: Morgan Kaufmann, 2005. 525p.
- YANG, Q.; SHAO, J.; SCHOLZ, M. & PLANT, C. Feature selection methods for characterizing and classifying adaptive Sustainable Flood Retention Basins. **Water Research**, 45:993–1004, 2011.
- ZHU, A.X.; HUDSON, B.; BURT, J.; LUBICH, K. & SIMONSON, D. Digital soil mapping using GIS, expert knowledge, and fuzzy logic. **Soil Science Society American Journal**, 65:1.463-1.472, 2001.

5. CAPÍTULO 4 - ATRIBUTOS DO TERRENO, CLIMÁTICOS E BALANÇO HÍDRICO NO MAPEAMENTO DIGITAL DE SOLOS DO ASSENTAMENTO RURAL SEZÍNIO FERNANDES, LINHARES, ESPÍRITO SANTO

5.1. RESUMO

O potencial de predição das variáveis e a facilidade de entendimento de suas relações com as classes de solos são importantes fatores considerados no mapeamento digital. Nesse aspecto, o objetivo deste trabalho foi avaliar variáveis climáticas, do balanço hídrico e atributos do terreno no mapeamento digital detalhado de solos. A área de estudo foi a do Assentamento Sezínio Fernandes, no município de Linhares, Espírito Santo. Foram utilizados algoritmos de árvore de decisão, denominados por SimpleCart e J48, rede neural artificial Multilayerperceptron e NaiveBayes. Dezenove variáveis preditoras foram utilizadas, entre elas diferentes atributos do terreno, climáticos e do balanço hídrico. Um subgrupo de variáveis, selecionadas por algoritmo baseado em correlação (CFS), foi comparado a outro composto por atributos do terreno. Os classificadores utilizados foram avaliados pelos resultados da validação cruzada e por comparação a um mapa de solo delineado manualmente. Os subconjuntos de atributos selecionados envolveram tanto variáveis climáticas e do balanço hídrico quanto os atributos do terreno. Nenhuma diferença nas exatidões das predições foi encontrada ao comparar os dois subconjuntos de variáveis e os quatro algoritmos classificadores. Os índices Kappa alcançaram valores em torno de 0,800. Embora as variáveis climáticas e do balanço hídrico tenham sido selecionadas, estas provavelmente não apresentam amplitudes de valores capazes de propiciar condições pedogenéticas diferenciadas na área de estudo. As maiores concordâncias com o mapa convencional foram obtidas para o algoritmo J48, seguido do NaiveBayes e do SimpleCart, todos com a utilização dos subconjuntos contendo somente atributos de terreno. Os valores de Kappa para estes três classificadores foram de 0,903, 0,889 e 0,840, respectivamente.

Termos para indexação: variáveis preditoras ambientais, redes neurais, árvore de decisão.

ATTRIBUTES OF TERRAIN, CLIMATE AND WATER BALANCE IN DIGITAL SOIL MAPPING IN SEZÍNIO FERNANDES RURAL SETTLEMENT, LINHARES, ES

5.2. ABSTRACT

The predictive potential of variables and the ease of understanding of their relationships with soil classes are considered important factors in digital mapping. In this respect, the aim of this study was to evaluate climatic and water balance variables and terrain attributes in detailed digital soil mapping. The study area was the Sezínio Fernandes Settlement, in Linhares, Espírito Santo. Decision tree, called by SimpleCart and J48, artificial neural network Multilayerperceptron and NaiveBayes algorithms were used. Nineteen predictor variables were used, including different terrain attributes, climate and water balance. A subset of variables selected by algorithm based on correlation was compared to another compound by terrain attributes. The classifiers used were assessed by results of cross-validation and by comparison to a soil map manually delineated. Subsets of selected attributes involved both climatic and water balance variables as many as the terrain attributes. No difference in the accuracies of the predictions was found when comparing the two subsets of variables and the four inductors. The Kappa coefficients reached values around 0.800. The greatest agreement with the conventional map was obtained for the algorithm J48, followed by NaiveBayes and SimpleCart, all with the use of subassemblies containing only terrain attributes. The kappa values for these three classifiers were 0.903, 0.889 and 0.840, respectively.

Index terms: environmental predictors, neural networks, decision tree.

5.3. INTRODUÇÃO

Atributos do terreno têm sido utilizados como as principais variáveis preditivas ambientais no mapeamento digital de solos. Em parte, isso se relaciona à grande capacidade de predição que estes atributos propiciam ao se utilizar os diferentes algoritmos de classificação disponíveis para o mapeamento digital de solos (McBratney et al., 2003; Ten Caten et al., 2012). Especialmente, a maior facilidade de se entender e de se estabelecer as relações entre esses atributos e os processos pedogenéticos e, por conseguinte às classes de solos. Sabe-se que os levantamentos são baseados em fatores

pedogenéticos e narelação solo-paisagem (Souza Junior & Demattê, 2008).

A maioria dos estudos sobre mapeamento digital de solos tem sido direcionada ao mapeamento detalhado (McBratney et al., 2003; Ten Caten et al., 2012). Isto provavelmente em resposta à grande demanda por este tipo de levantamento no Brasil (Silveira et al., 2013) e em muitos outros países. Em geral, nesses estudos detalhados os atributos do terreno tem permitido exatidões satisfatórias nas predições, inclusive nas comparações com mapas convencionais (Ten Caten et al., 2012).

Outras variáveis preditivas ambientais relacionadas aos processos pedogenéticos, como as climáticas têm sido utilizadas com maior restrição. Isto decorre em razão da falta de informações climáticas espacializadas em maior nível de detalhes, a qual está relacionada com a baixa densidade de postos pluviométricos e/ou de estações climatológicas (Hu, 2010). Outros fatores são as diferenças de precipitação pluvial e de temperaturas entre as condições de clima atual e as do passado (Ten Caten et al., 2012). Além desses fatos, como as escalas de trabalho em mapeamentos detalhados de solos são mais locais, seria de esperar que as variáveis climáticas e do balanço hídrico poderia não ser tão eficazes como preditivas ambientais de classes de solos.

Em situações de áreas maiores o clima pode auxiliar outras variáveis preditivas. O relevo, por exemplo, exerce maior influência sobre as características dos solos quando se considera pequenas áreas, ao passo que a variações em razão do clima são muito menores. Por outro lado, quando a área de estudo é mais extensa, o clima pode exercer influência mais discernível sobre a formação do solo, e tende a dominar o poder de influência dos atributos do terreno (Gonzalez et al, 2008). Baseado no exposto, o propósito foi avaliar a utilização de algumas variáveis climáticas, balanço hídrico e atributos do terreno no mapeamento digital detalhado de solos, com a utilização dos algoritmos de classificação de árvore de decisão, denominados por SimpleCart e J48, e por rede neural artificial Multilayerperceptron.

5.4. MATERIAL E MÉTODOS

O estudo foi desenvolvido no Assentamento Rural Sezínio Fernandes. A área total é de 20,806 km². Situa-se no município de Linhares, Estado do Espírito Santo, entre as coordenadas UTM 7843649 e 7849049 m N e 361485 e 367883 m E, zona 24S (Figura 1). O tipo de clima é Aw, conforme classificação de Köppen, com estação seca no

inverno e verão quente e chuvoso.

A geomorfologia é a de Tabuleiros Costeiros e o relevo predominante é de plano a ondulado. Os solos são formados de sedimentos do Terciário (Formação Barreiras) e Quaternários (fluviais) (Brasil, 1983).

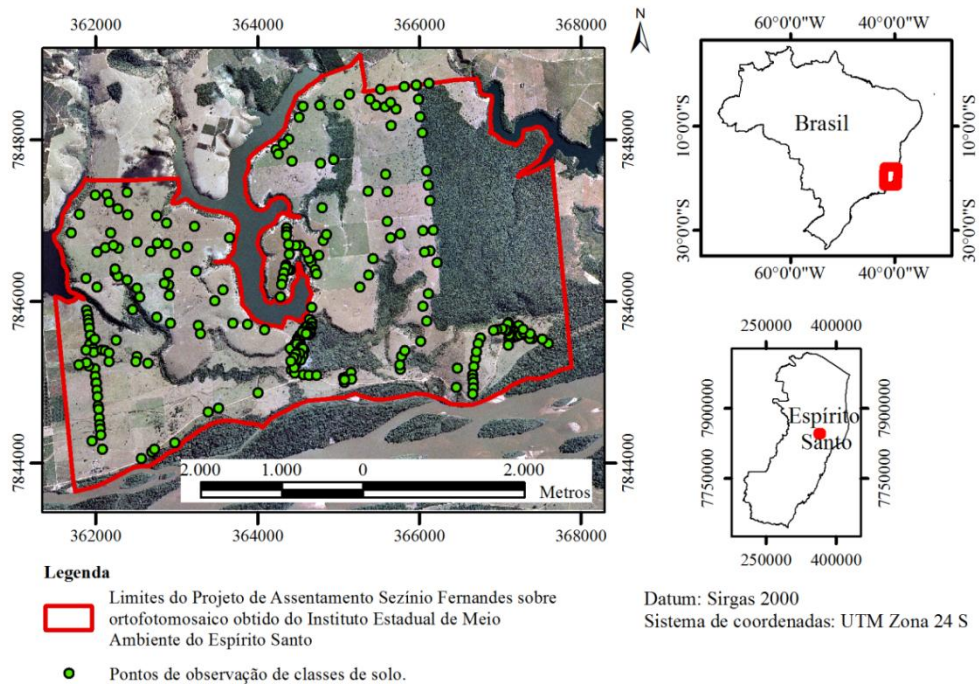


Figura 1. Localização da área de estudo: Assentamento Sezínio Fernandes, Linhares, ES.

A área do Assentamento foi percorrida e obtidas 259 observações, entre coletas de pontos de coordenadas (GPS de navegação, modelo Garmin, GPSMAP 76). Amostras de solos foram coletadas em 10 (dez) pontos selecionados previamente, nos horizontes A e B, para análises físicas e químicas. Nestes locais foram feitas as descrições morfológicas (Santos et al., 2005). As amostras foram submetidas à análise granulométrica (Embrapa, 1997, modificado por Ruiz, 2005) e às seguintes análises químicas, conforme Embrapa (1997): pH em H₂O e em KCl; bases trocáveis (Ca²⁺, Mg²⁺, K⁺, Na⁺); soma de bases (valor S); acidez; capacidade de troca de cátions (valor T); saturação por bases (valor V); saturação por Al e P disponível, e; C orgânico.

Os solos identificados na área de estudo e o número de amostras para treinamento dos modelos para árvore de decisão encontram-se no Quadro 1 (Embrapa, 2006)

No mapeamento digital de solos utilizaram-se dezenove variáveis discriminantes. Dentre elas, 12 (doze) foram relacionadas a atributos do terreno e 7 (sete) climáticas

(Quadro 2). As variáveis relacionadas aos atributos do terreno foram derivadas de modelo digital de elevação (MDE), com 10 m de resolução espacial. O MDE foi obtido com a opção “topo to raster” do programa ArcGIS 10.0 (ESRI, 2011) a partir de 7.315 pontos com cotas altimétricas, obtidas do Instituto Nacional de Colonização e Reforma Agrária – INCRA. As cotas altimétricas foram calculadas por nivelamento trigonométrico, utilizando-se duas estações totais, uma com precisão angular de 5” e linear de 3 mm +2 ppm e outra com precisão angular de 2” e linear de 2 mm +2 ppm. A área do Assentamento com existência de levantamento topográfico não abrangeu as áreas de mata, motivo pelo qual se limitou a área de estudo a 15,960 km².

Quadro 1. Classes de solos identificadas no Assentamento Sezínio Fernandes e número de instâncias para treinamento dos modelos por árvore de decisão

Unidade de mapeamento	Classe de solo identificada	Amostras (un.)
LAd	LATOSSOLO AMARELO Distrocoeso	104
CYbe + CYbd	CAMBISSOLO FLÚVICO Tb Eutrófico + CAMBISSOLO FLÚVICO Tb Distrófico	81
GXbd	GLEISSOLO HÁPLICO Tb Distrófico	74

Os dados climáticos, cujas escalas são regionais por natureza, foram propositalmente utilizados como variáveis discriminantes. Utilizaram-se as médias anuais da precipitação pluviométrica, e da temperatura do ar, e a radiação solar global calculada para o ano de 2012 com o uso de intervalos mensais. A espacialização da precipitação pluviométrica foi realizada por meio de cokrigagem dos dados dos postos pluviométricos localizados no Estado e áreas vizinhas, obtidos no sítio <http://hidroweb.ana.gov.br/> e junto ao Instituto Capixaba de Pesquisa e Extensão Rural - INCAPER. A cokrigagem foi realizada com dados de altitude (Shuttle Radar Topographic Mission – SRTM) e distância do mar.

Os dados de temperatura média do ar foram obtidos por meio de modelagem a partir de dados obtidos em estações meteorológicas distribuídas no Estado conforme equações de regressão desenvolvidas por Castro et al. (2010) para permitir estimativas de temperatura em função da altitude, latitude e longitude.

A partir dos dados de precipitação e da estimativa da temperatura média do ar para todos os locais dos postos, calculou-se a deficiência hídrica (DEF), evapotranspiração potencial (ETP), evapotranspiração real (ER), excedente hídrico (EXC) e disponibilidade hídrica (ER/ETP), pelo método proposto por Thornthwaite & Mather (1955). Os cálculos foram realizados com o auxílio do programa "BHnorm", elaborado

em planilha Excel por Rolim et al. (1998), assumindo-se uma Capacidade Máxima de Armazenamento de água no solo (CAD) igual a 100 mm. As superfícies de dados climáticos e de balanço hídrico foram gerados com resolução espacial de 30 metros.

Quadro 2. Variáveis preditivas ambientais utilizadas no mapeamento digital de solos do Assentamento Sezínio Fernandes, Linhares, Espírito Santo

Variável e unidade	Referência	Fator ambiental	Tipo	Média/ Intervalo	Amplitude
1. Excedente hídrico (mm)		*	Q	83,3 / 76,3 – 87,8	11,5
2. Evapotranspiração real		c	Q	1.134 / 1.129 – 1.139	10,0
3. Evapotranspiração potencial (mm)		c	Q	1.374 / 1.373 – 1.376	3
4. Deficiência hídrica (mm)		c	Q	231 / 222 – 237	15
5. Índice topográfico combinado	(Moore et al., 1993)	r	Q	8 / 2 – 15	13
6. Curvatura (°/m)	Moore & Grayson (1991)	r	C	3	
7. Dissecação de Martonne modificado	(Evans, 1972)	r	Q	0,45 / 0 – 1	1,0
8. Altitude (m)		r	Q	49,2 / 20,6 – 134,6	114,0
9. “Northness”	(Alexander et al., 2006)	r	Q	-0,06 / -1 - 1	2,0
10. Perfil de curvatura (°/m)	Moore et al., (1991)	r	C	3	
11. Plano de curvatura (°/m)	Moore et al., (1991)	r	C	3	
12. Precipitação pluvial média		c	Q	1.222 / 1.207 – 1.228	21,0
13. Radiação solar (MWh/m ²)	Fu & Rich (2002)	c	Q	1,7 / 1,3 - 1,8	0,5
14. Índice de Rugosidade do terreno (m)	(Riley et al. 1999)	r	Q	0,7 / 0 – 2,8	2,8
15. Declividade (°)	Burrough & McDonell (1998)	r	Q	4,7 / 0 - 38	38
16. Razão de elevação do relevo	(Pike & Wilson, 1971)	r	Q	0,48 / 0 - 0,9	0,9
17. Temperatura média anual (°C)	Castro (2010)	c	Q	24,9 / 24,3 – 25,1	0,8
18. Índice de corrente máximo de fluxo	(Moore et al., 1993)	r	Q	3,3 / -2,4 – 8,8	11,2
19. Índice capacidade de transporte de sedimentos	(Moore & Wilson, 1992)	r	Q	4,7 / 0 - 65,7	65,7

* Fator ambiental que melhor caracteriza as variáveis, considerando-se o modelo scorpan de McBratney et al. (2003), em que s = propriedade do solo em um ponto; c = clima; o = organismos; r = topografia; p = material originário; a = idade; n = posição espacial. C – Classe; Q – quantitativo.

A quantificação da variável radiação solar global foi realizada por meio da ferramenta de análise “área solar radiation” disponível no software ArcGis 10, a qual considera como dados de entrada o MDE.

Os valores das variáveis intervalares foram reescalados entre o intervalo 0 e 1. Este procedimento foi realizado com a ferramenta “raster calculator” do ArcGis 10.0, utilizando a expressão $x' = (x - x_{min}) / (x_{max} - x_{min})$, onde x é o valor da variável, x_{min} e x_{max} são respectivamente o menor e o maior valor que esta variável assume, e x' o novo

valor de x , normalizado. Nas variáveis com distribuição simétrica que apresentaram valores excessivamente elevados em um número muito reduzido de células (ruídos e outliers), os mesmos foram substituídos pelos valores correspondentes à média aritmética \pm três vezes o desvio-padrão. Para as variáveis com distribuição assimétrica os valores foram substituídos por aqueles correspondentes a 0,5 % dos maiores valores.

As variáveis preditivas utilizadas para as classes de solos foram selecionadas no software WEKA, Versão 3.6.8 (Witten & Frank, 2005), pelo algoritmo de seleção de atributos baseado em correlação, denominado “correlation-based feature selection” - “CFSSubsetEval” (Hall, 1998). Este algoritmo seleciona subconjuntos de atributos altamente correlacionados com a classe, mas os atributos têm baixa correlação entre eles. O método de busca utilizado foi o “SubsetSizeForwardSelection”, o qual foi ajustado com parâmetros padrão do software (-I -K 20 -T 0 -F 5 -S 1 -Z false -E) e a avaliação realizada sobre o conjunto de dados de treinamento. Para fins de comparação foi utilizado um conjunto de variáveis que continha exclusivamente atributos do terreno. Um ranqueamento dos atributos foi realizado por meio do algoritmo ReliefF (Kononenko, 1994) para se conhecer aqueles mais relevantes para a classificação.

As predições de classes de solos foram realizadas com a utilização de quatro algoritmos de aprendizagem disponíveis no software WEKA, Versão 3.6.8 (Witten & Frank, 2005): SimpleCart e J48 (classificadores em árvore de decisão), multilayer perceptron - MLP (rede neural artificial) e NaiveBayes NaiveBayes (classificador probabilístico baseado em inferência bayesiana). Cada um dos três classificadores foi aplicado a dois conjuntos de variáveis preditivas ambientais: um com variáveis selecionadas por meio da aplicação do algoritmo de seleção e outro com variáveis constituídas de atributos do terreno.

Diferentes arquiteturas foram testadas para utilização da rede neural (MLP). Utilizou-se a interface do Weka denominada “Experimenter”, na qual se configuram experimentos que executam algoritmos com diferentes parâmetros e conjuntos de dados e permite coletar estatística de desempenho e realizar teste T pareado ao nível de significância de 5%. Variou-se o número de neurônios na camada interna correspondente à quantidade de classes e ao número de classes mais a quantidade de variáveis ambientais utilizadas. Avaliou-se também diferentes valores para a taxa de aprendizagem (0,025 a 0,3) e para o momentum (0,1 a 1,0), todos contendo o mesmo número de neurônios na camada de saída, correspondente ao número de unidades de

mapeamento (10). Após essa avaliação, utilizou-se a interface “explorer” do Weka 3.6.8. Foram realizadas as classificações usando-se 8 e 15 neurônios nas camadas internas para os conjuntos de variáveis selecionadas e de atributos do terreno. Foram utilizadas taxas de aprendizagem de 0,3 e “momentum” de 0,2. Foram considerados 10 mil ciclos de aprendizagem. Os parâmetros foram avaliados com base no valor do índice de concordância Kappa apresentado com base nas classes preditas e observadas para o treinamento.

Para realizar as classificações pelo algoritmo SimpleCart e J48 foram previamente testadas variações no número mínimo de instâncias por folha (2 a 10). Após o teste, decidiu-se pela utilização do valor 2, em razão de ser o valor “default” e de que não houve diferenças significativas entre este e os demais valores testados. O método de poda da árvore utilizado gerado pelo SimpleCart foi o “minimal cost-complexity pruning”. Para o algoritmo J48, a poda foi desenvolvida (unpruned = false). O método de poda “reducedErrorPruning” não foi selecionado.

A avaliação dos resultados foi realizada por meio do índice de concordância Kappa, derivado de uma matriz de confusão (Congalton & Green, 2009). Essa medida estatística foi obtida por meio da validação cruzada, considerando o número de “folds” (subconjuntos de dados) igual a 10, onde cada um destes “folds” é utilizado para teste e o restante para treinamento.

A generalização, ou seja, a classificação das amostras que não foram utilizadas para treinamento foi realizada no software WEKA 3.6.8, com base nos modelos gerados. Essas amostras classificadas foram implementadas no ArcGis para gerar as imagens raster que se constituíram nos mapas de solos.

Antes do processo de mapeamento digital da área, foi elaborado um mapa de solos na escala 1:20.000 com a utilização de delineamento manual das unidades de mapeamento, buscando-se seguir a metodologia tradicional de levantamento (Embrapa, 1995). Na ausência de pares estereoscópicos de aerofotos, os limites das unidades de mapeamento foram delineados sobre um ortofotomosaico digital na escala 1:15.000 PEC “A”, de resolução espacial de 1 metro, elaborado a partir de um levantamento aerofotogramétrico na escala 1:35.000, realizado em maio de 2008. O ortofotomosaico foi obtido do Instituto Estadual de Meio Ambiente do Espírito Santo - IEMA. Para auxiliar o delineamento foram utilizadas curvas de nível extraídas do MDE de 10 m de resolução espacial a partir de pontos cotados, conforme descrito anteriormente. Em face

da inexistência de mapas de solos detalhados para a área de estudo, a finalidade desse mapeamento foi proporcionar uma comparação entre o mapeamento gerado pelos três classificadores e um mapa de solos de referência.

5.5. RESULTADOS E DISCUSSÃO

A seleção de atributos visa obter subgrupos com uma quantidade reduzida de variáveis relevantes sem que ocorra a perda na predição realizada pelos diferentes algoritmos de classificação, em comparação à utilização do grupo original de variáveis (Guyon & Elisseeff, 2003). Nos três subconjuntos de variáveis preditivas selecionadas, correspondentes àquelas mais relevantes a cada algoritmo classificador (SimpleCart, J48, MLP e NaiveBayes) estão envolvidas tanto as variáveis climáticas, quanto às do balanço hídrico e dos atributos do terreno (Quadro 3). Embora algumas das variáveis climáticas e do balanço hídrico tenham sido selecionadas para predição, suas aplicações carecem de fundamentação quanto à relação destas com as classes de solos. Conforme Gonzalez et al. (2008), em uma escala regional seria plausível a relevância dessas variáveis, uma vez que as variações de clima de uma área mais extensa podem apresentar influência mais discernível sobre a formação dos solos. No entanto, em pequenas áreas como a de estudo, as amplitudes de variações de clima no espaço são geralmente muito pequenas, o que não justificaria tais resultados.

As amplitudes dos valores encontrados para variáveis como excedente hídrico, evapotranspiração real, evapotranspiração potencial e deficiência hídrica foram de 11, 10, 3 e 15, respectivamente. Ao considerar toda a área do Estado do Espírito Santo, as amplitudes dos valores dos elementos do balanço hídrico mostraram-se muito mais elevadas (Cecílio et al., 2012). Este fato sugere que os valores encontrados para as amplitudes da área do Assentamento Sezínio Fernandes são pequenas demais para produzir efeitos pedogenéticos diferenciados e conseqüentemente se relacionarem com as diferentes classes de solos da área. Nesse sentido, observou-se que as variáveis climáticas e do balanço hídrico apresentaram comportamento diferenciado dos atributos do terreno quanto às relações entre os valores de amplitudes e as médias dos valores assumidos por essas variáveis (Quadro 2). Em geral, as variáveis de clima e do balanço hídrico apresentam amplitudes muito inferiores aos valores das médias, diferindo-se dos atributos do terreno, os quais apresentaram amplitudes mais elevadas que as médias.

O atributo mais relevante para a predição de classes de solos, dentre aqueles

selecionados pelo algoritmo CFS, foi a temperatura média anual (Quadro 3). Considerando-se a análise de todos os 19 atributos, a temperatura foi menos relevante apenas em relação à altitude, o que indica a semelhança de desempenho destes dois atributos em discriminar as classes de solos da área estudada. Essa relação com a altitude pode explicar em parte essa relevância da temperatura média do ar. Isto pode ter relação com o método de obtenção dessas variáveis, as quais foram obtidas por meio de modelagem. A temperatura, nesse caso, obtida por equações de regressão (Castro et al., 2010) está relacionada com a altitude em razão da utilização desta última como variável preditiva da primeira. Para as variáveis do balanço hídrico, que por natureza são de escalas regionais, a resolução espacial das imagens resultantes da interpolação não representa a real variabilidade das mesmas na área. Provavelmente a relevância se deva ao acaso.

Os resultados da seleção (Quadro 3) indicam que as inclusões das variáveis climáticas e do balanço hídrico são capazes de proporcionar previsões com o mesmo nível de índice Kappa que se produz ao utilizar os atributos do terreno (Quadro 4). A elevada exatidão da previsão, mostrada pelos altos valores de Kappa, pode ser atribuída à distribuição espacial pouco complexa das classes de solos na área e a pouca variabilidade do material de origem, restritos a materiais sedimentares aluvionares e da Formação Barreiras. Os valores de Kappa apresentados, em torno de 0,800, estão acima daqueles geralmente encontrados em outros estudos sobre mapeamento digital de classes de solos (Ten Caten et al., 2012), sejam obtidos por dados de validação de campo ou por comparação a mapas convencionais de referência.

Ao considerar os valores de Kappa, não houve diferenças significativas a 5% de probabilidade pelo teste Z entre os quatro algoritmos de classificação avaliados e nem ao se comparar a utilização de subconjuntos de variáveis selecionadas com subconjunto contendo somente atributos do terreno. Embora não tenha havido diferenças significativas entre os métodos, a vantagem no uso das árvores de decisão, conforme Giasson et al. (2011), está na possibilidade da análise das regras de decisão. Isto permite explicitar as relações entre as classes de solos e as variáveis preditivas ambientais utilizadas na modelagem. Menores tamanhos de árvores de decisão (números de nós) e com menores números de folhas (nó que indica uma classe), conforme Bui & Moran (2003), representam modelos mais simples e de mais fácil entendimento. Nesse aspecto, o algoritmo SimpleCart apresentou melhor desempenho que o J48.

Quadro 3. Atributos selecionados pelo algoritmo baseado em correlação – *CFSSubsetEval* e ranking da relevância dos atributos obtidos pela aplicação do algoritmo *ReliefF*

SimpleCart	Ranking* dos atributos selecionados pelo CFS			Ranking dos 19 atributos pelo ReliefF
	J48	MLP	NaiveBayes	
Temperatura média anual; Índice de corrente máximo de fluxo; Dissecação de Martonne; Razão de elevação de relevo; Deficiência hídrica, e; Excedente hídrico.	Temperatura média anual; Excedente hídrico; Índice de corrente máximo de fluxo; Razão de elevação de relevo; Dissecação de Martonne; Deficiência hídrica; Índice de rugosidade, e; Perfil de curvatura.	Temperatura média anual; Razão de elevação de relevo; Excedente hídrico; Radiação solar, e; Índice de capacidade de transporte de sedimentos.	Temperatura média anual; Razão de elevação de relevo; Deficiência hídrica; Radiação solar, e; Índice de capacidade de transporte de sedimentos.	Altitude; Temperatura média anual; Evapotranspiração potencial; Plano de curvatura; Dissecação de Martonne; Excedente hídrico; Razão de elevação de relevo; Índice de capacidade máxima de fluxo; Evapotranspiração real; Perfil de curvatura; Curvatura; Deficiência hídrica; Northness; Índice de rugosidade; Declividade; Precipitação anual média; Radiação solar; Índice topográfico combinado, e ; Índice de capacidade de transporte de sedimentos.

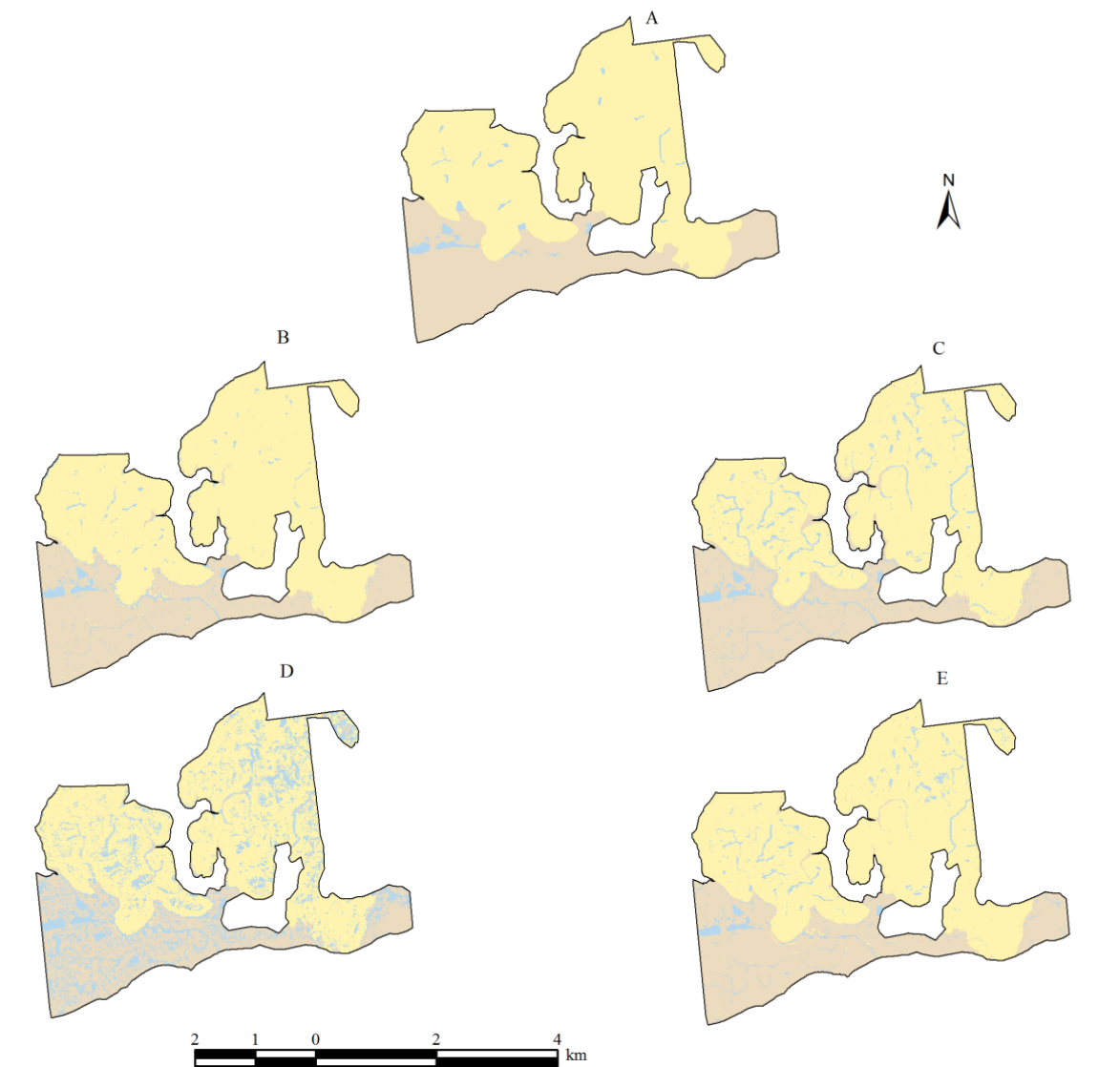
Nota: ranking de cima para baixo, com os mais relevantes na primeira linha do Quadro. * Ranking obtido pela aplicação do algoritmo ReliefF aos atributos previamente selecionados pelo algoritmo CFS.

Quadro 4. Matriz de significância dos índices Kappa, Variância do Kappa (vâr), número de folhas e tamanho da árvore (nº de nós) obtidos dos resultados da validação cruzada ao se utilizar os algoritmos SimpleCart, J48, MLP e NaiveBayes com dois subconjuntos de variáveis ambientais preditivas

Método/ atributos	CARTCFS	J48CFS	MLPCFS	BayesCFS	CARTAT	J48AT	MLPAT	BayesAT
Nós (und)	7	9			7	12		
Folhas (und)	4	5			4	7		
Kappa	0,805	0,792	0,822	0,811	0,804	0,775	0,753	0,822
Vâr (x 10 ⁻⁴)	9,64	10,23	8,97	9,50	9,69	10,93	11,87	9,03
CARTCFS	0,00	-0,28	0,40	0,14	-0,01	-0,67	-1,11	0,40
J48CFS		0,00	0,68	0,42	-0,27	-0,39	-0,83	0,68
MLPCFS			0,00	-0,26	-0,41	-1,06	-1,51	0,00
BayesCFS				0,00	-0,15	-0,80	-1,25	0,26
CARTAT					0,00	-0,66	-1,10	0,41
J48AT						0,00	-0,45	1,06
MLPAT							0,00	1,51
BayesAT								0,00

CARTCFS – SimpleCart e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; J48CFS – J48 e atributos selecionados pelo algoritmo de seleção CFSSubsetEval; MLPCFS - MLP e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; Bayes – NaiveBayes e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; CARTAT - SimpleCart e atributos do terreno; J48AT – J48 e atributos do terreno; MLPAT – MLP e atributos do terreno; BayesAT – NaiveBayes e atributos do terreno. * significativo a 5% de probabilidade pelo teste Z.

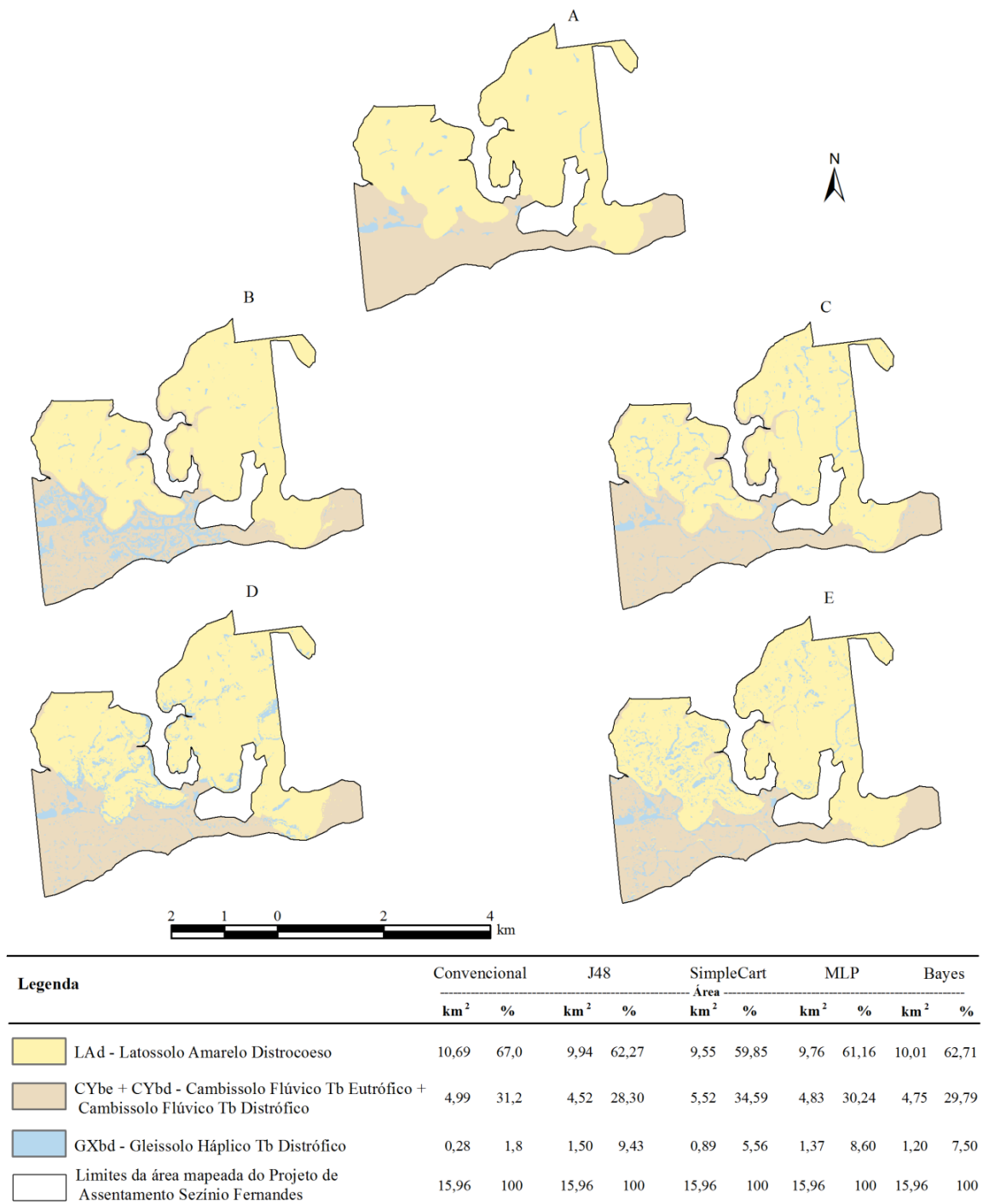
Baseado numa comparação visual dos mapas (Figura 2), considerando-se uma avaliação mais qualitativa, os mapas preditos (estimados) por árvores de decisão apresentaram forte semelhança entre si e entre estes e o mapa produzido pelo delineamento das unidades de mapeamento de forma manual. O mapa predito por MLP apresentou maior quantidade de área com Gleissolo que as demais, aparentemente representando melhor detalhamento, embora não confirmado pelos valores dos índices Kappa sem diferenças estatísticas verificadas. Quantitativamente, o mapa gerado por MLP apresentou área de Gleissolo mais de dez vezes maior que o encontrado no mapa convencional. Quantidades de áreas de Gleissolos mais próximas às encontradas no mapa convencional foram apresentadas pelos mapas estimados pelos algoritmos J48, SimpleCart e NaiveBayes. Estes resultados indicam uma provável superestimação de Gleissolos pela rede neural MLP, sobretudo nas áreas onde predominam os Latossolos Amarelos. Uma tendência de menor ocorrência de Gleissolos associados aos Latossolos Amarelos pode ser verificada ao se observar os mapas estimados pelos outros algoritmos de classificação, tanto com a utilização dos atributos do terreno quanto dos atributos selecionados pelo algoritmo de seleção CFS, que envolveu também os atributos climáticos como variáveis preditivas (Figura 3).



Legenda	Convencional		J48		SimpleCart		MLP		Bayes	
	km ²	%	km ²	%	km ²	%	km ²	%	km ²	%
LAd - Latossolo Amarelo Distrófico	10,69	67,0	10,51	65,9	9,79	61,3	8,95	56,1	10,36	64,9
CYbe + CYbd - Cambissolo Flúvico Tb Eutrófico + Cambissolo Flúvico Tb Distrófico	4,99	31,2	4,91	30,7	5,27	33,0	3,96	24,8	4,92	30,8
GXbd - Gleissolo Háplico Tb Distrófico	0,28	1,8	0,54	3,4	0,90	5,7	3,05	19,1	0,68	4,3
Limites da área mapeada do Projeto de Assentamento Sezinio Fernandes	15,96	100	15,96	100	15,96	100	15,96	100	15,96	100

J48 - Algoritmo J48; SimpleCart - algoritmo SimpleCart; MLP - algoritmo de redes neurais Multilayerperceptron; Bayes - algoritmo NaiveBayes.

Figura 2. Mapas convencional (A) e o elaborado por meio dos algoritmos J48 (B), SimpleCart (C), MLP (D) e NaiveBayes (E), utilizando-se atributos do terreno como variáveis preditivas ambientais.



J48 - Algoritmo J48; SimpleCart - algoritmo SimpleCart; MLP - algoritmo de redes neurais Multilayerperceptron; Bayes - algoritmo NaiveBayes.

Figura 3. Mapa convencional (A) e os elaborados por meio dos algoritmos J48 (B), SimpleCart (C), MLP (D) e NaiveBayes (E), utilizando-se atributos selecionados pelo algoritmo CFS como variáveis preditivas ambientais.

Em termos quantitativos, os resultados das análises de concordância entre os mapas digitais produzidos pelas árvores de decisão (Quadro 5) e o mapa de solos convencional mostraram os mais elevados valores de índice Kappa, com valores de concordância acima de 80%. Esses valores foram superiores àqueles encontrados na validação cruzada (Quadro 4). Nesse aspecto, a rede neural MLP apresentou comportamento diferente, com redução dos valores do índice Kappa quando considerada a comparação com o mapa convencional. Semelhantemente, Chagas et al. (2011) obtiveram maiores exatidões globais com a utilização de amostras de validação do que quando utilizado o mapa convencional de referência. Este fato pode ser interpretado por alguns autores (Zhu, 2000; Chagas, 2006) como um indicativo de que as redes neurais tendem a apresentar menor concordância com mapas convencionais, devido provavelmente a esse algoritmo classificador gerar mapas com mais detalhes espaciais que os convencionais.

Quadro 5. Matriz de significância dos índices Kappa dos mapas de solos gerados pelo SimpleCart, J48, MLP e NaiveBayes, utilizando-se atributos do terreno como variáveis preditivas, em comparação ao Mapa convencional de referência

Método/ atributos	CARTCFS	J48CFS	MLPCFS	BayesCFS	CARTAT	J48AT	MLPAT	BayesAT
Nós (und)	7	9			7	12		
Folhas (und)	4	5			4	7		
Kappa	0,812	0,741	0,806	0,821	0,840	0,903	0,632	0,889
Vâr (x 10 ⁻⁴)	2,00	2,34	1,97	1,88	1,75	1,19	2,46	1,33
CARTCFS	0,0	-109,1*	-9,5*	13,8*	45,7*	160,5*	-269,8*	132,3*
J48CFS		0,0	100,4*	123,9	156,1*	273,8*	-156,5*	244,8*
MLPCFS			0,0	23,5*	55,7*	172,0*	-261,9*	143,4*
BayesCFS				0,0	32,2*	148,1*	-286,5*	119,5*
CARTAT					0,0	115,5*	-320,8*	87,1*
J48AT						0,0	-448,5*	-28,5*
MLPAT							0,0	416,8*
BayesAT								0,0

CARTCFS – SimpleCart e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; J48CFS – J48 e atributos selecionados pelo algoritmo de seleção CFSSubsetEval; MLPCFS - MLP e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; Bayes – NaiveBayes e atributos selecionados pelo algoritmo de seleção CFSSubsetEval e método de busca SubsetSizeForwardSelection; CARTAT - SimpleCart e atributos do terreno; J48AT – J48 e atributos do terreno; MLPAT – MLP e atributos do terreno; BayesAT – NaiveBayes e atributos do terreno. * significativo a 5% de probabilidade pelo teste Z.

A maior concordância com o mapa de solos convencional foi obtida com a utilização do algoritmo J48 (Quadro 5 e Figura 4), com atributos do terreno. Desta maneira, este algoritmo apresentou o melhor desempenho preditivo, com valor do índice Kappa considerado por Landis & Koch (1977) como excelente. O valor de kappa mais

baixo foi verificado para o mapa gerado por MLP, tendo em vista a superestimação da extensão de área com Gleissolos. Quanto à comparação dos subconjuntos de atributos, aquele composto apenas de atributos do terreno resultou em geral nas previsões com maior concordância com o mapa convencional (Figuras 4 e 5).

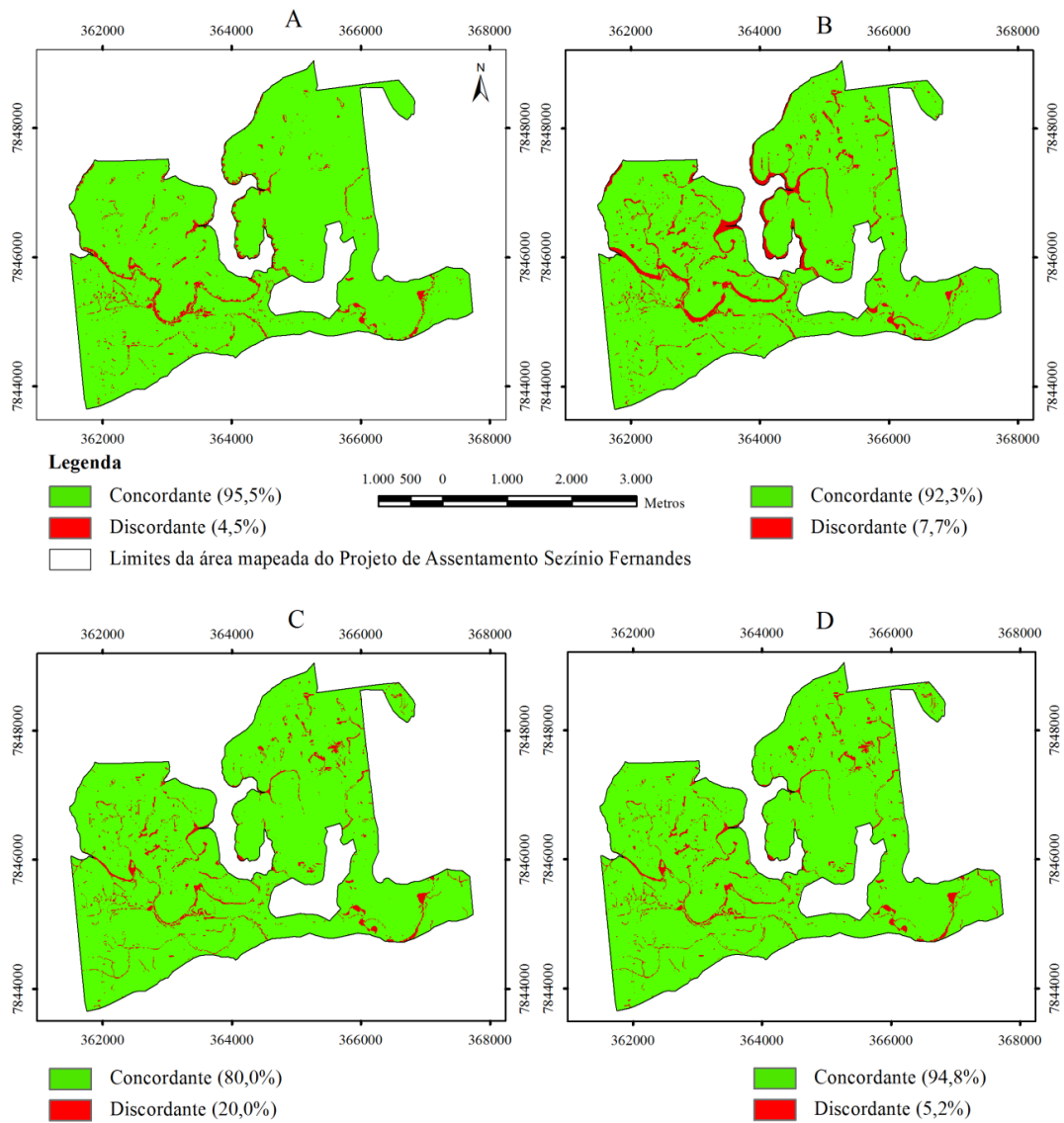


Figura 4. Concordância entre o mapa de solos convencional e os mapas elaborados com o pelos algoritmos J48 (A), SimpleCart (B), MLP (C) e NaiveBayes (D), utilizando os atributos do terreno como variáveis preditivas ambientais.

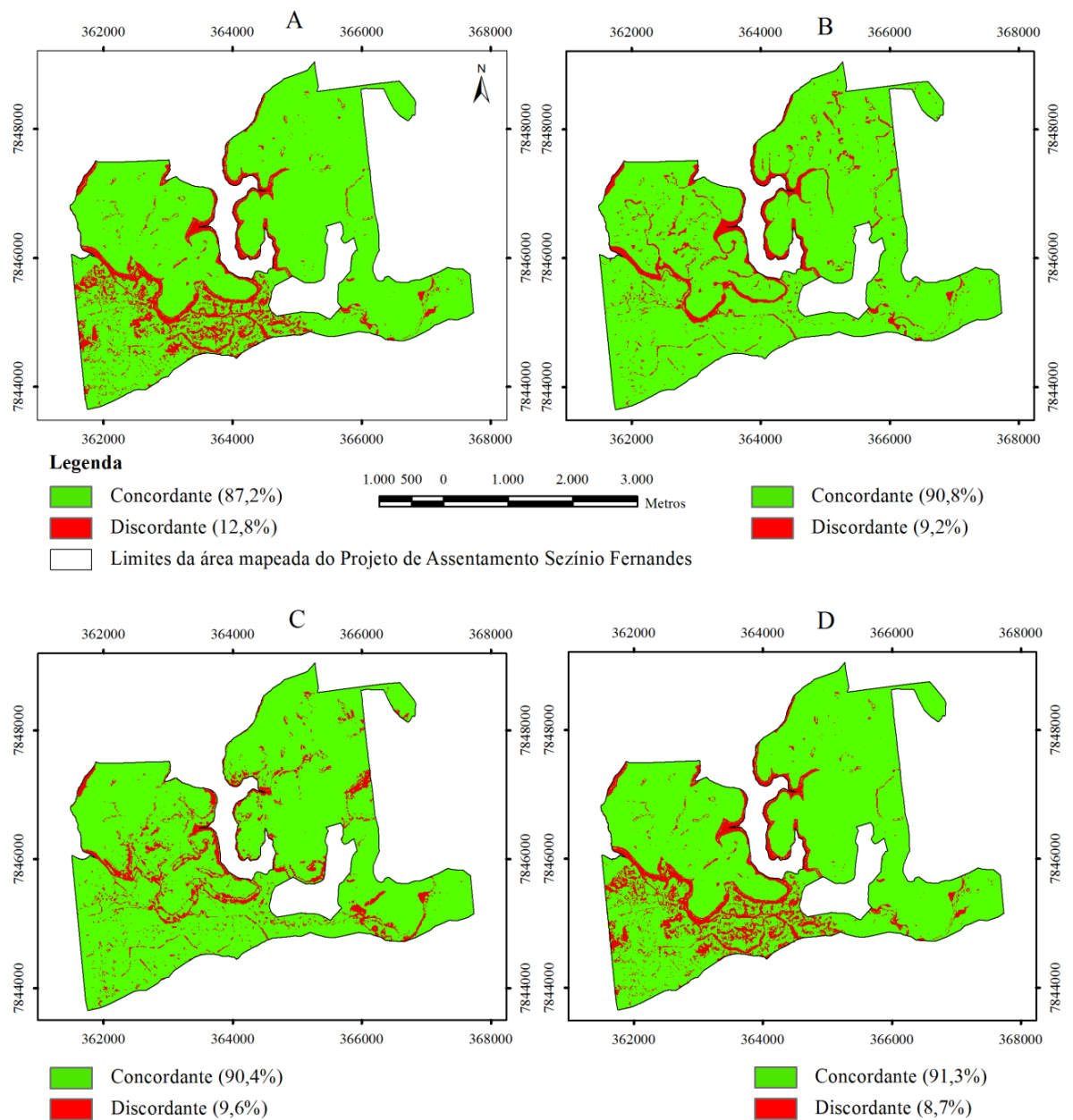


Figura 5. Concordância entre o mapa de solos convencional e os mapas elaborados com o pelos algoritmos J48 (A), SimpleCart (B), MLP (C) e NaiveBayes (D), utilizando os atributos selecionados pelo algoritmo CFS como variáveis preditivas ambientais.

O índice de exatidão do usuário indica a certeza que o usuário tem, ao analisar o mapa, que um ponto específico em certa unidade de mapeamento está classificado corretamente, enquanto o índice de exatidão do produtor representa a probabilidade de um

pixel de referência ser classificado corretamente (Congalton & Green, 2009). Ao analisar a exatidão das unidades de mapeamento individualmente, por meio dos índices de exatidão do usuário e do produtor, verificou-se que a unidade GX (Gleissolo Háptico) apresentou os menores índices (Quadro 6). Esses baixos índices resultam das confusões ocorridas na predição, com áreas de Latossolos e Cambissolos classificados como Gleissolos e vice-versa. Embora os Gleissolos e Latossolo sejam bastante distintos pedogeneticamente, as confusões se devem à ocorrência dessas classes lado a lado. Nesse caso, mostrando que as variáveis preditivas ambientais utilizadas não permitiram localizar com maior exatidão os limites entre essas classes.

Quadro 6. Matriz de confusão obtida na classificação pelo algoritmo J48, considerando-se a utilização dos atributos do terreno como variáveis ambientais preditivas, em relação ao mapa convencional de solos

		Classes no mapa convencional			Total	Exatidão do usuário
		LAd	CYbe + CYbd	GXbd		
Classes no mapa predito	LAd	103.49	1.027	559	105.078	98,45
	CYbe + CYbd	1.629	47.046	386	49.061	95,9
	GXbd	1.780	1.767	1.892	5.439	34,8
	Total	106.90	49.840	2.837		
Exatidão do produtor (%)		96,8	94,4	66,7		

5.6. CONCLUSÕES

As variáveis mais relevantes indicadas pelo algoritmo de seleção “CFSSubsetEval” envolveram tanto os atributos do terreno quanto as variáveis climáticas e dos elementos do balanço hídrico.

As variáveis climáticas e do balanço hídrico apresentaram bom potencial preditivo das classes de solo, embora devam ser evitadas em mapeamentos mais detalhados. Essas variáveis apresentaram pequenas amplitudes nos seus valores, o que torna improvável uma influência diferentemente marcante nos processos pedogenéticos de cada classe de solos da área de estudo. Nesse aspecto, no mapeamento digital de solos deve-se estar atento para que nem sempre sejam utilizadas todas as variáveis que resultem em elevada

exatidão na predição, mas aquelas que o pedólogo julgue que suas diferentes faixas de valores na área tenham real relacionamento com as diferentes classes de solo.

A rede neural MLP, NaiveBayes e as árvores de decisão SimpleCart e J48 não apresentaram diferenças na exatidão da predição, podendo-se utilizar qualquer um destes algoritmos no mapeamento digital de solos, mas preferencialmente as árvores de decisão por possibilitar mais fácil entendimento das relações entre as variáveis e as classes mapeadas.

O algoritmo J48 com a utilização de subconjunto composto somente de atributos do terreno foi o que apresentou a maior concordância com o mapa convencional de solos.

5.7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALEXANDER, S.M., LOGAN, T.B., & PAQUET, P.C. 2006. Spatio-temporal cooccurrence of cougars (*Felis concolor*), wolves (*Canis lupus*) and their prey during winter: a comparison of two analytical methods. **J. Biogeogr.** 33:2001–2012.
- BRASIL, Ministério das Minas e Energia, Secretaria Geral, Folhas SF 23/24 Rio de Janeiro/Vitória: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra / Projeto RADAMBRASIL, Rio de Janeiro, 1983, 775p.
- BUI, E.N. & MORAN, C. J. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. **Geoderma**, 111:21-44, 2003.
- BURROUGH, P. A. & MCDONELL, R. A. **Principles of Geographical Information Systems**. New York, Oxford University Press, 1998, 190p.
- CASTRO, F. S.; PEZZOPANE, J.R.M.; CECÍLIO, R.A.; PEZZOPANE, J.E.M. Uso de imagens de radar na espacialização da temperatura do ar, *Idesia*, 28:69-79, 2010.
- CECÍLIO, R.A.; SILVA, K.R.; XAVIER, A.C.; PEZZOPANE, J.R.M. Método para a espacialização dos elementos do balanço hídrico climatológico. **Pesquisa Agropecuária Brasileira**, 47:478-488, 2012.
- CHAGAS, C.S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. Viçosa, MG, Universidade Federal de Viçosa, 2006. 223p. (Tese de Doutorado)

- CONGALTON, R.G. & GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. 2.ed. Boca Raton, Taylor and Francis, 2009.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solo. **Manual de métodos e análises de solo**. Rio de Janeiro: 1997. 212p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 2. ed. Rio de Janeiro, 2006. 306p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Procedimentos normativos de levantamentos pedológicos**. Brasília, Embrapa - SPI, 1995. 101p.
- ESRI.ArcGIS 10.0.Environmental Systems Research Institute, Redlands, CA, USA. <http://www.esri.com>, 2011.
- EVANS, I.S. General geomorphometry, derivatives of altitude, and descriptive statistics. In Chorley, R.J. *Spatial Analysis in Geomorphology*. New York, Harper & Row, 1972.
- FU, P. & RICH, P.M.A geometric solar radiation model with applications in agriculture and forestry.**Computers and Electronics in Agriculture**, 37:25-35, 2002.
- GONZALEZ, J.P.; JARVIS, A.; COOK, S.E.; OBERTHÜR, T.; RINCON-ROMERO, M. BAGNELL, J.A. & M. DIAS, B. Digital Soil Mapping of Soil Properties in Honduras Using Readily Available Biophysical Datasets and Gaussian Processes. In: HARTEMINK A.E., MC BRATNEY A. & MENDONÇA-SANTOS M.L. (Org.). **Digital soil mapping with limited data**. Amsterdam, Elsevier, 2008. p.367-380.
- GIASSON, E.; SARMENTO, E.C.; WEBER, E.; FLORES, C.A.; HASENACK, H. Decision trees analysis for digital soil mapping on subtropical basaltic steepplands. **Scientia Agricola**, 68:167-174, 2011.
- GUYON, I. & ELISSEEFF, A.An introduction to variable and feature selection.**J. Mach. Learn. Res.**, 3:1157–1182, 2003.
- HALL, M. A. **Correlation-based feature selection for machine learning**. Hamilton, University of Waikato, 1998. (Tese de Doutorado).
- HU, YUMEI. **Mapping monthly precipitation in Sweden by using GIS**.Göteborg, Göteborg University, 2010. 56p. (Tese de mestrado). Disponível em:

http://www.gvc.gu.se/utbildning/avhandlingar_examensarbeten/master-and-bachelor-theses/. Acesso em 18/12/2011.

- KONONENKO, I. Estimating attributes: Analysis and extensions of relief. In: Proceedings of the seventh European Conference on Machine Learning, pages 171-182. Springer-Verlag, 1994.
- LANDIS J. & KOCH G. The measurement of observer agreement for categorical data. **Biometrics**, 33:159-174, 1977.
- MCBRATNEY, A. B., MENDONÇA-SANTOS, M. L., MINASNY, B. On digital soil mapping. **Geoderma**, 117:3-52, 2003.
- MOORE, I.D. & WILSON, J.P. Length-slope factors for the Revised Universal Soil Loss Equation: simplified method for estimation. **Journal of Soil and Water Conservation**, 47:423-428, 1992.
- MOORE, I.D.; GESSLER, G.A. & PETERSON, G.A. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, 57:443-452, 1993.
- MOORE, I.D.R.; GRAYSON, B. & LANDSON, A.R. Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. **Hydrological Processes**, 5:3-30, 1991.
- PIKE, R.J. & WILSON S.E. Elevation relief ratio, hypsometric integral, and geomorphic area altitude analysis. **Bull. Geol. Soc. Am.**, 82:1079-1084, 1971.
- RILEY, S. J., DeGLORIA, S. D. & ELLIOT, R. A terrain ruggedness index that quantifies topographic heterogeneity. **Intermountain Journal of Sciences**, 5:23-27, 1999.
- ROLIM, G.S.; SENTELHAS, P.C. & BARBIERI, V. Planilhas no ambiente EXCEL para os cálculos de balanços hídricos: normal, sequencial, de cultura e de produtividade real e potencial. **Rev. Bras. Agrometeorologia**, 6:133-137, 1998.
- RUIZ, H.A. Incremento da exatidão da análise granulométrica do solo por meio da coleta da suspensão (silte + argila). **R. Bras. Ci. Solo**, 29:297-300, 2005.
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C.; ANJOS, L.H.C. **Manual de descrição e coleta de solo no campo**. 5 ed. revisada e ampliada. Viçosa: Sociedade Brasileira de Ciência do Solo, 100 p. 2005.
- SILVEIRA, C. T.; OKA-FIORI, C.; SANTOS, L.J.C.; SIRTOLI, A.E.; SILVA, C.R.; BOTELHO, M.F. Soil prediction using artificial neural networks and topographic attributes. **Geoderma**, 195-196:165-172, 2013.
- SOUZA JUNIOR, J.G.A. & DEMATTÊ, J.A.M. Modelo digital de elevação na

- caracterização de solos desenvolvidos de basalto e material arenítico. **R. Bras. Ci. Solo**, 32:449-456, 2008.
- TEN CATEN, A.; DALMOLIN, R.S.D.; MENDONÇA-SANTOS, M.L.; GIASSON, E. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, 42:1990-1997, 2012.
- THORNTHWAITE, C.W. & MATHER, J.R. **The water balance**. Publications in Climatology. New Jersey: Drexel Institute of Technology, 104p. 1955.
- WITTEN, I.H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2nd ed. San Francisco: Morgan Kaufmann, 2005. 525p.
- ZHU, A. X. Mapping soil landscape as spatial continua: the neural network approach. **Water Resources Research**, 36:663-677, 2000.

6. CONCLUSÕES GERAIS

A utilização da interpolação de dados de precipitação pluvial e de balanço hídrico pelo método da cokrigagem forneceu superfícies que apresentaram resultados satisfatórios como variáveis ambientais preditivas ambientais de classes de solos. Isto demonstra o potencial de utilização desse método de interpolação na espacialização de outras variáveis de interesse no mapeamento digital de solos.

As variáveis climáticas e do balanço hídrico mostraram capacidade preditiva tão satisfatória quanto a outras variáveis geralmente utilizadas no mapeamento digital de solos, como os atributos do terreno, geologia e geomorfologia. Os resultados mostram ser promissora a utilização de variáveis climáticas no mapeamento digital de solos. No entanto, não se sabe até que nível de detalhes ou escala de mapeamento estas devem ser utilizadas. Há a necessidade ainda de avaliações em diferentes escalas de mapeamento, fontes de dados climáticos e condições ambientais diferentes em termos da relação solo e paisagem. O que ficou claro nesse estudo é que as variáveis climáticas e do balanço hídrico foram relevantes na predição, mesmo no mapeamento semidetalhado de solos. Embora, tendo em vista que por natureza são mais para estudos em escalas regionais, seja improvável uma relação plausível entre os valores dessas variáveis, especialmente considerando as suas amplitudes e as diferentes classes de solos.

Em geral, os classificadores por árvores de decisão (SimpleCart, principalmente) apresentaram desempenhos de predição de classes de solos semelhantes aos algoritmos de NaiveBayes e rede neural multilayer perceptron, com a vantagem da necessidade de utilização de menor quantidade de variáveis preditivas e de se apresentarem como geradores de modelos mais simples.

7. APÊNDICES

Perfil n° 1

Data: 11/05/2012

Classificação: NEOSSOLO FLÚVICO Tb Eutrófico gleissólico

Localização: Rodovia ES-289, de Muqui para Afílio Vivacqua, ES. Coordenadas UTM: 24, 260.563 E e 7.685.729 N (ponto 347 GPS).

Situação: descrito e coletado em barranco do leito maior do rio Muqui do Norte, com 1% a 2% de declive, sob pastagem.

Altitude: 211 metros.

Litologia: sedimentos Quaternários

Material de origem: sedimentos aluviais arenosos

Pedregosidade: não pedregosa

Rochosidade: não rochosa

Relevo regional: ondulado a forte ondulado e montanhoso

Relevo local e Declividade: plano (<3%)

Vegetação primária: Mata Atlântica

Drenagem: bem drenado

Uso atual: pastagem

Erosão: não aparente

Descrito e coletado por: João Luiz Lani e Alexson de Mello Cunha

DESCRIÇÃO MORFOLÓGICA

- A1 0 – 8 cm; bruno (7,5YR 4/4, úmido) e bruno claro (7,5YR 6/4, seco); franco-arenosa; pequena, laminar e fraca; macio, friável, ligeiramente plástico e ligeiramente pegajoso; transição plana e clara.
- C1 8 – 12 cm; bruno (7,5YR 4/4 úmido), bruno claro (7,5YR 6/4 seco); areia; grãos simples; solto, muito friável, não plástico e não pegajoso; transição ondulada e clara.
- C2 12 – 17 cm; bruno-escuro (7,5YR 3/4 úmido) e amarelo-avermelhado (7,5YR 6/4 seco); areia-franca; grãos simples; macio, muito friável, não plástico e não pegajoso; transição ondulada e clara.

- C3 17 – 29 cm; bruno-amarelado-escuro (10YR 4/4 úmido) e bruno-oliváceo-claro(10YR 5/4 seco); areia;grãos simples;solto, muito friável, não plástico e não pegajoso; transição ondulada e clara.
- C4 29 – 39 cm; bruno (7,5YR 4/4 úmido) ebruno (7,5YR 5/4 seco); areia;grãos simples;solto, muito friável, não plástico e não pegajoso; transição ondulada e clara.
- C5 39 – 56 cm; bruno (7,5YR 4/4 úmido) e amarelo-avermelhado (7,5YR 6/4 seco); franco-arenosa;médio/grande laminar/blocos subangulares; macio, firme, ligeiramente plástico e ligeiramente pegajoso; transição ondulada e clara.
- C6 56 – 73 cm; bruno (7,5YR 4/4 úmido) e amarelo-avermelhado (7,5YR 6/4 seco); franco-arenosa; grande/muito grande com blocos subangulares; solto, muito friável, ligeiramente plástico e ligeiramente pegajoso; transição ondulada e clara.
- C7 73 – 85 cm; bruno (7,5YR 4/4 úmido) e bruno (7,5YR 5/4 seco); areia; grãos simples; macio, muito friável, não plástico e não pegajoso; transição ondulada e clara.
- Cg1 85 - 100 cm+;bruno (7,5YR 4/4 úmido) e bruno (7,5YR 5/4 seco); mosqueado comum médio distinto, bruno-oliváceo-claro (2,5Y 5/3) e bruno-amarelado claro(2,5Y 6/3),areia;grãos simples; macio, firme, não plástico e não pegajoso; transição ondulada e clara.

Raízes: Comuns e finas no horizonte A1; poucas e finas nos horizontes C1, C2 e C3; raras e finas nos horizontes C4, C5, C6 e C7.

Fatores biológicos: presença de formigas pequenas em todo o perfil e presença de minhocas no horizonte C7.

OBSERVAÇÕES: mosqueado irregular com pontuações de Fe e Mn no horizonte Cg. Alta umidade até o horizonte C7, sobrejacente ao horizonte Cg. Presença de micas nos horizontes. Materiais de resíduos domésticos. Vale em “V”. Aparentemente rio sofreu retificação.

Perfil n° 2

Data: 24/05/2012

Classificação: GLEISSOLO HÁPLICO Tb Distrófico

Localização: Rodovia ES-490, Safra a Marataízes, ES. Usinas Paineiras, próximo a calha do rio Muqui do Norte. Coordenadas UTM: 24, 299.254 E e 7.676.376 N.

566 – ponto GPS

Situação: Descrito e coletado em trincheira aberta em área de várzea.

Altitude: 6 metros.

Litologia: sedimentos fluviais / Quaternário

Material de origem: sedimentos Aluviais argilo-arenosos

Pedregosidade: não pedregosa

Rochosidade: não rochosa

Relevo regional: plano a suave ondulado

Relevo local e Declividade: plano (< 3%)

Vegetação primária: campo de várzea

Drenagem: Imperfeitamente drenado

Uso atual: capim / pastagem Brachiária

Erosão: Não aparente

Descrito e coletado por: Alexson de Mello Cunha e Diego Lang Burak

DESCRIÇÃO MORFOLÓGICA

- A 0 - 13 cm; cinzento-brunado-claro (10YR 6/2, úmido) e acinzentado-muito-escuro (10YR3/2, seco); franco-argilo-arenosa; fraca pequenagranular; macio, muito friável, plástico e pegajoso; transição plana e gradual.
- ACg 13 - 29 cm; cinzento-brunado-claro (10YR 6/2 úmido), bruno-acinzentado (10YR 5/2); franco-argilo-arenosa; maciça, média a grande laminar; ligeiramente duro, muito friável, plástico e pegajoso; transição ondulada e abrupta.
- Cg1 29 - 37 cm; amarelo-acinzentado-claro (2,5Y 7/3 úmido) e amarelado-claro (2,5Y 8/3 seco); areia; grãos simples; solto e solto, não plástico, não pegajoso; transição ondulada e gradual.

Cg2 37 cm+; cinzento brunado-claro (2,5Y 6/2 úmido) e cinzento claro (2,5Y 7/2 seco); areia; grãos simples; solto e solto, não plástico, não pegajoso.

Raízes: Abundantes médias e grossas no horizonte A; poucas finas e médias no horizonte ACg; raras e muito finas nos horizontes Cg1 e Cg2.

Fatores biológicos: Minhocas aparentemente apenas no horizonte A.

OBSERVAÇÕES: Mosqueado Cg1 variando de 5Y 4/4 a 5YR 4/6.

Perfil n° 3

Data: 11/05/2012

Classificação atualizada: CHERNOSSOLO ARGILÚVICO Órtico abrupto

Fonte: Levantamento de reconhecimento dos solos do Estado do Espírito Santo, Perfil n° 62 – ES. SNLCS/EMBRAPA.

Localização: Município de Castelo, a 5 km de Castelo estrada para Venda Nova, lado direito.

Situação e declive: Corte de estrada, no terço anterior de uma elevação com 63% de declive.

Altitude: 120 metros

Litologia e formação geológica: Anfibólito. Pré-Cambriano Indiviso.

Material de origem: Anfibólito.

Relevo local: Montanhoso.

Relevo regional: Montanhoso, com elevações de topos angulosos, vertentes planas de centenas de metros e vales em “V” fechado.

Vegetação local: Capim-jaraguá.

Vegetação regional: Floresta subcaducifólia.

Drenagem: Bem drenado.

Uso atual: Pastagem de capim-jaraguá e culturas de milho e café.

Erosão: Laminar forte, em sulcos superfícies repetidos ocasionalmente e em voçorocas.

A 0 - 20 cm; vermelho escuro acinzentado (2,5YR 3/2, úmido), bruno avermelhado escuro (5YR 3/3, úmido amassado), bruno avermelhado escuro (5YR 4/2, seco) e bruno avermelhado (5YR 4/3, seco triturado); franco arenoso; moderada grande

granular; muitos poros pequenos; ligeiramente duro, friável, plástico e pegajoso; transição gradual e plana.

BA 20 - 45 cm; bruno avermelhado escuro (2,5YR 3/4 úmido); franco argilo arenoso; fraca média blocos subangulares; poros comuns pequenos; cerosidade pouca e moderada; ligeiramente duro, friável, plástico e pegajoso; transição clara e plana.

Bt1 40 - 90 cm; vermelho escuro (2,5YR 3/6 úmido); argila; forte grande blocos angulares; poros comuns pequenos; cerosidade comum e moderada; duro, firme, muito plástico e muito pegajoso, transição gradual e plana.

Bt2 90 - 126 cm, vermelho escuro (3,5YR 3/6, úmido); franco argiloso arenoso; fraca média blocos subangulares; poros comuns pequenos; duro, firme, plástico e pegajoso; transição gradual e plana.

C 126 - 195 cm+, bruno avermelhado escuro (4YR 3/4, úmido); franco argiloso; maciça; poucos poros pequenos; duro, firme, plástico e pegajoso.

Raízes - A₁, muitas fasciculares com diâmetros em torno de 1 mm.

BA, comuns fasciculares com diâmetro em torno de 1 mm.

Bt1, poucas fasciculares com diâmetros em torno de 1 mm.

Bt2, raras.

Observações: linha de pedras (rochas semi-intemperizada) entre os horizontes BAe Bt1 e ainda nos horizontes Bt2e C.

Perfil n° 4

Data: 12/06/1972

Classificação atualizada: ARGISSOLO AMARELO Distrocoesó típico.

Fonte: Levantamento de reconhecimento dos solos do Estado do Espírito Santo. Perfil n° 90-ES. SNLCS/EMBRAPA.

Localização: a 17,5 km de Maratáizes na estrada para Siri, lado esquerdo, na fazenda do Sr.Genildo. Coordenadas UTM aproximadas: 24, 306.879 E e 7.664.864 N.

Situação: trincheira em local plano com 0 a 2% de declividade.

Altitude: 30 metros.

Litologia e formação geológica: sedimentos. Formação Barreiras.

Material de origem: Sedimentos argilo-arenosos

Relevo regional: plano a suave ondulado

Vegetação primária: Floresta Estacional Semidecidual (Subperenifólia)

Drenagem: bem drenado

Uso atual: cultura de cana, abacaxi, mandioca e pastagem natural.

Erosão: não aparente

DESCRIÇÃO MORFOLÓGICA

- Ap 0 - 15 cm; bruno-escuro (10YR 3/3, úmido) e bruno-acinzentado-escuro (10YR 5/6, seco); areia-franca; fraca médiagranular; muitos poros pequenos e médios; ligeiramente duro, muito friável, não pegajoso; transição clara e plana.
- BA 15 - 24 cm; bruno-escuro (10YR 4/3 úmido), bruno (10YR 4/3); mosqueado pouco pequeno e difuso de cores amareladas; franco-arenosa; fraca, média blocos subangulares com aspecto maciço poroso pouco coeso in situ; poros comuns médios; duro muito friável, ligeiramente plástico e ligeiramente pegajoso; transição clara e plana.
- Bt1 24 - 40 cm; bruno-amarelado (10YR 5/5 úmido) ebruno-amarelado (10YR 5/4seco); mosqueado pouco pequeno e difuso, bruno-amarelado (10YR 5/8); franco-argilo-arenoso; fraca média blocos subangulares; poros comuns pequenos e médios; duro friável, plástico e ligeiramente pegajoso; transição gradual e plana.
- Bt2 40 - 88 cm; bruno-amarelado (10YR 5/6 úmido) e bruno-amarelado (10YR 5/5 seco); mosqueado pouco pequeno e difuso, bruno forte (7,5YR 5/6); argila-arenosa; fraca média blocos subangulares; poros comuns médios; muito duro, friável, plástico e pegajoso; transição difusa e plana.
- Bt3 88 - 190 cm; bruno-forte (8,5YR 5/6 úmido) e amarelo-brunado (10YR 6/6 seco); argila-arenosa; fraca média blocos subangulares; muitos poros pequenos; ligeiramente duro a duro, friável, plástico e pegajoso.

Raízes: comuns fasciculares de diâmetros em torno de 1 mm e poucas secundárias com diâmetro de entre 2 e 5 mm no Ap, poucas fasciculares com diâmetro em torno de 1 mm no AB e raras fasciculares com diâmetro de 1 mm no Bt1.

OBSERVAÇÕES: o terreno tem sido cultivado com culturas de mandioca e abacaxi e nunca foi adubado; aparecem algumas cerosidades nos horizontes BA e Bt1; constatou-se bastante atividade biológica no perfil nos horizontes Ap, BA, Bt1 e Bt2; na região aparecem muitos cupinzeiros que chegam a prejudicar um pouco a mecanização do solo.

Perfil n° 5

Data: 06/03/1970

Classificação atualizada: CAMBISSOLO HÁPLICO Tb Distrófico léptico.

Fonte: Levantamento de reconhecimento dos solos do Estado do Espírito Santo. Perfil n° 72-ES. SNLCS/EMBRAPA.

Localização: a 13 km de Alfredo Chaves na estrada para Matilde, lado direito. Município de Alfredo Chaves, ES. Coordenadas UTM aproximadas: 24, 319.337E e 7.722.198 N.

Situação: corte de estrada no terço inferior da elevação com 90% de declividade.

Altitude: 490 metros.

Litologia e formação geológica: biotita-gnaiss. Pré-Cambriano.

Material de origem: produtos da intemperização da rocha supracitada.

Relevo regional: montanhoso a escarpado, topos angulosos e vertentes planas.

Vegetação primária: Floresta Subperenifólia.

Drenagem: acentuadamente drenado

Uso atual: nenhum.

Erosão: laminar moderada.

DESCRIÇÃO MORFOLÓGICA

A 0 - 15 cm; bruno-acinzentado-muito-escuro (10YR 3/2); franco-arenosa; fraca pequena granular; muitos poros pequenos; muito friável, ligeiramente plástico e ligeiramente pegajoso; transição clara e plana.

Bi 15 - 70 cm; bruno-amarelado (9YR 5/8 úmido); franco-argilo-arenoso; fraca, pequena e média granular; poros comuns muito pequenos e pequenos; friável, ligeiramente pegajoso; transição abrupta ondulada.

R 70 cm +; biotita-gnaïsse.

Raízes: comuns fasciculares e secundárias com diâmetros em torno de 2 mm no A, poucas terciárias com diâmetro em torno de 3 mm e raras de diâmetro em torno de 20 mm no Bi.

OBSERVAÇÕES: o perfil está crivado de rochas e o solo ocupa os espaços entre as rochas.

Perfil n° 6

Data: 15/08/1966

Classificação atualizada: ARGISSOLO VERMELHO Eutrófico típico.

Fonte: Levantamento de reconhecimento dos solos do Estado do Espírito Santo. Perfil n° 03-ES. SNLCS/EMBRAPA.

Localização: Município de Cachoeiro de Itapemirim, a 10 km de Cachoeiro de Itapemirim na estrada para Castela. Coordenadas UTM aproximadas: 24, 275.785E e 7.702.214 N.

Situação: no terço superior de uma elevação com 30 metros de altitude relativa e 27% de declive.

Altitude: 80 metros.

Litologia e formação geológica: anfibólio biotita-xisto (?). Pré-Cambriano Indiviso.

Material de origem: anfibólio biotita-xisto.

Relevo regional: ondulado, com topos de morros arredondados, vertentes convexas de dezenas de metros e vales em manjedoura e em “V”.

Vegetação primária: Floresta tropical subcaducifólia.

Drenagem: bem drenado.

Uso atual: capim Jaraguá e capim-gordura.

Erosão: laminar ligeira e moderada e em sulcos superficiais reptidos ocasionalmente.

DESCRIÇÃO MORFOLÓGICA

- A 0 - 20 cm; bruno-avermelhado-escuro (2,5YR 3/4, úmido), bruno avermelhado (2,5YR 4/4, seco); franco-argilo-arenosa; moderadamédia granular e fraca pequena blocos subangulares; poros comuns pequenos e médios; ligeiramente duro, friável, plástico e pegajoso; transição clara e plana.
- Bt1 20 - 50 cm; vermelho-escuro (10R 3/6 úmido), vermelho (10R 4/6 úmido); argila fraca pequena blocos subangulares com aspecto maciço poroso “in situ”; poros comuns pequenos e pequenos; duro, friável, muito plástico e muito pegajoso; transição gradual e plana.
- Bt2 50 - 70 cm; vermelho-escuro (10R 3/6 úmido), vermelho (10R 4/6 seco); muito argilosa; fraca, pequena e médiablocos subangulares; poros comuns muito pequenos; cerosidade pouca e fraca; duro, friável, muito plástico e muito pegajoso; transição difusa e plana.
- Bt3 70 - 150 cm; vermelho-escuro (10R 3/6 úmido), vermelho (10R 4/8 seco);muito argilosa; moderada, pequena e média blocos subangulares e angulares; poros comuns muito pequenos; cerosidade comum e moderada; duro, friável, muito plástico e muito pegajoso; transição gradual e ondulada.
- Bt4 150 - 200 cm; vermelho (2,5YR 3/6 úmido), vermelho (2,5YR 4/8 seco);argila; moderada, pequena e média blocos subangulares; poros comuns muito pequenos; cerosidade comum e fraca; duro, friável, plástico e pegajoso; transição clara e ondulada.
- C 200 - 280 cm+; vermelho-amarelado (5YR 4/6 úmido), amarelo-avermelhado (5YR 6/ seco); franco-argilo-arenoso; estrutura com aspecto maciço poroso moderadamente coeso “in situ”, que se desfaz em agregados muito pequenos; muitos poros pequenos e pequenos; duro, muito friável, não plástico e não pegajoso.

Raízes: A e Bt1, muitas fasciculares verticais.

Bt2, comuns fasciculares verticais.

Bt3, poucas fasciculares verticais.

Bt4, raras e fasciculares.

Calhaus e matacões – Observa-se na região ocupada de 5 a 10% da área. No perfil observam-se calhaus e matacões que são restos de material de origem que ainda não estão decompostos, pois se trata de materiais resistentes à intemperização, como veios de quartzo, que são comuns nesta rocha.

OBSERVAÇÕES: No horizonte C, observa-se uma erosão moderada a forte formando pequenas voçorocas ou sulcos. Observa-se algumas vezes em alguns torrões um aumento de volume quando umedecido.

Perfil n° 7

Classificação: LATOSSOLO VERMELHO-AMARELO distrófico

Localização: Assentamento Monte Alegre, Lote 28, município de Muqui, ES.

Coordenadas UTM: 24, 249.081 E e 7.681.971 N.

Coletado por: João Luiz Lani

Fonte: Ferreira Neto et al. (2009): Estudos visando o licenciamento ambiental de assentamentos rurais do Estado do Espírito Santo

DESCRIÇÃO MORFOLÓGICA

Ap 0 – 20 cm; 5Y 4/2 (úmido); argila; moderada média granular; poros comuns médios e grandes; macio, muito friável, ligeiramente plástico e pegajoso; transição clara e plana.

Bw 100 cm; 5YR 5/6; franco-argilo-arenosa; fraca media granularcom aspecto maciço poroso pouco coeso "in situ"; muitos poros, muito pequenos e poros comuns pequenos; muito friável, plástico e pegajos; transição difusa e plana.

Perfil n° 8

Classificação: ARGISSOLO VERMELHO-AMARELO Eutrófico

Localização: Assentamento Monte Alegre, Lote 51, município de Muqui, ES.

Coordenadas UTM: 24, 249288 E e7681939 N.

Coletado por: João Luiz Lani

Fonte: Ferreira Neto et al. (2009): Estudos visando o licenciamento ambiental de assentamentos rurais do Estado do Espírito Santo

DESCRIÇÃO MORFOLÓGICA

A 0 – 20 cm; 5YR 3/3 (úmido); areia-franca; fraca grande granular; muitos poros pequenos; friável; plástico e pegajoso; transição clara e plana.

Bt 100 cm; 5YR 4/6; argilo-arenosa; moderada pequena blocos subangulares; poros comuns pequenos; cerosidade comum e moderada; firme; plástico e pegajoso; transição gradual e plana.

Perfil n° 9

Classificação: LATOSSOLO VERMELHO-AMARELO Distrófico

Localização: Estrada que liga a comunidade de Santa Cruz a Santo Antônio, nomunicípio de Atílio Vivácqua, ES. (ponto 500)

Coordenadas UTM: 24, 276.695 E e7.683.858 N

Coletado por: Alexson de Mello Cunha

DESCRIÇÃO MORFOLÓGICA

A 0 - 20 cm; 7.5YR 3/4; franco-argilo-arenosa; fraca, média subangular e fraca pequena granular.

B 20 – 40 cm; 7.5YR 5/6; argilo-arenosa; moderada, média a grande subangular e fraca pequena granular.

Perfil n° 10

Classificação: ARGISSOLO VERMELHO Eutrófico

Localização: Comunidade Candura, no município de Muqui, ES. (ponto 231)

Coordenadas UTM: 7677036 2654010

Descrito e coletado por: Alexson de Mello Cunha

DESCRIÇÃO MORFOLÓGICA

A 0 – 20 cm; 5YR 4/4; franco-argilo-arenosa; moderado, médio a grande subangular e fraca pequena granular.

B 20 – 40 cm; 2.5YR 4/8; argila; moderada, média a grande subangulares e fraca pequena granular.

Perfil n° 11

Classificação: ARGISSOLO VERMELHO eutrófico

Localização: Estrada Comunidade Independência-Safra, no município de Muqui, ES. (ponto 513)

Coordenadas UTM: 7682924 279339

Coletado por: Alexson de Mello Cunha

DESCRIÇÃO MORFOLÓGICA

A 0 – 20 cm; 5YR 4/3; franco-argilo-arenosa; fraca, média a grande blocosubangulares e fraca pequena granular.

B 20 – 40 cm; 2.5YR 4/8; argila; fraca, média a grande de blocos subangulares e fraca pequena granular.

Quadro 1. Resultados das análises químicas das amostras coletadas nos horizontes A e B dos solos na bacia do rio Muqui do Norte, ES

Horiz.	CO	pH		Δ pH	Ca ²⁺	Mg ²⁺	K ⁺	SB	Al ³⁺	H+Al ³⁺	T	t	V	m	P	P-rem
		H ₂ O	KCl													
dag Kg ⁻¹		cmol _c dm ⁻³										%		- mg dm ⁻³ -		
-----Perfil 1 - NEOSSOLO FLÚVICO Tb Eutrófico gleissólico-----																
A1	0,	6,35	5,15	-	2,29	0,98	0,09	3,50	0,00	1,3	4,80	3,50	73	0,00	16,8	50,9
C1	0,	6,71	5,78	-	0,96	0,43	0,04	1,48	0,00	0,8	2,28	1,48	65	0,00	13,3	53,2
C2	0,	6,37	5,33	-	2,13	0,88	0,08	3,18	0,00	1,4	4,58	3,18	69	0,00	19,9	43,0
C3	0,	6,64	5,55	-	1,23	0,57	0,07	1,94	0,00	0,6	2,54	1,94	76	0,00	20,1	50,5
C4	0,	6,67	5,40	-	1,49	0,64	0,05	2,24	0,00	0,6	2,84	2,24	79	0,00	13,0	60,0
C5	0,	6,79	5,33	-	2,84	1,04	0,05	4,02	0,00	1,3	5,32	4,02	76	0,00	11,4	43,6
C6	0,	6,46	5,16	-	3,38	1,18	0,05	4,70	0,00	1,6	6,30	4,70	75	0,00	6,8	31,7
C7	0,	6,70	5,71	-	0,87	0,32	0,02	1,24	0,00	0,5	1,74	1,24	71	0,00	12,9	57,5
C8	0,	6,47	5,36	-	1,56	0,52	0,03	2,21	0,00	0,8	3,01	2,21	73	0,00	10,6	53,8
Cg1	0,	6,48	5,56	-	1,62	0,57	0,03	2,36	0,00	0,8	3,16	2,36	75	0,00	16,3	49,6
-----Perfil 2 – GLEISSOLO HÁPLICO Tb Distrófico-----																
A	2,	6,42	5,44	-	4,60	1,64	0,43	6,81	0,00	2,9	9,71	6,81	70	0,00	29,2	29,5
ACg	0,	5,32	4,20	-	1,58	0,55	0,04	2,34	0,49	3,7	6,04	2,83	39	17,3	10,7	52,7
Cg1	0,	6,03	4,86	-	0,52	0,21	0,02	0,86	0,00	0,8	1,66	0,86	52	0,00	8,6	48,7
Cg2	0,	5,93	4,73	-	0,49	0,18	0,02	0,84	0,00	0,8	1,64	0,84	51	0,00	7,3	38,7
-----Perfil 3 – CHERNOSSOLO ARGILÚVICO Órtico Abruptico-----																
A	1,76	6,30	5,40	-	8,20	3,20	0,16	11,60	0	2,6	14,2	11,60	82	0,00	5	-
BA	0,83	6,60	5,30	-	7,80	3,20	0,07	11,20	0	2,0	13,2	11,20	85	0,00	1	-
Bt1	0,44	6,80	5,40	-	10,90	9,50	0,04	20,70	0	1,6	22,3	20,70	93	0,00	1	-
Bt2	0,32	7,00	5,40	-	9,70	13,6	0,04	23,60	0	0,0	23,6	23,60	100	0,00	15	-
C	0,27	7,10	5,30	-	9,70	16,2	0,04	26,30	0	0,0	26,3	26,30	100	0,00	78	-

Continua...

Continuação do Quadro 1.

Horiz.	CO dag Kg ⁻¹	pH		ΔpH	Ca ²⁺	Mg ²⁺	K ⁺	SB	Al ³⁺	H+Al ³⁺	T	t	V	m	P	P-rem - mg dm ⁻³ -
		H ₂ O	KCl													
					cmol _c dm ⁻³					%						
----- Perfil 4 - ARGISSOLO AMARELO Distrocoesio Típico -----																
A	0,49	5,4	4,5	-	0,90	0,40	0,10	1,40	0,10	2,1	3,5	1,50	40	7	1	-
BA	0,62	5,2	4,3	-	0,70	0,40	0,07	1,30	0,20	2,0	3,5	1,50	37	13	1	-
Bt1	0,40	5,0	4,1	-	0,80	0,80	0,07	0,90	0,50	2,1	3,5	1,40	26	36	1	-
Bt2	0,31	4,9	4,0	-	0,70	0,70	0,02	0,70	0,90	1,8	3,4	1,60	21	56	1	-
Bt3	0,26	4,6	4,0	-	0,50	0,50	0,02	0,60	0,90	1,6	3,1	1,50	19	60	1	-
----- Perfil 5 - CAMBISSOLO HÁPLICO Tb Distrófico léptico -----																
A	2,55	5,7	4,6	-	4,10	1,50	0,94	6,60	0,00	7,0	13,6	6,60	49	0	1	-
Bi	0,66	5,1	3,7	-	0,80	0,80	0,06	0,90	1,80	4,8	5,7	2,70	16	67	1	-
----- Perfil 6 - ARGISSOLO VERMELHO Eutrófico típico -----																
A	1,62	6,5	5,9	-	6,7	4,4	0,67	11,8	0	2,4	14,2	11,8	82	0	3	-
Bt1	0,51	7,2	6,4	-	5,8	2,3	0,11	8,3	0	0	8,3	8,3	100	0	1	-
Bt2	0,43	7,4	6,6	-	6,9	2,3	0,07	9,3	0	0	9,3	9,3	100	0	1	-
Bt3	0,31	7,2	6,5	-	4,1	3,3	0,05	7,5	0	0	7,5	7,5	100	0	2	-
Bt4	0,30	6,5	5,6	-	2,3	3,4	0,07	5,8	0	1,4	7,2	5,8	81	0	1	-
C	0,30	6,1	4,5	-	2,8	3,5	0,08	6,5	0,1	1,0	7,5	6,6	87	2	1	-
----- Perfil 7- LATOSSOLO VERMELHO-Amarelo distrófico -----																
A	1,	5,0	-	-	0,1	0,2	0,79	0,38	0,7	5,12	5,50	1,08	7	65	1,4	30,6
Bw	0,	4,9	-	-	0,0	0,0	0,13	0,01	0,7	3,63	3,64	0,71	0	99	1,9	10,9
----- Perfil 8 - ARGISSOLO VERMELHO-AMARELO Eutrófico -----																
A	1,	6,2	-	-	4,0	0,8	4,96	5,3	0,0	1,32	6,62	5,3	80	0	12,3	46,7
Bt	0,	6,4	-	-	2,2	0,4	0,31	2,63	0,0	0,99	3,62	2,63	73	0	8,2	22

Continua...

Continuação do Quadro 1.

Horiz.	CO	pH		Δ pH	Ca ²⁺	Mg ²⁺	K ⁺	SB	Al ³⁺	H+Al ³⁺	T	t	V	m	P	P-rem
		H ₂ O	KCl													
dag Kg ⁻¹		cmol _c dm ⁻³										%		- mg dm ⁻³ -		
----- Perfil 9- LATOSSOLO VERMELHO-AMARELO Distrófico -----																
A	1,	5,37	4,30	-	1,18	0,60	0,18	1,98	0,10	3,7	5,68	2,08	34,9	4,8	2,0	49,4
Bw	0,	5,11	4,38	-	0,92	0,32	0,06	1,32	0,20	2,3	3,62	1,52	36,5	13,2	1,5	33,5
----- Perfil 10 - ARGISSOLO VERMELHO Eutrófico -----																
A	1,	5,87	5,07	-	2,25	0,91	0,03	3,50	0,00	2,6	6,10	3,50	57,4	0,00	2,2	39,8
Bt	0,	6,22	5,55	-	1,87	0,83	0,50	3,20	0,00	1,6	4,80	3,20	66,7	0,00	1,6	21,9
----- Perfil 11 – ARGISSOLO VERMELHO Eutrófico -----																
A	1,	7,39	6,29	-	2,96	1,51	0,29	4,80	0,00	1,1	5,90	4,80	81,4	0,00	3,0	43,9
Bt	0,	6,01	4,94	-	1,78	0,60	0,08	2,49	0,00	1,8	4,29	2,49	58,0	0,00	1,2	34,6

CO – Carbono orgânico

Quadro 2. Resultados da análise granulométrica das amostras de solos da bacia do rio Muqui do Norte, ES

Horiz.	Areia Grossa	Areia Fina	Silte	Argila	Silte/Argila
----- dag kg ⁻¹ -----					
----- Perfil 1 - NEOSSOLO FLÚVICO Tb Eutrófico gleissólico -----					
A1	16	64	9	11	0,81
C1	47	48	2	3	0,67
C2	42	42	7	9	0,78
C3	64	31	2	3	0,67
C4	59	35	3	3	1
C5	23	53	10	14	0,71
C6	19	45	16	20	0,8
C7	55	41	2	2	1
C8	46	46	4	4	1
Cg1	41	50	4	5	0,8
----- Perfil 2 - GLEISSOLO HÁPLICO Tb Distrófico -----					
A	28	26	18	28	0,64
ACg	32	29	16	23	0,70
Cg1	53	44	1	2	0,50
Cg2	45	52	0	3	0
----- Perfil 3 - CHERNOSSOLO ARGILÚVICO Órtico Abrúptico -----					
A	34	28	20	18	1,11
BA	27	30	18	25	0,72
Bt1	14	14	20	52	0,38
Bt2	27	19	24	30	0,80
C	32	20	18	30	0,60
----- Perfil 4 - ARGISSOLO AMARELO Distrocóeso Típico -----					
A	73	12	4	11	0,36
BA	62	15	7	16	0,44
Bt1	58	14	5	23	0,22
Bt2	46	13	4	37	0,11
Bt3	47	13	4	36	0,11
----- Perfil 5 - CAMBISSOLO HÁPLICO Tb Distrófico léptico -----					
A	45	12	21	18	1,17
Bi	44	12	26	24	1,08
----- Perfil 6 - ARGISSOLO VERMELHO Eutrófico típico -----					
A	44	12	10	34	0,29
Bt1	31	9	7	53	0,13
Bt2	23	6	4	67	0,06
Bt3	23	6	3	68	0,04
Bt4	25	6	10	59	0,17
C	38	18	21	23	0,91
----- Perfil 7 - LATOSSOLO VERMELHO-AMARELO distrófico -----					
A	31	12	17	40	0,43
Bw	58	9	8	25	0,32
----- Perfil 8 - ARGISSOLO VERMELHO-AMARELO Eutrófico -----					
A	67	16	8	9	0,89
Bt	29	17	8	46	0,17
----- Perfil 9 - LATOSSOLO VERMELHO-AMARELO Distrófico -----					
A	50	11	8	31	0,26
Bw	23	9	12	56	0,21
----- Perfil 10 - ARGISSOLO VERMELHO Eutrófico -----					
A	48	11	7	34	0,21
Bt	40	8	7	45	0,16
----- Perfil 11 - CHERNOSSOLO ARGILÚVICO -----					
A	52	8	13	27	0,48
Bt	32	7	10	51	0,20

GXbd – Gleissolo Háptico Tb Distrófico; LAd – Latossolo Amarelo distrocóeso; CYbd – Cambissolo Flúvico Tb Eutrófico; CYBd - Cambissolo Flúvico Tb Distrófico.

Quadro 3. Resultados das análises químicas das amostradas coletadas nos horizontes A e B dos solos no Assentamento Sezínio Fernandes, Linhares, ES

Amostra	Classe	Horiz.	CO	pH		Δ pH	Ca ²⁺	Mg ²⁺	K ⁺	SB	Al ³⁺	H+Al ³⁺	T	t	V	m	P	P-rem
				H ₂ O	KCl													
			dag Kg ⁻¹	----- cmol _c dm ⁻³ -----						----- % -----			- mg dm ⁻³ -					
P1	GXbd	A	3,75	5,05	4,45	0,6	2,09	1,50	0,42	4,01	0,49	6,9	10,9	4,5	37	11	9	53,0
		Cg	0,66	5,18	4,34	0,84	0,68	0,82	0,06	1,56	0,88	3,5	5,1	2,4	31	36	2	46,0
P2	GXbd	A	3,24	5,44	5,01	0,43	3,68	0,97	0,34	4,99	0,00	5,6	10,6	5,0	47	0	7	51,7
		Cg	1,18	5,70	4,99	0,71	1,70	0,42	0,23	2,35	0,00	2,6	5,0	2,4	48	0	2	51,3
P3	LAd	A	1,47	5,23	4,35	0,88	1,35	0,73	0,12	2,20	0,39	4,8	7,0	2,6	31	15	2	54,2
		Bw	0,66	4,80	4,09	0,71	0,36	0,12	0,05	0,53	1,66	4,8	5,3	2,2	10	76	1	46,8
P4	LAd	A	2,42	5,30	4,45	0,85	0,86	1,10	0,47	2,43	0,39	5,6	8,0	2,8	30	14	2	52,7
		Bw	1,03	4,82	4,09	0,73	0,33	0,23	0,14	0,70	1,27	4,8	5,5	2,0	13	65	1	40,5
P5	LAd	A	3,75	4,84	4,40	0,44	2,31	1,40	0,20	3,91	0,29	7,9	11,8	4,2	33	7	5	52,7
		Bw	1,62	4,65	4,18	0,47	0,54	0,33	0,06	0,93	1,17	6,0	6,9	2,1	13	56	2	49,2
P6	LAd	A	2,94	5,50	4,72	0,78	1,41	1,19	0,46	3,06	0,10	5,0	8,1	3,2	38	3	2	54,1
		Bw	1,47	5,06	4,34	0,72	0,54	0,33	0,13	1,00	1,27	5,3	6,3	2,3	16	56	1	44,5
P7	LAd	A	2,42	4,87	4,12	0,75	0,26	0,27	0,14	0,67	1,76	8,7	9,4	2,4	7	72	1	36,3
		Bw	1,03	4,73	4,30	0,43	0,12	0,04	0,03	0,19	1,76	5,6	5,8	1,9	3	90	1	37,7
P8	CYbe	A	3,02	7,18	6,49	0,69	4,02	2,27	1,96	8,25	0,10	1,4	9,7	8,4	86	1	156	56,1
		Bi	0,74	7,58	6,78	0,8	2,24	0,91	2,36	5,51	0,00	0,5	6,0	5,5	92	0	6	49,2
P9	CYbe	A	1,69	7,75	4,82	2,93	1,55	0,89	0,21	2,65	0,00	3,9	6,6	2,7	41	0	2	49,7
		Bi	0,59	6,04	5,12	0,92	1,04	0,55	0,06	1,65	0,00	1,6	3,3	1,6	51	0	1	42,3
P10	CYbd	A	2,06	5,56	4,54	1,02	1,27	0,93	0,71	2,91	0,59	5,0	7,9	3,5	37	17	2	49,2
		Bi	0,74	4,82	4,14	0,68	0,27	0,16	0,17	0,60	1,85	5,5	6,1	2,5	10	76	1	34,7

GXbd – GLEISSOLO HÁPLICO Tb Distrófico; LAd – LATOSSOLO AMARELO distrocóico; CYbd – CAMBISSOLO FLÚVICO Tb Eutrófico; CYBd - CAMBISSOLO FLÚVICO Tb Distrófico. CO – Carbono orgânico

Quadro 4. Resultados da análise granulométrica das amostras coletadas nos horizontes A e B dos solos no Assentamento Sezínio Fernandes, Linhares, ES

Amostra	Classe	Horiz.	Areia	Areia	Silte	Argila	Silte/Argila
			Grossa	Fina			
----- dag kg ⁻¹ -----							
P1	GXbd	A	2	9	31	58	0,53
		Cg	1	26	32	41	0,78
P2	GXbd	A	39	9	6	46	0,13
		Cg	32	7	6	55	0,11
P3	LAd	A	32	20	11	37	0,30
		Bw	20	13	10	57	0,18
P4	LAd	A	30	13	11	46	0,24
		Bw	21	11	5	63	0,08
P5	LAd	A	25	9	12	54	0,22
		Bw	27	11	8	54	0,15
P6	LAd	A	32	9	7	52	0,13
		Bw	21	7	10	62	0,16
P7	LAd	A	33	7	7	53	0,13
		Bw	19	6	9	66	0,14
P8	CYbe	A	40	8	19	33	0,58
		Bi	17	3	25	55	0,45
P9	CYbe	A	46	6	17	31	0,55
		Bi	39	7	22	32	0,69
P10	CYbd	A	44	6	14	36	0,39
		Bi	20	4	16	60	0,27

GXbd – GLEISSOLO HÁPLICO Tb Distrófico; LAd – LATOSSOLO AMARELO distrocoeso; CYbd – CAMBISSOLO FLÚVICO Tb Eutrófico; CYBd - CAMBISSOLO FLÚVICO Tb Distrófico.